# ลำดับกรดอะมิโนและฤทธิ์ทางชีวภาพของโปรตีนจากขมิ้นอ้อย *Curcuma zedoaria* Rosc. และว่านคันทมาลา *Curcuma aromatica* Salisb.

นางสาวพรพิมล ทิพย์ธารา



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรดุษฎีบัณฑิต
สาขาวิชาเคมี ภาควิชาเคมี
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2549

# AMINO ACID SEQUENCES AND BIOLOGICAL ACTIVITIES OF PROTEINS FROM *Curcuma zedoaria* Rosc. AND *Curcuma aromatica* Salisb.

**Miss Ponpimol Tipthara**



**A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy Program in Chemistry**

**Department of Chemistry**

**Faculty of Science**

**Chulalongkorn University**

**Academic year 2006**

| | |
|---|---|
| **Thesis Title** | AMINO ACID SEQUENCES AND BIOLOGICAL ACTIVITIES OF PROTEINS FROM *Curcuma zedoaria* Rosc. AND *Curcuma aromatica* Salisb. |
| **By** | Miss Ponpimol Tipthara |
| **Filed of study** | Chemistry |
| **Thesis Advisor** | Associate Professor Amorn Petsom, Ph.D. |
| **Thesis Co-advisor** | Associate Professor Polkit Sangvanich, Ph.D. |

Accepted by the Faculty of Science, Chulalongkorn University in Partial Fulfillment of the Requirements for the Doctoral Degree

.....Dean of the Faculty of Science
(Professor Piamsak Menasveta, Ph.D.)

**THESIS COMMITTEE**

..... Chairman
(Professor Sophon Roengsumran, Ph.D.)

..... Thesis Advisor
(Associate Professor Amorn Petsom, Ph.D.)

..... Thesis Co-advisor
(Associate Professor Polkit Sangvanich, Ph.D.)

..... Member
(Assistant Professor Thumnoon Nhujak, Ph.D.)

..... Member
(Assistant Professor Warinthorn Chavasiri, Ph.D.)

..... Member
(Chantragan Phiphobmongkol, Ph.D.)

พรพิมล ทิพย์ธารา : ลำดับกรดอะมิโนและฤทธิ์ทางชีวภาพของโปรตีนจากขมิ้นอ้อย *Curcuma zedoaria* Rosc. และว่านคันทมาลา *Curcuma aromatica* Salisb. (AMINO ACID SEQUENCES AND BIOLOGICAL ACTIVITIES OF PROTEINS FROM *Curcuma zedoaria* Rosc. AND *Curcuma aromatica* Salisb.) อ. ที่ปรึกษา : รศ. ดร. อมร เพชรสม, อ.ที่ปรึกษาร่วม : รศ. ดร. พลกฤษณ์ แสงวณิช, 136 หน้า.

พืชตระกูลขิง-ข่ามีรายงานสารออกฤทธิ์ทางชีวภาพหลายชนิด เช่น เคอร์คิวมิน (Curcumin) จากขมิ้นอ้อยซึ่งสามารถฆ่าเซลล์มะเร็งได้ สารสกัดเมทานอลจากว่านคันทมาลามีฤทธิ์ต้านการอักเสบได้ อย่างไรก็ตามยังไม่มีรายงานการศึกษาองค์ประกอบของโปรตีนที่มีฤทธิ์ทางชีวภาพของขมิ้นอ้อยและว่านคันทมาลา ดังนั้นวัตถุประสงค์ของงานวิจัยในครั้งนี้คือเพื่อสกัดโปรตีนจากพืชทั้งสองสปีชีส์นี้ เพื่อทดสอบฤทธิ์ทางชีวภาพและเพื่อพิสูจน์เอกลักษณ์ของโปรตีนที่มีฤทธิ์ทางชีวภาพ โปรตีนผสมทำให้บริสุทธิ์ได้โดยเทคนิคโครมาโทกราฟีและเจลอิเล็กโทรโฟเรซิส ผลการทดลองแสดงว่าโปรตีนบริสุทธิ์จากขมิ้นอ้อยมีฤทธิ์ทำให้เม็ดเลือดแดงของกระต่ายเกิดปฏิกิริยาการจับกลุ่ม (hemagglutination) และสามารถยับยั้งการทำงานของเอนไซม์แอลฟากลูโคซิเดส (α-glucosidase inhibition) ขณะที่โปรตีนผสมจากว่านคันทมาลามีฤทธิ์ทำให้เม็ดเลือดแดงของกระต่ายเกิดปฏิกิริยาการจับกลุ่ม (hemagglutination) โดยการใช้เทคนิคทางแมสสเปกโทรเมทรีร่วมกับการสืบค้นฐานข้อมูลลำดับกรดอะมิโนของโปรตีน ระบุว่าโปรตีนบริสุทธิ์จากขมิ้นอ้อยมีน้ำหนักโมเลกุล 13,448 ดาลตัน และมีลำดับกรดอะมิโนคล้ายคลึงกับโปรตีนชนิดจับกับน้ำตาลแมนโนสจากกล้วยไม้สปีชีส์ *Epipactis helleborine* ขณะที่โปรตีนผสมจากว่านคันมาลามีน้ำหนักโมกุลอยู่ในช่วง 8 ถึง 36 กิโลดาลตัน และมีลำดับกรดอะมิโนคล้ายคลึงกับเลคติน (lectin) หลายสปีชีส์



ภาควิชา.....เคมี..... ลายมือชื่อนิสิต.....พรพิมล ทิพย์ธารา.....
สาขาวิชา.....เคมี..... ลายมือชื่ออาจารย์ที่ปรึกษา.....
ปีการศึกษา.....2549..... ลายมือชื่ออาจารย์ที่ปรึกษาร่วม.....

## 4473820823 : MAJOR CHEMISTRY

KEY WORD: AMINO ACID SEQUENCE / *Curcuma zedoaria* / *Curcuma aromatica*

PONPIMOL TIPTHARA : AMINO ACID SEQUENCES AND BIOLOGICAL ACTIVITIES OF PROTEINS FROM *Curcuma zedoaria* Rosc. AND *Curcuma aromatica* Salisb. THESIS ADVISOR : ASSOC. PROF. AMORN PETSOM, Ph.D. THESIS COADVISOR : ASSOC. PROF. POLKIT SANGVANICH, Ph.D., 136 pp.

Zingiberaceae plants have been reported to contain many biologically active compounds such as curcumin from *Curcuma zedoaria* Rosc. contain cytotoxicity to cancer cell, methanol extract from *Curcuma aromatica* Salisb has anti-inflammability. However, there has been no report on biological active proteins of *Curcuma zedoaria* Rosc. and *Curcuma aromatica* Salisb. Thus, the aims of this research are to extract proteins from these plants, to test their biological activity, and to identify biologically active proteins. The protein mixtures were purified by column chromatography and gel electrophoresis. Results showed that the purified protein from *Curcuma zedoaria* Rosc. has hemagglutinating activity to rabbit erythrocyte and inhibit the activation of α-glucosidase enzyme, while the protein mixture from *Curcuma aromatica* Salisb. has hemagglutinating activity to rabbit erythrocyte. Using the combination of mass spectrometry with protein sequence database searching, the purified protein from *Curcuma zedoaria* Rosc was identified to have the molecular weight of 13,448 Da, and to contain amino acid sequence similar to the mannose-specific lectin (Q39728) from *Epipactis helleborine* of Orchidaceae, while the protein mixtures from *Curcuma aromatica* Salisb was found to have molecular weights in a range of 8-36 kDa, and identified as lectins from various species.

Department…….Chemistry……..... Student's signature…Ponpimol Tipthara……

Field of study…..Chemistry…….. Advisor's signature……

Academic year…2006…….. Co-advisor's signature…Polkit Sangvanich……

# ACKNOWLEDGEMENTS

I would like to express my deepest appreciation and sincere gratitude to my advisor, Associate Professor Dr. Amorn Petsom and my co-advisor Associate Professor Dr. Polkit Sangvanich, for their invaluable suggestion and assistance throughout my studies and research at Chulalongkorn University. I wish to thank my graduate committee members, Professor Dr. Sophon Roengsumran, Assistance Professor Dr. Varintorn Chavasiri, Assistance Professor Dr. Thumanoon Nujak and Dr. Chantragan Phiphobmongkol for their valuable suggestions and comments.

I am very grateful to Professor Simon J Gaskell for his kindly accepting to be host at Michael Barber Centre for Mass spectrometry at University of Manchester for laboratory work and supported for mass spectrometric analysis. I would like to thank Dr. Claire Eyers, Dr. Sarah Harts and Dr. Isabel Riba Garcia for their kindly training to me on protein identificatrion techniques and Mass spectrometry. I am also thanking Sue McIntyr for her kindly gel electrophoresis technique and all members of Michael Barber Centre for Mass spectrometry. I would like to thank all members of Protein Unit at Research Center for Bioorganic Chemistry for their help and discussion on the experiments throughout this work.

I would like to thank TRF PhD golden Jubilee Research Fund by the Thailand Research Fund for financial support.

Finally, I wish to express extremely grateful to my mother and sister for their infinite love, understanding, support and looking forward to my graduation. Also, I thank my friends for their friendship and helping during my study.

# CONTENTS

Page

CHAPTER Page

CHAPTER Page

CHAPTER Page

# LIST OF FIGURES

Figure Page

Figure Page

Figure Page

# LIST OF TABLES

# CHAPTER I

# INTRODUCTION

The term proteome analysis or proteomics is well known in biological field. It was defined as the full complement of proteins expressed by a genome (*1*). The scale of a proteomic experiment varies according to its aim, such as determining the protein content of a whole organism, analyzing targeting proteins in a tissue or in cell.

In plant, proteome analysis approaches face specific challenges due to the rigidity of plant cell walls and the accumulation of large quantities of secondary compounds in the vacuole such as phenolic compound. Moreover, plant contains lower protein level than unicellular prokaryotes, animals and yeast. However, mass spectrometry (MS) overcome the problem of analysis less abundant protein due to it can analysis as level of femtomole ($1\times10^{-15}$ mole). Analysis of proteins by MS was revolutionized in the 1980's with the development of 'soft' ionization techniques such as electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI) (*2-4*). These techniques solved the difficult problem of generating ions from large, nonvolatile analytes such as proteins and peptides. Mass spectrometric data was done to characterize primary protein sequence information and relative and absolute quantification of protein abundance. Protein identifications are accomplished using several approaches, often in combination. MALDI-TOF is the first method employed; exact masses of peptides resulting from enzyme digestions such as trypsin, are used in conjunction with a number of data base searching routines (*5-9*). An alternative strategy is the use of tandem MS (MS/MS) to obtain mass spectra of peptide fragment ions from a particular peptide precursor ion. To identify the proteins, the observed low energy fragment ion spectrum is matched with one that is predicted for a peptide of the same precursor mass found in a data base (*10*). However, successful identification by any of the various data base searching depends on the presents of the proteins in a data base. Therefore, this approach was being limited when the goal is to identify protein from an organism that lacks a sequenced genome or an extensive sequence database. This is another factor limiting plant protein analysis due to the number of fully sequenced plant species is limited. Now, only three species including rice (*Oryza sativa*), poplar (*Populus balsamifera*), and Arabidopsis

(*Arabidopsis thaliana*) have yielded the complete genomic sequence. However, this has opened the door to a new era of plant research aimed towards an integrated understanding of biological processes. Identification of proteins has not presented in database and non-sequence genome will be done by *de novo* sequencing. This method is that amino acid sequences are interpreted from the data directly, and rely on the spectra themselves.

Nevertheless, the field of plant proteomics is rapidly growing. Until now, the plant proteomics methodology has been two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) in combination with MS. To identify separated proteins, the resolved gel spots are typically digested in-gel and then subsequently analyzed offline using MS or tandem MS (MS/MS). Although 2D-PAGE is robust and relatively straightforward, and allows for the separation of thousands of different proteins, it is a rather costly, slow, labor and time-consuming technique. Furthermore, the hydrophobicity, isoelectric point, and molecular weight range of proteins resolvable are resolution limits. In addition, low abundant proteins lie beyond detection limit of 2DE were unable detected. The introduction of non-gel-based proteomics approach called Multidimensional Protein Identification Technology (MudPIT) overcomes the restrictions of 2D-PAGE.

In this research, proteins from medicinal plant belonging to family Zingiberaceae were purified and identified. For chromatographic purification, hemagglutinating activity was used to guide each step and that the purified protein will be assumed as lectin. Also, purified protein and each protein containing fraction was subjected to α-glucosidase inhibition assay. Identification of purified protein is a challenge step due to Zingiberaceae has uncompleted genome and very little known of protein component, however, the identification was done by the combination of *de novo* sequencing and similarity searching. The results were described in Chapter IV by studying with two plant species including, *Curcuma zedoaria* Rosc. and *Curcuma aromatica* Salisb. The benefit of this study demonstrated that aside from the bioactive small molecule, Zingiberaceae also contains the bioactive proteins and the identification of them was done by a powerful mass spectrometric technique. In chapter V, the proteomic was used to studied plant biological system. In this study large-scale proteomic was described by preliminary study of the *Arabidopsis thaliana* guard cell abscisic acid (ABA) signalling network. The changes of protein expression from two states of guard cell extract, treated and untreated

with ABA, were compared by a novel reagent called iTRAQ (Isobaric Tag for Relative and Absolute Quantification). Also, the tandem affinity purification (TAP) method was used in combination with mass spectrometry in order to search the entire proteome that interacts with the well known protein, QST1, in abscisic acid signaling pathway.

The aims of this study are:

1) To purify and identify of a biological active proteins from plant in Zingiberaceae.

2) To utilize proteomic approaches to identify quantitative differences in guard cell enriched proteins between two states of extract from *Arabidopsis thaliana*.

# CHAPTER II

# LITERATURE REVIEW AND THEORY

**Literature review**

## 2.1 Plant proteomics

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 (*11)*, variations in response to physiological events (*12*). A number of review articles have already been published concerning the study of the plant proteome (*13-15*). Plant proteomics has long been associated with the techniques of two-dimensional gel electrophoresis (2DE) and mass spectrometry (MS). The identification of plant proteins by MS was performed in rice (*16*), Arabidopsis (*17*), maize (*18*), barrel medic (*Medicago truncatula*) (*19*), etc. The most comprehensive studies were performed in rice and barrel medic, which are model plants in cereals and legumes, respectively.

Protein expression varies depending on particular species, variety, growth stage, organs and cell organelle in particular environment. The expression profile is closely related to the function of proteins. The analysis of protein differential display, proteins are extracted from the cells under different conditions and are compared, have a number of studies. For example, the proteins which were up- or down-regulated by hormone treatment (*20*), disease (*21*) and stress such as low temperature (*22*), heat (*23*), drought (*24*) were investigated by 2-DE. These analyzes are essential to identify the proteins involved in growth, differentiation, disease resistance and stress tolerance of plants.

## 2.2 Zingiberaceae

Zingiberaceae is a large family consisting of 53 genera and more than 1200 species were found mainly in Asia, Africa and the Pacific Islands (*25*). *Curcuma* is one of genera consists of about 70 species in the world. They have been cultivated for widely use as spices (*26*), a food flavoring and coloring agent (e.g., in curry) as well as in

traditional medicinal applications (*27-30*). Also, some species was commonly used in industrial application (*31*). *Curcuma zedoaria* Rosc. (Kamin-oy in Thai name) (Figure 2.1A) and *Curcuma aromatica* Salisb. (Kanthamala in Thai name) (Figure 2.1B) are used as an anti-inflammatory. Moreover *Curcuma zedoaria* is used as antiseptic and aid to digestion and to relieve flatulence and colic. Also, their rhizomes has been reported to have hepatoprotective activities (*32*), antimicrobial (*33*), cytotoxic (*33*, *34*), antioxidant effects (*35*). In addition, polysaccharides from *Curcuma zedoaria* are reported to have antitumor, genotoxicity and anticlastogenic activities (*36*).

A B

**Figure 2.1** (A) Rhizomes of *Curcuma zedoaria* (Kamin-oy) (B) Rhizomes of *Curcuma aromatica* (Kanthamala)

Many bioactive secondary metabolites, such as curcumin and their derivatives, were extracted from rhizome of *Curcuma* species. In the contrast, protein components of this plant are a little known and report is scant.

**2.3 Lectin and hemagglutinating activity**

Hemagglutinating (lectin) activity was found in crude protein extract from various *Curcuma* species (*37*). Lectins or agglutinins are proteins that recognize and bind to sugar complexes attached to proteins and lipid. The specific agglutination properties of lectins are based on their defined recognition of and binding to sugar (*38*). They are widespread throughout the plant kingdom occurring in a number of plant species (*39*). Their functions

in plant tissue are defended against phytopathogenic microorganisms, phytophagous insects and plant-eating animals (*40*). The defensive role of lectins stimulates the research on possible application of lectins in crop protection. In addition, from the specificity of protein-sugar interaction, lectins could be used as carrier molecules to target drugs specifically to different cells and tissue (*41*). Understanding the expression patterns of lectins in biological systems and the physical nature of the protein–carbohydrate interaction provides important insights into biological information transfer for such diverse areas as microbiology, oncology, and plant pathology. Proteins that interact with sugars are grouped according to how the interaction takes place. Some proteins transport sugars across cell membranes to provide fuel for the cell. Others modify saccharides in metabolic processes or attach them covalently to lipids or other proteins.

Lectins are characterized by specific binding to carbohydrate residues, whether monosaccharides, disaccharides or polysaccharides. The sugar heads on the surface of the erythrocyte specify the different blood groups. Lectins, as an antigenic determinant of blood group, have come to be an important tool in the identification of different blood groups. Erythrocytes display a wide spectrum of glycoproteins on their cell surface. Lectins bind specifically to the terminal sugar residues present on the surface of the erythrocyte surface, causing them to clump, which is referred to as 'hemagglutinating' activity. The various different sugar heads that are present on the erythrocyte surface membrane, bound to either glycoprotein or glycolipids, specify blood groups. The carbohydrate binding property of lectins is thus extensively utilized for blood group typing or subgrouping.

**2.4 α-glucosidase inhibition**

Infection by human immunodeficiency virus (HIV) leads to the development of acquired immunodeficiency syndrome (AIDS). Inhibition of α-glucosidase is involved in HIV infection. At this time, almost all of the anti-HIV drug target reverse transcriptase or protease, two proteins encoded by the virus. A new approach to antiretroviral therapy is to design inhibitors that target discrete steps in the viral entry pathway. The glycoproteins expressed on the surfaces of enveloped viruses are known to be critical for virus-host cell interactions and offer an attractive target for antiviral intervention. The human immunodeficiency virus (HIV) envelope contains two glycoproteins, gpl20 and gp4l, that

are noncovalently linked in an oligomeric structure (*42, 43*). The external glycoprotein, gpl20, binds with a high affinity to the immune receptor protein, CD4. These receptor-binding events trigger changes in the envelope glycoproteins that lead to membrane fusion and virus entry. Approximately 50% of the 120 kDa molecular weight of gp120 is provided by carbohydrate, all of it N-linked that are synthesized by the addition of a preformed glycan glucose$_3$ mannose$_9$ N-acetyl glucosamine$_2$ ($G_3M_9GlcNAc_2$) to an asparagine-associated motif on the nascent polypeptide chain. Removal of the outer α-1,2-linked glucose residue by α -glucosidase I and the two inner α-1,3-linked glucose residues by α-glucosidase II is critical for posttranslational modification of glycoproteins and their subsequent transport to the Golgi region. Many proteins and lectins that bind HIV-1 gp120 envelope carbohydrates have already been characterized. Mannose binding lectin from Amaryllidaceaeand and Orchidaceae family has reported the anti-HIV infection (*44, 45*). In addition, Ng et al. (*46*) has studied to look for bioactive proteins in Chinese ginseng and found 28 kDa protein called Panaxagin can inhibit α-glucosidase with a lower potency (4.8% at 5mg/ml)

**2.5 Arabidopsis guard cell ABA-signalling network**

The surface of aerial parts of the vast majority of higher plants contains many microscopic apertures known as stomata. These are formed by a pair of guard cells which increase and reduce their turgor pressure as a result of osmotic pressure, leading to opening and closing of the stomatal pore. Stomata are involved in gas exchange between plant and atmosphere, particularly water vapour and carbon dioxide. In times of drought the plant closes its stomata in order to guard against excessive water loss. Stomatal closure is initiated by an intricate signaling network, and in drought response the phytohormone abscisic acid (ABA) is of great importance (*47*) (Figure 2.2). ABA is a plant stress hormone whose levels increase in response to various environmental factors including drought, cold and exposure to auxin herbicides. Despite many years of research the ABA receptor has yet to be isolated. Recently the protein kinase OST1 (for OPEN STOMATA1) was found to be a central component of the guard cell abscisic acid signalling pathway in the model organism *Arabidopsis thaliana*, although its precise mechanism of action is still not entirely understood. In particular little is known about the proteins that OST1 physically interacts with in order to exert its effect.

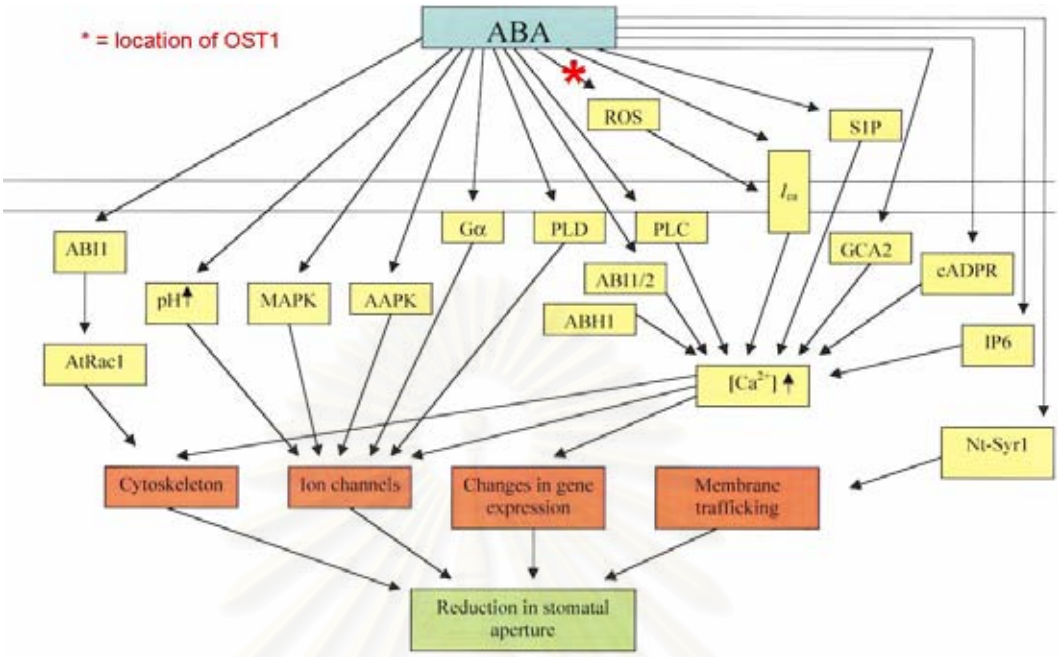



**Figure 2.2** A schematic overview of the ABA signalling network in guard cells. The horizontal bars represent the cell membrane. Yellow boxes represent signalling intermediates in the pathway, red boxes represent events leading to stomatal closure. This diagram emphasises how ABA signalling functions as an intricate network, with relationships between separate elements, rather then as a simple linear pathway. The red asterisk denotes the putative position of OST1 in the pathway.

*Abbreviations:* AtRac1 = a Rho family GTPase; ROS = reactive oxygen species; S1P = sphingosine-1-phosphate; $I_{Ca}$ = plasma membrane calcium permeable channel; MAPK = mitogen-activated protein kinase; AAPK = ABA activated protein kinase; $G_{\alpha}$ = heterotrimeric G-protein α subunit; Nt-Syr1 = syntaxin involved in membrane trafficking; PLC = phospholipase C; PLD = phospholipase D; cADPR = cyclic adenosine diphosphoribose; IP6 = inositol hexakisphosphate; ABH2 = mRNA cap-binding protein; ABI1 and ABI2 = "ABA-insensitive" type 2C phosphoprotein phosphatases; GCA2 = currently unidentified guard cell protein. (Reproduced from (*47*))

**Theory**

## 2.6 Genome and proteome

With the sequence information now available, MS-based proteomics has become an increasingly important key technology to analyze biological systems and to reveal their complex functions and interactions. Because cellular proteomes are complex and highly dynamic, may consist of more than 100,000 protein species with different chemical and physical properties. Furthermore, post-translational modifications increase the number of protein species several-fold compared to the number of genes. Generally, identification of the proteome is difficult 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 (*1*), to the complex, human (*48*), 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. 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.

## 2.7 Isolation of protein

The initial step of 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, 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.

During the initial stages of many protein purification procedures the protein of interest is present in dilute solution, thus large volumes of process liquid must be handles. It thus becomes necessary to concentrate such solution in order to render the extract volume for subsequent purification steps.

Protein precipitation can be promoted by agents such as neutral salts, organic solvents, high molecular mass polymers, or by appropriate pH adjustment. Ammonium sulphate is likely the most common protein precipitant utilized. This neutral salt is particularly popular due to its high solubility, inexpensiveness, lack of denaturing properties towards most proteins, and it stabilizing effect on many proteins. The addition of small quantities of neutral salts to a protein solution often increases protein solubility; the ‘salting in’ effect. However, increasing salt concentrations above an optimal level leads to destabilization of proteins in solution and eventually promotes their precipitation. This is known as ‘salting out’. At high concentrations, such salts effectively compete with the protein molecules for water of hydration. This promotes increased protein-protein interactions, predominantly interactions between hydrophobic patches on the surface of adjacent protein molecules. Such increased protein-protein interactions eventually result in protein precipitation.

## 2.8 Chromatography

Proteins are separated according to their different physiochemical properties. The composition of the mobile phase is manipulated to exploit differences in protein/peptide properties and therefore retention upon the stationary phase of the column. The mobile phase is controlled via a low or high performance liquid chromatography pump and proteins/peptides are detected using UV absorbance to produce a chromatogram.

### 2.8.1 Gel filtration chromatography (GFC)

GFC is unique in that fractionation is based on the relative size of the protein (Figure 2.3). In contrast to other techniques, none of the proteins are retained on the column and this feature has both advantages and disadvantages; the function of fragile proteins is undamaged by binding to a chromatographic support, but the absence of such binding limits the achievable resolution of the chromatography. GFC is performed using porous beads as the chromatographic support. A column constructed from such beads will have two measurable liquid volumes, the external volume, consisting of the liquid between the beads, and the internal volume, consisting of the liquid within the pores of the beads. Large molecules will equilibrate only with the external volume while smaller molecules will equilibrate with both the internal and external volumes depending on pore size. A mixture of proteins is applied to the top of the column and allowed to percolate through the column. Due to, GFC has low resolution, since none of the proteins are retained in the column and non-ideal flow occurs around the beads, therefore it may be best applied relatively late in a separation procedure when the numbers of other proteins are small and when the preceding step has fractionated the protein mixture on the basis of a different property. For example, fractions obtained from ion exchange chromatography will likely contain a mixture of proteins, each having similar net charge but a range of molecular weights.

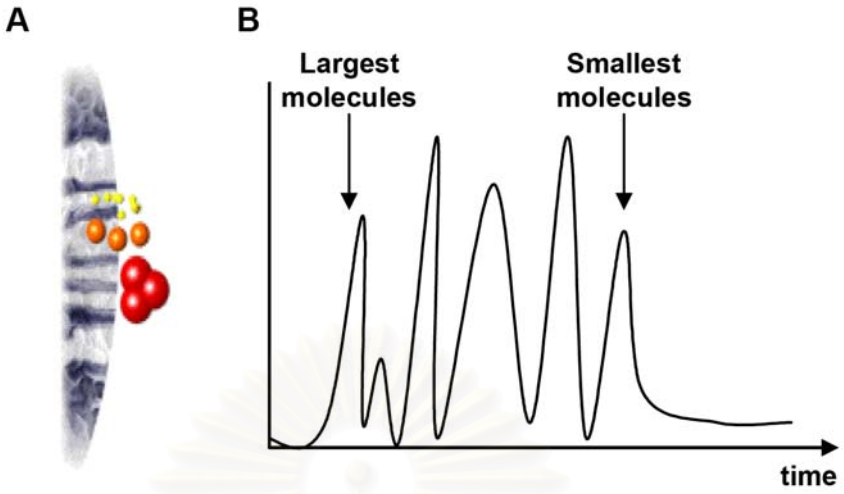



**Figure 2.3** Gel filtration chromatography. (A) Cartoon representing proteins of different sizes moving through immobilized porous beads, (B) Example of a GFC chromatogram. Small molecules will enter the pores whilst larger molecules will be excluded and elute more quickly.

### 2.8.2 Ion exchange chromatography (IEC)

Ion exchange chromatography (IEC) has been the most widely used technique for the isolation and purification of biological macromolecules. The hydrophilic sorbents, with fixed ionogenic functions, facilitate the separation of proteins and other biopolymers based on their relative charge without denaturation and with relatively high selectivity and resolution (Figure 2.4). Ion exchangers are insoluble solid matrices containing fixed ionogenic groups. The two major classes of ion exchangers are cation and anion exchangers which have negatively and positively charged functional groups, respectively. Strong Cation Exchangers (SCX) usually contains sulfonic groups and Strong Anion Exchangers (SAX) often has quaternary ammonium functions. The ionogenic functions in weak anion exchangers are amines, whereas weak cation exchangers usually contain carboxylic acid groups. The terms strong and weak reflect that weak ion-exchanger groups are ionized only in a narrower pH range than strong ion exchanger groups.

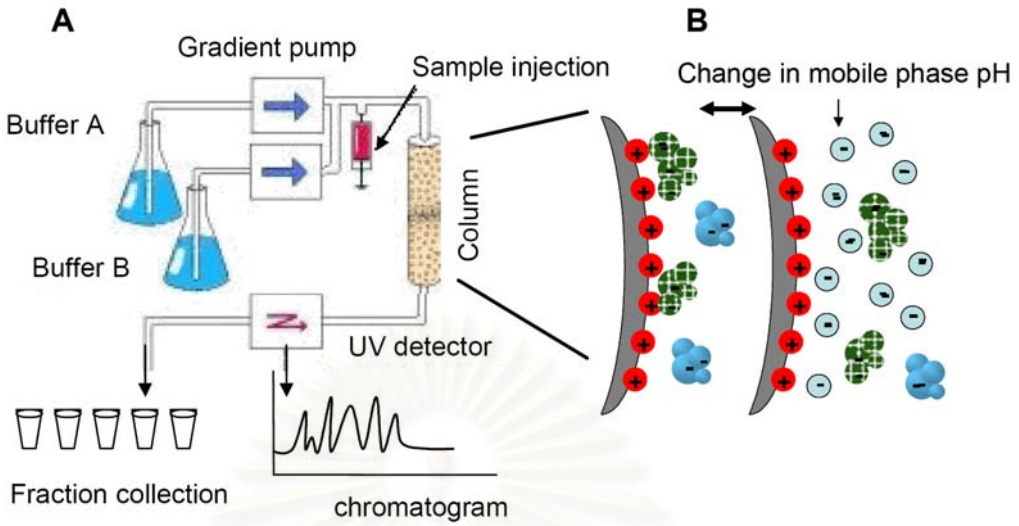



**Figure 2.4** (A) Schematic of a gradient IEC system and (B) Equilibrium between adsorption and desorption of charged molecules to the column ion exchanger.

Protein charge and retention in IEC is dependent on the pH of the mobile-phase relative to the p*I* of the protein. The net charge of a protein is zero at its isoelectric point. Therefore when the pH of the mobile phase equals the p*I* of the protein, theoretically, there should be no electrostatic interaction between the neutral protein and the chromatographic surface. When the mobile-phase pH decreases below the p*I* of the protein, the protein confers a net increase in positive charge which results in stronger retention of the protein on a cation exchanger. Similarly as the pH increases, the protein acquires more negative charge and retention to an anion exchanger increases. Manipulation of the mobile phase pH therefore represents an attractive method for regulating the degree of protein interaction (Figure 2.4).

The most frequently used mode of elution in IEC is a gradient of increasing salt concentration in the mobile phase. The sample mixture is introduced to the column as a pulse, and the components are separated as a consequence of their unequal partitioning between the mobile and stationary phases and appear at the column outlet as individual peaks. By increasing the salt concentration of the mobile phase during the

chromatographic run, proteins can be readily separated. This change in buffer can be either continuous (gradient) or stepwise (isocratic).

### 2.8.3 Reversed-phase chromatography (RP-LC)

Reversed-phase (RP) liquid chromatography separates proteins/peptides on the basis of their hydrophobicity. The popularity of RP-HPLC can be attributed to the excellent resolution that can be achieved for closely related, as well as structurally disparate substances, under a wide variety of chromatographic conditions. The RP-HPLC experimental system usually comprises an *n*-alkylsilica-based sorbent from which peptides or proteins are eluted with increasing concentrations of an organic solvent, such as acetonitrile containing an ionic modifier, e.g. trifluoroacetic acid (TFA). With modern instrumentation and columns, complex mixtures of proteins and peptides can be separated and low picomolar amounts of resolved components can be collected. Separations can be readily manipulated by changing the gradient slope, temperature, ionic modifier, organic solvent composition or alkyl length of the silica stationary phase, e.g. C4 *vs* C18. The specificity of RP-HPLC arises through the selective interactions between the proteins or peptides introduced into the column and the immobilized ligand on the stationary phase. Therefore, the practical widespread application of RP-HPLC has been accompanied by a significant improvement in the understanding of the molecular basis of retention and its impact on conformational stability. The mechanism by which peptide or protein solutes are retained in RP-HPLC depends on the hydrophobic expulsion of the peptide from a polar mobile phase and concomitant adsorption into the non-polar solvent. Under these conditions, peptides or proteins are retained to different extents depending on their intrinsic hydrophobicities, the conditions of the mobile phase and the nature of the sorbent ligands. The ability to manipulate solute resolution through changes in the composition of the mobile phase represents a powerful characteristic of RP-HPLC systems. At low pH values silica-based sorbents are chemically stable and the surface silanols fully protonated. TFA is the most popular of the acidic additives used because of its volatility; however, phosphoric, formic hydrochloric or acetic acid can also be used (*49, 50*). The three most commonly used organic solvent modifiers are acetonitrile, methanol or 2-propanol which all exhibit high optical transparency in the UV detection wavelengths used. Parameters which can influence the resolution of proteins in RP-HPLC

include the gradient time, the gradient shape, the mobile-phase flow rate and the operating temperature.

### 2.9 Mass spectrometry

Mass spectrometry is one of the most versatile and powerful techniques used for chemical characterization. It is essential in modern analytical chemistry (environmental pollutant and drug detection), biochemistry (protein identification and sequencing), and synthetic techniques (drug synthesis). A mass spectrometer creates charged particles (ions) from molecules. It then analyzed those ions to provide information about the molecular weight of the compound and its chemical structure. In this post-genomic era, mass spectrometry (MS) has become the central tool of proteomics.

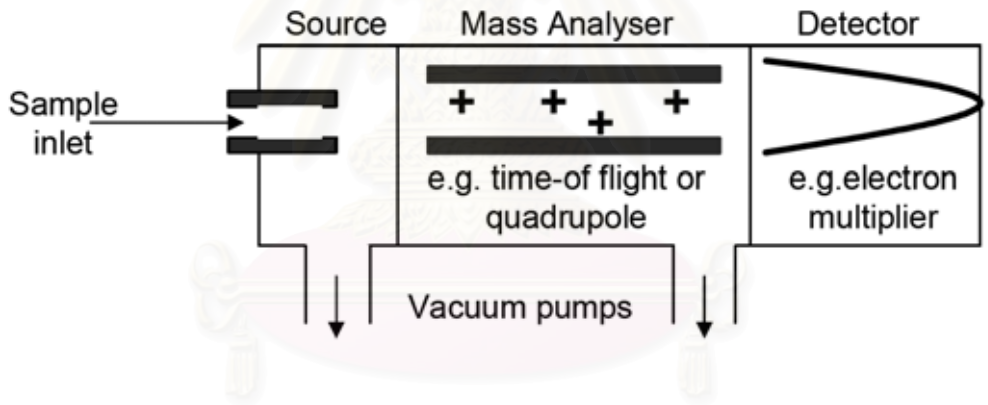



**Figure 2.5** Schematic representing the major components of a mass spectrometer.

#### 2.9.1 Components of a mass spectrometer

The five basic parts of any mass spectrometer (Figure 2.5) are: a vacuum system; a sample introduction device, an ionization source, a mass analyzer and an ion detector. Combining these parts, a mass spectrometer determines the molecular weight of chemical compounds by ionizing, separating and measuring molecular ions according to their mass-to-charge ratio (m/z). The ions are generated in the ionization source by inducing either the loss or gain of a charge (i.e. protonation, or deprotonation). Once the ions are

formed in the gas phase they are electrostatically directed into a mass analyzer, separated according to mass and finally detected. The result of ionization, ion separation, and detection is a mass spectrum that can provide molecular weight or even structural information.

**1) Ionization methods**

Ionization is the act of inducing a charge on a neutral molecule and a number of ionization methods are used in modern MS. These include, Matrix-Assisted Laser Desorption Ionization (MALDI), Electrospray (ESI), Atmospheric Pressure Chemical Ionization (APCI), Fast Atom Bombardment (FAB) and Electron Ionization (EI). High energy ionization methods such as EI are not suitable for the analysis of macromolecules as excessive fragmentation and thermal decomposition may occur during vaporization. Softer methods, such as MALDI and ESI, are used in the analysis of macromolecules as the rate of energy transfer and the internal energy of the generated ions is low. Minimal fragmentation of the ions will occur without the introduction of further internal energy, for example through gas phase or surface collision of analyte molecular ions.

**2) Mass analysis**

Immediately following ionization, gas phase ions enter a region of the mass spectrometer known as the mass analyzer. The mass analyzer is used to separate ions within a selected range of mass-to-charge (m/z) ratios. Ions are typically separated by magnetic fields, electric fields, or by measuring the time it takes an ion to travel a fixed distance. The analyzer is an important part of the instrument because of the role it plays in the instrument's accuracy and mass range.

**3) Detection**

The detector converts the ion population into a digitized measure of ion intensity. Multichannel plate detectors convert the signal into an electronic signal which is sent to a data processing system

### 2.9.2 Ionization techniques

#### 1) Matrix-Assisted Laser Desorption Ionization (MALDI)

MALDI, developed by Michael Karas and Franz Hillenkamp (*4*) has become a widespread analytical tool for peptides, proteins and most other biomolecules (oligonucleotides, carbohydrates, natural products, and lipids). The efficient and directed energy transfer during MALDI provides high ion yields of the intact analyte and allows measurement of compounds with high accuracy and sub-picomole sensitivity. In MALDI analysis, the analyte is first co-crystallized with a large molar excess of a matrix compound, usually a UV-absorbing weak organic acid. Pulsed laser radiation of this analyte-matrix mixture results in the vaporization of the matrix which carries the analyte with it (Figure 2.6). By absorbing the laser light energy, the matrix indirectly causes the analyte to vaporize. The matrix also serves as a proton donor and receptor, acting to ionize the analyte in both positive and negative ionization modes, respectively (*51*). Ions are then accelerated into the analyzer using a potential gradient.

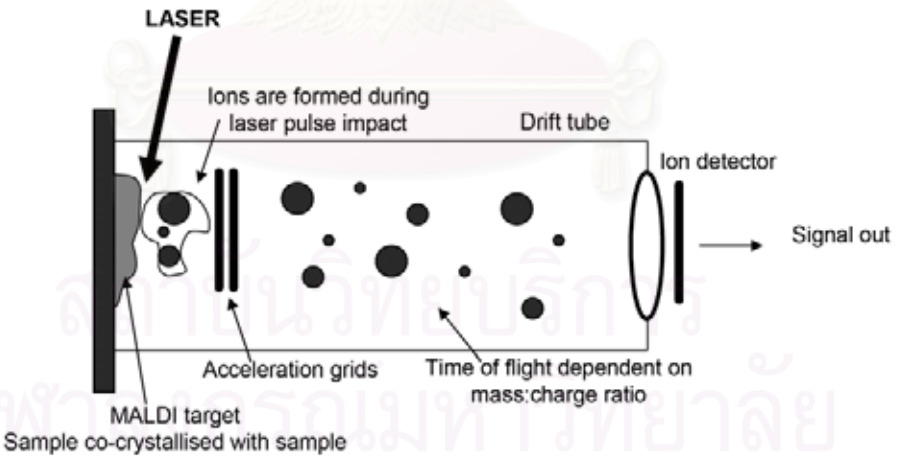



**Figure 2.6** Schematic of MALDI MS. The sample is loaded onto the matrix plate and subjected to a pulsed UV laser which ionizes the molecules. Smaller ions have a faster ‘time-of-flight’ through the drift tube allowing m/z values to be calculated.

The most commonly used matrices are listed in Table 2.1. Most MALDI mass spectrometers use N2 lasers which emit photons in the UV spectrum at 337nm.

**Table 2.1** Common matrices for UV-MALDI

| Matrix | Application |
|---|---|
| Sinapinic acid (3,5-Dimethoxy-4-hydroxy cinnamic acid) | Peptides and proteins >10,000 Da |
| α-cyano-4-hydroxycinnamic acid (CHCA) | • **Dried drop application**—Peptides/proteins <10,000 Da<br>• **Thin layer application**—Peptides <~3,000 Da |
| 2,4,6-Trihydroxy acetophenone (THAP) | Small oligonucleotides <3,500 Da, acidic carbohydrates, acidic glycopeptides, acid sensitive compounds |
| 3-hydroxypicolinic acid (3-HPA) in ammonium citrate | Large oligonucleotides >3,500 Da |
| 2,5-dihydroxybenzoic acid (2,5-DHB) | Peptides, neutral or basic carbohydrates, glycolipids (negative ions), polar and non-polar synthetic polymers, small molecules |
| 2,5-dihydroxybenzoicacid and 5-methoxysalicylic acid (DHBs) | Peptides and proteins >10,000 Da, glycosylated proteins |
| Dithranol and Ag TFA | Aromatic polymers |
| trans-3-indoleacrylic acid (IAA) | Non-polar polymers |

Although MALDI is a soft ionization technique, some fragmentation of ions will occur. This post source decay (PSD) can be exploited to provide sequence information for protein and peptide identification (*52*). MALDI has become the most widely used MS technique in non-specialist laboratories. It has the advantages of low level sample consumption, allowing multiple experiments on the same sample, and rapid acquisition of large amounts of data.

**2) Electrospray ionization (ESI)**

ESI was developed in 1988 by Fenn (*2*) and has had an enormous impact on modern proteomics. ESI generates ions directly from solution (usually an aqueous or aqueous/organic solvent system) by passing along a short length of capillary tube, the end of which is held at a high positive or negative electric potential (typically 3-5 kV) (Figure

2.7). As the droplet decreases in size, the electric charge density on its surface increases. The mutual repulsion between like charges on this surface becomes so great that it exceeds the forces of surface tension, and ions begin to leave the droplet by solvent evaporation or direct desorption through what is known as a ‘Taylor cone’. The ions are then electrostatically directed into the mass analyzer. Vaporization of these charged droplets results in the production of singly or multiply-charged gaseous ions. The number of charges retained by an analyte can depend on such factors as the composition and pH of the electrosprayed solvent as well as the chemical nature of the sample. For small molecules (< 2000 Daltons) ESI typically generates singly or doubly charged ions, while for larger molecules, the ESI process typically gives rise to a series of multiply-charged species.

