ANALYSIS OF PROTEINS EXTRACTED FROM THE LEAVES OF *Piper sarmentosum*

SHIM SIANG YIAN

UNIVERSITI SAINS MALAYSIA

2010

ANALYSIS OF PROTEINS EXTRACTED FROM THE LEAVES OF *Piper sarmentosum*

by

SHIM SIANG YIAN

Thesis submitted in fulfillment of the
requirements for the degree of
Master of Sceince

October 2010

ACKNOWLEDGEMENT

Firstly, I would like to extend my gratitude to my supervisor, Assoc. Prof. Dr Gam Lay Harn for her help and guidance throughout my study. I would also like to thank my co-supervisors Prof. Dr. Zhari for his assistance and experience during my study.

I would like to extend my gratitude to the staffs of the School of Pharmaceutical Sciences and the National Poison Centre for assisting me by providing facilities, equipment and assistance to my work.

I would like to thank Mr. Shumugan for Pusat Pengajian Kajihayat for plant identification.

Finally, I would like to thank my family members and my labmates for their assistance, advice and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

2-DE : Two-dimensional gel electrophoresis

ACN : Acetonitrile

AEBSF : 4-(2-Aminoethyl) benzenesulfonyl fluoride

APS : Ammonium persulfate

BSA : Bovine serum albumin

CHAPS : 3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate

CID : Collision induced dissociation

DTT : 1,4 – Dithiothreitol

ESI : Electrospray ionization

IEF : Isoelectric focusing

IPG : Immobilized pH gradient

kDa : kilo Dalton

LC-MS/MS : Liquid chromatography tandem mass spectrometry

MS/MS : Tandem mass spectrometry

MS : Mass spectrometry

m/z : Mass to charge ratio

NaCl : Sodium chloride

$Na_2HPO_4.2H_20$: Disodium hydrogen phosphate dihydrate

$NaH_2PO_4.2H_20$: Sodium dihydrogen phosphate dihydrate

NH_4HCO_3 : Ammonium bicarbonate

PAGE : Polyacrylamide gel electrophoresis

PBS : Phosphate buffered saline

pI : Isoelectric point

RP-HPLC : Reverse-phase high performance liquid chromatography

SDS : Sodium dodecyl sulfate

TCA : Trichloroacetic acid

TEMED : N, N, N’ N’ - tetramethylethylenediamine

TLB : Thiourea lysis buffer

TRIS : Tris buffer

Tris : Tris(hydroxymethyl)aminomethane

LIST OF APPENDICES

PENYELIDIKAN ESTRAKAN PROTEIN DARIPADA DAUN *PIPER SARMENTOSUM*

ABSTRAK

Suatu kajian proteomik berdasarkan penggunaan gel telah berjaya diusahakan ke atas tisu daun *P. sarmensotum*. Suatu kaedah pengekstrakan protein yang dikenali sebagai pengesktrakan kesemua protein diikuti dengan pengendapan trikloroacetik asid (TCA)/ aseton telah digunakan dan telah dibuktikan menghasilkan ekstrak tisu daun yang berkualiti tinggi dan serasi dengan pemisahan dua-dimensi elektroforesis gel. Profil protein tisu daun *P. sarmenstosum* telah menghasilkan sejumlah 242 tompok protein dan seterusnya 283 protein telah dikenalpasti dengan menggunakan analisa LCMS/MS. Kebanyakan protein ini adalah terlibat dalam metabolisme dan protein-protein yang berkaitan dengan destinasi dan simpanan protein, tenaga dan pengangkutan, pertahanan dan tekanan. Dalam analisis komparatif (perbandingan) proteomic ke atas daun muda dan dewasa yang diperolehi dari tujuh tempat yang berbeza, 84 tompok protein telah dikesan dalam semua duan ekstrak samada daripada daun muda, daun dewasa mahupun kawasan geografi. Daripada 84 tompok protein, 21 tompok protein masing-masing daripada daun muda dan dewasa telah dikenalpasti sebagai protein ekspresi berbeza di mana perbezaan antara keamatan tompok tertinggi dan tompok terendah dari tujuh tempat yang berlainan adalah sebanyak 2 kali ganda. Tambahan lagi, 4 tompok protein telah dikesan sebagai tompok unik dan 21 tompok protein dikesan sebagai tompok tak teratur. Secara keseluruhannya, pemetaan proteome tisu daun *P. sarmentosum* telah

berjaya mendedahkan komposisi protein dalam tisu daun *P. sarmentosum*, suatu tanaman herbal yang nilai ubatannya terkandung dalam daun.

ANALYSIS OF PROTEINS EXTRACTED FROM THE LEAVES OF *PIPER SARMENTOSUM*

ABSTRACT

A gel based proteomic method has been developed for the leaf tissues of *P. sarmentosum*. A suitable protein extraction method termed as total protein extraction followed by trichloroacetic acid (TCA)/acetone precipitation was applied and has been shown to produce good leaf tissue extract that was compatible with 2-DE separation. The protein profiling of leaf tissues of *P. sarmentosum* showed a total of 242 protein spots and from which 283 proteins were identified by LCMS/MS analysis. Most of these identified proteins were involved in metabolism (including photosynthesis), protein destination and storage, energy and transport, and defense and stress related protein. In the comparative proteomic analysis on young and mature leaf tissues of *P. sarmentosum* obtained from seven geographical areas, 84 protein spots were found common in all the leaf extracts of *P. sarmentosum* regardless of young or mature leaf or geographical areas. Among these 84 common protein spots, 21 protein spots respectively in the young and mature leaves of *P. sarmentosum* were detected as differentially expressed proteins with a 2-fold change in intensity between the highest intense spot and the least intense spot when comparing between the seven geographical areas. In addition, 4 protein spots were detected as unique spots and 21 protein spots were detected as irregular protein spots. The proteome of leaf tissues of *P. sarmentosum* has been successfully mapped which revealed the protein composition of the leaf tissue of *P. sarmentosum*, a traditionally known leaf-based medicinal plant.

