PROTEOME ANALYSIS OF MALE GAMETOPHYTE DEVELOPMENT IN RICE ANTHERS

A thesis submitted for the degree of
Doctor of Philosophy
at

THE AUSTRALIAN
NATIONAL UNIVERSITY

By

Tursun Kerim

Research School of Biological Sciences

November 2003

DECLARATION

The research presented in this thesis is my own work and does not contain results that have been generated by other persons, except where due reference and acknowledgement have been made. The work presented in this thesis has not been previously submitted for any degree at any institution.

Tursun Kerim

ACKNOWLEDGEMENTS

I would like to thank my supervisors Barry Rolfe and Jeremy Weinman for their guidance and support throughout my PhD study. I also acknowledge the other members of my supervisory panel, Charles Hocart and Michael Djordjevic, for their advice during my experiment. I whish to express my greatest gratitude to Barry Rolfe and Charles Hocart for their valuable suggestions to my thesis. I also thank my colleague and best friend, Nijat Imin, for his pioneering work on rice anther proteomics. My special thanks also go to all the previous and current members of the GIG group, in particular to my officemate, Giel van Noorden, for his company and interesting chats between experiments.

I am grateful to the Research Centre for Sustainable Rice Production for providing PhD tuition fee scholarship, to the Australian National University for ANU PhD scholarship, to the Australian Proteome Analysis Facility for PMF analysis, and to the Biomolecular Resource Facility of ANU for *N*-terminal sequencing analysis.

I would also like to thank my family friends who helped us in many ways to make our life in Canberra more meaningful: Mahira Nizamidin, Jillian Craven (Vim), Ceciliya Gutierrez, and Vanessa Young. I especially thank Vanessa for her proof reading of my thesis.

My special thanks also go to my parents and other family members back in far western China. On this happy moment I also remember with great sadness my brother and best friend, Nabijan Kerim, who perished in a massive fire accident to rescue school children.

At last, my greatest thanks go to my loving wife Aynur Ahmat who brought me all the happiness and our lovely daughter, Nabila, into my life. Without her support this thesis will never have happened. I love you!

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

2-DE	two-dimensional gel electrophoresis
BCIP	5-bromo-4-chloro-3-indolyl phosphate
BLAST	basic local alignment search tool
BSA	bovine serum albumin
CAM	carboxyamidomethyl
CAPS	3-[cyclohexylamino]-1-propanesulphonic acid
CBB	coomassie brilliant blue
CHAPS	3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulphonate
DAPI	4,6-diamidino-2-phenylindole
DIGE	two-dimensional difference gel electrophoresis
DTT	dithiothreitol
EBN	early binucleate
ECL	enhanced chemiluminescence
ESI	electrospray ionisation
EST	expressed sequence tag
EYMS	early young microspore
FAA	formalin-acetic-alcohol
HEPES	N-(2-hydroxyethyl)-piperazine-N'-(2-ethanesulfonic acid)
HPLC	high performance liquid chromatography
IAA	iodoacetamide
ICAT	isotope-coded affinity tag
IEF	isoelectric focusing
IPG	immobilised pH gradient
kDa	kilodaltons
LBN	late binucleate
LC	liquid chromatography
MALDI	matrix-assisted laser desorption/ionisation
Mbp	Megabase pair
Milli-Q	high purity water (resistance greater than18 $M\Omega.cm^{-1}$)
min	minute
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MudPit	multidimensional protein identification technology
MW	molecular weight
MYMS	middle young micropspore

NBT	nitro blue tetrazolium
NCBI	national centre for biotechnology information
OD	optical density
ORF	open reading frame
OsGI	*Oryza sativa* gene index
PAGE	polyacrylamide gel electrophoresis
PAM	propionamide cysteine
p*I*	isoelectric point
PMF	peptide mass fingerprinting
PMC	pollen mother cell
ppm	parts per million
PSD	post-source decay
PTM	posttranslational modification
PVDF	polyvinylidene difluoride
RP	reverse phase
rpm	revolutions per minute
s	second
SDS	sodium dodecyl sulphate
SCX	strong cation exchange
TBS	tris buffered saline
TBST	tris buffered saline with tween
TC	tentative consensus
TCA	trichloroacetic acid
TFA	trifluoroacetic acid
TIFF	tagged image file format
TIGR	The Institute for Genomic Research
TOF	time-of flight
Tris	tris-(hydroxymethyl)-aminomethane
Vh	volthours
v/v	volume to volume
w/v	weight to volume

ABSTRACT

In this thesis, two-dimensional electrophoresis (2-DE)-based comparative proteomics were applied to monitor the global changes in protein expression during the male gametophyte development in anthers of the Australian rice cultivar Doongara, with the aim of providing a protein-level insight into the molecular mechanism underlying this important reproductive developmental process. At the same time, this thesis also evaluates the potential application of 2-DE based proteomics to other aspects of plant developmental biology.

In order to collect sufficient amount of homogenous anther populations which represent a number of discrete cellular events encompassing the process of microsporegenesis, the cytological examination of developing anthers was done to establish the allometric relationships between a number of growth parameters and anther developmental stages. This provided a base for the quick and nondestructive assessment of microspore developmental stages. From this study, a strategy for the collection of rice anthers for six developmental stages was established using a combination of anther length, auricle distance and days before heading. The findings of the cytological analysis opened up the possibility of establishing rice plants as a new model system for male gametophyte research of plants.

Anther proteome maps were established for six microspore developmental stages within the pH ranges of 4 to 7 and 6 to 11. Over 3,500 protein spots were reproducibly resolved in the combined pH range of 4 to11. Comparison of proteome maps of six developmental stages resulted in the detection of 150 differentially displayed protein spots at various stages. Putative identities were predicted for 49 out of 155 protein spots which were subjected to peptide mass fingerprinting (PMF) analysis. Eight low molecular weight protein spots were matched to putative translation products of rice expressed sequence tags (EST) by *N*-terminal terminal sequencing followed by homology searches. This verified the translation of these small open reading frames (ORF) and revealed the presence of some post translational modifications of these proteins. By integrating the information about the functions of identified proteins and their temporal regulation patterns, three developmentally regulated metabolic pathways

were identified and the significance of these pathways in relation to male gametophyte development was discussed.

Based on the *N*-terminal sequencing data, three isoforms of rice homologues of grass group II pollen allergens (Ory s 2) were identified and further characterized using bioinformatics and immunochemical techniques. Polyclonal antibodies were produced against Ory s 2 isoforms using gel-separated proteins as the antigen. Immunoblot analysis revealed that Ory s 2 proteins are pollen specific and accumulated to high abundance at mature pollen, indicating their possible involvement in fertilization process. Immunochemical analysis also showed that rice group II allergens do not possess cross-reactivity with group II allergens of other grasses.

This study produced valuable molecular data to provide some insight into the global changes of protein expression accompanying pollen development, and identified some developmentally regulated protein markers which have potential practical application to other research projects. From the promising results of this proteomic study it can be expected that our understanding of complex biological processes in plant development will be enhanced with the availability of a fully annotated rice genome and the application of integrated systems biology research approaches.

CHAPTER 1 GENERAL INTRODUCTION

1.1 INTRODUCTION

The life cycle of plants alternates between a diploid sporophytic generation and a haploid gametophytic generation. In higher plants, the sporophytic cells undergo meiosis to produce haploid spores, which subsequently develop into multicellular male gametophytes (pollen) and female gametophytes (embryo sac) by cell proliferation and differentiation. Certain cells in the male and female gametophytes differentiate into gametes (sperm and egg cells). Fusion of sperm cells with egg cells during fertilization gives rise to the diploid sporophytes, thereby completing the full life cycle of plants (Raven *et al.* 1992). In angiosperms, the successful completion of the above-mentioned reproductive development provides fruits and seeds that make up the bulk of the world's food supply.

Central to the study of plant sexual reproduction is the angiosperm male gametophyte development, which has been a subject of cytological, cytochemical, biochemical and modern molecular biological investigations (Mascarenhas 1990; Chasan and Walbot 1993). In spite of the extensive research effort, a limited number of male gametophyte specific genes have been identified in several model plants, and the molecular mechanism underlying this important developmental process is still poorly understood (Yang and Sundaresan 2000). With the rapid advancement in protein separation and identification technology and in bioinformatics, it has become possible to perform large-scale analysis of the protein complement of cells or tissues (Anderson and Anderson 1998; van Wijk 2001). This process of looking at the complete array of protein products expressed by a cell population or tissue under a predefined physiological condition or at a particular developmental stage is known as proteomics (Wilkins *et al.* 1996a).

Rice, *Oryza sativa*, is one of the most important cereal crops providing staple food for about half of the world's population (Sasaki and Burr 2000), with an annual production of approximately 600 million tons (http://www.irri.org/). In addition to its economical importance, rice has also become an attractive model system for cereal genomic research because of its relatively small genome (440 Mb) (Bevan and Murphy 1999), a high degree of genomic synteny with other cereal crop plants and amiability for

genetic transformation (Goff 1999)(Sasaki and Burr 2000). Rice is also the first crop plant whose genome is sequenced (Goff *et al.* 2002; Yu *et al.*). The above-mentioned features of rice also made this model plant an ideal system for the investigation of male gametophyte developmental process. In this study we employed the state-of-the art proteomics technology to the study of plant male gametophyte developmental process in rice, making an attempt to generate comprehensive knowledge at protein level about the molecular mechanism of angiosperm male gametophyte developmental process. The cold sensitive rice cultivar, Doongara, has been chosen for use in this project because the knowledge which is resulted from this study can directly be compared and utilized for the other research project of our group, which investigates the molecular mechanism of cold-induced male sterility in rice using the same cultivar.

This review is in three parts. The first part provides an overview about the male gametophyte developmental process in angiosperms and the progress made in male gametophyte research. The second part provides an overview of the proteomics including principles, instrumentation, technology, recent advances and other related global expression analysis. The third part summarizes the application of proteomics in plant science.

1.2 PLANT MALE GAMETOPHYTE DEVELOPMENT

In this section, the reproductive development of angiosperms is described with the primary focus on male gametophyte development in rice. Different techniques to study male gametophyte and the progress made in male gametophyte research in recent years are also reviewed.

1.2.1 Life cycle of angiosperms

The angiosperms comprise about 235,000 species, by far the largest number of species of any plant group. They are divided into two main classes: the Monocotyledons (monocots) and Dicotyledons (dicots). The monocots and dicots exhibit a very similar life cycle in spite of some morphological differences. In Figure 1.1, the life cycle of angiosperms are illustrated for rice, a representative monocot plant.

Cultivated rice is a semi-aquatic annual grass, although in the tropics it can also survive as a perennial. The life cycle of the rice plant is 3-6 months, depending on the

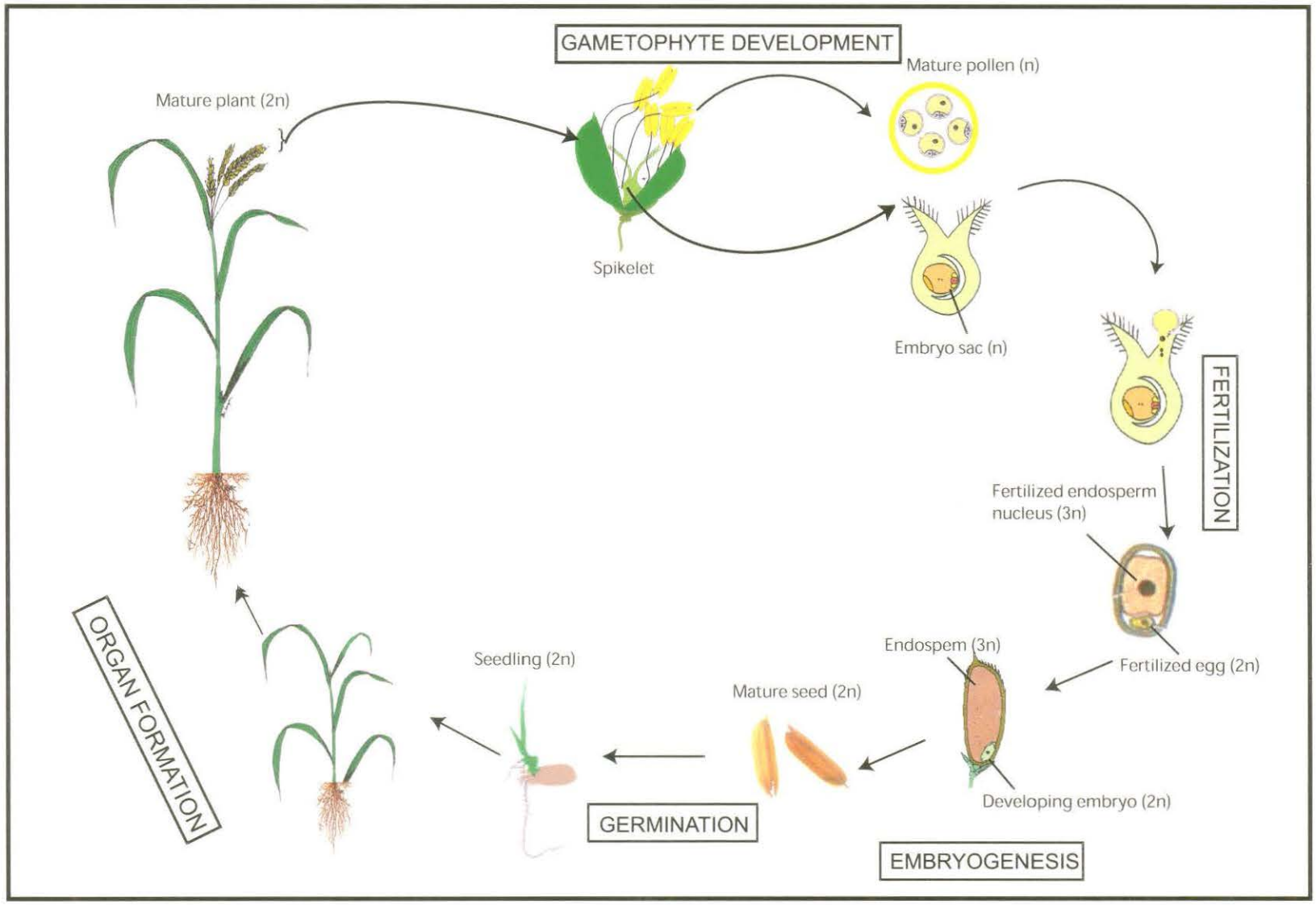


Figure 1.1. Schematic diagram of life cycle of a rice plant.

variety and environment under which it is grown. During the growth cycle, rice completes two distinct phases: vegetative and reproductive. As in other angiosperms, the vegetative growth starts with germination of a rice seed and ends with panicle initiation. The vegetative phase can also be subdivided into germination, seedling and tillering stages. During vegetative growth, the plant develops into a mature plant, which possesses the necessary structures for subsequent reproductive development.

Transition of the shoot apical meristem from producing vegetative structures to producing inflorescence branches and floral bracts marks the initiation of the reproductive development phase. During this phase, the floral meristem cells eventually differentiate into gametophytic organs: stamen and pistil (carpel). A male gametophyte, or pollen grain, arises from a microspore produced within stamen while a megaspore develops within ovules of pistil into a female gametophyte, or embryo sac. The process of male gametophyte development in rice will be reviewed in detail in section 1.2.3. Pollination in angiosperms is indirect in that pollen is deposited on the stigma of the pistil rather than directly on the ovule as is the situation in gymnosperms. The pollen grain on a receptive stigma initiates growing pollen tube, which delivers two sperms to the embryo sac. Ultimately, one sperm nucleus (n) enters the egg cell (n), resulting in the formation of a diploid zygote (2n). The other sperm nucleus (n) enters the central cell (2n), where it fuses with the two polar nuclei to form a triploid (3n) primary endosperm nucleus. This fertilization process is called double fertilization and it represents one of the unique characteristics of angiosperms. In addition to the sporophyte and the gametophyte, angiosperms have a third life stage, embryogenesis. This stage is also called ripening. After double fertilization, the zygote undergoes cell divisions and develops into an embryo with shoot and root apical meristems and cotyledons for subsequent development. It is with the formation of cotyledons that a distinction first emerges between dicot and monocot embryos: the dicots develop two cotyledons whereas monocots produce only one. The fertilized primary endosperm cell, in the meantime, divides to become a terminally differentiated mass of tissue that provides nutrition to the growing embryo. After the successful accomplishment of the fertilization process, the tissue of the ovary develops into a fruit and ovule into a seed.

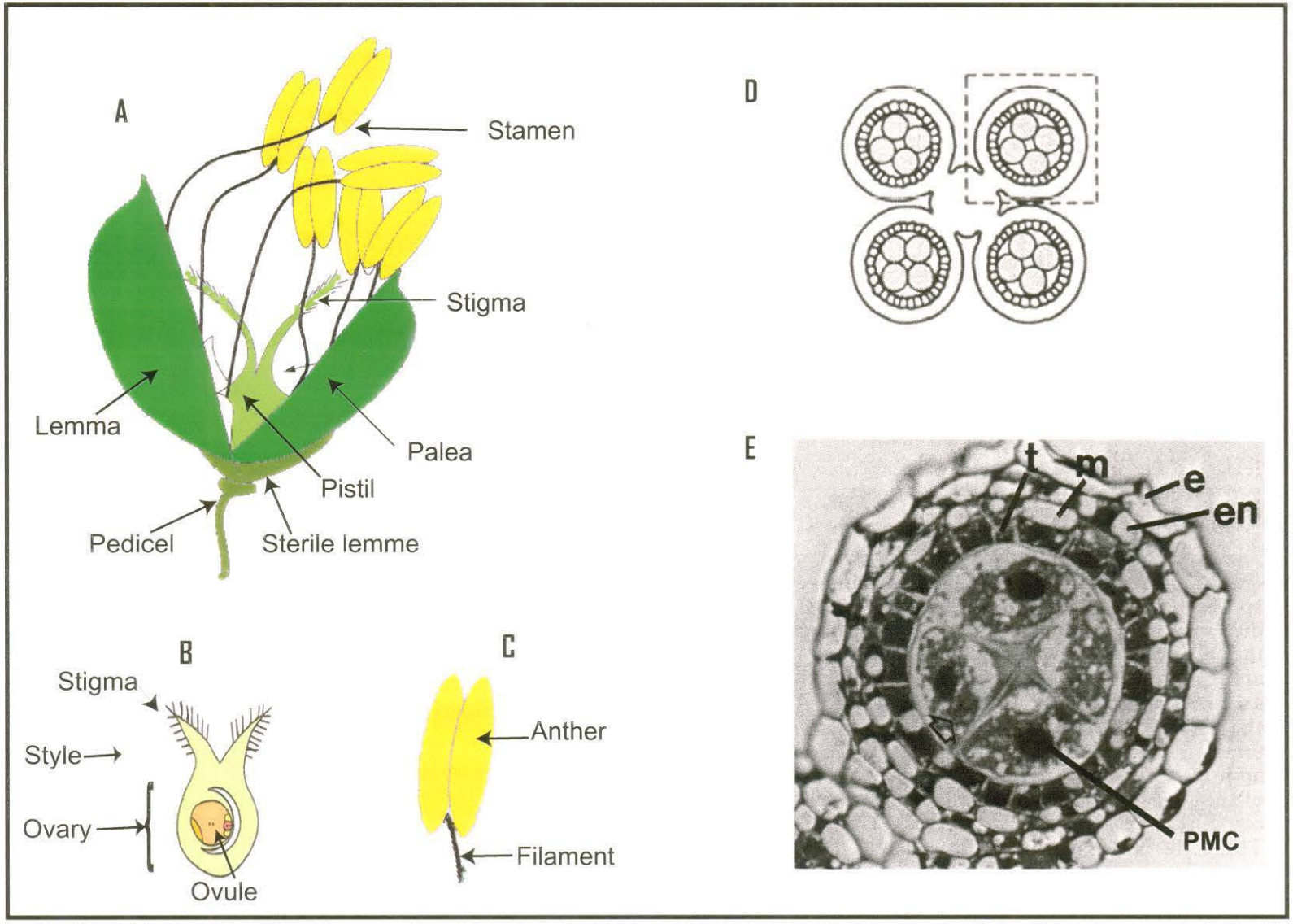
1.2.2 Anatomy of rice flower

The rice flower is called a spikelet. The spikelets occur in a group on a loose and irregular compound flower cluster called a panicle. Figure 1.2 illustrates the structure of a rice spikelet. Each individual spikelet contains a set of flower parts flanked by the palea and lemma, which are hardened and modified stems that protect the floral organs. The pedicel connects each spikelet to the secondary rachis of the panicle. The spikelet consists of six stamens and a pistil. Each stamen comprises an anther borne on a slender filament. The filament is a tube of vascular tissue that connects the anther to the flower and serves as a conduit for water and nutrients. Each anther has four elongated macrosporangia, or anther loculi. Four anther wall layers (the epidermis, the endothecium, the middle layer and the tapetum) enclose a locule where the microsporegenesis took place (Figure 1.2). The pistil consists of an ovary, a style and two feathery stigmas. The female gametes are produced inside the ovary. The stigmas receive pollen grains from stamens and conduct it down to the ovary via the pollen tube, where the double fertilization occurs.

1.2.3 Overview of male gametophyte development in rice

As mentioned in the introduction, the main subject of this study is the male gametophyte developmental process in a model monocot plant, rice. As the structural details of male gametophyte development are quite uniform in most angiosperms, the following account of male gametophyte developmental process in rice anther illustrated in Figure 1.3 is of general applicability to other angiosperms.

At the end of vegetative growth, the flag leaf primordium differentiates into the panicle primordium. The initiation of panicle formation consumes the shoot apical meristem and prevents a stem from undergoing further vegetative growth. The apical meristem (AM) of the panicle primordium gives rise to a primary branch primordium and the AM of which, in its turn, differentiates into a secondary branch primordium and a pedicel primordium. The AM of the pedicel primordium developes a spikelet primordium and ultimately forms a spikelet with a differentiated stamen and a pistil. At this point, the flower initiation process come to an end and the development of male and female gametophyte begins.



(A) Outline of a rice spikelet.
(B) Diagram of a rice pistil.
(C) Diagram of a rice anther
(D) Cross-section of a rice anther
(E) A micrograph of an anther locule at PMC.

PMC- pollen mother cell
en- endothecium
e- epidermis
m- middle layer
t- tapetum

Figure 1.2 Schematic representation of a rice flower. Pictures A and B were modified from Izawa et al (1996); D was adapted from Bedinger (1992), and E was adapted from Raghavan (1988).

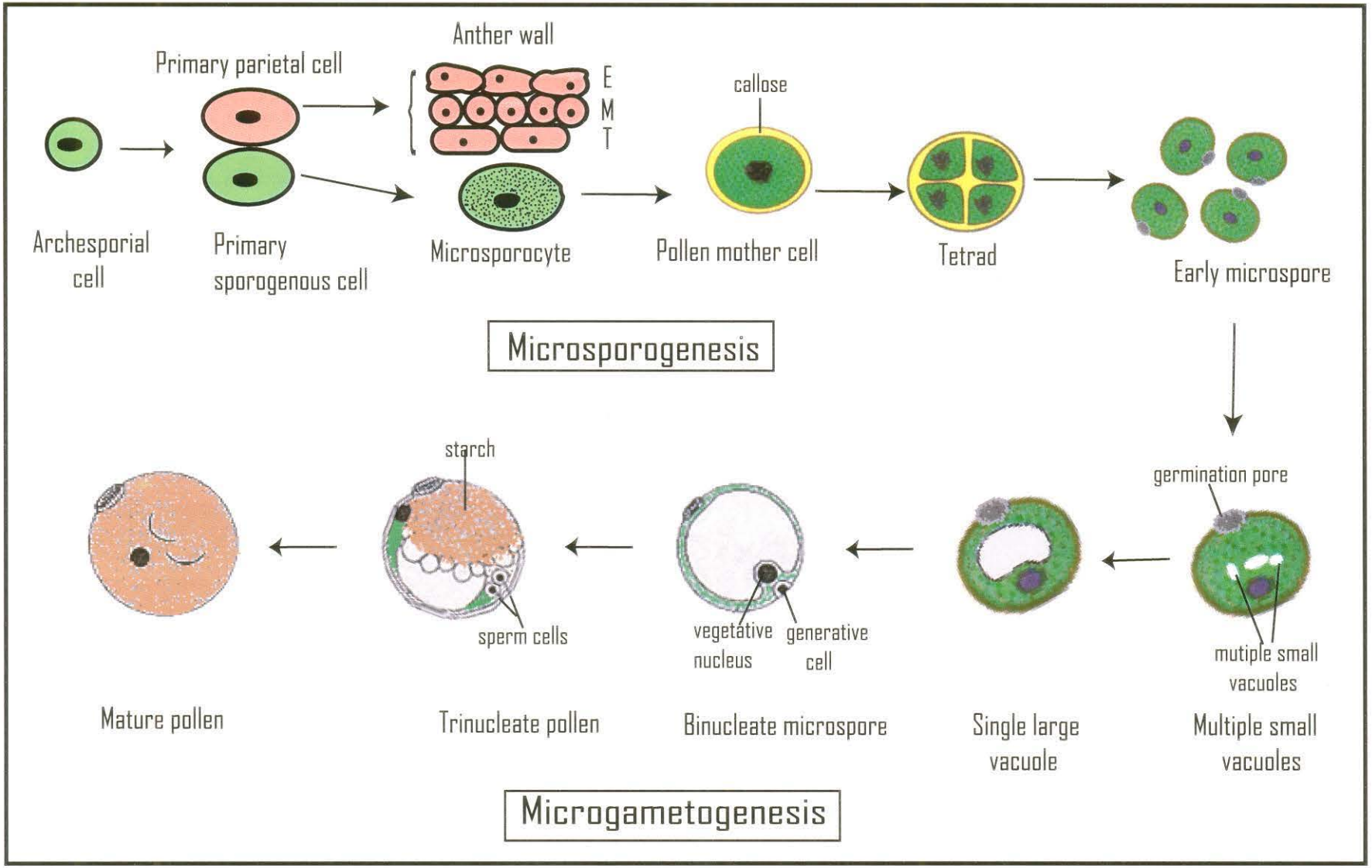


Figure 1.3. Schematic diagram depicting the sequential cell events of male gametophyte development in rice

The plant male gametophyte development process can be divided into two consecutive phases: microsporogenesis and microgametogenesis. Although these two terms are used interchangeably in general use, the former indicates the formation of microspores within an anther locule while the latter points to the development of a microspore into a fully developed microgametophyte, or a pollen grain. Each of these two phases progress through of a series of characteristic stages involving both gametophytic and somatic cells (Figure 1.3).

1.2.3.1 Microsporogenesis

Pollen mother cell formation stage: In a newly differentiated spikelet, the anther consists of a mass of undifferentiated meristematic cells, which are surrounded by a partially differentiated epidermal cell layer. As development progresses, the archesporial cells appear at the four corners of the young anther and further differentiate into primary parietal cells and primary sporogenous cells. Later the primary parietal cells differentiate into endothecium, middle layer and tapetum that envelop the core of the anther locule, whereas the primary sporogenous cells develop into diploid pollen mother cells (PMC). The pollen mother cell secretes a thick wall of callose, a β(1,3)-glucan, between the cell membrane and cell wall. The callose wall acts as a barrier between the pollen mother cells and the rest of the anther locule. The innermost tapetum layer of the anther locule provides the sporogenous cells with enzymes, nutrients and structural materials necessary for development.

Meiotic stage: Upon establishment, each pollen mother cell undergoes two rounds of meiosis in a short time period, producing a tetrad of four haploid cells that are still encased inside the callose wall. At this time point, the four haploid daughter cells are called microspores. Meiosis occurs synchronously throughout the hundreds of pollen mother cells in the six anthers within a spikelet. Immediately after the end of the second meiotic division, the pollen wall formation begins.

Young microspore stage: The callose wall is dissolved by callase, β(1,3)-glucanase), which is secreted into the anther locule by the tapetal cells, and the young microspore encased inside the callose wall will be released into the anther locule. The released young microspores arrange themselves as a ring around the anther locule in close proximity to the tapetum. Thereafter, an exine wall layer starts to establish on the

surface of the microspores from sporopollenin, a highly chemically resistant polymer secreted by tapetal cells. At the time of release from the tetrad, the microspores have already attained one-fourth the size of mature pollen grains and the major features of pollen grains are established. Each microspore contains one centrally placed nucleus. At this time point, microsporegenesis is completed with the formation of a single celled young microspore, and subsequently the microgametogenesis starts.

1.2.3.2 Microgametogenesis

Vacuolated microspore stage: The microspores steadily increase in size and the newly formed exine wall of the microspore thickens. The germination pore also becomes visible. Unlike dicots, which usually form multiple germination pores, rice pollen form only one germination pore on the pollen surface. The centrally located haploid nucleus migrates toward the cell periphery to a position opposite the pollen pore. At the same time the young microspore fill with multiple small vacuoles that eventually combine into a single large vacuole, which will compress the cytoplasm into a small region opposite the germination pore.

Bicellular pollen stage: At the end of the vacuolated microspore stage, the nucleus of the microspore undergoes an asymmetric mitosis (first haploid mitosis) to produce vegetative and generative nuclei. An arched cell wall cuts off a small amount of condensed cytoplasm around the generative nucleus to form the generative cell, while the rest of the microspore cytoplasm, including a large single vacuole and the vegetative nucleus, forms the vegetative cell. This "cell inside a cell" bicellular product of first haploid pollen mitosis is called a pollen grain. At the same time the intine, a final layer of the pollen wall, is synthesized from polysaccharides and proteins secreted by the pollen protoplast.

Trinucleate pollen stage: The nucleus of the vegetative cell becomes larger and moves closer to the germination pore. Then the generative cell, which has already detached from the pollen wall, also moves closer to the vegetative cell in the central part of the pollen. Starch accumulation (pollen engorgement) begins in the cytoplasm near the germination pore and the vacuole size will gradually be reduced by accumulating fresh cytoplasm. At this stage, the generative cell produces two sperm cells by second haploid mitosis. In nearly 70% of plant families this second mitosis takes place while

the pollen tube grows though the style tissue. By the time of anther dehiscence, the pollen grains become fully engorged with starch granules and attain their full size. With the formation of mature starchy type pollen the microgametogenesis process will come to an end.

1.2.4 Different approaches to investigate male gametophyte development in plants

Plant reproductive development has been one of the most exciting fields of plant research and male gametophyte development has been the main focus in this field because of the ready availability of large numbers of pollen grains or microspores (Knox *et al.* 1993). Compared with the sporophyte, the structure of a microgametophyte is much simple and is readily accessible for cytological and molecular analysis. A series of unique cell events that occur during the process of pollen development typify all of the major aspects of plant development such as organogenesis, differential cell division, mitosis, tissue degeneration, cellular differentiation and alteration in gene expression. For these reasons, male gametophyte represents an attractive model system for the study of plant development. Different approaches including cytological, molecular and protein related research tools have been applied to number of model plants in an attempt to elucidate the molecular processes that are responsible for flower initiation, gametophyte development and fertilization.

1.2.4.1 Nucleic acid related methods

Underlying the male gametophyte development is precise spatial and temporal coordination of cell-specific gene expression within anthers. Large-scale RNA-excess DNA/RNA hybridization experiments on tobacco pollen anthers show that around 26,000 diverse genes are expressed in tobacco anthers and approximately 11,000 of these are anther specific (Kamalay and Goldberg 1984). Similar studies on pollen and shoot tissues of *Zea Mays* and *Tradescantia* also estimated that about 10% and 20% of the total sequences expressed in *Zea Mays* and *Tradescantia pollen*, respectively, might be pollen specific (Willing and Mascarenhas 1984;Willing *et al.* 1988). According to their expression pattern, anther specific genes can be classified into "early" genes that become active soon after the completion of meiosis, and "late" genes whose transcripts are first detectable after first pollen mitosis (Mascarenhas 1990). The accumulation and

expression of "late" genes during the latter part of pollen maturation also highlights the major functions of these genes during pollen germination and pollen-stigma interaction. These findings suggest that transcriptional activation of male gametophyte-specific genes in a temporal manner is required for the establishment and maintenance of differentiated cell types and functions during anther development (Koltunow 1990).

Identification of stamen specific genes is an important step in understanding the development of male gametophyte development at the molecular level since these genes can provide clues about the metabolic pathways that are active during the process. Therefore, a number of male gametophyte-, or flower- specific genes have been isolated. There are several strategies that can be used for the identification of genes which are predominantly expressed in male gametophyte. Differential hybridization of pollen or anther cDNA libraries is one of the most often-used methods to isolate flower specific genes (Gasser 1991). In this approach the replica filters of pollen or anther cDNA libraries are separately hybridized with a probe prepared from pollen mRNA and with a probe made from mRNA from vegetative organs of the same plant. The intensity of the signal produced by each clone on the filters reflects the steady state level of mRNA from the corresponding gene. Clones that produce strong signals with the pollen probe and no signals with the vegetative probe represent the male gametophyte specific genes. This approach has been applied successfully in the isolation of stamen and pollen specific genes from tobacco (Kamalay and Goldberg 1980; Goldberg 1988; Koltunow 1990; Thangavelu *et al* 1993; Tebbutt *et al.* 1994; Oldenhof *et al.* 1996; Rogers *et al.* 2001), from *Brassica napus* (Albani *et al.* 1990; Hird *et al.* 1993; Treacy *et al.* 1997), from *Brassica campestris* (Theerakulpisut *et al.* 1991), from tomato (Twell *et al.* 1989; Ursin *et al.* 1989; Smith *et al.* 1990;), and from *Zea mays* (Stinson *et al.* 1987; Hanson *et al.* 1989; Turcich *et al.* 1993). This method has also been used to isolate genomic clones of differentially expressed genes directly from *Arabidopsis thaliana* (Simoens *et al.*; 1988Roberts *et al.* 1993b).

The "subtractive hybridization" method, an alternative to the method described above, relies on producing a population of RNA or cDNA from which the commonly expressed sequences have been removed. The cDNA produced from the male gametophyte is hybridized in solution to an excess of RNA from vegetative tissues and passed over a hydroxylapatite column on which the hybridized sequences are retained. The cDNA populations in the flow-through fraction represent the male gametophyte

specific genes. By using this method, Kobayashi *et al* have isolated cDNA clones which are preferentially expressed at the meiotic phase of microsporogenesis in *Lilium longiflorum* (Kobayashi *et al.* 1994). Although the above-mentioned male gametophyte specific genes demonstrated certain degrees of sequence similarity to some known proteins in databases, their exact biological functions during the male gametophyte development is largely unknown. Genetic analyses of male gametophytic mutants have also started to provide insights into the function of these genes (Yang and Sundaresan 2000).