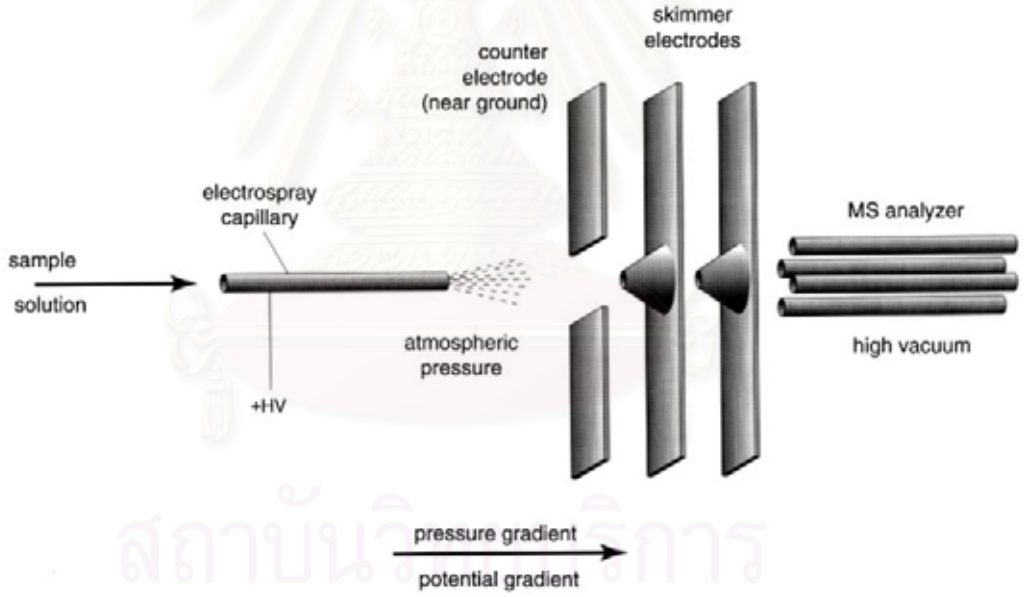



**Figure 2.7** Essential features of the electrospray interface. (Reproduced from (*52*))

Electrospray is typically performed in either the infusion mode, the nanoelectrospray format, or in combination with high-performance liquid chromatography. Nanospray uses low flow rates (ηl/min), to introduce peptides in solution via an electrospray needle. Alternatively, RP-HPLC can be directly coupled to the MS instrument via an electrospray capillary. In this instance, MS analysis of the

components of the sample takes place on-line as they elute from the chromatography column. ESI is a soft, low energy process which produces multiply-charged intact molecular ions.

### 2.9.3 Mass analyzers

#### 1) Time of flight (ToF) mass analyzer

A time-of-flight (ToF) analyzer is one of the simplest mass analysis devices and is commonly used with MALDI. These analyzers use the acceleration of a packet of ions with the same amount of energy to a detector. Smaller ions traverse the fixed distance to the detector first, because of their greater velocity, and the larger ions take longer. The arrival time of an ion at the detector is dependent upon the mass, charge, and kinetic energy of the ion. Ions released from a pulsed source will have the same kinetic energy. Since kinetic energy (Ek) is equal to $mv^2/2$ (where *m* is mass, *v* is velocity), ions will travel a given distance, within a time which is dependent upon their m/z. Theoretically, the mass range detectable using ToF is unlimited, but in practice only a limited region can be detected at high resolution (*53*). Resolution of mass analyzers is calculated by comparing the expected m/z value with the observed m/z and is expressed in parts per million (ppm). Linear ToF resolution can be improved by focusing the ions using a reflectron instrument.

#### 2) Quadrupole mass analyzer

Quadrupole mass analyzers separate ions based on their different m/z-dependent trajectories within a radio-frequency/direct current field. They are made up of four precisely parallel rods with a direct current (DC) voltage and a superimposed radiofrequency (RF) potential (Figures 2.8). Quadrupole mass analyzers have been used in conjunction with electron ionization sources since the 1950s and are the most common instruments in existence today. Under a particular set of RF/DC values, ions of a particular m/z ratio will have stable trajectories through the quadrupole cell, other ions will collide with the quadrupole rods. Therefore, in order to obtain a spectrum across a range of m/z values quadrupole instruments scan over a range of RF/DC values. Quadrupole analyzers are tolerant of relatively poor vacuums (~5 x 10-5 torr), which

makes them well-suited to electrospray ionization, since the ions are produced under atmospheric pressure conditions.

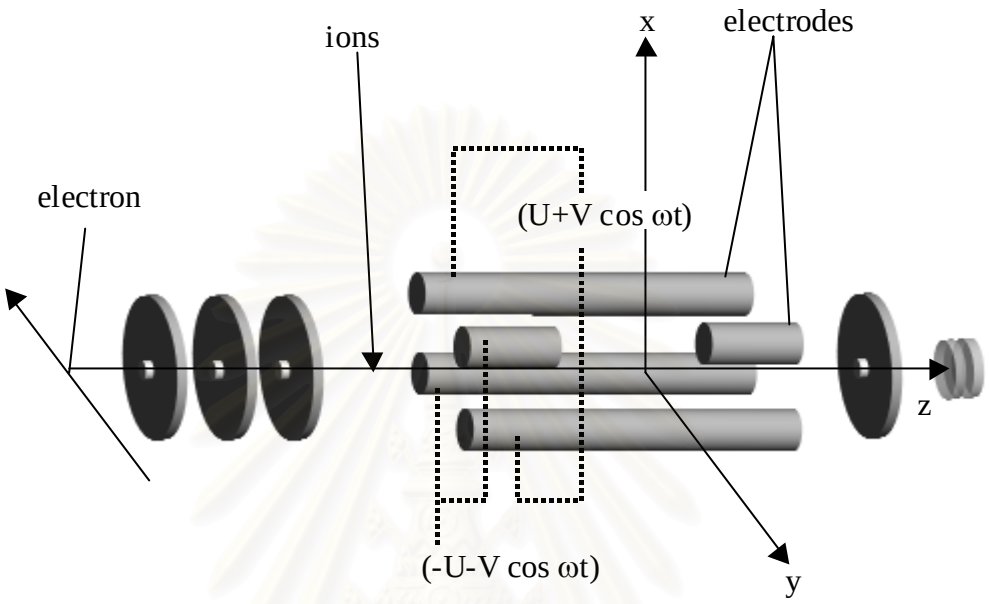



**Figure 2.8** Schematic of quadrupole rods.

The trajectory of ions will pass in the z-direction between the four rods (Figure 2.8). The Mathieu equations of ion motion lead to the definition of two parameters:

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

where, ze = charge of ion

U = direct current (DC) potential

(V + cos ωt) = radio frequency potential

and $r_o$ = ½ of the distance between opposite rods

In these equations, $r_o$ and $\omega$ are constants but some further constraint must be made on the instrument in order to determine m/z as a single variable. Generally, a and q are chosen to be proportional to one another, so the relation between a and q is

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

Then, since a, q, U and V are interrelated, the m/z of transmitted ions varies with U and V. This is particularly convenient since U and V are both easily controllable by electronic circuitry.

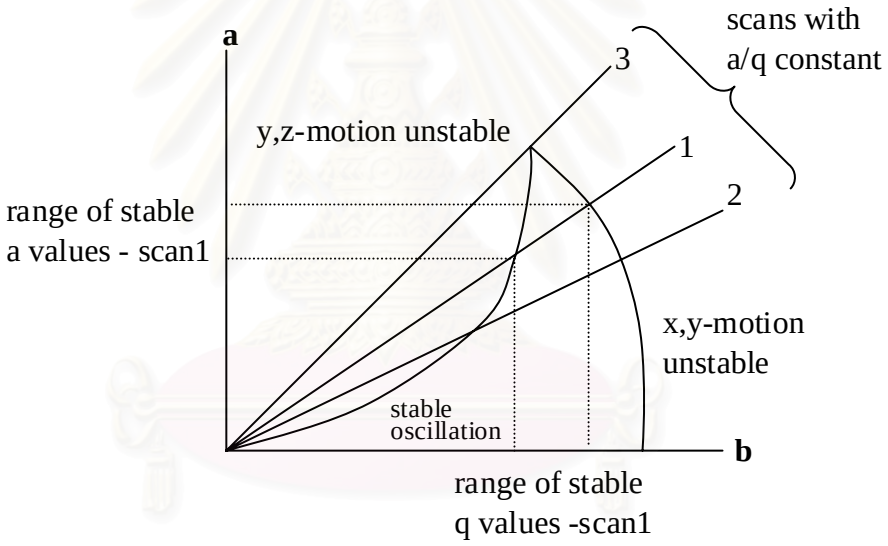



**Figure 2.9** A portion of the stability region for ion trajectories in the quadrupole mass filter.

For the graph in Figure 2.9, 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 2.9 are called scan lines and correspond to variation of U and V such that U/V (a/q) remain constants.

The actual scan line used experimentally is determined by the desired mass resolution. In mass spectrometry, resolution is defined as M/ΔM, where M is the mass of the ion and ΔM is the smallest increment of mass that can be distinguished by the analyzer. For the quadrupole mass analyzer, ΔM is commonly about one dalton over the entire mass range. In theory, the mass resolution of a quadrupole can be increased to a high value by operating with an a/q value close to the apex of the stability area (Figure 2.9). In practice, however, the attainable resolution depends upon the initial ion velocities in the x and y directions and upon the position at which the ion enters the filter.

In the quadrupole analyzer, the resolution is thus dependent on the slope of the scan line. The closer the line comes to the apex of the stability boundary, the greater the resolution. The scan-line defines the ranges of values for both a and q that allow stable ion motion through the quadrupole at any instant (Figure 2.9). Since a and q are both functions of m/z, all ions within the range of m/z values defined by the allowable values for a and q traverse the rods at the same time. In Figure 2.9, the highest resolution is achieved with scan line 3. During quadrupole mass analyzer scanning, there are changing voltages (U,V) and a fixed frequency (ω) and ratio of a/q. By continuously increasing U and V, ions of all of m/z will successively pass along the axis of the quadrupole rods to enable recording of mass spectra.

### 2.9.4 Tandem Mass Spectrometry

Tandem MS (MS/MS) is any method involving at least two stages of mass analysis, either in conjunction with a dissociation process or a chemical reaction that causes a significant change in the mass or charge of an ion. The most commonly used MS/MS experiment involves selection of a precursor ion which undergoes fragmentation. This can be achieved by inducing ion/molecule collisions by a process known as collision-induced dissociation (CID). CID is accomplished by selecting an ion of interest within a mass analyzer and introducing that ion into a collision cell. The selected ion then collides with a collision gas (typically argon or helium) resulting in fragmentation. The fragments are then analyzed to obtain a product ion spectrum.

Collision energy is categorized into two groups: high-energy (keV) and low-energy (up to 200 eV). In the classical CID process, the fragmentation event is usually

single collision decay. However, one has to be aware that during the passage of the ions through the collision cell, multiple collisions with gas molecules and therefore multiple fragmentation events may take place. Due to the fact that the ions are accelerated using a potential of several keV, the CID process in MALDI-MS instruments is always a high energy process. This has significant consequences on the appearance of the MS/MS fragment ion spectrum. In classical CID on QqQ, QToF or ion trap type instruments will usually observe y- or b-ions and their respective neutral losses ($H_2O$, $NH_3$). When looking at high energy CID spectra, additional ions are observed which originate from side chain fragmentations.

Quadrupole instruments can be programmed for a number of different scan modes owing to the presence of two independent quadrupole sections. The first scanning mode, used for precursor ion fragmentation, uses the selection of ions with a chosen m/z ratio in the first quadrupole. The selected ions undergo gas phase collisions in the central quadrupole collision cell. The fragments and reaction products are analyzed by the third quadrupole; this is termed the product ion scan. The second scanning mode consists of focusing the third quadrupole on a selected ion while scanning the masses using the first quadrupole. All of the ions that produce the ion with the selected mass through fragmentation are thus detected. This method is referred to as a precursor scan because the precursor ions are detected. In the third scan mode, both quadrupoles are scanned together with a constant mass offset between the two. This allows the detection of ions which have yielded a fragment ion with a specific mass shift from the precursor and is called neutral loss scanning.

A tandem MS experiment can be conceived in two ways; in space by the coupling of two physically distinct instruments (e.g. hybrid Q-ToF), or in time by performing a sequence of events in an ion storage device (e.g. ion trap, FT-ICR). Modern mass spectrometers often combine ionization sources and mass analyzers to allow optimization of the type of mass analysis to be carried out and minimize the limitations of single analyzers. Tandem MS instruments such as the triple quadrupole, ion-trap and hybrid quadrupole-time-of-flight (Q-ToF) are routinely applied in LC MS/MS or nanospray experiments with ESI to generate peptide fragment ion spectra suitable for protein identification by database searching. Mass accuracy and sensitivity for database searching and *de novo* sequencing have also been recently improved by coupling a MALDI

ionization source to a hybrid QToF instrument (*54, 55*). This configuration provides opportunities for automation and high-throughput applications and the possibility to archive samples for future re-analysis.

**1) Quadrupole-Time-of-Flight (Q-ToF)**

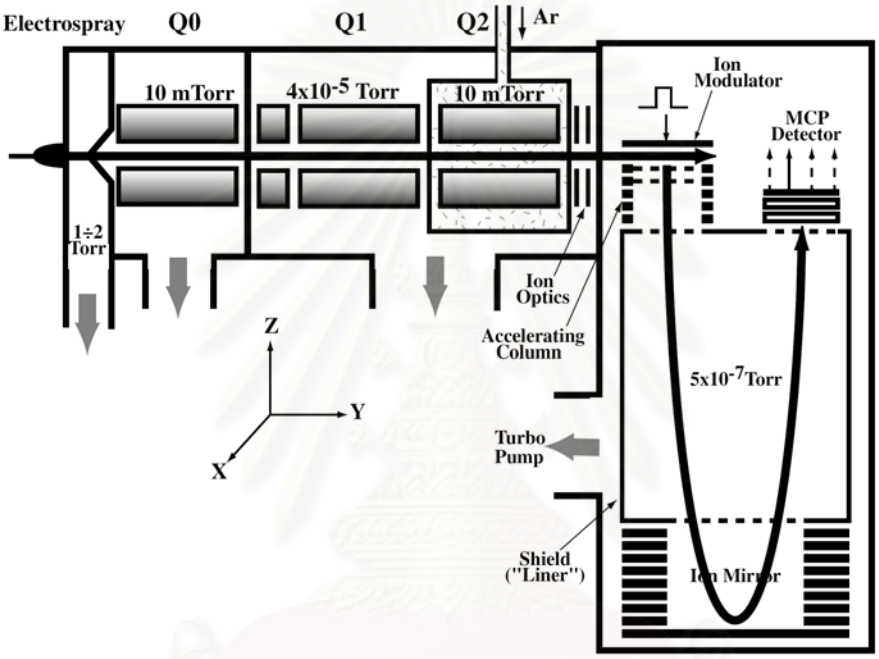



**Figure 2.10** Schematic diagram of the tandem Q-ToF mass spectrometer. (Reproduced from (*56*))

Historically, the development of the Q-ToF followed closely after the development of the so-called ESI-ToF technique The configuration can be regarded either as the addition of a mass-resolving quadrupole and collision cell to an ESI-TOF, or as the replacement of the third quadrupole (Q3) in a triple quadrupole by a TOF mass spectrometer. From either viewpoint, the benefits that accrue are high sensitivity, mass resolution and mass accuracy of the resulting tandem mass spectrometer in both precursor (MS) and product ion (MS/MS) modes (*57).*

The instrument consists of three quadrupoles, Q0, Q1 and Q2, followed by a reflecting TOF mass analyzer with orthogonal injection of ions (Figure 2.10). For single MS, the mass filter Q1 is operated in the r.f.-only mode so that it serves as a transmission element, collision energy is kept below 10 eV to avoid fragmentation in Q2, while the ToF analyzer is used to record spectra. Ions are sampled from a high-pressure electrospray or APCI ion source through an r.f. ion guide Q0 into Q1. This additional quadrupole Q0 is used for collisional cooling and focusing of the ions entering the instrument. For MS/MS, Q1 is operated in the mass filter mode to transmit only the precursor ion of interest. The ion is then accelerated to an energy of between 20 and 200 eV before it enters the collision cell Q2, where it undergoes collision induced dissociation (CID) with neutral gas molecules (usually argon or nitrogen). The resulting fragment ions are collisionally cooled and re-accelerated to the required energy and focused into a parallel beam that continuously enters to the ToF analyzer.

**2) MALDI-ToF/ToF**

MALDI-ToF/ToF instruments are characterized by a co-linear arrangement of two ToF mass analyzers (Figure 2.11), each equipped with an ion source that allows acceleration and focusing of the ions. In TOF1 the analyte ions are accelerated, precursor ions are selected and fragmented. Fragment ions are allowed to proceed to the “source” of ToF2, where they are accelerated and mass analyzed (Figure 2.11). High-energy collisions to fragment peptides in a ToF/ToF mass spectrometer are mainly approach to overcome disadvantages of PSD (*58*). This technique can produce fragment ion spectra that show the backbone ions a-, b-, y- and i-, internal ions and the side chain d- and w-ions. This technique can also analyze the same sample “stored” on a target for two types of analyzes without the need for further sample preparation.

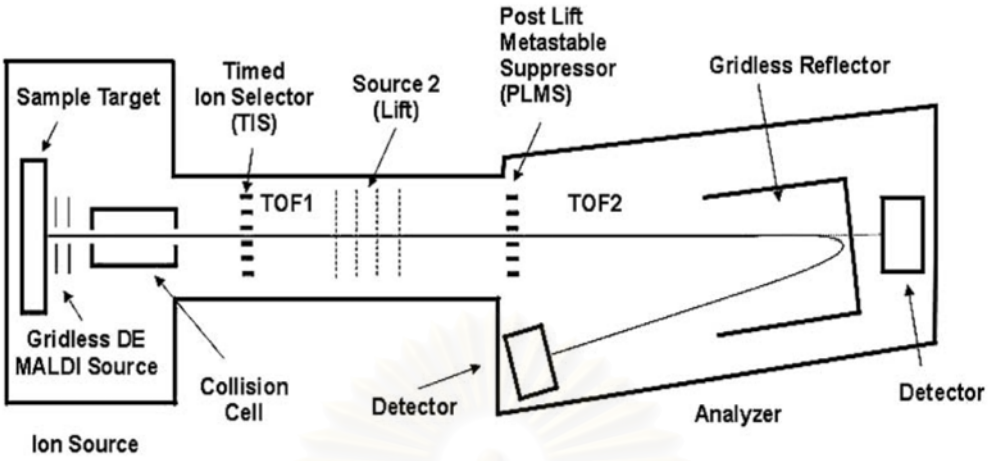



**Figure 2.11** Schematic diagram of the LIFT-TOF/TOF mass spectrometer (Ultraflex ToF/ToF, Bruker). ToF1 ranges from the MALDI ion source to the LIFT cell, ToF2 from the second accelerator stage in the LIFT cell to the reflector. (Reproduced from (*59*))

## 2.10 Protein identification

Information from recently derived genomic databases can now be used to predict protein sequences and therefore sequencing of entire proteins is no longer necessary for confident identification of proteins. The combination of genome data with mass spectrometric approaches permits the identification of proteins by matching experimentally measured properties such as peptide masses with those expected from the databanks. Underlying this approach are the biological databases of DNA and protein information, which are constantly evolving. These databases are collated, stored and managed by dedicated laboratories and can be accessed via the Internet (e.g. National Centre for Biotechnology Information (NCBI)).

### 2.10.1 Proteolysis

A key component to mass mapping using MS is the cleavage of proteins into predictable fragments. Endoproteinases are now used to exploit their site specific cleavage-properties which can provide a subset of peptides for analysis by MS for protein

identification. The most commonly used protease in modern proteomics is trypsin, which cleaves proteins at the C-terminal side of arginine and lysine residues (*51*). Each protein will have a distinct set of peptide masses, termed a peptide mass fingerprint. The conditions for proteolysis are dependent on the type of experiment and test sample and can be performed on pre-stained spots excised from a gel or in a mixture of soluble of proteins.

### 2.10.2 Peptide mass mapping

Peptide mass mapping is a method used to identify proteins contained within a sequence database using an algorithm to match a set of peptide-masses generated from protein specific enzymatic cleavage. The probability of success depends on the purity of the sample, the number of fragments and the resolution of mass accuracy used to acquire the data. The accuracy obtained in the measurement of peptide mass strongly influences the specificity of the search (*8*). When high mass accuracy (10 to 50 ppm) is achieved, as a rule at least five peptide masses need to be matched to the protein and 15% of the protein sequence needs to be covered for an unambiguous identification. Peptide mass fingerprinting using MALDI has been successful in many different settings including yeasts (*60*), human cell lines (*61*). However, in general, peptide mass fingerprinting is used for the rapid identification of a single protein component and it is not feasible to apply this technique to complex mixtures.

### 2.10.3 Protein identification by tandem MS

Over the last few decades, two different approaches for Tandem MS-based protein identification have been developed. The top down approach starts with an intact protein and cleaves the protein in the gas phase rather than in solution. The protein is fragmented inside the mass spectrometer to create a ‘ladder’ of ions indicative of the sequence. This approach relies on very high resolution instruments such as FT-ICR.

The bottom-up approach identifies proteins by tandem MS analysis of peptides derived from the digestion of mixtures of intact proteins. The fragmentation pattern obtained from each peptide is indicative of the amino acid sequence of the peptide. This pattern can then be used to search sequence databases. There are several advantages to

using tandem MS data to identify proteins. There is a higher level of certainty in the identification of proteins since the method relies on several pieces of highly specific information, it is also easier to identify sites of post-translational modification and proteins can be identified from mixtures (*62*).

The mechanism of peptide fragmentation depends on the ionization method used to generate the ions, the method of coupling the ion source to the MS, and the charge state and amino acid composition of the peptide ions. ESI most commonly produces doubly charged ions; the two protons on a typical peptide ion allow even fragmentations along the peptide backbone, producing a spectrum composed of predominantly N- and C-terminus ions. The nomenclature currently used to identify these ions was modified by Biemann (*63*); fragments containing the N-terminus of the peptide are denoted as *a*, *b* and *c* and the C-terminal containing fragments as *x*, *y* and *z* (Figure 2.12). A product ion spectrum containing a representative *b* and *y* ions series can be interpreted manually or using database searching. The mass shift between adjacent *b* and *y* ions along the peptide backbone corresponds to the molecular weight of the amino acid components. MALDI ionization typically produces singly-charged ions which fragment badly to give some *b* and *y* ions but also *a* ions and other internal fragments (*64)*. The proposed mechanism of formation of b and y-type ions was shown in Figure 2.13.

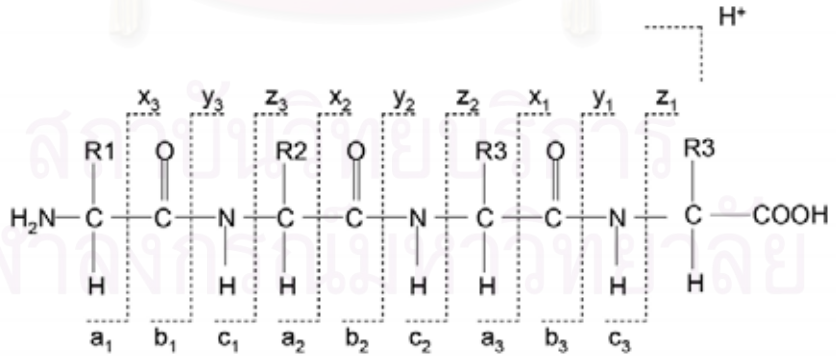



**Figure 2.12** Peptide fragment ion formation. Fragments will only be detected if they carry at least one charge. If this charge is retained on the N terminal fragment, the ion is classed as either *a*, *b* or *c*. If the charge is retained on the C terminal, the ion type is either *x*, *y* or *z*. A subscript indicates the number of residues in the fragment.

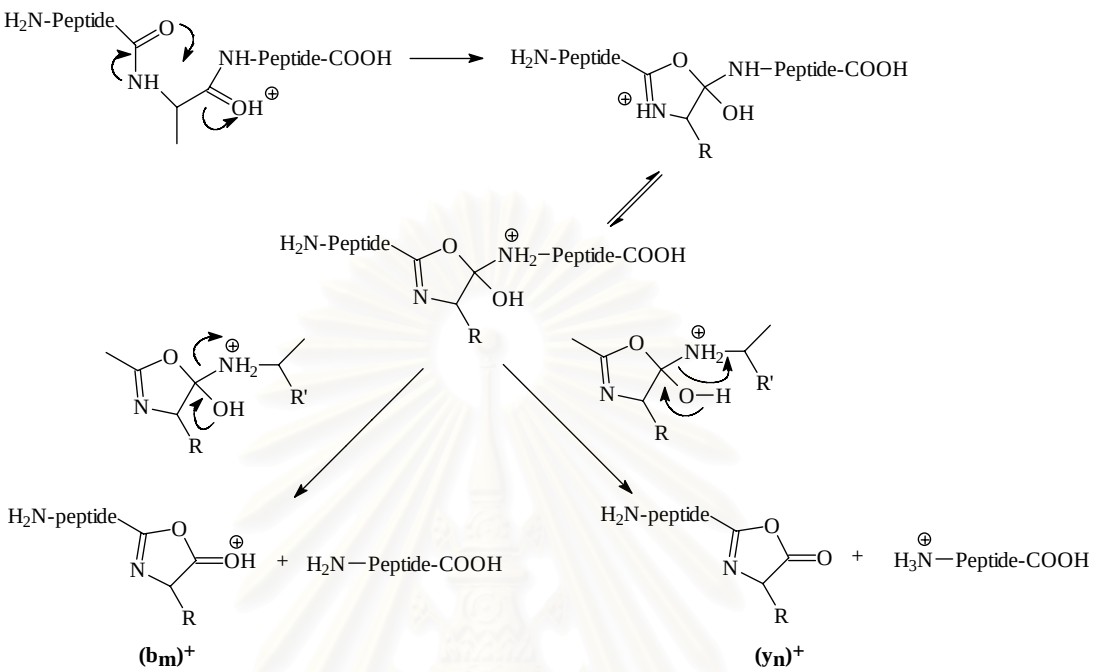



**Figure 2.13** The proposed mechanism of formation of b and y-type ions (*65*).

Other types of ions are also present that are derived solely from cleavage of the peptide backbone: internal fragment ions and immonium ions. Immonium ions have the structure $H_2N=CHR^+$, derived from individual amino acid residues are labelled using the single letter code for the amino acid concerned.

### 2.10.4 Database searching

Most database searching programs allow the user to adjust the parameters of any given search; these will depend on the type of data, instrumentation and desired stringency of the search. For example, the measured error in ppm will vary between types of instruments, increasing the error margin in a search may increase the number significant results. In addition, denatured proteins exposed to trypsin often undergo imperfect proteolysis. Therefore allowing for ‘missed cleavage’ in the search will increase the size of the database searched and affect the significance of the results. A further consideration is that of post-translational modifications, either inherent (e.g.

acetylation) or introduced during the preparation process (oxidation of methionine residues). These can usually be inputted into the search algorithm as fixed or variable modifications. A list of peptides with good correlation to the theoretical fragmentation pattern is returned to the user with the top-scoring peptide being considered the best candidate. These database-searching algorithms can also support post-translational modifications of specific amino acids. For example, the database search can differentially include the change in mass associated with a modification to specific amino acid residues encountered in the database. A number of search engines exist to facilitate protein identification from MS data, these are listed in below.

**Peptide mass mapping database matching tools**

Mascot http://www.matrixscience.com
Mowse http://www.hgmp.mrc.ac.uk/Bioinformatics/Webapp/mows
MS-Fit http://wwww.prospector.ucsf.edu
Pepsea http://www.pepsea.protana.com
ProFound http://www.proteometrics.com

**Tandem mass spectrum database matching tools**

Mascot http://www.matrixscience.com
MS-Tag http://www.prospector.ucsf.edu
Pepsea http://www.pepsea.protana.com
SEQUEST http://www.fields.scripps.edu/sequest
Sonar http://www.proteometrics.com

### 2.10.5 *De novo* peptide sequencing and similarity search

A database-searching approach is only good for the identification of peptides present in a protein database. When an appropriate database is not available, *de novo* sequencing is the only way to identify the peptide. Besides the ability of identifying peptides without a database, *de novo* sequencing also has the advantages that: (i) *de novo* sequencing results can be used for homology-based database searches to identify peptide homologues and modifications, and (ii) *de novo* sequencing results can also be used to

validate the database search results. Significant similarity between the database search result and the *de novo* sequencing result could be taken as evidence that the database-derived sequence is correct (*66*).

Sequence similarity search methods represent an attractive alternative to cloning because the identification of unknown proteins can be achieved without further “wet” biochemistry experiments. However, mass spectrometry and sequence similarity searches are difficult to combine. Conventional database search algorithms like BLAST (*67*) or FASTA (*68*) are optimized for accurate sequence queries that are longer than 35 amino acid residues. Usually peptide sequences obtained by tandem mass spectrometry do not exceed the length of a tryptic peptide, typically comprising 10–15 amino acid residues, and therefore the statistical significance of retrieved hits is often ambiguous. Recently, several database searching approaches were reported that accommodate specific requirements of tandem mass spectrometric sequencing (*69-71*). Shevchenko *et al.* developed a MS BLAST (*70*) that does not allow gaps within individual peptides, while gaps between peptides are not penalized and can be of arbitrary length. Therefore all peptide sequences obtained by the interpretation of acquired tandem mass spectra are assembled into a single searching string in arbitrary order. The basic principle of MS-BLAST is the software searches for sequence stretches in the database which are homologous to entered short sequences. For normal BLAST searches, need long sequence stretches to get meaningful results. MS-BLAST overcomes this problem by using multiple sequences.

## 2.11 Analysis of complex mixtures

Proteomic analysis of complex mixtures presents some specialized analytical problems with the dynamic range (relative protein abundance) and the diversity of protein expression. The dynamic range problem is most commonly tackled by either increased separation power or prefractionation to enrich for lower abundance proteins.

### 2.11.1 Gel electrophoresis

A combination of two-dimensional gel electrophoresis (2DE), using isoelectric focusing (IEF)/SDS-PAGE gel for the separation, detection and quantification of individual proteins with mass spectrometry (MS) and sequence database searching for the identification of the separated proteins has been used to identify and catalogue large numbers of proteins in complex samples (Figure 2.14). MALDI is the most attractive MS method to identify proteins from 2DE because of the potential for highthroughput analysis.

Shortly after it was discovered that proteins had unique molecular weights it was demonstrated that they could also be separated on the basis of electrophoretic mobility.

Modern day polyacrylamide gel electrophoresis has evolved over a series of one and two dimensional systems to produce increased resolution of protein spots in complex biological samples. Even under optimal conditions, any of the current 1-DE methods is only capable of resolving about 100 distinct zones (*72*). The coupling of isoelectric focusing (IEF) with polyacrylamide gel electrophoresis in the presence of the anionic detergent sodium dodecyl sulphate (SDS-PAGE) in the second dimension resulted in a 2D method that separated proteins according to two independent parameters, charge and size.

Proteins are visualized by staining the gel with hydrophobic dyes such as Coomassie, silver ions or fluorophores. The stains adhere principally to the proteins, fixing their position in the gel and allowing the determination of p*I* and molecular weight. The intensity of the staining also provides a semi-quantitative measure of the intensity of the protein.

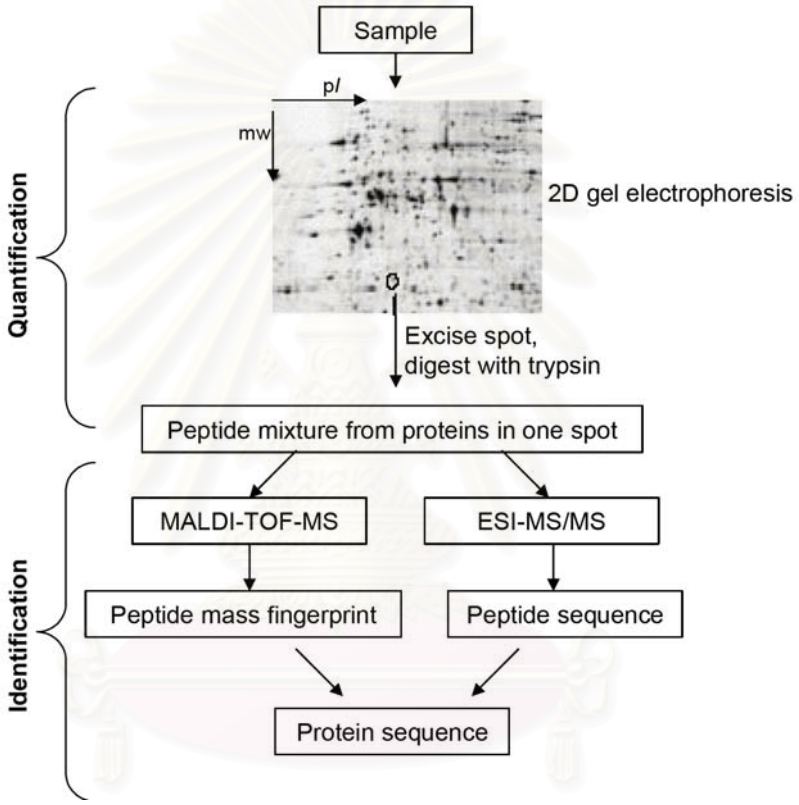



**Figure 2.14** Example of a workflow for protein identification using a combination of 2DE and mass spectrometry (Reproduced from (*73*)).

Although 2D gel electrophoresis (2DE) revolutionized protein separation techniques there are a number of inherent problems with this technique. Some of issues related to reproducibility, poor representation of low abundant proteins, highly acidic/basic proteins, or proteins with extreme size or hydrophobicity, as well as difficulties in automation of the gel-based techniques.

### 2.11.2 Liquid chromatography-mass spectrometry (LC-MS)

When mixtures are extremely complex, on-line RP-HPLC is used to concentrate and separate the peptides before sequencing by tandem MS. In a typical LC-MS/MS experiment, the analyte is eluted from a reversed-phase column, which separates the peptides by hydrophobicity, and is ionized and transferred with high efficiency into the mass spectrometer for analysis (Figure 2.15).

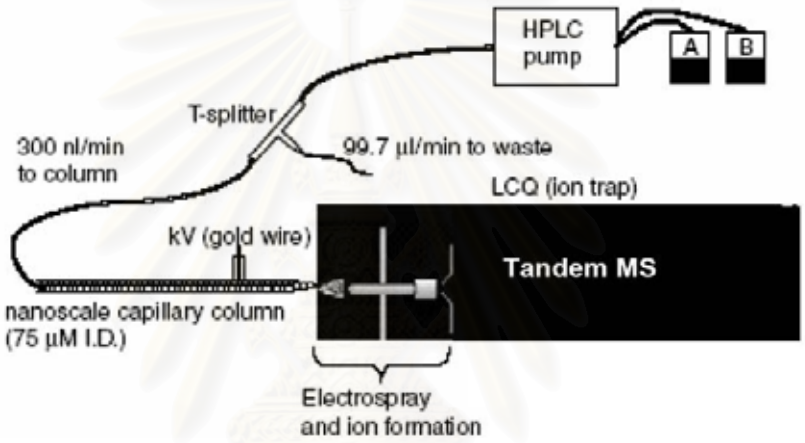



**Figure 2.15** Schematic of an on-line-LC/MS.

An on-line capillary LC-MS/MS system consists of conventional HPLC pumps, transfer tubing, a pre-column flow splitter, a liquid junction, a reversed-phase microcapillary column and a tandem mass spectrometer. The pumps and mass spectrometer are controlled by the same software to allow coupling between chromatography and ion detection. The flow-rate at which conventional pumps operate (~100 µl/ min) and the optimum flow-rate for the microcapillary column (~300nl/min) are vastly different. A flow restrictor (Tsplitter) permits the gradient to form quickly and be distributed to the microcapillary column at acceptable flow-rates. A conductive liquid junction (e.g. gold wire) to a high voltage source (1–2 kV) is needed to promote electrospray. Although numerous methods for coupling liquid chromatography to MS have been explored, electrospray ionization has transformed LC-MS/MS into a routine laboratory procedure sensitive enough to analyze peptides and proteins at low

concentrations. Currently, detection limits of a few femtomoles of peptide material loaded on the column make this technique compatible with silver-stained, fluorescently labelled, or faintly stained Coomassie gel bands and capable of detecting proteins and peptides present at a low copy number per cell. In addition, the high resolution chromatography step makes this technique attractive for the analysis of peptides derived from complex mixtures which are reduced, denatured and alkylated prior to proteolytic digestion. The resulting peptide mixture is then lyophilized and resuspended in the LC buffer for analysis. A sample can be loaded and eluted on the capillary column in different ways depending on its peptide concentration and volume. Commonly, the sample is loaded on to a pre-column trap at µl flow-rates for concentration and fast desalting and then eluted to the reversed-phase column for separation. The sequence of events in a typical LC-ESI-MS/MS experiment is shown in Figure 2.16.

High efficiency RP-LC-MS/MS is capable of a protein identification dynamic range of approximately four orders of magnitude (*74*) and substances expressed at levels of 5 µg/ml have been detected using this technique (*75*). A number of investigators have increased the protein identification dynamic range by combining strong cation exchange (SCX) chromatography with high efficiency RP-LC. This multidimensional chromatography can be either on-line (directly coupled to the RP-LC), or offline (fractions collected separately). On-line techniques such as Multidimensional

Multidimensional Protein Identification Technology (MudPIT), have been used to improve coverage of the yeast proteome (*76*). By using a specially packed biphasic column with a SCX support as the initial phase and a RP as the second phase, peptides can be stepped from the SCX onto the RP phase in a series of salt steps that increase in concentration. Since each phase separates peptides by orthogonal chemical properties, a high resolution separation is achieved.

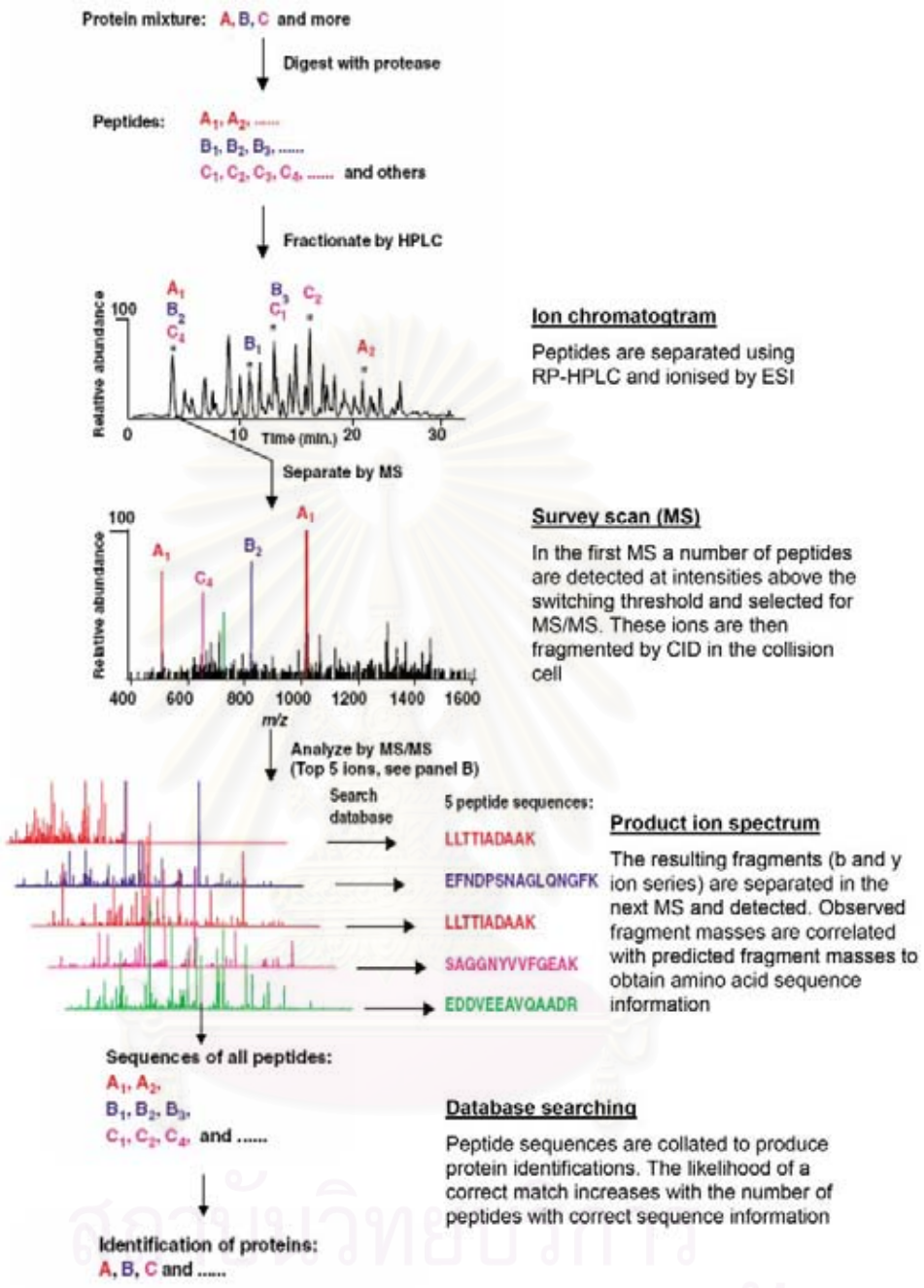



**Figure 2.16** Direct analysis of complex peptide mixtures by on-line reversed-phase liquid chromatography tandem mass spectrometry (LC-MS/MS). During automated peptide sequencing by LC-MS/MS analysis, five MS/MS scans representing the sequence analysis of five different peptide ions occur in 11 seconds with no user input. During a 30 min analysis, more than 800 sequencing attempts (tandem mass spectra) are acquired.

**2.12 Quantitative proteome analysis**

2DE-based proteome analysis provides information about protein abundance at the gel level by comparing staining intensities. However, when peptide mixtures are analyzed directly by LC-MS/MS techniques, the original quantitative information is lost. Recently, quantitative proteome analysis techniques have been introduced incorporating stable isotope labeling. Stable isotope dilution involves the addition to the sample of a chemically identical form of the analyte(s) containing stable heavy isotopes (e.g. 2H, 13C, 15N, etc.) as an internal standard. Ionization efficiencies are highly variable for peptides and therefore the best-suited internal standard for a candidate peptide is the same peptide labelled with stable isotopes. Differential protein expression is then assessed by analysis of the two protein mixtures, where one serves as the reference sample. This sample contains the same proteins as the other sample but at different abundances and labelled with stable isotopes. In theory, all the peptides present in the mixture will exist as analyte pairs of identical sequence, but with differing masses. The peptide pairs have the same physicochemical properties and behave similarly under any conceivable isolation or separation step. Thus, the ratio between the intensities of the lower and upper mass components of these pairs of peaks provides an accurate measure of the relative abundance of the peptides (and hence the protein) in the original mixture (*77*). An example of this technique is the isotope coded affinity tag (ICAT) strategy (*78*). Proteins from two mixtures are labelled with either a heavy or a light reagent and then mixed together. Following tryptic digestion, the tagged peptides can be enriched using biotin-affinity chromatography and subsequently separated and analyzed by LC-MS/MS. Peptide masses derived from the two differentially-labelled samples are separated by 9 Da due to the mass difference between the heavy and light tags. The relative abundance of a protein in the two samples can thus be determined by comparing the intensities of the ICAT light- and heavy-labelled versions of each peptide. More recently, an alternative strategy has been developed called isobaric tags for relative and absolute quantification (iTRAQ) (*79*).

## 2.13 iTRAQ reagent

iTRAQ reagents are a multiplexed set of four isobaric reagents (Figure 2.17) which specifically react with primary amine groups. The isobaric nature of the reagents gives them identical m/z values in single MS mode, but in tandem MS generates strong, diagnostic, low mass MS/MS signature ions. These reporter ions allow quantification of up to four different samples simultaneously. The reporter ions have m/z values of 114, 115, 116 and 117 Th and as such are visible in a portion of the product ion spectrum which typically contains few other product ions. Differentially labelled samples from different patient groups or different experimental conditions are mixed and analyzed using LC-MS/MS. Assuming identical digestion of the two samples, relative differences in intensities of these reporter ions are used to determine differences in protein expression.