CHAPTER 1

INTRODUCTION

1.1 Traditional Medicinal Plants

A medicinal plant or herb is defined as a plant or plant part valued for its medicinal, savory, or aromatic qualities (Kowalchik *et al.*, 1998). Medicinal plants are widely used in prevention and treatment of disease. They are also used as an ingredient in health care supplements. The usage of the medicinal plants is normally based on traditional knowledge, where most of them have not been scientifically proven. There are a few drawbacks in the usage of medicinal plants; failure of the treatment includes mis-usage, and its potential side effects resulted in the therapeutic usage being pressured by World Health Organization (Caniato and Puricelli, 2003). However, 70-80% of world population still rely on traditional herbal medicine as a primary source of disease treatment (Caniato and Puricelli, 2003; Hamilton, 2004).

The demand of herbal medicine is still growing continually due to the exponential increase in human population, less side effects and the low cost of traditional medicine (Hamilton, 2004; Patwardhan *et al*., 2005). This phenomenon is noted in the significant increase of the market for Ayuvedic and Chinese medicines (Patwardhan *et al.*, 2005). Today, more than 35, 000 herbal plant species in the world are utilized appropriately in various parts of the world (Philip *et al.*, 2009).

The systematic studies on medicinal plant are significantly increased recently, where medicines derived from plants for the treatment of different diseases such as cancer, microbial, fungal and viral infection has been developed (Hamilton, 2004; Philip

et al., 2009). Moreover, the need of traditional medicines as an alternative therapy for drug resistance strain of microbial and fungal has been documented (Najib Nik A. Rahman *et al.*, 1999; Patwardhan *et al.*, 2005). Many pharmaceutical companies have started to focus on natural product chemistry for discovery and development of natural product drug. As a result, more than 1500 medicine plants were developed and processed into various forms of botanical products such as dietary supplements and herbal teas (Patwardhan *et al.*, 2005).

1.2 *Piper*

Genus *Piper*, under the family Piperaceae consists of more than 700 species habitually distributed in tropical and subtropical regions of the world (Rukachaisirikul *et al.*, 2004). The *Piper* species include *Piper nigrum*, *Piper longum* L., *Piper ribesoides*, *Piper amalago*, *Piper brachystachyum*, *Piper sarmentosum* and more (Parmar *et al.*, 1996; Scott *et al.*, 2007). The genus *Piper* shares similar characteristics. They are erect, creeping and small terrestrial tree chiefly possessing flower and fruit. The leaves, fruits, stems and roots of the *Piper* species are commercially, economically and medicinally valuable in the worldwide market (Scott *et al.*, 2007).

The value of the fruit of *P. nigrum* is economically high, where products of white and black pepper are used as spices in many tropical countries (Parmar *et al.*, 1996). In the regions of Latin America, West Indies and India, the genus *Piper* plant is well known for their medicinal properties. *Piper hispidum* and *Piper aduncum* were used by Jamaican to alleviate stomach aches. In the Ayuvedic medicine, *Piper sylvaticum* is an excellent antidote for snakebite. Extensive phytochemical studies have revealed the presence of various class of physiologically active compound such as lignans, steroids,

alkaloids, amides, pyrones, flavones, flavanones, in *Piper* species (Rukachaisirikul *et al.*, 2004; Scott *et al.*, 2008). These physiologically active compounds were proven to be important in plant chemical defense (de Morais *et al.*, 2007). A number of species of Piper contain piperamides, potential botanical insecticides that is used as pest controls to prevent the damage of stored crops (Scott *et al.*, 2008).

1.3 *Piper sarmentosum*

Piper sarmentosum, locally known as kadok, is mainly found in Northeast India, South China, Indonesia, Thailand, and Malaysia (Rukachaisirikul *et al.*, 2004). It is a very popular and widely used herb in Malaysia as food flavouring agent and traditional medicines (Nirwana *et al.*, 2009). *P. sarmentosum* is used traditionally for treatment of fever, coughs, flu, diabetes, hypertension rheumatism and joint aches. The plant appears as glabrous, creeping, terrestrial herbs with about 20 cm to 50 cm tall (Rukachaisirikul *et al.*, 2004). Its flowers are either bisexual or unisexual. Its fruits are small and dry with several rounded bulges. *P. sarmentosum* has a characteristic of pungent odour.

Studies have shown that *P. sarmentosum* contains antioxidant property (Subramaniam *et al.*, 2003). A naturally occurring antioxidant, superoxide scavenger, Naringenin was isolated from the methanolic leave extracts of *P. sarmentosum*. It is believed that frequently consuming the fruits or leaves of *P. sarmentosum* could scavenge access free-radicals and superoxide anion (O_2^-) to prevent certain conditions such as DNA damage, cardiovascular disease, stroke, hypertension and cancer (Subramaniam *et al.*, 2003). The positive effect of *P. sarmentosum* in treating caecal amoebiasis in mice has also been reported (Sawangjaroen *et al.*, 2004). Two amides isolated from *P. sarmentosum*, namely sarmentine and 1-piperettyl pyrrolidine possess

antituberculosis and antiplasmodial activities (Rukachaisirikul *et al.*, 2004). Pellitorine, guineensine, brachyamide, sarmentosine and 1-(3, 4-methylenedioxyphenyl)-1E-tetradecene extracted from the fruits of *P. sarmentosum* were reported to contain antituberculosis activities (Rukachaisirikul *et al.*, 2004). According to Peungvicha *et al.* (1998), the water extract of *P. sarmentosum* has been tested to display hypoglycemic effect on the streptozaotocin-diabetic rats.