The development of methods to isolate sperm cells from pollen (Blomstedt *et al.* 1996; Faure *et al.* 1994; Kranz and Dresselhaus 1996; Kranz and Lorz 1993) have opened up a new phase in plant reproductive research which aims to isolate sperm or generative cell-specific genes from the total pollen cDNA population. Using a differential hybridization approach, Xu *et al.* have isolated and characterized several generative cell-specific genes from Lilium (Xu *et al.* 1999a; 1999b). More recently, a generative cell-specific clone has been isolated from the isolated plastid-deficient generative cells of the same plant (Mori and Tanaka 2000).

1.2.4.2 Protein related methods

A number of protein-based approaches have also been applied to the male gametophyte study. One of the protein-based approaches to study male gametophyte-specific gene expression involves the isolation of such genes with the aid of information already available on known gametophyte specific proteins. In such an approach, pollen-specific proteins are purified by a series of standard protein purification steps including centrifugation, ion exchange chromatography and HPLC. Then, degenerative primers, which are designed from partial amino acid sequences of purified pollen specific proteins, can be used for the cloning of corresponding genes by PCR- based cloning techniques. Cloning of Lol p II, a group 2 pollen allergen of *Lolium perenne*, and several other pollen allergens exploited such an approach (Sidoli *et al.* 1993; Asturias *et al.* 1997; Marknell DeWitt *et al.* 2002). As an alternative way to this approach, monoclonal or polyclonal antibodies that were raised against the purified proteins can be used for the immunoscreening of cDNA expression libraries prepared from pollen mRNA to isolate cDNA clones encoding corresponding proteins (Lifschitz 1988; Toriyama *et al.* 1998). By using the serum of pollen allergic patients, this approach has

led to the cloning of increasing numbers of mRNAs encoding pollen allergens (Twell 1994). A heterologous sheep sperm antibody was also used to identify an anther specific *Brassica* protein, which is conserved between the animal and plant kingdom (Blomstedt *et al.* 1997).

The second protein-based approach involves the identification of proteins of unknown function based simply on their specificity in a flower organ. Such proteins can be identified by the comparison of proteins from pollen and vegetative organs on one-dimensional, or two-dimensional gels (Gasser 1991). By using two-dimensional gel electrophoresis (2-DE), followed by autoradiographic protein detection, Schrauwen *et al* investigated the changes of gene expression during the microspore development of lily and tobacco by comparing the polypeptides synthesized *in-vitro* from the accumulated mRNAs at four discrete stages of microspore development (Schrauwen *et al.* 1990). Even though the identities of the 2-DE separated peptides were not established in this study, the differential protein expression profiles indicate that these proteins have important physiological functions during the microspore development. In other studies, comparison of both one- and two- dimensional electrophoresis profiles of vegetative organs with those of developing anthers of *Lilium* at different stages led to the detection of 10 anther organ-enriched and 17 anther specific proteins (Wang *et al.* 1992b). Further characterization of two of those proteins, using 2-DE, western blot and immunolocalization analyses with polyclonal antibodies raised against gel purified proteins, revealed several isoforms of the same proteins and demonstrated their tissue specificity in the Lilium anther (Wang *et al.* 1992a, *1*993). The contribution of 2-DE based protein analysis technology to microspore research of maritime pine has been reviewed (Bahrman *et al.* 1997). Recently, the analysis of *Arabidopsis thaliana* pollen coat proteome using advanced proteome analysis technology resulted in the identification of two clusters of proteins involved in lipid metabolism, indicating the involvement of lipid molecules in pollen stigma interactions (Wolters-Arts *et al.* 1998; Mayfield *et al.* 2001). More recently, Kalinowski *et al* employed 2-DE electrophoresis to study the changes in the pollen protein composition of parent forms and amphiploids of *Aegilops kotschyi* and *Ae. variabilis* with *Secale cereale* (Kalinowski *et al.* 2001; Kalinowski *et al.* 2002). Two-dimensional electrophoresis also has been employed to study one of the most important biological processes, protein phosphorylation, using microspore embryogenesis in *Brassica napus* as a model system (Cordewener *et al.*

1994; Cordewener *et al.* 2000). Unfortunately, the identities of most of the proteins were not established in the above-mentioned studies.

1.2.4.3 Rice as a new model system for male gametophyte development

Rice, *Oryza sativa*, is one of the most important cereal crops providing staple food for about half of the world's population (Sasaki and Burr 2000). Because of its economical importance, many anatomical features of both the vegetative and reproductive parts of the rice plant have been extensively studied (Matsuo and Hoshikawa 1993; Raghavan 1988). In addition to its economical importance, rice has also become an attractive model system for cereal genomic research and its genome sequence is now available (Goff *et al.* 2002). The above nucleic acid- and protein-related methods have been applied to isolate a number of rice anther-specific proteins and genes (Tsuchiya *et al.* 1992; Tsuchiya *et al.* 1994; Zou *et al.* 1994; Chung *et al.* 1995; Xu *et al.* 1995a; Xu *et al.* 1995b; Hihara *et al.* 1996; Kang *et al.* 1997; Jeon *et al.* 1999; Zheng *et al.* 2000). Taking advantage of the available genomic sequence, identification of rice genes is also aided by the application of advanced proteomic technology as a functional genomics tool (reviewed in section 1.4). Apart from the relatively small size of rice flowers compared to other model plants (Lilium, Brassica, and tobacco), other exceptional features of rice, including sequenced small genome, genomic synteny with other cereal crop plants and ease of genetic transformation, also made this model plant an ideal system for the investigation of male gametophyte development. In an attempt to contribute to the knowledge of molecular mechanisms underlying the male gametophyte developmental process in crop plants, we employed advanced proteomics technology to identify and characterize stage-specific anther proteins using rice as a model plant.

1.3 CURRENT PROTEOME ANALYSIS TECHNOLOGY

1.3.1 Functional genomics in the post-genomic era

Over the past two decades, the biological sciences have witnessed enormous progress in characterizing the genomes of a wide variety of organisms. As a result of the coordinated effort of the international scientific community, complete genomes of a number of organisms have been deciphered (for a list of completed genomes visit NCBI

website at http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Genome). In addition to the completed genomic sequences, the number of EST sequences of different organisms is also accumulating very rapidly in databases (http://www.tigr.org/tdb/tgi/). The complete genomes of two angiosperms, model plant *Arabidopsis thialana* (Kaul *et al.* 2000) and model crop plant rice (Goff *et al.* 2002; Yu *et al.* 2002), have also been sequenced.

Genome sequences, however, can only provide a blueprint for all the putative open reading frames in a genome. Merely having a complete genome sequence is not sufficient to elucidate the function of a complex biological system. Therefore, the main task researchers face in the post-genomic era is predicting and assigning functions to possible genes in a genome to bridge the gap between genotype and phenotype (Bork *et al.* 1998). Putative open reading frames in a genome can be predicted using computer based comparative genomic tools. Because of the inadequacy of probabilistic or pattern based gene prediction programs, some small genes (<300 bp) can be easily overlooked and incorrect predictions can run as high as 30% (Pandey and Lewitter 1999; Bork 2000). Furthermore, gene prediction in eukaryotos is more complicated due to low coding density and the presence of introns (Mann and Pandey 2001). Once genes are correctly predicted from a genome, the functions of the predicted genes may be deduced by sequence alignment algorithms that compare DNA or deduced protein sequences of predicted genes with genes or proteins with known function in databases (Mann and Pandey 2001). Unfortunately, not all the genes can be annotated in this way since about a third of the predicted genes from a genome encode for proteins that display no significant homology to proteins of known functions in databases. Furthermore, similar gene sequences do not always imply a similar protein structure (Eisenstein *et al.* 2000). However, even elucidating the function of every single gene in a genome cannot provide insight into the ways in which an organism may modify and co-ordinate its pattern of gene expression in different cell states or at different times in the course of development. Therefore, in the post-genomic era, the reductionist way of studying one gene and one mutant at a time should give way to high-throughput multidimensional approaches of studying all genes and proteins in a biological system. Such systematic analyses of the possible functions of genes in the post-genomic era can take place at the oligonucleotide or protein level using functional genomics tools (Andersen and Mann 2000).

Functional genomics represents a new phase of genome analysis at different levels of gene expression, and is characterized by high-throughput and large scale experimental methodologies combined with statistical and computational analysis of the experimental results (Hieter and Boguski 1997). One of the newly-developed key technologies in functional genomics is high density DNA microarrays (chips) used for global transcriptional profiling or trascriptomics (de Saizieu *et al.* 1998). The array contains up to several thousands of gene-specific oligonucleotides or cDNA sequences on a slide or a chip that can be probed with fluorescently labeled or biotinylated cDNA obtained by reverse transcribing mRNA from samples of interest. Different methods for the production and hybridization of microarrays are extensively discussed in a number of reviews (Eisen and Brown 1999; Lockhart and Winzeler 2000). Important technical aspects of mRNA based microarray technologies include their high degree of sensitivity and ability to amplify the signal, which allows the analysis of low-abundance mRNAs (Holtorf *et al.* 2002). This technology allows the simultaneous quantification of transcript levels of all the genes represented on the chips, and eventually leads to the differential display of biological samples at the transcriptional level, or transcriptomics. However, the transcriptional expression patterns revealed by DNA microarray analysis are still insufficient to describe the underlying mechanism of cell activities resulting from differential gene expression. First, the expression levels of mRNA do not necessarily predict the levels of corresponding proteins. Separate studies show correlation coefficient between levels of mRNA and corresponding protein abundance levels can be as low as 0.48 (Anderson and Seilhamer 1997) and 0.356 (Gygi *et al.* 1999b). It was demonstrated in these studies that levels of protein expression coded for by some mRNA with comparable abundance varied between 30-50 fold. Secondly, transcriptional profiling does not provide any information on the process of post-translational modifications by which a single gene can give rise to more than one protein entity in a cell. These modified proteins comprise a substantial fraction of any cellular system, and many biological events taking place in cell networks are dependent on different types of post-translational modifications for regulating their activity (Cozzone 1998). Finally, protein-protein interactions involving the dynamic reorganization of cellular proteins into complex signaling assemblies cannot be predicted from mRNA expression data. Nor can you predict the identity or quantity of metabolites present in a cell from the mRNA expression data.

For realization of their true value, however, genomic sequences and transcriptional expression data should be related to the proteins they encode using protein analysis tools (Yee *et al.* 2003). Compared to genetic approaches, protein analysis tools have the advantage of being closest to the function of a gene in view of the fact that it is the proteins, which are expressed by the genes that ultimately perform most of the biological activities that take place within a cell. Therefore, during the last ten years there has been renewed interest in proteins. Researchers have developed and optimized new high-throughput protein separation and analysis tools that are suitable for genome-scale protein analysis. As a result of these efforts, proteomics, a new systematic research approach aiming to complement genomic study in protein levels, has emerged.

1.3.2 Proteome - a linguistic equivalent of genome

The term proteome, coined in analogy to the term 'genome' by Wilkins and colleagues in 1994 (Wasinger *et al.* 1995; Wilkins *et al.* 1996b), is used to describe the complete set of proteins that is expressed, and modified following expression, by the entire genome in the lifetime of a cell. Again by analogy to the term of 'genomics', proteomics is the systematic analysis of the proteins expressed by a genome. With its capability of analyzing thousands of gene products in parallel, proteomics has emerged as a complementary research tool to genomics in the post-genome era. The context of functional proteomics is systems biology, which characterizes the qualitative and quantitative behavior of all the components of a biological system rather than the behavior of every single component.

Three key technological developments in biological sciences have paved the way for proteomics technology. First, the technological advances in protein sample preparation and two-dimensional gel electrophoresis (2-DE) technology made it possible to reproducibly separate and visualize the complex protein mixture of a cell or tissue. Secondly, the development of mass spectrometry ionization techniques applicable to biopolymers (proteins, nucleic acids and carbohydrates) has increased the sensitivity, speed and robustness of protein identification in post-separation proteomic analysis. Thirdly, the accumulation of completely sequenced genomes in databases has provided proteomics technology with a 'blueprint' of possible gene products that are the focal points of proteomic analysis. Figure1.4 illustrates the key steps in standard 2 -DE based proteome analysis techniques.

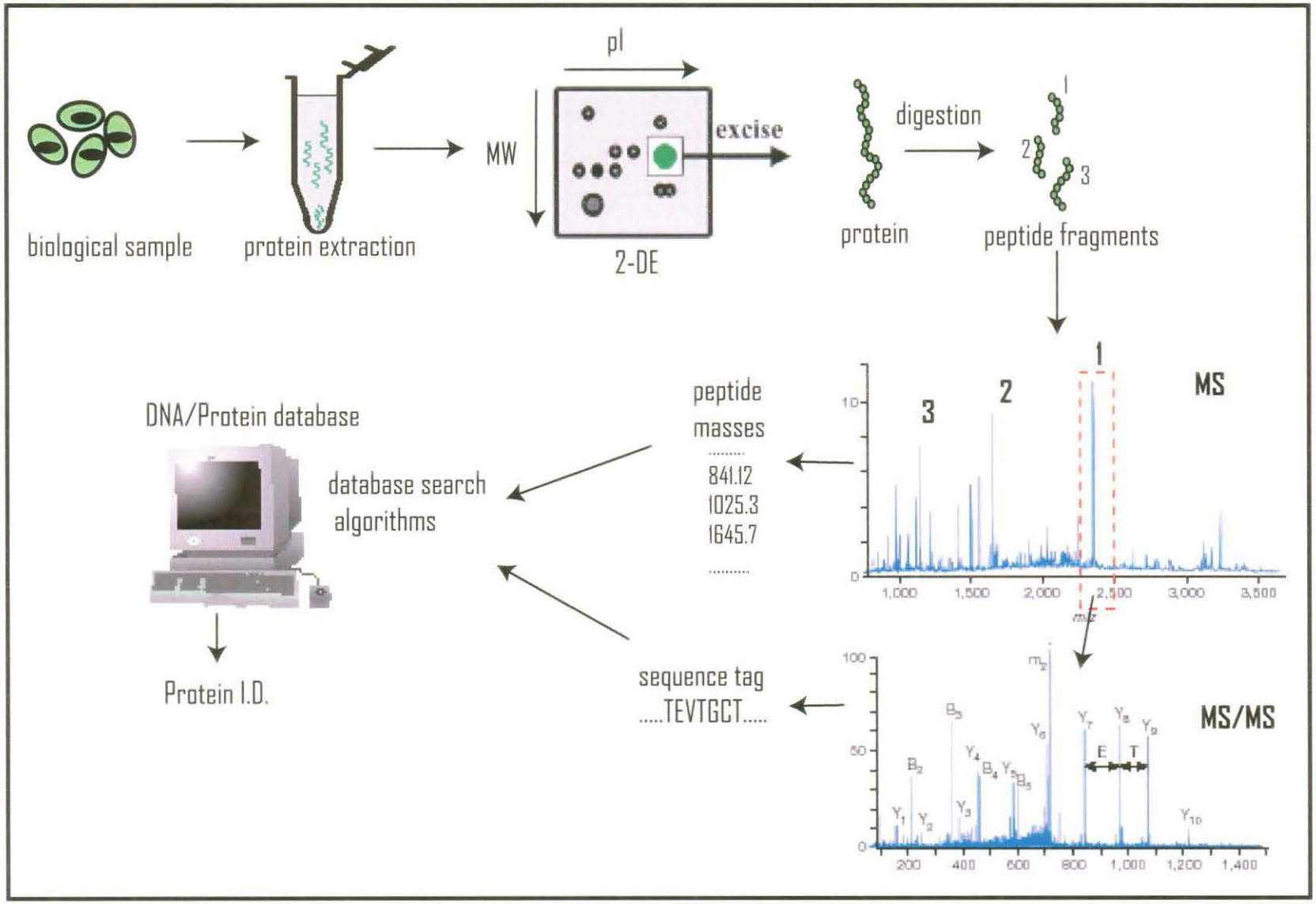


Figure 1.4. Schematic illustration of 2-DE and mass spectrometry based proteome analysis procedure.

1.3.3 Separation of proteins by two dimensional gel electrophoresis

For separation of complex mixtures of the protein content of cells or tissues, the most powerful technique available is two-dimensional polyacrylamide gel electrophoresis. The 2-DE technology was developed independently in 1975 in two separate laboratories by O'Farrel (1975) and Klose (1975). Although the 2-DE is now almost synonymous with proteomics, the ideas of building protein databases (such as the Human Protein Index) utilizing the high-resolution power of 2-DE long predates the term 'proteomics' (Anderson and Anderson 1982).

Essentially, two-dimensional electrophoresis separates proteins according to two independent physiochemical properties of proteins: the charge and size of protein molecules. In the first dimension, the proteins are resolved on the basis of isoelectric points by isoelectric focusing (IEF). Then, the isoelectrically resolved proteins are further separated in a perpendicular second dimension by SDS-PAGE according to their molecular weight. The combination of these two modes of separation enables 2-DE to resolve up to several thousand protein species in a single experiment (Klose and Kobalz 1995). In addition to its high-resolution power, the ability of 2-DE to deliver accurate and detailed information on p*I*, molecular weight, relative quantity, post-translational modifications, and solubility of the separated proteins, leaves it unsurpassable by any other currently available separation technique.

1.3.3.1 Sample preparation

Protein extraction and solubilization are critical steps for successful 2-DE experiments. Sample preparation in proteome analysis must be optimized to recover a maximum amount of proteins from the biological samples with the least amount of handling procedure. So far, there is no universal protocol for sample preparation for 2-DE, but generally such procedures involve the physical or chemical disruption of cells to release the protein content of the cell, the separation of proteins from other cell components which can interfere with the subsequent iso-electric focusing of the proteins, and the solubilisation of proteins in a suitable buffer containing detergents and other chemicals that are compatible with IEF (Rabilloud 1999). Next, the non-covalent interactions, the main forces holding proteins together and allowing binding to other non-protein cell compounds, need to be disrupted. This can be achieved using nonionic

or zwitterionic detergents such as CHAPS and NP-40 in a solubilization buffer to eliminate hydrophobic interactions, one of the major non-covalent interactions between proteins. The hydrogen bonds can be effectively disrupted by the denaturing power of a high concentration of urea in the solubilization buffer (Herskovits *et al.* 1970). It is also necessary to break disulfide bonds to obtain individual polypeptides. This can be achieved by adding reducing agents, usually mercaptoethanol and dithiothreitol (DTT), into the solubilization buffers. The standard urea-reducer-detergent solubilization buffer performs well with most biological samples. The optimal solubilization protocol, however, should be determined experimentally depending on the type of biological samples analyzed (e.g., animal or plant samples) and depending on the particular requirements of the experiment, e.g. targeting the membrane proteins or cytosol proteins. Advances in protein solubilization for 2-DE are reviewed elsewhere (Herbert 1999; Shaw and Riederer 2003).

1.3.3.2 Two-dimensional SDS-PAGE

In the early days of 2-DE, the first dimensional isoelectric focusing of proteins was carried out in polyacrylamide tube gels containing carrier ampholytes, which were a mixture of aliphatic polyaminocarboxylic acid compounds that form a stable pH gradient across the length of the IEF gel when voltage is applied. This traditional approach to IEF has several limitations. Apart from the relative technical difficulty of performing IEF in fragile tube gels, the unstable pH gradient generated by carrier ampholytes tends to drift towards the cathode over time. Thus, adding the batch-to-batch variation of the carrier ampholytes, makes it very difficult to generate highly reproducible 2-DE gels for proteome analysis (Righetti and Macelloni 1982). The introduction of immobilized pH gradient (IPG) chemistry, in which the pH gradient is fixed in the gels by ampholytes that are covalently linked to the polyacrylamide matrix, overcame the above-mentioned problems of isoelectric focusing (Bjellqvist *et al.* 1982). Since then, a basic protocol of 2-DE electrophoresis that was introduced by Görg *et al.* and the availability of pre-cast IPG strips from commercial suppliers have greatly improved the loading capacity, resolution and reproducibility of 2-DE analysis (Gorg *et al.* 1988). It has also expanded the possible pH gradient range of iso-electric focusing (pH 3-12), which enabled researchers to separate and analyze proteins with extreme p*I* values. These improvements in IEF technology have made it possible to differentially

display protein arrays and to compare proteome maps between the laboratories (Gorg *et al.* 2000).

After the first dimensional focusing step, the strips are treated with a buffer that contains sodium dodecyl sulphate (SDS) and reducing agents to coat protein molecules with net negative charge. Then the strips are placed on SDS-PAGE gels and the equilibrated proteins in the strips are resolved according to their molecular size on SDS-PAGE slab gels. Both the vertical and the horizontal SDS-PAGE system can be used for the second dimension. The availability of large format pre-cast SDS slab gels with plastic backing simplifies the gel handling and also improves the reproducibility and standardization of 2-DE procedure in research laboratories (Gorg *et al.* 1995).

1.3.3.3 Visualization and detection of proteins separated by 2-DE

Once separated, the protein spots need to be visualized and detected , before being subjected to subsequent comparative and analytical studies. There are a number of visualization strategies available for post-separation detection of proteins on 2-DE gels including single step staining with organic dyes, multiple step staining with metal salts and fluorescent dyes. When choosing a visualisation strategy for 2-DE gels, the compatibility of a staining method with subsequent microchemical characterization must be taken into consideration in addition to the sensitivity, linear dynamic range and reproducibility of the chosen visualisation technique.

Coomassie Brilliant Blue (CBB) dyes, which are the most commonly used organic dyes, are easy to use, low cost and fully compatible with down stream mass spectrometry and *N*-terminal sequencing analysis. The major limitations with CBB R-250 and CBB G-250 dyes are the poor detection sensitivity (30-100 ng and 8-10 ng, respectively) and considerably small linear dynamic range (10-30 fold) resulting in the under representation of low abundant proteins (Neuhoff *et al.* 1988; Patton 2002).

In terms of sensitivity and cost, the commonly used silver stain is the method of choice. It can detect as little as one nanogram of protein. However, the silver staining method has several disadvantages. The linear dynamic range of silver staining is restricted to only a 10-fold range. The staining process is also quite complex and gel-to-gel reproducibility is not very high (Quadroni and James 1999). Because the most commonly used silver staining methods use aldehyde-based chemicals in the initial gel

fixative step, this prevents the subsequent microchemical analysis by Edman sequencing or mass spectrometry. Compatibility of silver staining with microchemical analysis can be improved by employing an additional destaining procedure (Gharahdaghi *et al.* 1999), or by using a modified less-sensitive silver staining protocol that is compatible with mass spectrometry analysis (Shevchenko *et al.* 1996; Mortz *et al.* 2001).

Another staining method gaining popularity in proteomic analysis is fluorescent staining. Fluorescence stains such as the recently marketed SYPRO Ruby have the same sensitivity as the most sensitive silver staining method. Unlike the silver stain, it is endpoint staining and it requires a simple single step staining procedure. In terms of linear dynamic range and compatibility with mass spectrometry analysis, the 1000-fold dynamic range SYPRO Ruby surpasses both silver staining and CBB (Lopez *et al.* 2000). The only problem with SYPRO Ruby staining is that the stained protein spots can only be detected under the UV transilluminator or special imaging systems equipped with a laser source. An Australian biotechnology company, FLUOROtechnics, recently marketed a new fluorescent dye, the Lightning Fast, which was developed by scientists at the Macquarie University in Australia. The Lightning Fast is ultra sensitive (<60 pg) and has a similar linear dynamic range as SYPRO Ruby staining. The stain is also fully compatible with subsequent post-electrophoretic MS, Edman and C-terminal sequencing analyses (Ferrari *et al.* 2003).

After staining, the information in 2-DE gels can be interpreted visually, or by using specialized computer hardware and software which treat the gels as digital images. The most commonly used image digitization devices are charge coupled device (CCD) camera, document scanners and laser scanners (Horgan and Glasbey 1995). After image-acquisition, commercially available 2-DE gel analysis software packages can be used to detect and quantify protein spots on computer images for documentation, or to quantitatively compare 2-DE images to find differentially displayed proteins.

1.3.4 Post-separation protein identification

Protein identification is based on the development of three technological platforms; (i) a fast and sensitive method to generate sufficient and unambiguous structural information from gel purified proteins, (ii) availability of protein, EST (expressed sequence tag) or preferably the complete genomic sequence database, and (iii)

computer algorithms capable of interrogating the databases using various structural information (Gevaert and Vandekerckhove 2000). The main strategies of protein identification currently being employed at proteomic laboratories are immunoaffinity identifications using specific antibodies against known proteins, amino acid composition analysis, *N*-terminal or internal protein sequencing using Edman chemistry, peptide mass fingerprinting (PMF) analysis and tandem mass spectrometry (MS/MS) analysis. With improvements in the sensitivity and automation in Edman sequencing and the development of new ionization techniques in mass spectrometry, the last three techniques have become the most useful and rapid protein identification tools in current proteomic research.

1.3.4.1 N-terminal microsequencing and database searching

N-terminal and internal protein sequencing uses the chemistry developed by Per Edman in 1949 (Edman 1949). Whilst the basic principles of Edman chemistry have remained unchanged for five decades, many improvements have been made both in instrumentation and automation to increase the sensitivity and speed of protein analysis. Two major breakthroughs in instrumentation including the introduction of an automated spinning cup protein sequencer in 1967 (Edman and Begg 1967) and the development of a gas-liquid solid phase protein sequencer in 1980, with a ~100 increase sensitivity (Hewick *et al.* 1981). This transformed Edman sequencing into the most extensively used standard protein identification tool. The advent of sequencer stable membranes for protein-chemical analysis, together with the improvement in 2-DE separation and protein blotting technologies (Towbin *et al.* 1992), made it possible for *N*-terminal sequencing to obtain the sequence of up to ten amino acids using low picomole and sub-picomole amounts of protein from polyvinylidene difluoride (PVDF) membrane blots (Matsudaira 1987; Guerreiro *et al.* 1997). The sequence information obtained then can be compared to protein or translated DNA sequences in databases using computer searching algorithms to determine putative identity based on sequence homology. In addition to sequence data, *N*-terminal sequencing and homology searches can also provide information about the signal peptide cleavage sites.

One of the limitations with *N*-terminal sequencing is that it is incompatibile with proteins which are *N*-terminally blocked by acetylation, formylation or pyroglutamyl formation during the biosynthesis. The problem posed by *N*-terminal blockage can be

bypassed using the internal *N*-terminal sequencing approach developed by Aebersold *et al.* in 1987 (Aebersold *et al.* 1987). In this approach, Edman sequencing is conducted on each of the proteolytically derived peptide fragments from an in-gel or on-membrane protein digestion. Because of the additional steps involved, the internal sequencing approach is slower and less sensitive than direct *N*-terminal sequencing. Application of the Edman sequencing approach to high-throughput large-scale proteomic analysis is still limited by the high cost, relatively low sensitivity and time-consuming nature of this technique (one hour per amino acid residue). However, *N*-terminal sequencing is still extremely useful for generating amino acid sequence information where no database correlation is available (Wasinger and Corthals 2002).

1.3.4.2 Application of mass spectrometry to proteome analysis

Protein identification has been revolutionized by new enabling MS ionization techniques for large biopolymers – proteins, nucleic acids and carbohydrates. These are Matrix Assisted Laser Desorption Ionization (MALDI) (Karas and Hillenkamp 1988) and Electrospray Ionization (ESI) (Fenn *et al.* 1989). These two ionization techniques are coupled to any of the three types of commonly used mass separation methods. These mass separation methods are: separation on the basis of time-of-flight (TOF), separation by quadrupole electric fields (quadrupole MS), and separation by selective ejection of ions from a three dimensional trapping field (ion trap MS or fourier transform ion cyclotron, FT-ICR MS) (Martin *et al.* 2000). For structural analyses of peptides, two steps of mass spectrometry are performed in tandem (MS/MS), which can be done by employing the same MS separation principle twice (triple quadruple instrument, see Figure 1.5) or by combining two different MS separation principles (quadruple / time-of-flight, QTOF instrument) (Mann *et al.* 2001).

In principle, these mass spectrometers consist of three essential parts (Figure 1.5). The first is the source, which produces gas phase ions from the sample. The second is the mass analyzer, which separates the ions according to their mass to charge (*m/z*) ratio. The resolved ions are then detected by a third part, the detector (Yates 2000). The combination of 2-DE and MS provided both higher sensitivity and higher throughput than is possible with Edman sequencing and have essentially replaced Edman sequencing as the main protein identification tool in proteomic analysis (Pandey and

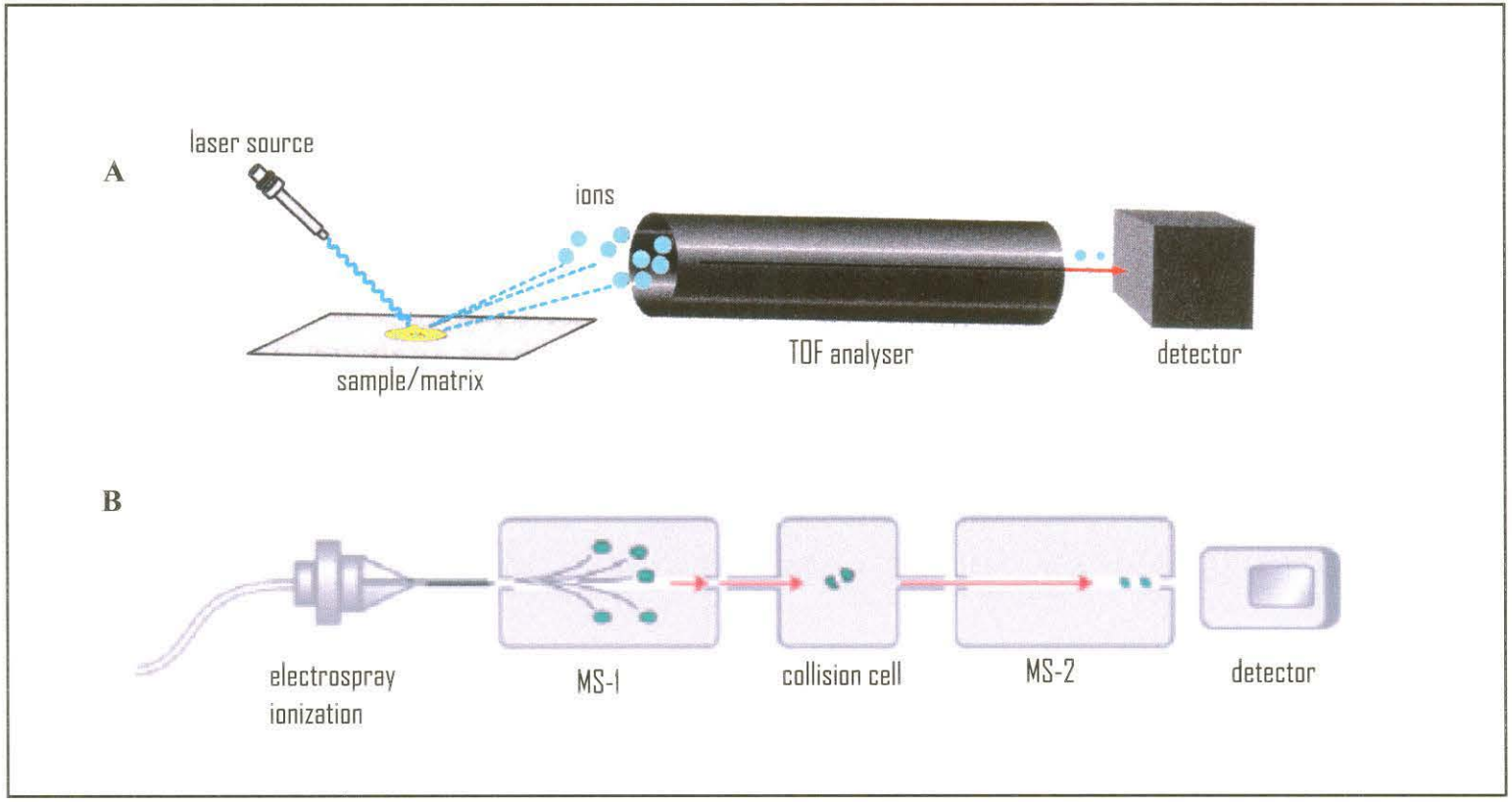


Figure 1.5. Schematic representation of a MALDI-TOF mass spectrometer (A) and a ESI-tandem mass spectrometer (B). The picture A was modified from Liebler (2002) and B was modified from Nelson and Cox (2000).

Mann 2000). The typical steps involved in 2-DE-MS based proteomic analysis are illustrated in Figure 1.4.

a. MALDI-TOF analysis and peptide mass fingerprinting

The first part (MALDI) refers to the ionization source, whereas TOF refers to the Time of Flight mass analyzer, which can measure the flight time of an ion to move through a field free region from an ion source to a detector. The resulting flight time spectrum is directly convertible into a mass spectrum, since the flight time of an ion is proportional to the square root of its *m/z* value:

$$t = D\,(m/z\ 2eEs)^{1/2}, \quad \text{so that} \quad m/z = 2\,eEs\,(t/D)^2$$

where *m/z* is the mass to charge ratio, *t* is the flight time of the ion, D is the length of the flight tube and *Es* is the applied potential (Lay Jr. and Holland 2000).

In a MALDI-TOF mass spectrometer, the sample to be analyzed is mixed with a laser energy-absorbing matrix, typically an aromatic acid such as α-cyano-4-hydroxycinnamic acid (CHCA), sinapinic acid (SA) and dehydrobenzoic acid (DHB). The mixture of sample and matrix are spotted onto a target and are co-crystallized. In the MALDI source, the crystalline mixture is converted into gas phase ions when the target is irradiated with a laser beam (often a pulsed nitrogen laser at 337 nm) and singly charged peptide ions are accelerated down a flight tube to a detector. All ions receive the same accelerating energy, so lighter ions travel faster. Then a timing mechanism can be calibrated to yield ion mass values for the peptide sample. The advantages of MALDI-TOF instruments over protein sequencing are that they are relatively simple to use, they are compatible with high-throughput automation, with high sensitivity (low femtomole or attomole) and some tolerance to the common components of biological buffers (Pandey and Mann 2000). Apart from providing peptide masses of intact proteolytic peptide ions, some high-end MALDI-TOF instruments can also analyze fragmentation of peptide ions by post-source decay (PSD) and others can fragment the peptide ions in a collision cell to generate MS / MS data (TOF / TOF mechanism). Both approaches yield information on peptide sequence and on post-translational and chemically induced modifications (Chaurand *et al.* 1999; Medzihradszky *et al.*; 2000 Liebler 2002).