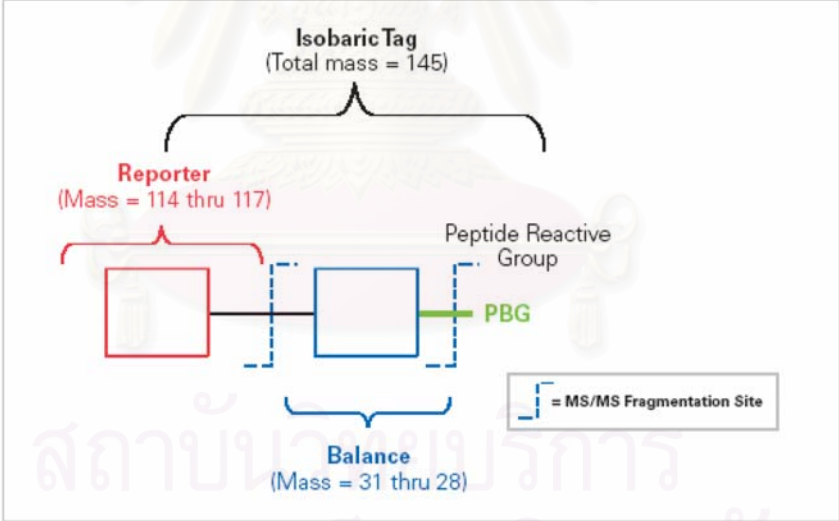



**Figure 2.17** Schematic of the iTRAQ reagent**.** iTRAQ consist of a reporter group, a balance group and a peptide reactive group (Adapted from Applied Biosystems iTRAQ™ Reagents, Chemistry Reference Guide, www.appliedbiosystems.com)

The resultant product ion spectra can be searched using a web-based program such as Mascot to identify proteins within the mixture. Changes in protein expression between samples can therefore be assessed using relative quantification of several peptides from the same protein, which improves the accuracy of the quantification results. Quantification using iTRAQ has been shown to be highly reproducible (*80*).

Although conceptually similar to ICAT, the two techniques differ in a number of areas. iTRAQ reagents bind to primary amine groups on tryptic peptides and therefore will tag most, if not all peptides. In contrast, the ICAT method relies on the tagging of cysteine residues which creates dependence on relatively non-abundant cysteine containing peptides. Other features of the iTRAQ technique are that relative quantification is performed via MS/MS and that there are four possible tags, which allows multiplexing of up to four samples within a given experiment.

Quantification is performed via the differences in abundance of the product ions that are cleaved from one of the four possible tags (Figure 2.18). The tags have an identical mass, a result of differences in the balance region of the isobaric tag. The importance of this feature is that an identical peptide in the four samples will have an identical mass and LC retention time after tagging. This strategy simplifies the analysis and has the potential to increase analytical accuracy and precision.

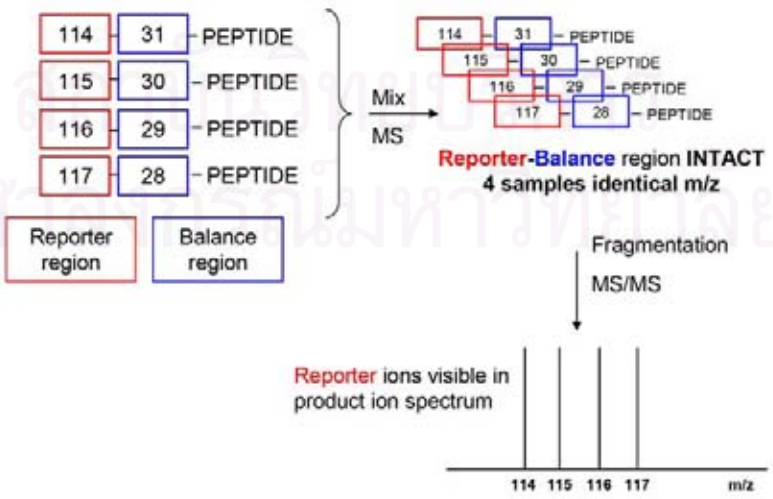



**Figure 2.18** iTRAQ reagent structures.

# CHAPTER III

# EXPERIMENTAL

## 3.1 Plant materials

Fresh rhizomes of *Curcuma zedoaria* and *Curcuma aromatica* were purchased from local market in Bangkok, Thailand. They were stored at room temperature until use.

*Arabidopsis thaliana* guard cell extracts were received from Dr. Julie E Gray (Molecular Biology, University of Sheffield, UK). *Arabidopsis thaliana* were grown under fluorescent light with a 16 h photoperiod, 25° C day/22° C night. 8 week old plants were sprayed with ABA and MeOH for ABA-treated and untreated samples, respectively, and then incubated for 1 h. Epidermal Fragments of guard cell enrichment were extracted by the blender method (*81*). Protein was extracted from epidermal fragments as described in (*82*) using YeastBuster$^{TM}$ reagent.

## 3.2 Chemicals

Materials for protein purification, Q-Sepharose, DEAE-cellulose, SP-Sepharose and Superdex 75 were purchased from GE Healthcare (Uppsala, Sweden). All chemicals, protein and peptide standards were obtained from Sigma, formic acid (for Mass spectrometry) was from Fluka, Trifluoroacetic acid (TFA) (spectrophotometric grade) was from Aldrich. HPLC-grade water, acetonitrile (ACN) and methanol were purchased from Sigma-Aldrich.

### 3.3 Instrumentation

#### 3.3.1 ÄKTA prime system

All mobile phases and samples were filtered through a 0.45 μm filter (Millipore, MA, USA) prior to use.

**Figure 3.1** ÄKTAprime™ plus (GE Healthcare), Use for simple purification of Proteins (Figure from www1.amershambiosciences.com)

Protein extracts of *Curcuma zedoaria* Rosc. and *Curcuma* aromatica Salisb. were purified by liquid chromatography with low pressure system. They were including anion exchange, cation exchange and size exclusion.

**1) Anion exchange chromatography**

Q-Sepharose Fast Flow and DEAE-Cellulose were packed manually in XK 16/20 column (GE Healthcare) according to manufacturer's instruction to give 1.6 ×15 cm column size. The columns were then used on an ÄKTA prime system for purification of protein at a flow rate of 1 ml/min. Mobile phase A was 20 mM phosphate buffer, pH 7.0.

Mobile phase B was as A with an additional 0.35 M NaCl. Large sample volumes (>5 ml) were applied through the system pump. Columns temperature was maintained at 4°C while running sample. The eluate was monitored at 280 nm and 8 ml fractions were collected.

**2) Cation exchange chromatography**

SP-Sepharose Fast Flow was used as media. Mobile phase A was 20 mM Sodium acetate buffer, pH 5.0. Mobile phase B was as A with an additional 0.5 M NaCl. All other processes were as above.

**3) Size exclusion chromatography**

Superdex 75 was packed manually in XK 16/70 column (GE Healthcare) according to manufaturer's instruction to give 1.6 ×56 cm column size. Sample (1-2 ml) was applied through the sample loop by manual filling with a syringe. 3 ml fractions were collected. All other processes were as above.

**3.3.2 High performance liquid chromatography (HPLC)**

All mobile phases and samples were filtered (0.45 µm nylon filter; Millipore, MA, USA) prior to use.

**1) Reversed phase (RP)**

The last step of purification of *Curcuma zedoaria* sample was purified by RP-HPLC. A 250 x 4.6 mm $C_8$ BDS column (5 µm particle size) was used on an HP1100 LC system (Agilent) at a flow rate of 1 ml/min. Mobile phase A was 0.01% (v/v) TFA/water; mobile phase B was 0.01% (v/v) TFA/ACN. Gradient condition was described in chapter IV.

**2) Strong cation exchange (SCX)**

iTRAQ labelled peptides of Arabidopsis guard cell extracts were fractionation using SCX. A 200 x 4.6 mm polysulfoethyl aspartamide column (5 µm particle size, 30 nm pore size) was used on an HP1100 LC system (Agilent) at a flow rate of 1 ml/min. Mobile phase A was 5 mM $KH_3PO_4$, 25% acetonitrile, pH 3. Mobile phase B was as A with an additional 250 mM KCl.

### 3.3.3 MALDI-ToF

MALDI experiments were performed in the positive reflectron mode, with delayed extraction using a Voyager DE STR (PE Biosystems, Framingham, USA) fitted with a 337 nm $N_2$ laser. The operating pressures were typically around $7x10^{-8}$ torr (source chamber) and $7x10^{-9}$ torr (analyzer). The accelerating voltage was set to 20 kV, the reflectron grid voltage was 75-76% of the acceleration voltage and the guide wire voltage was 0.001-0.002%. The extraction delay time was 260-325 ns and the low mass gate was set at 400 m/z. Data manipulation was performed using Data Explorer software (PE Biosystems, Framingham, USA). External calibration of the ToF analyzer was performed prior to sample analysis using the monoisotopic m/z ratios of the peptide standards: des-$Arg^1$ bradykinin, angiotensin I and II, neurotensin, adrenocorticotropic hormone fragment 18-39 and oxidized insulin chain B, with m/z ratios of 904.468, 1046.542, 1296.685, 1672.918, 2318.283 and 3494.651, respectively.

Sample (1 µl) was mixed 1:1 with 10 mg/ml α-cyano-4-hydroxycinnamic acid (dissolved in 50% acetonitrile, 0.1% TFA). 1 µl of this sample:matrix mixture was spotted onto a stainless steel target and allowed to dry. Spectra were obtained by summing 50 laser shots. In order to determine the molecular mass of the protein, 0.5-1 µL of protein was spotted onto the target plate, followed by an equal volume of saturated sinapinic acid in 0.1% (v/v) TFA, 50% (v/v) ACN.

### 3.3.4 MALDI-ToF/ToF

LIFT mass spectra were acquired on a Bruker Ultraflex II ToF/ToF mass spectrometer operated in the positive ion mode. Metastable fragmentation was induced by a nitrogen laser (337 nm) without the further use of collision gas. Precursor ions were accelerated to 8 kV and selected in a timed ion gate. In the LIFT-cell the fragments were further accelerated to 19 kV. The reflector potential was 29 kV.

### 3.3.5 LC-ESI-q-TOF-MS/MS

ESI-MS was carried out on a Q-TOFI (Figure 2.2) instrument (Micromass, Manchester, UK). Mass spectrometric data were acquired either manually or using automated protocols within version 3.4 of MassLynx™ (Waters) operating software. ToF mass analyzer was calibrated at the start of each session using the product ions formed from the collisional activation of [$Glu^1$]-fibrinopeptide B (Sigma; F3261) [$M+2H^+$] ions. (monoisotopic m/z ratios: 72.081, 246.157, 333.189, 480.257, 684.347, 813.389, 942.432, 1056.475, 1171.502, 1285.545). The calibrant (1.27 fmol/µl in 50% acetonitrile/0.1% formic acid) was pressure-driven into the spray capillary (fused silica capillary, 20 µm i.d., 90 µm o.d.) using a syringe driver (Harvard Apparatus, MA, USA) at a flow rate of 0.5µl/min. A nitrogen gas stream aided desolvation of ions. The spray tip voltage was 3.5 kV. Manual data acquisition was carried out over a mass range of 400-1600 Da, with a scan time of 1s and typical cone voltage values of 48-52 volts.

HPLC peptide separation was performed on an UltiMate HPLC system (LC Packings, Netherlands) coupled to the Q-TOF instrument. Mobile phase A was 2% acetonitrile, 0.06% formic acid, mobile phase B was 95% acetonitrile, 0.05% formic acid. Samples were loaded using a Famos autosampler and desalted using a C18 precolumn (µ-Precolumn Cartridge PepMap C18; 5 µm particle size, 10 nm pore size, 300 µm x 5 cm; LC Packings) at 100% mobile phase A (30 µl/min), diverting the eluent to waste before switching to the analytical column (C18 PepMap, 75µm x 15cm, 10 nm pore size; LC Packings) into line. The flow rate through the analytical column was 0.18 µl/min. Peptides were separated using a reversed phase (RP) linear gradient of 0-60% mobile phase B over 60 minutes.

For LC MS sample analysis, sample was sprayed from a distal-coated fused-silica PicoTip emitter needle (10 µm i.d. tip; New Objective, MA, USA), with a spray tip voltage ranging from 1800-2800 V. Automated selection of peaks was carried out using the automatic function-switching mode in MassLynx, which enables automated instrument operation and selection of suitable ions for collision-induced dissociation (CID) according to their peak intensity and isotope distribution. Peaks with an intensity >10 counts/s and a charge-state of 2+, 3+ or 4+ triggered switching from MS to MS/MS mode. Switching back from MS/MS to MS occurred after 6 s or by a decrease in fragment peak intensity to <2 counts/s. Automatic generation of peak lists for database searching of MS/MS data was carried out using the ProteinLynx™ module within MassLynx software.

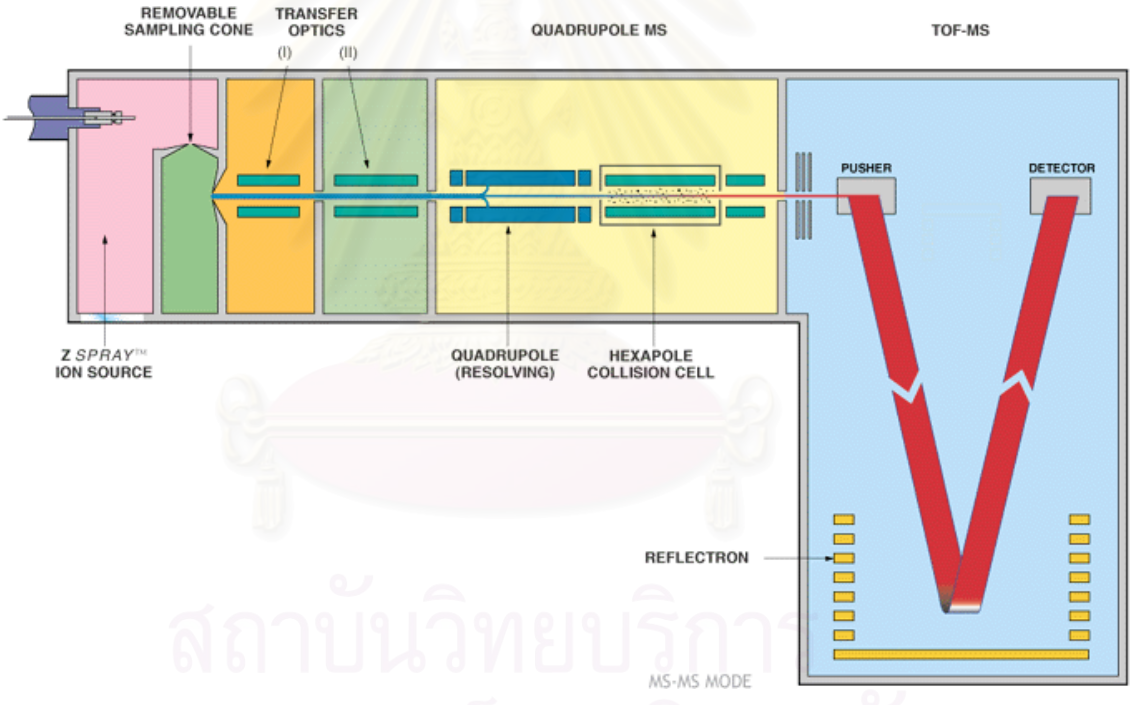



**Figure 3.2** The electrospray QToF mass spectrometer, operating in protduct ion mode. (Figure provided by Micromass, UK)

## 3.4 Methods

### 3.4.1 Protein extraction and $(NH_4)_2SO_4$ precipitation

Fresh rhizomes (1.5 kg) of *Curcuma zedoaria* and *Curcuma aromatica* were washed, cut into small pieces and homogenized in a blender with 50 mM phosphate buffer, pH 7.0, containing 0.1 M NaCl and 1.0 mM EDTA (3 ml/g). The mixture was stirred for 1 h and filtered through cheesecloth. NaCl was added to 2% (w/v), stirred for 2 h and the suspension was centrifuged at 15,000 g at 4°c for 30 min. The supernatant was brought to 20% saturation with ammonium sulphate, stirred for 1 h and then centrifuged at 15,000 g for 30 min. The supernatant was retained and ammonium sulphate was added to 60% saturation, adjusted to pH 7.2 with NaOH and stirred for 3 h at 4°c. The resulting suspension was centrifuged, the supernatant discarded, and the precipitate resuspended in 20 mM phosphate buffer, pH 7.0. The solution was then dialyzed over 24 h (using dialysis tubing with molecular weight cut off 3,500 Da) against the same buffer containing 1 mM EDTA. The dialysate was centrifuged at 12,000 g for 30 min at 4°c to remove any insoluble material and was loaded onto chromatography column.

### 3.4.2 SDS polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE was performed using standard methods on the Bio-Rad Mini-Protean II system. It is a discontinuous system with 12% separating gel (pH 8.8) and 4% stacking gel (pH 6.8) with a mini size (7 cm × 10 cm × 1cm). Protein samples were mixed 1:1 with sample buffer [1.4 mL of distilled water, 2.0 mL of 0.5 M Tris-HCl, pH 6.8, 2.0 mL of 10% (w/v) SDS, 2.0 mL of glycerol, and 0.4 mL of 0.05% (w/v) bromophenol blue]. Reducing conditions were obtained by adding dithiothreitol (DTT) to a final concentration of 10 mM and heating the samples for 5 min at 100 °C. To each lane was applied 15 µl of protein sample, and gels were run at at 10 mA/gel.

For the molecular weight markers, SigmaMarker$^{TM}$ wide molecular weight range (Sigma, Saint Louis, USA) was used according to the instructions of the manufacturer.

**1) Coomassie blue staining**

Gels were washed with water (~100 ml; 3 x 5 min), stained with Brilliant Blue R concentrate (Sigma-Aldrich) for 30 min and destained in 50% (v/v) methanol, 5% (v/v) for 30 min or until bands appeared.

**2) Silver staining**

A silver-staining protocol compatible with mass spectrometry was followed (*83*). The gel was fixed in 40% ethanol, 10% acetic acid for 1 h, followed by washing with 30% ethanol (2 x 20 min) and water (20 min). The gel was sensitized in 0.02% $Na_2S_2O_3$ (w/v) for 1 min and then washed with water (3 x 20 s). The gel was incubated in chilled 0.1% $AgNO_3$ (w/v; 20 min, 4°C). The gel chamber was changed and the gel washed with water (1 min). The gel was developed in 3% $Na_2CO_3$ (w/v), 0.05% formaldehyde and then washed with water (20 s). Staining was terminated with 5% acetic acid, followed by washing with water (3 x 10 min).

**3.4.3 Tryptic digestion of proteins**

**1) In-solution digestion**

Proteins in ammonium bicarbonate (25-50 mM) were reduced using dithiothreitol (DTT; 10 mM) for 1 h at 56°C and then alkylated using iodoacetamide (55 mM) for 45 min at room temperature in the dark. The reaction was quenched with the addition of DTT (55 mM). Proteomics grade trypsin (T6567-20 µg; Sigma) was added at a ratio of 1:50 or 1:100 (weight trypsin:weight of protein) and hydrolysis allowed to occur at 37°C overnight (~18 h). The reaction was quenched by addition of 0.5% formic acid.

**2) In-gel digestion**

Bands were excised, divided into 3 (~1 $mm^3$ cubes) and destained. Coomassie bands were destained with 30% acetonitrile (15 min incubation with agitation) followed by 50% acetonitrile, 25 mM ammonium bicarbonate (15 min). This cycle was repeated if necessary until gel pieces were destained. Silver-stained bands were destained in 15 mM

$K_3Fe(CN)_6$, 50 mM $Na_2S_2O_3$, 25 mM ammonium bicarbonate. Destained gel pieces were dried (vacuum centrifugation; 5 min) and rehydrated in 10 mM DTT, 25 mM ammonium bicarbonate. Gel pieces were reduced at 56°C for 45 min, the liquid was removed and replaced with 55 mM iodoacetamide, 25 mM ammonium bicarbonate. Gel pieces were incubated at room temperature in the dark for 45 min and then washed with 25 mM ammonium bicarbonate (10 min, with agitation) and 50% acetonitrile, 25 mM ammonium bicarbonate (2 x 5 min, with agitation). Trypsin (T6567-20 µg; Sigma) was resuspended in 200 µl of water to give 0.1 µg/µl. Trypsin solution (0.1 µg/µl) was diluted in 25 mM ammonium bicarbonate to give a 10 ng/µl solution. Gel pieces were then dried and rehydrated in 10 ng/µl trypsin, 25 mM ammonium bicarbonate. Excess trypsin solution was removed after ~10 min, the gel pieces were covered with 25 mM ammonium bicarbonate and hydrolysis was allowed to occur overnight (~18 h) at 37 °C. Formic acid was added to ~0.5% and the supernatant was removed. Remaining peptides were extracted from the gel pieces using 50% acetonitrile with vigorous agitation and brief sonication. This was repeated, followed by a final neat acetonitrile extraction. All supernatants were pooled and the volume reduced by vacuum centrifugation.

### 3.4.4 Peptide desalting

Peptides were desalted using a reverse phase $C_{18}$ peptide trap (Peptide MacroTrap: bed volume 50 µl, sample capacity 200 µg; Michrom Bioresources). The trap was washed with 70% acetonitrile, 0.1% TFA (500 µl), conditioned with 0.1% TFA (500 µl). The sample was loaded and then washed with 0.1% TFA (500 µl). Elution was with 70% acetonitrile, 0.1% TFA (200 µl).

For desalting of small volumes of peptide solutions prior to MALDI mass spectrometry C18 ZipTips (Millipore) were employed. ZipTips were moistened with 50% acetonitrile (10 µl) and equilibrated with 0.1% trifluoroacetic acid (3 x 10 µl). Acidified sample was loaded onto the ZipTip, washed with 0.1% trifluoroacetic acid (3 x 10 µl) and eluted with 50% acetonitrile, 0.1% trifluoroacetic acid.

### 3.4.5 Protein quantification

#### 1) Bradford

Bovine serum albumin (Sigma) was used as a standard and Coomassie Plus-the better Bradford™ assay reagent (Pierce). Prepare five standards solution (1mL each) containing 0, 10, 20, 30, 40 and 50 μg/ml BSA. To a 1.4 ml plastic cuvelte, add;

0.2 ml Protein assay solution (BioRad).

0.8 ml of one of the protein standard solution or sample to be assayed (containing <100 μg of protein for <50 μg/ml standards)

Cover with Parafilm and gently invert several times to mix. Record the absorbance spectrum of sample at 595 nm and repeat the steps above for each of the protein standards and for samples to be assayed. Plot standard curve of Absorbance at 595 nm vs. protein concentration for the protein standard. To determine the protein concentration of sample from it absorbance, use the standard curve to find the concentration of standard that would have the sample absorbance as the sample.

#### 2) BCA

Bicinchoninic Acid Protein Assay Kit (Sigma) was used in this assay. Reagents provided in kit including, Bicinchoninic Acid Solution**,** Reagent A (1,000 ml solution containing bicinchoninic acid, sodium carbonate, sodium tartrate, and sodium bicarbonate in 0.1 N NaOH (final pH 11.25)). Copper(II) Sulfate Pentahydrate 4% Solution**,** Reagent B (25 ml solution containing 4% (w/v) copper(II) sulfate pentahydrate and protein standard BSA Solution (1.0 mg/ml BSA in 0.15 M NaCl with 0.05% sodium azide as a preservative). The BCA Working Reagent is prepared by mixing 50 parts of Reagent A with 1 part of Reagent B. Mix the BCA Working Reagent until it is light green in color. Prepare standards of different BSA concentrations (range from 200-1,000 mg/ml). Add 2 ml of the BCA Working Reagent to 0.1 ml of each BSA protein standard, blank, and unknown sample. Vortex gently for thorough mixing. The total liquid volume in the test tube is 2.1 ml. Incubate at 37 °C for 30 minutes and allows the tubes to cool to room temperature. Transfer the reaction solutions into a cuvet. Measure the absorbance of the solution at 562 nm. Create a standard curve to by plot the absorbance at 562 nm versus the BSA protein standard concentrations.

### 3.4.6 Ultrafiltration

Vivaspin centrifugal concentrators (Vivascience AG, Hannover, Germany) 500 µl, MWCO 5 kDa was used to exchange Arabidopsis guard cell extract dissolved in YeastBuster$^{TM}$ reagent to 0.5 M triethylammonium bicarbonate, (TEAB) pH 8.5. Initially, wash membrane by fill concentrator with 500 µl of 0.5M TEAB and centrifuge at 7500 rpm for 10 min, empty the filtrate container. Refill concentrator with a protein sample 500 µl, centrifuge and empty filtrate container. Refill concentrator with (3×500 µl) of 0.5M TEAB, centrifuge and empty filtrate container. Recover the concentrated, de-salted sample from the bottom of the concentrate pocket with a pipette.

### 3.4.7 iTRAQ labelling protocol

iTRAQ reagents and buffers were supplied by Applied Biosystems (Warrington, UK) and peptides were labelled according to the manufacturer's standard protocol with minor modifications. Briefly, 100µg of protein was diluted in 20µl dissolution buffer (0.5M triethylammonium bicarbonate, (TEAB) pH 8.5), reduced with 2µl reducing reagent (Tris (2-carboxy ethyl) phosphine (TCEP) (40-60mM)) and incubated at 60°C for 1 hour. Cysteine groups were blocked with 1µl methylmethane thiosulphonate in isopropanol (MMTS) (180-220mM) and incubated at room temperature for 10 minutes. Proteins were digested with trypsin (1:10, enzyme:substrate (w/w)) at 37°C for 16-18 hours. Digested samples were dried using a SpeedVac (Thermo Savant, UK) to a final volume of <10µl and made up to 25µl final volume with 0.5M TEAB. iTRAQ reagents were suspended in 70µl ethanol and added to the digested samples and incubated at room temperature for 1 hour. On completion of the labelling protocol, differentially labelled samples were mixed and dried. Prior to the labelling of plasma samples, the protocol was validated using a standard peptide and protein mixture.

### 3.4.8 Assay for Hemagglutinating activity

A serial two-fold dilution of the protein in micro titer U-plates (50 µL) was mixed with 50 µL of a 2% (v/v) suspension of rabbit erythrocytes in phosphate buffered saline, pH 7.2, at room temperature. The results were read after 1 h when the blank had fully sedimented. The hemagglutination titer, defined as the reciprocal of the highest dilution

exhibiting hemagglutination, was defined as one hemagglutination unit. Specific activity is the number of hemagglutination units per mg lectin.

### 3.4.9 Hemagglutination inhibition by carbohydrates

The hemagglutination inhibition tests were performed in a manner analogous to the hemagglutination test. Serial two-fold dilutions of sugar samples were prepared in phosphate buffered saline and mixed with an equal volume (25 µL) of the protein solution with eight hemagglutination units. The mixture was incubated for 30 min at room temperature and then mixed with 50 µL of 2% (v/v) rabbit erythrocyte suspension. The minimum concentration of the sugar in the final reaction mixture which completely inhibited eight hemagglutination units of the lectin preparation was calculated.

### 3.4.10 α-glucosidase inhibition

α-glucosidase from baker's yeast (EC.3.2.1.20) were purchased from Sigma (St. Louis, MO). The inhibitory effect of each compound on α -glucosidase activity was measured according to the literature procedure (*84*). α-glucosidase from baker's yeast was assayed using 0.1M phosphate buffer at pH 6.9, and 1mM p-nitrophenyl-α –D glucopyranoside (PNP-G) was used as a substrate. The concentration of the enzymes was 1 U/mL in each experiment. α -Glucosidase (40 µl) was incubated in the absence or presence of various concentrations of transcinnamic acid derivatives (10 µl) at 37° C. The preincubation time was specified at 10 min and PNP-G solution (950 µl) was added to the mixture. The reaction was carried out at 37 °C for 20 min, and then 1ml of 1 M $Na_2CO_3$ was added to terminate the reaction. The assays were performed in duplicate. The colored product was detected in a Bio Rad model 3550 microplate reader at 405 nm.

# CHAPTER IV

# PROTEINS IDENTIFICATION FROM NON-MODEL PLANTS: RHIZOME PROTEINS FROM ZINGIBERACEAE

## 4.1 Introduction

Protein identification may be achieved by a large variety of mass spectrometric techniques; however it ultimately requires that the acquired mass spectra be accurately matched to protein sequences from the corresponding database entries. Therefore, the availability of a complete genome or at least a substantial part of the cDNA sequences is importance (*85*). However, if the protein of interest is not present in databases, the sequence of a highly homologous protein from another species is required. Enzymatic digestion of the protein of interest would then be expected to yield some peptides identical to the ones present in the known protein homologue. In most case, peptide mass mapping is insufficient to produce a statistically reliable hit upon database searching. The approach to allow the identification is sequencing by tandem mass spectrometry that requires manual interpretation of MS/MS data and careful inspection of the match. Sequence similarity searches have been also employed to identify proteins via their known homologues in other species. If sequence similarity is still too low, or if the gene is new, tryptic peptides must be sequenced *de novo*.

Many previous reports of plants belong to family Zingiberaceae are concerned with their secondary metabolites, mainly extracted from rhizome, and a variety of biological properties of these compounds such as Curcumin and derivatives (*29*). In China and Japan many medicines were extracted from the rhizome of *Curcuma longa*, *Curcuma wenyujin*, *Curcuma phaeocaulis*, *Curcuma kwangsiensis*, *Curcuma zedoaria* and *Curcuma aromatica* (*86, 87, 88*). Moreover, they have been used for invigorating blood circulation, promoting digestion to alleviate pain, as a cholagogic, stomachic, and a health food. Although plants in this family have been used for a long time as a component of oriental traditional medicine, there are few reports on the biological activities of its proteins. Proteins from Zingiberaceae family have been studied mainly in ginger (*Zingiber officinale*) that used widely in the world. Ginger proteases from ginger rhizome

were studied by Choi and Laursen in 2000 (*89*). Recently, first mannose-binding lectin was cloned from ginger (*Zingiber officinale* Roscoe) rhizomes (*90*). However, 366 proteins of Zingiberaceae in NCBI protein database reveal that only maturase-like polypeptide and maturase K were found in *Curcuma* species.

## 4.2 Aim

The aim of this research is to identify biological active proteins from plants with incomplete genome sequence, *Cucurma zedoaria* (Kamin-oy) and *Curcuma aromatica* (Kanthamala) by a combination of tandem mass spectrometry and a modified BLAST searching protocol called 'MS BLAST' (mass spectrometry-driven BLAST searching).

## 4.3 Methods

### 4.3.1 Purification of *Curcuma zedoaria* protein

Protein extract from *Curcuma zedoaria* rhizomes was fractionated by $NH_4(SO_4)_2$ precipitation as described in Section 3.4.1. Proteins solution (300 ml) was chromatographed on Q Sepharose Fast-flow column. An unbound protein was removed by washing with mobile phase A. Bound proteins were then eluted with a linear gradient 0-100% mobile phase B over 15 column volumes. All collected fractions were investigated hemagglutinating assay and α-glucosidase inhibition. The hemagglutinating containing fraction was lyophilized, dissolved in water and dialyzed against 0.1 M $NH_4HCO_3$, pH 7.8. Insoluble proteins were removed by centrifugation at 13,000 g at 4 °C for 30 min and the supernatant was further purified by Superdex 75 column. The hemagglutinating containing fraction was lyophilized and finally purified by reversed-phase HPLC on a BDS $C_8$ column. The column was developed with a linear gradient of 10 to 60% mobile phase B (0-60 min) and 60 to 90% mobile phase B (60-90 min). For biological assay and sequencing studies, reversed phased HPLC were repeated ten times. The hemagglutinating containing peaks were pooled and lyophilized.

### 4.3.2 Purification of *Curcuma aromatica* protein

Protein extract from *Curcuma aromatica* (Kanthamala) was fractionated by $NH_4(SO_4)_2$ precipitation as described in Section 3.4.1. Proteins solution (300 ml) was chromatographed on DEAE-Cellulose. An unbound protein was removed by washing with mobile phase A. Bound proteins were eluted with a isocratic gradient of 100% mobile phase B. The fraction was then investigated hemagglutinating assay, lyophilized, dissolved in water and dialyzed against a 20 mM sodium acetate buffer at pH 5.0. Insoluble proteins were removed by centrifugation at 13,000 g at 4 °C for 30 min and the supernatant was further purified by SP-Sepharose Fast Flow column. Unbound protein fraction was collected and bound fraction was eluted with 0-100% linear gradient of buffer B over 15 column volumes. The activity containing fraction was lyophilized, dissolved in water and dialyzed against 0.1 M $NH_4HCO_3$ at pH 7.8. Insoluble proteins were removed by centrifugation at 13,000 g at 4 °C for 30 min and the supernatant was applied to Superdex 75 column. Protein fractions were lyophilized for further characterization by gel electrophoresis. (See all buffers in experimental chapter).

## 4.4 Results and disscussion

### 4.4.1 Mannose binding lectin from *Curcuma zedoaria* (Kamin-oy)

#### 1) Purification

Proteins precipitating between 20% and 60% ammonium sulphate saturation were chromatographed using a Q-Sepharose column. Three peaks in UV intensity were eluted with a linear NaCl gradient (0–0.35 M) in 20 mM phosphate buffer, pH 7.0 over 20 column volumes. These were designated Q1, Q2, and Q3, respectively (Figure 4.1).

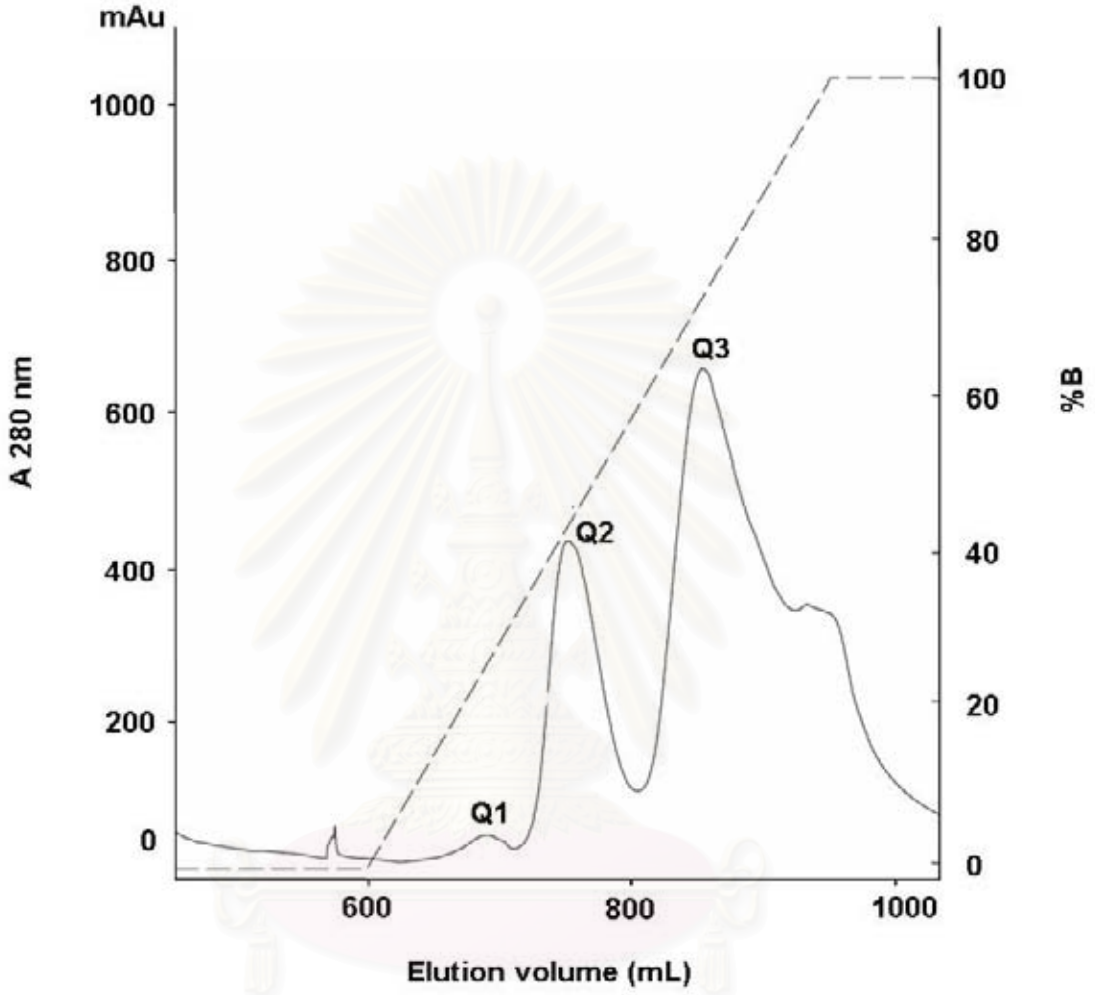



**Figure 4.1** Ion exchange chromatography of protein solution from *Curcuma zedoaria* on Fast-flow Q-Sepharose. Mobile phase A was 20 mM phosphate buffer, pH 7.0. Mobile phase B was as A with an additional 0.35 M NaCl. The column was washed with 5 column volumes mobile phase A and developed with 0-100% linear gradient of mobile phase B over 15 column volumes. (The solid line indicates UV intensity and the dash line indicates the 0–100% mobile phase B).

The ion exchange chromatography was chosen to be a first step of purification because its property can be loaded with a large volume of sample. Large sample volumes (>5 ml) were applied through the system pump.

The hemagglutinating assay was used to determine the activity of active fraction in each purification step. The small peak called Q1 was shown to have activity at 249.61 titer/mg (Table 4.1). The fraction was then pooled, lyophilized and loaded onto a Superdex 75 gel filtration column following protein complexity was investigated by 1D SDS-PAGE. It revealed a major band of low mass protein (Figure 4.2).

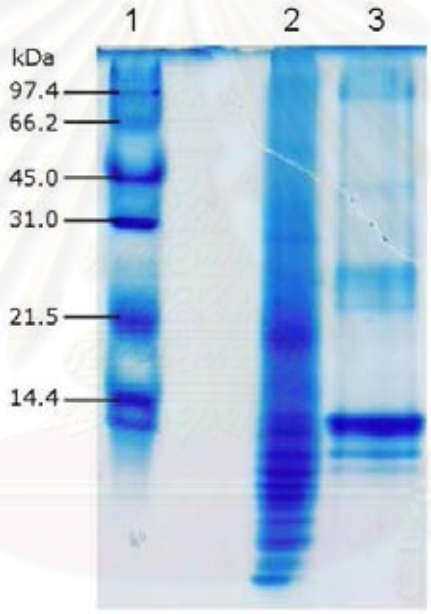



**Figure 4.2** 1D SDS-PAGE of *Curcuma zedoaria* protein separated by Q-Sepharose**.**

| | | |
|---|---|---|
| Lane 1 | Molecular weight marker of protein standard | |
| | Phosphorylase b | 97.4 kDa |
| | Albumin | 66.2 kDa |
| | Ovalbumin | 45.0 kDa |
| | Carbonic anhydrase | 31.0 kDa |
| | Soybean trypsin inhibitor | 21.5 kDa |
| | α-Macroglobulin | 14.4 kDa |
| Lane 2 | Crude 20-60 % $(NH_4)_2SO_4$ | 30 μg of protein |
| Lane 3 | Fraction Q1 from Q-Sepharose | 20 μg of protein |

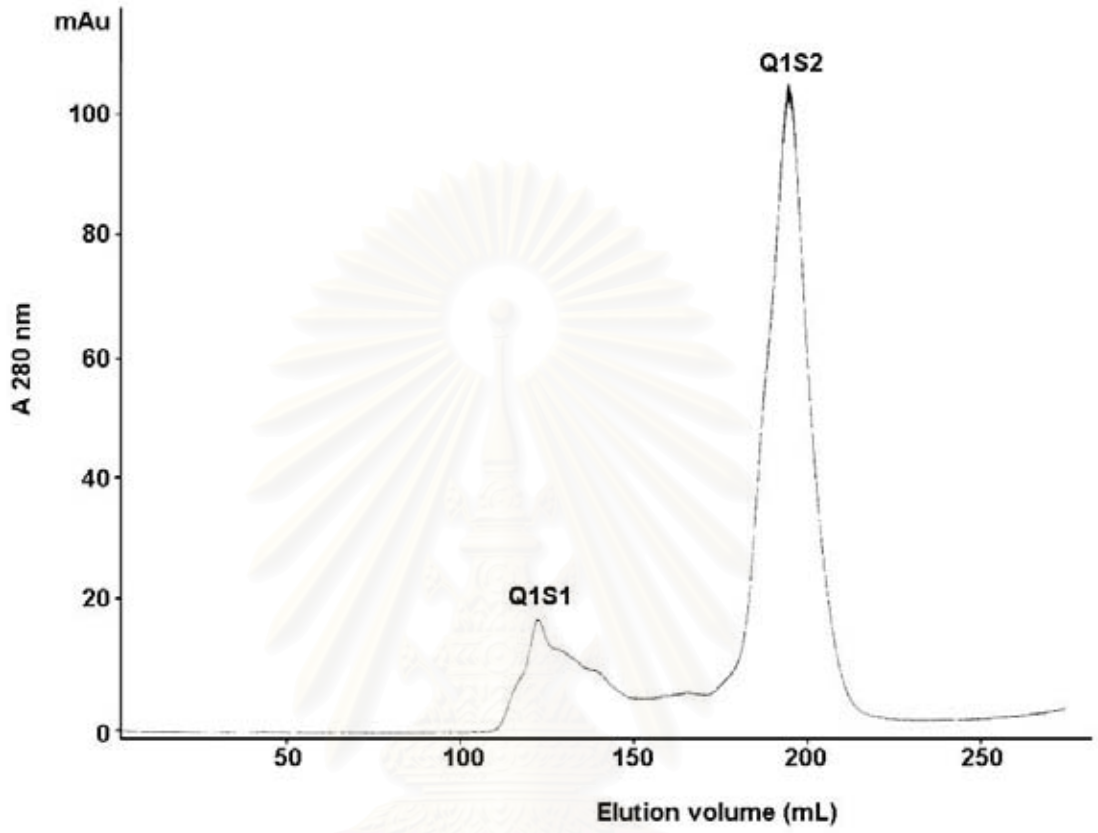



**Figure 4.3** Size exclusion chromatography of fraction Q1 on Superdex 75 in 0.1 M $NH_4HCO_3$, pH 7.8 at flow rate of 0.5 ml/min.

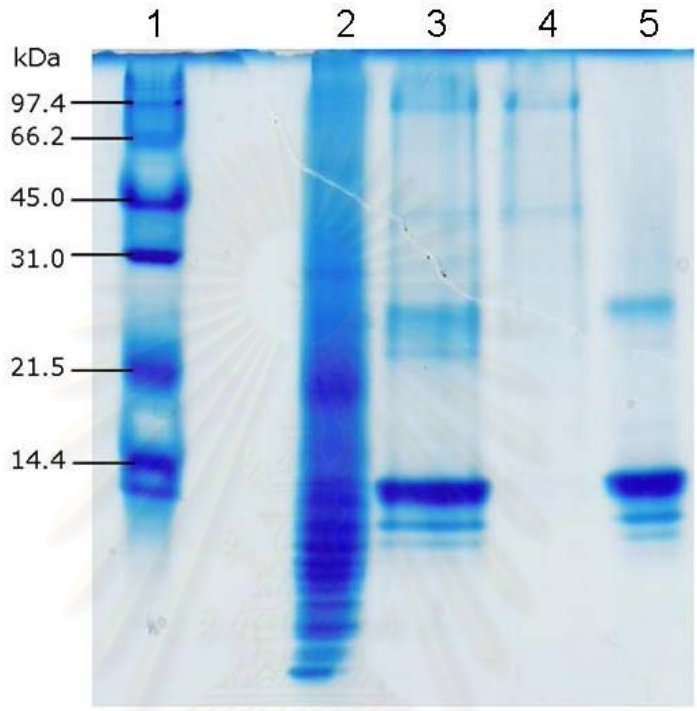



**Figure 4.4** 1D SDS-PAGE of *Curcuma zedoaria* protein separated by Q-Sepharose and Superdex 75.

| | | |
|---|---|---|
| Lane 1 | Molecular weight marker of protein standard | |
| Lane 2 | Crude 20-60 % $(NH_4)_2SO_4$ | 30 μg of protein |
| Lane 3 | Fraction Q1 from Q-Sepharose | 20 μg of protein |
| Lane 4 | Fraction Q1S1 from Superdex 75 | 5 μg of protein |
| Lane 5 | Fraction Q1S2 from Superdex 75 | 10 μg of protein |

The hemagglutinating containing fraction, Q1S2, contained major low mass protein (Figure 4.4). It was then subjected to HPLC on reversed phase column, BDS $C_8$. This purified step was repeated ten times to accumulate protein for hemagglutinating assay and sequencing studies.