1.4 Protein, polypeptide, peptide and amino acid

Protein is a polymeric compound composed of a set of monomers called amino acid (Matthew *et al.*, 2000). An amino acid contains a central carbon atom attached to an amino group, a carboxylate group, a hydrogen atom and a side chain R. There are 20 amino acids found in human biological system, namely alanine (A), arginine(R), asparagine(N), aspartic acid(D), cysteine(C), glutamic acid(E), glutamine(Q), glycine(G), histidine(H), leucine(L), isoleucine(I), lysine(K), methionine(M), proline(P), phenylalanine(F), serine(S), threonine(T), tryptophan(W), tyrosine(Y), and valine(V) (Matthew *et al.*, 2000; McKee, 2008).

The structure of amino acid enables it to behave as acid and base simultaneously by donating a proton from its carboxylate group and accepting a proton to its amino group at pH of 7. This phenomenon is called amphoteric. A condition with the presence of both positive and negative charges on an amino acid is termed zwitterionic form. The classes of amino acid are determined according to the solubility of the amino acid (McKee and McKee, 2008).

Based on the amino acid solubility, the groups of amino acids were classified into four groups: nonpolar, polar, acidic and basic. The side chain of nonpolar amino

acid mainly constitute of either aliphatic or aromatic hydrocarbon R group. The weakly interaction of nonpolar amino acid with water is required to maintain three dimensional structure of protein. Glycine, alanine, valine, leucine, isoleucine, proline, phenyalanine, tryptophan, cystein and methionine are categorized as nonpolar amino acids. Polar amino acid is referred as water loving amino acid. The hydroxyl group within the amino acid enables the interaction between amino acid and water to take place through the linkage of hydrogen bond. Serine, threonine, tyrosine, asparagine, and glutamine are grouped as polar amino acids. Acidic amino acids carry negative charge on its side chain at pH7 (Matthew *et al.*, 2000). There are only two amino acids in acidic form: aspartic acid and glutamic acid. In contrary, basic amino acids carry positive charge on its side chain at pH 7. The group of basic amino acid includes histidine, lysine and arginine.

Amino acids linked together covalently to form peptide by removing a water molecule between carboxyl group from one amino acid and amino group from other amino acid (Matthew *et al.*, 2000). The bond formed is termed as peptide bond. Amino acid residues combined into a long chain is called polypeptide. Protein has one or more polypeptide fold into three-dimensional structure to perform biological function. Post-translation modifications, such as phosphorylation, glycosylation are common in protein. These modifications play an important role in explicating the biological functions of proteins (Seo and Lee, 2003). Proteins that are involved in a variety of biological functions in living organisms, including catalyst, provide structural support, help in transportation and movement, regulate stress response and cellular mechanism pathway, storage of essential nutrient, and have protective function against mechanical and chemical injury.

Proteins have four levels of structures: primary, secondary, tertiary and quaternary structure (Branden and Tooze, 1999). Protein primary structure is described as a polypeptide chain with defined amino acid sequence. Protein secondary structure is the local arrangement of the backbone of a linear polypeptide to form a repeating structure such as α-helix and β-sheet structure. Extensive folding structure in a linear polypeptide to form three-dimensional shape is referred as protein tertiary structure. A protein quaternary structure is formed by two or more folded polypeptide chain to achieve highest complexity. Depending on the types of amino acid that made up the protein, protein will carry certain chemical property resulted from the synergistic interaction of its amino acids component.

1.5 Proteomics

Proteome is depicted as entire protein complement of a particular biological system (Cagney *et al.*, 2003). Consequently, proteomics is related with systematic analysis of protein population synthesized in cells (Boguski and Mclntosh, 2003; Cagney *et al.*, 2003). The completion of genome sequencing projects had urged the rapid progress of proteomics to evaluate the genome expression after mRNA editing, alternative splicing and post-translational modification. The emergence of proteomics is greatly boosted by the emerging of new technologies such as the sensitive and high throughput of mass spectrometry and high resolution of two-dimensional gel electrophoresis (2-DE), which are the standard procedure for quantitative proteome analysis (Boguski and Mclntosh, 2003).

The proteome is a challenging, complex and dynamic system influenced by diverse physiological and pathological factor (Gromov and Celis, 2000). The transcripted and translated genome varies in different type of cells. Usually only low quantity of genome is transcripted and translated within a cell. Therefore, after undergoing mRNA editing, alternative splicing, and post-translational modification, the proteome variability is greatly different from cell to cell, and eventually species to species (Gromov and Celis, 2000). The proteomics reveal the expression of gene products, characterize protein composition, functions, activities, interactions and distribution at different cellular stage (Gromov and Celis, 2000; Ong and Mann, 2005). However, about 80-90% of total protein is monopolized by housekeeping protein with the different expression patterns in majority of cells (Gromov and Celis, 2000).

1.6 Plant Proteomics

Plant proteomics which started in early 1980s, has progressed rapidly over the few years with the completion of first plant genome sequence of *Arabidopsis thaliana* (thale cress) in year 2000 (Canovas *et al.*, 2004). *Arabidopsis* is a premier dicot model species composed of five chromosomes with a total size of approximately 135 Megabase (MB) (Rossignol *et al.*, 2006). Small genome of Arabidopsis with a relative short life cycle has made the species a popular tool in molecular biology studies such as flower development. Plant proteomics is still in the beginning stage. Proteomic study is mainly depending on availability of database. The slow growth in plant proteomics is due to the scarcity of plant protein database (Nam *et al.*, 2003; Katam *et al.*, 2009).