MALDI-TOF instruments are extensively used in peptide mass fingerprinting (PMF) analysis of proteins (Henzel *et al.* 1993). In the PMF approach, the protein spots that are separated by 2-DE are excised from the gel and are digested with a site-specific protease. Then the mass-to-charge ratios of proteolytic peptide mixture that are eluted from the gel matrix are measured by MALDI-TOF mass spectrometry to generate peptide mass maps of the protein being analyzed. Subsequently, the experimentally observed peptide mass map of the unknown protein is compared with the theoretical peptide mass maps that are generated by virtual digestion of protein sequences in a database with the same site-specific protease to look for possible matches (Yates 1998). The protein is identified as the one with most theoretical peptide masses that match to the experimental peptide masses of the unknown protein. As the value of a single peptide mass does not necessarily correspond to a specific amino acid sequence, matching of multiple peptides (usually more than 3) that are derived from the same protein must be used to confidently identify a protein. However, the PMF approach also has its own limitation. Application of this approach is limited to the species, which have a complete or near complete genomic sequences available, as the successful protein identification by PMF analysis strongly depends on the availability of protein or gene sequence in the database (Gevaert and Vandekerckhove 2000).

b. Tandem mass spectrometry (MS/MS) analysis

If a protein cannot be positively identified by PMF, then the proteolytic peptide mixtures can be analysed by MS/MS to generate peptide sequence tags. In tandem mass spectrometry analysis, the ions are produced in the source by electrospray ionisation. Electrospray ionisation refers to the process in which a solution containing the peptides is sprayed across a high potential difference in the source. Then the multiply charged peptide ions are separated from solvents and transferred into mass analysers (e.g. QTOF, Ion traps, linear ion traps, FT-ICR, triple quad (for review of instrumentation see Liebler 2002).

Figure 1.5 B shows the schematic diagram of a typical triple quadrupole MS/MS instrument set up, which uses two mass analysers to carry out the MS/MS experiments. From a mixture of peptide ions generated at the ESI source, the first mass analyser (MS-1) selects a single peptide. The selected peptide ion is then further fragmented in a collision cell between the two mass analysers, and the *m/z* for each charged fragment is

measured in the second mass analyser (MS-2). From the mass differences between the series of successive y- ion or b-ion spectra, the whole or partial amino acid sequence of the peptide can be interpreted (Yates 1998).

There are two approaches to identify proteins using a peptide MS/MS spectrum data. In the first approach, the "uninterpreted" MS/MS spectra and the m/z value of the precursor peptide ion are directly compared with the theoretical peptide masses and MS/MS spectra generated by virtual digestion and fragmentation of protein or translated DNA sequences in the databases using computer algorithms like SEQUEST (Yates *et al.* 1995). In the second approach, a "peptide sequence tag" which is assembled from the precursor peptide ion mass, the interpreted partial amino acid sequences and distances in mass to *N*-and C-terminus of the peptide, is used to search a database for possible matches (Mann and Wilm 1994; Jensen *et al.* 1999). An alternative method in the second approach is to obtain the amino acid sequence of a peptide from MS/MS data followed by BLAST searching of protein databases for sequence similarity using the peptide sequences (Shevchenko *et al.* 2001).

The capability of the electrospray ionisation technique to interface with liquid-chromatographic separation techniques makes it possible to separate and assign structural identities to individual proteins in a complex biological sample. At the same time, greatly increased sensitivity of MS/MS analysis (single femtomole) is achieved with the development of a nanoelectrospray ionisation sources (Wilm and Mann 1996). The combination of automated microcolumn-LC-MS/MS with automated database search algorithms provided a rapid and high throughput protein identification and sequence analysis tool for large-scale proteomic analysis (McCormack *et al.* 1997).

1.3.4.3 Data interpretation and bio-informatics in proteomics

Proteomics and genomics are information sciences. Current success in proteomics would never have been achieved without concurrent progress in bio-informatics. In comparative proteomics analysis, spot detection, gel comparison and relative quantification of protein spots on 2-DE gel requires specialized image analysis tools. For this purpose, there are many commercially available software packages as shown in Table 1.1. Inter-library gel matching over the internet has also been accelerated by the online Flicker program (Lemkin 1997).

Protein identification from MS experimental data requires computer algorithms that correlate MS and MS/MS spectra with protein or genomic sequence databases. Manufacturers of mass spectrometers have already included or integrated such computer algorithms as part of their software platforms. Some of the free web based tools for protein identification from MS experimental data are listed in Table 1.1, whilst more complete listing can be found in a recent review (Vihinen 2001). Another major development in bio-informatics is the establishment of a number of 2-D-PAGE proteome databases that containing information on identified proteins by a number of large-scale proteome projects. So far, there are over 100 2-D-PAGE databases established for a number of model organisms and they are indexed on the ExPASY molecular biology server (http://expasy.org/ch2d/2d-index.html).

1.3.5 Limitations in 2-DE-based proteomics

A major challenge remaining in current proteomics technology relates to both the inherent dynamic nature of the proteome and the limitations on currently available

Table 1.1 Software tools used most often in proteomics

Application	Program	Website or company name
2-DE image analysis		
	Melanie 4	GeneBio, Geneva, Switzerland
	PDQuest	Bio_Rad, Hercules, CA, USA
	Imagemaster ®	Amersham Biosiences, Upsala
	Progenesis	Nonlinear Dynamics Limited, UK
	Z3	Compugen, Tel-Aviv, israel
Peptide mass fingerprinting		
	Mascot	http://www.matrixscience.com
	MS-Digest	http://prospector.ucsf.edu/ucsfhtml4.0/msdigest.htm
	MS-Fit	http://prospector.ucsf.edu/ucsfhtml4.0/msfit.htm
	ProFound	http://prowl.rockefeller.edu/cgi-bin/ProFound
	PepIdent	http://www.expasy.ch/tools/peptident/
	MOWSE	http://www.hgmp.mrc.ac.uk/Bioinformatics/Webapp/mowse/
	ProteinLynx	Micromass, Manchester, UK
	PepSea	http://195.41.108.38/PA_PepSeaForm.html
MS/MS analysis	Mascot	http://www.matrixscience.com
	MS-Seq	http://prospector.ucsf.edu/ucsfhtml3.4msseq.htm
	MS-Tag	http://prospector.ucsf.edu/ucsfhtml3.4mstagfd.htm
	MultIdent	http://www.expasy.ch/tools/multident/
	PepFrag	http://prowl.rockefeller.edu/prowl/
	TurboSEQUEST	Thermo Finnigan, San Jose, CA, USA
	PepSea	http://195.41.108.38/PA_PeptidePatternForm.html

separation and detection techniques. Unlike genomics, which deals with one static genome, proteomics has to deal with varying ratios and levels of protein turnover and changing patterns of expression resulting from changing developmental status and responses to external signals (Tyers and Mann 2003). Internal biological processes such as cell differentiation, division and activation, and external stimuli such as infection and abiotic stresses can all significantly change the proteome of a species at any given time (Haynes *et al.* 1998). This dynamic nature of a proteome makes it very difficult to solubilize, separate, detect and quantify the "complete set of proteins that is expressed by a genome", a task that is necessary for the comprehensive analysis of a biological system.

With its unsurpassing ability to resolve thousands of proteins simultaneously on a single gel in a differential display format, the 2-DE currently serves as a principal separation technique in most proteomics projects. However, the technique is not free of limitations such as, for example, the relative incompatibility of some proteins with extreme pI with the first dimension IEF separation. Although the separation of very basic or very acidic proteins and hydrophobic plasma membrane proteins on 2-DE gels has been made possible with the introduction of new detergents and low and high pH range IPG strips, the hydrophobic properties and charges still have a strong impact on resolution of these extreme proteins on 2-DE gel (Huber 2003; Santoni *et al.* 2000). In addition, separation and visualization of proteins larger than 100 kDa and smaller than 8 kDa in standard second dimension SDS-PAGE is also difficult (Harry *et al.* 2000).

The detection of low abundance proteins is the greatest challenge for 2-DE. Several studies have shown that the majority of proteins identified in 2-DE are the more abundant and more long-lived proteins in cells (Haynes and Yates 2000). Proteomic studies on simple unicellular organism yeast show that despite high sample load and extended electrophoretic separation, proteins from genes with codon bias index values of < 0.1 could not be detected on 2-DE gels, even though one-half of all yeast genes have codon bias values of < 0.1 (Gygi *et al.* 2000). Codon bias value of a gene is a measure of the probability that only one of several possible codons will be used to translate that gene (Kurland 1991). A codon bias value is a measure of protein abundance because highly expressed proteins generally have large codon bias values (Garrels *et al.* 1997). With a wide dynamic protein expression ranges of more than 6 orders of magnitude in a cell (Anderson and Anderson 1998), small populations of the

most abundant "house keeping" proteins can swamp these gene products with low codon bias values on standard 2-DE gels, unless these proteins are enriched by pre-sample fractionation or affinity purification (Huber 2003). Unfortunately, there is no protein equivalent of PCR and gene cloning to enrich proteins. Apart from that, due to the limitation in detection sensitivity and the linear dynamic range of currently used staining methods such as silver staining, quantification and differential display analysis of very high or low abundant proteins is still problematic (Ong and Pandey 2001).

Another shortcoming of 2-DE is that this technique is not amenable to high-throughput automation. This prevents current proteomics from having the same degree of high-throughput automation as is available with DNA sequencing (Ong and Pandey 2001). There is a clear need more technical improvements and optimization of currently available techniques before proteomics can reach its goal of studying all the proteins encoded by a genome.

1.3.6 New emerging technologies in proteomics

In less than ten years since the term "proteome" was coined, new developments in protein staining, labeling and mass spectrometric instrumentation have been made and both the speed (automation) and sensitivity of proteomics analysis have been significantly improved. Some recent technological developments in proteomics analysis are outlined in the following sections.

1.3.6.1 Two-dimensional difference gel electrophoresis (DIGE)

Differential display proteomics is based on the comparison of protein profiles of samples from different biological conditions. Experimental variability between 2-DE gels is such that spot quantification and comparison needs to be made using multiple (~6) 2-DE experiments (Wasinger *et al.* 2000). If comparisons of protein profiles of two sample states can be achieved on a single gel,the number gels can be greatly reduced. The 2-D DIGE technology achieves this goal by using spectrally resolvable fluorescent dyes (Unlu *et al.* 1997). The protocol outline for 2-D DIGE (Two-dimensional difference gel electrophoresis) is illustrated in Figure 1.6 A. Proteins extracted from as many as three samples are separately labeled with different amine reactive fluorescent dyes such as Cy2, Cy3 and Cy 5. Then the samples can be combined and run together

on a single 2-DE gel. The 2-DE image for each sample is acquired by scanning the same gel with different wavelength filters. The images are then compared by a computer-assisted image superimposing method to detect differentially expressed proteins (Wasinger and Corthals 2002). 2-D DIGE technology allows the separation of two or three samples under identical electrophoretic conditions, simplifying the process of registering and matching of gel spots.

1.3.6.2 Isotope-coded affinity tags (ICAT)

Relative quantification of proteins is required in comparative proteomics, which investigates changes in levels of protein expressions in cells in response to different treatments. The currently used 2-DE- based protein quantification method is sequential, labor intensive and difficult to automate. Gygi *et al.* introduced a novel method for quantitative and qualitative protein profiling based on isotope-coded affinity tag (ICAT) reagents (Gygi *et al.* 1999a). The ICAT reagent consists of a biotin affinity tag, a cysteine (thiol)-specific alkylating group and a linker group. In this method (Figure 1.6 B), post-isolation protein samples from two cell states are differentially labeled with two forms of ICAT reagents, a light form (d_0) and a heavy form (d_8) that differ by eight mass units. The labeled protein samples are then combined and subjected to proteolysis. Subsequently, the ICAT-labeled (cysteine-containing) peptides in the digest mixture are isolated by avidin-affinity chromatography and subjected to LC- ESI-MS/MS analysis. Tandem mass spectrometry analysis provides relative MS quantification of labeled peptide pairs based on signal intensities of the heavy and light isotopically labeled peptide masses, whilst subsequent MS/MS analysis and database searching provides the identities of proteins from which the peptides originated (Gygi and Aebersold 2000) (Ashcroft 2003). N-terminal isotopic labeling of proteins by d_0/d_4 nicotinylation, which does not rely on the presence of a specific amino acid, has also been described (Munchbach *et al.* 2000). A similar technique for biological labeling of proteins in cell culture, called stable isotope labeling with amino acids (SILAC), has been developed and applied for the quantification of protein expression levels (Ong *et al.* 2002)

1.3.6.3 Multidimensional protein identification technology (MudPIT)

A new trend in proteomic analysis is the utilization of microcolumn liquid chromatography (μLC) for the separation of peptide mixtures. Recently, a gel independent "shotgun" proteomic approach, named multi-dimensional protein identification technique (MudPIT), has been developed to directly interface protein and peptide separation to mass spectrometers (Link *et al.* 1999). In this technique, a reduced and denatured protein mixture is digested with a site-specific protease and the resulting complex peptide mixture is loaded onto a biphasic micro LC column containing strong cation exchange (SCX) resin upstream of reversed phase (RP) resin. The peptides retained on the SCX column are step-wisely eluted based on their charge by a step gradient of increasing salt concentration. Each elution step releases a group of peptides, which then pass onto a down-stream RP column where each peptide in that group is further separated based on their hydrophobicity. From the RP column the peptides are directly eluted into the tandem mass spectrometry instrument for MS/MS analysis (Wolters *et al.* 2001). MudPIT has a dynamic range of 10000:1, which allows the identification of low abundance proteins. In addition, all experimental procedures can be run in a fully automated system.

1.3.7 Utility of proteome analysis in biological research

The primary application of proteomic analysis in biological research is to identify and catalogue large numbers of proteins present in a complex biological system and to present them in a proteome database. This kind of application is known as descriptive proteomics or proteome profiling and is a major driving force behind other proteomic analysis (Yanagida 2002). Protein profiling projects result in the production of a simple list of identified proteins detected in particular a cell state. However, this kind of information is insufficient to solve biological questions. To be biologically useful, proteome-profiling experiments should interface with other possible applications of MS-based proteomics tools (described below) (Aebersold and Mann 2003).

Another important application of proteomics research is its use as a global discovery tool to detect dynamic changes in the proteome of a cell or tissue in response to external or internal stimuli. This is often referred to as "differential display proteomics" or "comparative proteomics". Because the detection of dynamic changes requires relative quantification of each protein component, it is also called as "quantitative proteomics"

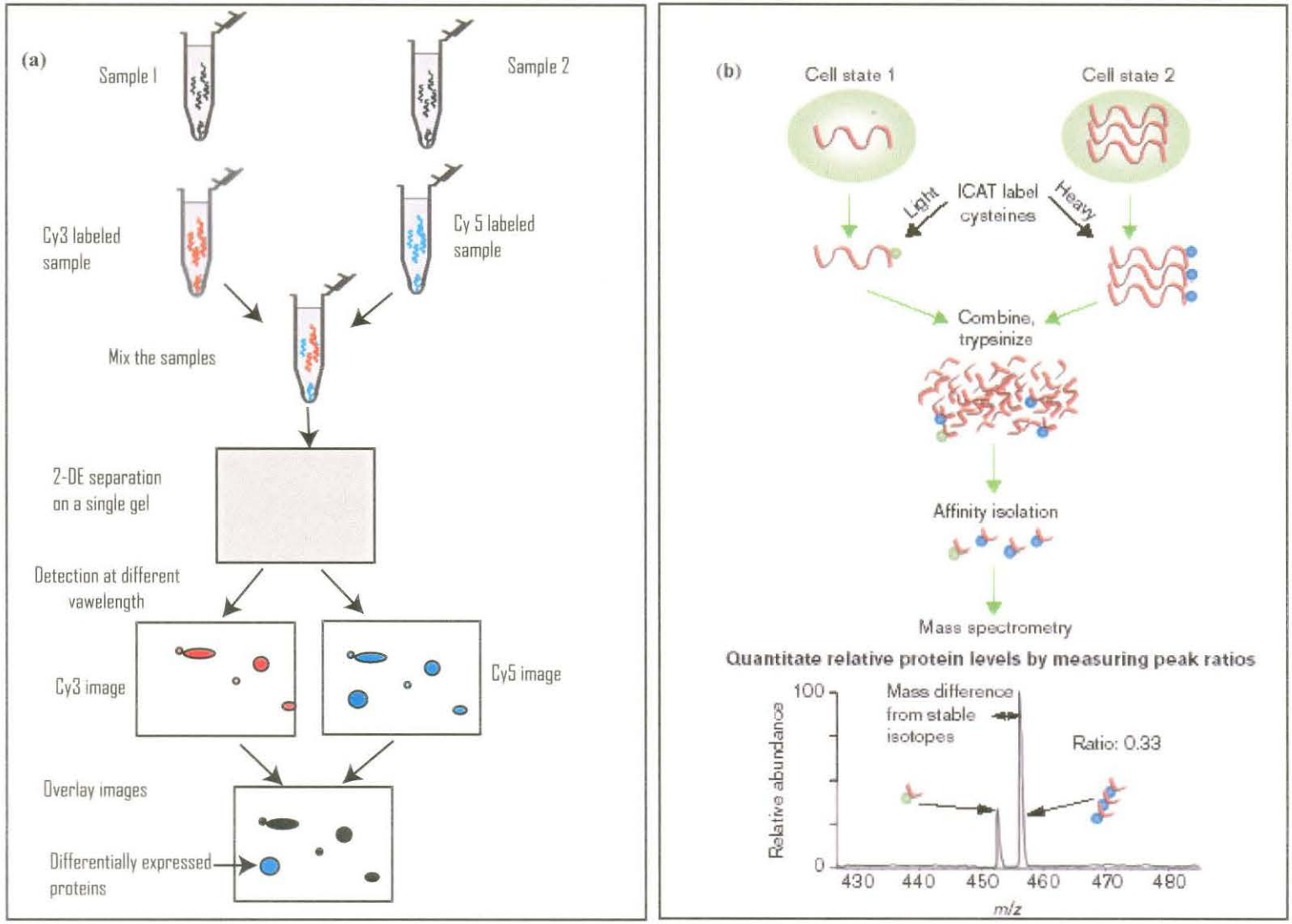


Figure 1.6. Schematic representations of (a) two-dimentional difference gel electrophoresis (DIGE) technology and (b) isotope coded affinity tags (ICAT) technology. (b) was adapted from Gygi et al. (2000).

(Gygi and Aebersold 2000). By comparing the proteome profiles of healthy and diseased tissues, marker proteins associated with specific diseases can be identified for use in clinical diagnosis. Pharmacology studies frequently employ comparative proteomic analysis to identify drug targets and to understand the mechanism of action of drug candidates (Pandey and Mann 2000).

One of the biggest challenges for proteomics and one of its greatest potential applications is studying post-translational modifications (PTMs) of proteins. After translation, a large part of cellular proteins are covalently modified by PTMs leading to the attainment of their proper functions. There are more than 300 post-translational modifications reported (see the RESID database: http://pir.georgetown.edu/pirwww/dbinfo/resid.html). It is by co- and post-translational modifications that a single gene locus gives rise to multiple proteins. The post-translational modifications of proteins are essential for biological functions of proteins in a cell, because the PTMs can determine the activity state, localization, turnover of proteins and their interactions with other proteins. However, post-translational modifications of proteins cannot be predicted from genomic sequence analysis. Recent developments in mass spectrometry-based proteomics technologically have made it feasible to systematically map certain types of PTM(s) on all the proteins present in a sample (Aebersold and Mann 2003). General proteomic strategies for mapping PTMs in a purified single protein or on a proteome-wide scale have recently been reviewed (Mann *et al.* 2002; Mann and Jensen 2003).

Another major aspect of proteomic analysis is its best suitability to study protein-protein interactions in biological systems. Most cellular functions in cells are not performed by individual proteins but by a number of interacting proteins in multi-protein complexes. Therefore, functions of individual proteins should be best described in the context of their interaction with other proteins or biological molecules. Based on the assumption that the proteins in a multi-protein complex are involved in the same biological pathways, the biological function of unknown proteins (corresponding genes) can be assigned according to proteins of known functions in the same protein complex. Detection and identification of individual members of a multi-protein assembly will give insight into the function and regulation of some biological pathways. However, this kind of information cannot be deduced from DNA sequence analysis alone (Huber 2003). Recently, a new high-throughput strategy for the analysis of multiprotein

complexes in their natural level under native conditions using tandem affinity purification (TAP) has been developed (Rigaut *et al.* 1999). The above-mentioned different fields of current proteomics can be applied in parallel to the same biological system to generate more comprehensive and integrated knowledge about the structure and dynamics of a proteome.

1.4 CURRENT PROTEOMIC STUDIES IN RICE RESEARCH

Proteomic tools have been employed actively in plant sciences. Although the concept of proteomics is quite new, the main underlying technology of proteomics, the 2-DE technique, had already been employed in many different aspects of plant research for many years for various applications including differential display, mutant analysis, plant development, variety identification, plant microbe interaction, allergen identification, and these studies that were undertaken before 1988 have been reviewed in detail (Damerval *et al.* 1988). More recently, several comprehensive reviews have extensively covered the achievements of current mass spectrometry based proteomics in different fields of plant research (Thiellement *et al.* 1999; Zivy and de Vienne 2000; Jacobs *et al.* 2000; Dubey and Grover 2001; Rossignol 2001; van Wijk 2001; Kersten *et al.* 2002; Roberts 2002).

One of the successful applications of proteomics in rice is its use for cataloguing tissue and organ- specific proteomes of rice plants. As early as 1987 Komatsu *et al.* initiated a study on the construction of a 2-DE protein database of seed embryo, endosperm, and leaves of three rice cultivars (Komatsu *et al.* 1993). In this study, more than 600 embryo, 100 endosperm, and 150 leaf protein spots were detected on 2-DE gels by CBB staining. Sequence information was obtained for 61 protein spots out of 85 proteins subjected to Edman sequencing. A data-file of rice proteins containing MW, p*I*, amino acid sequences, identity and glycosylation was constructed. Later, the same research group added to this database the rice root protein data-file (Zhong *et al.* 1997). Almost in parallel with Komatsu *et al.*, Tsugita and colleagues reported 2-DE analysis of rice proteins from nine different tissues (Tsugita *et al.* 1994). A total of 4892 protein spots were resolved, of which 2.8% were analyzed by N-terminal sequencing. They also reported a 2-DE comparative study of proteins isolated from different tissues of rice and *Arabidopsis thaliana* (Tsugita *et al.* 1996). Large-scale protein identifications in these studies were impeded because of the unavailability of high-throughput MS based

protein analysis tools at that time. Also, because the complete rice genome sequence was unavailable at that time, the success rate for protein identification was not high.

More recently, our laboratory reported a comprehensive proteomic study of rice anther proteins expressed at the young microspore stage (Imin *et al.* 2001). In combination with 2-DE separation and MS analysis, over 4000 anther proteins that represent 10% of the total genomic output of rice were resolved in silver stained gels. The protein identification results were presented in a 2-DE protein database, which can be accessed at http://semele.anu.edu.au/2d/2d.html. In a similar way, rice leaf sheath proteins have also been profiled (Shen *et al.* 2002). Another report representing the most comprehensive proteome exploration to date in rice was presented recently, in which both 2-DE and MudPIT technologies were used to identify a total of 2528 unique proteins (Koller *et al.* 2002). Most recently, a proteomic study providing an extensive picture of mitochondrial functions for monocot plants has been reported on rice (Heazlewood *et al.* 2003). In addition to 149 proteins identified by 2-DE and LC-MS/MS analysis in this study, several multi protein complexes in the mitochondrial membrane fraction and their constituents were analyzed by blue native (BN) SDS-PAGE. In a more recent large-scale rice proteomics study (Andon *et al.* 2003), carbohydrate-binding proteins were fractionated from rice leaf, root and seed tissue extracts by column affinity chromatography using α-D-mannose as the ligand. The protein fractions were then separated by SDS-PAGE followed by LC-MS / MS analysis of the separated proteins. This multiplexed technology resulted in the isolation and identification of 136 distinct mannose binding proteins, nearly 15% of which were matched to gene products with no previously known functions. This research indicated the potential of this multiplexed proteomic approach in assigning preliminary functions to novel genes in a high-throughput fashion.

The differential display property of proteomic analysis has been a useful means to investigate the protein profile responses of different rice tissues to the effects of global signaling molecules (Moons *et al.* 1997; Rakwal *et al.* 1999; Rakwal and Komatsu 2000; Shen *et al.* 2003), and chemical and environmental stresses (Agrawal *et al.* 2002; Hajduch *et al.* 2001; Rakwal *et al.* 2003; Salekdeh *et al.* 2002a; Salekdeh *et al.* 2002b). Some metabolic pathways that were disturbed under different stress conditions were identified. In several instances, this approach has led to the isolation of novel stress-

responsive cDNAs, providing clues to the function of uncharacterized genes (Salekdeh, Siopongco *et al.* 2002a). From the examples mentioned above, it can be expected that the application of high-throughput proteomics technology in the post- genomic era of rice research will lead to a dramatic increase in knowledge of growth, development and stress responses in rice plants. Once extrapolated to other plants, this kind of knowledge will permit major breakthroughs in plant biotechnology and will lead to further improvement in crop breeding.

1.5 SCOPE OF THIS THESIS: APPLICATION OF PROTEOMICS APPROACH TO PLANT DEVELOPMENTAL BIOLOGY

The basic hypothesis of this thesis is whether the molecular mechanism underlying the male gametophyte developmental process of cereal crops could be efficiently studied in model crop plant rice using modern post genomic investigation techniques.

The main aims of this thesis then are:

- To evaluate the potential application of proteomics tools for plant developmental biology
- To provide an insight into the network of metabolic pathways involved in the sequential cellular events that take place during the course of pollen development in monocot plants by examining the global protein profiles of rice anthers at different developmental stages
- To exploit the wealth of rice genomic and EST sequence information that are currently available in the public domains in order to complement the rice functional genomics study with proteomic data at the protein level
- To assist the ongoing research projects aimed at understanding the molecular mechanism of abnormal development of rice anthers by providing information about proteins involved in the normal development of rice anther
- To identify and characterize a range of micropore stage-specific protein markers, which have potential applications as indicators of essential pollen developmental stages in breeding programs in rice as well as in other cereal crops.

CHAPTER 2 GENERAL MATERIALS AND METHODS

2.1 PLANT GROWTH AND SAMPLING

2.1.1 Rice seeds

Seeds of the Australian rice cultivar Doongara (*O. sativa* L. cvs Doongara) were obtained from Yanco Agricultural Institute, New South Wales, Australia. Seeds were stored in airtight containers at 4°C until use.

2.1.2 Plant growth

Soil used for plant growth was from the experimental paddy field at the Yanco Agricultural Institute, New South Wales. The soil was sterilised by steaming before being used for plant growth. The rice plants were grown in controlled environment facilities at the Research School of Biological Sciences. The rice seeds (10 seeds per pot) were sown in individual pots (height 15 cm and diameter 8.3 cm) filled with the sterilised soil and the pots were placed in the plastic tubs that were filled with water to cover ~10 cm of the bottom parts of the pots for germination. The plants were grown until the panicle initiation stage in a glasshouse under the following growth conditions: 30°C / 20°C (day/night), 70% relative humidity and natural day length. Once the plants were well established they were reduced to one plant per pot. Four weeks after sowing, urea (15 g/m^2) was applied as nitrogen fertiliser and the plants were flooded. Just before panicle initiation stage the plants were transferred to a growth chamber maintained at a day/night temperature of 30 °C / 20°C (12 / 12h), 70% relative humidity and a photon flux density of 330 $\mu mol.m^{-2}.s^{-1}$.

2.1.3 Anther sampling

Different microspore developmental stages were predicted according to auricle distance and the days before heading (Refer to Chapter 3). The top spikelets of the top three primary branches were dissected with fine forceps and scalpels under a dissection microscope to remove the enclosed anthers. In order to take more synchronous anther samples, only the main tillers of the rice plants were used for the experiments. Anthers within the same anther length range were collected into a eppendorf tube and immediately frozen in liquid nitrogen and stored at –80°C until use. Anther samples

belonging to the same anther developmental stage were pooled to obtain a 100 mg of starting sample for protein extraction. For studies on early stage anthers, around 3000 anthers were collected to obtain 100 mg material, whereas for late stage anthers only 700 anthers were needed to obtain 100 mg material.

2.2 PROTEIN EXTRACTION

2.2.1 Solutions

All solutions were prepared using analytical or electrophoresis grade reagents (a list of reagent and their sources is given in section) and made up in high purity Milli-Q® water (Millipore Co, MA, USA) with a resistance of higher than 18 MΩ.

TCA precipitation solution:

TCA (10%, w/v)	5.0 g
DTT (0.07%, w/v)	0.035 g

made up to 50 mL with milli-Q water.

Acetone washing solution:

DTT (0.07%, w/v)	0.035 g

made up to 50 mL with Acetone.

Bromophenol blue stock solution:

Bromophenol blue (1%, w/v)	0.1 g

made up to 10 ml with milli-Q water and filtered through 0.45 μm size membrane filters.

Solubilization buffer:

Urea (9 M)	5.4 g
DTT (1%, w/v)	0.1 g
CHAPS (4%, w/v)	0.4 g
Bio-lyte ampholytes (0.8%, v/v)	250 µL
Tris (35 mM)	0.0424 g

made up to 10 ml with milli-Q water and filtered through 0.45 µm size membrane filters. The solution was prepared fresh every time or small aliquots were stored at – 80°C for up to 2 months. The thawed buffers were not refrozen again.

Sample buffer:

Urea (9 M)	5.4 g
DTT (1%, w/v)	0.1 g
CHAPS (4%,w/v)	0.4 g
Bio-lyte ampholytes (0.8%, v/v)	250 µL
Tris (35 mM)	0.0424 g

200 µL of 1% bromophenol blue added.

made up to 10 ml with milli-Q water and filtered through 0.45 µm size membrane filters. The solution was prepared fresh every time or small aliquots were stored at – 80°C for up to 2 months. The thawed buffers were not refrozen again.

2.2.2 TCA/acetone protein extraction for 2-DE analysis

The anther proteins were extracted using the TCA-acetone precipitation method (Kamo *et al.* 1995; Natera *et al.* 2000). Prior to protein extraction the anther samples belonging to the same developmental stages were pooled in liquid nitrogen to make

about 100 mg starting material. The pooled anther materials were disrupted by vigorous grinding in a liquid-nitrogen cooled mortar and pestle with a small amount of glass powder for at least 10 min until the very fine powder was obtained. The resultant powder was transferred to pre-chilled (in dry ice) centrifuge tubes and suspended in 5 mL of TCA/acetone precipitation solution (pre-chilled to -20°C). While being kept on dry ice, the sample was sonicated 6 x 10 s with 50 s intervals using a MSE probe ultrasonicator (MSE 100, Thomas Optical and Scientific Co Ltd., Australia). A peak-to-peak amplitude setting of 7 microns was used. After a minimum incubation of one hour at -20 °C, the sample was centrifuged at 35000 x g for 15 min at 4°C and the supernatant was discarded. The pellet was resuspended in 5 mL of pre-chilled (-20°C) acetone washing solution and incubated at -20°C for 30 min before being pelleted by centrifugation at 35000 x g. This washing step was repeated two or three more times in order to remove the TCA in the sample. After the last washing step the pellet was lyophilised for 5 min (Flexi-Dry MP, FTS Systems Inc., Stone Ridge, NY, USA).

2.2.3 Protein solubilization

The dried acetone-free pellet was suspended in 250 µL of solubilisation buffer using pipette tips. In order to maximize the protein solubilisation, the resuspended sample was sonicated in a water bath ultrasonicator (Elma® Transonic 460, John Morris Scientific, Willoughby, NSW, Australia) for 6 x 10s with 30 s of vortexing between each sonication. The sample was left for an hour at room temperature. Then, undissolved materials in the sample were removed by centrifugation at 12000 x g for 15 min at room temperature using a desktop centrifuge. The clean supernatant was collected and was stored in 100 µg aliquots at –80°C until use.

2.2.4 Bradford protein assay

The protein concentration of the anther protein extract was determined using a Bradford protein assay kit (Bio-Rad Laboratories, CA, USA) in order to standardize the protein loading for first dimensional isoelectric focusing. A standard microtiter plate assay procedure was used with some modifications to reduce the amount of protein sample used for the assay. The dye concentrate was diluted 1 in 4 in milli-Q water and filtered through 0.45 µm size membrane filters to remove particulates. The protein samples to be assayed were diluted 1 in 2, 1 in 5, and 1 in 10 in solubilization buffer.

Series dilutions (0 to 1 mg/mL) of bovine serum albumin (BSA) were prepared using the same buffer. Then 5 µL of each diluted and undiluted protein samples or BSA standard solutions were mixed with 245 µL of diluted dye reagents in microtiter plate wells. After a 5 min incubation at room temperature, absorbance at 600 nm was measured with a microtitre plate reader (LabSystems Multiskan RC, Helsinki, Finland). A linear regression line, or a standard curve, was generated from optical density readings and concentrations of BSA standards using InStat ™ (GraphPad Software) or Microsoft® Excel 2000 software packages. The protein concentrations of the diluted and undiluted samples were calculated by comparing their optical density readings to the BSA standard curve.

2.3 TWO-DIMENSIONAL GEL ELECTROPHORESIS

2.3.1 Solutions

0.5 M Tris-HCl (pH 6.8) stock solution:

Tris base	15.25 g
10 M HCl	10-12 mL

The Tris base was first dissolved in 75 mL of milli-Q water and the pH of the solution was adjusted to 6.8 with 10 M HCl. The volume was made up to 250 mL and filtered through a 0.45 µm Sterile Acrodisc® syringe filter (Gelman Sciences, MI, USA). The solution was stored at 4°C.

IPG strip rehydration solution:

Urea (8 M)	9.6 g
DTT (0.2%, w/v)	0.04 g
CHAPS (0.5%, w/v)	0.1 g
Bio-lyte ampholytes (0.52%, v/v)	260 µL
1% bromophenol blue	120 µL

made up to 20 mL with milli-Q water and filtered through a 0.45 μm size syringe filter prior to use.