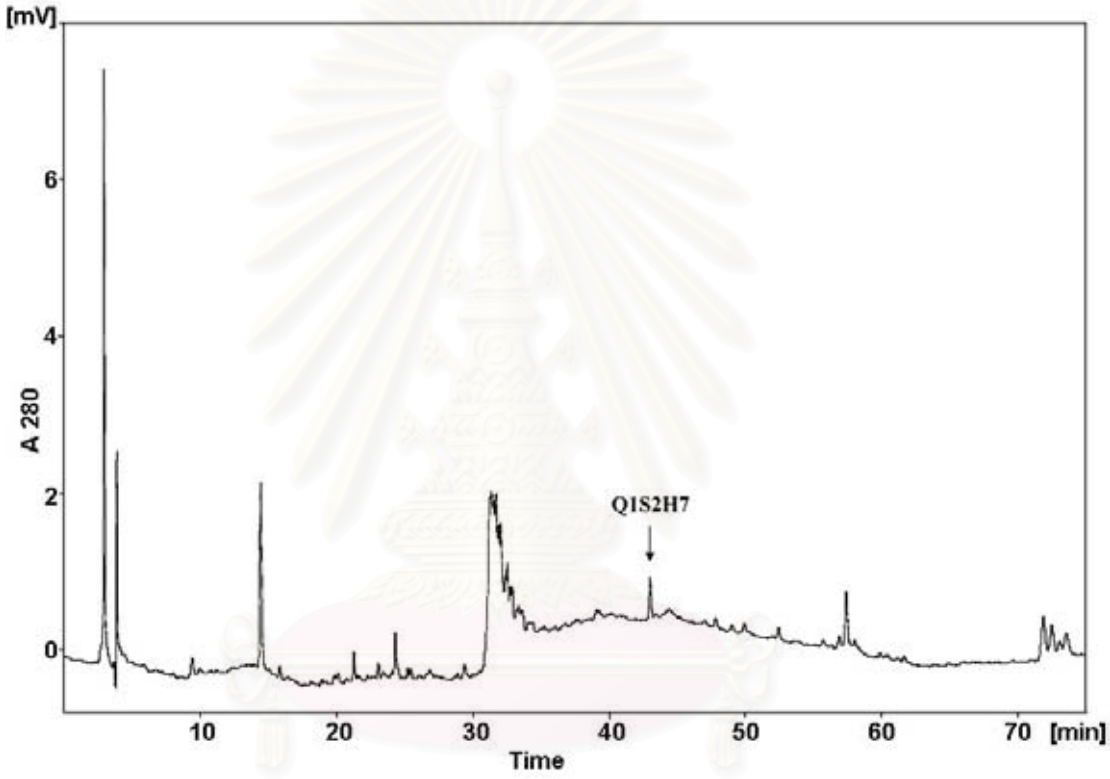



**Figure 4.5** Reversed phase HPLC of fraction Q1S2 from gel filtration on the $C_8$ column (4.6 × 250 mm). The column was developed with a linear gradient of 10 to 60% (v/v) mobile phase B (0-60 min) and 60 to 90% (v/v) mobile phase B (60-90 min) at a flow rate of 1 ml/min. Mobile phase A is 0.1% (v/v) trifluoroacetic acid/water, mobile phase B is 0.1% (v/v) trifluoroacetic acid/acetronitrile.

The hemagglutinating activity was found in a small peak called fraction Q1S2H7 eluted at retention time of 42.8 min (Figure 4.5). Consequently, SDS-PAGE with Coomassie Blue staining shows single band of purified protein (Q1S2H7) with molecular weight of 13 kDa (Figure 4.6).

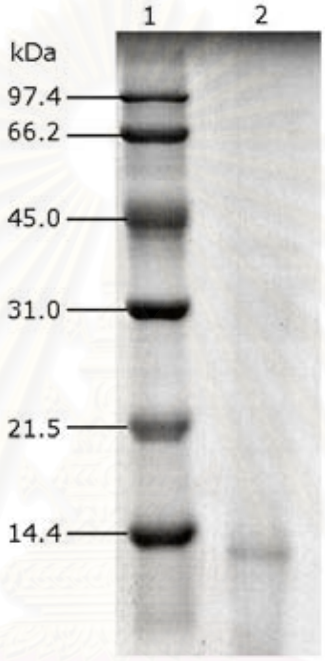



**Figure 4.6** 1D SDS-PAGE of purified protein from *Curcuma zedoaria*
Lane 1 Molecular weight marker of protein standard
Lane 2 Purified protein

The accurate molecular mass of protein was record as 13,448.7 Da by time-of-flight mass spectrometric analysis with sinapinic acid as the matrix (Figure 4.7).

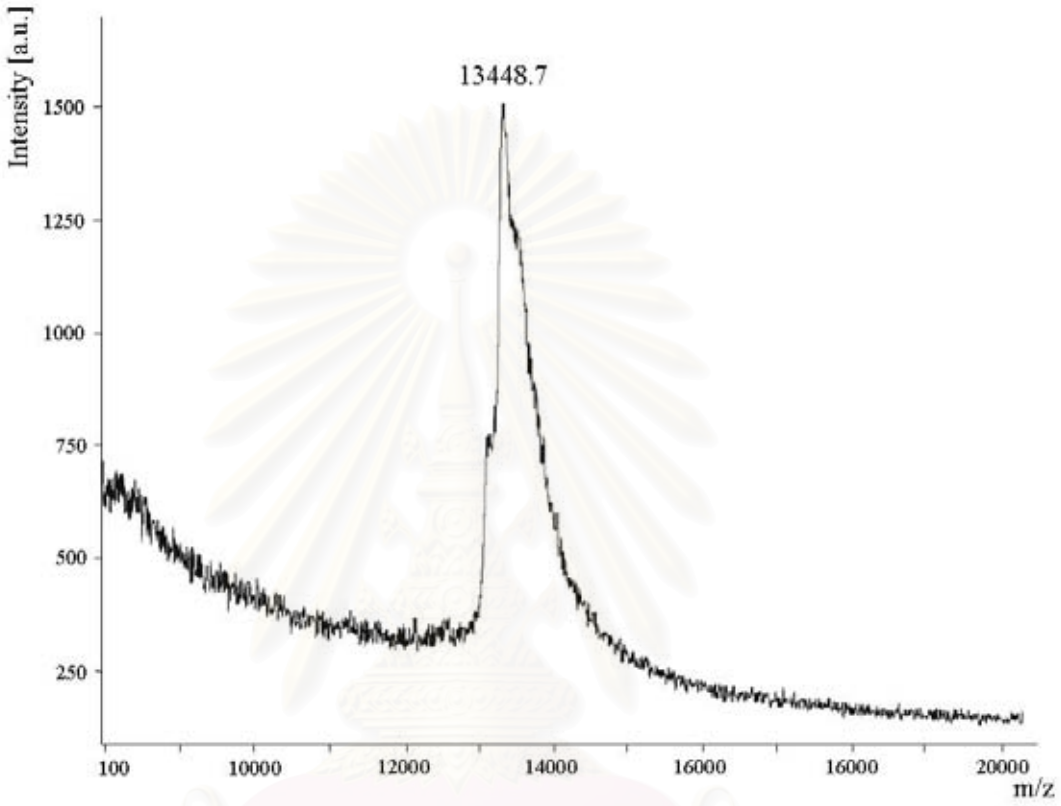



**Figure 4.7** Molecular mass of purified protein called Q1S2H7 by MALDI-ToF

The hemagglutinating assay was used to determine the activity of active fractions in each purification step. The calculated values; protein yield, specific hemagglutinating activity and fold of purification for each purification step were shown in Table 4.1.

**Table 4.1** Hemagglutinating activity during purification of *C. zedoaria* lectin from 1.5 kg rhizomes

| Fraction | Protein yield | Specific hemagglutinating activity (titera/mg) | Folds of purification |
|---|---|---|---|
| 20-60% $(NH4)_2SO_4$ | 1100.00 | 75.58 | 1 |
| Q1 | 5.80 | 249.61 | 3.3 |
| Q1S2 | 2.15 | 656.60 | 8.69 |
| Q1S2H7 | 0.23 | 5820.72 | 77.01 |

[a]Titer was defined as the reciprocol of the highest dilution with the hemagglutination with 2%(w/v) rabbit erythrocyte in phosphate buffer saline (PBS).

From hemagglutinating activity assay, it is suggested that an increase in specific activity of approximately seventy fold was obtained when the crude extract was subjected to the various purification steps reported (Table 4.1).

This purified protein was classified as lectin because it contains hemagglutination activity. To investigate sugar specificity of this lectin, comparing the ability of various sugars to inhibit its hemagglutinating activity was determined. Concentration of lectin was eight hemagglutination units, and 2% (w/v) rabbit erythrocyte was used in this assay. A solution of the lectin with eight hemagglutinating units of activity was inhibited by 31 mM D(+) mannose and was not inhibited by other sugars (Table 4.2).

**Table 4.2** Test of inhibition of lectin-induced hemagglutination by various sugar.

| Sugars | Minimal inhibitory concentration (mM) |
|---|---|
| D-mannose | >31 |
| D-glucose | >125 |
| D-galactose | >125 |
| D-moltose | >125 |
| D-lactose | >125 |
| D-fructose | >125 |
| D-xylose | >125 |
| D-sorbital | >125 |

Note Concentration of lectin was eight hemagglutination units, and 2% (w/v) rabbit erythrocyte was used in this assay.

**2) Identification of Protein by tandem MS using MALDI- ToF/ToF**

Following tryptic digestion, peptide mixtures were analyzed using MALDI-ToF/ToF (Bruker Ultraflex II, Bruker daltonics). Twenty product ion spectra were obtained. Peptide sequences and protein database searches were performed via BioTools Version 2.2 with the RapiDeNovo extension (Bruker).

Peptide sequences were determined by a *de novo* sequencing algorithm to derive a short list of possible sequence candidates which served as query sequences in a subsequent homology-based database search. These sequences were searched against non redundant protein database using a MS-BLAST to find likely protein candidates. The results from MS-BLAST searching showed that query sequence LNTGDFLTEGEFLFLMK (m/z of 1,973.98) have similar sequence to the previously isolated mannose-specific lectins (Q39728) from *Epipactis helleborine* (*91*) at amino acid residues 34 to 50 (Figure 4.8). This sequence producing high-scoring segment pair (HSP) equal 77, which is higher than the threshold of statistical significance for a single matched HSP and positive identities percentage is 76% (Figure 4.8). The tandem mass spectrum of

this precursor ion was shown in Figure 4.9. Moreover, the percentage of sequence coverage was increased from two product ion spectra m/z of 1014.52 with sequence DGHVVIYGR (Figure 4.10) and m/z of 771.46 with sequence LVLYHK (Figure 4.11). It reveled peptide sequence similar to the same protein (Q39728) even though these fragments were not hit by MS BLAST searching.

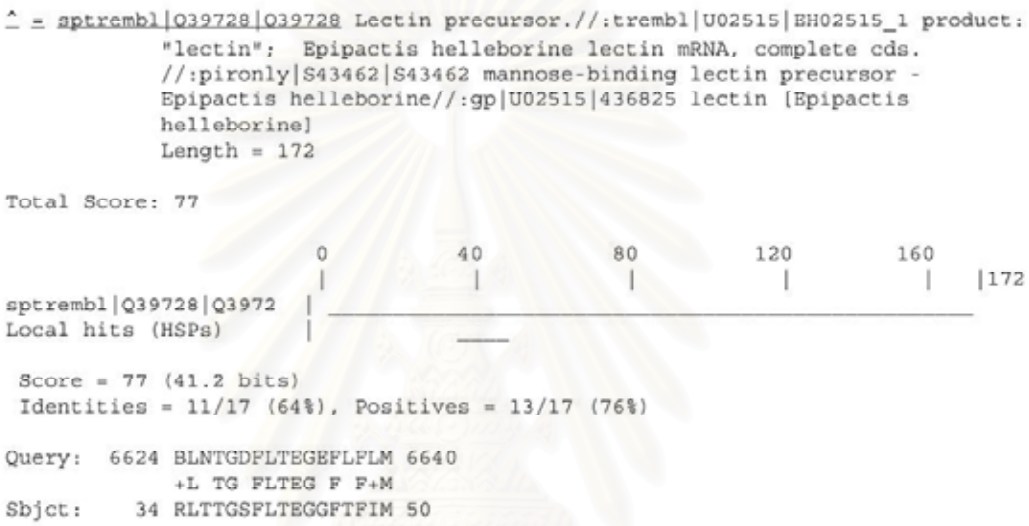



**Figure 4.8** MS-BLAST result showed the scoring alignments of the queried peptide sequence and corresponding homologous peptides from the database.

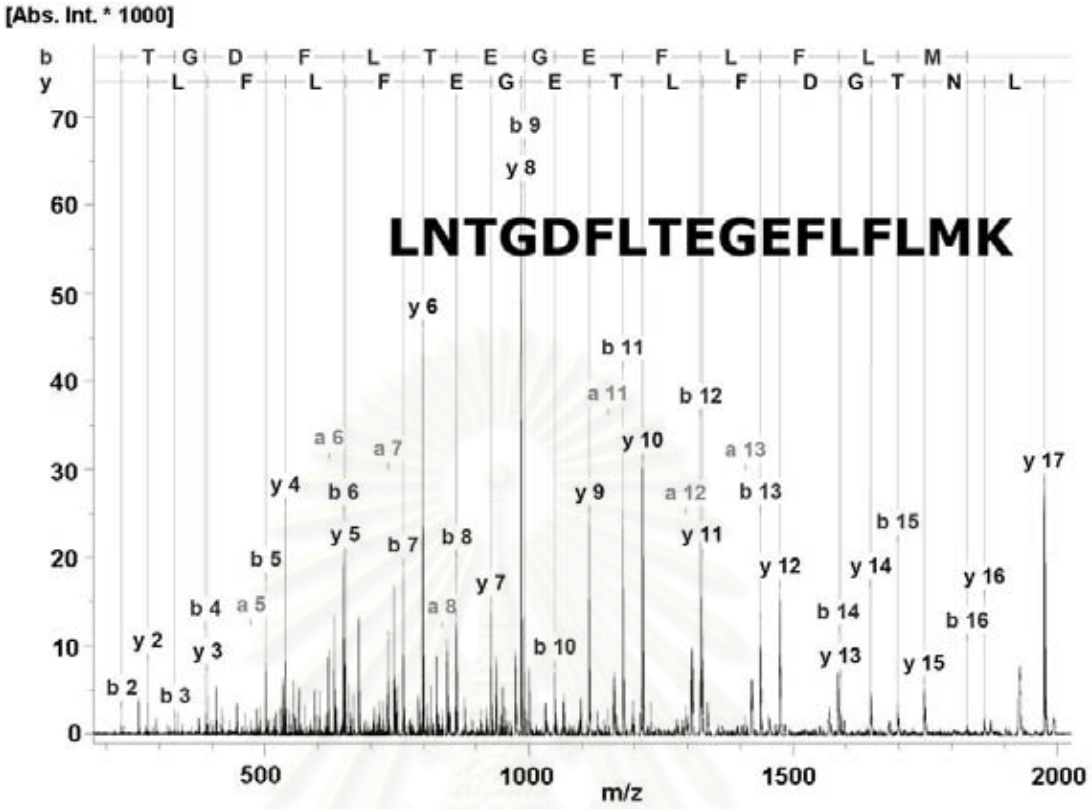



**Figure 4.9** Tandem spectrum of precursor ion m/z of 1973.98

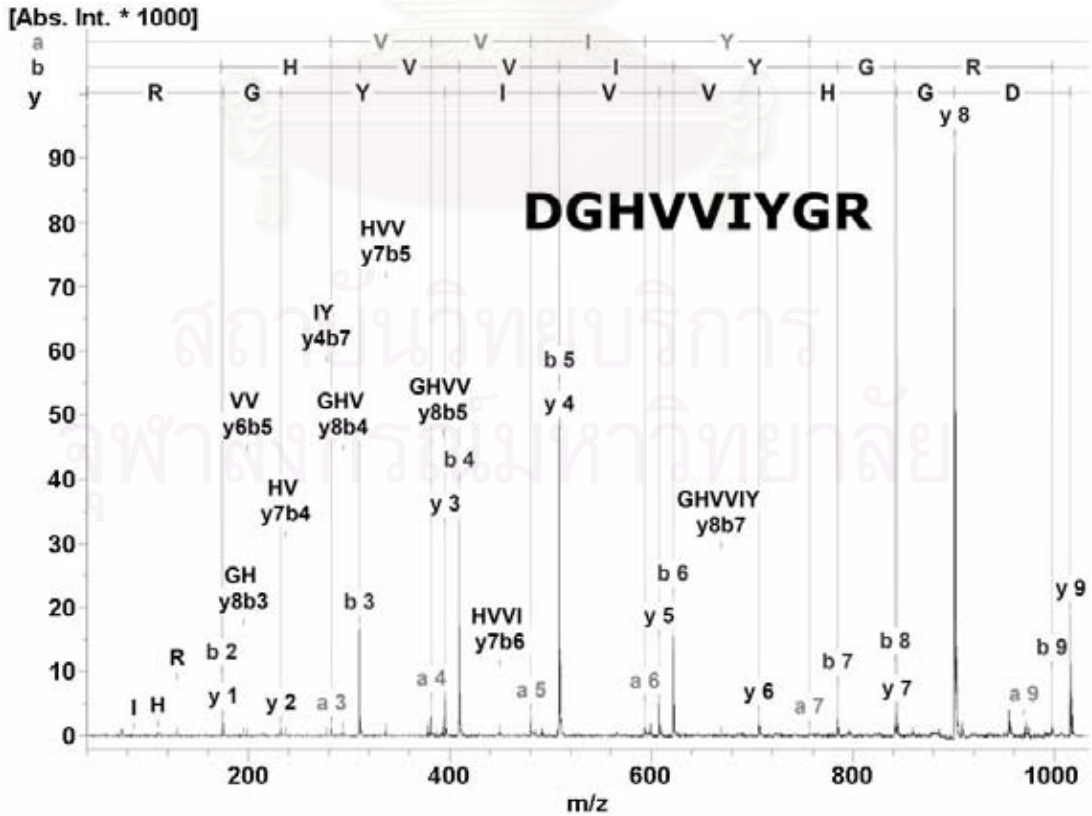



**Figure 4.10** Tandem spectrum of precursor ion m/z of 1014.52

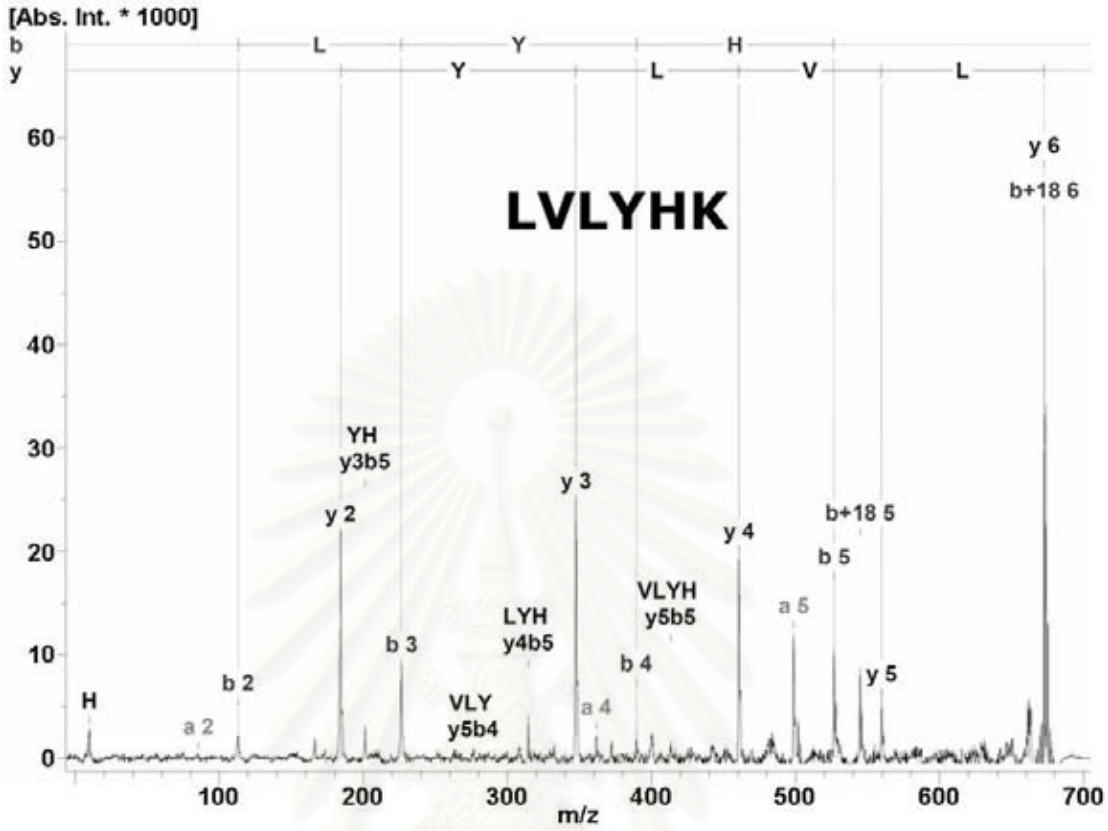



**Figure 4.11** Tandem spectrum of precursor ion m/z of 771.46

The amino acid alignment compared between Q1S2H7 from *Curcuma zedoaria* and Q39728 was shown in Figure 4.12


**Figure 4.12** Amino acid sequence of the purified protein, Q1S2H7, from *C. zedoaria* align with known amino acid sequences of *Epipactis helleborine*

### 4.4.2 α-glucodidase inhibitory activity of protein from *Curcuma zedoaria*

In this study, crude protein and protein containing fractions from Q-Sepharose of *Curcuma zedoaria* (0.2mg/ml) were used to determine their inhibitory activity against α-glucosidase. Inhibition >50% was taken as significant. The results in Table 4.3 revealed that fraction Q1S1 contained potent inhibitors (94.3%).

**Table 4.3** α-Glucosidase inhibition

| Fraction | % inhibition (n=2) |
|---|---|
| Crude (20-60%) | 80.3 |
| Q-unbound | 43.1 |
| Q1 | 29.5 |
| Q2 | 26.4 |
| Q3 | 21.8 |
| Q1S1 | 94.3 |
| Q1S2 | 56.3 |

Values of %inhibition are calculated from duplicate measurements.

For identification, fraction Q1S1 was digested in-solution by trypsin. The peptide mixture was then analyzed by LC-MS/MS using Q-ToF. However, MASCOT searching result was not revealed any identified protein. The MS/MS spectra of each precursor ion from MS survey scan (Figure 4.13) were manually inspected and *de novo* peptide sequence was submitted to MS-BLAST. However, with 12 MS/MS spectra (see appendix C) submitted, MS-BLAST searching was not revealed any protein from database.

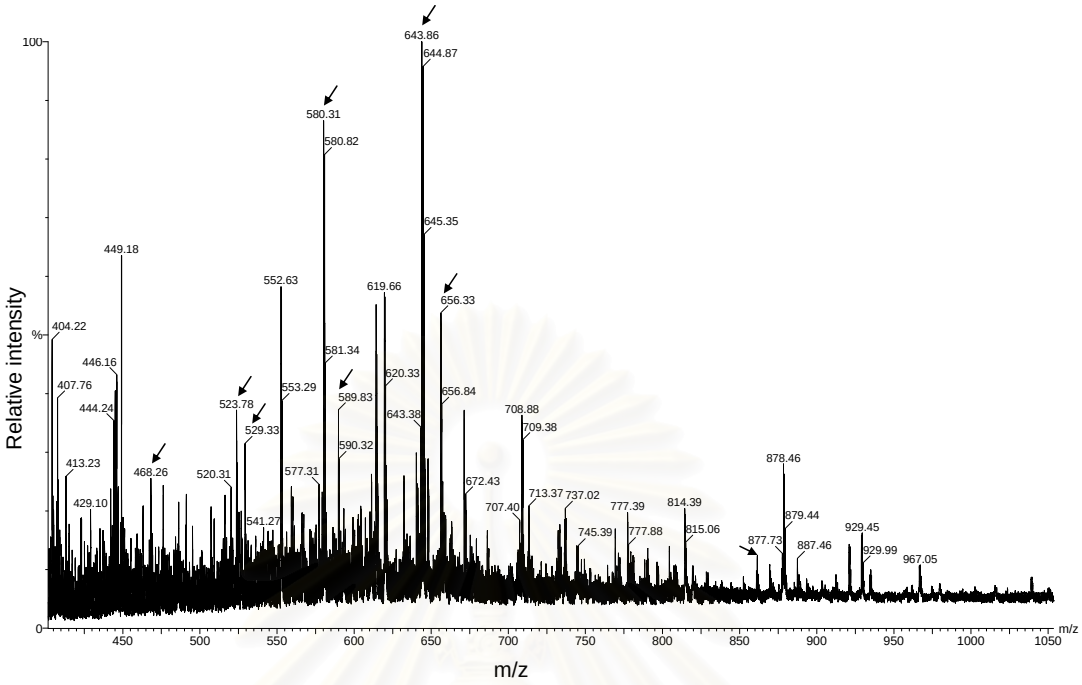



**Figure 4.13** MS survey scan 400-2000 m/z of digests Q1S1.

### 4.4.3 Hemagglutinating activity and corresponding putative sequence identity from *Curcuma aromatica* (Kanthamala)

#### 1) Protein isolation and SDS-PAGE

Protein from *Curcuma aromatica* (Kanthamala in Thai) rhizomes was extracted and precipitated by $NH_4(SO_4)_2$ as described in the experimental chapter. Ion exchange chromatography was chosen to be a first step of purification because its property can be loaded with a large volume of sample. Nevertheless, most of protein components were not bound with SP-Sepharose column (Figure 4.14).

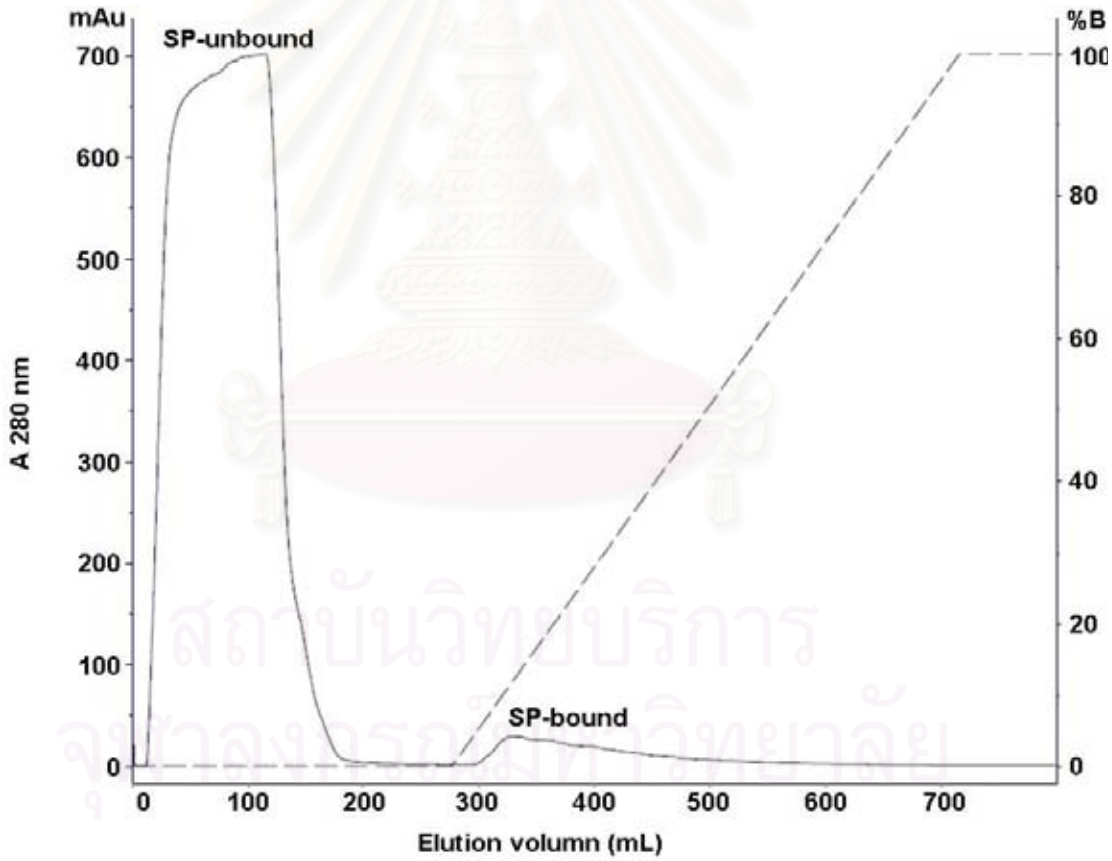



**Figure 4.14** Ion exchange chromatography of crude *Curcuma aromatica* (Kantamala) on Fast-flow SP-Sepharose. The dash line indicates the 0–100% linear gradient of eluent B (NaCl 0.35 M) in 20 mM NaOAc buffer, pH 5.0.

This unbound fraction containing hemagglutinating activity toward rabbit erythrocyte at 0.55 mg/ml was then lyophilized and further purified by gel filtration chromatography on a Superdex 75 column. The isolated proteins yielded three peaks called S1, S2 and S3 (Figure 4.15).

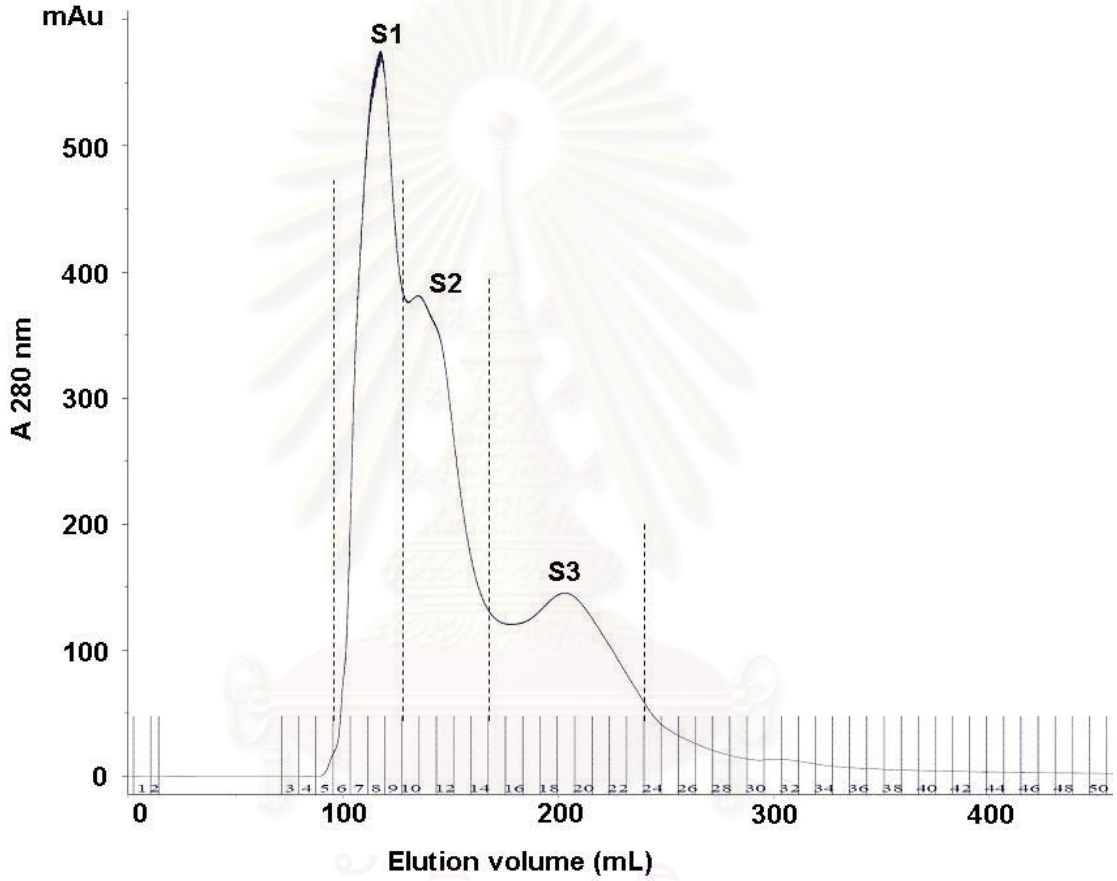



**Figure 4.15** Size exclusion chromatography of unbound fraction on Superdex 75 in 0.1 M $NH_4HCO_3$, pH 7.8.

Hemagglutinating activity was located in the large peak S1 (pooled fraction 6-9) at 0.19 mg/ml and peak S2 (pooled fraction 10-14) at 0.14 mg/ml. Consequently, SDS-PAGE with Coomassie blue staining method was used to separate both the activity containing peaks. The Coomassie blue staining method was used to visualize protein band

because of its compatibility with mass spectrometry analysis. S1 and S2 protein fractions resolved in SDS-PAGE are shown in Figure 4.16.

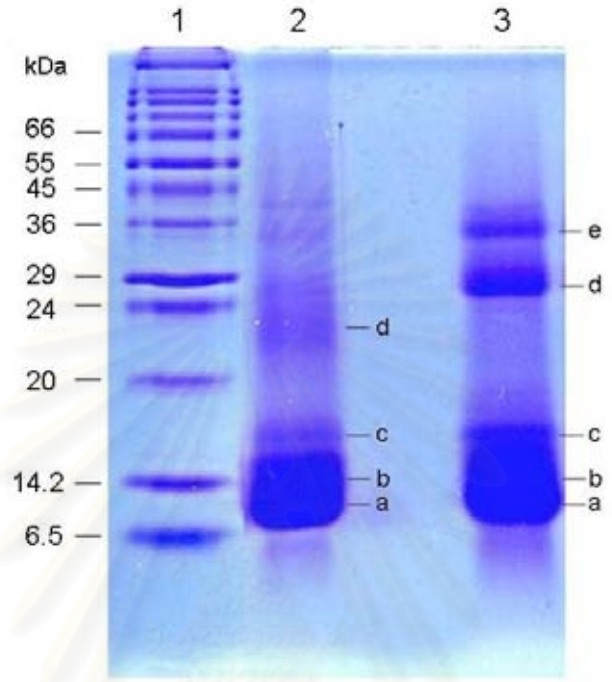



**Figure 4.16** 1D SDS-PAGE of *Curcuma zedoaria* protein separated by Superdex 75

| | | |
|---|---|---|
| Lane 1 | Molecular weight marker of protein standard | |
| | Albumin | 66.0 kDa |
| | Glutamic dehydrogenase | 55.0 kDa |
| | Ovalbumin | 45.0 kDa |
| | Glyceraldehyde-3-phosphate dehydrogenase | 36.0 kDa |
| | Carbonic anhydrase | 29.0 kDa |
| | Trypsinogen | 24.0 kDa |
| | Trypsin inhibitor | 20.0 kDa |
| | α-Lactalbumin | 14.2 kDa |
| | Aprotinin | 6.5 kDa |
| Lane 2 | Fraction S1 | 25 μg of protein |
| Lane 3 | Fraction S2 | 25 μg of protein |

From gel electrophoresis result, only one distinct region was apparent on S1 lane and two distinct regions on S2 lane. The dominant bands marked in Figure 4.16 were excised, four bands excised from S1 lane (called a, b, c, d) and five bands excised from S2 lane (called a, b, c, d, e). Excised gel bands were subjected to in-gel digestion, and the tryptic peptide hydrolyzed at the carboxyl side of lysine and arginine residues by trypsin were dried under vacuum prior to analyze by LC-MS/MS.

**2) LC-MS/MS and protein identification**

Following protein fractionation, proteins were then analyzed by LC-MS/MS and identified by mascot and MS BLAST as experimental workflow in Figure 4.17. LC-MS/MS was performed using a Quadrupole Time-of Flight (Q-ToF) as described in the experimental chapter.

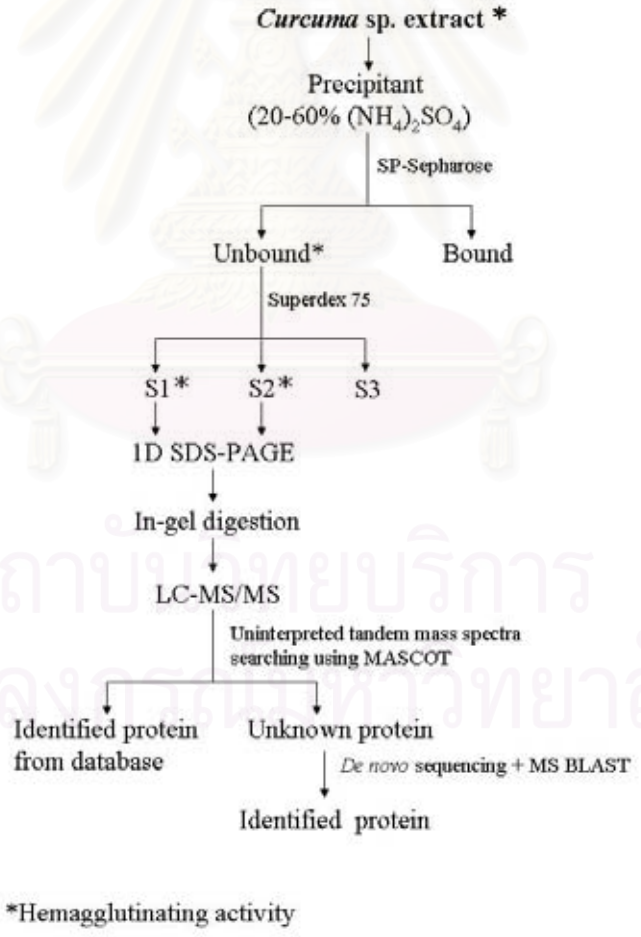



**Figure 4.17** Schematic of approaches for protein fractionation and gel-based identification from *Curcuma aromatica*.

The initial step of protein identification, all MS/MS spectra were used for database searches with Mascot software (*7*) against all sequences in the National Center for Biotechnology Information (NCBI) protein non-redundant database at the Matrix Science Ltd. server (http://www.matrixscience.com/).

The following parameters were set for Mascot: Trypsin with up to one missed cleavage allowed; fixed modifications: carbamidomethyl (C); variable modifications: oxidation (M); peptide tolerance ±200 ppm; MS/MS tolerance ±200 mmu; peptide charge +2, +3 (monoisotopic); instrument type: ESI-QUAD-TOF. If separated proteins do not match to any protein in database, *de novo* sequencing was performed manually using the Peptide Sequencing tool implemented in the MassLynx (Micromass) and the resulting peptides were searched against the nonredundant database at modified mass spectrometry-driven BLAST searching protocol (MS-BLAST) (*70*) using the European Molecular Biology Laboratory (EMBL) web-interface (http://dove.emblheidelberg.de/Blast2/msblast.html) and default settings.

Protein bands, which can be detected by Coomassie blue, are generally present in sufficient quantity to permit identification by tandem mass spectrometry. This technique generally requires at least 1-2 pmol of protein and it is required to identify peptides from organism with incomplete genome sequence. The approach to allow the identification of protein from organism lack genome or proteome information is determination partial or complete amino acid sequences by manual or automated *de novo* sequencing. An amino acid sequence was interpreted directly from spectra without comparison to database. The sequence candidates are then submitted to protein database to recognize with protein from other species using similarity searching. In this study, nine protein bands from two hemagglutinating active containing fractions, S1 and S2, from the *Curcuma aromatica* (Kanthamala) were identified (Table 4.4, 4.5). Only band S2/d and S2/e were identified by mascot as Cysteine proteinase GP-I (P82473) that found previously in ginger (*Zingiber officinale*) (Table 4.4). Peptide sequence CGIAISPSYPIK was hit significantly with mascot ion score for both of them. Peptides from the rest of protein bands failed to give a significant match with Mascot. The example of the identification of S2/d gel band by LC-MS/MS was shown in Figure 4.18.

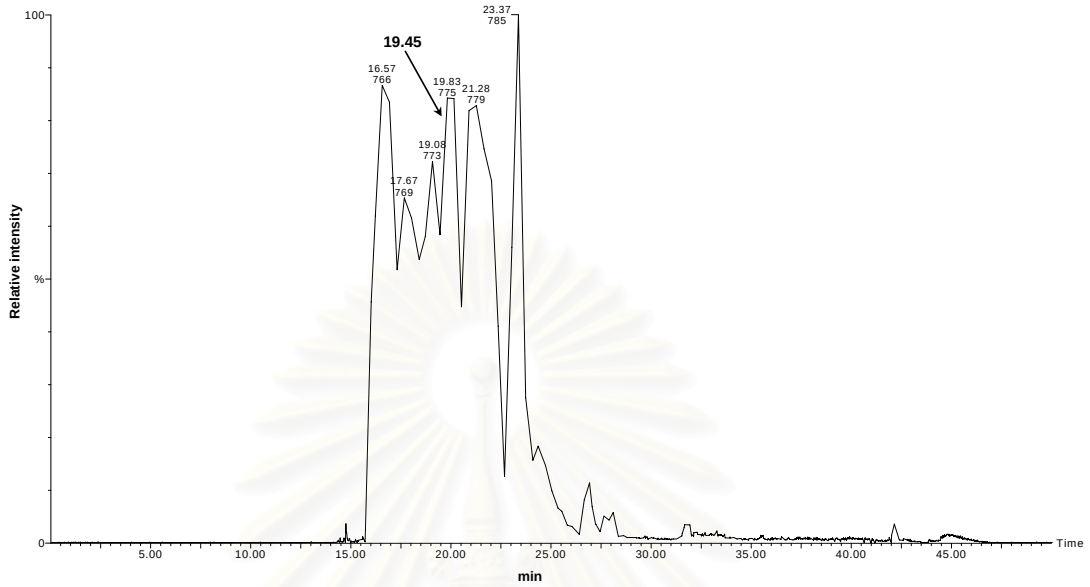
**Table 4.4** Protein bands S2/d and S2/e were identified as Cysteine proteinase GP-I by mascot

| Band | Accession number | Org | Mass (Da) | (M+2H)2+ | Sequence | Mr (expt) | Mr (calc) | Δ M (Da) | Ion score |
|---|---|---|---|---|---|---|---|---|---|
| S2/d | P82473 | *ZO* | 24628 | 653.3276 | CGIAISPSYPIK | 1304.6406 | 1304.6798 | -0.0392 | 49 |
| S2/e | P82473 | *ZO* | 24628 | 653.3455 | CGIAISPSYPIK | 1304.6764 | 1304.6798 | -0.0034 | 59 |

*ZO* is *zingiber officinale*

All identification met statistical confidence criteria according to Mascot with $p<0.05$. (*7*).