Most proteomic studies are focus on *Arabidopsis* and *Oryza sativa* (rice). This is mainly due to the completion of the genome sequence of these plant species, which

resulted in the availability of the protein databases in the public domain that facilitate the process of protein identification using MS data. In the past few years, the complete genome sequence of the grape, *Populus trichocarpa* (black cottonwood) and *Physcomitrella patens* have been published (Quirino et al., 2010), which enhance the progress of the plant proteomic research. Recently, the plant proteomics were further extended to plants in which the genome sequence is available in significant number in the forms of genomic DNA and Expressed Sequence Tag (EST) sequences. These include *Medicago truncatula* (legume barrel medic) along with *Solanum lycopersicum* (tomato), *Zea mays* (maize), *Solanum tubersum* (potato), sorghum and soybean, *Quercus ilex* (holm oak), *Spartium junceum* (Spanish broom), *Panax ginseng* (ginseng), *Pisum sativum* (pea), *Pinus radiate* (Radiate pine), *Pinus halepensis* (Aleppo pine), *Ananas comosus* (pineapple) *Phoenix dactylifera* (Date palm) and, *Embothrium coccineum* (Notro) (Jorrin *et al.*, 2007; Jorrin-Novo *et al.*, 2009). The scarcity database entries of other plants group is compensated for by protein identification through *de novo* MS sequencing and BLAST searches.

Since plant proteomics have generated a huge number of data, some of the researchers have made an effort on the database organization. The examples of current available database are Plant Proteome DataBase (PPDB), http://ppdb.tc.cornell.edu/ mainly for *Arabidopsis* and *Z. mays*; Soybean proteome Database, http://proteome.dc.affrc.go.jp/Soybean/; The Plant Organelles Database 2 (PODB2), http://podb.nibb.ac.jp./Organellome/; Protein Mass Spectra Extraction (ProMex), http://promex.mpimp-golm.mpg.de/home.shtml; Rice Proteome Database, http://gene64.dna.affrc.go.jp/RPD/, etc. The ProMex, is a MS/MS spectral database

derived from *Arabidopsis, Chlamydomonas reinhardtii*, *Medicago truncatula*, *Solanum lycopersicum*, *Solanum tubersum* and other plants.

In proteomic, the plant tissue parts that had been analyzed include seeds, roots, stems, leaves, pistils, xylem or phloem saps, pollens and whole seedlings (Jorrin *et al.*, 2007). Generally, the plant proteomic studies focus on protein profile of plant tissue parts in order to examine the developmental changes of plant and influence of environmental factors. The main focus of plant proteomic is on *Arabidopsis, O. sativa* and *Z. mays*. Studies on the seed germination of *Arabidopsis* by 2-DE technique have shown 1300 proteins in seeds, 74 of the presence of the proteins showed expression changes during the germination process. A protein named gibberellins (GAs) was reported to involve in the initial stage of germination (Gallardo *et al.* 2001; 2002). A study on proteome of root, leaf, and seed tissues of *O. sativa* was carried out by using 2-DE and LC-based separation methods (Koller *et al.*, 2002). The authors reported that only 7.5% of a total of 2528 proteins were expressed in all three plant tissues, showing the differential expression of specific proteins in different parts of the plant. In 2001, Porubleva *et al*. studied the maize leaf proteome where a total of 1100 proteins were detected on 2-DE gels. The protein spots were then subjected to the MS analysis and subsequently EST databases search. The outcome summarized that over 50% of proteins cannot be identified from the database.

Studies on *P. sativum* and *M. truncatula* (Schiltz *et al.*, 2004) for identification of the nitrogen mobilization from leaves to seed filling in *P. sativum* resulted in establishment of proteome reference maps of mature leaves and stems. The result showed the importance of Rubisco in nitrogen mobilization and 14-3-3 like protein as a potential regulator during the massive remobilization of nitrogen. Study of proteome

patterns was performed in leaves, stem, roots, flowers, seed pods and cell suspension cultures of model plant *M. truncatula* (Watson *et al.*, 2003). Three hundred and four proteins out of 551 proteins were identified using peptide mass fingerprinting through matrix-assisted laser desorption ionization time-of-flight mass spectrometry. When the expression levels of the identified proteins were compared to the mRNA levels through EST counting, approximately 50% of the proteins can be associated with their corresponding mRNA levels. A similar study on Cannabis sativa to compare the proteome of leaves, flower and glands was carried out (Raharjo *et al.*, 2004). In this study, the authors found that less than half of proteins expressed in flowers and glands as compared to the leaves.

Recently, the proteomic of specific protein function in plant subcellular compartments of chloroplast and mitochondria was investigated (Thiellement *et al.*, 2002; Jorrin *et al.*, 2007). By sub-fractionating into hydrophilic and hydrophobic condition for the better understanding of proteins functional roles, 81 proteins were identified from the mitochondria of *Arabidopsis* (Millar *et al.*, 2001). A series of study on the chloroplast proteins of Arabidopsis and spinach was carried out on the envelope and thylakoids (Rossignol *et al.*, 2006), lumenal (Schubert *et al.*, 2002), ribosomal (Yamaguchi and Subramanian, 2000) and stromal proteins (Ytterberg et al., 2006). In addition, Peltier *et al.* (2001) have identified a chloroplastic protease complex from *Arabidopsis*. Recently, the development of proteome on the cell walls, membranes, nucleus and vacuoles of plant cells were sequentially introduced to reveal the knowledge of protein location and function in order to understand the plant biological processes including plant development and growth, seed germination, signal transduction and the responses of plants to symbionts, biotic and abiotic stresses (Rossignol *et al.*, 2006;

Jorrin *et al.*, 2007). Since the last decade, many reports have shown the potential of plant proteins in inhibition of fungal growth. In 2000, Giudici *et al.* had successfully isolated a 16 kDa protein named SAP 16 from *Helianthus annus*. SAP 16 functions as trypsin inhibitor that restricts the ascospore germination of *Sclerotinia sclerotiorum* and the growth of mycelia. Other antifungal proteins (AFPs) reported were glucanases (Ng and Ye, 2003), chitinases (Kitajima *et al.*, 2010), thaumatin-like proteins (Wang and Ng, 2002) and several types of basic cysteine rich proteins and proteinase inhibitors (Joshi *et al.*, 1998) that were found in various plant species. Recently, the studies on the proteomic of subcellular plant has been shifted from the technique 2-DE to a non-gel based approaches which have been used in *Arabidopsis* species. In another study, Morel *et al.* (2006) used multidimensional protein identification technology (MudPIT) to study the proteome of detergent resistant membrane from plant tobacco. MudPIT is a technique introduced by Yates laboratory for protein separation and identification using two-dimensional liquid chromatography (2-D LC) coupled with a mass spectrometer (Link *et al.*, 1999; Lohrig and Wolters, 2009).