Equilibration solution for 2 x 11 cm strips:

0.5 M Tris-HCl, pH 6.8 (50 mM)	5 mL
Urea (6 M)	18 g
Glycerol (40%, v/v)	15 mL
SDS (2%, w/v)	1 g

made up to 50 mL. The solution was then divided into 2 x 25 mL. To solution A, 0.5 g of dithiothreitol (DTT, 2%) was added. To solution B, 1 g of iodoacetamide (IAA, 4%) and 125 μL of 1% bromophenol blue were added.

Equilibration solution for 2 x 18 cm strips:

0.5 M Tris-HCl, pH 6.8 (50 mM)	8 mL
Urea (6 M)	28.8 g
Glycerol (40%, v/v)	24 mL
SDS (2%, w/v)	1.6 g

made up to 80 mL. The solution was then divided into 2 x 40 mL. To solution A, 0.8 g of dithiothreitol (DTT, 2%) was added. To solution B, 1.6 g of IAA,(4%) and 200 μL of 1% bromophenol blue were added.

2.3.2 Isoelectric Focusing (First dimension)

2.3.2.1 Rehydration of immobilized pH gradient (IPG) strips

Immobiline™ DryStrip Gels (Amersham Biosciences, Uppsala, Sweden) were used throughout the experiments. Linear gradient 11 cm and 18 cm strips covering the pH range of 4-7 and 6-11 were rehydrated overnight using either a Immobiline™ DryStrip reswelling tray or a reswelling cassette (both from Amersham Biosciences, Uppsala, Sweden) according to the manufacturer's instructions. For the rehydration of

strips in a reswelling tray, 220 µL and 360 µL of rehydration solution was used for each 11 cm and 18 cm strips respectively.

2.3.2.2 *First dimension set up*

The first dimensional isoelectric focusing experiment was carried out on a Multiphore II electrophoresis unit (Amersham Biosciences, Uppsala, Sweden). The components of the Immobiline™ DryStrip kit (Amersham Biosciences, Uppsala, Sweden) were thoroughly washed and dried. Following overnight rehydration, the IPG strips were gently blotted on filter paper for a few seconds to remove any excess rehydration solution, which can interfere with isoelectric focusing. Then the strips were immediately aligned into the groves of plastic aligner, which was placed inside the tray. The tray was assembled onto the kerosene-wetted cooling plate of the Multiphore II electrophoresis unit according to the manufacturer's instruction.

2.3.2.3 Sample application

Throughout the experiment the samples were applied to the IPG strips by a cup-loading sample application method. For analytical 2-DE analysis followed by silver staining, 100 µg of protein samples that were diluted in sample buffer to a final loading volume of 50 µL was applied into a plastic cup placed on the anodic end of the 18 cm IPG strips. For micropreparative runs, each strip was loaded with 1000 µg of protein sample in a final sample volume of 200 µL using same sample application method. All samples that had been stored frozen were sonicated in a water bath ultrasonicator for 6 x 10s with 30 s of vortexing between each sonication. Prior to loading, the samples were centrifuged for 5 min in a desktop eppendorf centrifuge at maximum speed to remove any insoluble material which may cause protein precipitation on the gel surface.

2.3.2.4 Running condition for first dimensional isoelectric focusing

Multiphor EPS 3500XL power packs (Amersham Biosciences, Uppsala, Sweden) were used to apply electric current for isoelectric focusing. The isoelectric focusing temperature was maintained at 20°C using a MultiTemp III thermostatic circulator (Amersham Biosciences, Uppsala, Sweden) in order to minimize the variation of isoelectric focusing patterns between experiments. The voltage was increased gradually

Table 2.2 Running conditions for 18 cm, pH 4-7 and pH 6-11 linear IPG strips (both analytical and preparative gels)

Step	Voltage (V)	Current (mA)	Power (W)	Volt hours (Vh)	Time (h)
1	150	1	5	19	0:15
2	150	1	5	75	0:30
3	300	1	5	1	0:01
4	300	1	5	1800	6
5	3500	1	5	9500	5
6	3500	1	5	189000	54
Total				200000	65:45 h

Table 2.1 Running conditions for IPG pH 4-7 linear, 11 cm preparative gels

Step	Voltage (V)	Current (mA)	Power (W)	Volt hours (Vh)	Time (h)
1	150	1	5	19	0:15
2	150	1	5	75	0:30
3	300	1	5	1	0:01
4	300	1	5	1800	6
5	3500	1	5	9500	5
6	3500	1	5	64800	18:30
Total				76000	30:15h

from 150 V to 3500 V to improve the sample entry into the IPG strips. After the completion of the isoelectric focusing, the strips were either subjected to the second dimension SDS-PAGE immediately, or sealed in plastic wrap and stored at -8 °C until required for the second dimension analysis.

2.3.3 Second dimension analysis

2.3.3.1 Equilibration of first dimension IPG strips

The first dimension IPG strips were placed in individual test tubes with the support film towards the wall of the test tubes and equilibrated in Equilibration solution A for 10 min with gentle shaking on an orbital shaker. Then the strips were further equilibrated in the Equilibration solution B in new test tubes with gentle shaking. After

the second equilibration step, the strips were placed on blotting paper with the gel face up and gently blotted with Kimwipes® tissue (Kimberly-Clark Australia Pty. Ltd., Milsons Point, Australia) to remove excess equilibration solution which may cause streaking in the second dimension.

2.3.3.2 Second dimension set up

The second dimension SDS-PAGE was performed on pre-cast ExcelGel SDS gels (T = 12 to 14% acrylamide, 245 x 180 x 0.5 mm, Amersham Biosciences, Uppsala, Sweden) using a Multiphore II electrophoresis unit according to the manufacturer's instructions. The gel was positioned on the kerosene-wetted cooling plate of the Multiphore II electrophoresis unit. Great care was taken to avoid any spillage of kerosene on the gel surface as this may cause the buffer strips to slide on the gel surface during the electrophoresis. Any air bubbles trapped between the cooling plate and the plastic gel backing were squeezed out. Then, pre-cast anode and cathode buffer strips (Amersham Biosciences, Uppsala, Sweden) were positioned onto the appropriate ends of the gel surface. Any air bubbles trapped between the gel and buffer strips were removed by gently stroking the buffer strips with a pair of forceps. The equilibrated and blotted IPG strips were placed gel face down onto the SDS gel at a maximum distance of 1 mm away from the cathode buffer strip. IEF sample application pieces were placed at each end of the IPG strips in close contact with the edge of the IPG strip gels to absorb buffer solution which may cause protein spot streakings on the gel. Finally, the plastic backings of the IPG strips were gently stroked with a pair of forceps to remove any air bubbles caught between the IPG strips and the SDS gel. A 5 μL volume of molecular weight markers containing 100 ng (for analytical gel) and 400 ng (for micropreparative gel) was applied onto extra sample application pieces aligned beside the IPG strips.

2.3.3.3 Second Dimension Running Conditions

The second dimension SDS-PAGE temperature was maintained at 6°C using a MultiTemp III thermostatic circulator. After step 1, the electrophoresis was paused to remove the IPG strips and sample application pieces before proceeding with step 2. After step 2, the cathodic buffer strips were moved to cover the region where the IPG strips originally had been placed and step 3 electrophoresis was continued for 4 hours or

Table 2.3 Second dimension running conditions (per gel)

Step	Voltage (V)	Current (mA)	Power (W)	Time (h)
1	200	10	30	1:25 h
2	200	10	30	0:10 h
3	600	30	60	4:00 h

until the bromophenol blue dye reaches the anodic buffer strips. Then, the SDS gels were subjected to either staining or electroblotting.

2.3.4 Protein visualization

Table 2.4 Silver nitrate staining protocol according to Rabilloid et al. (1994)

Solution	Composition	Time
Fixation	40% (v/v) ethanol 10% (v/v) acetic acid	3 x 30 mins
Sensitisation	30% (v/v) ethanol 2.5 g/L potassium tetrathionate 68 g/L sodium acetate trihydrate 10 mL of 50% gluteraldehyde made up to 1000 mL with Milli-Q water	16 h
Rinse	1L of milli-Q water	6 x 20 mins
Silver	0.624 g/L HEPES 2 g/L silver nitrate 700 µL/L formaldehyde (37%) Formaldehyde was added just prior to use	120 mins
Rinse	1 L/gel water	15 seconds
Development	30 g/L potassium carbonate 10 mg/L sodium thiosulfate pentahydrate 250 µL/L formaldehyde (37%) Formaldehyde was add just prior to use	4 - 12 mins
Stop	50 g/L Tris base 20 mL/L acetic acid	10 mins
Storage	Several changes of water. The gel was sealed in a plastic bag with some milli-Q water for long term storage	

2.3.4.1 Silver staining

After separation by 2-DE, the proteins on analytical gels loaded with 100 µg of total protein sample were visualized using the silver nitrate-tetrathionate staining method (Rabilloud *et al.* 1994). The detailed silver staining protocol is shown in Table 2.4. Each gel was stained in a clean glass or plastic trays with a minimum solution of 500 mL/gel. The entire staining procedure was performed inside a fume hood. All the solutions were prepared fresh, 30-45 min before use. Fresh formaldehyde stock, which is less than three months old, was used throughout the experiment.

2.3.4.2 Colloidal Coomassie staining

The micropreparative gels were stained with colloidal coomassie, which is fully compatible with subsequent mass spectrometry analysis of the protein spots excised from the gels. The gels were stained in sealed plastic bags containing a minimum of 200 mL of staining solution for 24 hours, with gentle shaking on an orbital shaker. The gels were destained in milli-Q water and stored in sealed plastic bags at 4°C.

Colloidal coomassie solution:

Ammonium sulphate (17%, w/v)	170 g
85% Phosphoric acid (3%, w/v)	36 mL
Coomassie G-250 (0.1%, w/v)	1 g
Methanol (34%, v/v)	340 mL

To make the solution, the ammonium sulphate, methanol and phosphoric acid were mixed completely with magnetic stirring and heating. Then, the Coomassie G-250 was added, made up to 1 L with milli-Q water and the dye was dissolved completely by magnetic stirring. The used dye was discarded after each use.

2.3.5 Image analysis of 2-DE gels

After being stained by silver, the analytical gels were scanned at 600 dpi (dots per inch) resolution using the transparent scanning mode of a UMAX Astra-2400S scanner (UMAX Technologies, Fremont, CA, USA) run under Photoshop 5.5 (Adobe,

Mountain View, CA). For coomassie-stained preparative gels, the gels were scanned at the same resolution using reflective mode, while the back of the gels were covered with a piece of white paper to obtain an optimal reflective image. The resultant gray scale images was saved as TIFF (tagged image file format) files.

Image analysis, including spot detection, quantification and gel matching was performed using MELANIE 3.0 software package (Swiss institute of bioinformatics, Geneva, Switzerland). For the purpose of comparison, all gels were adjusted to the same colour levels and the automatic spot detection was carried out using the same spot detection parameters that were optimised for all gels. Some protein spots that were common to the compared gels were marked as 'landmark' proteins to assist the alignment of gel images. Gel comparison was carried out both by superimposing the gels on a light box using the naked eye and by comparing the relative spot volume (% volume) of protein spots, which were measured and calculated by the MELANIE 3.0 program.

The experimental molecular mass of the proteins was determined from the molecular weight of protein standards co-migrated on the same gel. The experimental p*I*s of spotswas determined from measuring the distance between pH units along the linear IPG strips.

2.4 ELECTRO-BLOTTING AND *N*-TERMINAL SEQUENCING

2.4.1 Solutions

100 mM 3-[Cyclohexylamino]-1-propanesulfonic acid (CAPS) stock:

22.13 g of 3-(Cyclohexylamino)-1-propanesulfonic acid was made up to 1 L in milli-Q water and stored at 4°C after being filtered through a 0.45 μm pore size filter.

10 mM CAPS transfer buffer:

100 mM CAPS stock (10 mM)	100 mL
Methanol (10%, v/v)	100 mL
SDS (0.05%, w/v)	0.5 g

made up to 950 mL with milli-Q water and the pH was adjusted to 11 with 5 M NaOH. The volume was then made up to 1 L and chilled at 4°C. The buffer solution was made fresh every time one night before use.

PVDF membrane staining solution:

Coomassie brilliant blue, R-250 (0.1%, w/v)	0.5 g
Methanol (40%, v/v)	200 ml

made up to 500 mL with milli-Q water and filtered through 0.45 μm size filter.

PVDF membrane destaining solution:

Methanol (50%, v/v)	500 mL
Acetic acid (5%, v/v)	50 mL

made up to 1 L with milli-Q water and filtered through 0.45 μm size filter.

2.4.2 Preparation of PVDF membrane

PVDF membrane (TransBlot® Transfer Medium, 0.2 μm, BioRad, Richmond, CA, USA) was cut to the same dimension as the separation gel (160 mm x 240 mm for an ExcelGel SDS 12-14) using a clean blade. The membrane was soaked in 100% methanol for 15 s and rinsed in Milli-Q water for 2-3 min. The membrane was equilibrated in CAPS transfer buffer for at least 5 min prior to use. The membrane was handled at the edges with flat-tipped forceps or with gloved hands. Twelve pieces of filter paper (Munktell Filter AB, Grycksbo, Sweden) were also cut to the dimension as the PVDF membrane.

2.4.3 Semi-dry electroblotting

Semi dry protein transfer was performed using a Nova-Blot electrophoretic transfer unit (Amersham Biosciences, Uppsala, Sweden). Following the micropreparative 2-DE analysis, the top 4 cm part of the gel (the stacking gel) was sliced off using a clean surgical blade. The remainder part of the gel was rinsed for 5 sec in milli-Q water and equilibrated in chilled CAPS transfer buffer for 5 min. Then the gel was placed onto a film remover (Amersham Biosciences, Uppsala, Sweden) and the cutting wire was

slowly pulled along the gel length in order to separate the gel from the support film. Any folds and wrinkles which formed were removed by wetting the gel surface with CAPS transfer buffer and gently rolling a clean glass pipette along the gel. The PVDF membrane was then aligned and lowered slowly onto the gel surface. To prevent smearing, care was taken not to move the membrane once in contact with the gel.

Then the six filter papers that had been soaked in chilled CAPS transfer were placed onto the wetted graphite anode plate of Nova-Blot electrophoretic transfer unit. Any air bubbles trapped between the filter papers were pressed out using a roller. Then the gel sandwich (membrane, gel and support film), with membrane face down, was lowered onto the filter paper. The support film was removed from the assembly by slowly peeling from one side. The air bubbles trapped between the gel and the PVDF membrane were removed by slowly rolling a glass pipette along the gel while keeping the gel surface wet with CAPS transfer buffer. The assembly of the transfer sandwich was completed by placing another six pieces of filter paper which were pre-soaked in chilled CAPS buffer on the top of the gel. Then, the graphite cathode plate wetted with the transfer buffer was placed on the top of the transfer sandwich and the whole assembly was placed into the Multiphor II electrophoresis unit.

Electrophoretic transfer of proteins from the gel to the PVDF membrane was carried out at a constant current of 0.8 $mA.cm^{-2}$ and maximum voltage of 35 V for 90 min at room temperature. Upon completion of protein transfer, the membranes were either stained with coomassie blue stain (see below) or used for Western immuno-detection.

2.4.4 Visualization of proteins on PVDF membranes

After transfer, the PVDF membrane was rinsed in milli-Q water for 1-2 min to remove any contaminants before staining with Coomassie Brilliant Blue R-250. The membrane was stained in PVDF membrane staining solution for 5 mins and destained in several changes of PVDF membrane destaining solution over a period of 10 min or until the background became light blue. The membrane was then rinsed in several changes of milli-Q water over 30 min and air-dried overnight between blotting paper. The dried blot was scanned at 300 dpi using the reflective mode for future reference. The membrane was placed between blotting papers and stored in sealed plastic bags at -20°C until use.

2.4.5 N-terminal sequencing and database searching

Selected spots were excised from the dried membrane using separate sterile scalpel blades. In order to reduce the cross-contamination from neighbouring protein spots, the target protein spots were excised within the spot of interest or as close to the spot as possible. Each PVDF piece was put into an individual Eppendorf tube and stored at -20°C before being sent for *N*-terminal sequence analysis. The *N*-terminal micro sequencing was done on a PROCISE 494-01 sequencer system (Perkin-Elmer Applied Biosystems, Foster City, CA) at the ANU Biomolecular Resource Facility. Depending on the amount of protein sample in the membrane, up to 15 cycles of Edman degradation reaction were performed on each sample.

The identification of protein spots by *N*-terminal amino acid sequence homology was done in two steps. First, the *N*-terminal amino acid sequences homology search was done with the online FASTA program (http://www2.ebi.ac.uk/fasta3/) (Pearson and Lipman 1988) and BLASTP program (http://www.ncbi.nlm.nih.gov/BLAST/) (Altschul *et al.* 1997) in order to screen non-redundant SWISS-PROT, TREMBL and NCBInr protein databases to find possible matches. If no significant matches were found, then a tBLAST search was conducted using the theoretical six-frame translation of rice EST sequences stored at the TIGR Rice Gene Indices database. The translated amino acid sequences of the matching candidate ESTs were analysed for possible signal peptide cleavage sites using the on-line signal peptide prediction program SignalP (Nielsen *et al.* 1997), available at http://www.cbs.dtu.dk/services/SignalP/. The theoretical molecular weight and pIs of the matching sequences, excluding the signal peptide, were calculated using the on-line Compute pI/Mw tool available at the ExPASy Molecular Biology Server (http://expasy.org/). The agreement between the theoretical isoelectric points and molecular masses of the matched sequences and those experimentally observed was taken into consideration during *N*-terminal homology searches.

2.5 PEPTIDE MASS FINGERPRINTING ANALYSIS

2.5.1 Preparation of samples for MALDI-TOF

The target protein spots were excised from micropreparative gels stained with colloidal coomassie (section 2.3.4.2) using a separate sterile surgical blade for each protein spot. Each excised gel piece containing the protein of interest was placed in one

of the 96 wells of a polypropylene microtitre plate (Medos, Langenselbold, Germany). Each well contained 10 μL of 10% (v/v) methanol which asisted in getting the gel piece into the well without static repulsion. The plates were sealed with adhesive plate seals (Advanced Biotechnologies Ltd., Surrey, UK) and sent to the Australian Proteome Analysis Facility (APAF, Macquarie University, Sydney, Australia) for MALDI-TOF MS analysis.

2.5.2 Tryptic digestion and MALDI-TOF MS analysis

The excised gel pieces containing the proteins were destained with 100 mM ammonium bicarbonate, pH 7.8, and spun-dry with a speed vac (Thermo Savant, Holbrook, NY, USA), and then treated for 16 hours of tryptic digestion with 8 μL of 15 ng/μL sequencing-grade modified trypsin in 25 mM NH_4CO_3, pH 7.8 (Promega, Madison, USA) at 37°C. The resulting peptides were extracted from the gel pieces with a 50% (v/v) acetonitrile, 1% (v/v) triflouroacetic acid (TFA) solution. A 1 μl aliquot was spotted onto a sample plate with 1 μl of sample matrix containing 8mg/ml α-cyano-4-hydroxycinnamic acid in 50% (v/v) acetonitrile and 1% TFA solution and air-dried. A blank gel piece from the protein-free region of the gel was treated in parallel as a control experiment.

The mass spectrometry analysis of tryptic peptides was performed on a Micromass TofSpec 2E Time of Flight Mass Spectrometer (Waters Corporation, Milford, MA, USA). The mass spectra were acquired in a reflectron mode over a mass range of 600 to 3500 Da. Two trypsin auto-digestion peaks at 842.51 Da and 2211.1 Da were used for internal calibration of each spectrum. Monoisotopic mass peaks with a minimum relative intensity of at least 5% were extracted from the mass spectrum of each protein spot for peptide mass finger printing analysis.

2.5.3 Database searching using PMF data

The PMF peak list was used as query masses to search against the SWISS-PROT and NCBI non-redundant public protein database using the search engine Profound available at http://129.85.19.192/profound_bin/WebProFound.exe (Zhang and Chait 2000). The taxonomic category of *Oryza sativa* was searched first. If no significant match was found, then the whole Viridiplantae (green plants) database was searched. During the searches all the peptide masses were assumed to be monoisotopic and a mass

accuracy of ±100 ppm (part per million) was used. A maximum of one missing cleavage site and modification of cysteines to carboxyamidomethyl cysteine (Cys_CAM) by iodoacetamide or to propionamide cysteine (Cys_PAM) by acrylamide monomers were considered during the searches. Confidence in peptide mass finger printing matches was based on multiple criteria including the origin of the matching protein, the number of matching peptides and missed cleavages. The agreement between the isoelectric points and molecular masses of the matched proteins and those experimentally observed for the analysed protein spots was also taken into consideration during PMF homology searches.

At the same time, the PMF data were also used to search against the TIGR (The Institute for Genomic Research) Rice Gene Indices database available at http://www.tigr.org/. The Tentative Consensus (TC) and EST sequences (OsGI Release 7.0 - May 29, 2001) were downloaded and translated in six reading frames to generate all possible translation products of the EST sequences. The FLIP program developed by the Organelle Genome Mega-sequencing Program (OGMP) was used to translate the EST sequences, available at the OGMP website http://megasun.bch.umontreal.ca/ogmp/ogmpid.html. Then PMF homology searches of this custom database were done using MassLynx software version 3.4 (Micromass, Waters, Milford, MA, U.S.A.). The deduced protein sequences of the matching ESTs were then used for a BLAST homology search against the NCBI non-redundant protein database.

2.6 SOURCES OF CHEMICALS AND REAGENTS

The chemicals used in the experiments and their sources are listed in the table next page. Only the analytical or electrophoresis chemicals were used unless otherwise indicated.

Chemical or Reagent	Source
Acrodisc® filters, sterile	Gelman Sciences, MI, USA
Anti-rabbit IgG AP conjugate	BoehringerMannheim, Mannheim, Germany
Anti-rabbit IgG HRP conjugate	Amersham Biosciences, Uppsala, Sweden
Acetic acid	BDH, Poole, UK
Bio-lyte® 3-10 ampholytes	BioRad Laboratories, CA, USA
Bovine serum albumin, BSA	Sigma Chemical Co., St. Louis, MO, USA
Bradford Dye Concentrate	BioRad Laboratories, CA, USA
Bromophenol blue	Sigma Chemical Co., St. Louis, MO, USA
CAPS	Sigma Chemical Co., St. Louis, MO, USA
CHAPS	Sigma Chemical Co., St. Louis, MO, USA
Coomassie Brilliant Blue R-250	BioRad Laboratories, CA, USA
Coomassie Brilliant Blue G-250	BioRad Laboratories, CA, USA
DAPI dihydrochloride	Sigma Chemical Co., St. Louis, MO, USA
EDTA	Ajax Chemicals, Sydney, Australia
Ethanol, 99.5%	BDH, Poole, UK
Formaldehyde, 37% aqueous solution	Fluka, Neu-Ulm, Germany
Gelatine (cold fish)	Sigma Chemical Co., St. Louis, MO, USA
Glutaraldehyde, 50% aqueous solution	Sigma Chemical Co., St. Louis, MO, USA
Glycerol	ICN Biomedical Inc., Aurora, Ohio, USA
HEPES	Research Organic Inc., Cleveland, OH
Historesin embedding kit	Leica Instruments, Heidelberg, Germany
Iodoacetamide, IAA	Sigma Chemical Co., St. Louis, MO, USA
Mannitol	APS Chemicals, NSW, Australia
Methanol	BDH, Poole, UK
NBT/BCIP stock solution	BoehringerMannheim, Mannheim, Germany
Paraffin oil, light or Ondina 18 oil	Shell, Australia
Paraformaldehyde	Electron microscopy sciences, FT. Washington, PA, USA
Peroxidase, horseradish	Pharmacia-Biotech, Uppsala, Sweden
Ponceau S	Sigma Chemical Co., St. Louis, MO, USA
Potassium tetrathionate	Fluka, Neu-Ulm, Germany
Protein standards, LMW	Pharmacia-Biotech, Uppsala, Sweden
SDS	Amresco, Solon, OH, USA
Silver nitrate	ICN Biomedical Inc., Aurora, Ohio, USA
Sodium acetate trihydrate	BDH, Poole, UK
Sodium thiosulphate	BDH, Poole, UK
Thiourea	Fluka, Neu-Ulm, Germany
Toluidine blue	BDH, Poole, UK

CHAPTER 3 DETERMINATION OF POLLEN DEVELOPMENTAL STAGES IN RICE CULTIVAR DOONGARA

[Parts of the research presented in this chapter, together with chapter 4, has been published: Kerim, T., Imin, N., Weinman, J. J., Rolfe, B. G. (2003) Proteome analysis of male gametophyte development in rice anthers. *Proteomics* 3 (5): 738-51.]

SUMMARY

The developmental stages of male gametogenesis in the rice cultivar Doongara were cytologically examined for the period from pollen mother cell to panicle heading stage in plants grown under the strictly-controlled growth condition. The allometric relationships between the anther developmental stages and different growth parameters such as anther length, auricle distance and the days before heading were established to provide a set of growth measurements for the precise prediction of anther developmental stages. The measurement of anther length proved to be one of the most reliable growth measurements for the prediction of the early stages, while measurement of the auricle distance and the days before heading provided a nondestructive method for the assessment of pollen development at later stages. Based on the allometric relationships, an anther sampling strategy involving the combined use of anther length, auricle distance and the days before heading was devised to collect near homogenous anther population for six discrete anther developmental stages for protein profiling of these stages using proteomics tools.

3.1 INTRODUCTION

Microsporegenesis in angiosperms is a continuous developmental process encompassing a chain of cellular events taking place in growing anthers. Any nucleic acid or protein-based experimental approaches to study this reproductive process requires the collection of a homogenous population of developing pollens or anthers at a precisely defined developmental stage. Anther or pollen-based haploid tissue culture experiments also requires the culturing of anthers at a precisely defined stage (Zapata *et*

al. 1983). Therefore, determination of allometric relationships between growth parameters of flowers or other vegetative parts of a model plant system and cytological stages of pollen development is absolutely necessary for sample collection, molecular cytology and tissue culture of anther or pollen. Such an allometry has been studied in a number of model plants for male gametophyte study including *Lilium longiflorum* (Erickson 1948; Gould and Lord 1988), *Brassica napus* (Fan *et al.* 1988; Scott *et al.* 1991), *Brassica campestris* (Singh *et al.* 1985; Theerakulpisut *et al.* 1991), maize (Bedinger and Edgerton 1990) and rice (Kowyama *et al.* 1994; Raghavan 1988). Floret length and anther length are among the most often-used growth measurements to determine the anther and pollen developmental stages with some degree of accuracy because a considerable level of correlation between anther and floret length and pollen developmental stages was demonstrated in the above-mentioned studies.

Establishment of an allometric relationship between growth parameters such as anther and floret length and pollen developmental stages in rice is complicated by two intrinsic features of rice flowers. First, the flowering pattern of the rice flower belongs to a multiple inflorescence in which rice flowers occur in a group on a tiller, in contrast to the more general type of single inflorescence (as in Lilium). Developmental stages of spikelets in the same panicle are not synchronous and the age of spikelets in a panicle usually differs by more than 7 days (Satake and Hayase 1970). Therefore, developmental status of a spikelet in a certain position in a panicle cannot represent the developmental stages of other florets. To make the assessment more complicated, the panicle emerges from inside the flag leaf sheath only shortly before anther dehiscence, making the flowers inaccessible for the assessment. Second, the size of rice anthers is relatively small compared to those of the above-mentioned model plants (mature rice anthers attain only about 5% of the lengths of mature anthers of lily), which made it very hard to measure the length of anthers without the aid of a microscope. In the same way as for the panicle, rice anthers are enclosed inside the palea and lemmas until anther dehiscence occurs at heading stage. Owing to these features, the length of anther and floret cannot be measured nondestructively without sacrificing the entire tiller at early stages of anther development. For these reasons, it becomes necessary to use other nondestructive growth measurements in addition to anther and floret length while predicting cytological stages of developing anthers in rice.

There are several other growth markers that can be measured to determine anther developmental stages. The "leaf number index", which is calculated as the percentage of the total number of leaves existing on a main stem at a given time in relation to the total number of leaves produced on the main stem from germination to heading, is a simple nondestructive method used to assess the age of a rice plant. Panicle length and internode elongation can also be used for such purpose, but it requires the opening of leaf sheath that leads to the death of plants. Auricle distance (AD), the distance between the ligules of the flag leaf and penultimate leaf, can also be used as a nondestructive indicator of anther development. Correlation between the above-mentioned growth measurements and anther development is not as high as for anther or floret length, but still can be used to estimate the approximate age of a panicle in order to choose rice plants for dissection.

In addition to the above growth measurements, the number of days from the time of panicle initiation or meiosis to the time of panicle heading (or days before heading, DBH) can also be used in the prediction of the developmental status of anthers .The main reason for using the number of days before heading as growth indicator is that the number of days from initiation of panicles to heading is relatively constant (about 30 days) and this number does not vary too much with different varieties. According to a study on a mid-term *japonica* rice variety conducted by Kowyoma *et al.*, anther dehiscence occurs 13-14 days after the onset of microspore meiosis (Kowyama *et al.* 1994). Although the DBH value varies slightly with different cultivars differing in maturity and also with growing conditions, for a given variety that has been grown under the strictly controlled same growth conditions, the value should be constant.

According to Matsushima, *et al.* discrete developmental stages of panicle development can be divided into 21 stages (Matsushima and Manaka 1956). Although normal panicle development goes through all of these stages, the duration of each stage varies due to different varieties, varying times of cultivation and different conditions in which the plants are grown. Therefore, allometry established on a rice cultivar under the defined nutritional and photoperiodic growth conditions is not generally applicable to other cultivars or to the same cultivar grown under different growth conditions. This necessitates the establishment of an allometric relationship between growth measurements and anther developmental stages for each cultivar used for the

experiment and for the growth condition in which the cultivar is grown in order to collect more synchronous anther samples for predefined anther developmental stages.

3.2 MATERIALS AND METHODS

3.2.1 Plant material

Plants were grown as described in sections 2.1.1 and 2.1.2.

3.2.2 Naming scheme of rice spikelets

The naming scheme for the spikelets was adopted from Nishiyama (Nishiyama 1982) to distinguish spikelets from different locations of a panicle. Figure 3.1 illustrates the structure of a rice panicle and the naming scheme of spikelets at different position in a panicle. In the double-digit number assigned for each of the spikelets, the first digit indicates the position of a panicle branch in the whole panicle, and the second digit specifies the location of a spikelet on that branch.

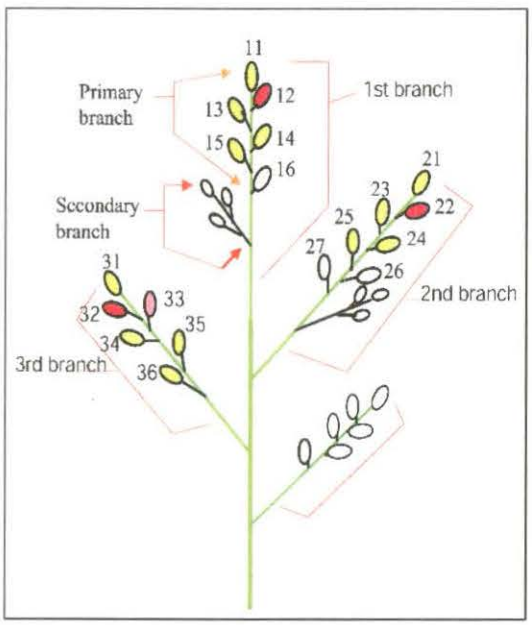


Figure 3.1. Schematic diagram of a rice panicle and naming scheme of spikelets. Numbering of only the top three primary branches is shown. Spikelets in yellow colour are used in our experiment.

3.2.3 Measurement of auricle distance and anther length

A few days before the panicle booting stage (when the auricle distance was around - 40 to 70 mm) the rice plants were dissected and cytological studies were used to predict the correlation between early microspore developmental stages and vegetative

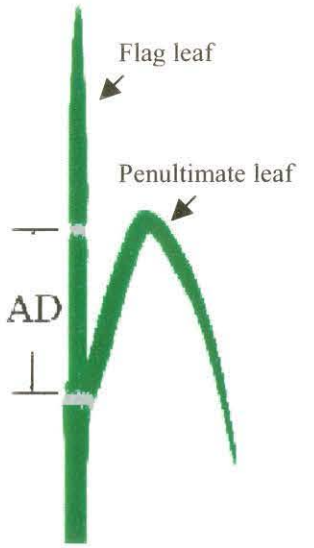


Figure 3.2. Auricle distance

growth measurements such as anther length and auricle distance. First, the auricle distance of each plant was measured with a student ruler and recorded. The auricle distance is expressed in terms of three values: (-), (0) and (+). When the auricle of a flag leaf was still inside the sheath of an immediate lower leaf the AD was assigned a negative (-) value, when the auricle of a flag leaf has already emerged out from the sheath of the immediate lower leaf the AD was assigned a positive value (+). When the two auricles coincide, the AD was assigned a value of 0. Then the lengths of the anthers of the top five spikelets of the top three primary branches were measured with a binocular dissection microscope equipped with an ocular micrometer using 2× and 4× objectives. Because of small variations in length of anthers from the same spikelet, the value of the longest anther in each spikelet was recorded. The developmental stages of the fresh anthers were determined either directly by cytological techniques or the anthers were fixed in a fixation solution containing 50% ethanol (v/v) and 5% (v/v) acetic acid for future cytological analysis.

3.2.4 Aceto-carmine anther squash technique

The developmental stage of each anther was subsequently determined by the aceto-carmine anther squash technique as previously described (Jong 1997). The anthers were placed on a drop of staining solution containing 0.4% aceto-carmine and 45% acetic acid on a microscope slide and were gently squashed with a glass rod to extrude enclosed microspores. The anther wall debris was removed from the solution using a dissection needle and a cover slip was applied. The excess staining solution was absorbed with a piece of filter paper before the slide was observed under a Nikon Optiphot light microscope. Photographs were taken on Fujichrome 400 colour slide film using a Nikon FX-35 A camera.

3.2.5 DAPI staining

The auricle distance and days before heading were recorded for each plant before the plant was dissected. DAPI staining was used to visualise the pollen nuclei to determine the late microspore developmental stages as described before (Vergne *et al.* 1987). DAPI was stored at 4°C as a 1 mg/ml stock solution in distilled water. One microliter of this stock solution was diluted in 1 ml of citrate phosphate buffer (pH 4) containing 1% Triton X-100 and fresh anthers, collected every morning from plants at various days before heading stage, were dissected in this solution to release the microspores. The microspores were incubated in this solution for 30 min under dark conditions and observed under a Nikon Optiphot light microscope using UV filter (main wavelength 365 nm). Photographs were taken on Fujichrome 400 colour slide film using a Nikon FX-35 A camera.