**A.**




**B.**

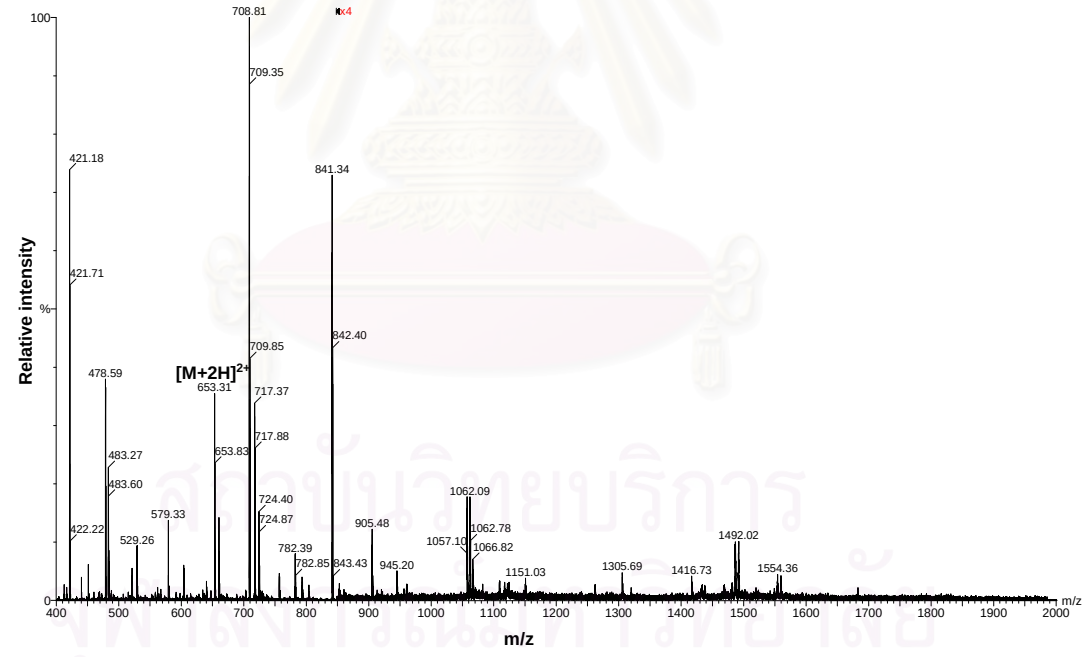
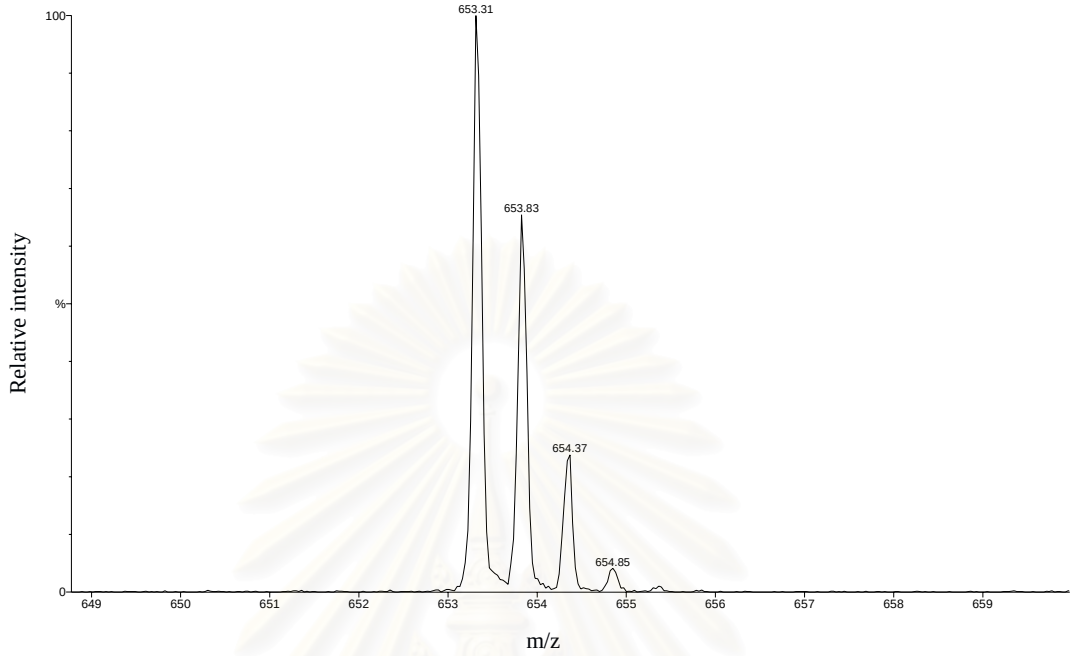



**Figure 4.18** Example of S2/d identified by mass spectrometry. A) Base peak ion chromatogram B) Mass spectrum of precursor ion eluted at 19.45 min scan from 400-2000 Da. C) Isotope distribution of doubly charged precursor ion m/z of 653.31. D) MS/MS spectrum of precursor ion m/z of 635.31.

**C.**




**D.**

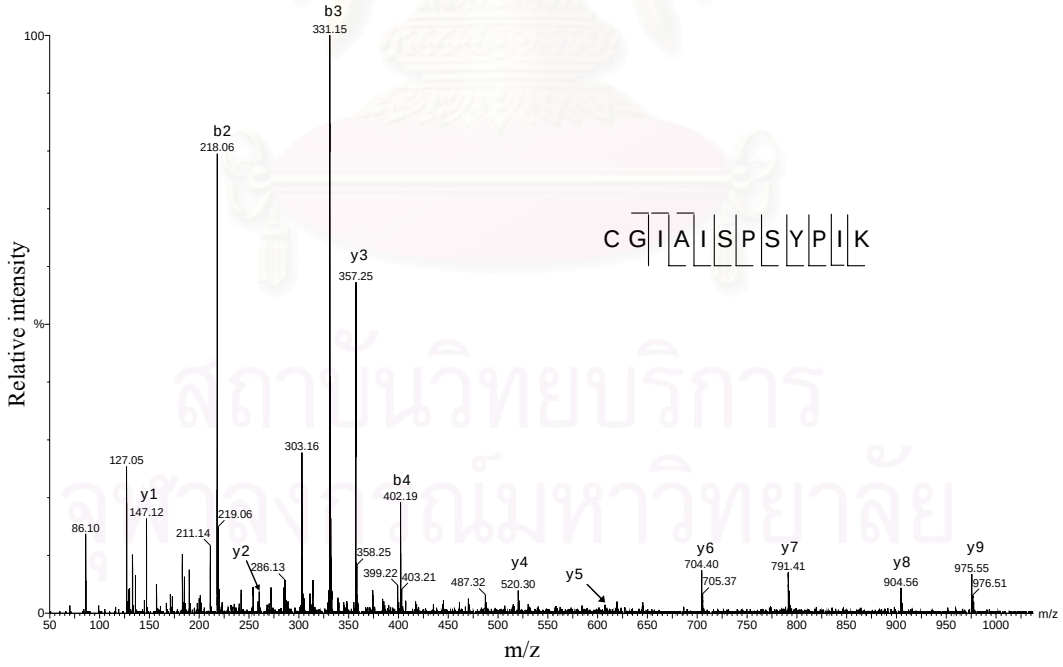



**Figure 4.18** Continued.

To overcome this problem, peptides were further sequenced *de novo* and matched for homology-based identification with modified mass spectrometry-driven BLAST searching method (MS BLAST). This search uses an alternative scoring scheme, based on threshold scores that are set conditionally on the number of retrieved high-scoring segment pairs and the total number of fragmented precursors.

The list of positive hit of identified *Curcuma aromatica* proteins by MS BLAST searching were shown in Table 4.5. They are significant with respect to the threshold values. The score of the high scoring segment pair (HSP) of the protein hit was compared with the threshold score.

**Table 4.5** Protein identification of 1D gel bands from *Curcuma aromatica* by MS BLAST

| Band[a] | Accession number | Protein | Organism | Mass (kDa) | MW (expt) | Query / Subject | HSP score[b] | %Pos[c] |
|---|---|---|---|---|---|---|---|---|
| S1/a | Q40417 | Dimeric mannose specific lectin | *Narcissus hybrid cultivar* | 18.5 | 1606.644 | NEALWSSDTDGLGSR<br>N A+W S+TDG<br>NKAIWASNTDG | 57 | 81 |
| | | | | | 1555.644 | MQTDGNFLVFNDR<br>MQTDGN V+<br>MQTDGNLVVY | 54 | 80 |
| | | | | | 1421.704 | TVTNNNLLYAGDK<br>+N+LY GD<br>DNILYSGD | 44 | 87 |
| | | | | | 1606.644 | DQALWSSDTDGLGSR<br>D +W +T GL S<br>DKPIWATNTGGLSS | 42 | 57 |
| S1/b | Q7NTB6 | Probable mannose binding lectin | *Chromobacterium violaceum* | 8.84 | 1454.584 | MQTDGNFLLFN<br>MQTDGNF L++<br>MQTDGNFVLYD | 66 | 90 |
| | | | | | 1454.664 | MQTDGNFLLFGGR<br>MQ DGN ++<br>MQDDGNLVIY | 40 | 70 |
| S1/c | Q41624 | Mannose-binding lectin | *Tulipa hybrid cultivar* | 19 | 1403.715 | ETEGNFVLVLQR<br>EGN+VLVLQ<br>EGNYVLVLQ | 56 | 100 |
| | | | | | 1620.638 | QDALWSSDTDQLSR<br>LW +TDQ S<br>LWATNTDQFS | 46 | 70 |
| | | | | | 1898.839 | VVLYGNPLFTLPNAEPR<br>VV+YG<br>VVIYG | 35 | 100 |
| S1/d | P92932 | Lectin related protein | *Allium sativum* | 32.8 | 1512.724 | MQTDGNFLLFDGR<br>MQ DGNF ++<br>MQVDGNFVIY | 54 | 80 |
| | | | | | 1606.704 | DQALWSSDTDGATLK<br>+A+WSS TD<br>HAIWSSHTD | 48 | 88 |
| | | | | | 1580.076 | DGHVVLFPEDV<br>DG VV++ DV<br>DGNVVIYGPDV | 39 | 72 |
| S2/a | Q39906 | Lectin | *Galanthus nivalis* | 17.3 | 1898.984 | VVLYGPARFTLPGQEPR<br>VV+YGPAR<br>VVIYGPAR | 57 | 100 |
| | | | | | 1454.584 | MQTDGNFLLFNR<br>MQTDGN +N<br>MQTDGNLVVYN | 55 | 72 |

**Table 4.5** (Continued)

| Band[a] | Accession number | Protein | Organism | Mass (kDa) | MW (expt) | Query<br>Subject | HSP score[b] | %Pos[c] |
|---|---|---|---|---|---|---|---|---|
| S2/b | Q40232 | Mannose-binding protein | *Listera ovata* | 17.7 | 1260.584 | AVWASNTDGKGR<br>AVWAS T+G G<br>AVWASGTNGRG | 56 | 81 |
| | | | | | 1260.584 | GLWASNTDGKGR<br>+WASNT+<br>IWASNTN | 47 | 100 |
| | | | | | 1562.624 | QEALWSSGTDGLGSR<br>A+W GT+<br>AIWATGTN | 39 | 75 |
| | | | | | 1085.924 | GHYTLVLQR<br>G Y L+LQR<br>GNYYLILQR | 38 | 77 |
| | | | | | 1555.704 | MQTDGNFLVFNDR<br>MQ DGN ++<br>MQRDGNLVIY | 36 | 70 |
| S2/c | Q8U2Y6 | Prismane protein | *Pyrococcus furiosus* | 51.3 | 1901.084 | YLGAKPPLFTAPNLFPR<br>YLG KPP F PN F<br>YLGPKPPEFLTPNVF | 63 | 66 |
| S2/d | P82473 | Cysteine proteinase GP-I | *Zingiber officinale* | 24.6 | 1404.584 | ENDHVVSLDSYK<br>EN HVVS+DSY<br>ENAHVVSIDSY | 71 | 90 |
| | | | | | 1432.724 | GCLALSPSYPLKK<br>+A+SPSYP+K<br>IAISPSYPIK | 66 | 100 |
| | | | | | 1099.384 | FWGDSGYLR<br>WG+SGY+R<br>WGESGYIR | 56 | 100 |
| S2/e | Q7XYU7 | Senescence-associated cysteine protease | *Anthurium andraeanum* | 50.5 | 1304.664 | NCLALSPSYPLK<br>+A+ PSYP+K<br>IAIEPSYPIK | 56 | 90 |
| | | | | | 1276.567 | DDLPDXXDWR<br>DDLP DWR<br>DDLPQKVDWR | 56 | 70 |
| | | | | | 1099.504 | FWGDSGYRL<br>WG+ GY<br>WGEAGY | 37 | 83 |
| | | | | | 1511.885 | QGGAAAPVSVLGGCNVL<br>GA AP+ GGC<br>GAVAPIKDQGGC | 29 | 66 |

[a]Band name according to posotion on 1D gel (Figure 4.16).

[b]MS BLAST match was defined as statistically significant if the score of the HSP was higher than the threshold value that scoring scheme are described in (*70*)

[c]%positive

MS BLAST result revealed that sequence coverage of S2/d can be increased by two peptide fragments that sequences were generated by *de novo* sequencing (Figure 4.19, 4.20). The score of the top-ranked HSP was 71 (Used 13 peptides; threshold score 1HSP=66).

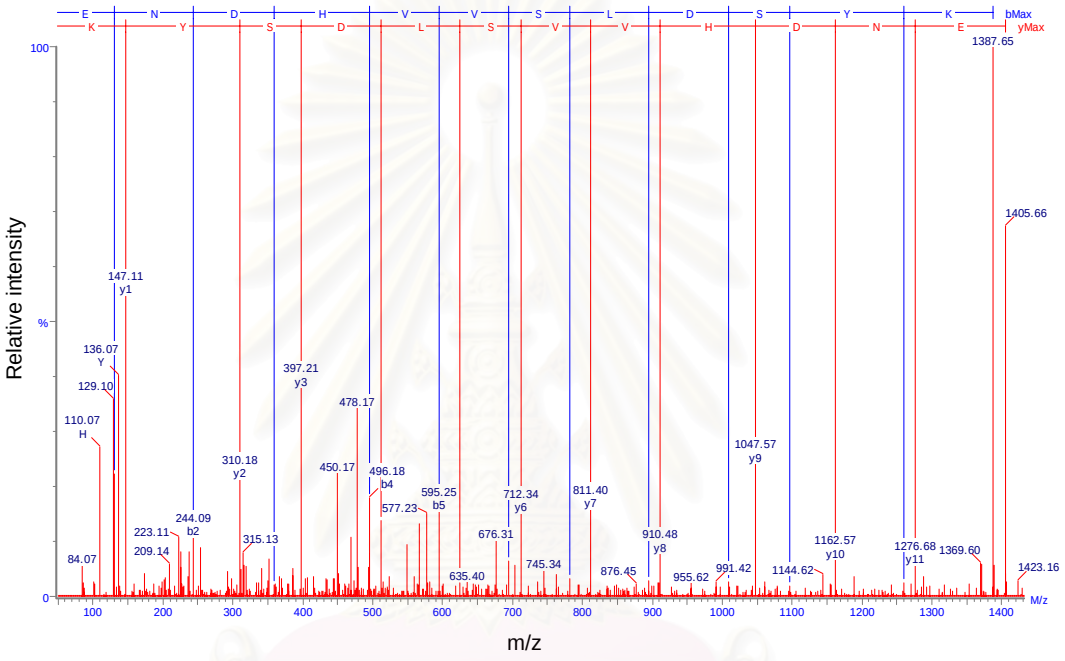



**Figure 4.19** MS/MS spectrum of doubly charged ion m/z of 703.30. *De novo* peptide sequence ENDHVVSLDSYK was performed manually using the Peptide Sequencing tool implemented in the MassLynx (Micromass).

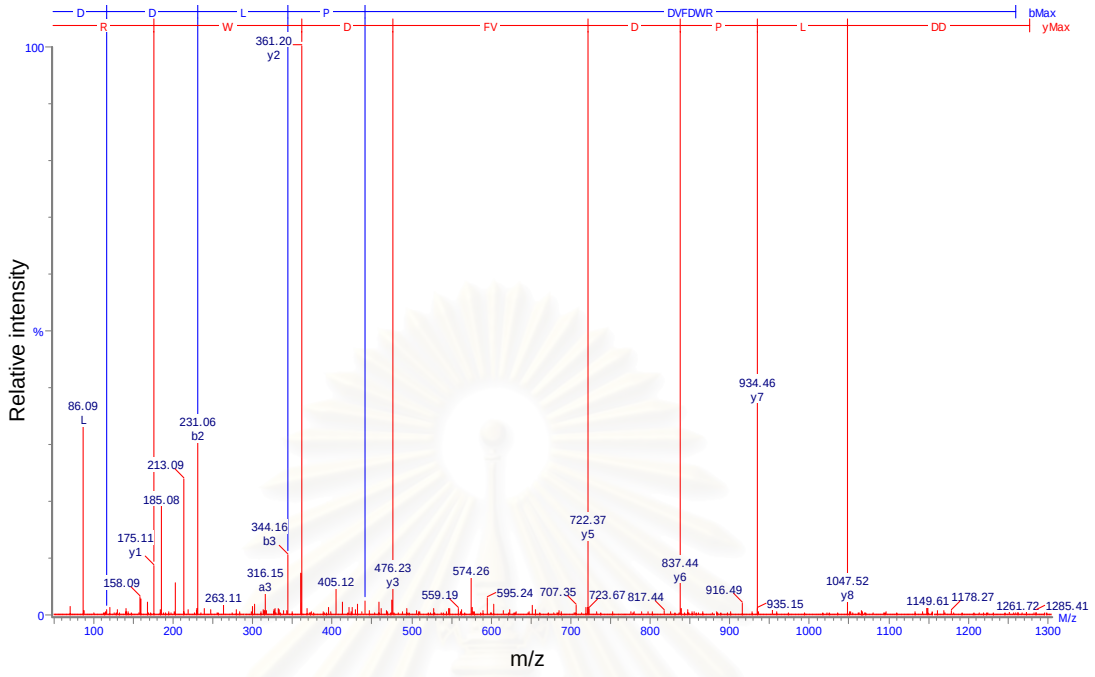



**Figure 4.20** MS/MS spectrum of doubly charged ion m/z of 550.70. *De novo* peptide sequence FWGDSGYLR was performed manually using the Peptide Sequencing tool implemented in the MassLynx (Micromass)

However, S2e was identified as senescence-associated cysteine protease (Q7XYU7) from *Anthurium andraeanum* with the sum of three top-scoring HSPs; 149 (Used 4 peptides; threshold score 1HSP=68, 2HSP=112, 3HSP=138) (Table 4.4). It was claimed as this protein rather than cysteine proteinase GP-I (P82473) due to sequence coverage can be increased by three peptide fragments (Figure 4.21-4.23) and their molecular weight consistent with S2/e position on 1D gel (Figure 4.16).

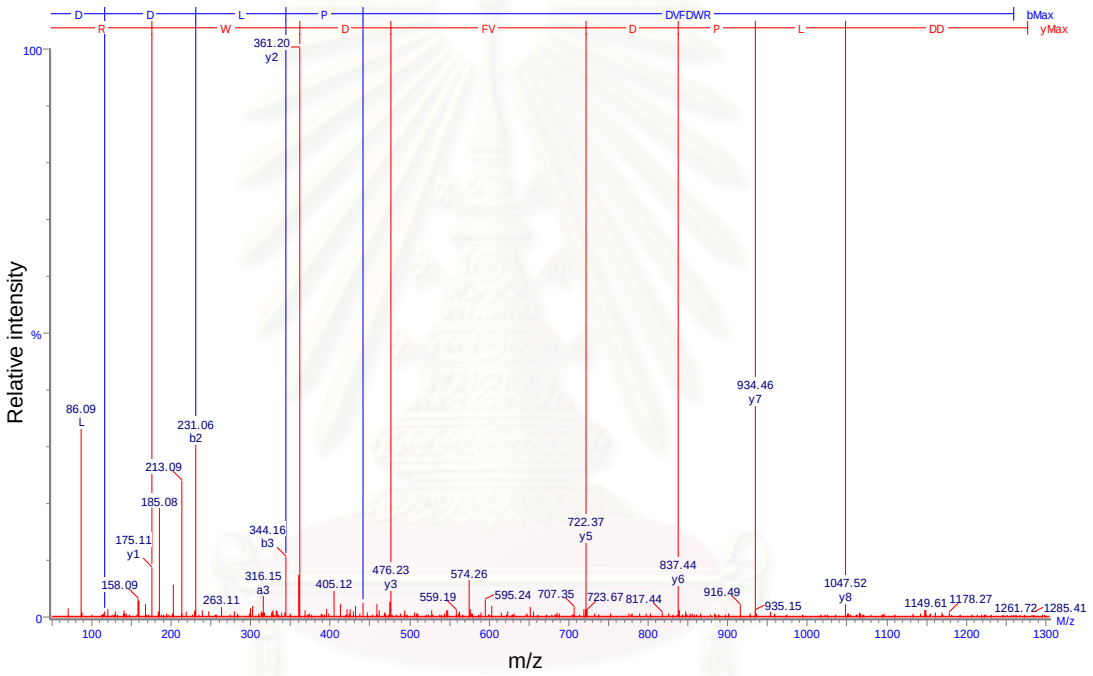



**Figure 4.21** MS/MS spectrum of doubly charged ion m/z of 639.29. *De novo* peptide sequence DDLPDVFDWR was performed manually using the Peptide Sequencing tool implemented in the MassLynx (Micromass).

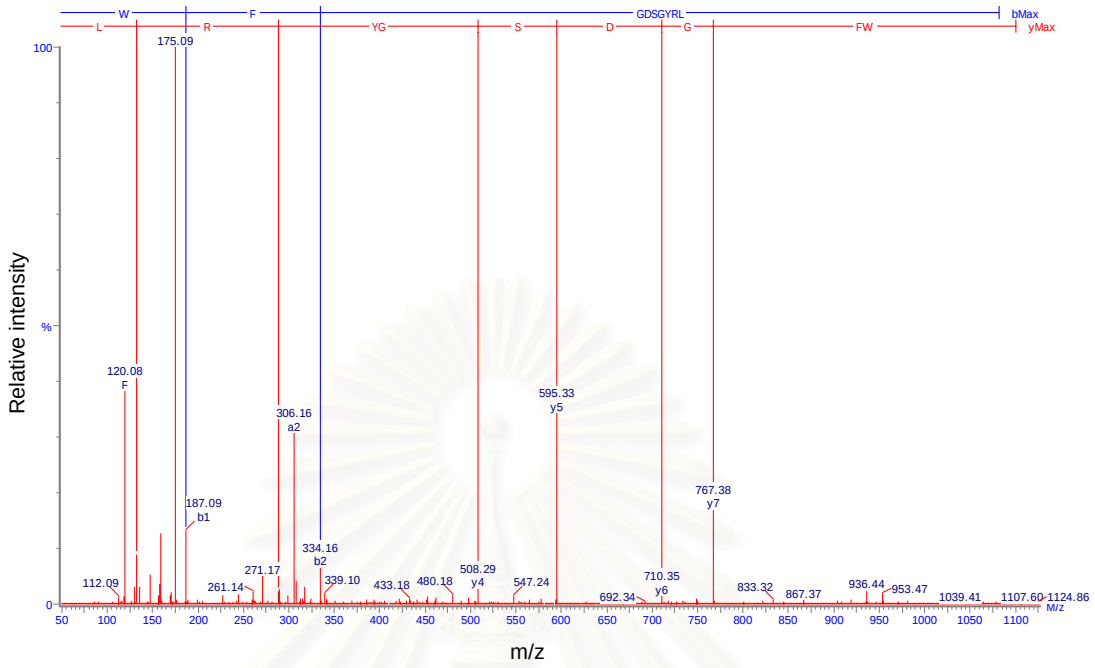



**Figure 4.22** MS/MS spectrum of doubly charged ion m/z of 550.76. *De novo* peptide sequence FWGDSGYRL was performed manually using the Peptide Sequencing tool implemented in the MassLynx (Micromass)

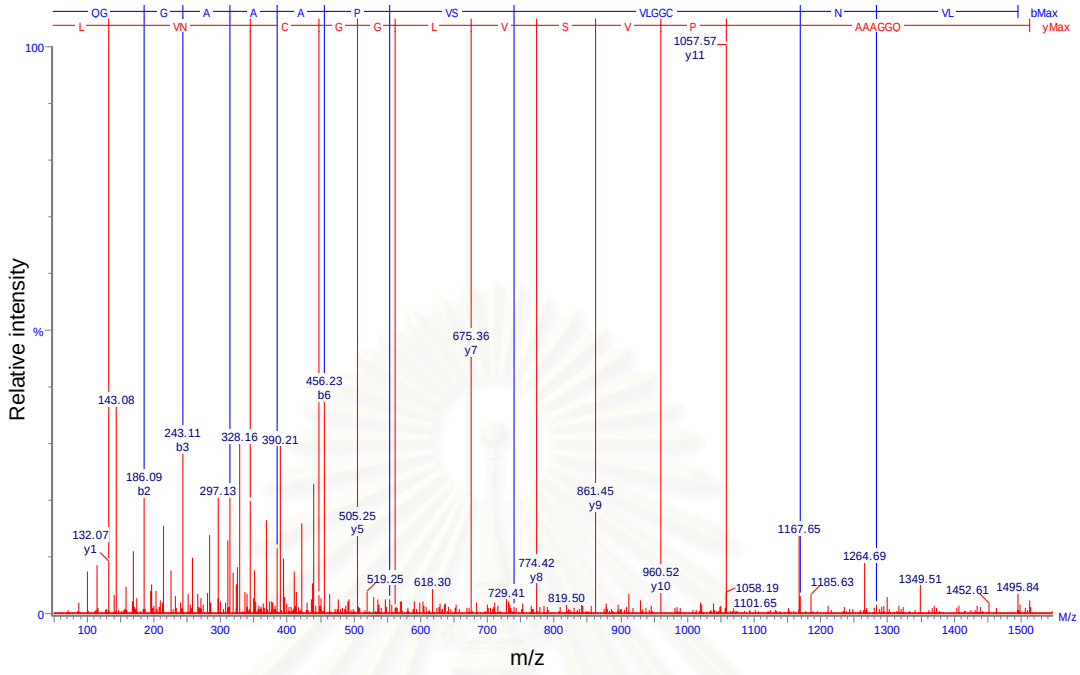



**Figure 4.23** MS/MS spectrum of doubly charged ion m/z of 756.95. *De novo* peptide sequence QGGAAAPVSVLGGCNVL was performed manually using the Peptide Sequencing tool implemented in the MassLynx (Micromass).

S1/a was identified as dimeric mannose specific lectin from *Narcissus hybrid cultivar*. The score of the top HSP peptide sequence, 57, is low than the threshold of statistical significance for a single matched HSP. The score of the second-ranked HSP sequence that is 54 was then added and compared the sum with the threshold for two matched HSPs that are equal 106. Now the score of the sum is 111 that significantly exceed the threshold and positive identification can be mentioned. (Used 20 peptides; threshold score 1HSP=67, 2HSP=106). In the same way, S1/b was identified as probable mannose binding lectin from *Chromobacterium violaceum* with the score of the top-ranked HSP; 66 (Used 11 peptides; threshold score 1HSP=65).

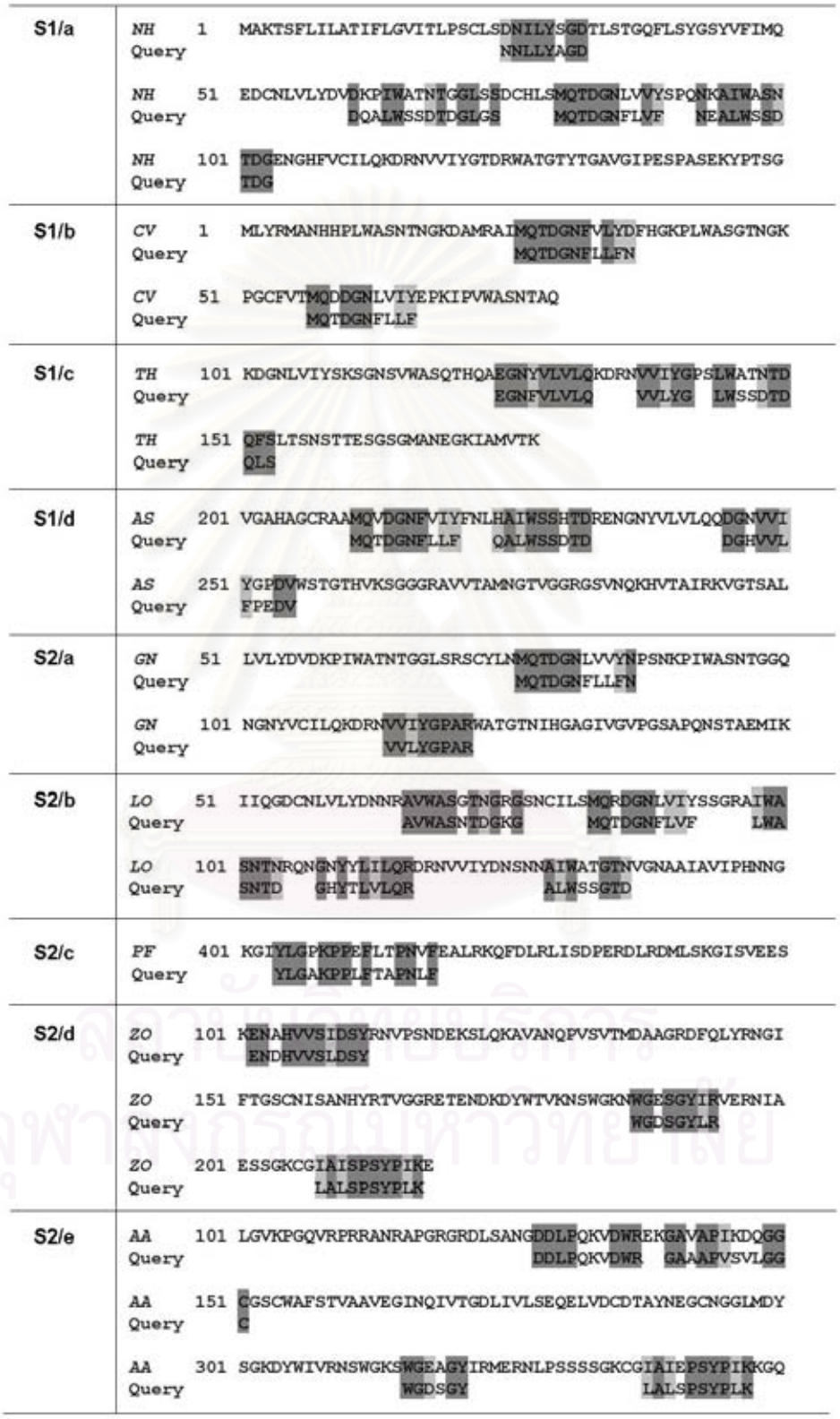
S1/c was identified as mannose-binding lectin from *Tulipa hybrid cultivar* (*92*) with the sum of two top-scoring HSPs; 102 (Used 12 peptides; threshold score 1HSP=66, 2HSP=102). S1/d was identified as lectin related protein from *Allium sativum* (*93*) with

the sum of two top-scoring HSPs; 102 (Used 11 peptides; threshold score 1HSP=65, 2HSP=101).

S2/a was identified as lectin from *Galanthus nivalis* (common snow drop) with the sum of three top-scoring HSPs; 112 (Used 14 peptides; threshold score 1HSP=66, 2HSP=103). S2/b was identified as mannose-binding protein from *Listera ovata* (*91*) with the sum of two top-scoring HSPs; 103 (Used 12 peptides; threshold score 1HSP=66, 2HSP=102). S2/c was identified as prismane protein from *Pyrococcus furiosus* with the score of the top-ranked HSP; 63 (Used 1 peptide; threshold score 1HSP=60). The comparison of amino acid sequences of the query from MS BLAST and from database was shown in Table 4.6.

This result showed six gel bands were identified as lectin that is consistent with the hemagglutinating activity contained in fraction S1 and S2.

**Table 4.6** Comparison of the amino acid sequences of query peptide from *Curcuma aromatica* and sequence from database




**Note**: Identities are highlighted on dark grey and similarities are highlighted on grey background.

## 4.5 Conclusion

This study demonstrated the identification of proteins from non-model plant, *Curcuma zedoaria* and *Curcuma aromatica*, was achieved by mass spectrometry method. Tandem mass spectra were exploited to gain some partial sequence information to allow sequence to be compared with MS BLAST algorithms to search for homology between a *de novo* determined peptide sequence and protein sequences available in database.

# CHAPTER V

# RELATIVE QUANTIFICATION OF PROTEIN EXPRESSION CHANGES IN ARABIDOPSIS GUARD CELLS FOLLOWING ABSCISIC ACID (ABA) TREATMENT

## 5.1 Introduction

There is an interest for understanding how plants respond to the changing environmental conditions. In the leaf epidermis, pairs of guard cells form stomatal pores, which provide $CO_2$ intake conducts for photosynthesis and sites for controlling transpirational water loss in plants. The guard cell system is a good model for studying the transduction of environmental and endogenous signals in plants because guard cells respond to various environmental factors such as blue light, temperature, $CO_2$, drought, abscisic acid (ABA), and other hormones. Environmental signals control the aperture of the stomatal pore by activating guard cell-signalling pathways that result in alterations to guard cell tugor. One of the best understood of these pathways is the abscisic acid (ABA)-signalling pathway (*47*). ABA builds up in the leaves during drought causing stomatal closure, and this enables the plant to conserve its existing supplies of water. In this study, changes of protein expression in guard cell enriched epidermal fragment of Arabidopsis were analyzed by two dimensional liquid chromatography method using iTRAQ reagent.

## 5.2 Aim

To evaluate the iTRAQ technique as a method of determining differences in protein expression in guard cell enriched protein from Arabidopsis with ABA-treated compared to ABA-untreated protein sample.

## 5.3 Methods

### 5.3.1 iTRAQ labelling of standard protein mixtures

Protein mixtures consisting of equimolar amounts (2.5 nmol) of alcohol dehydrogenase, α-casein, lysozyme, BSA, hexokinase (all from Sigma) were performed reduction, alkylation and digestion with trypsin (1:20 w/w, 50 mM triethylammonium bicarbonate (TEAB), 37 °C, 18 h) as described in section 3.4.7. Two equal aliquots (80µl, 100µg) were labelled with iTRAQ reagent 114 and 116. The two reactions were combined in various proportions (1:1, 1:2, 1:5, 1:10), dried by SpeedVac, resuspended in 10µl of 0.1% formic acid and each component were then analyzed by LC-MS/MS.

### 5.3.2 Sample preparation and labelling

Following the extraction step, guard cell dissolved in YeastBuster$^{TM}$ reagent (section 3.1) was exchanged to 0.5 M triethylammonium bicarbonate by ultrafiltration (MWCO 5 kDa). Protein concentration was determined by BCA assay (section 3.4.5.2). A 100 µg protein sample from each of the two extracts was reduced, alkylated, digested with trypsin and labelled individually with iTRAQ reagent 117 and 115 as described in section 3.4.7 and then combined in one mixture. The combined peptide mixture was dried by SpeedVac.

### 5.3.3 SCX Fractionation

Sample was resuspended in 1 ml of mobile phase A. (adjusted pH 2.5-3.3 by 10% phosphoric acid) and fractionated by strong cation exchange (SCX) chromatography as describe in section 3.3.2.. Sample was loaded and washed isocratically for 10 min at 1.0 ml/min to remove excess reagent. Peptides were eluted with a linear gradient of 0-40% mobile phase B over 80 min; 40-60% B over 20 min; 60-100% B over 5 min at a flow rate of 200 µl/min. Fractions were collected at one minute intervals (200 µl) and subsequently pooled where necessary. The collected fractions were dried by SpeedVac, dissolved in 10µl of 0.1% formic acid and analyzed by LC-MS/MS.

### 5.3.4 Nano-LC ESI MS/MS Analysis of SCX fractions

Individual or pooled SCX fractions were analyzed by LC-MS/MS. HPLC peptide separation was performed on an UltiMate HPLC system coupled to the Q-ToF instrument and samples were loaded as described previously in section 3.3.5. Peptides were separated using a linear gradient of 0-60% mobile phase B (95% ACN; 0.05% FA) over 30-105 minutes, then 60-95% B in 5 min. For MS/MS analysis, the collision energy range was ~20% higher than that used for unlabelled peptides to promote complete fragmentation of the labelled precursor ion.

### 5.3.5 Protein identification and data analysis

Data was processed using the ProteinLynx 1.0 module in MassLynx to generate peak list. Processed data was searched against SwissProt using an in-house MASCOT Server (Matrix Science). Quantification was analyzed by iTRACKER software (*94*). Briefly, the software takes as input non-centroided mass spectra and returns the relative ratios of each reporter ion.

Search parameters were as follows:

| | |
|---|---|
| Type of search: | MS/MS Ion Search |
| Enzyme: | Trypsin |
| Fixed modifications: | iTRAQ (K), iTRAQ (N-term), MMTS (Cys) |
| Variable modifications: | Oxidation (M) |
| Mass values: | Monoisotopic |
| Protein Mass: | Unrestricted |
| Peptide Mass Tolerance: | ± 200 ppm |
| Fragment Mass Tolerance: | ± 200 mmu |
| Max Missed Cleavages: | 1 |
| Instrument type: | ESI-QUAD-TOF |

## 5.4 Results and discussion

### 5.4.1 iTRAQ labelling of a standard protein mixture

In preliminary studies, proteins digest mixtures were used as simple model system to validate the labelling protocol and the usage of MS/MS signature ions for quantification. All proteins (alcohol dehydrogenase, α-casein, lysozyme, BSA, hexokinase) in the mixture were identified using Mascot searching. iTRAQ was submitted as a variable modification to assess the extent of completion of the labelling reaction. None of unmodified peptides were identified in the Mascot search indicating the reaction was complete.

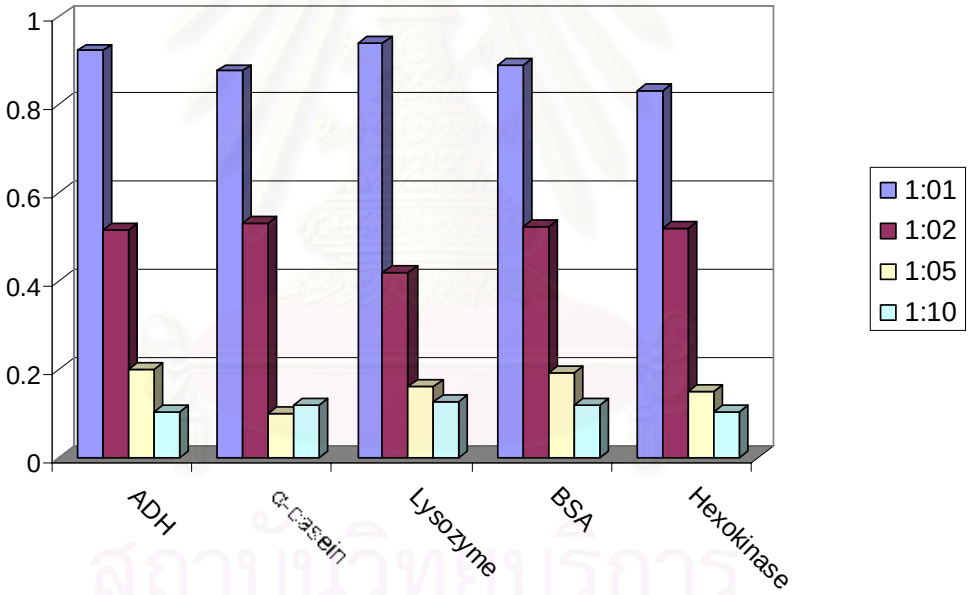



**Figure 5.1** The bar graph represents relative protein measurements for mixtures of a five protein digest. The ratio of ion counts for reporter ions 114 and 116 were expressed with four proportions, 1:1, 1:2, 1:5 and 1:10, respectively.

**Table 5.1** Ratio of five protein mixtures in four proportions

| Ratio | Protein | | | | | Mean | SD | CV (%) |
|---|---|---|---|---|---|---|---|---|
| | ADH | α-casein | Lysozyme | BSA | Hexokinase | | | |
| 1:1 | 0.922 | 0.876 | 0.941 | 0.889 | 0.831 | 0.892 | 0.038 | 4.3 |
| 1:2 | 0.514 | 0.530 | 0.419 | 0.522 | 0.519 | 0.501 | 0.041 | 8.2 |
| 1:5 | 0.200 | 0.100 | 0.162 | 0.191 | 0.150 | 0.161 | 0.036 | 22.1 |
| 1:10 | 0.104 | 0.119 | 0.125 | 0.118 | 0.102 | 0.114 | 0.009 | 7.8 |

A comparison of protein digest mixtures of known proportions (Figure 5.1) gave accurate ratios compared to the expected values and the coefficient of variation (CV) < 23% (Table 5.1) for each of the five different proteins analyzed.

A representative product ion spectrum of a peptide derived from BSA is shown in Figure 5.2.

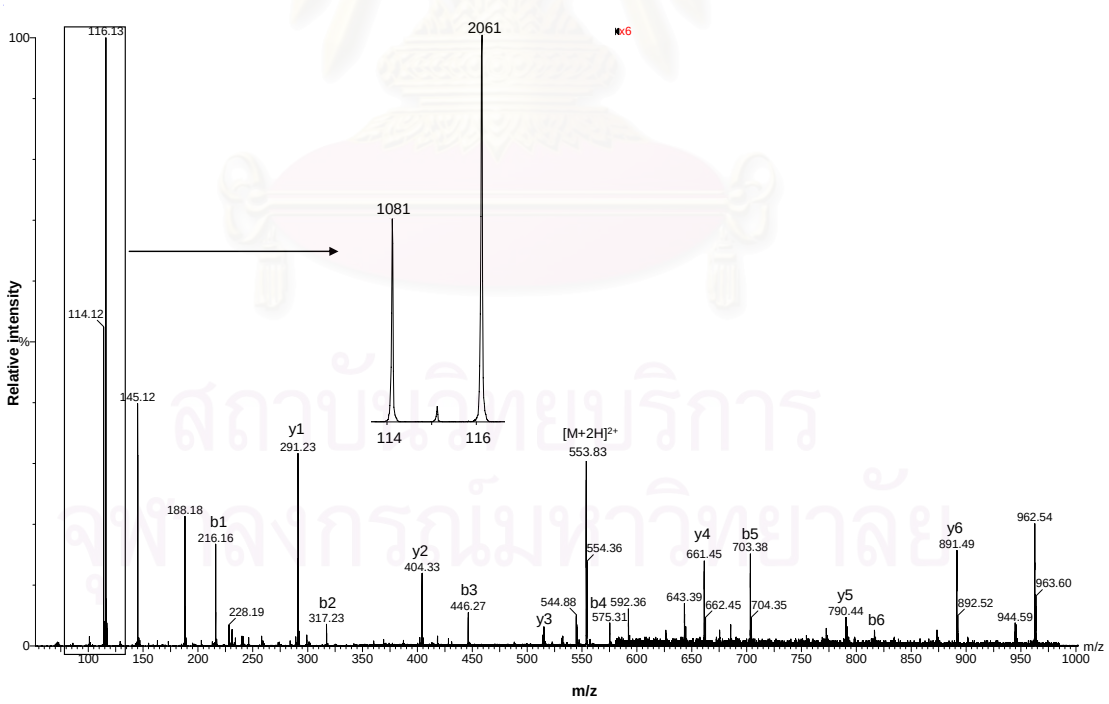



**Figure 5.2** MS/MS spectrum of precursor ion m/z of 553.83. The reportor ion intensities of iTRAQ reagents 114 and 116 were 1081 and 2061 ion counts, respectively.

### 5.4.2 iTRAQ labelling of guard cell enriched proteins

Initially, the complexity of guard cell enriched protein was investigated by 1D SDS-PAGE (Figure 5.3). The multiple band of protein was visualized in wide rang of molecular weight. Major proteins appeared in low molecular weight (< 30 kDa).

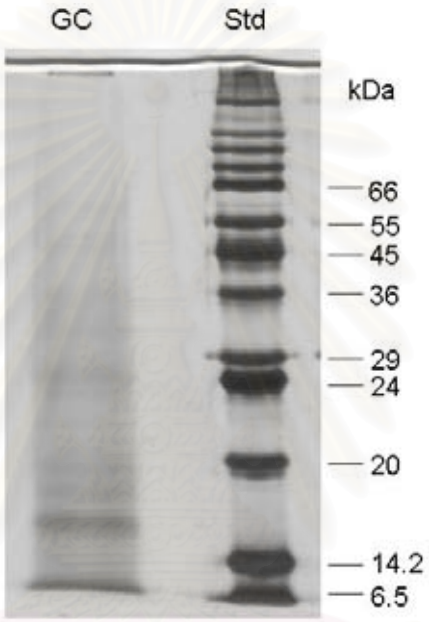



**Figure 5.3** 1D SDS-PAGE of guard cell enriched protein. (GC) is guard cell enriched protein. (Std) is standard marker protein same as in Figure 4.16.

Then, the two states of guard cell enriched protein were digested and labelled parallel with two iTRAQ reagent. A summary of the experimental scheme comparing guard cell extract prepared by treated with abscisic acid and treated with MeOH is diagrammed in (Figure 5.4).

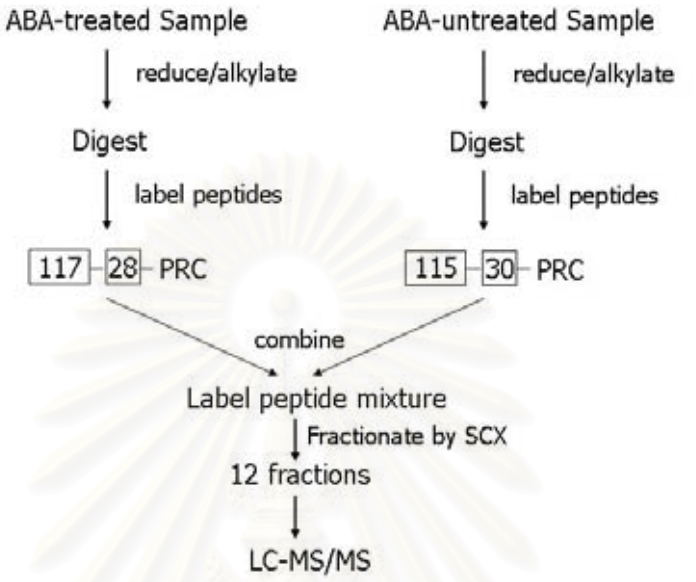



**Figure 5.4** The parallel workflow of experimental when using iTRAQ reagent.

After combining samples, the labelled peptide mixtures were fractionated by SCX into 12 fractions to reduce the peptide complexity and increase the capacity of peptide identification by mass spectrometry (Figure 5.5).