Malaysia is a country gifted with valuable types of plants that had been used traditionally for their medicinal values. Although many research have been carried out on these plants, most of them concentrated on the small molecule compounds instead of the valuable macromolecule compounds of these plants.

1.7 Sample preparation

Sample preparation is the key step to determine the success of the experiment. In plant proteomics, protein extraction is a great challenge because plant tissues normally contain low amount of protein and high quantity of protease, phenolic compounds,

secondary metabolite, lipids and other non-protein compounds (Cristina-Maria and Katja, 2006; Isaacson *et al.*, 2006; Weiss and Gorg, 2007). Thus, choosing an appropriate method is critical in producing a high quality and reproducible result. Several protein extraction methods have been developed due to the varied nature of plant tissue. Three critical aspects that need to be considered in plant protein extraction namely (1) tissue disruption, (2) removal of interfering compounds and (3) protein solubilization (Wang *et al.*, 2008):

(1) Plant cells are surrounded by a thick complex cell walls polysaccharide, cellulose containing hemicellulose and pectin which are difficult to disrupt (Shewry and Fido, 1996; Jimenez *et al.*, 2001; Isaacson *et al.*, 2006). Plant cell wall provides rigidity strength and protective structure against mechanical stress to the plant cells. As a consequence, a proper tissue disruption step is required to completely destroy the cell wall. The most widely used technique to break down the plant cell walls is by freezing the plant tissues with liquid nitrogen and then grinding the plant sample using mortal and pestle (Wittmann-Liebold, 2006; Wang *et al.*, 2006; 2008). Liquid nitrogen, a cryogenic liquid, possesses boiling point as low as -196 °C which can cause rapid freezing on the plant tissue (Chawla, 2002). The employment of liquid nitrogen with cryogenic property gives the extra advantage in prevention of protein degradation and proteolysis. Quartz sand can be added to assist the grinding process (Wang *et al.*, 2008).

(2) Plant tissues have comparatively low protein content while high contaminants of secondary metabolites accumulate inside the vacuoles of plant cells (Cristina-Maria and Katja, 2006; Isaacson *et al.*, 2006; Weiss and Gorg, 2007). A vacuole contains tonoplast that enclosed the inorganic and organic molecules such as polysaccharides, proteases,

toxic compounds, phenolic compounds, alkaloids, flavonoid pigments, tannins, lignins, etc (Taiz, 1992; Weiss and Gorg, 2007). The presence of all these interfering compounds in the vacuoles varies with the development stage of a plant and varies from species to species. Generally, these compounds were found rich in adult plant tissues than in young tissues (Wang *et al.*, 2008). The removal of secondary metabolites in extraction step is absolutely important to obtain a high quality 2-DE result. The existence of secondary metabolites such as pigments, lipid and polysaccharides will interfere the performance of 2-DE, especially the phenolics. Phenolics contain aromatic ring with one or more hydroxyl groups, they are mainly derived from the pentose phosphate, shikimate and phenylpropanoid pathways (Balasundram *et al.*, 2005). Phenolics which form a complex with the protein irreversibly can cause the 2-DE gels streaking and artifactual spots (Xu *et al.*, 2008). The application of the 10% (w/v) trichloroacetic acid (TCA)/acetone precipitation on plant extract is an effective way to remove the secondary metabolites, either before or after the protein extraction. The resulting pellet from the TCA/acetone precipitation should be white or light yellowish in color. The addition of polypyrrolidone (PVPP) in extraction buffer can precipitate the phenolic compounds through the formation of PVPP-phenolic complex and the elimination of the complex can easily be done through centrifugation (Wang *et al.*, 2008).

(3) Solubility of protein in plants varied from species to species, depending on the nature of the plant tissue. The protein solubilization normally includes the breaking down of the macromolecular interactions in order to produce individual polypeptide chains via disruption of disulfide bonds and non-covalent interactions, for example ionic bonds, hydrogen bonds and hydrophobic interactions inside the proteins and non-covalent interactions between proteins and non-proteinaceous compounds such as lipids,

carbohydrate and nucleic acids (Rabilloud, 2002). The ideal protein extraction condition is to maintain all the extracted polypeptides in the structure similar to their exact living state. Thus, artifact modifications of the polypeptides have to be prevented by the appliance of low temperature condition and protease inhibitor cocktail to the plant tissues. This step is important to inactivate all the proteases that are involved in protein modification (Lopez, 2007). Protease inhibitor and reducing agent can also be used to improve protein solubilization (Wang *et al.*, 2008).

1.8 Protein Extraction buffer

In plant proteomics, a strong extraction buffer is required to extract total protein from the tissues. The most common extraction buffer used in the plant proteomics, especially for extraction of leave proteins always contains chaotrope, denaturant, detergent and reductant (Cristina-Maria and Katja, 2006; Lopez, 2007).