3.2.6 Measurement of Days Before Heading (DBH)

The batch of rice plants that were transferred to a controlled environmental facility was monitored closely until they encountered the booting stage (when the flag leaf starts to emerge from the penultimate leaf). When the auricle distance of the rice plants was approaching 0 mm (or within the range of –10 to +10 mm) the auricle distance of each plant was measured as described in section 3.2.3 and each plant was tagged with the auricle distance, current time and date. Then, the auricle distance of each plant was measured and recorded at the same time every day until panicle heading occured. At the same time, the developmental stages of anthers of top spikelets in that auricle distance range were determined using one or two plants by cytological methods mentioned in sections 3.2.4 and 3.2.5. The progressive increase in auricle distance in every 24-hour interval and the number of days from the time of auricle distance 0 to the time of panicle heading were calculated, and the relationship between DBH and AD was analyzed by a scatter plot analysis.

3.2.7 Data analysis

Data analysis was carried out using the Microsoft® Excel 2000 program.

3.3 RESULTS

3.3.1 Correlation between anther length and microspore developmental stages

In order to relate cytological stages of microspore development to changes in anther length, 200 rice plants representing pollen mother cell stage to panicle heading stage were dissected and the cytological analyses were carried out using the anthers from the top five spikelets of the top three primary branches of the rice panicles. Cytological examination of anthers from individual spikelets dissected from different panicle positions showed that the developmental stages of anthers from individual spikelets in a primary panicle branch are not synchronous. The order of spikelet maturation in a panicle branch is: n1→ n5 → n4→n3→ n2 (refer to section 3.2.1 for the definition of spikelet numbers). This order of spikelet maturation is also reflected in the gradual increase of length of anthers from individual spikelets in the same branch except for the fact that the average length of anthers from fifth spikelets (n5) is up to 6% longer than the apical spikelets (n1) (data not shown). The difference in developmental stages between a apical spikelet (n1) and second spikelet (n2) is very significant (more than two days in terms of days before heading), while this value is around half a day or less for other spikelets.

The length of anthers and their corresponding cytological developmental stages were recorded in Table 3.1 and the histological observations of rice microspore development are shown in Figure 3.3. About 85% of the microscopically examined anthers within the anther length range of 625-900 μm were at pollen mother cell stage (Figure 3.3-A). The subsequent tetrad stage is very transitory (lasting less than a day), and the two rounds of meiosis occurred within the anther length range of 825-1000 μm (Figure 3.3-B, C). The microspores enclosed in tetrads were released to form early young microspores in anthers with a length of 925-1225μm. At this stage the germination pores of the young microspores were hardly visible and the microspores were not vacuolated (Figure 3.3-D). The free young microspores increased in volume rapidly and had a large nucleus, which was located in the cell periphery position opposite the pollen pore (Figure 3.3-E). These middle young microspores were dominant in the anther length range of 1300-1500μm. At the early binucleate microspore stage, the nucleus of the vacuolated middle microspore produced two asymmetric nuclei as a result of first pollen mitosis, which occurred in anthers with a

length of 1650-1800 μm (Figure 3.3-F). The results of cytological analyses mentioned above showed that microspore development during the early stages was accompanied by a progressive rapid increase in anther length. Therefore, the microspore developmental stages can be correctly predicted by measuring the length of anthers at these early stages.

The earlier phase of rapid elongation of anther length was followed by a later phase of slow growth in length at later stages. Approaching the end of the early binucleate microspore stage, it became very difficult to correlate microspore developmental stages to anther length as the anthers attain almost full length (1800-2000μm). The binucleate pollen grains steadily increased in size and the nuclei moved to the central part of the pollen after detaching from the inner pollen wall (Figure 3.3-G). After another round of pollen mitosis, the trinucleate pollen grains were formed by the production of two sperm nuclei from the generative nucleus. Based on the above-mentioned observations, seven cytologically distinguishable unique developmental markers listed in Table 3.2 were adopted in order to divide the whole process of microspore development (after pollen mother cell stage).

3.3.2 The relationship between auricle distance and microspore developmental stages

At the same time as the cytological analysis was undertaken, the auricle distances of the dissected plants were also recorded. The relative dominance of each of the early microspore developmental markers on the top five spikelets of the top three primary panicle branches at each auricle distance point were studied. The results are expressed in Figure 3.4 as the percentage of each of the developmental stages present at each auricle distance length. More than 80% of the examined anthers were at pollen mother cell stage when the auricle distance was –30 mm. The auricle distance range of –28 to –20 mm has covered three early stages where the appearance of tetrad stage overlaps with those of pollen mother cell and early young microspore stages. The early young microspore and middle microspore stages were dominant in the top spikelets at auricle distances of –10 to 0 mm and 35 to 55 mm respectively.

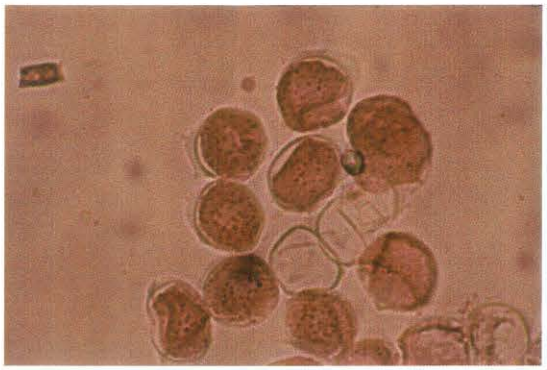
The graph shows that auricle distance can predict with some degree of accuracy the stage of panicle development at early stages without destruction of plants. However,

Table 3.1. Record of anther length and the corresponding stages

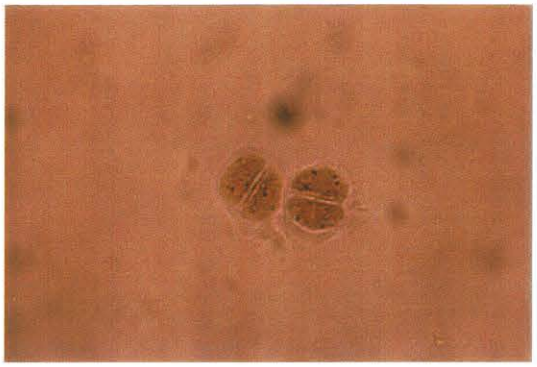
Range of anther length									
<700	700-800	800-900	900-1000	1000-1100	1100-1200		1200-1300	1300 -1500	1400-1500
P 650	P 775	T 825	T 925	E 1000	E 1125	M 1175	M 1200	M 1325	1520
P 625	P 750	P 825	E 925	E 1000	E 1100	E 1150	E 1225	M 1350	1500
P 650	P 750	P 800	P 925	E 1025	E 1125	M 1150	M 1200	M 1375	1500
P 625	P 750	P 850	T 900	E 1000	E 1100	M 1150	M 1250	M 1300	1550
P 650	P 750	P 850	T 925	E 1050	E 1125	E 1125	M 1250	M 1325	1550
P 650	P 775	P 850	T 900	E 1000	E 1100	E 1075	M 1250	M 1350	1525
P 675	P 700	T 875	P 900	E 1025	E 1125	E 1100	M 1200	M 1325	1574
	P 700	T 850	T 975	E 1025	E 1150	E 1150	M 1225	M 1350	1625
		P 850	T 975	E 1050	E 1075	E 1150	E 1200	M 1475	1575
		T 850	E 975	E 1050	E 1075	M 1175	M 1225	M 1475	
		P 875	T 975	M 1100	E 1075	E 1175	M 1275	M 1400	
			P 975	E 1100	E 1100			M 1450	
			T 975	E 1050	E 1075			M 1425	
			T 950	E 1075	E 1050				
			T 975	E 1050					
			T 925	E 1075					
			T 900	T 1000					

Notice: P - pollen mother cell
T - tetrad
E - early young microspore
M- middle young microspore

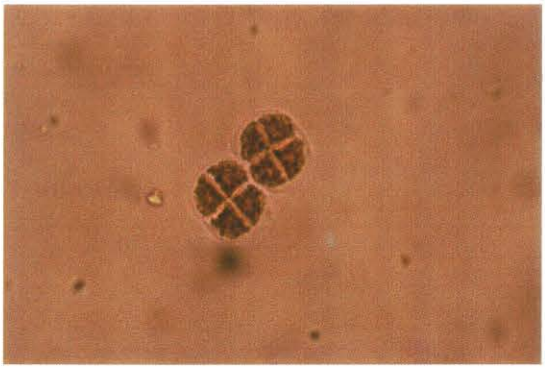
Figure 3.3. Microspore development and pollen maturation in rice cultivar Doongara. The developmental stages of the fresh rice anthers, which were collected at the different time points of microspore development, were examined under a light microscope using anther squash technique followed by staining either with aceto-carmine (**A-D**) or with DAPI (**E-H**). The micrographs were taken for the developing microspores at: **A**-pollen mother cell stage, **B**- first meiotic division, **C**-tetrad stage, **D**-early young microspore stage, **E**-middle young microspore stage, **F**- early binucleate microspore stage, **G**- late binucleate microspore stage, and **H**- heading (trinucleate pollen grain) stage.



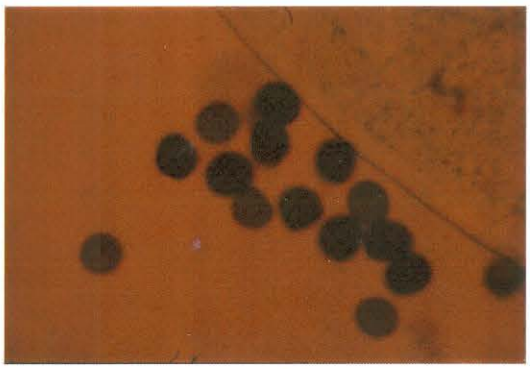
A



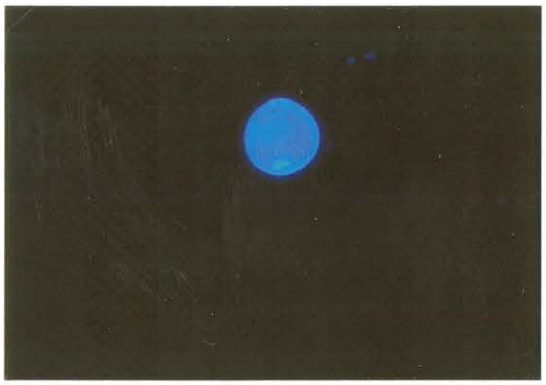
B



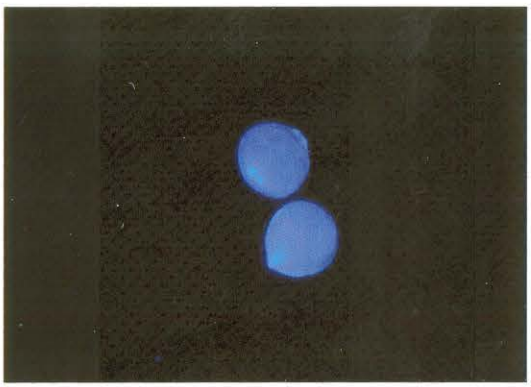
C



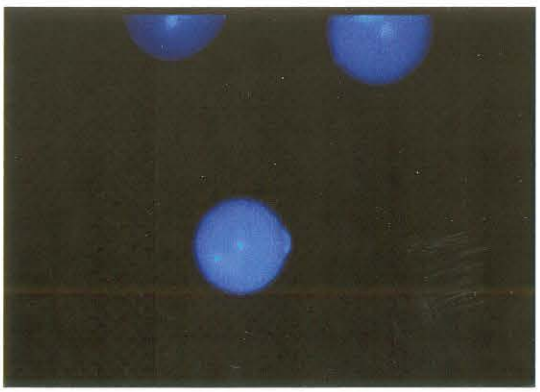
D



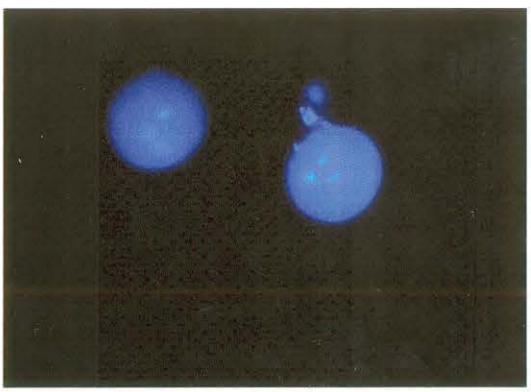
E



F



G



H

Table 3.2. The microspore developmental stages in rice cultivar Doongara

Stage of development	**Range of anther lengths** (μm)	**Range of auricle distance** (mm)	**DBH**[a)] (days)	**Developmental criteria**
Pollen mother cell	700-900	-40 to -30	14	Relatively large in size, callose wall starts to form.
Tetrad	800-950	-35 to -25	13-13.5	Four microspores enclosed in callose wall.
Early young microspore	1000-1200	-10-0	11	Non-vacuolate microspores just released from tetrad, germination pore is not visible.
Middle young microspore	1350-1500	40-55	8	Vacuolate microspore, germination pores are visible.
Early binucleate	1650-1800	70-95	5	Binucleate, pollen starts to accumulate starch.
Late binucleate	1800-2000	80-125	3	Pollen starch grains begin to accumulate, pollen become engorged.
Heading stage	1800-2000	110-150	0	Trinucleate, pollen grains completely starch filled, stigma become very pink in colour.

Microspore developmental stages and the corresponding anther length, auricle distance and the approximate days before heading are shown. a) DBH- days before heading.

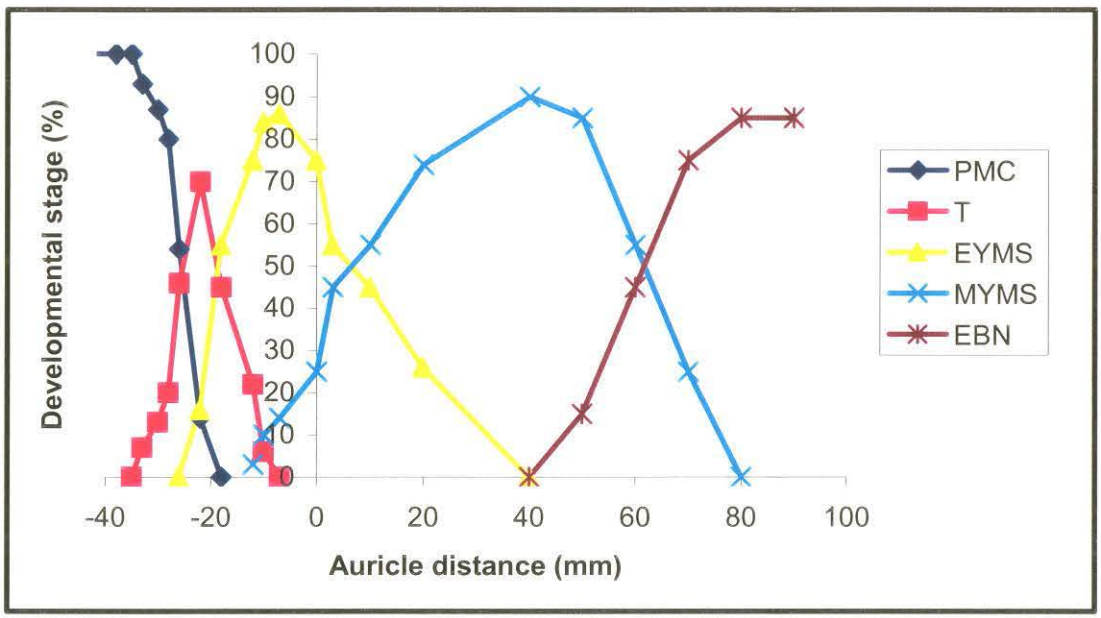


Figure 3.4. The relationship between auricle distance and developmental stages of microspores from top three primary panicle branches in rice cultivar Doongara. The graph indicates the relative dominance of anthers at a particular developmental stage at a given auricle distance range. PMC- pollen mother cell stage; T-tetrad stage; EYMS- early young microspore stage; MYMS- middle young microspore stage; EBN- early binucleate microspore stage.

depending on the results of our experiments, the quality and consistency of this kind of relationship is largely dependent on growth conditions such as growing season, light intensity, day length and temperature. Slight changes in growth conditions can change significantly the allometric relationship between auricle distance and anther developmental stages. The total tiller number of the plant and the tiller type (main or side tiller) also affect the correlation. While growing a batch of rice plants, the elevated light intensity of the growth chamber resulting from erroneous set up of growth condition resulted in a completely different correlation between auricle distance and developmental stages (data not shown).

3.3.3 The days before heading and microspore developmental stages

The correlation between the number of days before heading and anther developmental stages was studied in two batches of rice plants. Instead of using the panicle initiation stage as a starting point, the auricle distance of 0 was marked as a starting point as it is easily distinguishable. The auricle distance of plants was measured and recorded everyday. The developmental stages of anthers of top spikelets were determined by cytological analysis in combination with using anther length and auricle distance. Although some variations were found between two batches of plants, more than 90% of the profiled rice plants commenced panicle heading an average of eleven days after auricle distance was zero (0). The relationship between the progressive increase of auricle distance and the days before heading are presented in Figure 3.5. The approximate date of occurrence of major cytological markers is also shown. At the starting point of profiling (11 days before heading) the plants were at young microspore stage. At 8-9 day before heading, the majority of the anthers of the top spikelets were observed to be in middle young microspore stage. The first and second pollen mitoses were observed in anthers sampled at around 5 days and 3 days before the heading stage respectively.

The plot of auricle distance to number of days before heading in Figure 3.5 demonstrated the concomitant increase of auricle distance as the pollen maturation gradually progressed. At the same time, the standard deviation of the average auricle distance value also significantly increased (up to 8%) as the plants were gradually approaching the heading stage.

3.4 DISCUSSION

The study of male gametophyte development at the molecular level requires the collection of homogenous populations of pollen or anther materials at specific developmental stages. Unlike pollen development in some higher plants, the pollen development in rice is asynchronous; the developmental stage of microspores in anthers from the different spikelets in the same panicle can very greatly. In order to collect enough synchronous anther samples for proteome analysis of each discrete developmental stage, it was essential to establish a reliable, non-destructive growth measurement to enable the assessment of the precise cytological stage of microspore development within the inflorescence of an Australian rice cultivar, Doongara. The synchrony of development among the microspore populations within a spikelet and the strong allometric relationship that exists between a number of growth parameters and anther developmental stages has made it possible to achieve this goal. The cytological stages of rice male gametophyte development were cytologically examined with special attention given to differentiation and development of anthers from individual spikelets. The results of histological observation as shown in Figure 3.3 are in general agreement with descriptions of microgametogenesis in an *indica* rice cultivar, IR-30 (Raghavan 1988) and in a *japonica* rice cultivar, Aichi-Asahi (Kowyama *et al.* 1994). The study shows that the development of spikelets n1, n5, n3 and n4 in the same panicle branch demonstrated reasonable synchrony whereas the spikelet n2 showed a significant level of asynchronous development from other spikelets in the same branch. Therefore, the n2 spikelets were excluded while collecting anther samples in order to increase the homogeneity of the anther population sampled for a particular developmental stage. Aceto-carmine staining of anther squash preparation used in this study provided a quick and reliable method for the determination cytological markers of anthers at early stages. As the observation of cytological markers of later stages involves the visualization of nuclei of the developing pollen grains, which cannot be achieved by aceto-carmine staining, the florescence dye (DAPI) was used for the visualization of multiple nuclei of developing pollen and proved to be an effective staining method in the determination of late stage-specific cytological markers.

Establishment of an allometric relationship between anther developmental stages and different growth parameters of the plant requires the construction of a set of

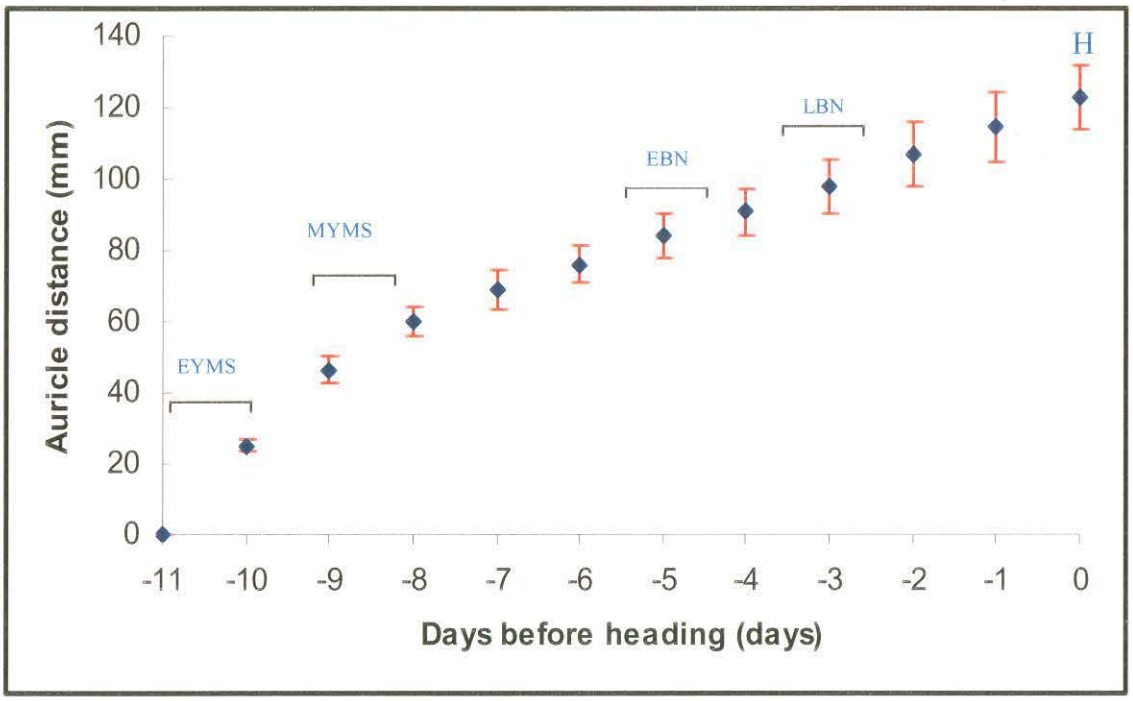


Figure 3.5. The relationship between auricle distance (AD) and the number of days before panicle heading (DBH). The microspore developmental stages that occur at specific DBH are indicated in blue letters. The days before heading are given in negative values. Error bars represent the standard deviation in auricle distance. PMC-pollen mother cell stage; T-tetrad stage; EYMS-early young microspore stage; MYMS-middle young microspore stage; EBN- early binucleate microspore stage; LBN- late binucleate microspore stage; H- heading stage.

cytological "markers" for the anther development based on the cytological events of developing microspores or somatic anther tissues (Scott *et al.* 1991). As the tapetum of rice anther changes very little during early stages of microspore development and the cytological changes in tapetum and other somatic tissue cells cannot be visualized by anther squash preparations used in this study, the cytological events which occurred in those somatic cells were not used as markers in defining the distinctive microspore developmental stages. Based on the developmental criteria described in similar studies on different rice varieties and on our initial cytological analysis results, the rice male gametophyte developmental process was divided into seven microscopically distinguishable stages (Table 3.2). The early archesporial cell differentiation stages that lead to the formation of pollen mother cells were not included in this study because of the difficulty of collecting sufficient anther material for protein extraction due to the small size of rice anthers at these early stages.

The results (Table 3.1) show that the progression of early anther developmental stages is accompanied by a concomitant increase in anther length. Therefore, the developmental status anthers can be assessed by simply measuring the length of the anthers at the early phase of anther elongation. However, the later stage of pollen development could not be effectively predicted based on the anther length because of the slow elongation rate and relatively longer duration time (2-3 days) of later stages. In this regard the calculation of the days before heading can be used for the determination of later anther developmental stages based on the correlation data presented in Figure 3.5. From the data presented in Figure 3.4 it can be seen that the top branches of a panicle are dominated to some degree by anthers belonging to a particular developmental stage at a given auricle distance range, even though the development of individual spikelets in the same branch varies. This enabled us to initially choose a plant, which has a most enriched anther population for a particular development stage, for the dissection based on the auricle distance range. As shown in Figure 3.5, the significant increase in the standard deviation of auricle distance value approaching the later stages indicates the divergence of auricle distance elongation among the individual plants at panicle maturation, which makes auricle distance not very suitable for precise determination of anther development at these stages. It is important to note that the auricle distance and the days before heading are not absolute measures of the anther developmental stages since they vary depending on the varieties and growing

conditions. Slight changes in the growth season can also alter the correlation between developmental stages and auricle distance. Therefore, the correlation was checked and calibrated (if necessary) by a quick cytological analysis on each new batch of plants before the start of anther sampling.

Among the three growth parameters investigated in this study, the anther length proved to be the most reliable and direct growth measurement that can be used for the precise determination of early stages of anther development. However, it is a destructive method that requires the opening of the flag leaf sheaths and spikelets whereas the other two measurements are nondestructive. Therefore, the plants can be chosen for anther sampling of a particular stage based on the auricle distance and the days before heading (DBH) as nondestructive measurements. The length of the anthers can be monitored under the microscope while anther sampling, to collect anthers within a certain anther length range representing to a particular stage. As the tetrad stage is very transitory and there is also a significant level of overlap in anther length between the tetrad and pollen mother cell stage (Table 3.1 and Figure 3.4), collecting separate anther samples for the tetrad stage was not pursued. Instead, the tetrad and pollen mother cell stages were combined for anther sampling in this study.

In conclusion, a strategy for the precise assessment of anther developmental stages using the combination of three different growth parameters, i.e. anther length, auricle distance and days before heading has been established in this work to collect homogenous anther population for proteomic analysis in rice cultivar Doongara.

CHAPTER 4 IDENTIFICATION OF STAGE-SPECIFIC RICE ANTHER PROTEINS BY PROTEOME ANALYSIS

[The research presented in this chapter has been published: Kerim, T., Imin, N., Weinman, J.J., Rolfe, B.G. (2003) Proteome analysis of male gametophyte development in rice anthers. Proteomics 3 (5): 738-751]

SUMMARY

Two-dimensional electrophoresis-based proteomics tools were employed to investigate the changing patterns of protein expression during pollen development in rice anthers collected from rice plants grown under strictly controlled growth conditions. The allometric relationship that was determined between the various growth measurements and the pollen developmental stages allowed the collection of homogenous anther populations representing six discrete pollen developmental stages. Proteins were extracted from the anther samples and separated by two-dimensional gel electrophoresis to produce proteome maps. The anther proteome maps of different developmental stages were compared and 150 protein spots, which were changed consistently during the development, were analysed by MALDI-TOF mass spectrometry to produce peptide mass fingerprint (PMF) data. Database searches using these PMF data enabled 49 of the protein spots to be identified. These 49 proteins represent 42 unique gene products. Eight protein spots that could not be identified by PMF analysis were analysed by *N*-terminal micro sequencing. Multiple charge-isoforms of vacuolar acid invertase, fructokinase, β-expansin and profilin were identified. These proteins are closely associated with sugar metabolism, cell elongation and cell expansion, all of which are cell activities that are essential to pollen germination. The existence of multiple isoforms of the same proteins suggests that during the process of pollen development some kind of post-translational modification of these proteins occurs.

4.1 INTRODUCTION

In higher plants the development of the male gametophyte is a well-programmed and elaborate process that plays a crucial role in plant reproduction (for reviews see Bedinger 1992; McCormick 1993; Scott 1991). The whole developmental process of

micro-gametogenesis is controlled by coordinated gene expression in both somatic and gametophytic cells. Although many anatomical features of the male gametophyte developmental process have been extensively studied in a number of model plants as outlined in Chapter 1, many molecular and biochemical aspects of this important reproductive developmental process remain poorly understood and the dynamic change in protein synthesis throughout male gametophyte development is one of them. A temporal and spatial regulation of gene expression has been reported during male gametophyte development in higher plants (Gasser 1991). A number of research programs focused on the molecular biology of anther differentiation have identified several anther-specific genes, which are preferentially expressed in anthers of rice and certain other plants. However, the exact biological function of these anther-specific gene products and the roles of these proteins in pollen developmental pathways are not clear.

Rice (Oryza sativa L.) is an important crop, which provides staple food for about half of the world's population (Sasaki and Burr 2000). It has a relatively small genome (~430 Mb), and shares a high degree of synteny with other cereal crop plants. Rice is the easiest of the cereal plants to transform (Gale and Devos 1998). Rice is also the first crop plant whose genome has been sequenced. Recently, two groups simultaneously published drafts of the rice genome sequences of the two major rice subspecies, *indica* and *japonica* (Goff, Ricke *et al.* 2002; Yu, Hu *et al.* 2002). These unique features of the rice plant make it an ideal system for studying the molecular mechanisms of plant developmental processes including male gametophyte development, in spite of the physical limitations imposed by the small size of the rice anthers.

The availability of a complete rice genome sequence plus a large amount of EST sequence data has started to shift the experimental emphasis from a "reductionist" approach of studying genes and their products in isolation, to the "systems biology" approach of studying the structure and dynamics of cellular and organismal function of a whole system at different gene expression levels (Hood 2003; Raikhel and Coruzzi 2003). As the rice genome sequence has become publicly available, the prediction of possible open reading frames (ORF) by bioinformatics and the understanding of the functions of gene products are the first steps of exploiting this wealth of genome sequence information. However, there is not a strict linear relationship between genes and their protein products. The prediction of an open reading frame in genomic data by

sophisticated gene prediction programs does not necessarily imply the existence of a functional gene or the translation of a protein (Pandey and Mann 2000). Therefore, the verification of the existence of a gene product in the protein complement of a living organism is an important step in understanding the function of a gene. In this respect, 2-DE based-proteomics analysis is currently the best research tool to complement genomics. The application of 2-DE-based proteomics to study a living organism with a sequenced genome can provide information about the expression patterns of proteins coded by the genome, their possible functions and interactions, and their post-translational modifications. As previously described in Chapter 1, this kind of functional information obtained by proteomic analysis cannot be predicted directly by genome sequencing or by transcriptional profiling.

As described in Chapter 1, the proteomics tools have been applied to different aspects of rice research. In the field of plant male gametophyte development research, the 2-DE based proteomic approach has been employed to differentially display the proteins extracted from microspores of maize, which were isolated by sucrose gradient centrifugation, at three developmental stages (Bedinger and Edgerton 1990). Differential protein synthesis and protein phosphorylation patterns of *Brassica napus* microspores during microspore embryogenesis have been investigated by 2-DE and *in vivo* radiolabelling (Cordewener, Bergervoet *et al.* 2000). The pollen coat proteins of self-incompatible and self-fertile Festuca pratensis were also profiled by the same technology (Kalinowski, Winiarczyk *et al.* 2002; Kalinowski, Winiarczyk *et al.* 2001). Unfortunately, no large-scale protein identification work has been carried out in either of the studies mentioned above. In this study, homogenous anther populations for six discrete developmental stages were collected based on the allometric relationship determined by cytological study as described in Chapter 3. The changes in protein expression patterns at these six different male gametophyte developmental stages have been profiled using state-of-the art proteomics technology in an attempt to characterise the signalling and metabolic pathways involved in the male gametophyte developmental process in rice anthers. The temporal changes in protein synthesis and post-translational modifications of some proteins during pollen development in rice demonstrated by this study has enabled us to obtain some insight into the possible pathways and protein interaction networks underlying the complex cellular process of pollen development.

4.2 MATERIALS AND METHODS

4.2.1 Anther sampling

The anther samples were collected for pollen mother cell (including tetrad) , early young microspore, middle young microspore, early binucleate, late binucleate and heading stages by the micro dissection method as described in section 2.1.3 of Chapter 2. The allometric relationships between the growth parameters and microspore developmental stages determined in Chapter 2 have been used for the prediction of microspore developmental stages. The allometric relationship presented in Table 3.2 was always checked by a quick cytological analysis on each new batch of plants before the start of anther sampling. The second spikelets (n2) were excluded from anther sampling.

4.2.2 Protein extraction and 2-DE

Protein extraction for each microspore developing stage was started with at least 100 mg of anther material. Protein extractions, 2-DE analysis, gel staining and database searching was carried out essentially as described in Chapter 2.

4.2.3 Database searching using PMF data

Database searching using the generated peptide peak lists was carried out in two steps as described in the section 2.5.3 of Chapter 2. In addition to these two-step searching procedures, the same PMF data was also used to search against the functionally annotated full-length cDNA database of *japonica* rice that was released recently (Kikuchi *et al.* 2003). The data was downloaded from the KOME (Knowledge-based Oryza Molecular Biology Encyclopedia) website available at http://cdna01.dna.affrc.go.jp/cDNA/, and the Mascot program (Matrix Science Ltd, London, UK), which is running on our local server, was used for the interrogation of the downloaded cDNA database. The parameters used for the database searching were the same as those described in section 2.5.3 of chapter 2.

4.3 RESULTS

4.3.1 Proteome maps of rice anthers at six developmental stages

The proteins were extracted from anthers at pollen mother cell stage, early young microspore stage, middle young microspore stage, early binucleate stage, late binucleate stage and heading stage. An approximate protein yield of 2% of the starting weight of the anthers was achieved when the TCA-acetone precipitation method was used. The extracted proteins (100 μg for each stage) were separated by 2-D electrophoresis using the pH 4 to 7 and the pH 6-11 IPG strips in the first dimension isoelectric focusing step. The established 2-DE maps for the above six stages within the pH 4 to 7 range are shown in the Figures 4.1- 4.6, and the proteome maps within the pH 6 to 11 range are shown in the Figures 4.7 to 4.12. Image analysis revealed over 2500 protein spots on each gel (stained with silver) within the pH range of 4 to 7 and the size range of 6 to 120 kDa. Over 1000 silver stained protein spots were resolved within the pH range of 6 to 11 and the size range of 6 to 120 kDa. The overlap between the pH 6 to 7 ranges allowed the two gels for each analyzed step to be aligned and also provided greater resolution within the pH 6 to 7 range. The 2-DE analysis in these two pH ranges provided over 320 mm of effective gel distance for resolving the anther proteins in each developmental stage.