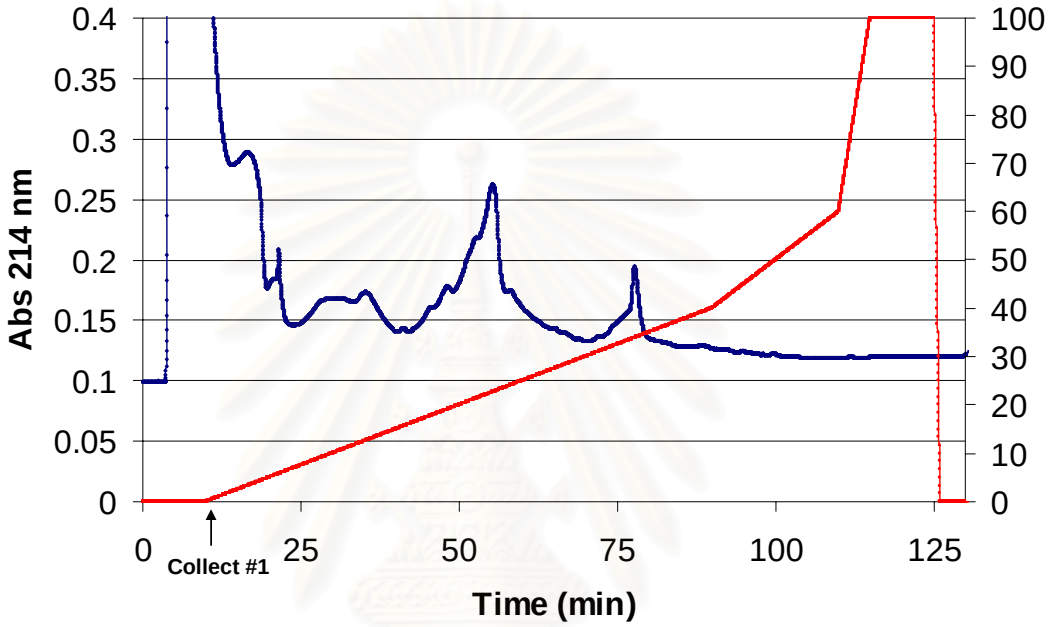



**Figure 5.5** SCX chromatogram demonstrating separation of iTRAQ labelled peptides of guard cell enriched proteins from Arabidopsis with and without ABA-treatment. The collection of fraction started at starting gradient; the salt gradient is represented by the red line.

The SCX profile (Figure 5.5) shows medium resolving power of the peptides from the labelled sample digest. A strong absorbance in the flow-through component of the chromatography indicated that residual iTRAQ reagent was removed prior to elution of peptide containing fractions.

For each SCX fraction analyzed by LC-MS, reporter ion intensities were extracted from the pkl file format using iTRACKER software (*94*). To date, published relative quantification work using iTRAQ reagents has been conducted using ProQuant and ProGroup software (Applied Biosystems), which is compatible with tandem MS data

generated from a QStar™ instrument, also from Applied Biosystems. The instrument used for LC-MS/MS in these experiments is a Q-TOF instrument, which generates similar tandem MS data, but in a different format. Peak lists (.pkl files), which are used to match peptide fragment ion masses to protein databases such as Mascot, are automatically generated within MassLynx Software.

In this experiment, 47 proteins were identified and quantified from the two treatment states of guard cell extract. These are listed in Table 5.2

**Table 5.2** Proteins and expression ratios of two states of guard cell enriched protein.

| No. | Accession No. | Protein description | Mass (Da) | Matches | 117:115 | SD |
|---|---|---|---|---|---|---|
| 1 | O03042 | Ribulose bisphosphate carboxylase large chain precursor | 56938 | 10 | 0.978 | 0.116 |
| 2 | P37702 | Myrosinase precursor | 66741 | 3 | 0.845 | 0.052 |
| 3 | P10795 | Ribulose bisphosphate carboxylase small chain 1A | 22306 | 1 | 0.820 | |
| 4 | P93819 | Malate dehydrogenase | 38994 | 5 | 0.990 | 0.061 |
| 5 | P10796 | Ribulose bisphosphate carboxylase small chain 1B | 22520 | 1 | 0.944 | |
| 6 | Q9T076 | Early nodulin-like protein 2 precursor | 38016 | 1 | 0.870 | |
| 7 | O49203 | Nucleoside diphosphate kinase III | 28404 | 2 | 0.867 | 0.025 |
| 8 | P25855 | Glycine cleavage system H protein 1 | 20045 | 1 | 1.141 | |
| 9 | Q9XI01 | Probable proteindisulfide-isomerase 1 precursor | 62570 | 1 | 0.933 | |
| 10 | P48491 | Triosephosphate isomerase | 30074 | 1 | 1.254 | |
| 11 | P46283 | Sedoheptulose-1,7-bisphosphate | 46744 | 1 | 0.795 | |
| 12 | Q9M1X0 | Ribosome recycling factor | 35107 | 2 | 1.865 | 0.212 |
| 13 | Q9SMU8 | Peroxidase 34 precursor | 40374 | 4 | 1.124 | 0.152 |
| 14 | P50318 | Phosphoglycerate kinase | 55462 | 2 | 0.854 | 0.004 |
| 15 | Q42547 | Catalase 3 | 60872 | 1 | 0.467 | |
| 16 | O22126 | Fasciclin-like arabinogalactan protein 8 precursor | 46887 | 1 | 1.039 | |
| 17 | Q42589 | Nonspecific lipid-transfer protein 1 precursor | 13216 | 1 | 1.246 | |
| 18 | O64903 | Nucleoside diphosphate kinase II | 27906 | 1 | 1.127 | |
| 19 | P10896 | Ribulose bisphosphate carboxylase/oxygenase activase | 56593 | 1 | 0.635 | |
| 20 | P94040 | Germin-like protein subfamily 3 member 1 precursor | 22934 | 2 | 0.535 | 0.482 |
| 21 | Q9LTS3 | Cytokinin dehydrogenase 3 precursor | | 1 | 0.876 | |
| 22 | P23321 | Oxygen-evolving enhancer protein 1-1 | 39484 | 2 | 1.782 | 0.195 |
| 23 | Q41932 | Oxygen-evolving enhancer protein 3-2 | 27078 | 1 | 1.704 | |
| 24 | P25858 | Glyceraldehyde-3- phosphate dehydrogenase | 41594 | 1 | 0.711 | |
| 25 | P25819 | Catalase 2 | 60774 | 1 | 0.469 | |
| 26 | P28493 | Pathogenesis-related protein 5 precursor | | 1 | 0.795 | |
| 27 | P27140 | Carbonic anhydrase | | 2 | 0.764 | 0.063 |
| 28 | Q9LHB9 | Peroxidase 32 precursor | | 1 | 0.766 | |
| 29 | O48737 | Thioredoxin M-type 1 | | 1 | 1.377 | |
| 30 | P59259 | Histone H4 | | 1 | 1.420 | |
| 31 | P59226 | Histone H3 | 17335 | 1 | 1.070 | |
| 32 | Q9SSK1 | Asparaginyl-tRNA synthetase, cytoplasmic 3 | | 1 | 1.030 | |
| 33 | P59263 | Ubiquitin | | 1 | 2.290 | |
| 34 | P29510 | Tubulin alpha-2/alpha-4 chain | | 1 | 2.290 | |
| 35 | P04777 | Chlorophyll a-b binding protein 165/180 | 30128 | 1 | 1.440 | |
| 36 | P56778 | Photosystem II 44 kda reaction centre | | 1 | 2.480 | |
| 37 | Q42029 | Oxygen-evolving enhancer protein 2-1 | 30954 | 1 | 2.450 | |
| 38 | Q96520 | Peroxidase 12 precursor | | 1 | 2.580 | |
| 39 | Q9XF89 | Chlorophyll a-b binding protein CP 26 | 719.41 | 1 | 2.080 | |
| 40 | P53496 | Actin 11 | 44714 | 1 | 1.550 | |
| 41 | P49107 | Photosystem I reaction centtre subunit N | 21097 | 1 | 2.100 | |
| 42 | Q42029 | Oxygen-evolving enhancer protein 2-1 | 30954 | 1 | 2.440 | |
| 43 | Q9S7H1 | Photosystem I reaction center subunit II-1 | 25224 | 1 | 1.000 | |
| 44 | P13905 | Elongation factor 1-alpha | 56952 | 1 | 2.300 | |
| 45 | Q9XFT3 | Oxygen evolving enhancer protein 3-1 | 26663 | 1 | 1.000 | |
| 46 | Q42534 | Pectinesterase 2 precursor | 71031 | 3 | 2.937 | 0.272 |
| 47 | P17745 | Elongation factor Tu | P17745 | 2 | 2.59875 | 0.479 |

Note number of matches refers to the number of peptides corresponding to the same protein.

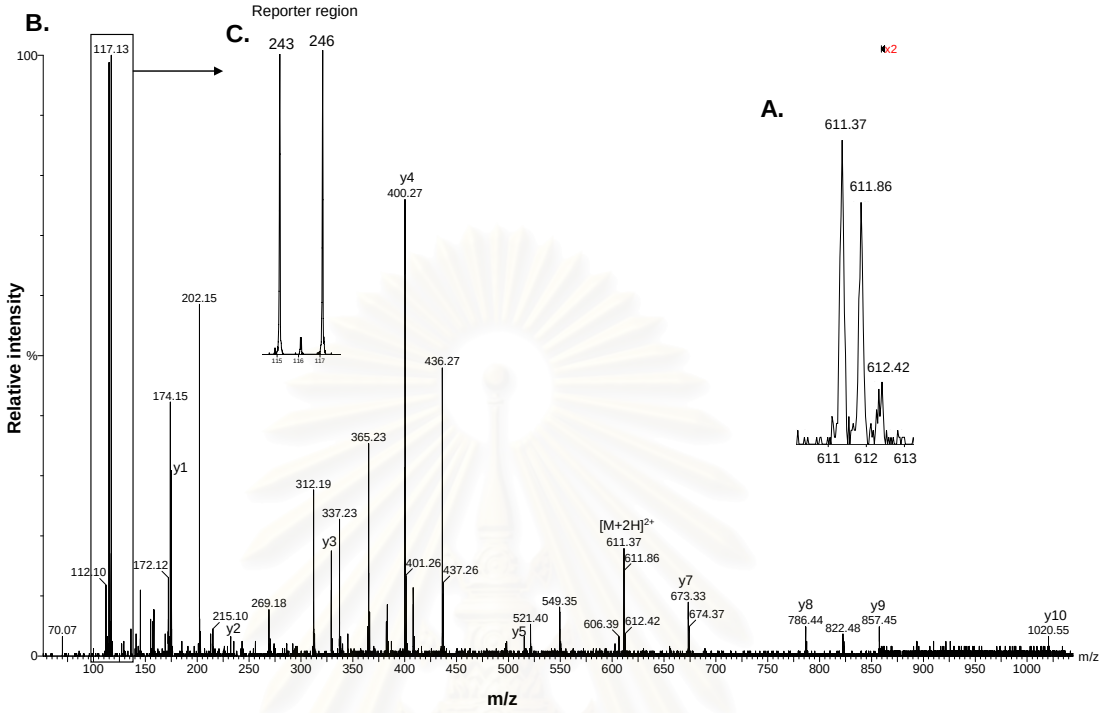



**Figure 5.6** MS and MS/MS spectra from a duplex sample labelled with two iTRAQ reagents showing **A.** doubly charged precursor ion m/z of 611.37. **B.** MS/MS spectrum corresponding to GYALGTDAPGR **C.** Two diagnostic reporter ions 115 and 117.

Figure 5.6 illustrates an example of a tandem mass spectrum of iTRAQ reagent labelled peptide for identification of the Myrosinase precursor (P37702). The reporter ions m/z of 115 and 117 represent the relative levels of this peptide for the ABA-treated and untreated samples with ratio 117:115 equal 246: 246. Since two isotopic labels are isobaric, the precursor is observed as a single component. In MS/MS spectra, the y and b ions were enhanced since they correspond to sum of ion current from the combination of two samples. This represents the benefit for analysis of large-scale proteome that normally contain wide range of peptide concentrations.

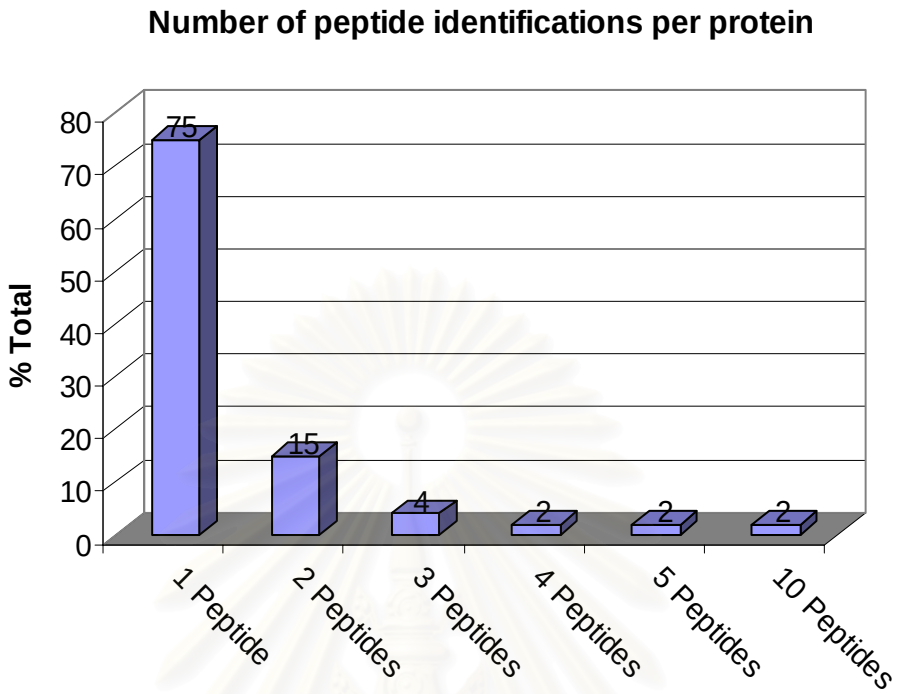



**Figure 5.7** Distribution of Arabidopsis proteins identifications as a function of the number of confidently identified peptides.

When precursor ions with associated reporter ion intensities of <10 counts were found, it has been excluded from this reported dataset. Then, protein coverage was examined (Figure 5.7) and revealed that from 47 identified proteins, 75% of these proteins were identified by a single significant peptide per protein and the rest were identified by at least two significant peptides of the same protein. It is necessary to increase peptide coverage to high confidential level for identification in future work.

All of the chloroplast located proteins are most likely to come from mesophyll cell contamination that needs to study the influence of guard cell extraction in future work. Myrosinase precursor (P37702), Nucleoside diphosphate kinase III (O49203) Glyceraldehyde-3-phosphate dehydrogenase (P25858) and Chlorophyll a-b binding protein CP 26(Q9XF89) are known to be more highly expressed in guard cells, based upon previously-published transcriptome data (*95*).

### 5.4.3 Reproducibility of reporter ion ratios

Reproducibility of reporter ion ratios for iTRAQ labelled peptides is important for the determination of relative protein expression. In this section, the reproducibility of reporter ion ratios for iTRAQ labelled peptides within the same LC-MS analysis and between LC-MS analysis were investigated. The coefficient of variation (CV) of reporter ion ratios was calculated for:

1) Peptides derived from the same protein, analyzed during the same LC-MS run (Table 5.3)
2) Identical peptides analyzed during different LC-MS runs (Table 5.4)

**Table 5.3** Coefficient of variation of reporter ion ratios corresponding to different peptides from the same proteins

| Protein | SCX # | Precursor | Charge state | Intensity 117.12 | Intensity 115.11 | Ratio | Mean | SD | CV (%) |
|---|---|---|---|---|---|---|---|---|---|
| Ribulose bisphosphate carboxylase large chain precursor | 25 | 522.286 | 2 | 415 | 435 | 0.95 | | | |
| | | 583.338 | 2 | 3553 | 3805 | 0.93 | | | |
| | | 625.357 | 2 | 230 | 203 | 1.13 | 0.94 | 0.14 | 14.8 |
| | | 686.910 | 2 | 417 | 441 | 0.94 | | | |
| | | 697.438 | 2 | 40 | 54 | 0.74 | | | |
| Peroxidase 34 | 25 | 474.309 | 2 | 146 | 137 | 1.06 | | | |
| | | 559.361 | 2 | 590 | 549 | 1.07 | 1.03 | 0.07 | 7.0 |
| | | 590.337 | 2 | 76 | 81 | 0.94 | | | |
| | 28 | 444.277 | 2 | 399 | 333 | 1.20 | | | |
| | | 474.262 | 2 | 486 | 516 | 0.94 | 1.14 | 0.18 | 15.7 |
| | | 590.299 | 2 | 74 | 58 | 1.29 | | | |
| Malate dehydrogenase | 31 | 509.281 | 2 | 123 | 134 | 0.92 | | | |
| | | 553.310 | 2 | 263 | 247 | 1.07 | 1.00 | 0.08 | 7.7 |
| | | 616.352 | 2 | 515 | 503 | 1.02 | | | |
| | 34 | 430.783 | 2 | 530 | 537 | 0.99 | | | |
| | | 506.280 | 2 | 215 | 223 | 0.96 | 0.96 | 0.03 | 3.2 |
| | | 553.303 | 2 | 277 | 299 | 0.93 | | | |

**Table 5.4** Coefficient of variation for reporter ion ratios corresponding to the same peptide from proteins analyzed in different LC-MS runs.

| Protein | SCX # | Precursor | Charge state | Intensity 117 | Intensity 115 | Ratio 117:115 | Mean | SD | CV (%) |
|---|---|---|---|---|---|---|---|---|---|
| Ribulose bisphosphate carboxylase large chain precursor | 25 | 522.286 | 2 | 415 | 435 | 0.95 | | | |
| | 28 | 522.246 | 2 | 1332 | 1387 | 0.96 | 0.92 | 0.07 | 7.6 |
| | 31 | 522.242 | 2 | 450 | 537 | 0.84 | | | |
| | 25 | 583.350 | 2 | 480 | 445 | 1.08 | | | |
| | 28 | 583.288 | 2 | 1175 | 1285 | 0.91 | 1.01 | 0.09 | 8.5 |
| | 31 | 583.299 | 2 | 39 | 38 | 1.04 | | | |
| Peroxidase 34 | 25 | 474.309 | 2 | 146 | 137 | 1.06 | 1 | 0.09 | 8.5 |
| | 28 | 474.262 | 2 | 486 | 516 | 0.94 | | | |
| | 25 | 590.337 | 2 | 76 | 81 | 0.94 | 1.12 | 0.25 | 21.9 |
| | 28 | 590.299 | 2 | 74 | 58 | 1.29 | | | |
| Malate dehydrogenase | 28 | 616.349 | 2 | 106 | 115 | 0.92 | 0.97 | 0.07 | 7.4 |
| | 31 | 616.352 | 2 | 515 | 503 | 1.02 | | | |
| | 28 | 509.273 | 2 | 352 | 389 | 0.91 | 0.91 | 0.01 | 0.9 |
| | 31 | 509.281 | 2 | 123 | 134 | 0.92 | | | |

Significant peptides from Mascot hits for Ribulose bisphosphate carboxylase large chain precursor, Peroxidase 34 and Malate dehydrogenase were used to calculate the CV. In this example, the CV value for a number of peptides from the same protein (Table 5.3) was highest when the low intensity product ion spectra were increased. The CV for identical peptides analyzed in different LC-MS experiments run on the same or different days was below 22% (Table 5.4). This is comparable to published studies of Choe et al. (*80*) and demonstrates a high level of reproducibility.

In this study, the important issues for the number of protein identifications and differential reporter ion ratios were described:

1. Identical peptides appeared in adjacent SCX fractions reflects the limited resolution of the SCX column.

2. The co-elution of numerous different peptides in analyte mixture from the RP column reflects a proportion of the precursor ions in each.

3. The automatic switching function of the Q-TOF selects the three most abundant ions for fragmentation in each MS survey scan. Therefore, high abundance peptides which elute from the RP column will be selected repeatedly for MS/MS. An attempt to

reduce this was made by increasing the exclusion time within the Mass Lynx software. This setting prevents the same precursor ion m/z value being selected for MS/MS.

4. The reporter ion intensity is related to the concentration of the peptide at the time point it is selected for MS/MS. For instance, peptides selected for MS/MS at the peak of the ion chromatogram will display better ion statistics than peptides selected at the tail end of the elution peak for that peptide.

5. There were some product ion spectra which contained either very poor intensity reporter ions or were deficient of reporter ions, but had a sufficient *b* and *y* ion series to be matched to a peptide sequence by Mascot. Despite appropriate identification, relative quantification was not possible for these precursor ions. A Mascot database search is a qualitative analysis and does not account for the intensity of the product ion spectrum. Therefore, a sequence match from a low intensity product ion spectrum is still possible despite the absence of relative quantification data from the iTRAQ reporter ions.

## 5.5 Conclusion

The validation experiments described in this chapter demonstrate the successful application of the iTRAQ labelling protocol to a standard protein mixture digest. Within this work, we were able to identify and quantify 47 proteins from the two extracts; however the number of protein identifications with significant Mascot scores should be increased as discussed above. Also, this study revealed iTRAQ labelling combined with tandem mass spectrometry is a powerful method to identify and quantify the changes of protein expression. In addition, the benefit of iTRAQ technology is suitable for high-throughput protein identification and quantification of multiplexed protein samples in a single mass spectrometric experiment.

# CHAPTER VI

# CONCLUSIONS AND FUTURE WORK

The overall aim of the work in this thesis was to analyse plant protein using tandem mass spectrometry. The different approaches to the problem were taken in two chapters: a small-scale, targeted approach and a large-scale, proteomic approach. Chapter IV described the characterization of bioactive protein from plant of interest, *Curcuma* species. Chapter V described the development of isobaric reagent (iTRAQ) and two-dimensional liquid chromatography (2D-LC) approaches for large-scale protein identification. In this section the results from each chapter will be summarized and future work will be outlined.

## 6.1 Identification of protein from Curcuma plant

The objective of this research was the characterization of biological active protein from interest medicinal plant, *Curcuma* species. No previous work reported the active protein or protein component from this plant and also it lack of genome database. In this study, mannose binding lectin with hemagglutinating activity to rabbit erythrocyte was purified from *Curcuma zedoaria* (Section 4.4.1). For *Curcuma aromatica* (Kanthamala), protein mixture in hemagglutinating containing fraction; S1 and S2 (Section 4.4.2) was identified following separated by 1D SDS-PAGE revealed that they compose of protein with similar sequence to various lectins. The purification of these fractions was needed in further study.

In addition, high mass protein in fraction Q1S1 revealed the containing of potent inhibitory activity to α-glucosidase. However, the identification of this protein was not complete and further study is required.

## 6.2 Relative quantification of protein expression changes in Arabidopsis guard cells following abscisic acid (ABA) treatment

The other aspect of this research was to analyze changes in guard cell enriched proteins of Arabidopsis before and after ABA treatment, using iTRAQ. This technique is based upon derivatization of primary amino groups (N-termini, Lysine residues) with a specific isobaric tag following proteolysis with trypsin. These tags fragment during tandem MS of the derivatized peptide generating a low molecular mass reporter ion that is unique to the tag used to label each of the digests. The isotope-labelled product ions are present in proportions characteristic of the initial analyte concentrations. Measurement of the intensity of these reporter ions enables relative quantification of the peptides in each digest and hence the proteins from where they originate. ABA-induced changes in the relative expression levels of the proteins identified following database searching of the tandem MS spectra from Arabidopsis guard cells was reported in this study.

However, for improve the statistic reliable of this approach the further work was suggested by duplicate the experiment by labelled ABA-treated sample (A) with iTRAQ reagent 114, 115 and ABA-untreated sample (C) with 116, 117. Theoretically, the reporter ion ratio for the duplex labelled samples (114:115 (A) and 116:117 (C)) ratio should always equal 1.

# REFERENCES


[1] Wasinger, V.C., Cordwell, S.J., Cerpa-Poljak, A., Yan, J.X., Gooley, A.A., Wilkins, M.R., Duncan, M.W., Harris, R., Williams, K.L. and Humphery Smith, I. Progress with gene-product mapping of the Mollicutes: *Mycoplasma genitalium*. *Electrophoresis* 16 (1995): 1090-1094.

[2] Fenn, J.B., Mann, M., Meng, C.K., Wong S.F. and Whitehouse, C.M. Electrospray ionization for mass spectrometry of large biomolecules. *Science* 246 (1989): 64-71.

[3] Cole, R.B. *Electrospray Ionization Mass Spectrometry: Fundamentals, Instrumentation and Applications.* New York: John Wiley & Sons, 1997.

[4] Karas, M. and Hillenkamp, F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Analytical Chemistry* 60 (1988): 2299-2301.

[5] Mann, M., Hoejrup, P. and Roepstorff, P. Use of mass spectrometric molecular weight information to identify proteins in sequence databases. *Biological Mass Spectrometry* 22 (1993): 338–345.

[6] Zhang, W. and Chait, B. ProFound: An expert system for protein identification using mass spectrometric peptide mapping information. *Analytical Chemistry* 72 (2000): 2482–2489.

[7] Perkins, D., Pappin, D.J.C., Creasy, D.M. and Cottrell, J.S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20 (1999): 3551–3567.

[8] Clauser, K.R., Baker, P. and Burlingame, A.L. Role of accurate mass measurement (+/-10 ppm) in protein identification strategies employing MS or MS/MS and database searching. *Analytical Chemistry* 71 (1999): 2871-2882.

[9] Egelhofer, V., Bussow, K., Luebbert, C., Lehrach, H. and Nodhof, E. Improvements in protein identification by MALDI-TOF-MS peptide mapping. *Analytical Chemistry* 72 (2000): 2741–2750.

[10] Eng, J.K., McCormack, A.L. and Yates, J.R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *Journal of the American Society for Mass Spectrometry* 5 (1994): 976–989.

[11] Mo, M., Tse, Y.C. and Jiang, L. Organelle identification and proteomics in plant cells. *Trends in Biotechnology* 21 (2003): 331–332.

[12] Gallardo, K., Job, C., Groot, S.P.C., Puype, M., Demol, H., Vandekerckhove, J. and Job, D. Proteomic analysis of *Arabidopsis thaliana* seed germination and priming. *Plant Physiology* 126 (2001): 835–849.

[13] Kersten, B., Burkle, L., Kuhn, E.J., Giavalisco, P., Konthur, Z., Lueking, A., Walter, G., Eickhoff, H. and Schneider, U. Large-scale plant proteomics. *Plant Molecular Biology* 48 (2002): 133–141.

[14] Newton, R.P., Brenton, A.G., Smith, C.J. and Dudley, E. Plant proteome analysis by mass spectrometry: principles, problems, pitfalls and recent developments. *Phytochemistry* 65 (2004): 1449–1485.

[15] Rossignol, M. Analysis of the plant proteome. *Current Opinion in Plant Biotechnology* 12 (2001): 131–134.

[16] Fukuda, M., Islam, N., Woo, S.H., Yamagishi, A., Takaoka, M. and Hirano, H. Assessing matrix assisted laser desorption/ionization-time of flight-mass spectrometry as a means of rapid embryo protein identification in rice. *Electrophoresis* 24 (2003): 1319–1329.

[17] Gallardo, K., Job, C., Groot, S.P., Puype, M., Demol, H., Vandekerckhove, J. and Job, D. Proteomics of Arabidopsis seed germination. A comparative study of wild-type and gibberellin deficient seeds. *Plant Physiology* 129 (2002): 823–837.

[18] Touzet, P., Riccardi, F., Morin, C., Damerval, C., Huet, J.C., Pernollet, J.C., Zivy, M., and DeVienne, D. The maize two dimensional gel protein database: towards an integrated genome analysis program. *Theoretical and Applied Genetics* 93 (1996): 997–1005.

[19] Watson, B.S., Asirvatham, V.S., Wang, L. and Summer, L.W. Mapping the proteome of barrel medic (*Medicago truncatula*). *Plant Physiology* 131 (2003): 1104–1123.

[20] Shen, S., Sherma, A. and Komatsu, S. Characterization of proteins responsive to gibberellin in the leaf-sheath of rice (*Oryza sativa* L.) seedling using proteome analysis. *Biological & Pharmaceutical Bulletin* 26 (2003): 129–136.

[21] Konishi, H., Ishiguro, K. and Komatsu, S. A proteomics approach towards understanding blast fungus infection of rice grown under different levels of nitrogen fertilization. *Proteomics* 1 (2001): 1162–1167.

[22] Tafforeau, M., Verdus, M.C., Charlionet, R., Cabin-Flaman, A. and Ripoll, C. Two-dimensional electrophoresis investigation of short-term response of flax seedlings to a cold shock. *Electrophoresis* 23 (2002): 2534–2540.

[23] Majoul, T., Bancel, E., Triboi, E., Ben H. J. and Branlard, G. Proteomic analysis of the effect of heat stress on hexaploid wheat grain: Characterization of heat-responsive proteins from total endosperm. *Proteomics* 3 (2003): 175–183.

[24] Salekdeh, G.H., Siopongco, J., Wade, L.J., Ghareyazie, B. and Bennett, J. Proteomic analysis of rice leaves during drought stress and recovery. *Proteomics* 2 (2002): 1131–1145.

[25] Kress, W.J., Prince, L.M. and Williams, K.J. The phylogeny and a new classification of the gingers (Zingiberaceae): Evidence from molecular data. *American Journal of Botany* 89 (2002): 1682-1696.

[26] Foster, S. Ginger your food is your medicine [online]. Available from: http://www.stevenfoster.com/education/monograph/ginger.html.

[27] Wu, W.Y., Xu, Q., Shi, L.C. and Zhang, W.B. Inhibitory effects of *Curcuma aromatica* oil on proliferation of hepatoma in mice. *World Journal of Gastroenterology* 6 (2000): 216-219.

[28] Brinkhaus, B., Hentschel, C., Von Keudell, C., Schindler, G., Lindner, M., Stutzer, H., Kohnen, R., Willich, S.N., Lehmacher, W. and Hahn, E.G. Herbal medicine with curcuma and fumitory in the treatment of irritable bowel syndrome: A randomized, placebo-controlled, double-blind clinical trial. *Scandinavian Journal of Gastroenterology* 40 (2005): 936-943.

[29] Chearwae, W., Wu, C.P., Chu, H.Y., Lee, T.R., Ambudkar, S.V. and Limtrakul, P. Curcuminoids purified from turmeric powder modulate the function of human multidrug resistance protein 1 (ABCC1). *Cancer Chemotherapy and Pharmacology* 57 (2006): 376-388.

[30] Jang, M.K., Sohn, D.H. and Ryu, J.H. A curcuminoid and sesquiterpenes as inhibitors of macrophage TNF-alpha release from *Curcuma zedoaria*. *Planta Medica* 67 (2001): 550-552.

[31] Kim, J.K., Jo, C., Hwang, H.J., Park, H.J., Kim, Y.J. and Byun, M.W. Color improvement by irradiation of *Curcuma aromatica* extract for industrial application. *Radiation Physics and Chemistry* 75 (2006): 449-452.

[32] Matsuda, H., Ninomiya, K., Morikawa, T. and Yoshikawa, M. Inhibitory effect and action mechanism of sesquiterpenes from Zedoaria rhizoma on D-galactosamine/lipopolysaccharide-induced liver injury. *Bioorganic & Medicinal Chemistry Letters* 8 (1998): 339-344.

[33] Lai, E.Y.C., Chyau, C.C., Mau, J.L., Chen, C.C., Lai, Y.J., Shih, C.F. and Lin, L.L. Antimicrobial activity and cytotoxicity of the essential oil of *Curcuma zedoaria*. *American Journal of Chinese Medicine* 32 (2004): 281-290.

[34] Syu, W.J., Shen, C.C., Don, M.J., Ou, J.C., Lee, G.H. and Sun, C.M. Cytotoxicity of curcuminoids and some novel compounds from *Curcuma zedoaria*. *Journal of Natural Products* 61 (1998): 1531-1534.

[35] Mau J.L., Lai E.Y.C., Wang N.P., Chen C.C., Chang C.H. and Chyau C.C. Composition and antioxidant activity of the essential oil from *Curcuma zedoaria*. *Food Chemistry* 82 (2003): 583-591

[36] Kim, K.I., Kim, J.W., Hong, B.S., Shin, D.H., Cho, Y., Kim, H.K. and Yang, H.C. Antitumor, genotoxicity and anticlastogenic activities of polysaccharide from *Curcuma zedoaria*. *Molecules and Cells* 10 (2000): 392-398.

[37] Sangvanich P., Kaeothip S., Srisomsap C., Thiptara P., Petsom A., Boonmee A. and Svasti J. Hemagglutinating activity of *Curcuma* plants. *Fitoterapia* 78 (2007): 29-31.

[38] Van Damme, E.J.M. *Handbook of plant lectins properties and biomedical applications.* Chichester: John Wiley & Sons, 1998.

[39] Van Damme, E.J.M., Peumans, W.J., Barre, A. and Rouge, P. Plant lectins: A composite of several distinct families of structurally and evolutionary related proteins with diverse biological roles. *Critical Reviews in Plant Sciences* 17 (1998): 575-692.

[40] Vasconcelos, I.M. and Oliveira, J.T.A. Antinutritional properties of plant lectins. *Toxicon* 44 (2004): 385-403.

[41] Christiane, B., Claus, M.L. and John, F.W. Lectin-mediated drug targeting: history and applications. *Advanced Drug Delivery Reviews* 56 (2004): 425-435.

[42] Earl, P.L., Doms, R.W. and Moss, B. Oligomeric structure of the human immunodeficiency virus type 1 envelope glycoprotein. *Proceedings of the National Academy of Sciences of the United States of America* 87 (1990): 648-652.

[43] Pinter, A., Honnen, W., Shermaine, A.T., Bona, C., Zaghouani, H., Gorny, M.K. and Pazner, S.Z. Oligomeric structure of gp4l, the transmembrane protein of human immunodeficiency virus type 1. *The Journal of Virology* 63 (1989): 2674-2679.

[44] Balzarini, J., Schols, D., Neyts, J., Van Damme, E., Peumans, W. and De Clercq, E. α-(1–3)- and α-(1–6)-D-mannose-specific plant lectins are markedly inhibitory to human immunodeficiency virus and cytomegalovirus infections *in vitro*. *Antimicrobial Agents and Chemotherapy* 35 (1991): 410-416.

[45] Balzarini, J., Neyts, J., Schols, D., Hosoya, M., Van Damme, E., Peumans, W. and De Clercq, E. The mannose-specific plant lectins from *Cymbidium hybrid* and *Epipactis helleborine* and the (*N*-acetylglucosamine) n-specific plant lectin from *Urtica dioica* are potent and selective inhibitors of human immunodeficiency virus and cytomegalovirus replication *in vitro*. *Antiviral Research* 18 (1992): 191–207.

[46] Ng, T.B. and Wang, H. Panaxagin, a new protein from Chinese ginseng possesses anti-fungal, anti-viral, translation-inhibiting and ribonuclease activities. *Life Sciences* 68 (2001): 739–749.

[47] Hetherington, A.M. Guard cell signaling. *Cell* 107 (2001): 711-714.

[48] McPherson, J.D., Marra, M., Hillier, L., Waterston, R.H., Chinwalla, A., Wallis, J., et al. A physical map of the human genome. *Nature* 409 (2001): 934–941.

[49] Poll, D.J. and Harding, D.R. Formic acid as a milder alternative to trifluoroacetic acid and phosphoric acid in two-dimensional peptide mapping. *Journal of Chromatography* 469 (1989): 231-239.

[50] Thevenon, G. and Regnier, F.E. Reversed-phase liquid chromatography of proteins with strong acids. *Journal of Chromatography* 476 (1989): 499-511.

[51] Kaufmann, R. Matrix-assisted laser desorption ionization (MALDI) mass spectrometry: A novel analytical tool in molecular biology and biotechnology. *Journal of Biotechnology* 41 (1995): 155-175.

[52] Gaskell, S. J. Electrospray: Principles and Practice. *Journal of Mass Spectrometry* 32 (1997): 677-688.

[53] Hoffmann, E.D. and Stroobant, V. *Mass Spectrometry Principles and Applications.* Chichester, UK: John Wiley & Sons, 2002, 63-132.

[54] Krutchinsky, A.N., Zhang, W. and Chait, B.T. Rapidly switchable matrix-assisted laser desorption/ionization and electrospray quadrupole-time-of-flight mass spectrometry for protein identification. *Journal of the American Society for Mass Spectrometry* 11 (2000): 493-504.

[55] Shevchenko, A., Loboda, A., Ens, W. and Standing, K.G. MALDI quadrupole time-offlight mass spectrometry: A powerful tool for proteomic research. *Analytical Chemistry* 72 (2000): 2132-2141.

[56] Loboda, A.V., Thomson, B.A. and Chernushevich, I.V. An introduction to quadrupole–time-of-flight mass spectrometry. *Journal of Mass Spectrometry* 36 (2001): 849-865.

[57] Morris, H.R., Paxton, T., Dell, A., Langhorne, J., Berg, M., Bordoli, R.S., Hoyes, J., and Bateman, R.H. High sensitivity collisionally-activated decomposition tandem mass spectrometry on a novel quadrupole/orthogonal-acceleration time-of-flight mass spectrometer. *Rapid Communications in Mass Spectrometry* 10 (1996): 889-896.

[58] Bienvenut, W.V., Deon, C., Pasquarello, C., Campbell, J.M., Sanchez, J.C., Vestal, M.L. and Hochstrasser, D.F. Matrix-assisted laser desorption/ionization-tandem mass spectrometry with high resolution and sensitivity for identification and characterization of proteins. *Proteomics* 2 (2002): 868–876.

[59] Suckau, D., Resemann, A., Schuerenberg, M., Hufnagel, P., Franzen, J. and Holle A. A novel MALDI LIFT-TOF/TOF mass spectrometer for proteomics. *Analytical and Bioanalytical Chemistry* 376 (2003): 952–965.

[60] Shevchenko, A., Jensen, O.N., Podtelejnikov, A.V., Sagliocco, F., Wilm, M., Vorm, O., Mortensen, P., Boucherie, H. and Mann, M. Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. *Proceedings of the National Academy of Sciences of the United States of America* 93 (1996): 14440-14445.

[61] Gharbi, S., Gaffney, P., Yang, A., Zvelebil, M.J, Cramer, R., Waterfield, M.D. and Timms, J.F. Evaluation of two-dimensional differential gel electrophoresis for proteomic expression analysis of a model breast cancer cell system. *Molecular & Cellular Proteomics* 1 (2002): 91-98.

[62] Cottrell, J.S. Protein identification by peptide mass fingerprinting. *Journal of Peptide Research* 7 (1994): 115- 124.

[63] Biemann, K. Contributions of mass spectrometry to peptide and protein structure. *Biomedical and Environmental Mass Spectrometry* 16 (1988): 99-111.

[64] Cramer, R. and Corless, S. The nature of collision-induced dissociation processes of doubly protonated peptides: Comparative study for the future use of matrix-assisted laser desorption/ionization on a hybrid quadrupole time-of-flight mass spectrometer in proteomics. *Rapid Communications in Mass Spectrometry* 15 (2001): 2058-66.

[65] Arnott, D., Kottmeier, D., Yates, N., Shabanowitz, J. and Hunt, D.F. Proceedings of the 42rd ASMS on mass spectrometry and allied topics. *The American society for mass spectrometry* (1994): 470.

[66] Shadforth, I., Crowther, D. and Bessant, C. Protein and peptide identification algorithms using MS for use in high-throughput, automated pipelines. *Proteomics* 5 (2005): 4082–4095.

[67] Altschul, S. F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research* 25 (1997): 3389–3402.

[68] Pearson, W.R. and Lipman, D. J. Improved tools for biological sequence comparison. *Proceedings of the National Academy of Sciences of the United States of America* 25 (1988): 2444–2448.

[69] Taylor, J.A., and Johnson, R.S. Implementation and uses of auto-mated de novo peptide sequencing by tandem mass spectrometry. *Analytical Chemistry* 73 (2001): 2594–2604.

[70] Shevchenko, A., Sunyaev, S., Loboda, A., Shevchenko, A., Bork, P., Ens, W. and Standing, K.G. Charting the proteomes of organisms with unsequenced genomes by MALDI-Quadrupole Time-of-Flight mass spectrometry and BLAST homology searching. *Analytical Chemistry* 73 (2001): 1917–1926.

[71] Mackey, A.J., Haystead, T.A.J. and Pearson, W.R. Getting more from less: Algorithms for rapid protein identification with multiple short peptide sequences. *Molecular & Cellular Proteomics* 1 (2002): 139–147.

[72] Dunn, M. J.; Corbett, J. M.; Two-dimensional polyacrylamide gel electrophoresis. *Methods in Enzymology* 271 (1996): 177-203.

[73] Peng, J., and Gygi, S.P. Proteomics: The move to mixtures. *Journal of Mass Spectrometry* 36 (2001): 1083-1091.

[74] Shen, Y., Tolic, N., Masselon, C., Pasa, T.L., Camp, D.G., Hixson, K.K., Zhao, R., Anderson, G.A. and Smith, R.D. Ultrasensitive proteomics using high-efficiency online micro-SPE-nanoLC-nanoESI MS and MS/MS. *Analytical Chemistry* 76 (2004): 144-154.

[75] Petersen, S.V., Valnickova, Z. and Enghild, J.J. Pigment-epithelium-derived factor (PEDF) occurs at a physiologically relevant concentration in human blood: purification and characterization. *Biochemical Journal* 374 (2003): 199-206.

[76] Peng, J., Elias, J.E., Thoreen, C.C., Licklider, L.J. and Gygi, S.P. Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LCMS/MS) for large-scale protein analysis: the yeast proteome. *Journal of Proteome Researsch* 2 (2003): 43-50.

[77] Oda, Y., Huang, K., Cross, F.R., Cowburn, D. and Chait, B.T. Accurate quantitation of protein expression and site-specific phosphorylation. *Proceedings of the National Academy of Sciences of the United States of America* 96 (1999): 6591-6596.

[78] Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H. and Aebersold, R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnology* 17 (1999): 994-999.

[79] Ross, P.L., Huang, Y.N., Marchese, J.N., Williamson, B., Parker, K., Hattan, S., et al. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Molecular & Cellular Proteomics* 3 (2004): 1154–1169.

[80] Choe. L.H., Aggarwal, K., Franck, Z. and Lee, K.H. A comparison of the consistency of proteome quantitation using two-dimensional electrophoresis and shotgun isobaric tagging in Escherichia coli cells. *Electrophoresis* 26 (2005): 2437-2449.

[81] Kopka, J., Provart, N.J. and Mueller, R.B. Potato guard cells respond to drying soil by a complex change in the expression of genes related to carbon metabolism and turgor regulation. *The Plant Journal.* 11 (1997): 871-882.

[82] Hunt, L., Mills, L.N., Pical, C., Leckie, C.P., Aitken, F.L., Kopka, J., Mueller-Roeber, B., McAinsh, M.R., Hetherington, A.M. and Gray, J.E. Phospholipase C is required for the control of stomatal aperture by ABA. *The Plant Journal* 34 (2003): 47-55.

[83] Mortz, E., Krogh, T. N., Vorum, H. and Gorg, A. Improved silver staining protocols for high sensitivity protein identification using matrix-assisted laser desorption/ionization-time of flight analysis. *Proteomics* 1 (2001): 1359-1363.

[84] Matsui, T., Yoshimoto, C., Osajima, K., Oki, T. and Osajima, Y. *Bioscience Biotechnology and Biochemistry* 60 (1996): 2019-2022.