Urea is a neutral chaotrope which acts as a denaturant to solubilize and unfold proteins via breaking their hydrogen bonds into a random conformation in which all ionizable amino acid side chains are exposed to buffer solution (Monribot and Boucherie, 2000; Rabilloud, 2002). Urea normally functions at concentration as high as 7 M in the presence of thiourea for further solubilization improvement, this is especially true for the hydrophobic proteins (Rabilloud, 2002). At high concentration, urea in the structure of organized channels can directly bind to linear alkyl chains molecules to form a complex called inclusion compound. These inclusion compounds are less soluble and the precipitation is often induced in the presence of nonionic or zwitterionic detergents with basic linear alkyl chains structure. Therefore, the choice of the detergents is limited to non-linear alkyl tails, for example: Tritons, Nonidet P40, 3-[(3-cholamidopropyl)

dimethylammonio]-1 propanesulfonate (CHAPS) and octyl glucoside. Urea can easily break down to form isocyanate through a process called protein carbamylation. For this reason, the use of protein extraction buffer containing urea should be completely avoided at temperature above 37°C in order to reduce the formation of artifactual spots in 2-DE gels (Rabilloud, 2002). Other precautions are to avoid using of low purity urea or adding of cyanate scavenger into a urea solution.

Nonionic or zwitterionic detergents are required to solubilize protein completely while preventing protein aggregation through hydrophobic interactions (Lopez, 2007). In several studies, ionic detergent such as sodium dodecyl sulfate (SDS) was added in protein extraction. The presence of SDS as an anionic surfactant assists the protein solubilization by disrupting non-covalent interactions and subsequently produces electrostatic repulsion in order to separate polypeptides (Monribot and Boucherie, 2000; Rabilloud, 2002). However, the utilization of SDS is incompatible with IEF due to its ionic nature (Mechin *et al.*, 2007). Therefore, SDS is always used in a diluted form in extraction solution with high concentrations of urea and nonionic or zwitterionic detergents (Rabilloud, 2002). Only low amount of SDS at a concentration lower than 0.25% or at a ratio at least 8:1 (detergent to SDS) is tolerated with the use of IEF (Righetti *et al.*, 2001).

Reducing agents are used to reduce protein by breaking its disulfide bonds during the protein extraction process (McGettrick and Worrall, 2003). Dithiothreitol (DTT) and 2-mercaptoethanol are universal reductants being used (Rabilloud, 2002). However, the usage of DTT is preferable due to formation of artifacts by 2-

mercaptoethanol. Moreover, high concentration of 2-mercaptoethanol is needed compared to DTT, which DTT only used at concentration ranging from 20 to 100 mM.

1.9 Protein Precipitation and Purification

The application of the 10% (w/v) trichloroacetic acid (TCA)/acetone precipitation can be performed either before the protein extraction or after the protein extraction (Wang *et al.*, 2008). In the former usage, plant tissue powder is directly subjected to TCA/acetone precipitation before the protein extraction. In contrary in the latter usage, the tissue powder is extracted with aqueous buffer and the extract is subjected to TCA/acetone precipitation. The main purpose for TCA/acetone precipitation is to purify and concentrate the protein before running the 2-DE. This protein purification and precipitation technique, primarily described by Damerval *et al.* (1986) is more effective than the purification using trichloroacetic acid or acetone alone. 10% TCA can effectively remove majority of secondary metabolites such as lipid, pigments and phenolic compounds under acidic and hydrophobic conditions (Wang *et al.*, 2008). Moreover, the use of TCA in acetone can inhibit the activities of proteases, phenoloxidases and peroxidases by preventing proteolytic degradation and thus, loss of high molecular weight proteins (Mechin *et al.,* 2007). The lyophilized pellet from TCA/acetone precipitation cannot be over-dried to ensure the efficiency of resolubilization.

1.10 Protein Assay

The protein assay used in this study is a colorimetric assay for protein quantitation based on the Lowry method (Lowry *et al.*, 1951). The assay employs two

different reactions. First, the protein reacts with copper ion to form a reduced copper in alkaline solutions. Second, the copper-amide bond complex from first reaction reduces Folin-Ciocalteu reagent (phosphomolybdate and phosphotungstate) (Olson and Markwell, 2007). The protein amount is spectrophotometrically detected via the formation of reduced Folin-Ciocalteu reagent, which is blue in color with the absorption in the range of 500 to 750 nm wavelength. The Lowry assay is sensitive, however, the sensitivity of Lowry assay is reduced by some interferent such as chlorophyll, detergents, carbohydrates, glycerol, Tricine, EDTA, Tris, potassium compounds, sulfhydryl compounds and disulfide compounds (Berges *et al.*, 1993; Olson and Markwell, 2007). Several modified Lowry assay as well as commercialized ones have been developed to eliminate the interferents. The commercialized protein assay has been modified to be reducing agent and detergent-compatible.

1.11 Two-Dimensional Gel Electrophoresis

Two-dimensional gel electrophoresis is a powerful technique used to separate and analyze proteins according to isoelectric point (*pI*) in the first dimension and molecular weight in the second dimension (Stochaj *et al.*, 2002; Xu *et al.*, 2008). The use of two-dimensional gel electrophoresis in proteome studies is first described by O' Farrell and J. Klose in 1975 (Stochaj *et al.*, 2002). However, the original technique is not commonly used due to handling problem and the complexity of result.

Recently, the two-dimensional gel electrophoresis has been successfully improved in the aspects of reproducibility, resolution, and separation of highly acidic and basic proteins (Weiss and Gorg, 2007). The new immobilized pH gradients were introduced by Gorg and colleagues to replace the old carrier ampholyte-generate pH

gradients. Moreover, the availability of second generation software and powerful computer assisted software facilitate the complicated protein analysis process (Gygi *et al*., 2000; Wittmann-Liebold, 2006). Two-dimensional gel electrophoresis is able to separate amount of protein as low as 1 ng of protein content per spot, with detection of more than 5000 proteins per gel by the combination use of highly sensitive protein detection method (Weiss and Gorg, 2007). The deliverable result is presented as a map that characterizes the protein profile of a biological system, which reveals the protein expression level, isoforms or post-translation modification for a target protein (Barnouin, 2004).