4.3.2 Comparison of proteome maps

The proteome map of the pollen mother cell stage was designated as our reference map and the proteome maps of other developing stages were compared to the reference map to detect the proteins differentially expressed during the rice anther development. The results of the manual and computer assisted gel comparison of the three replicate experiments showed that while the expression levels of a majority of the anther proteins remained unchanged, the expression levels of over 150 protein spots (less than 5% of the total proteins resolved by 2-DE) were found to change by more than two-fold. The enlargements of the selected gel regions of the six proteome maps, where significant and reproducible spot differences were detected, are shown in Fig 4.14 and Fig 4.15 . The comparison of proteome maps of the pollen mother cell and the heading stages, which are the earliest and latest two stages being investigated in this study, revealed that more than 50 proteins were newly induced at the heading stage. The stage-specific

expression patterns of a number of changed proteins that were identified by PMF analysis are presented in Table 4.1. Approximately half of the identified protein spots are absent at the early stages of microspore development and appear or increase in amount at the later stages.

4.3.3 Identification of differentially displayed proteins by peptide mass fingerprinting analysis

In order to identify the differentially expressed proteins, 155 unique protein spots were excised from Coomassie blue G-250 stained gels and analysed by MALDI-TOF-MS to generate peptide mass fingerprinting data. Some differentially displayed protein spots, which can be detected on silver gels but cannot be visualized in the preparative Coomassie stained gels due to the different staining properties and low abundance levels, could not be analyzed. Of the 155 protein spots analyzed by mass spectrometry, no usable spectra were derived from 10 proteins, possibly due to the low abundance of the protein sample, or because of ion suppression in the MALDI source.

Using the three-step database interrogation, putative identities of 49 protein spots (31.6% of the total protein spots analyzed by MALDI-TOF) were assigned with high confidence. The identified proteins were circled and assigned arbitrary numbers as shown in Figures 4.1-4.6. The protein identification results by PMF analysis are shown in Table 4.1. In the first step, the PMF homology searching of SwissProt and NCBI non-redundant public protein databases using the online search engine Profound revealed the identities of 12 proteins spots (24% of the identified proteins). In the second approach, the peptide lists were searched against the six reading frame translations of the annotated rice TC sequences, which were downloaded from the TIGR Gene Indices database, using the MassLynx program running on the local server. The search resulted in the identification of 38 proteins (77% of the total identified proteins), including the 12 protein spots that were previously identified by the first searching step. In the third step, the PMF data was searched against the functionally annotated full-length cDNA database, which was released more recently, using the Mascot program. This search revealed the identities of further 11 proteins, which were not identified by the previous two approaches. This Mascot search also identified 30 proteins which were identified by the previous two searches. The rest of the 155 protein spots did not display

Figure 4.1. 2-DE reference proteome map of anther proteins from rice cultivar Doongara at the pollen mother cell (PMC) stage. Isoelectric focusing in the first dimension was carried out on a linear pH 4-7, 18 cm IPG strip loaded with 100 μg of anther protein. In the second dimension a 12-14% SDS-PAGE gel was used. Proteins were visualised by silver staining. The identified gel spots are circled and the arbitrary identifiers used correspond to those listed in Table 4.1. The molecular weights of the protein standards are denoted on the left (kDa) and the isoelectric points are denoted at the bottom.

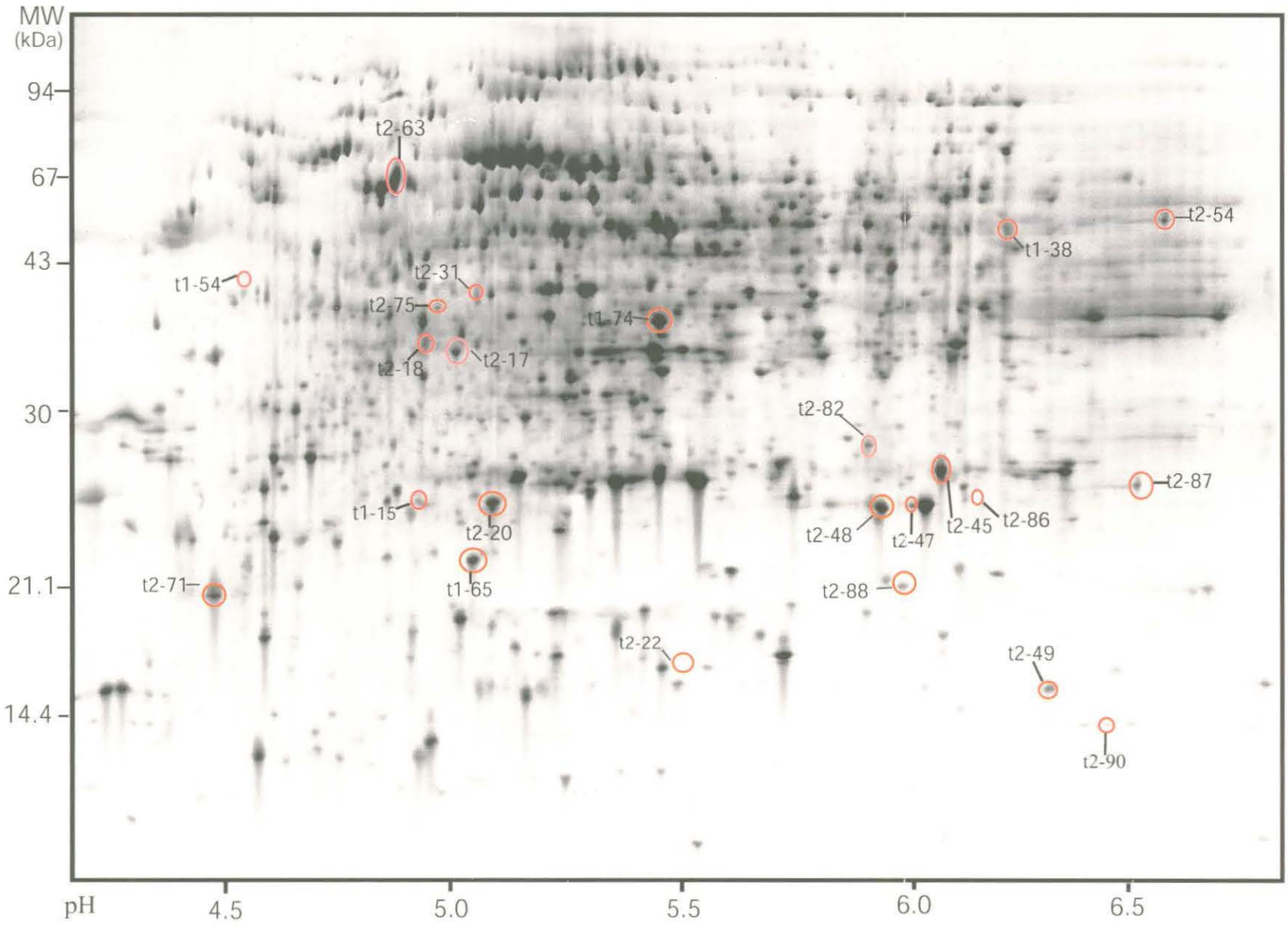

MW
(kDa)
94
67
43
30
21.1
14.4
pH
4.5
5.0
5.5
6.0
6.5
t2-63
t1-54
t2-31
t2-75
t1-74
t2-18
t2-17
t1-38
t2-54
t2-82
t2-87
t1-15
t2-20
t2-48
t2-47
t2-45
t2-86
t2-71
t1-65
t2-88
t2-22
t2-49
t2-90

Figure 4.2. 2-DE proteome map of anther proteins from rice cultivar Doongara at the early young microspore (EYMS) stage. The anthers were collected about 11 days before heading (11 DBH). Isoelectric focusing in the first dimension was carried out on a linear pH 4-7, 18 cm IPG strip loaded with 100 μg of anther protein. In the second dimension a 12-14% SDS-PAGE gel was used. Proteins were visualised by silver staining. The protein spots identified by PMF analysis are circled and the arbitrary identifiers used correspond to those listed in Table 4.1. The molecular weights of the protein standards are denoted on the left (kDa) and the isoelectric points are denoted on the bottom.

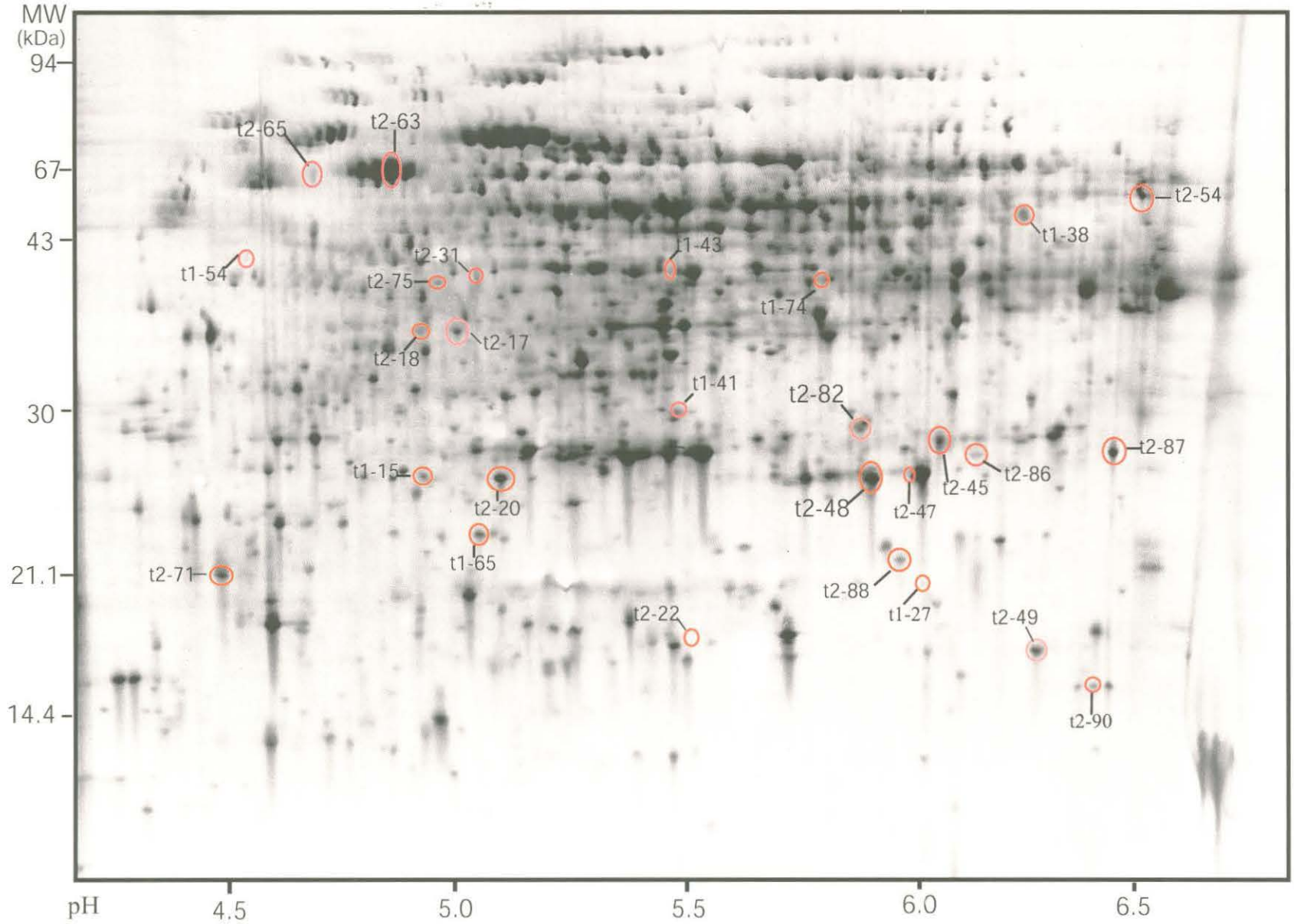

MW
(kDa)
94
67
43
30
21.1
14.4
pH
4.5
5.0
5.5
6.0
6.5
t2-65
t2-63
t2-54
t1-38
t1-54
t2-31
t1-43
t2-75
t1-74
t2-17
t2-18
t1-41
t2-82
t2-87
t1-15
t2-86
t2-45
t2-20
t2-48
t2-47
t1-65
t2-71
t2-88
t1-27
t2-22
t2-49
t2-90

Figure 4.3. 2-DE proteome map of anther proteins from rice cultivar Doongara at the middle young microspore (MYMS) stage The anthers were collected about 8 days before heading (8 DBH). Isoelectric focusing in the first dimension was carried out on a linear pH 4-7, 18 cm IPG strip loaded with 100 μg of anther protein. In the second dimension a 12-14% SDS-PAGE gel was used. Proteins were visualised by silver staining. The protein spots identified by PMF analysis are circled and the arbitrary identifiers used correspond to those listed in Table 4.1. The molecular weights of the protein standards are denoted on the left (kDa) and the isoelectric points are denoted at the bottom.

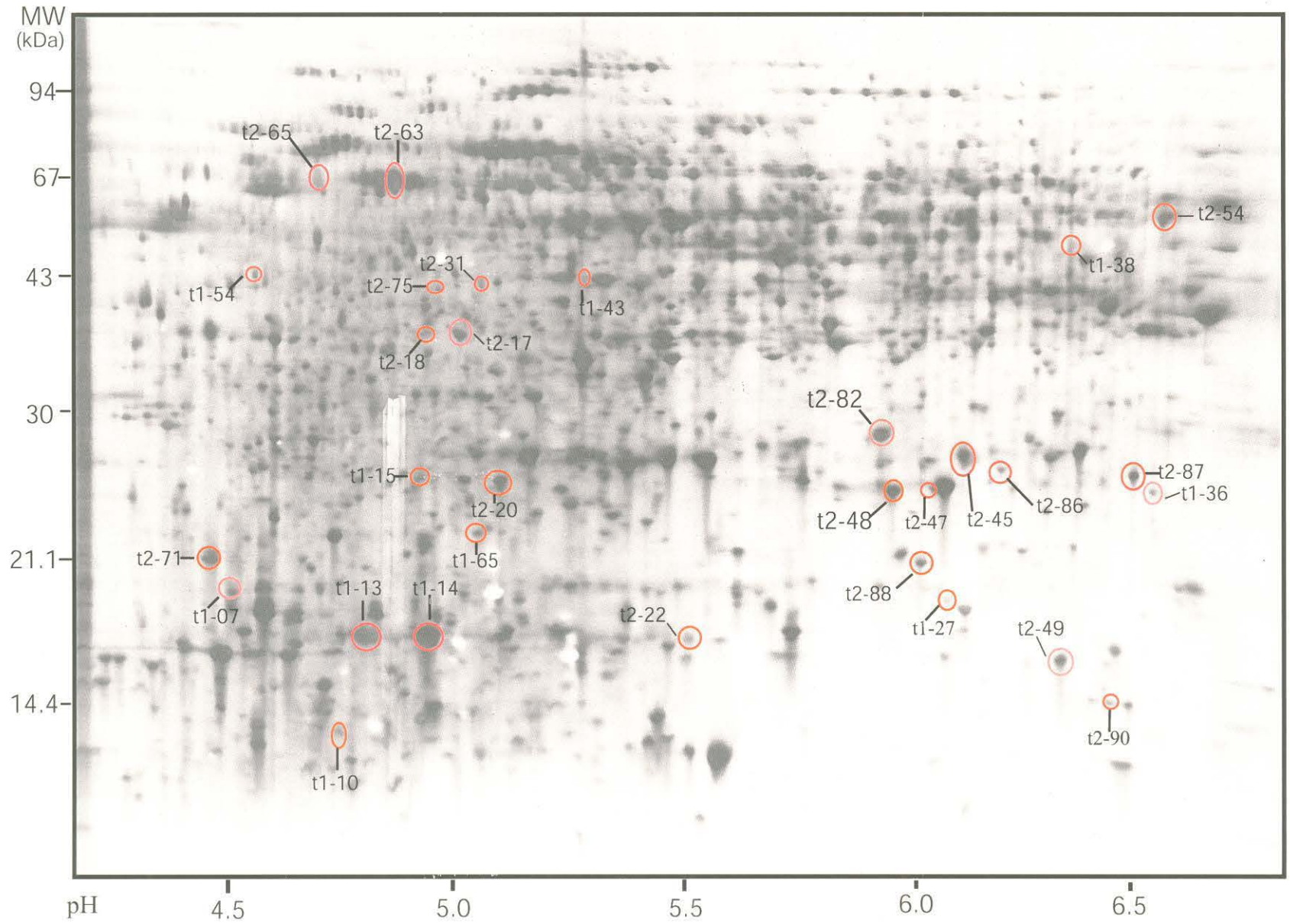

MW
(kDa)
94
67
43
30
21.1
14.4
pH
4.5
5.0
5.5
6.0
6.5
t2-65
t2-63
t2-54
t1-38
t1-54
t2-31
t2-75
t1-43
t2-17
t2-18
t2-82
t1-15
t2-20
t2-87
t1-36
t2-48
t2-47
t2-45
t2-86
t2-71
t1-65
t2-88
t1-07
t1-13
t1-14
t2-22
t1-27
t2-49
t2-90
t1-10

Figure 4.4. 2-DE proteome map of anther proteins from rice cultivar Doongara at the early binucleate (EBN) microspore stage. The anthers were collected about 5 days before heading (5 DBH). Isoelectric focusing in the first dimension was carried out on a linear pH 4-7, 18 cm IPG strip loaded with 100 μg of anther protein. In the second dimension a 12-14% SDS-PAGE gel was used. Proteins were visualised by silver staining. The protein spots identified by PMF analysis are circled and the arbitrary identifiers used correspond to those listed in Table 4.1. The molecular weights of the protein standards are denoted on the left (kDa) and the isoelectric points are denoted at the bottom.

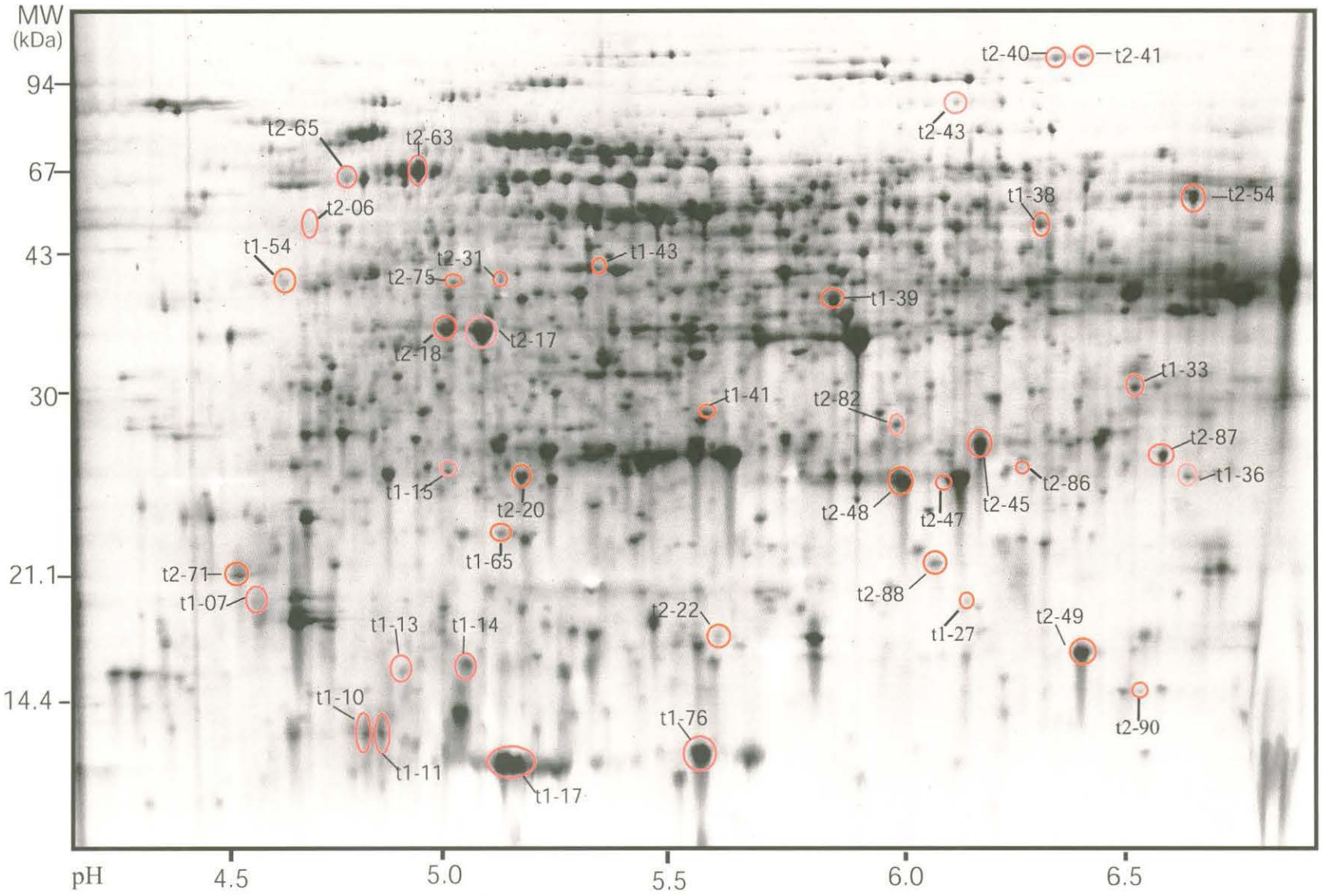

MW
(kDa)
94
67
43
30
21.1
14.4
pH
4.5
5.0
5.5
6.0
6.5
t2-40
t2-41
t2-43
t2-65
t2-63
t2-06
t1-38
t2-54
t1-54
t2-31
t1-43
t2-75
t1-39
t2-18
t2-17
t1-33
t1-41
t2-82
t2-87
t1-36
t1-15
t2-20
t2-86
t2-48
t2-47
t2-45
t1-65
t2-71
t1-07
t2-88
t2-22
t1-27
t2-49
t1-13
t1-14
t1-10
t1-76
t2-90
t1-11
t1-17

Figure 4.5. 2-DE proteome map of anther proteins from rice cultivar Doongara at the late binucleate (LBN) microspore stage. The anthers were collected about 3 days before heading (3DBH). Isoelectric focusing in the first dimension was carried out on a linear pH 4-7, 18 cm IPG strip loaded with 100 μg of anther protein. In the second dimension a 12-14% SDS-PAGE gel was used. Proteins were visualised by silver staining. The protein spots identified by PMF analysis are circled and the arbitrary identifiers used correspond to those listed in Table 4.1. The molecular weights of the protein standards are denoted on the left (kDa) and the isoelectric points are denoted at the bottom.

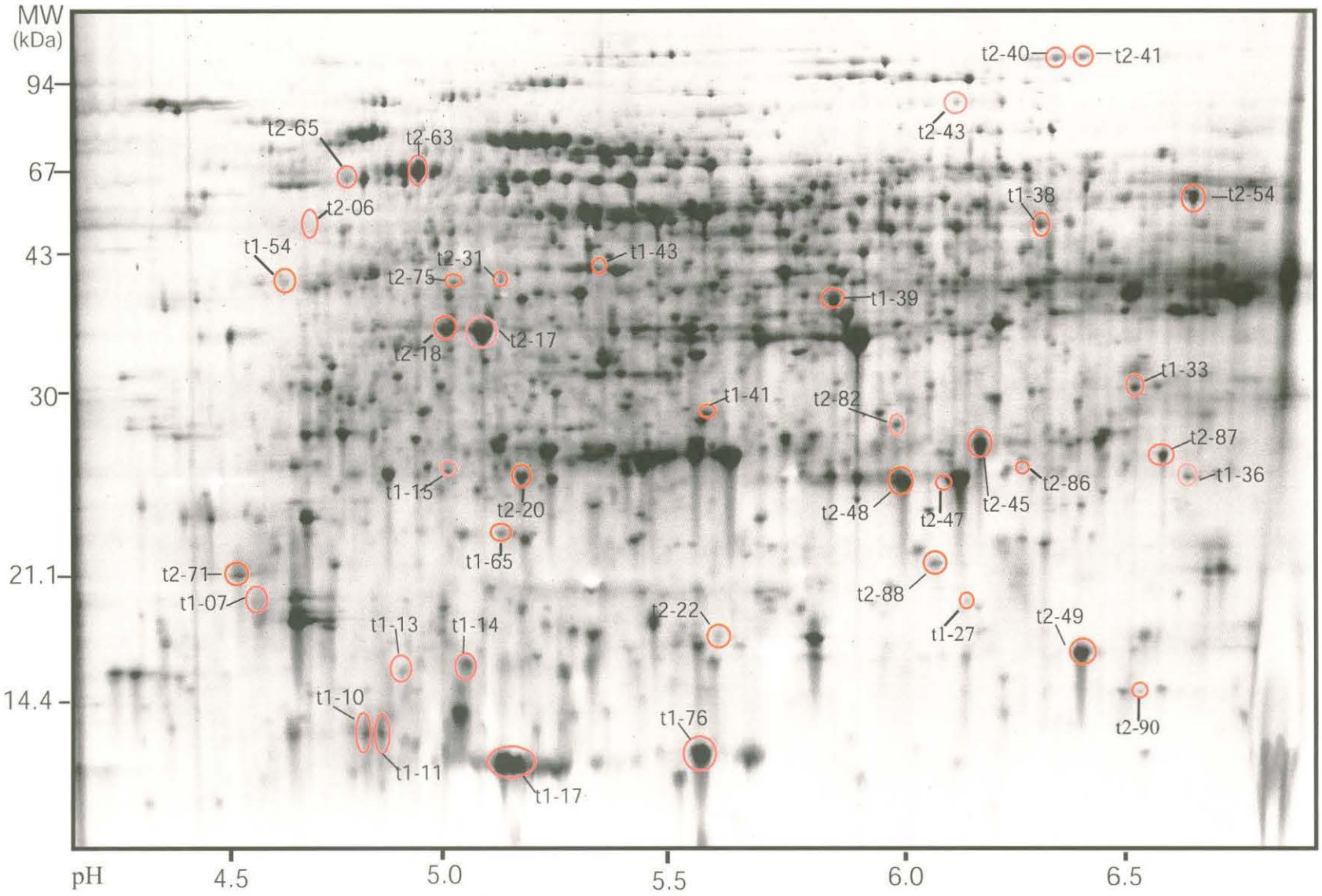
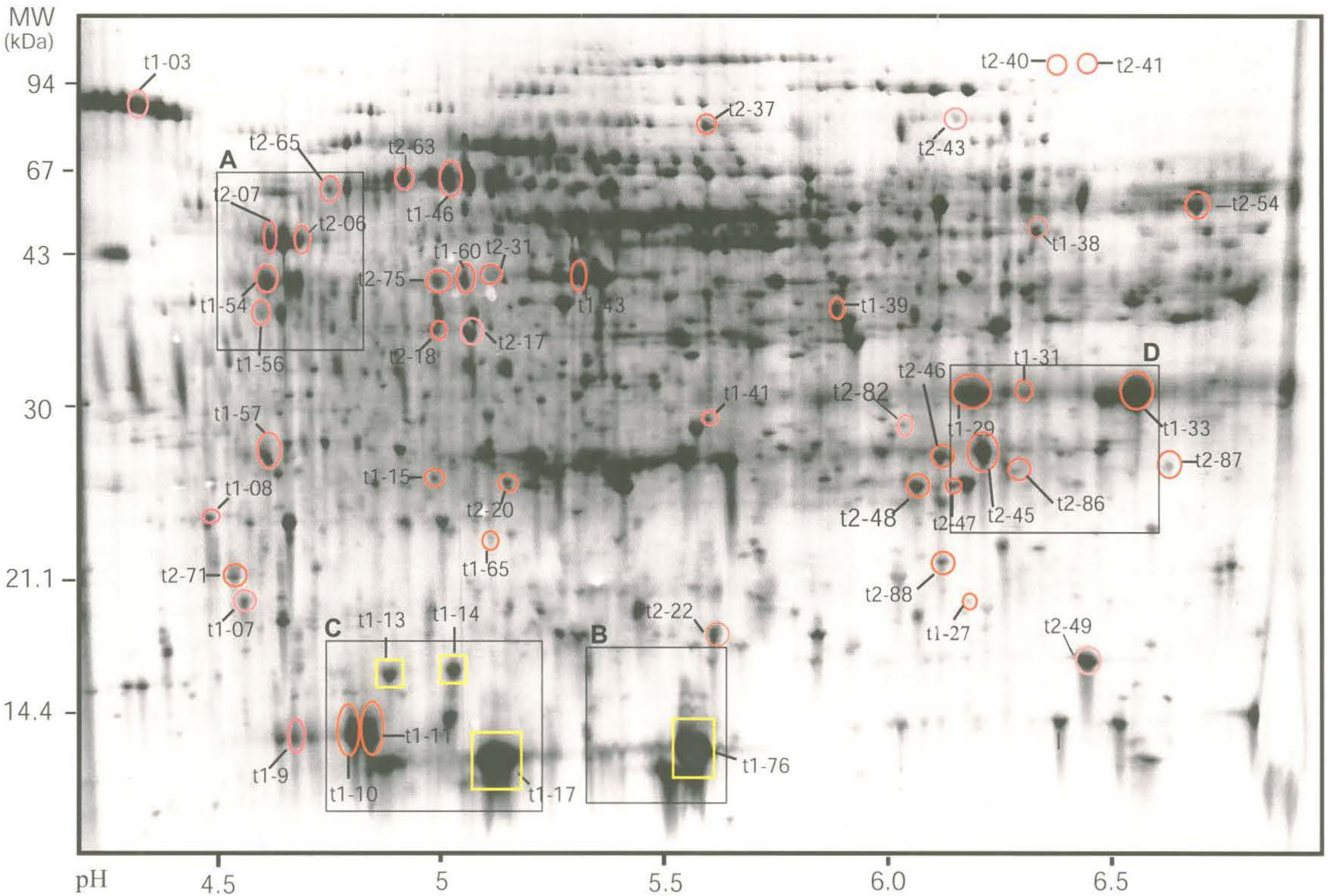

MW
(kDa)
94
67
43
30
21.1
14.4
pH
4.5
5.0
5.5
6.0
6.5
t2-40
t2-41
t2-43
t2-65
t2-63
t2-06
t1-38
t2-54
t1-54
t2-31
t1-43
t2-75
t1-39
t2-18
t2-17
t1-33
t1-41
t2-82
t2-87
t1-36
t1-15
t2-20
t2-86
t2-48
t2-47
t2-45
t1-65
t2-71
t1-07
t2-88
t2-22
t1-27
t2-49
t1-13
t1-14
t1-10
t1-76
t2-90
t1-11
t1-17

Figure 4.6. 2-DE proteome map of anther proteins from rice cultivar Doongara at the heading (H) stage. The anthers were collected shortly after the panicle had emerged from flag leaf sheath. Isoelectric focusing in the first dimension was carried out on a linear pH 4-7, 18 cm IPG strip loaded with 100 μg of anther protein. In the second dimension a 12-14% SDS-PAGE gel was used. Proteins were visualised by silver staining. The protein spots identified by PMF analysis are circled and the arbitrary identifiers used correspond to those listed in Table 4.1. The protein spots which were subjected to N-terminal sequencing were boxed in yellow rectangles and the arbitrary identifiers used correspond to those in Table 4.2. The enlargements of the framed gel regions (**A,B,C** and **D)** are shown in Figures 4.14 and 4.15 The molecular weights of the protein standards are denoted on the left (kDa) and the isoelectric points are denoted at the bottom.



MW
(kDa)
94
67
43
30
21.1
14.4
pH
4.5
5
5.5
6.0
6.5
A
B
C
D
t1-03
t2-65
t2-63
t2-07
t2-06
t1-46
t2-37
t2-43
t2-40
t2-41
t2-54
t1-38
t1-60
t2-31
t2-75
t1-54
t1-56
t1-43
t1-39
t2-18
t2-17
t1-41
t2-82
t2-46
t1-31
t1-29
t1-33
t2-87
t1-57
t1-15
t2-20
t1-08
t2-48
t2-47
t2-45
t2-86
t1-65
t2-71
t1-07
t2-88
t1-27
t2-22
t1-13
t1-14
t2-49
t1-9
t1-10
t1-11
t1-17
t1-76

Figure 4.7. 2-DE proteome map (pH 6-11) of basic anther proteins from rice cultivar Doongara at the pollen mother cell (PMC) stage. Isoelectric focusing in the first dimension was carried out on a linear pH 6-11, 18 cm IPG strip loaded with 100 μg of anther protein. In the second dimension a 12-14% SDS-PAGE gel was used. Proteins were visualised by silver staining. The gel spots excised for analysis are circled and the arbitrary identifiers used correspond to those listed in Table 4.1. The molecular weights of the protein standards are denoted on the left (kDa) and the isoelectric points are denoted at the bottom.