[85] Pandey, A. and Mann, M. Proteomics to study genes and genomes. *Nature* 405 (2000): 837-846.

[86] Kojima, H., Yanai, T. and Toyota, A. Essential oil constituents from Japanese and Indian *Curcuma aromatica* rhizomes. *Planta Medica* 64 (1998): 380-381.

[87] Tohda, C., Nakayama, N., Hatanaka, F. and Komatsu, K. Comparison of anti-inflammatory activities of six Curcuma rhizomes: A possible curcuminoid-independent pathway mediated by *Curcuma phaeocaulis* extract. *Evidence-Based Complementary and Alternative Medicine* 3 (2006): 255-260.

[88] Pitasawat, B., Choochote, W., Tuetun, B., Tippawangkosol, P., Kanjanapothi, D., Jitpakdi, A. and Riyong, D. Repellency of aromatic turmeric *Curcuma aromatica* under laboratory and field conditions. *Journal of Vector Ecology* 28 (2003): 234-240.

[89] Choi, K. H. and Laursen, R. A. Amino-acid sequence and glycan structures of cysteine proteases with proline specificity from ginger rhizome *Zingiber officinale*. *European Journal of Biochemistry* 267 (2000): 1516-1526.

[90] Chen, Z. H., Kai, G.Y., Liu, X.J., Lin, J., Sun, X.F. and Tang, K.X. cDNA cloning and characterization of a mannose-binding lectin from *Zingiber officinale* Roscoe (ginger) rhizomes. *Journal of Biosciences* 30 (2005): 213-220.

[91] Van Damme, E.J.M., Smeets, K., Torrekens, S., Vanleuven, F. and Peumans, W.J. Characterization and molecular-cloning of mannose-binding lectins from the Orchidaceae species *Listera ovata*, *Epipactis helleborine* and *Cymbidium hybrid*. *European Journal of Biochemistry* 221 (1994): 769-777.

[92] Van Damme, E.J.M., Brike, F., Winter, H.C., Van Leuven, F., Goldstein, I.J. and Peumans, W.J. Molecular cloning of two different mannose-binding lectins from tulip bulbs. *European Journal of Biochemistry* 236 (1996): 419-427.

[93] Smeets, K., Van Damme, E.J.M., Verhaert, P., Barre, A., Rouge, P., Van Leuven, F. and Peumans, W.J. Isolation, characterization and molecular cloning of the mannose-binding lectins from leaves and roots of garlic (*Allium sativum* L). *Plant Molecular Biology* 33 (1997): 223-234.

[94] Shadforth, I.P., Dunkley, T.P.J., Lilley, K.S. and Bessant, C. i-Tracker: For quantitative proteomics using iTRAQ™ *BMC Genomics* 6 (2005): 145.

[95] Leonhardt, N., Kwak, J.M., Robert, N., Waner D., Leonhardt, G. and Schroeder, J.I. Microarray expression analyses of Arabidopsis guard cells and isolation of a recessive abscisic acid hypersensitive protein phosphatase 2C mutant. *The plant cell* 16 (2004): 596-615.

# APPENDICES

# Appendix A

Solution in 1D SDS-PAGE

**A. 30% T, 2.6 %C monomer stock solution**

| | |
|---|---|
| Acrylamide | 30 mg |
| *N*,*N*'-methylenebisacrylamide | 0.8 g |
| Double distilled $H_2O$ | to 100 ml |

Filter solution through a 0.45 μm filter. Store at 4 °C in the dark

**B. Resolving buffer**

| | |
|---|---|
| Tris base | 4.5 g |
| Double distilled $H_2O$ | 70 ml |
| HCl | adjust to pH 8.8 |
| Double distilled $H_2O$ | to 100 ml |

Filter solution through a 0.45 μm filter. Store at 4 °C

**C. 10% SDS**

| | |
|---|---|
| SDS | 5.0 g |
| Double distilled $H_2O$ | to 50 ml |

Filter solution through a 0.45 μm filter. Store at room temperature

**D. 10% ammonium persulfate**

| | |
|---|---|
| Ammonium persulfate | 0.1 g |
| Double distilled $H_2O$ | 1 ml |

Prepare just prior to use.

**E. SDS electrophoresis buffer**

| | |
|---|---|
| Tris-base | 30.3 g |
| Glycine | 144.0 g |
| SDS | 10.0 g |
| Double distilled $H_2O$ | to 10 l |

Store at room temperature

**F. Coomassie Blue staining solution**

| | |
|---|---|
| Coomassie Brilliant Blue G-250 | 0.1 g |
| Methanol A.R. grade | 40 ml |
| Acetic acid | 10 ml |
| Double distilled $H_2O$ | 50 ml |

Dissolved Coomassie Brilliant Blue G-250 in methanol until it is completely dissolved. Add Double distilled $H_2O$ and acetic acid. Prepare just prior to use.

**G. Destaining solution**

| | Destaining solution I | Destaining solution II |
|---|---|---|
| Methanol A.R. grade | 40 ml | 10 ml |
| Acetic acid | 10 ml | 5 ml |
| Distilled $H_2O$ | to 100 ml | to 100 ml |

Store at room temperature

**Table 1A** Amino acid residue masses

| Amino acid | Single letter code | Residue mass (Da) | Immonium ions (Da) |
|---|---|---|---|
| Alanine | A | 71.04 | 44 |
| Arginine | R | 156.1 | 129 |
| Asparagine | N | 114.04 | 87 |
| Aspartic acid | D | 115.03 | 88 |
| Carbamidomethylcysteine | $C^*$ | 160.03 | 133 |
| Cysteine | C | 103.01 | 76 |
| Glutamic acid | E | 129.04 | 102 |
| Glutamine | Q | 128.06 | 101 |
| Glycine | G | 57.02 | 30 |
| Histidine | H | 137.06 | 110 |
| Isoleucine | I | 113.08 | 86 |
| Leucine | L | 113.08 | 86 |
| Lysine | K | 128.09 | 101 |
| Methionine | M | 131.04 | 104 |
| Oxidized Methionine | $M_{Ox}$ | 147.04 | 120 |
| Phenylalanine | F | 147.07 | 120 |
| Proline | P | 97.05 | 70 |
| Propionamidocysteine | $C^a$ | 174.04 | 147 |
| Serine | S | 87.03 | 60 |
| Threonine | T | 101.05 | 74 |
| Tryptophane | W | 186.08 | 159 |
| Tyrosine | Y | 163.06 | 136 |
| Valine | V | 99.07 | 72 |

**Table 2A** $b_2$-ion masses lookup table.

| | G | A | S | P | V | T | C | L/I | N | D | Q/K | E | M | H | $M_{Ox}$/F | R | $C^*$ | Y | $C^a$ | W |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| G | 115 | | | | | | | | | | | | | | | | | | | |
| A | 129 | 143 | | | | | | | | | | | | | | | | | | |
| S | 145 | 159 | 175 | | | | | | | | | | | | | | | | | |
| P | 155 | 169 | 185 | 195 | | | | | | | | | | | | | | | | |
| V | 157 | 171 | 187 | 197 | 199 | | | | | | | | | | | | | | | |
| T | 159 | 173 | 189 | 199 | 201 | 203 | | | | | | | | | | | | | | |
| C | 161 | 175 | 191 | 201 | 203 | 205 | 207 | | | | | | | | | | | | | |
| L/I | 171 | 185 | 201 | 211 | 213 | 215 | 217 | 227 | | | | | | | | | | | | |
| N | 172 | 186 | 202 | 212 | 214 | 216 | 218 | 228 | 229 | | | | | | | | | | | |
| D | 173 | 187 | 203 | 213 | 215 | 217 | 219 | 229 | 230 | 231 | | | | | | | | | | |
| Q/K | 186 | 200 | 216 | 226 | 228 | 230 | 232 | 242 | 243 | 244 | 257 | | | | | | | | | |
| E | 187 | 201 | 217 | 227 | 229 | 231 | 233 | 243 | 244 | 245 | 258 | 259 | | | | | | | | |
| M | 189 | 203 | 219 | 229 | 231 | 233 | 235 | 245 | 246 | 247 | 260 | 261 | 263 | | | | | | | |
| H | 195 | 209 | 225 | 235 | 237 | 239 | 241 | 251 | 252 | 253 | 266 | 267 | 269 | 275 | | | | | | |
| $M_{Ox}$/F | 205 | 219 | 235 | 245 | 247 | 249 | 251 | 261 | 262 | 263 | 276 | 277 | 279 | 285 | 295 | | | | | |
| R | 214 | 228 | 244 | 254 | 256 | 258 | 260 | 270 | 271 | 272 | 285 | 286 | 288 | 294 | 304 | 313 | | | | |
| $C^*$ | 218 | 232 | 248 | 258 | 260 | 262 | 264 | 274 | 275 | 276 | 289 | 290 | 292 | 298 | 308 | 307 | 321 | | | |
| Y | 221 | 235 | 251 | 261 | 263 | 265 | 267 | 277 | 278 | 279 | 292 | 293 | 295 | 301 | 311 | 320 | 324 | 327 | | |
| $C^a$ | 232 | 246 | 262 | 272 | 274 | 276 | 278 | 288 | 289 | 290 | 303 | 304 | 306 | 312 | 322 | 331 | 335 | 338 | 349 | |
| W | 244 | 258 | 274 | 284 | 286 | 288 | 290 | 300 | 301 | 302 | 315 | 316 | 318 | 324 | 334 | 343 | 347 | 350 | 361 | 373 |