1.11.1 Isoelectric focusing

Isoelectric focusing is the first dimension of two-dimensional gel electrophoresis. It is carried out by using an immobilized pH gradients (IPG) strip (Stochaj *et al.*, 2002). The separation method of protein is based on their isoelectric point under an electric field (Garfin, 2003). Isoelectric point of a protein is the pH of the local environment when protein charge is zero (Rabilloud, 2002). The net charge of a protein is a total of its positive and negative charges. It is a very environment dependent value. A protein carries a negative charge when the environment pH is above its pI value, and conversely, it carries a positive net charge when the surrounding pH is below its pI value (Lopez, 2007). Under a pH gradient, a positively charged protein migrates to the cathode with decreasing charge density until its gain a zero net charge (Garfin, 2003; Barnouin, 2004). The pI value is determined by the characteristic of side chains and prosthetic groups of the protein; an acidic protein has a low pI value, while the basic protein has a high pI value. The resolution of isoelectric focusing depends on the strength of electric field and

the range of pH gradient. High voltage and narrow pH gradient will give high resolution isoelectric focusing (Stochaj *et al.*, 2002).

In 1982, Bjellqvist developed a new immobilized pH gradient for isoelectric focusing instead of carrier ampholyte-generated pH gradient. The covalent fixed pH gradient has resolved the problem of cathodic drift generated by carrier ampholyte-generated pH gradient. In the method of immobilized pH gradient, Bjellqvist had introduced the use of acrylamide buffer into polyacrylamide gel by copolymerize acrylamide with bisacrylamide monomers. This is the present standard approach in proteomics, which effectively improved the reproducibility, increase loading capacity and broader pH range separation for protein. After isoelectric focusing, disulfide bridges in protein are reduced and the sulfhydryl group is alkylated using dithiothreitol and iodoacetamide respectively, before being subjected to the second dimension, which is gel electrophoresis separation (Weiss and Gorg, 2007).

1.11.2 Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Electrophoresis is a technique of separating charged particle such as proteins, peptides and nucleic acids in an electric field (Eswara-Reddy and Jacobs, 1997). The gels used in an electrophoresis usually can be made of a variety of sources including starch, agarose, polyacrylamide and cellulose acetate.

Polyacrylamide gel is composed of nonionic polymerization of acrylamide and crosslinking agent, N, N'-methylene-bis-acrylamide, the polymer is chemically stable over a wide range of pH, temperature and ionic strength (Walker *et al.*, 2000). The concentration of acrylamide and cross-linker can be altered easily to produce different

pore size of the gel, allowing proteins to separate in a particular range of molecular weights (Stochaj *et al.*, 2002).

Sodium dodecyl sulfate polyacrylamide gel electrophoresis is a general technique used to separate a mixture of protein based on their molecular weight. The presence of sodium dodecyl sulfate in polyacrylamide gel acts as an anionic surfactant by surrounding denatured protein in an approximate ratio 1.4:1 (sodium dodecyl sulfate: protein) to form a net negative charged complex (Dunn and Bradd, 1993; Stochaj *et al.*, 2002; Lopez, 2007). In an applied electric field, the negative chargedly sodium dodecyl sulfate-protein complex migrated towards the anode based on their molecular weights; a low molecular weight protein moved faster than a high molecular weight protein (Walker *et al.*, 2000). In 2-D gel electrophoresis, a sodium dodecyl sulfate polyacrylamide gel electrophoresis is based on Laemmli discontinuous Tris-chloride/ Tris-glycine system (Stochaj *et al.*, 2002). However, no stacking gel is necessary due to the presence of pre-separated step by isoelectric focusing. Protein resolution of two-dimensional gel can be influenced by gel size. A larger gel is capable of detecting higher number of protein resulting in a complex profile analysis. The appropriate acrylamide concentration for two-dimensional gel electrophoresis ranges from 10-12.5% for separation of protein size range of 10-200 kDa (Dunn and Bradd, 1993). Lower percentages of acrylamide decrease the resolution of two-dimensional maps, while the higher percentages complicate the later analysis.

1.12 Staining

There are diverse staining methods with different sensitivities available for protein spots detection in two-dimensional gels, these staining methods include

Coomassie Brilliant Blue, silver and fluorescence staining (Wirth and Romano, 1995; Stochaj *et al.*, 2002). A selected staining method in an analysis has to fulfill several criteria including sensitive enough for low abundant protein detection, allow further proteomic analysis and compatible with mass spectrometry. However, neither one of the staining techniques is perfect. None of the techniques is able to stain all the present proteins in a gel. Thus, different protein staining patterns may be obtained from different staining techniques. Coomassie Brilliant Blue is the widely use, economical and simple organic-based stain for post-electrophoretic protein detection (Weiss *et al.*, 2009). The narrow dynamic range of Coomassie Brilliant Blue dye only enables it to detect approximately 0.5-2.0 μg of protein concentrations range (Stochaj *et al.*, 2002). Therefore, low abundance proteins below the sensitivity limit is precluded (Weiss *et al.*, 2009). Coomassive blue staining is an appropriate and reliable method for further quantitative analysis coupled with computer software programs. Moreover, the dye can be completely removed from the stained protein making it compatible with the subsequent mass spectrometry analysis (Shevchenko *et al.*, 2007).

1.12.1 Coomassie Brilliant Blue-R

Coomassie Brilliant Blue-R (reddish hue) is a non-polar, sulfated aromatic dye that complexs with the basic amino acids, arginine, lysine, histidine as well as tyrosine in order to visualize the presence of protein in a gel, with a limit of protein detection about 30-100 ng (Wirth and Romano, 1995; Stochaj *et al.*, 2002). Regressive staining approach is used to destain Coomassie Brilliant Blue-R stained gel using a similar solution as staining solution devoid of dye (Patton, 2002). The staining solution is a mixture of dye with an aqueous solution containing methanol and acetic acid. During the

destaining step, the stained gel background is gradually diminished, leaving the well labeled protein band due to the higher affinity of proteins to the dye molecules than the gel matrix.