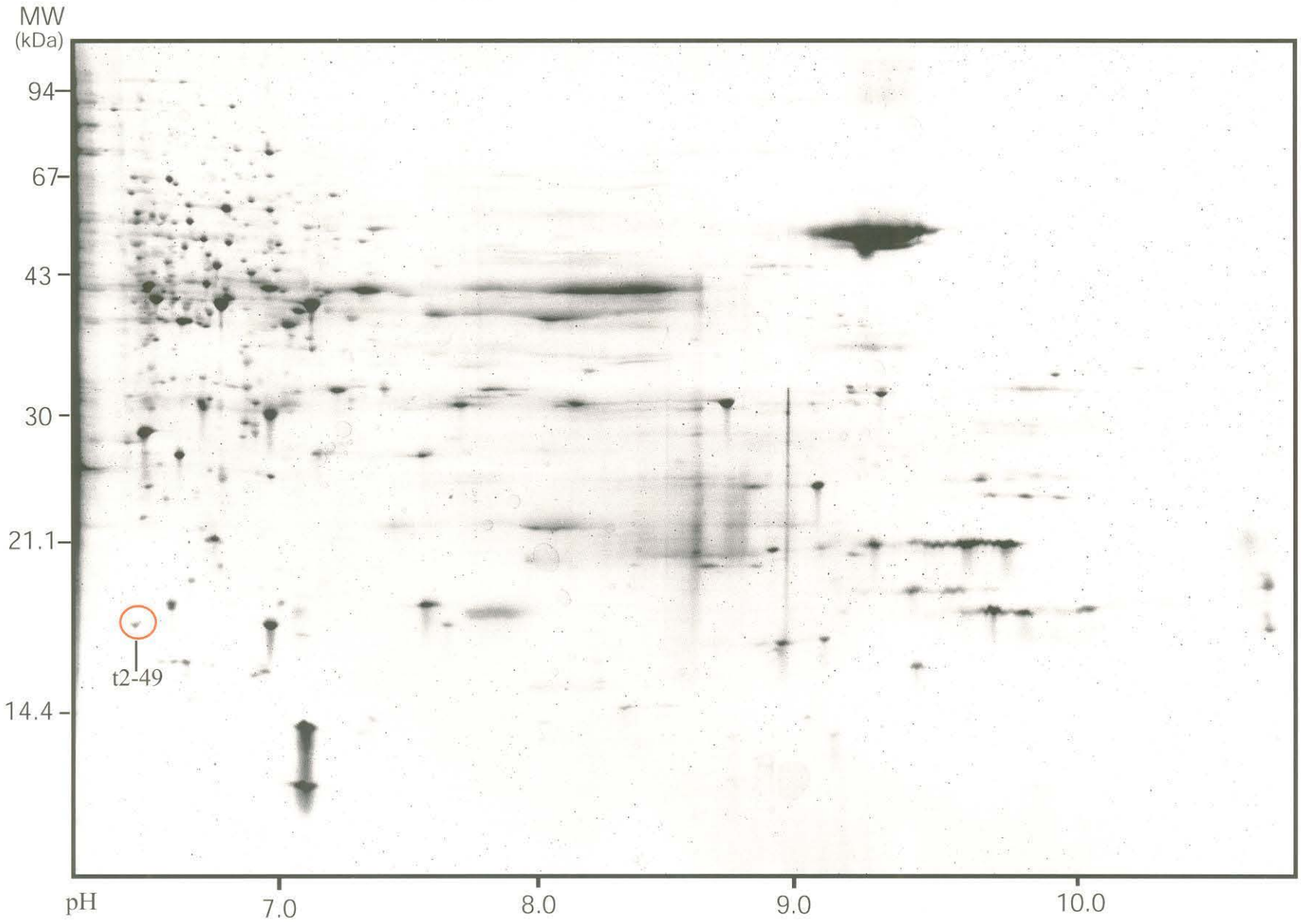

MW
(kDa)
94
67
43
30
21.1
14.4
t2-49
pH
7.0
8.0
9.0
10.0

Figure 4.8. 2-DE proteome map (pH 6-11) of basic anther proteins from rice cultivar Doongara at the early young microspore (EYMS) stage. The anthers were collected about 11 days before heading (11 DBH). Isoelectric focusing in the first dimension was carried out on a linear pH 6-11, 18 cm IPG strip loaded with 100 μg of anther protein. In the second dimension a 12-14% SDS-PAGE gel was used. Proteins were visualised by silver staining. The protein spots identified by PMF analysis are circled and the arbitrary identifiers used correspond to those listed in Table 4.1. The molecular weights of the protein standards are denoted on the left (kDa) and the isoelectric points are denoted at the bottom.

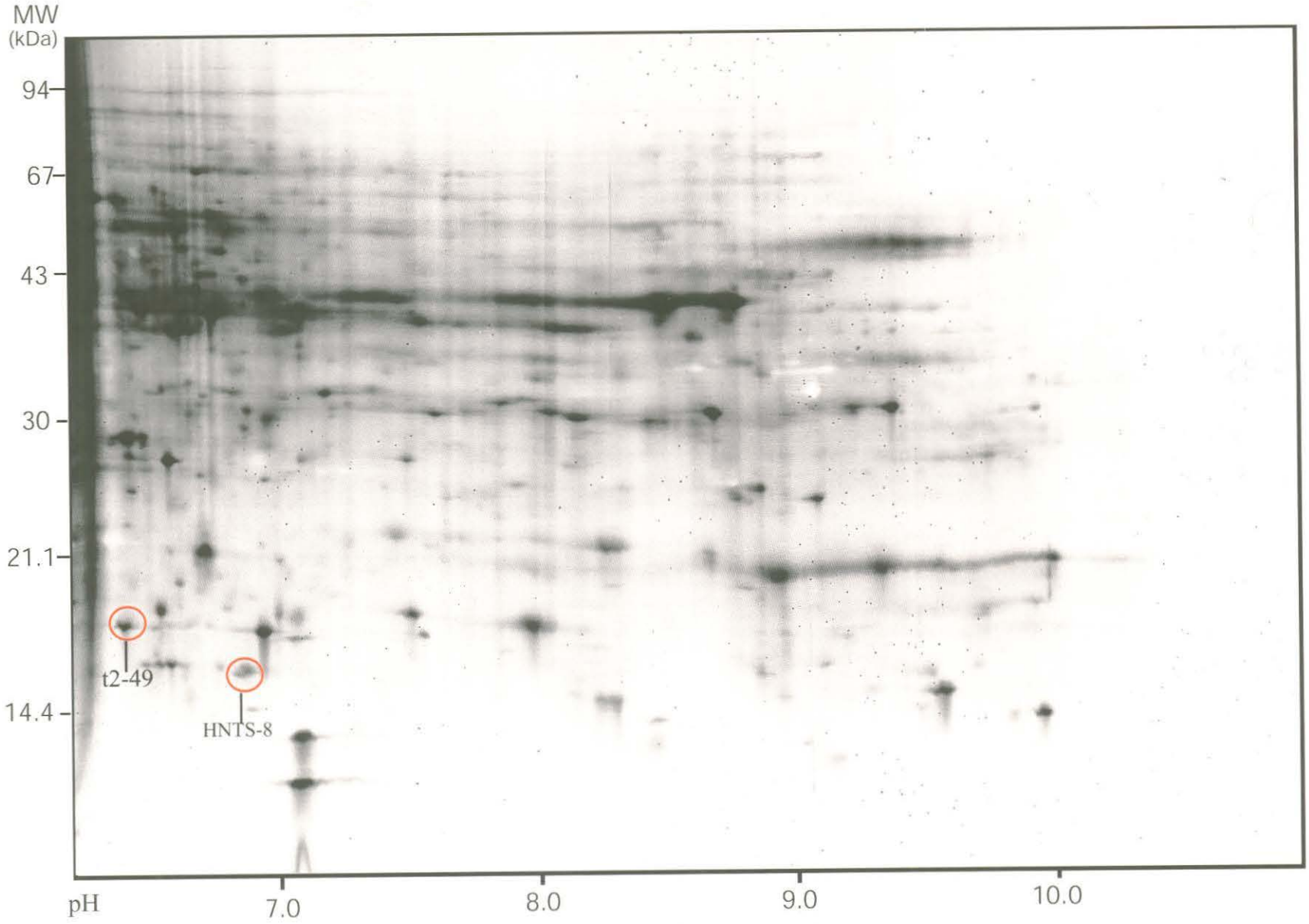

MW
(kDa)
94
67
43
30
21.1
14.4
t2-49
HNTS-8
pH
7.0
8.0
9.0
10.0

Figure 4.9. 2-DE proteome map (pH 6-11) of basic anther proteins from rice cultivar Doongara at the middle young microspore (MYMS) stage. The anthers were collected about 8 days before heading (8 DBH). Isoelectric focusing in the first dimension was carried out on a linear pH 6-11, 18 cm IPG strip loaded with 100 μg of anther protein. In the second dimension a 12-14% SDS-PAGE gel was used. Proteins were visualised by silver staining. The protein spots identified by PMF analysis are circled and the arbitrary identifiers used correspond to those listed in Table 4.1. The molecular weights of the protein standards are denoted on the left (kDa) and the isoelectric points are denoted at the bottom.

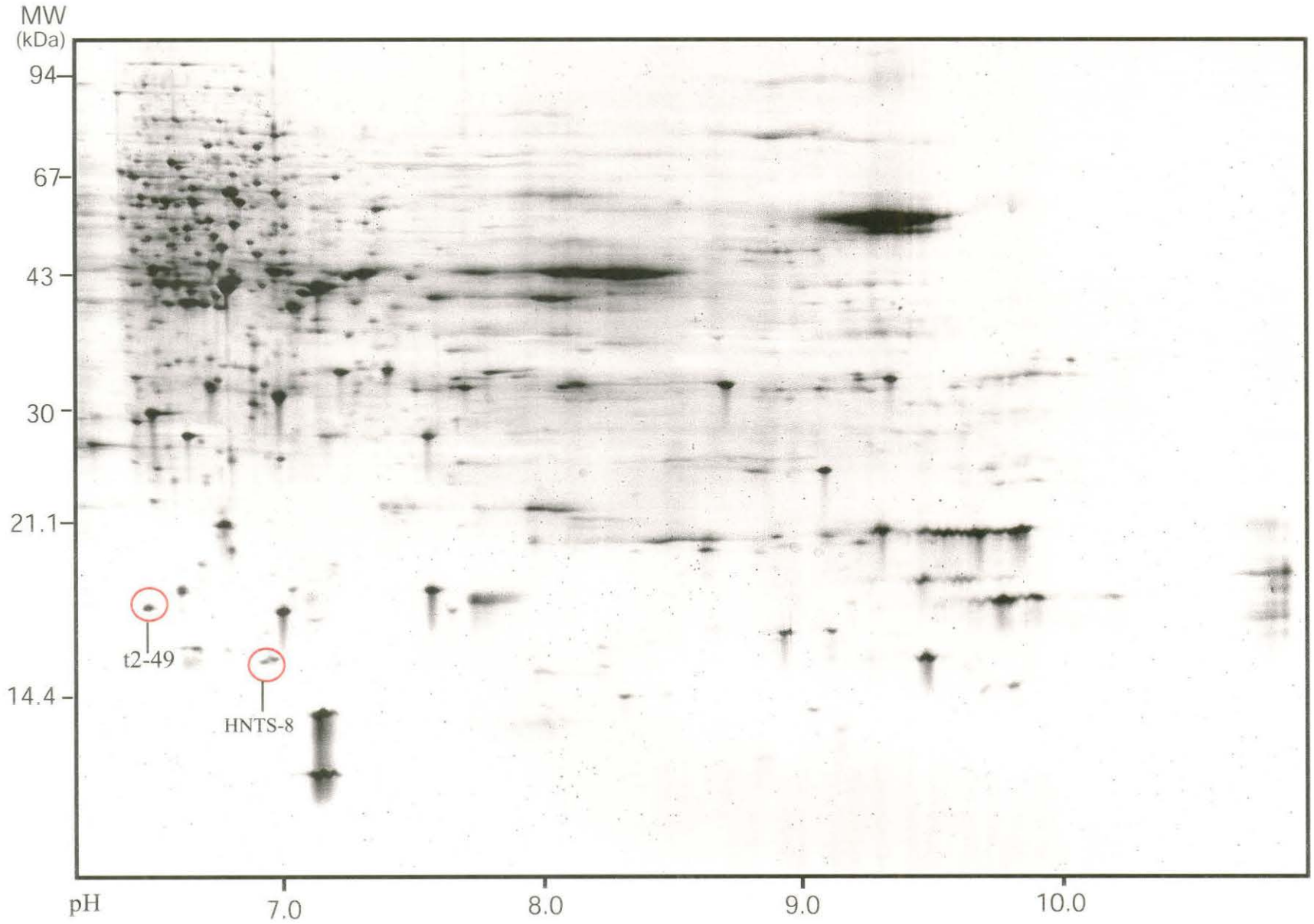

MW
(kDa)
94
67
43
30
21.1
14.4
t2-49
HNTS-8
pH
7.0
8.0
9.0
10.0

Figure 4.10. 2-DE proteome map (pH 6-11) of basic anther proteins from rice cultivar Doongara at the early binucleate (EBN) microspore stage. The anthers were collected about 5 days before heading (5 DBH). Isoelectric focusing in the first dimension was carried out on a linear pH 6-11, 18 cm IPG strip loaded with 100 μg of anther protein. In the second dimension a 12-14% SDS-PAGE gel was used. Proteins were visualised by silver staining. The protein spots identified by PMF analysis are circled and the arbitrary identifiers used correspond to those listed in Table 4.1. The molecular weights of the protein standards are denoted on the left (kDa) and the isoelectric points are denoted at the bottom.

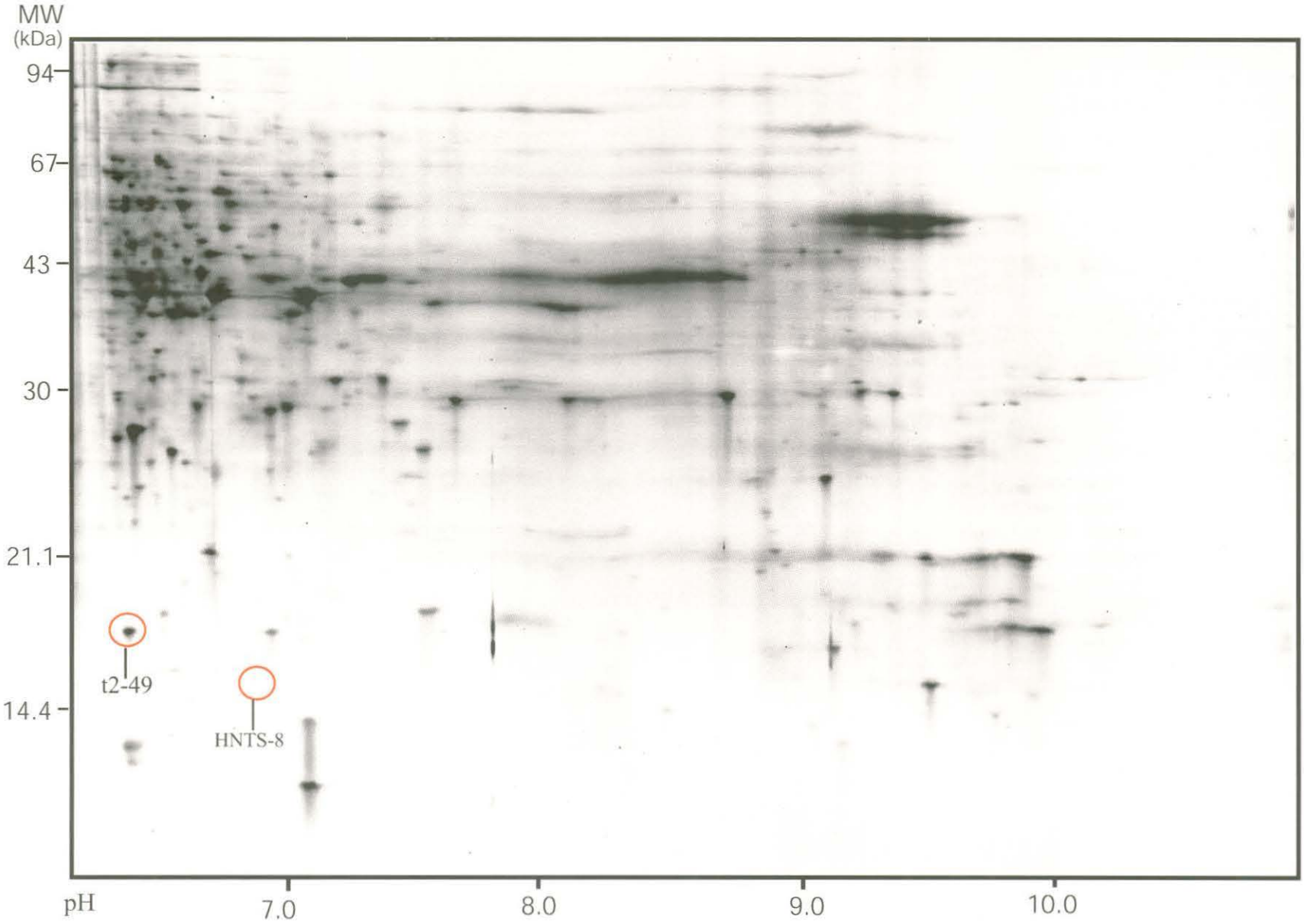

MW
(kDa)
94
67
43
30
21.1
14.4
t2-49
HNTS-8
pH
7.0
8.0
9.0
10.0

Figure 4.11. 2-DE proteome map (pH 6-11) of basic anther proteins from rice cultivar Doongara at the late binucleate (LBN) microspore stage. The anthers were collected about 3 days before heading (3DBH). Isoelectric focusing in the first dimension was carried out on a linear pH 6-11, 18 cm IPG strip loaded with 100 μg of anther protein. In the second dimension a 12-14% SDS-PAGE gel was used. Proteins were visualised by silver staining. The protein spots identified by PMF analysis are circled and the arbitrary identifiers used correspond to those listed in Table 4.1. The molecular weights of the protein standards are denoted on the left (kDa) and the isoelectric points are denoted at the bottom.

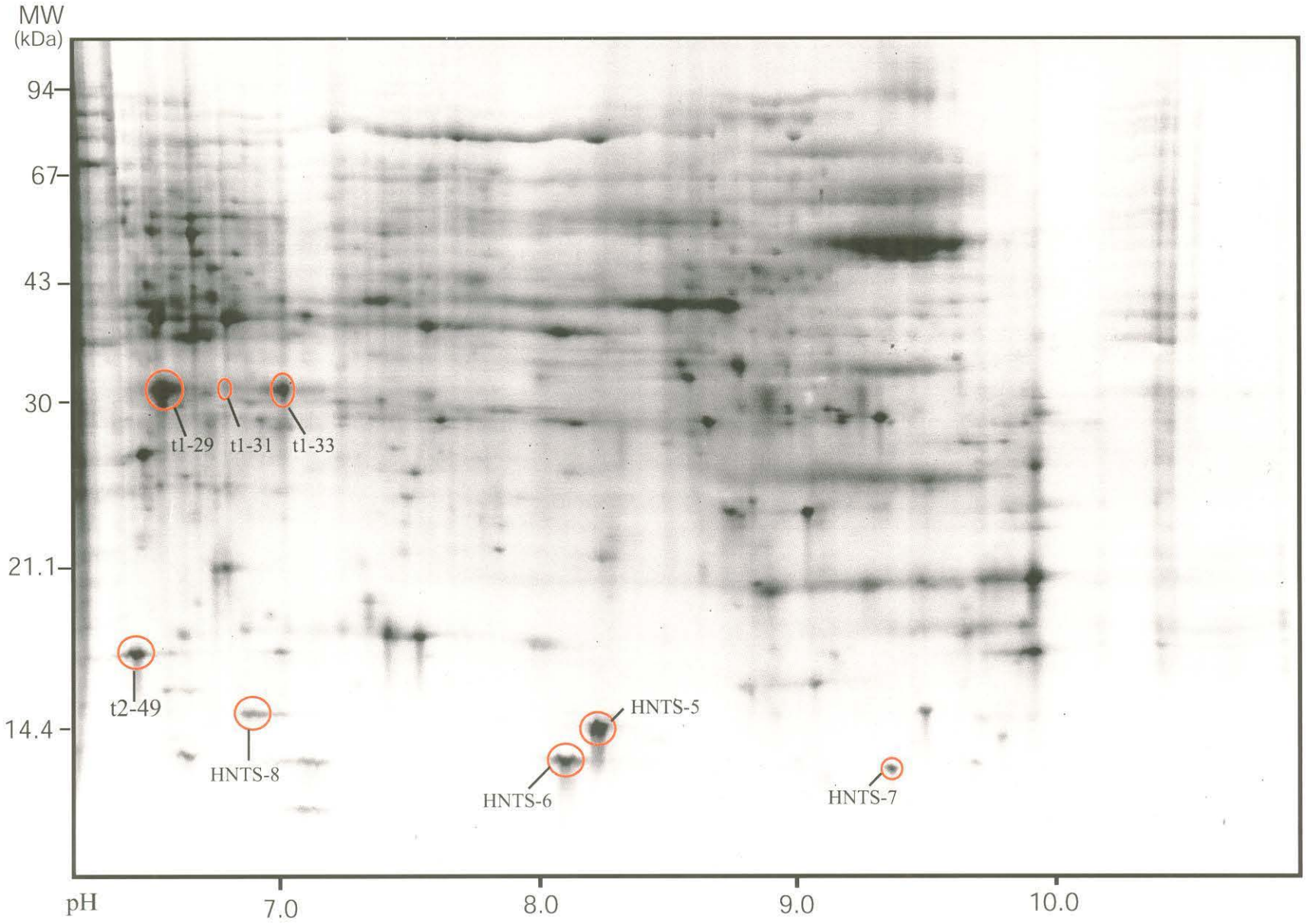

MW
(kDa)
94
67
43
30
21.1
14.4
t1-29
t1-31
t1-33
t2-49
HNTS-8
HNTS-5
HNTS-6
HNTS-7
pH
7.0
8.0
9.0
10.0

Figure 4.12. 2-DE proteome map (pH 6-11) of basic anther proteins from rice cultivar Doongara at the heading (H) stage. The anthers were collected shortly after the panicle had emerged from the flag leaf sheath. Isoelectric focusing in the first dimension was carried out on a linear pH 6-11, 18 cm IPG strip loaded with 100 μg of anther protein. In the second dimension a 12-14% SDS-PAGE gel was used. Proteins were visualised by silver staining. The protein spots identified by PMF analysis are circled and the arbitrary identifiers used correspond to those listed in Table 4.1. The protein spots which were subjected to *N*-terminal sequencing were boxed in yellow rectangles and the arbitrary identifiers used correspond to those in Table 4.2. The molecular weights of the protein standards are denoted on the left (kDa) and the isoelectric points are denoted at the bottom.

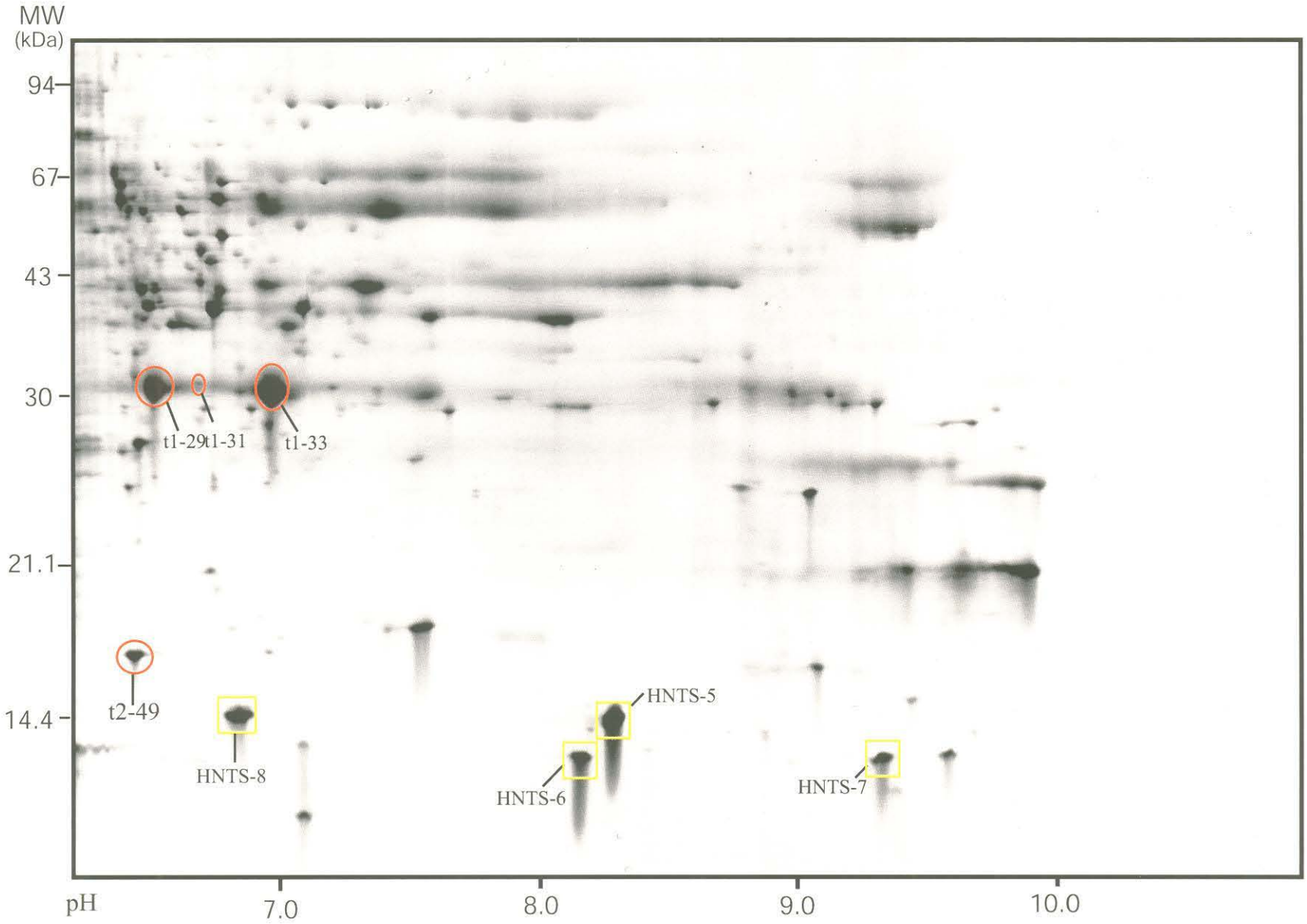

MW
(kDa)
94
67
43
30
21.1
14.4
t1-29
t1-31
t1-33
t2-49
HNTS-8
HNTS-6
HNTS-5
HNTS-7
pH
7.0
8.0
9.0
10.0

Figure 4.13. An example of protein identification by *N*-terminal sequencing and database search. **A-** Nucleotide sequence of an open reading frame of TC115250, to which the *N*-terminal amino acid sequence of the protein spot t1-13 was matched, is shown. The deduced amino acid sequence is indicated by the single amino acid letter code below the corresponding nucleotide sequence. The experimental *N*-terminal tag in the deduced protein sequence is denoted in bold blue letters. **B-** Graphical representation of the signal peptide prediction analysis using the deduced amino acid sequence. The Signal IP program predicted a signal peptide cleavage site between amino acid 21 and 22 with a mean signal peptide score of 0.912.

A.

```
atggcccagaacaagaccattgcggtggccctgctccttgccacccttgtggcggtcatg
 M  A  Q  N  K  T  I  A  V  A  L  L  L  A  T  L  V  A  V  M
ggaaaggagccagagacgctagaggaggcagtgagagctgggtgcaaagaggaatgcagt
 G  K  E  P  E  T  L  E  E  A  V  R  A  G  C  K  E  E  C  S
gagcagaaaaagaaagcccccatcgacgagaaacaatgtgaagacttctgcttcatcaag
 E  Q  K  K  K  A  P  I  D  E  K  Q  C  E  D  F  C  F  I  K
accaagtctatatttgaggcccataagggagtgaaggatctaaaggcggatagattcata
 T  K  S  I  F  E  A  H  K  G  V  K  D  L  K  A  D  R  F  I
gatttttgcaacaatgagtgtaatgcagtatataaggaggatcccgccacaagcaagaag
 D  F  C  N  N  E  C  N  A  V  Y  K  E  D  P  A  T  S  K  K
tgtgcagaaagttgcgaagctgatgctaaggaggccgaggtattccttgataaagtcgtc
 C  A  E  S  C  E  A  D  A  K  E  A  E  V  F  L  D  K  V  V
gcttatattcaaacaacgaaacaagcatg
 A  Y  I  Q  T  T  K  Q  A
```

B.

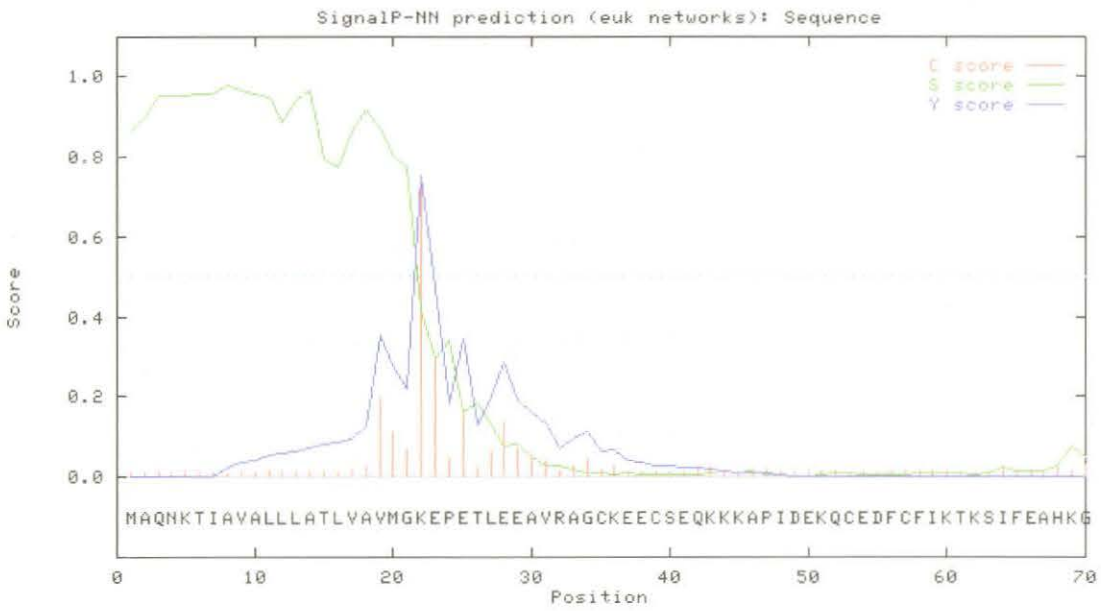


Figure 4.14. Enlargements of the corresponding gel regions that were framed in Figure 4.6 (frame **A** and **B**). The same areas in Figures 4.1 to 4.6 are enlarged. The red arrows indicate the locations of some differentially displayed proteins. The arbitrary identifiers correspond to those in Table 4.1 and Table 4.2. **PMC**-pollen mother cell stage; **EYMS**-early young microspore stage; **MYMS**-middle young microspore stage; **EBN**-early binucleate microspore stage, **LBN**- late binucleate microspore stage; **H**- heading stage.

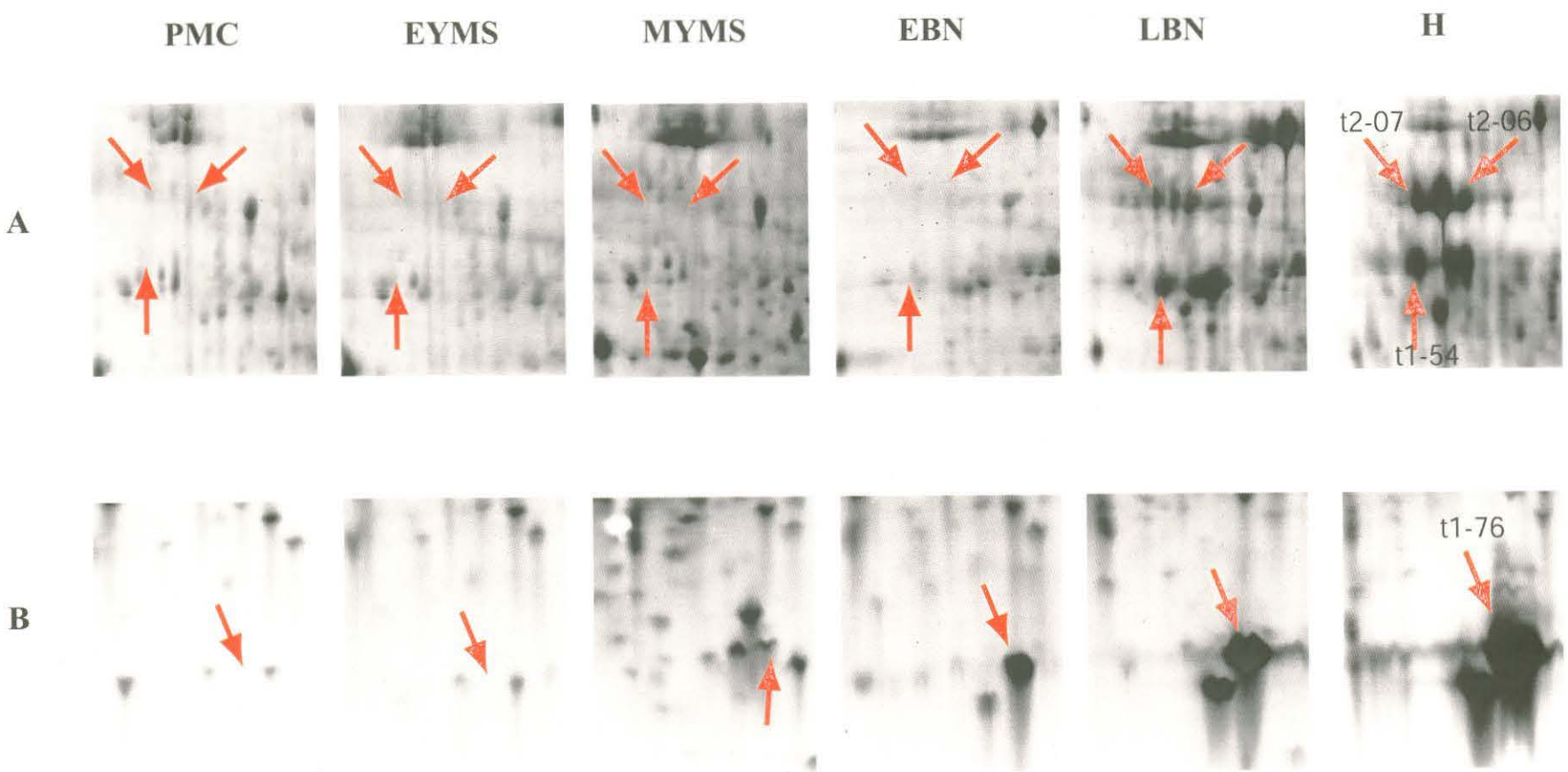

PMC
EYMS
MYMS
EBN
LBN
H
A
B
t2-07
t2-06
t1-54
t1-76

Figure 4.15. Enlargements of the corresponding gel regions that were framed in Figure 4.6 (frame **C** and **D**). The same areas of Figures 4.1 to 4.6 are enlarged. The arrows indicate the locations of some differentially displayed proteins. The arbitrary identifiers correspond to those in Table 4.1 and Table 4.2. **PMC**-pollen mother cell stage; **EYMS**-early young microspore stage; **MYMS**-middle young microspore stage; **EBN**-early binucleate microspore stage, **LBN**- late binucleate microspore stage; **H**- heading stage.