# Appendix B

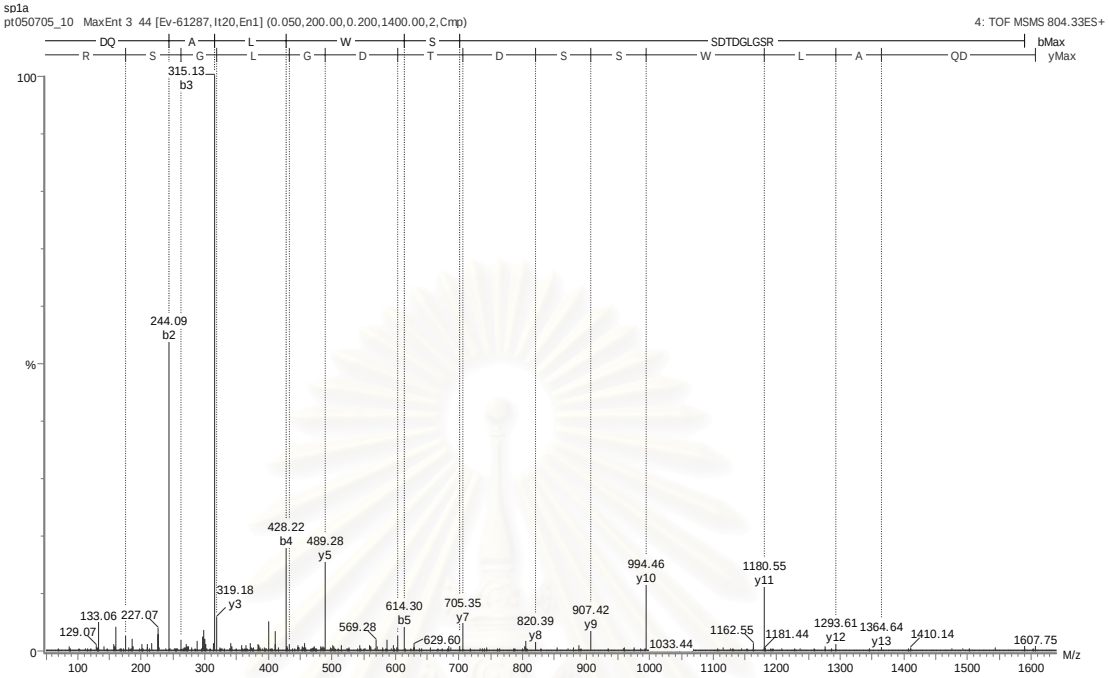



**Figure 1B** MS/MS spectrum of doubly charged precursor ion m/z of 804.33 (MW 1606.64) from gel band S1a

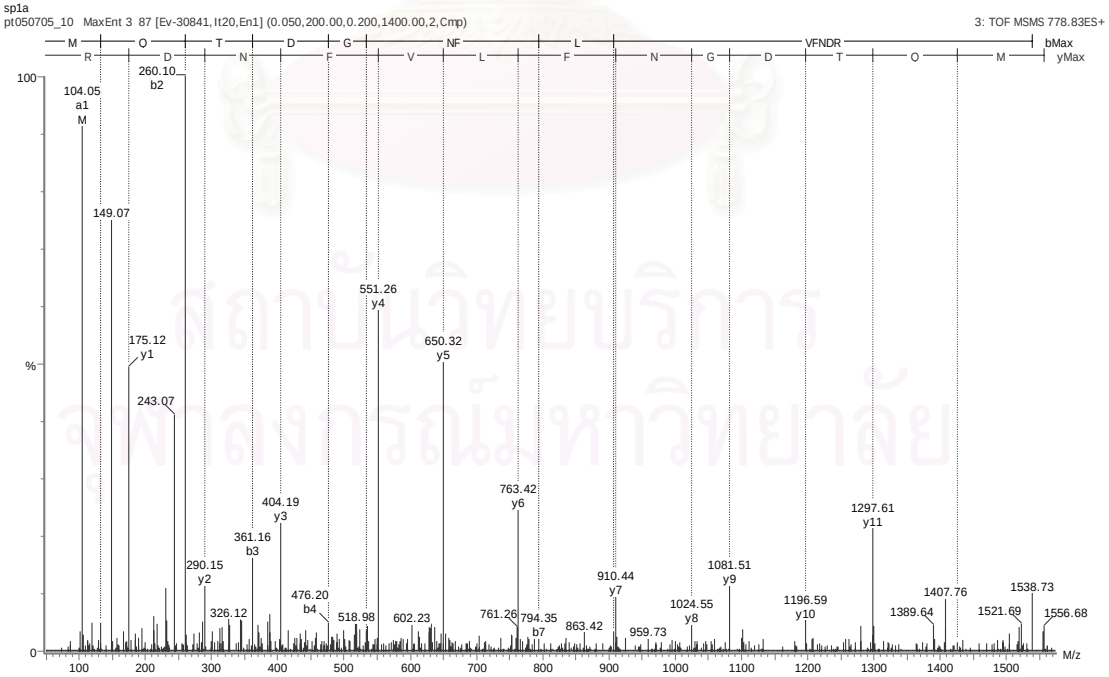



**Figure 2B** MS/MS spectrum of doubly charged precursor ion m/z of 778.83 (MW 1555.64) from gel band S1a

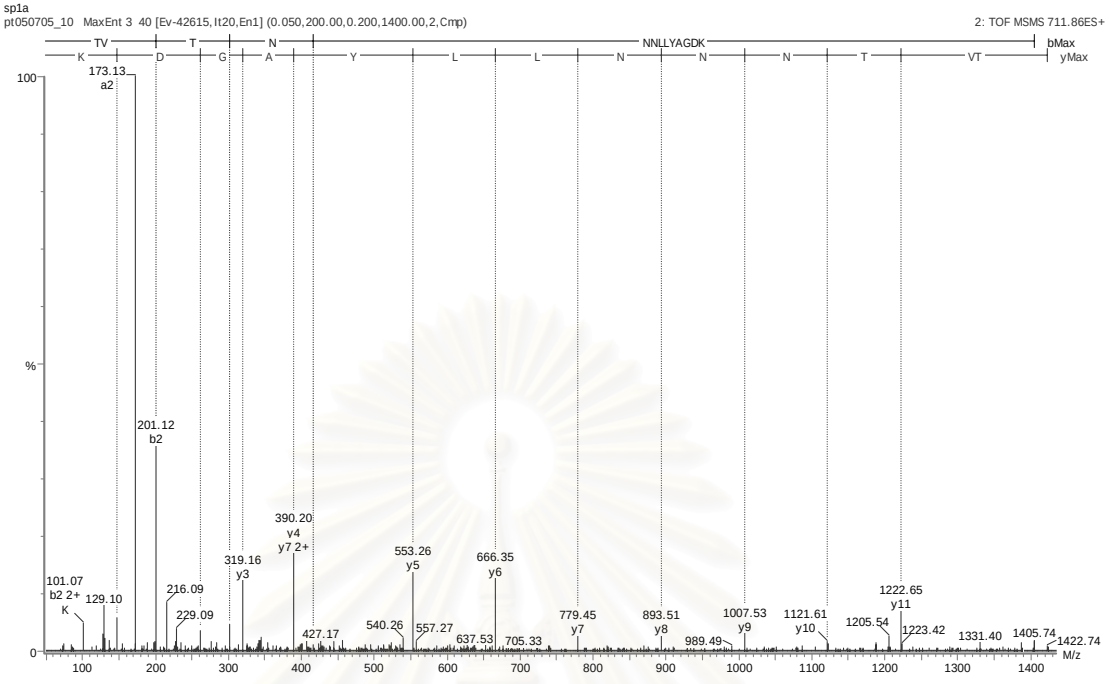



**Figure 3B** MS/MS spectrum of doubly charged precursor ion m/z of 711.86 (MW 1421.70) from gel band S1a

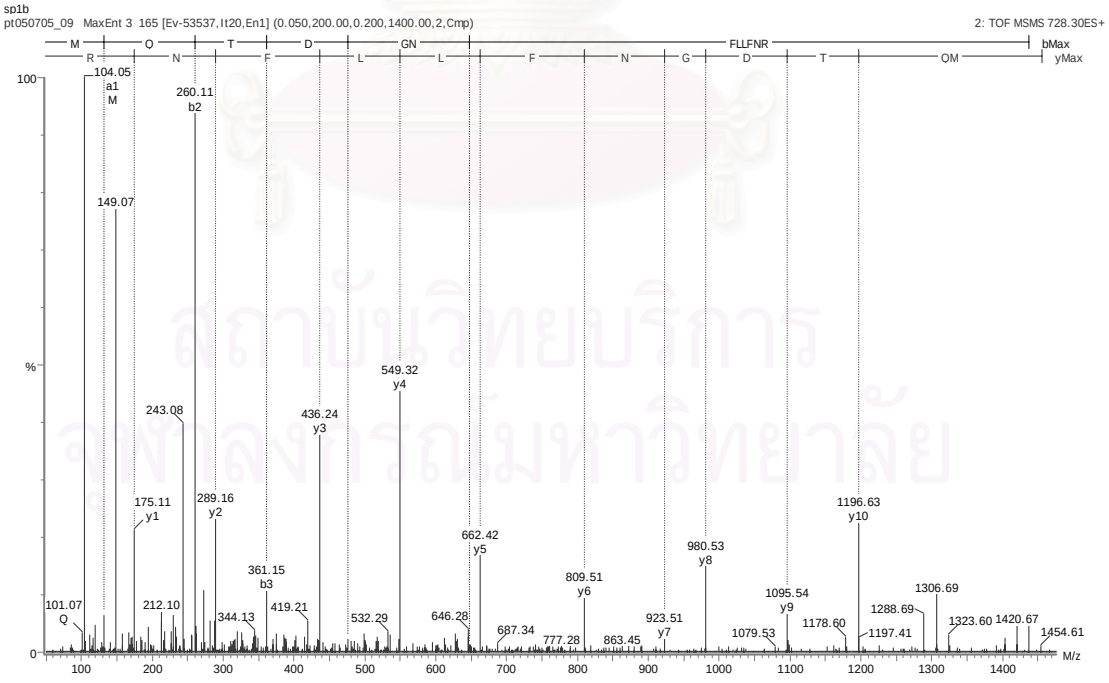



**Figure 4B** MS/MS spectrum of doubly charged precursor ion m/z of 728.30 (MW 1454.584) from gel band S1b

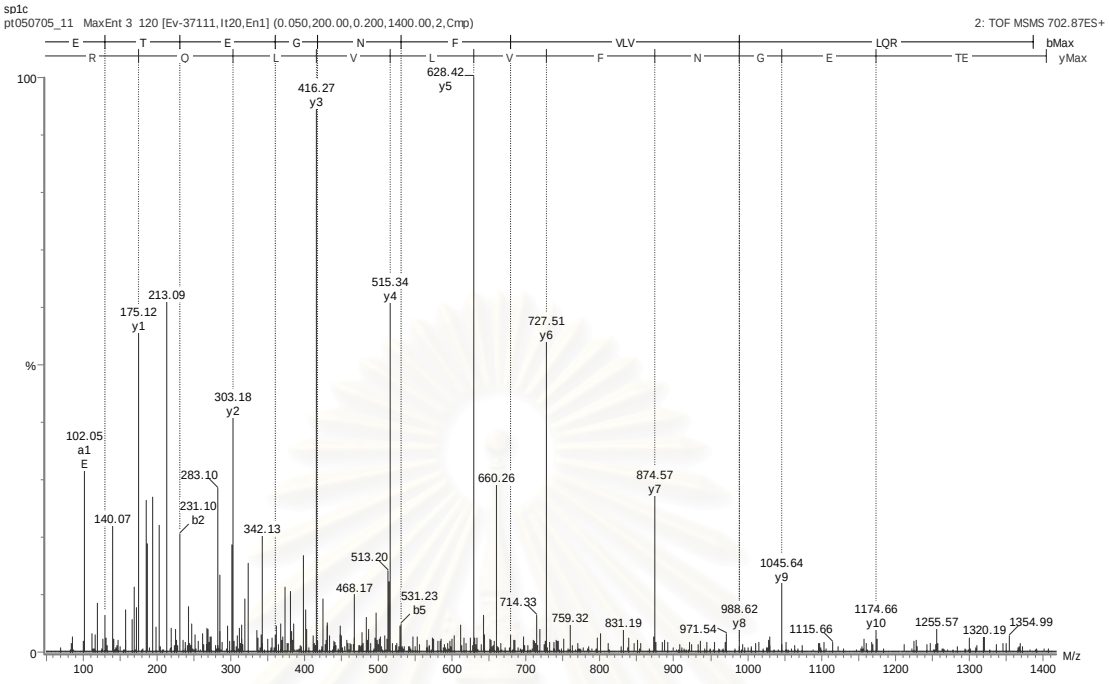



**Figure 5B** MS/MS spectrum of doubly charged precursor ion m/z of 702.87 (MW 1403.72) from gel band S1c

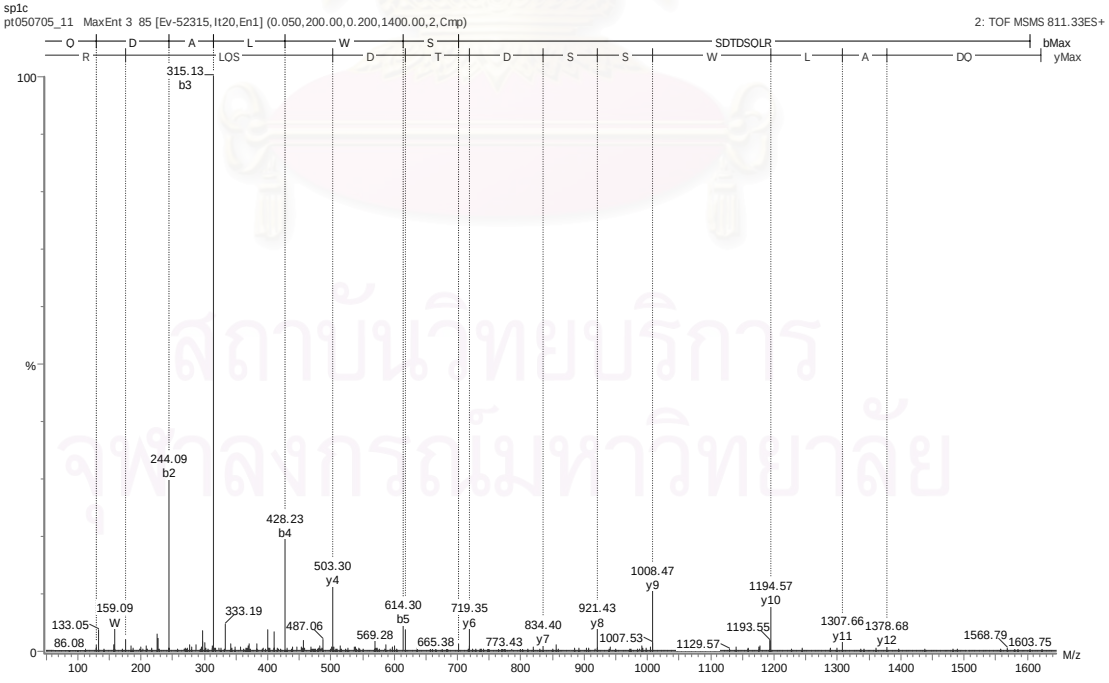



**Figure 6B** MS/MS spectrum of doubly charged precursor ion m/z of 811.33 (MW 1620.64) from gel band S1c

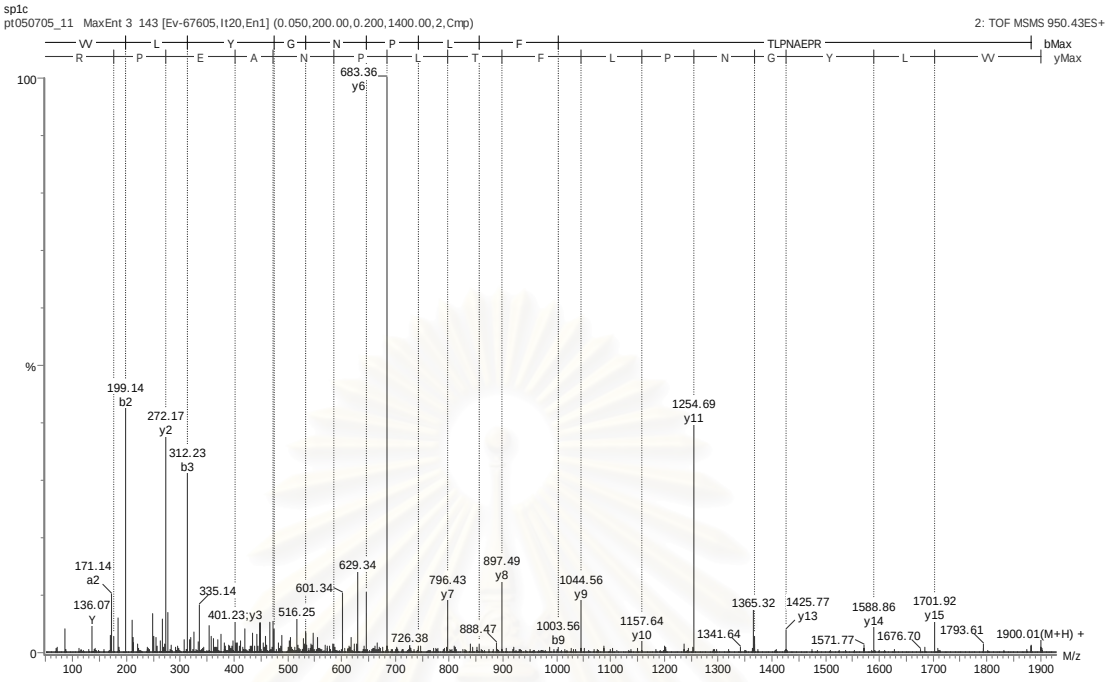



**Figure 7B** MS/MS spectrum of doubly charged precursor ion m/z of 950.43 (MW 1898.84) from gel band S1c

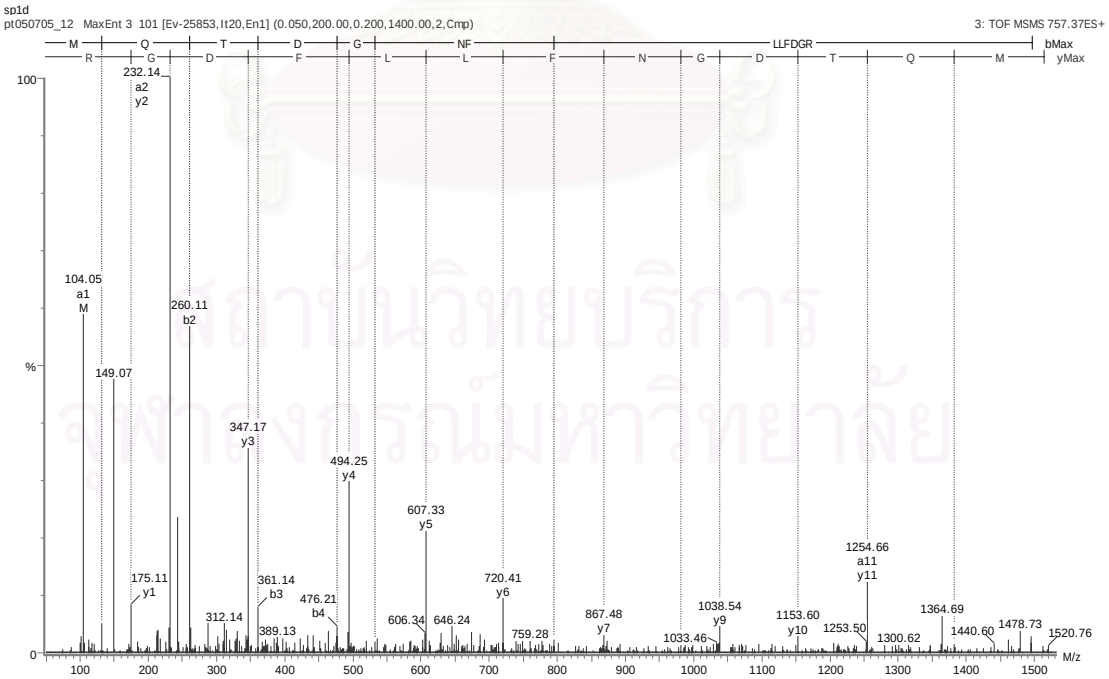



**Figure 8B** MS/MS spectrum of doubly charged precursor ion m/z of 757.37 (MW 1512.72) from gel band S1d

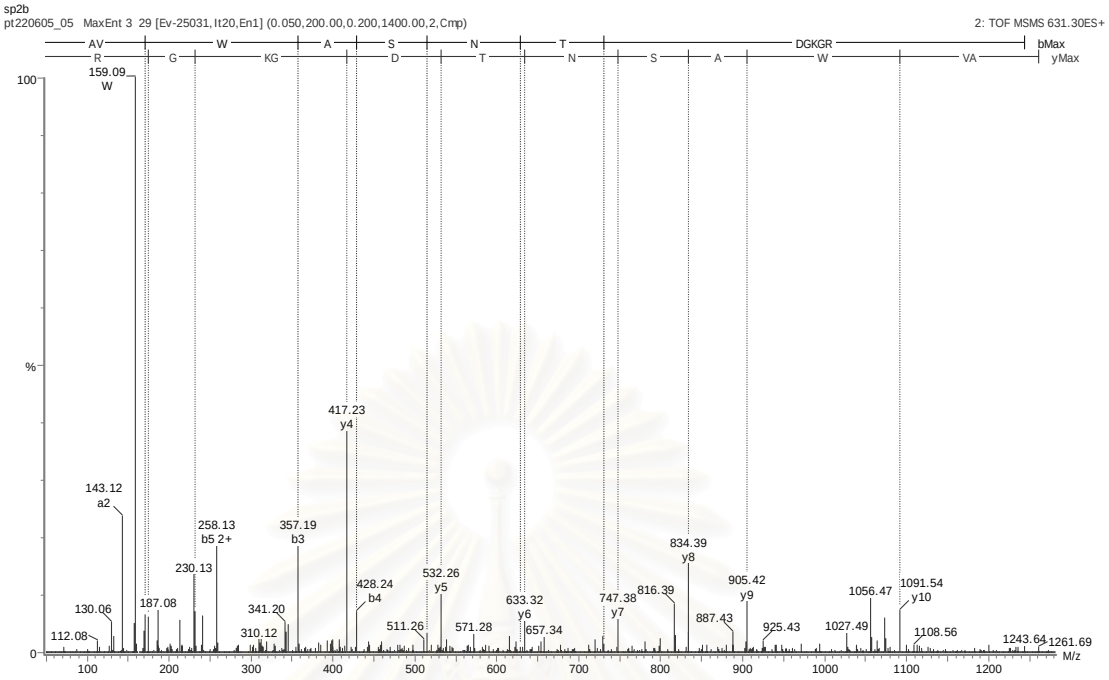



**Figure 9B** MS/MS spectrum of doubly charged precursor ion m/z of 631.30 (MW 1260.58) from gel band S2b

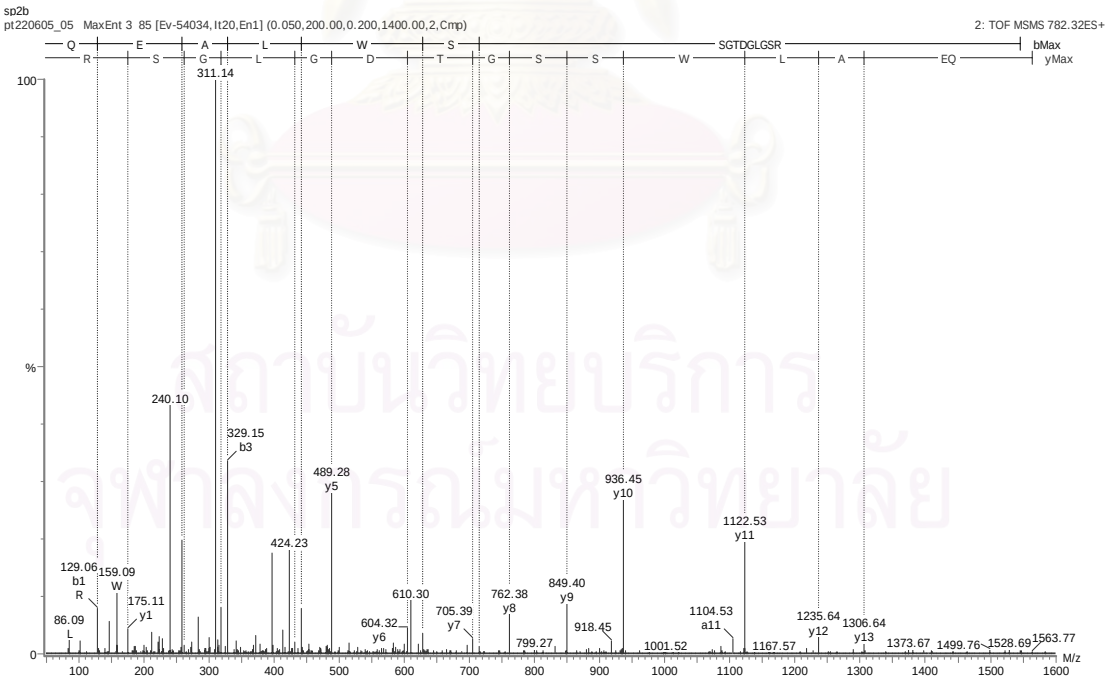



**Figure 10B** MS/MS spectrum of doubly charged precursor ion m/z of 782.32 (MW 1562.62) from gel band S2b

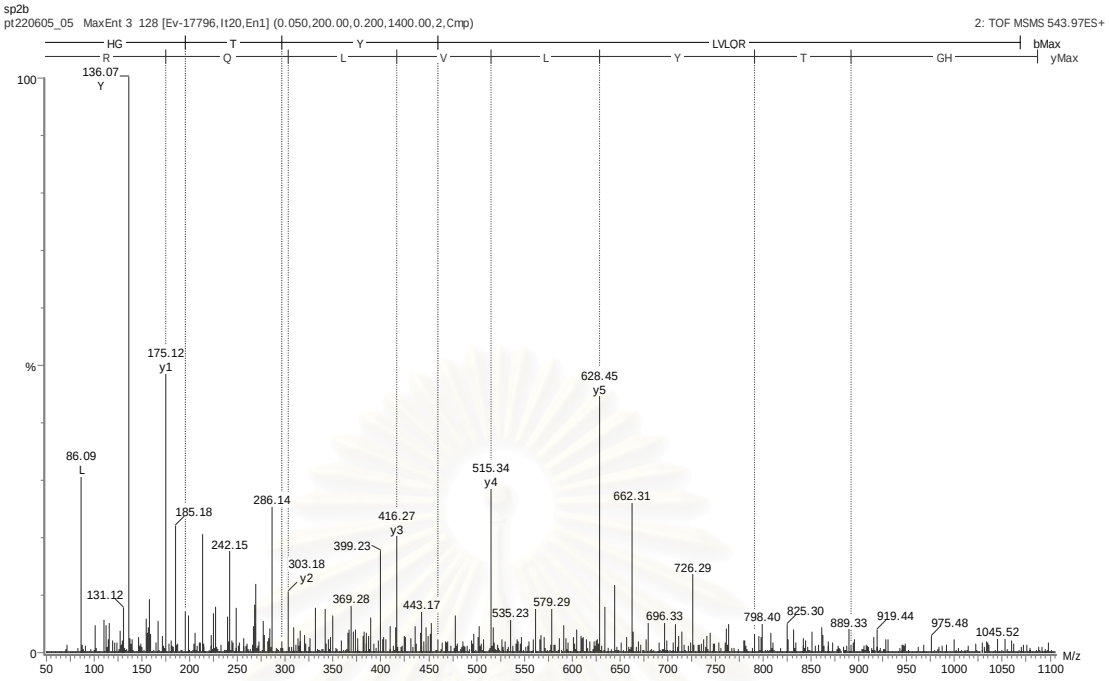



**Figure 11B** MS/MS spectrum of doubly charged precursor ion m/z of 543.97 (MW 1085.92) from gel band S2b

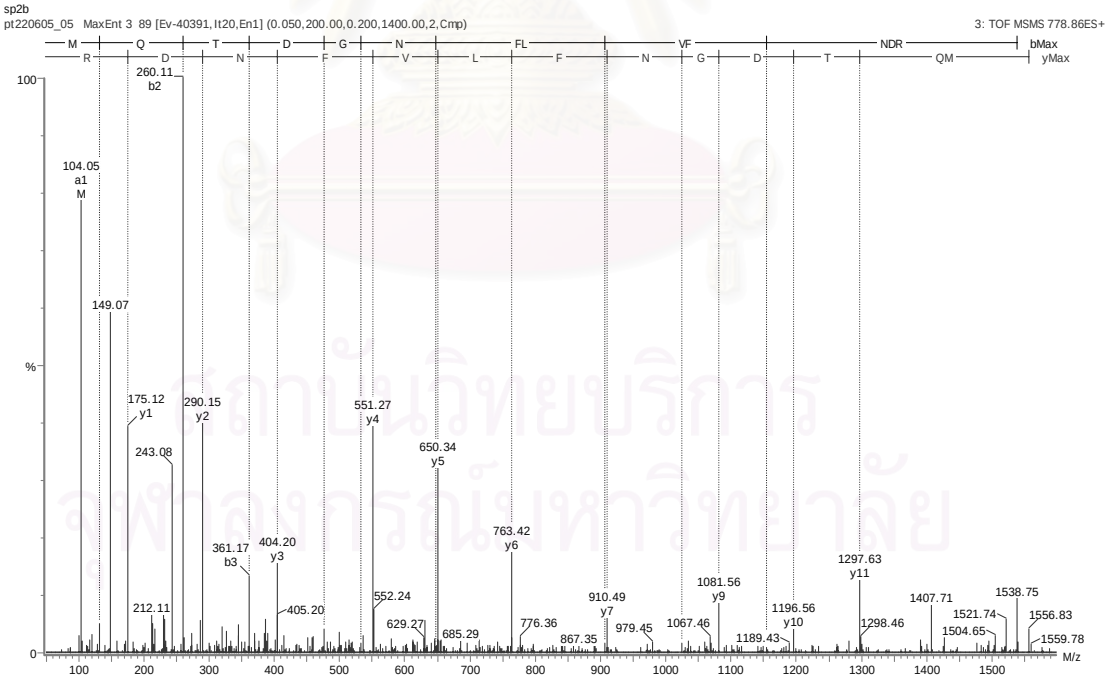



**Figure 12B** MS/MS spectrum of doubly charged precursor ion m/z of 778.86 (MW 1555.70) from gel band S2b

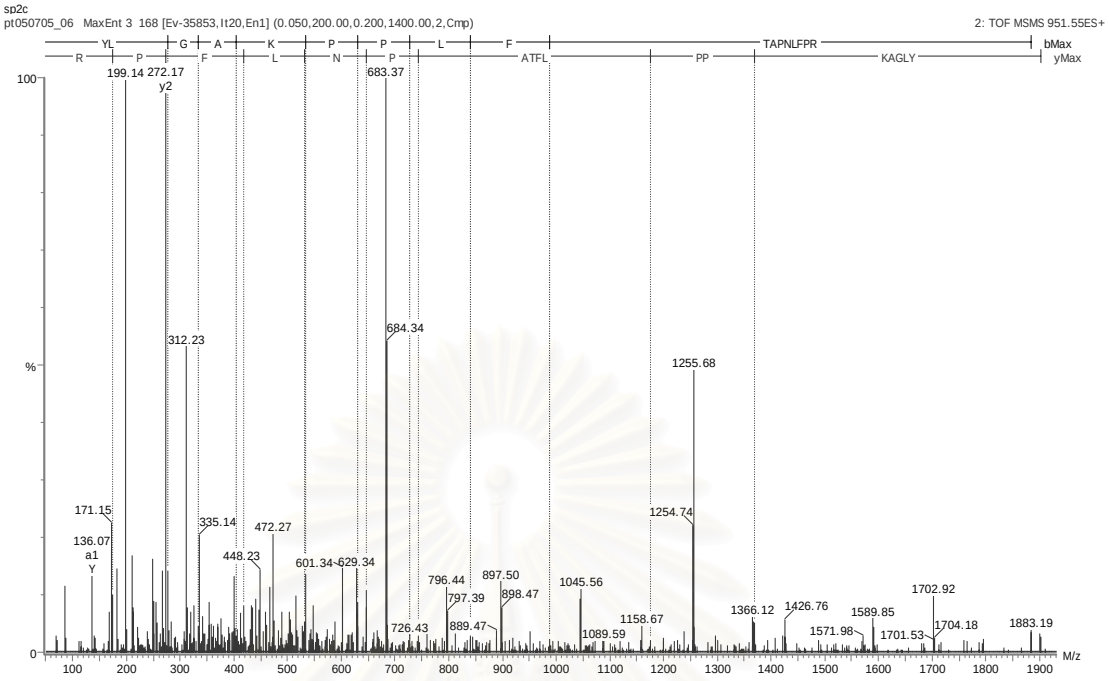



**Figure 13B** MS/MS spectrum of doubly charged precursor ion m/z of 951.55 (MW 1901.08) from gel band S2c

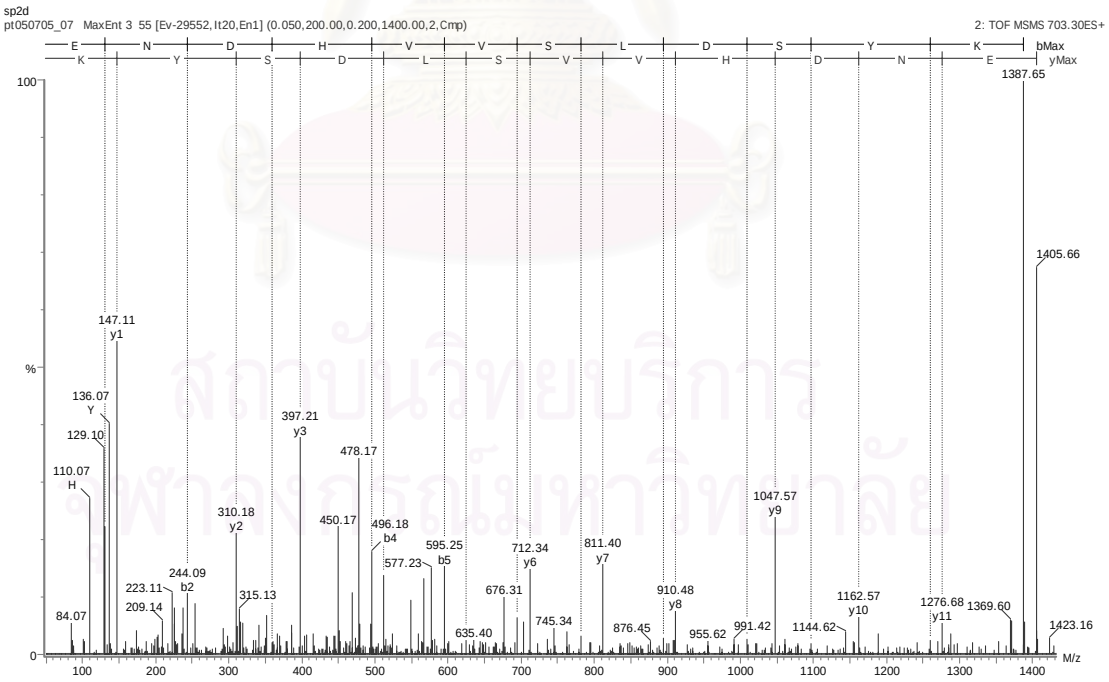



**Figure 14B** MS/MS spectrum of doubly charged precursor ion m/z of 703.30 (MW 1404.58) from gel band S2d

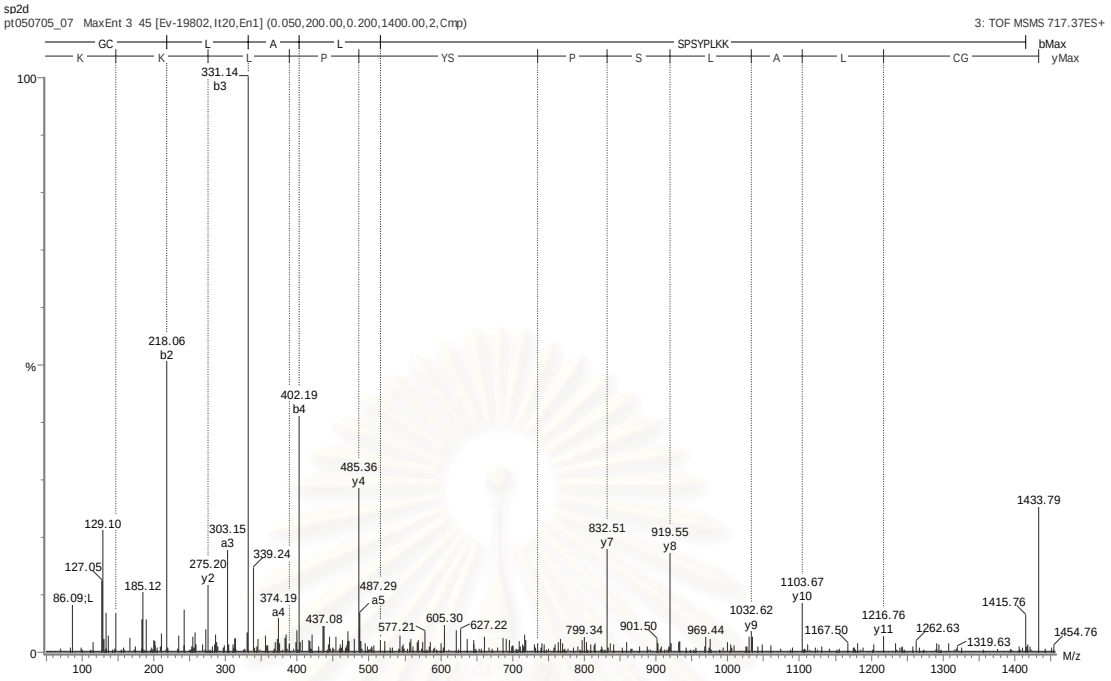



**Figure 15B** MS/MS spectrum of doubly charged precursor ion m/z of 717.37(MW 1432.724) from gel band S2d

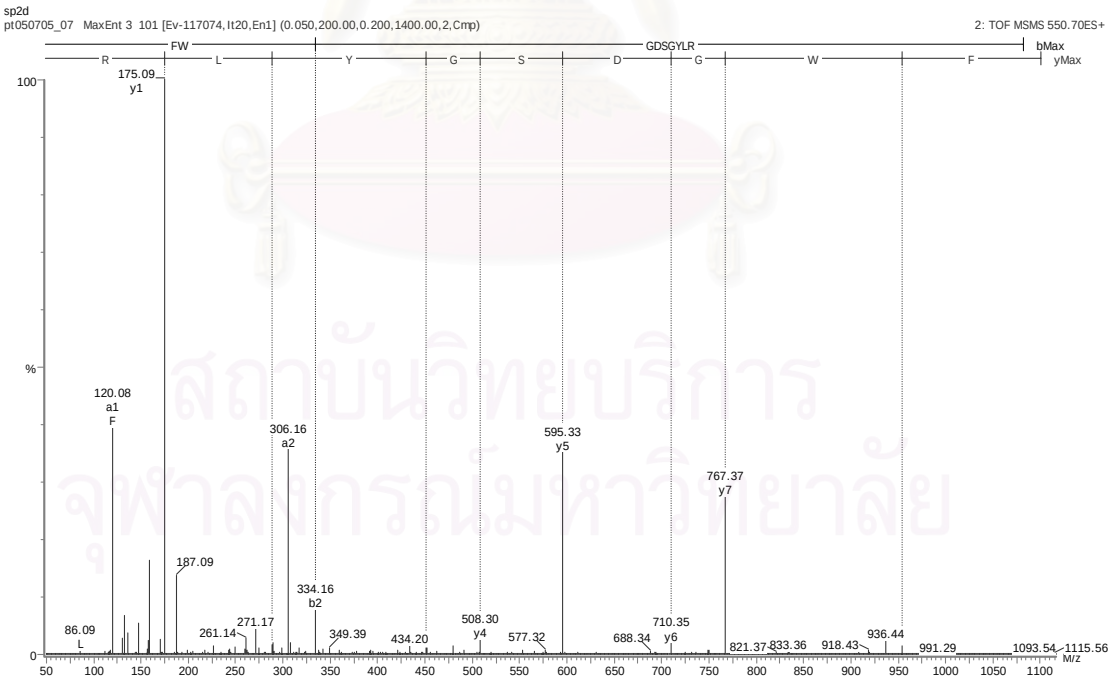



**Figure 16B** MS/MS spectrum of doubly charged precursor ion m/z of 550.70(MW 1099.384) from gel band S2d

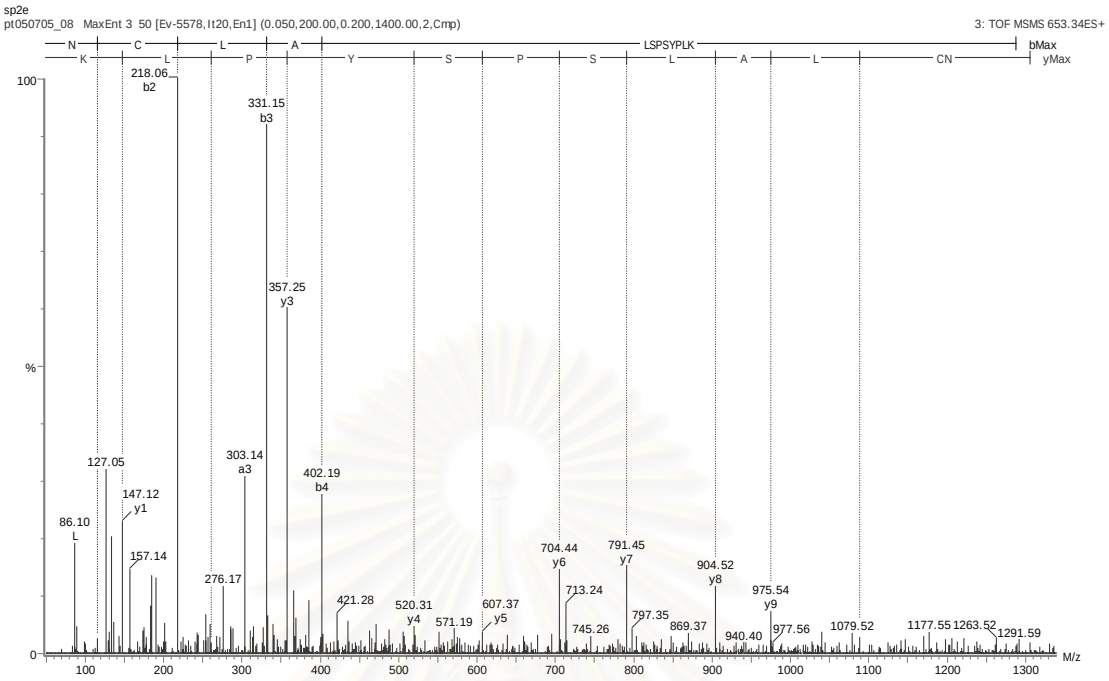



**Figure 17B** MS/MS spectrum of doubly charged precursor ion m/z of 653.34(MW 1304.664) from gel band S2e

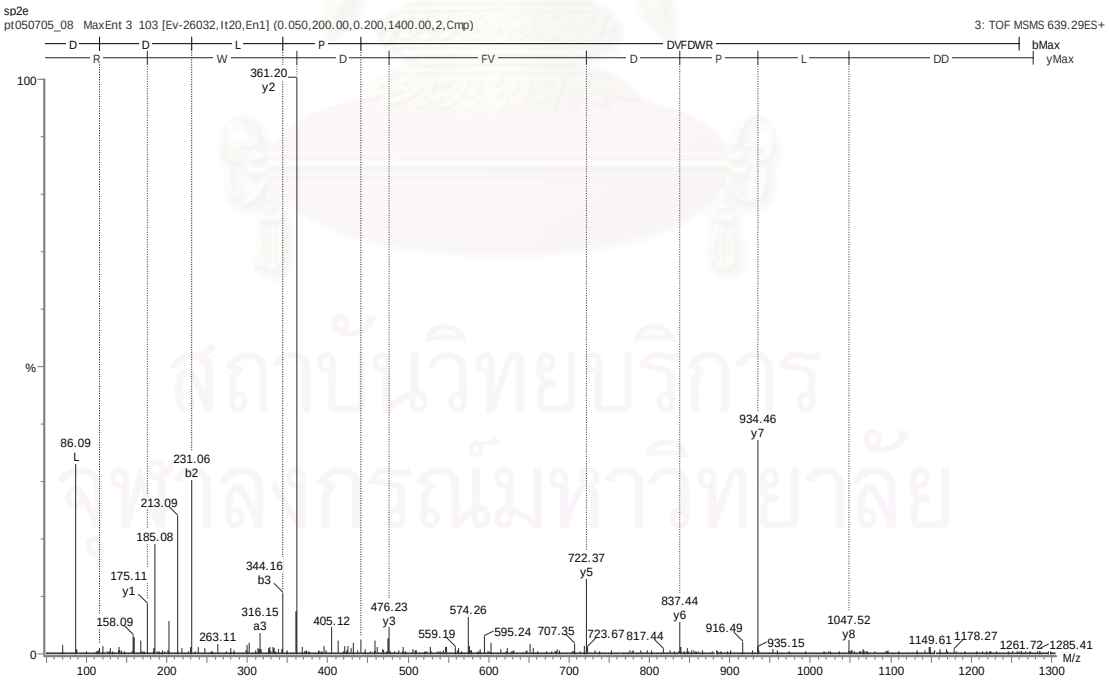



**Figure 18B** MS/MS spectrum of doubly charged precursor ion m/z of 639.29 (MW 1099.504) from gel band S2e

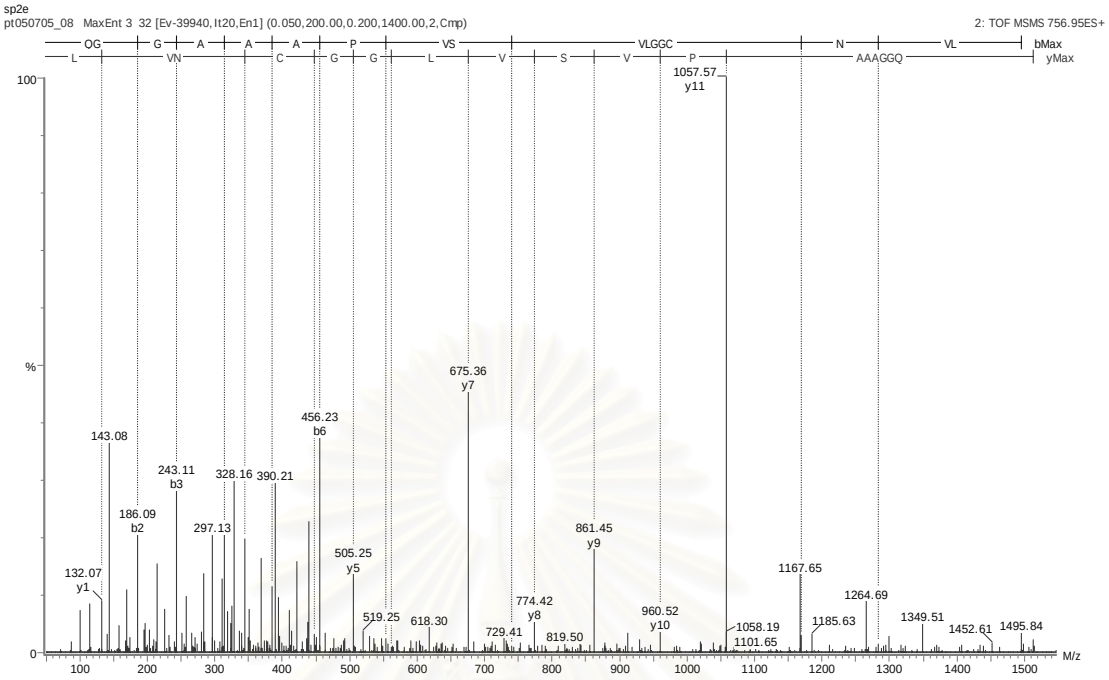



**Figure 19B** MS/MS spectrum of doubly charged precursor ion m/z of 756.95 (MW 1511.885) from gel band S2e

# Appendix C

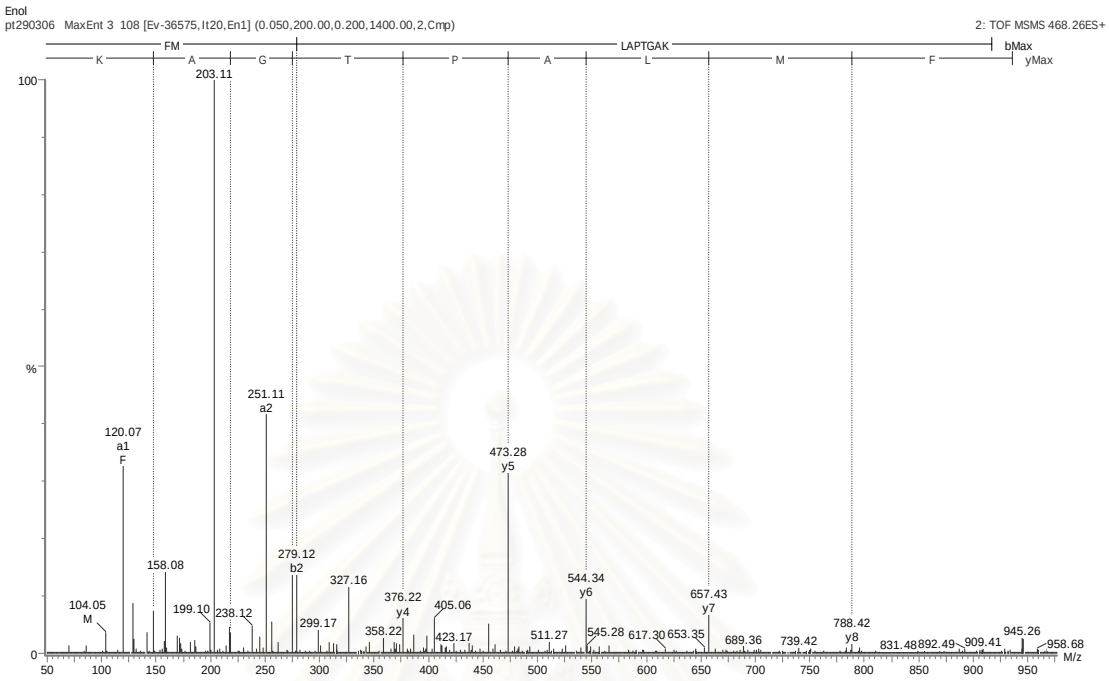



**Figure 1C** MS/MS spectrum of doubly charged precursor ion m/z of 468.26 from Q1S1

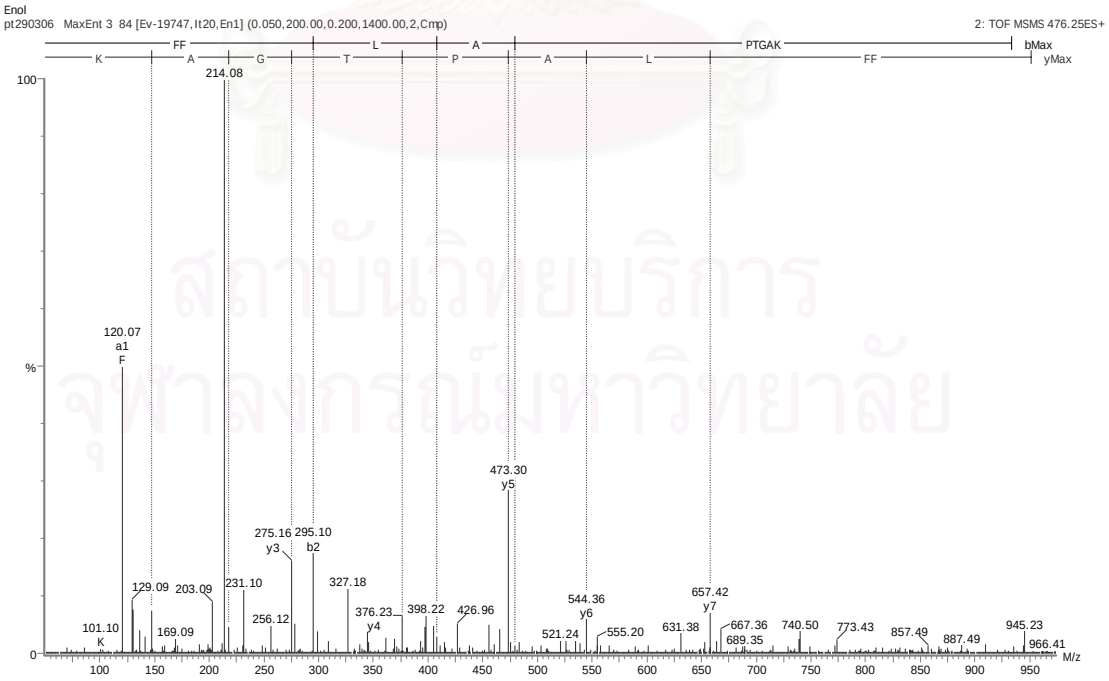



**Figure 2C** MS/MS spectrum of doubly charged precursor ion m/z of 476.25 from Q1S1

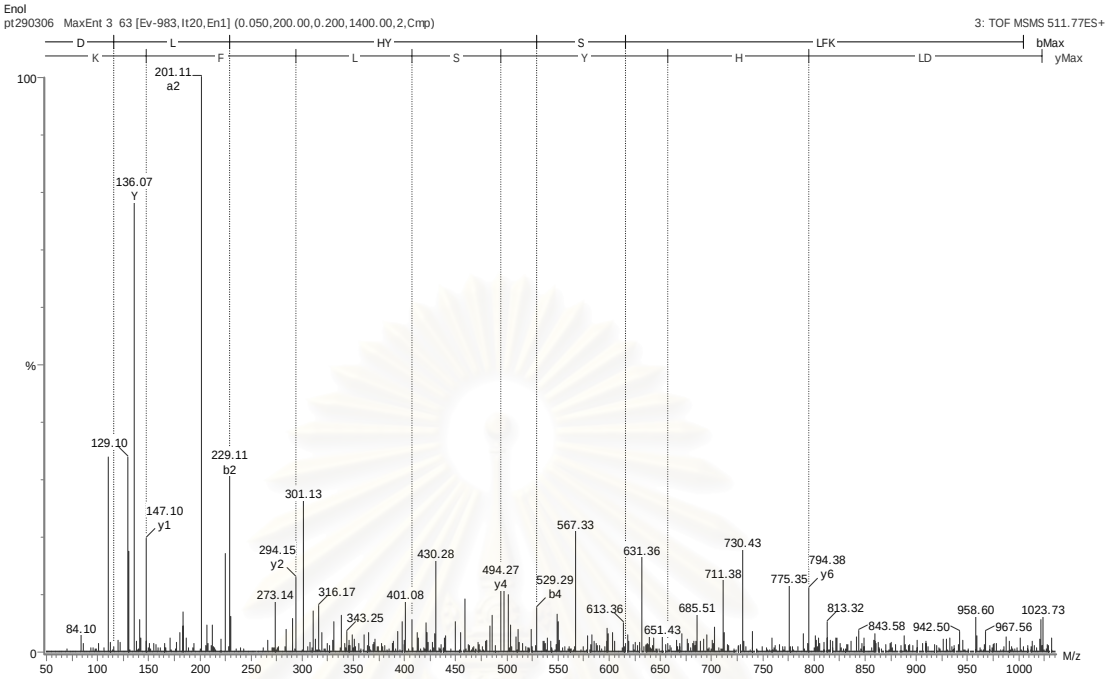



**Figure 3C** MS/MS spectrum of doubly charged precursor ion m/z of 511.77 from Q1S1

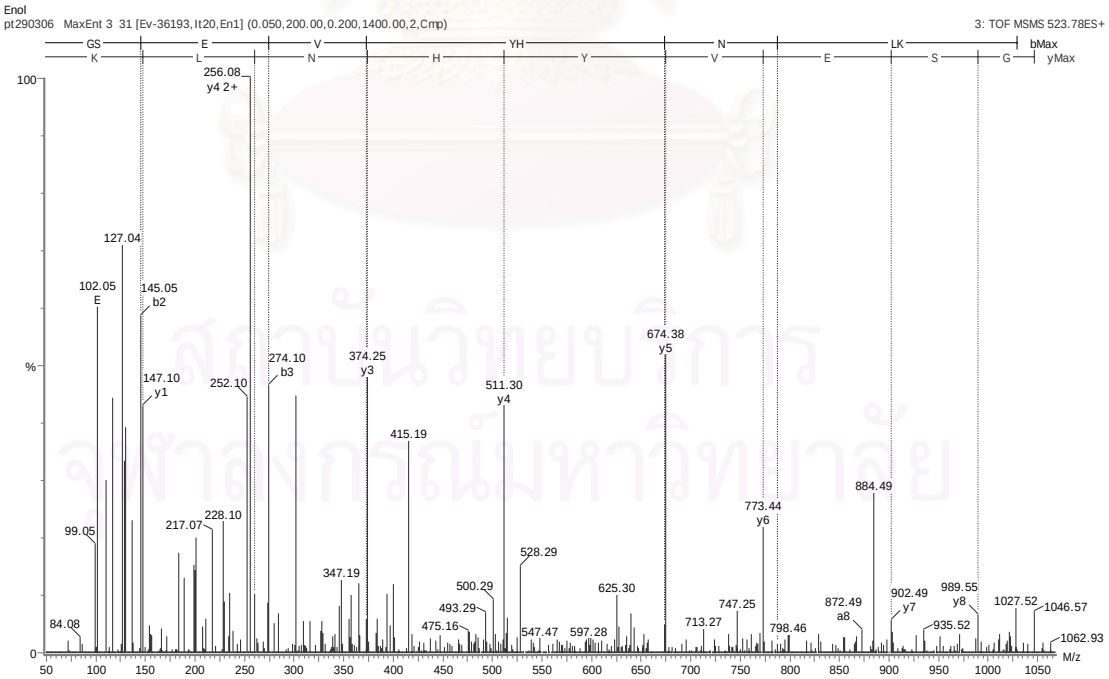



**Figure 4C** MS/MS spectrum of doubly charged precursor ion m/z of 523.78 from Q1S1

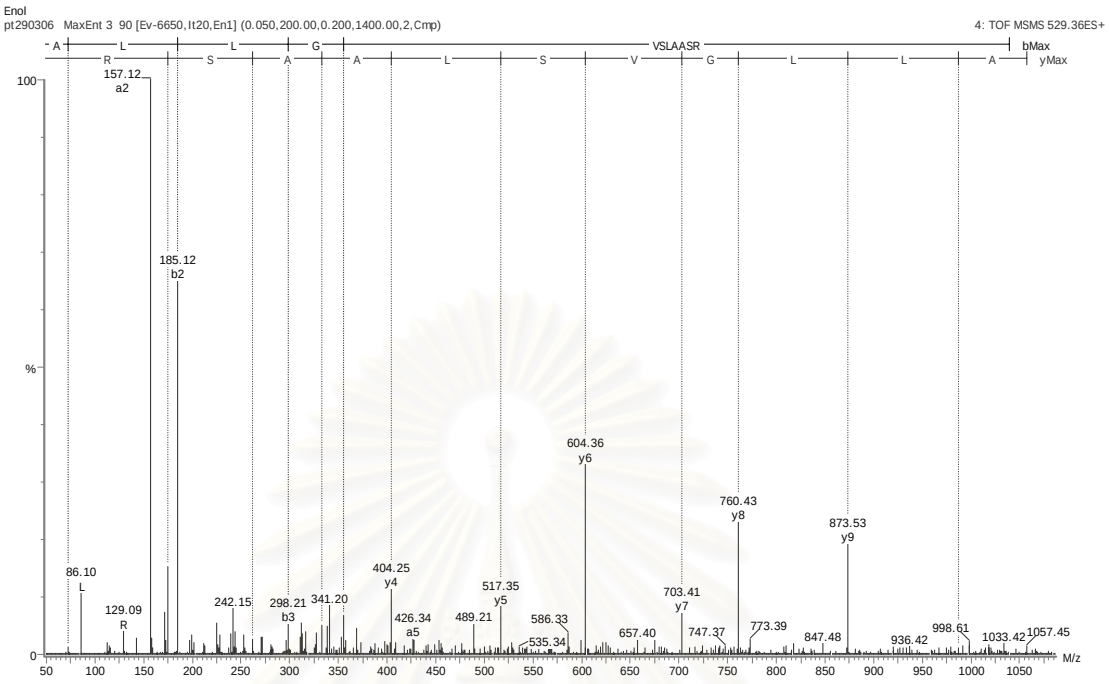



**Figure 5C** MS/MS spectrum of doubly charged precursor ion m/z of 529.36 from Q1S1

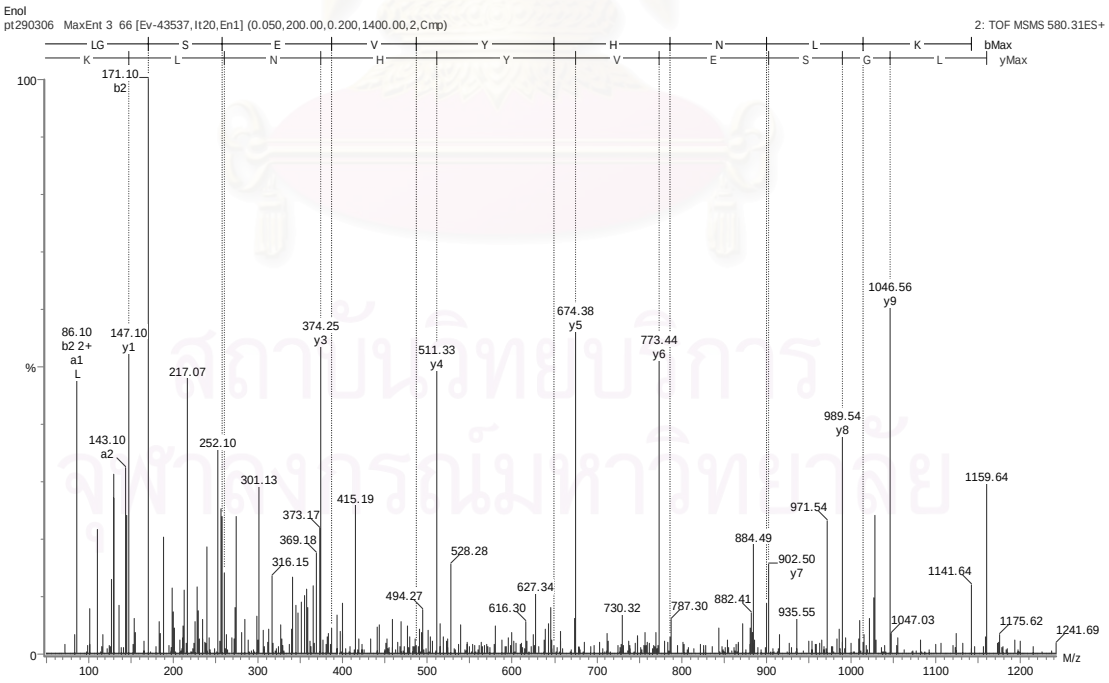



**Figure 6C** MS/MS spectrum of doubly charged precursor ion m/z of 580.31 from Q1S1

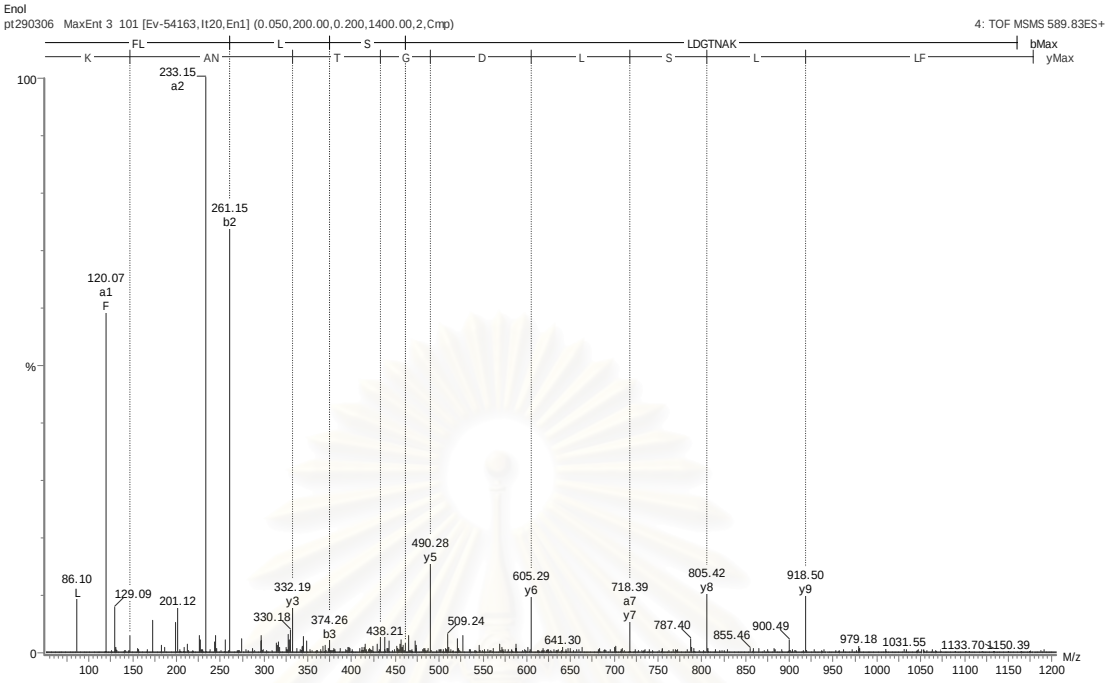



**Figure 7C** MS/MS spectrum of doubly charge precursor ion m/z of 589.38 from Q1S1

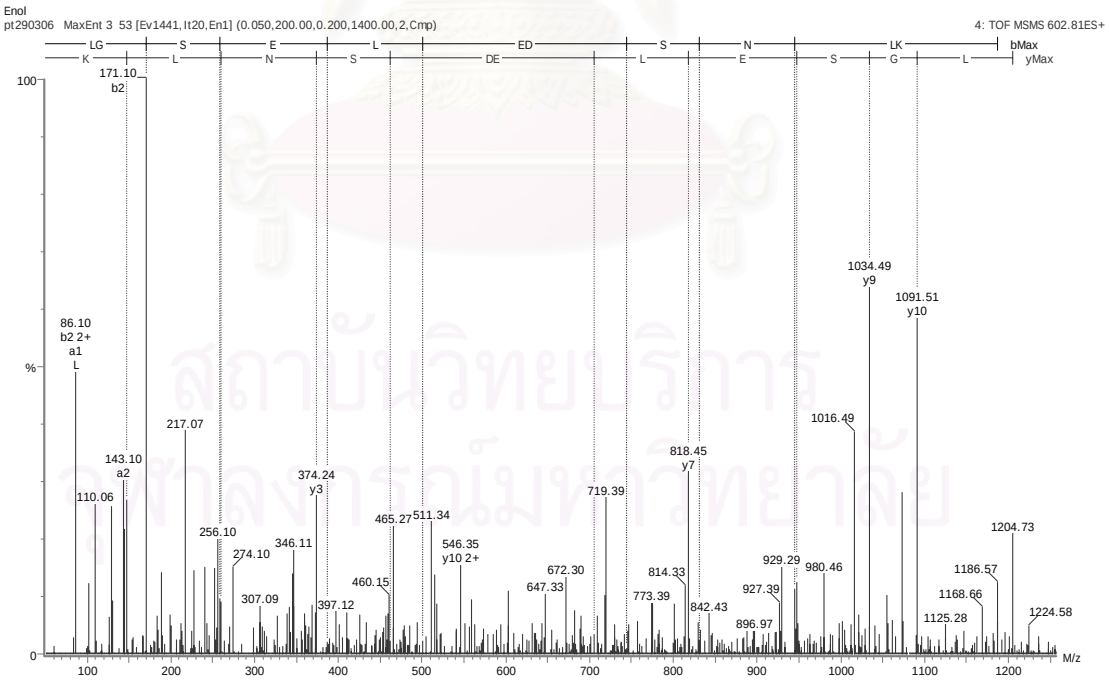



**Figure 8C** MS/MS spectrum of doubly charged precursor ion m/z of 602.81 from Q1S1

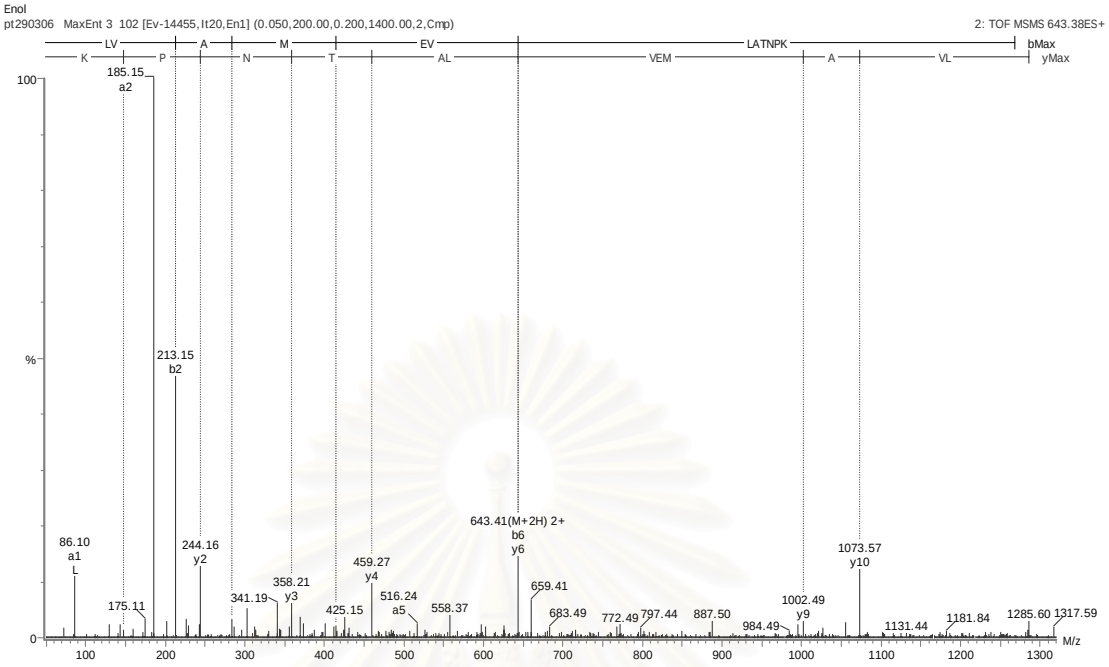



**Figure 9C** MS/MS spectrum of doubly charged precursor ion m/z of 643.38 from Q1S1

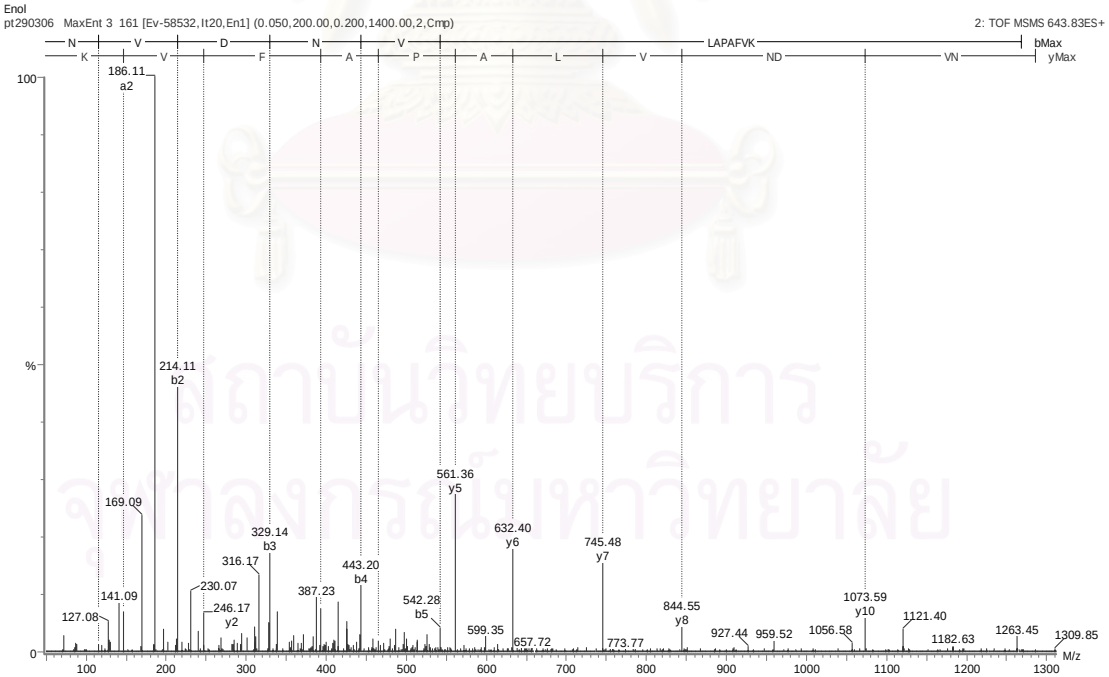



**Figure 10C** MS/MS spectrum of doubly charged precursor ion m/z of 643.83 from Q1S1

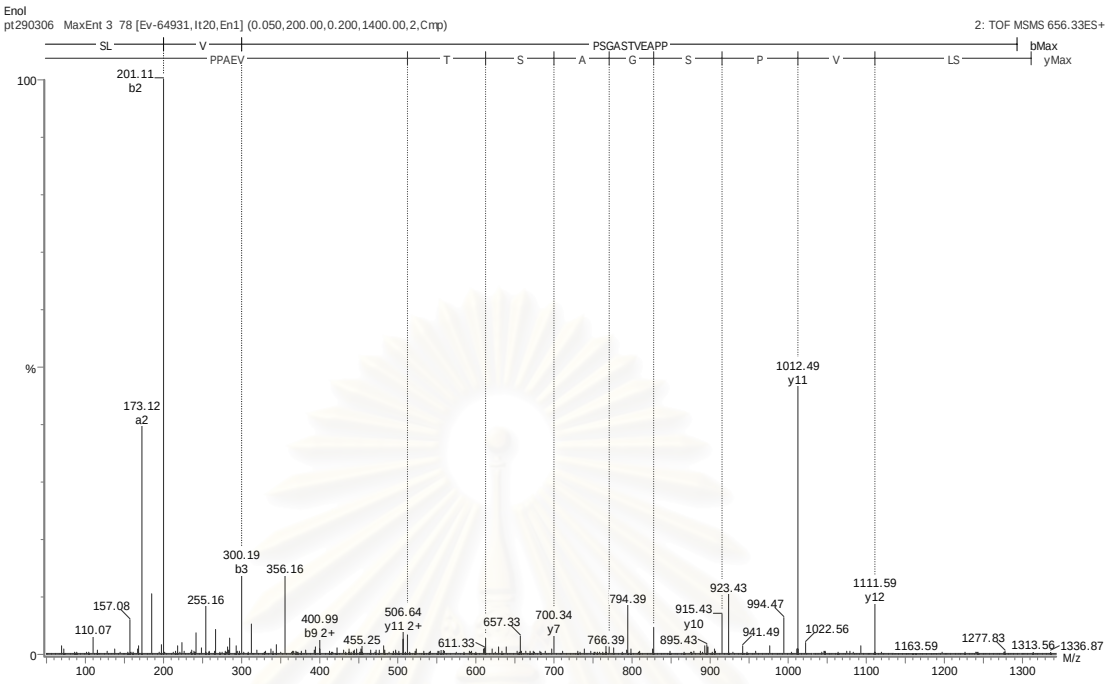



**Figure 11C** MS/MS spectrum of doubly charged precursor ion m/z of 656.33 from Q1S1

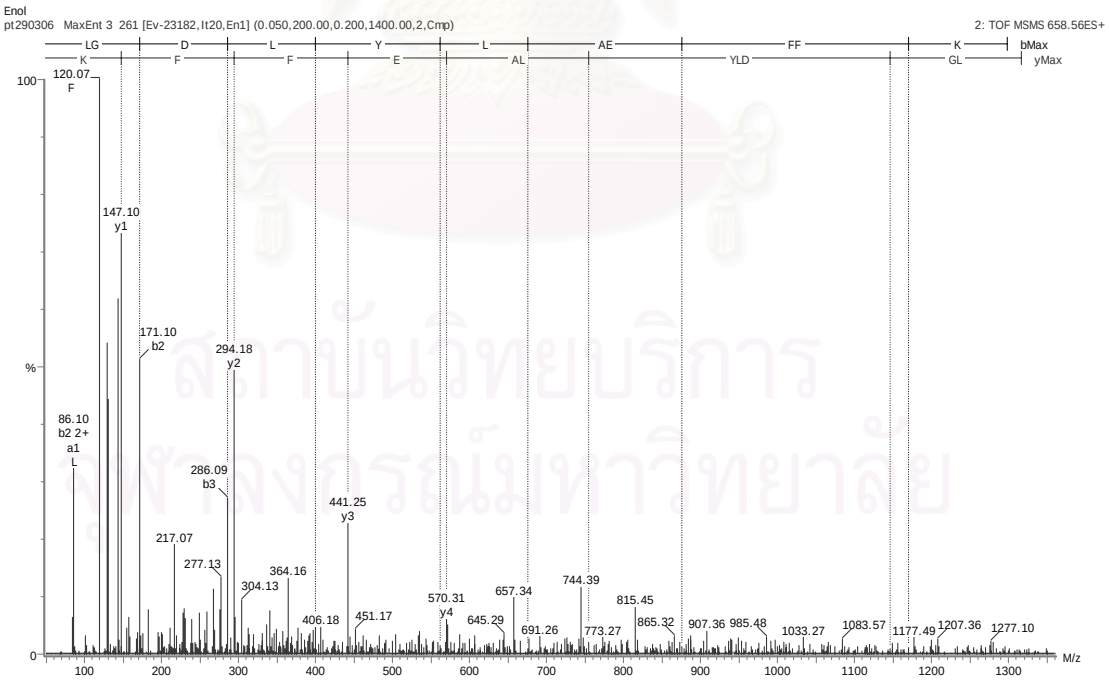



**Figure 12C** MS/MS spectrum of doubly charged precursor ion m/z of 658.56 from Q1S1

# VITA

Miss Ponpimol Tipthara was born on June 6, 1979 in Bangkok. She obtained a Bachelor's degree of Science in Chemistry in 2000 from the Faculty of Science, Chulalongkorn University. Since 2001, she has studied for a Doctor of Philosophy Degree in Organic Chemistry at the Department of Chemistry, the Faculty of Science, Chulalongkorn University. She had received financial support from the Thailand Research Fund in the program of the Royal Golden Jubilee Ph.D. scholarship. This gave her a good opportunity to receive an international experience of doing research at the University of Manchester in UK for a year. Also, she had received financial supporting from the Graduate School, the Faculty of Science, Chulalongkorn University. She completes her Ph.D. degree in 2006.