1.12.2 Coomassie Brilliant Blue-G

Coomassie Brilliant Blue-G binds to the basic amino acids of protein. Unlike the Coomassie Brilliant Blue-R, Coomassie Brilliant Blue-G contains colloidal dye particles, as well as free dispersed dye in solution (Patton, 2002). During a staining equilibrium, the low concentration of free dye preferentially penetrate and stain the protein, while the colloidal dye particle of Coomassie Brilliant Blue-G (greenish hue) is incapable of gel matrix penetration. Thus, a clear background can be obtained in a short period of destaining step (Wirth and Romano, 1995; Weiss *et al.*, 2009). The amount of protein for detection range of Coomassie Brilliant Blue-G is approximately 8-50 ng, which is more sensitive compared to Coomassie Brilliant Blue-R (Choi and Yoo, 2002; Stochaj *et al.*, 2002).

1.13 In-gel Digestion

In-gel digestion is a pre-step for mass spectrometric analysis. It is purposed to extract, digest and degrade proteins into peptide form through an enzymatic cleavage activity (Granvogl *et al.*, 2007). The in-gel digestion method was initially established by Rosenfeld *et al*. (1992) and several alterations have sequentially been proposed to assure the high peptide yield and quality of mass spectrometric analysis (Granvogl *et al.*, 2007). Enzymatic cleavage with a specific protease results in a set of characteristic peptides

with diverse molecular masses. The molecular masses of peptides can be used to give the identity of protein upon mass spectrometry.

Ammonium bicarbonate and Tris/hydrochloric acid are commonly used as inorganic buffer salts in the process of in-gel digestion. Unlike the Tris/ hydrochloric acid, ammonium bicarbonate is basic and therefore no pH adjustment is needed. Moreover, ammonium bicarbonate can be completely removed and evaporated by vacuum concentrator into ammonia gas, carbon dioxide and water. The continually applied inorganic buffer salts and acetonitrile respectively in the washing and dehydration steps efficiently destained the gel pieces by swelling and shrinking of the gel pieces (Shevchenko *et al.*, 2007). Finally, white dried gel pieces are obtained, which was then saturated with a specific protease for protein digestion purpose. After an incubation period, the peptide digestion product is extracted with solution for peptide recovery (Kumarathasan *et al.*, 2005).

1.13.1 Trypsin

Among the commercially available enzymes, the serine endopeptidase trypsin is mostly used (Smith and Wheeler, 1996). Trypsin hydrolyse the polypeptide chain specifically at arginine and lysine on the side of carboxyl terminal (Granvogl *et al.*, 2007). The hydrolysis rate is reduced if acidic or prolyl residues occur next to the cleavage site. Typically, trypsin can function actively between pH 7 and 9 with an optimum pH approximately at 8. At extremely acidic or basic condition, the activity of trypsin can be reversible inactivated (Smith and Wheeler, 1996).

1.14 Liquid Chromatography- Mass Spectrometry

The coupling of mass-spectrometry to liquid chromatography is a very sensitive and reliable technique that can be applied to a wide range of biological samples. This coupling technique progressed sluggishly in the early age. The employ of liquid chromatography-mass spectrometry is limited due to the incompatibility of the early mass spectrometric ion source with the liquid chromatographic system. In the 1980s, Fenn developed an interface termed as electrospray ionization source to remove the chromatographic mobile phase before further analysis by mass-spectrometry. This development had carried a great impact on proteomic studies (Pitt, 2009).

Liquid chromatography is used to separate a complex protein mixture (Simpson, 2002). However, identity of an unknown analyte cannot be determined based on its similarity to reference retention time alone (Ardrey, 2003).

Mass spectrometry is powerful technique for compound identification and molecular structure study (Larsen and Roepstorff, 2002). The application of mass spectrometry is incapable to obtain single mass spectrum for an analyte of the mixture. The data for all compound ions is reported in a mass spectrum leads to complicated analysis. Thus, coupling of high performance liquid chromatography with mass spectrometry enables definitive identification through mass determination and the quantitative analysis of the compound of interest.

1.14.1 Reverse-phased High Performance Liquid Chromatography (RP-HPLC)

The International Union of Pure and Applied Chemistry (IUPAC) defines chromatography as “a physical method of separation in which the components to be separated are distributed between two phases, one of which is stationary (the stationary

phase), while the other (the mobile phase) moves in a definite direction." The mobile phase for the liquid chromatography is a mixture of solvent which distributes on a column (stationary phase) in a definite direction (Ardrey, 2003). High performance liquid chromatography which carries out a reverse phase chromatographic separation utilizes a polar mobile phase with a non-polar stationary phase (Simpson, 2002).

The elution time for an analyte to move out from the chromatographic column is termed as retention time (t_{an}). However, it is influenced by column length and mobile phase flow rate. Thus, capacity factor (k') is used to demonstrate the retention time of an analyte and the unretained compounds (Simpson, 2002, Ardrey, 2003).

In proteomic studies, the high performance liquid chromatography separates protein based on the interaction of the hydrophobic binding between amino acid side chain of a protein and hydrophobic surface of chromatographic column (Simpson, 200). Typically, a mixture of solvent is use in the high performance liquid chromatography mobile phase. The initial condition of mobile phase always started with a highly aqueous solvent followed by modification of subsequent mobile phase composition to produce a gradient condition along the chromatographic separation process. Thus, a highly polar protein is eluted earlier than the less polar protein. The high performance liquid chromatography consists of five components: pump, injector, mobile phase, column and detector (Ardrey, 2003). However, the separated samples are further analysed and then detected in the mass spectrometry. Thus, the detector in the high performance liquid chromatography is not being used.