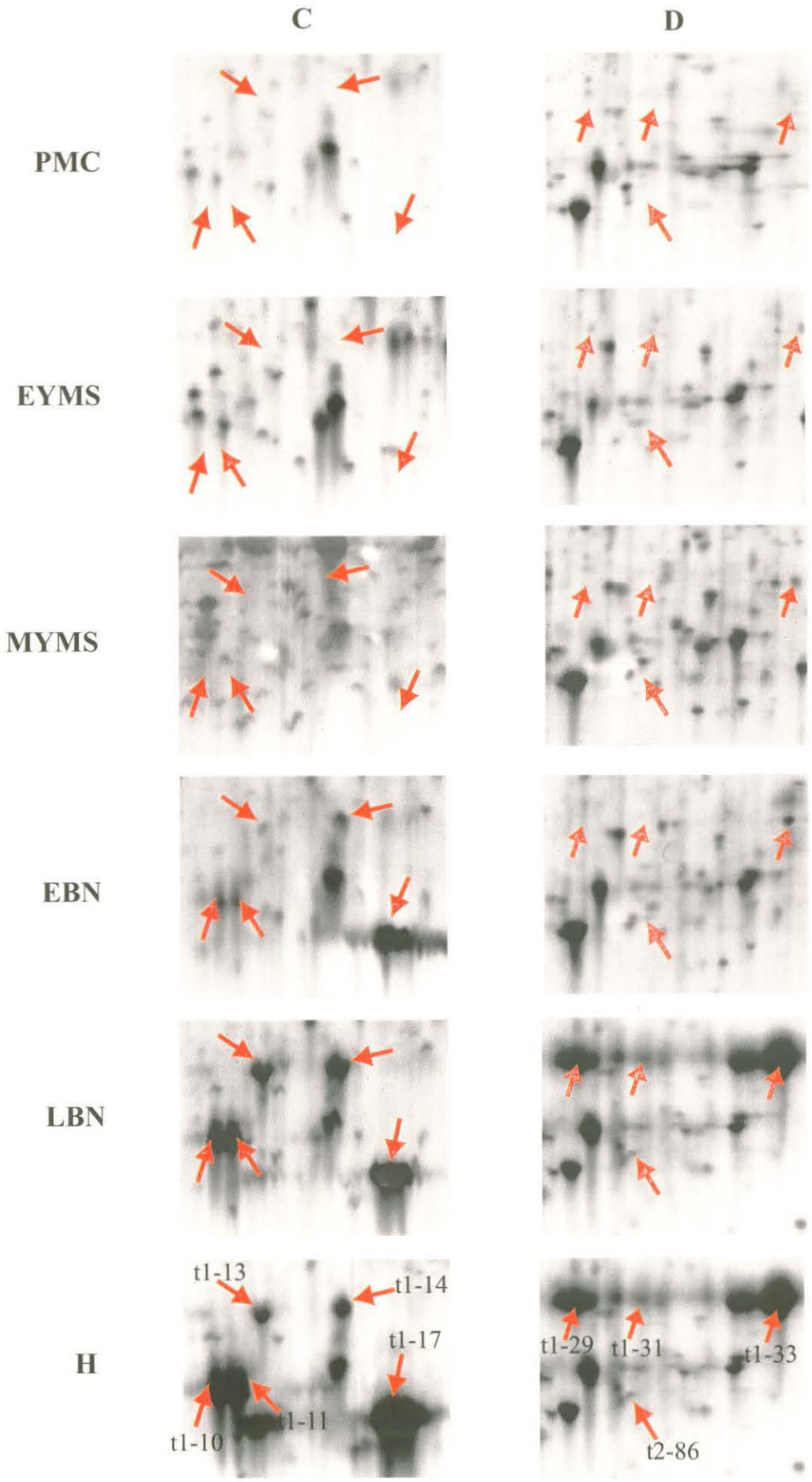

C
D
PMC
EYMS
MYMS
EBN
LBN
H
t1-13
t1-14
t1-17
t1-11
t1-10
t1-29
t1-31
t1-33
t2-86

Table 4.1. Results of protein identification by peptide mass fingerprinting analysis

Spot no.	Protein match to	Organism	Access. No.	Peptide number	Coverage	MW/pI [a)]	Spot regulation[b)]					
							PMC	EYM	MYMS	EBN	LBN	H
t2-06	Vacuolar acid invertase	*Oryza sativa*	AAF87245	8	16.87%	71.3/5.1	-	-	-	+	+	+++
t2-07	Vacuolar acid invertase	*Oryza sativa*	AAF87245	10	21.00%	71.3/5.1	-	-	-	-	+	+++
t2-17	Putative fructokinase II	*Oryza sativa*	AAL26573	14	49.00%	35.9/5.0	+	+	+	++	+++	+
t2-18	Putative fructokinase II	*Oryza sativa*	AAL26573	17	49.00%	35.9/5.0	+	+	+	++	+++	+
t2-20	Submergence induced protein 2A	*Oryza sativa*	AAC19375	6	37.00%	23.65/5.1	++	++	++	++	++	+
t2-22	Hypothetical protein	*Oryza sativa*	BAA87854.1	5	40.68%	N/A	+	+	+	+	+	++
t2-27	Hypothetical protein	*Oryza sativa*	J013000M11	5	26.00%	21/5.4	-	+	+	+	++	+
t2-31	Jab1 protein	*Oryza sativa*	AAC33765	9	38.61%	40.0/5.2	+	+	+	+	+	++
t2-37	Putative reductase	*Oryza sativa*	AAL58200	15	34.00%	82.18/5.9	-	-	-	-	-	+
t2-40	Putative beta-galactosidase	*Oryza sativa*	AAL31090	14	24.00%	92.72/6.0	-	-	-	+	++	-
t2-41	Putative beta-galactosidase	*Oryza sativa*	AAL31090	14	24.00%	92.72/6.0	-	-	-	+	++	+
t2-43	Poll-like DNA polymerase	*Oryza sativa*	J013009A21	8	17.00%	59.7/6.8	-	-	-	+	++	+
t2-45	Glutathione S-transferase II	*Oryza sativa*	AAC64007	5	28.84%	23.9/6	+	+	+	+	++	++
t2-46	Putative soluble inorganic pyrophosphatase	*Oryza sativa*	AAK98675	5	32.00%	22.2/5.7	-	-	-	-	+	+
t2-47	putative quinone oxidoreductase	*Oryza sativa*	J023054M12	7	50.00%	21.5/6.0	++	+	+	+	+	++

Table 4.1. Continued

Spot no.	Protein match to	Organism	Access. No.	Peptide number	Coverage	MW/pI [a)]	Spot regulation[b)]					
							PMC	EYM	MYMS	EBN	LBN	H
t2-48	beta 1 subunit of 20S proteasome	*Oryza sativa*	BAA96834	9	43.90%	26.2/5.7	++	++	+	++	++	+
t2-49	Nucleoside diphosphate kinase	*Oryza sativa*	Q07661	4	32.89%	16.8/6.30	+	+	+	++	++	++
t2-54	Putative mitochondrial processing peptidase alpha-II chain precursor	*Oryza sativa*	BAA99436	10	23.39%	90.2/8.7	+	+	+	+	++	++
t2-63	Protein disulfide isomerase	*Oryza sativa*	BAA92322	13	52.35%	33.4/4.8	+	++	++	++	+	+
t2-65	Hypothetical protein	*Oryza sativa*	001-123-B09	8	13.00%	65.9/5.3	-	+	+	+	++	+
t2-71	Translationally controlled tumour protein homologue	*Oryza sativa*	P35681	8	58.33%	18.9/4.5	+	+	+	++	++	+
t2-75	ubiquitin	*Oryza sativa*	J013126L13	9	28.00%	45.6/5.0	+	+	+	+	+	+++
t2-82	Beta-ketoacyl-ACP synthase	*Hordeum vulgare*	CAA84022	5	22.63%	53.1	+	+	++	+	+	+
t2-86	GSH-dependent dehydroascorbate reductase1	*Oryza sativa*	BAA90672	7	53.99%	23.5/5.9	+	+	+	+	++	+
t2-87	Putative glutathione S-transferase	*Oryza sativa*	AAG32476	4	30.77%	24.1/6.6	++	++	++	++	+	+
t2-88	Unknown protein	*Oryza sativa*	BAA90363	5	36.87%	20.9/8.3.	+	+	+	+	++	+
t2-90	putative translational inhibitor protein	*Oryza sativa*	006-304-F08	5	38.00%	18.8/9.0	+	+	++	+	+	-
t1-3	Hypothetical protein	*Oryza sativa*	TC50154	4	10.78%	85.9	-	-	-	-	+	+

Table 4.1. Continued

Spot no.	Protein match to	Organism	Access. No.	Peptide number	Coverage	MW/pI [a)]	Spot regulation[b)]					
							PMC	EYM	MYMS	EBN	LBN	H
t1-07	cold shock protein-like protein	*Oryza sativa*	J033050L02	4	40.00%	18.6/6.2	-	-	+	+	+	+
t1-08	calmodulin-like protein	*Oryza sativa*	J023089J12	7	32.00%	33.1/6.1	-	-	-	-	-	+
t1-9	Profilin	*Oryza sativa*	Q9FUD1	3	44.00%	14.24/4.9	-	-	-	-	+	++
t1-10	Profilin	*Oryza sativa*	Q9FUD1	3	44.00%	14.24/4.9	-	-	+	+	++	+++
t1-11	Profilin	*Oryza sativa*	Q9FUD1	3	44.00%	14.24/4.9	-	-	-	+	++	+++
t1-15	Lycopersicon esculentum resistance complex protein I2C-2 like protein	*Oryza sativa*	BAB07969.1	11	12.15%	164.6	+	+	+	++	++	++
t1-29	Beta expansin	*Oryza sativa*	AAF72983.1	5	29.00%	29.12/6.4	-	-	-	-	+	+++
t1-31	Beta expansin	*Oryza sativa*	AAF72983.1	4	15.00%	29.12/6.4	-	-	-	-	+	+
t1-33	Beta expansin	*Oryza sativa*	AAF72983.1	8	32.00%	29.12/6.4	-	-	-	+	+	+++
t1-36	Putative cholinephosphate cytidylyltransferase	*Oryza sativa*	TC49579	4	25.76%	37.5	-	-	+	+	+	-
t1-38	putative aminotransferase	*Oryza sativa*	J033090A04	10	22.00%	74.3/6.1	+	+	+	+	++	++
t1-39	Putative legumin-like protein	*Oryza sativa*	BAB63836.1	8	42.18%	40.2/5.6	-	-	-	+	++	+
t1-41	putative 6 - phosphoglucono lactonase	*Oryza sativa*	J023028P13	9	39.00%	29/5.4	-	-	-	-	-	+
t1-43	Actin	*Oryza sativa*	P13362	6	24.00%	42.18/5.3	-	+	+	++	+++	+++
t1-46	Hypothetical protein	*Oryza sativa*	J033048H06	6	14.00%	74.7/5.7	-	-	-	-	-	+
t1-54	Hypothetical protein	*Oryza sativa*	BAB40057.1	8	22.44%	56.2	+	+	+	+	++	++++
t1-56	peroxidase	*Oryza sativa*	J023065P13	4	20.00%	38.7/7.5	-	-	-	-	-	+

Table 4.1. Continued

Spot no.	Protein match to	Organism	Access. No.	Peptide number	Coverage	MW/pI [a)]	Spot regulation[b)]					
							PMC	EYM	MYMS	EBN	LBN	H
t1-57	Unknown protein	*Oryza sativa*	BAA96146	6	46.33%	23.59/4.7	-	-	-	-	-	+
t1-60	Unknown protein	*Oryza sativa*	BAB39912.1	5	12.48%	67.6	-	-	-	-	-	+
t1-65	deoxyuridine triphosphatase	*Oryza sativa*	001-120-F09	6	43.00%	24.0/6	+++	++	++	++	++	+
t1-74	Lipase-like protein	*Oryza sativa*	CAC39051.1	6	21.32%	58.8	+	+	-	-	-	-

a) molecular weight and pI of matched protein

b) - indicates that the protein spot is absent, + indicates that the spot is present, ++ indicates more than two fold increase, +++ indicates more than five fold increase and ++++ indicates more than ten fold increase.
PMF- pollen mother cell, **EYMS** - early young microspore, **MYMS** - middle young microspore, **EBN** - early binucleate **LBN** - late binucleate, **H** - heading

any significant PMF homology to proteins currently available in the three databases that were used in this study.

The 49 matching database entries represent 42 unique proteins. Some of the identified proteins matched to the same database entries, indicating that they are orthologous isoforms. These proteins are two isoforms of vacuolar acid invertase (spots t2-06 and t2-07), two isoforms of fructokinase II (spots t2-17 and t2-18), two isoforms of beta-galactosidase (spots t2-40 and t2-41), three isoforms of profilin (spots t1-09, t1-10 and t1-11) and three isoforms of beta expansin (t1-29, t1-31 and t1-33). Except for spot t2-82, which was matched to a protein sequence of *Hordeum vulgare*, all the other matching entries are of rice origins. A total of nine matching entries are hypothetical or unknown rice proteins, and the deduced amino acid sequences of these proteins did not display any significant homology to proteins in the nonredundant protein database (SWISSPROT and TREMBL) by BLAST protein homology searching. The identified proteins are mainly involved in carbohydrate metabolism, lipid metabolism, signal transduction, cytoskeleton and cell wall formation, nucleotide and amino acid metabolism and stress responses.

4.3.4 Protein identification by *N*-terminal sequencing and database searching

In order to establish the identities of some protein spots whose expression levels changed significantly at the late anther developmental stages and could not be identified by PMF analysis, eight protein spots were selected for further analysis. These proteins are marked in yellow rectangles and assigned arbitrary identifiers as shown in Figures 4.6 and 4.12. The *N*-terminal amino acid sequences of the selected proteins were analyzed after they had been electroblotted onto PVDF membrane and visualized with Coomassie brilliant blue. A sequence data of 15 amino acid residues was obtained for each of the eight proteins with initial sample peak yields between 0.5 to 1.0 picomole. The *N*-terminal microsequencing analysis of the protein spots t-76, t1-13, HNTS-5, HNTS-6, HNTS-7 and HNTS-8 yielded an unequivocal sequence data of 15 amino acid residues.

For the protein spot t1-14, two sets of *N*-terminal sequences were obtained. One set of the sequences is 100% identical to the *N*-terminal sequence of spot t1-13, which possesses the same molecular weight but has a different p*I* value, indicating that spot t1-

14 is possibly a mixture of a charge isoform of spot t1-13 and another protein. A polymorphism was observed at residue 14 for protein spot t1-17, both serine and threonine being detected in a 0.5 to 0.5 ratio. This indicates the presence of two proteins with similar biochemical properties in the position of spot 17 (further discussed in detail in Chapter 5). The N-terminal sequencing analysis revealed that the protein spots HNTS-5 and HNTS-8, which possess the same molecular weight and different p*I*s, have an identical N-terminal amino acid sequence. This indicated that they are possibly the charge isoforms of an unknown protein encoded by TC124876.

The FASTA sequence alignment searches of a nonredundant protein database (SWISSPROT and TREMBL) using the *N*-terminal sequences did not result in any significant matches. Then the tBLASTn (translated BLAST) searches of the six-reading frame translation products of the rice tentative consensus (TC) sequences, which are stored at the TIGR Rice Gene Index database, were conducted. The search results revealed that the *N*-terminal amino acid sequences of the seven proteins exhibited 100% sequence identities to the deduced protein sequences of eight TC sequences in the database. One protein (spot HNTS-7) exhibited 93% (14 in 15 amino acid residues) sequence identity to the protein product that is encoded by an ORF in TC sequence (Table 4.2). In all database matches, the experimental *N*-terminal tags of the eight protein spots analyzed are aligned 21 to 23 amino acid residues downstream from the methionine initiation site of the protein sequences encoded by the corresponding ORFs. In agreement with these, potential signal peptide cleavage sites were detected with high probability, just before the matching *N*-terminal sequences for all eight theoretical translation products, by analyzing the deduced protein sequences using the SignalP program. One example of the database searching results is shown in Figure 4.13. The *N*-terminal sequence of the protein spot t1-13 displayed 100% sequence identity to the frame +3 translation product of TC115250. The sequence alignment showed that the first amino acid residue of the *N*-terminal tag started at residue 22 of the deduced protein sequence. SignalP analysis of the deduced protein sequence identified with high probobality a putative signal peptide cleavage site between residues 21 and 22. Therefore, it is estimated that the protein spot t1-13 is most likely periplasmic or membrane-located.

The primary structures of the deduced protein sequences from the corresponding TC sequences were further analyzed using computer software tools. As shown in Table 4.2,

Table 4.2. Proteins analysed by *N*-terminal microsequencing and Fasta search

Spot I.D.	Mr/pI[a)]	N-terminal Sequence	Homology (% identity)	Organism	MW/pI[b)]	Putative identitiy	TIGR accession number
t1-13	16.5/5.0	KEPETLEEAVRAGCK	100%	*Orysa sativa*	12205/ 5.03	Not assigned	TC115250
t1-14	16.5/4.9	TEEKDIEEAVCSEHC	100%	*Orysa sativa*	11371 / 4.85	Not assigned	TC126007
		KEPETLEEAVRAGCK	100%	*Orysa sativa*	12205/ 5.03	Not assigned	TC115250
t1-76	11/5.6	TEVTFKVGEGSSGKS	100%	*Orysa sativa*	10085 / 5.45	putative group II allergen	TC114263
t1-17	11/5.2	TELTFKVAEGSSASS	100%	*Orysa sativa*	10140 / 5.12	putative group II allergen	TC125017
	11/5.2	TELTFKVAEGSSATS	100%	*Orysa sativa*	10096 / 5.12	putative group II allergen	TC129660
HNTS-5	13.5 / 8.5	DAPVEKSFNKALLAP	100%	*Orysa sativa*	11702 / 6.52	Not assigned	TC124876
HNTS-6	11 / 8.2	TEVTFKIGESSTTST	100%	*Orysa sativa*	10346 / 7.86	Not assigned	TC121540
HNTS-7	11 / 9.5	TELTFKVGPGSSTTS	93%	*Orysa sativa*	10239 / 9.48	Not assigned	TC130372
HNTS-8	13.5 / 8.5	DAPVEKSFNKALLAP	100%	*Orysa sativa*	11702 / 6.52	Not assigned	TC124876

[a)] The observed molecular mass and pI of the protein spots.

[b)] The predicted molecular mass and isoelectric point of the matched peptides calculated from the ORFs after excluding the predicted signal peptide sequences.

the experimentally determined molecular weights and p*I*s of the protein spots t1-17, t1-76, HNTS-6 and HNTS-7 are in close agreement with the calculated molecular weights and p*I*s of the translation products of the matching TC sequences. Up to 4 kDa of differences were observed between the experimental molecular weights of the protein spots t1-13, t1-14, HNTS-5 and HNTS-8 and the predicted molecular weights of the deduced matching proteins. However, the experimental and theoretical p*I*s of these four protein spots are still in close agreement. The discrepancies between the observed and predicted molecular weights of these protein spots indicated that these proteins most likely have undergone some kind of posttranslational modifications which increased the molecular weights of the translation products of the corresponding genes.

A BLAST homology searching of the non-redundant protein databases using the theoretical translation products of the matching TC sequences showed that the spots t1-17, t1-76, HNTS-6 and HNTS-7 display some degree of sequence similarity to the group-II pollen allergens of grasses (further discussed in Chapter 5). No significant homology was found for the protein spots t1-13, t1-14, HNTS-5 and HNTS-8 in the database, indicating that they may be novel proteins.

4.4 DISCUSSION

Differential display with 2-DE has made it possible to array and compare the protein contents of microspores and pollen grains at the different time point of development. In order to resolve and analyse only the male gametophyte specific proteins, we collected and used anthers instead of using the whole flower. The whole flower of rice, or the panicle, contains pollen grains belonging to different developmental stages. In addition, a panicle contains many different types of somatic tissues and the female gametophyte organs. By excluding all the non anther-specific parts of the panicle, the scope of our study was focused only on the anther-specific proteins, and at the same time some anther-specific and low abundant proteins were enriched by this sample pre-fractionation step. The anther samples collected for six discrete developmental stages covered the majority of the male gametophyte developmental process, ranging from the pollen mother cell to the mature pollen stage. The comparison of the 2-DE proteome maps revealed over 150 proteins whose expression levels are developmentally regulated. By integrating information about the regulation and biological function of the

identified proteins, we were able to detect some of the metabolic and signalling pathways that are active in the developing rice anthers.

4.4.1 Ubiquitin / proteosome pathway

In our study we identified the three main regulators of light-mediated signal transduction in plants including the homologues of Jab1 protein, the beta 1 subunit of the 20S proteosome and an ubiquitin-like protein. The protein spot t2-31 was identified as Jab1, or c-Jun activating binding protein, of animals. Jab1 is also known as COP9 (Constitutive Photomorphogenesis 9) signalosome subunit 5 (CSN5) in *Arabidopsis*, which is a core component of the COP9 signalosome regulatory complex (Chamovitz and Segal 2001; Hellmann and Estelle 2002). The COP9 signalosome is composed of eight subunits that are highly conserved between plants and animals. Until recently the original names for the COP9 signalosome subunits were largely unrelated. A unified subunit nomenclature was adopted recently and the eight subunits are named CSN1 to CSN8 (Deng *et al.* 2000). The COP9 signalosome is a pleiotropic regulator that is involved in numerous pathways. In plants, the COP9 signalosome is a master repressor of photomorphogenesis. It can repress the expression of light-induced genes in the absence of light by interacting with 26S proteasome (Karniol and Chamovitz 2000). The presence of the COP9 signalosome is necessary for the 26S proteasome to recognise and degrade HY5, a positive regulator of light induced genes, in an ubiquitin-dependent manner. The association of the 19S regulatory protein complex with the 20S proteasome, of which the beta 1 subunit was identified in this study, creates the 26S proteasome. The other subunits have been identified in rice anthers in our previous report (Imin *et al.* 2001). The protein spot t2-75, which was up regulated at late stages, was matched to an ubiquitin-like protein in the database. The conserved domain searching results revealed that this protein has a conserved ubiquitin domain (UBQ, NCBI accession: cd00196.1), a retroviral aspartyl protease domain (rvp, pfam 00077.8), and an ubiquitin-associated domain (UBA, NCBI accession: cd00194.1). It is known that the ubiquitin-mediated proteolysis is part of the regulated turnover of proteins required for controlling the cell cycle progression in plants. The ubiquitin / proteosome pathway is well connected to almost all aspects of plant biology, including the cell cycle, embryogenesis, photomorphogenesis, hormone signaling, disease resistance and senescence (Vierstra 2003). Comparison of the proteome maps showed that the expression of the Jab1/ CSN5 homologue and the ubiquitin-like protein in rice anthers is

up regulated at the panicle heading stage, indicating the possible role of COP9 signalosome-mediated light signal sensing in anther dehiscence at the flowering stage.

4.4.2 Carbohydrate metabolism

We identified three key enzymes of the starch and sucrose metabolism pathway: vacuolar acid invertase (EC:3.2.1.26), fructokinase (EC:2.7.1.4) and β-galactosidase (EC:3.2.1.23). The gel comparison results show that the two charge isoforms of vacuolar acid invertase (spots t2-6 and t2-7) are expressed at a very low level and not visualised in coomassie stained gels (data not shown) until 5 days before heading. Then they start to accumulate and reach a very high level just before the heading stage (Figure 5). A similar pattern of up-regulation of spots t2-17 and t2-18, which match to fructokinase with very high confidence, was also observed starting from 5DBH continuing to the 3DBH stage. The up-regulation of these two key enzymes coincides with increased starch accumulation activity in the developing pollen grains. The developing anther is a photosynthetically inactive organ (sink). Sucrose, which is channelled from the photosynthetically active organs (source) to the anther tissues, is the main source of both the carbon and energy for the starch synthesis in pollen grains (Nguyen-Quoc and Foyer 2001; Ylstra *et al.* 1998). The vacuolar acid invertase hydrolyses the transported sucrose into the two monosaccharides (sucrose + $H_2O \rightarrow$ glucose + fructose). The enzyme fructokinase catalyses the irreversible reaction of fructose phosphorylation (fructose + ATP $\rightarrow$ fructose 6-phosphate + ADP). The phosphorylated fructose and the glucose are either converted into glucose 6-phosphate, from which the starch is synthesised in amyloplasts, or energy is produced by the glycolysis pathway. These two reactions are very important in sink tissues where sucrose assimilation and its conversion to starch are in progress. Beta-galactosidase (spots t2-40 and t2-41) is also found to be up- regulated during the late stages. Beta-galactosidase produces galactose by hydrolysing the terminal galactosyl residues from carbohydrates, glycoproteins and galactolipids. The free galactose may enter the glycolysis pathway and release energy (Smith and Gross 2000).

In addition to the major sucrose-partitioning role of the vacuolar acid invertase, the enzyme also has a proposed function in osmoregulation and cell enlargement in plants (Sturm 1999). The increased osmoregulation activity of the vacuolar acid invertase may play a crucial role in anther dehiscence, mediated by changing water potential in anther

tissues. The cell enlargement function is important for pollen grain engorgement and subsequent pollen tube growth after pollen release. Plant beta-galactosidase also has important roles in seed germination, fruit softening and senescence by hydrolysing the galactose-containing cell wall polymers, such as arbinogalactan, that contribute to the rigidity of the plant cell wall (Smith and Gross 2000; Triantafillidou and Georgatsos 2001). It is possible that increased beta-galactosidase activity during the late stage of pollen development plays important roles in pollen wall loosening essential for pollen engorgement and in anther dehiscence. In our study at least two charge isoforms for each of the three enzymes mentioned above were detected on the 2-DE gels. Slight differences were observed between the peptide mass spectra profiles of each pair of isoforms. The types and the positions of the post-translational modifications giving rise to the isoforms, and the physiological effects of formation of these isoforms in the complex carbohydrate metabolism pathways, needs to be further investigated.

4.4.3 Cell wall and cytoskeleton associated proteins

The temporal regulation patterns of some cell cytoskeleton and cell wall- associated proteins clearly show that cell expansion activity increases at late microspore developmental stages. We identified three isoforms of profilin, three isoforms of β-expansin and one isoform of actin. Actin is the major component of microfilaments in eukaryotic cells. The remodelling of the cell cytoskeleton by polymerisation and depolymerisation of actin filaments is the major driving force of cytoplasmic streaming, cytokinesis, cell expansion and development (McCurdy *et al.* 2001; Ramachandran *et al.* 2000). Profilin is a small actin monomer-binding protein (usually 12-15 kDa) which can have two opposing effects on actin polymerisation. By sequestering G-actin, profilin can cause the depolymerisation of actin microfilaments, or it can promote actin polymerisation by stimulating the exchange of ADP for ATP on G-actin (Kovar *et al.* 2000). Recent studies have shown that profilins are involved in the signalling pathways that regulate pollen tube growth by modulating protein kinase and phosphatase activity in pollen (Clarke *et al.* 1998; Vidali and Hepler 1997; Vidali and Hepler 2001). Ramachandran *et al.* also reported that the expression levels of profilin could affect flowering time in *Arabidopsis* (Ramachandran, Christensen *et al.* 2000). Many research findings indicate the importance of profilin in pollen development. In our study, the accumulation of profilin isoforms (spots t1-9, t1-10 and t1-11) gradually increased

starting from the first pollen mitosis and reached the highest level at the heading stage. The temporal regulation patterns of profilin and actin are paralleled by cell activities, such as cell division and cell elongation, in which the reorganisation of the actin cytoskeleton is crucial. This suggests the active involvement of these proteins in cell cytoskeleton remodelling during the two rounds of pollen mitosis, cell enlargement and pollen tube growth during the pollen germination. The chemical nature of the different profilin isoforms and their potential role in signal transduction pathways during male gametophyte development needs to be further studied.

We also identified three isoforms of β-expansin (spots t1-29, t1-31 and t1-33), which are expressed only at late stages. It has been demonstrated that β-expansins, also known as group-1 grass allergens, can induce the cell wall loosening necessary for plant cell growth, cell division and cell separation (Cosgrove 1997; Cosgrove *et al.* 1997). However, there is still controversy over the cell wall loosening mechanism of β-expansins (Cosgrove 2000; Grobe *et al.* 2002; Li and Cosgrove 2001). In developing pollen, the β-expansins may play two roles. First, the expression of the β-expansins are up-regulated to meet the increasing demand of cell wall loosening during the mitosis and the pollen engorgement during the late stage of pollen development. Second, the accumulated β-expansins can aid the penetration of the growing pollen tube through the stigma and style by loosening the style cell walls after the pollen release. We observed a very similar temporal regulation pattern for both the β-expansins and profilins. It is possible that the pollen cytoskeleton re-modelling and the accompanied cell wall expansion activities during the pollen tube growth are controlled by the coordinated regulation mechanism of profilin and β-expansin. Baluska *et al.* has reported the association of these two proteins in tip growth of root hair formation (Baluska *et al.* 2000).

4.4.4 Protein isoforms identified by PMF analysis

One of the potential applications of 2-DE based proteomics is identifying different isoforms of a protein. There are several post-translational processes by which a gene gives rise to several isoform proteins. These processes include post-translational splicing of the same gene product, co- and post-translational modifications (such as phosphorylation, deamidation, glycolization and acetylation), endogenous protein

degradation and oligomerization (Harry, Wilkins *et al.* 2000). As demonstrated by this work and other plant proteomics studies which were conducted in our group (Imin, Kerim *et al.* 2001; Mathesius *et al.* 2002; Mathesius *et al.* 2001; Natera, Guerreiro *et al.* 2000), PMF analysis provides a quick and economical method of identifying protein isoforms on 2-DE gels.

Of the 49 proteins identified by PMF analysis in this study, 12 protein spot (25%) were isoforms of the five proteins. In each case, the isoforms matched to the same gene product, and the isoform proteins have the same molecular weight and slightly different p*I*s on the 2-DE gels. This indicated that these isoforms were formed by the biological co- and post-translational modifications of the gene products. It is also possible that chemical modifications such as deamidation and cys-acrylamidation of peptides during the sample preparation have also contributed to the formation of these isoforms. However, the nature of the post-translational modifications, which gave rise to the isoform formation, was not revealed by this study. Elucidation of the nature of post-translational modifications using MS/MS-based approaches will provide insights into the metabolic and signal transduction pathways in which those proteins are involved during male gametophyte development.

4.4.5 New ORFs are defined

Currently the new EST and the genomic sequence information databases are expanding very fast. Although the advances in bioinformatics and the accumulation of vast genomic sequence information make it possible for comparative genomics to predict and annotate the ORFs from EST or genomic sequences, the success rate of correct prediction and annotation is not very high (Pandey and Mann 2000). This is particularly evident in the case of small ORFs. Even if a gene is successfully predicted and annotated by bioinformatics tools, the validation of the prediction by other protein-based experimental approaches is necessary. In our study, the *N*-terminal amino acid sequences obtained for four low molecular weight proteins matched to four unannotated EST sequences in the public database. These ORFs were not correctly predicted, possibly because of their small size. The confirmation of hypothetical proteins by PMF analysis verified the correct prediction of these genes and their expression in the anther tissue. The results of our *N*-terminal micro-sequencing and data analysis clearly

demonstrate that 2-DE- based proteomics is an ideal complementary tool to the comparative genomic analysis in gene prediction.

4.4.6 The N-terminal sequencing as a reliable method for the identification of low molecular weight proteins

The two different post-separation protein identification techniques, the *N*-terminal sequencing and MALDI-TOF MS-based PMF matching, were used to assign the putative identities to the proteins of interest in this study. The MALDI-TOF analysis of protein spots t1-17, t1-76, t1-13, t1-14, HNTS-5, HNTS-5, HNTS-5 and HNTS-5 produced good peptide spectra. The three-step database searches for PMF homology using this PMF data did not result in any significant PMF matches in the databases. However, all of the eight *N*-terminal sequence tags, which were obtained by the subsequent *N*-terminal analysis of the above eight protein spots, displayed 100% sequence similarities to the deduced protein sequences of TC sequences in the same database as searched by the PMF analysis. This indicated that the under representation of the candidate protein or nucleotide sequence entries in the databases was not the reason that these proteins were not identified by PMF approach. Low molecular weight proteins produce only a few peptides in the spectra–acquiring mass range (600-3500 Da) of MALDI-TOF mass spectrometry analysis after tryptic digestion. Furthermore, not all the tryptic peptides generated can necessarily be ionised and detected by mass spectrometry analysis. These factors reduce the number of experimental peptide masses available to produce confident identification of small proteins by PMF analysis. From the results of PMF and *N*-terminal sequencing analysis of these eight protein spots, we concluded that *N*-terminal sequencing, is more suitable for the identification of low molecular size proteins than PMF analysis.

4.4.7 Conclusion

In this study we successfully employed proteome analysis as a tool to investigate the global protein expression profiles of male gametophyte development in rice. Our results have provided useful information about the protein complement of rice anthers and its temporal regulation during pollen development. It is almost certain that some low-abundant proteins, which may play very important signalling and regulatory roles during the male gametophyte development, were not visualised in our 2-DE gels and

therefore not included in our study. With rapid advances in protein fractionation and separation technology we will be able in the future to visualize and analyze these important proteins. It is possible that we could not identify all of the differentially displayed proteins by PMF analysis because the genes coding for all of these proteins are probably not yet represented in the databases. Overall, the success rate of protein identification (32%) by PMF analysis in this study is very close to the success rates of PMF analysis of cell wall proteins (Chivasa *et al.* 2002) and nuclear proteins in Arabidopsis (Bae *et al.* 2003). However, the 32% success rate still compares to the 59% success rate of protein identification by PMF in maize leaf proteome analysis (Porubleva *et al.* 2001) and the 55% of success rate by *N*-terminal sequence matching in wheat endosperm proteome analysis (Skylas *et al.* 2000). The draft version of the rice genome has been released but the annotation of the genome is still in progress. The completion and the release of the fully annotated rice genome should enable us to identify most of the anther specific proteins and eventually it will lead to a better understanding of the molecular mechanism underlying male gametophyte development in plants.

CHAPTER 5 IDENTIFICATION AND CHARACTERIZATION OF RICE HOMOLOGUES OF GRASS GROUP II POLLEN ALLERGENS

[The research presented in this chapter has been published : Kerim, T., Imin, N., Weinman, J. J. and Rolfe, B. G (2003). Proteome analysis reveals developmentally expressed rice homologues of grass group II pollen allergens, Functional Plant biology, 30 (8): 843-852.]

SUMMARY

Three isoallergens of Ory s 2, homologues of grass group II pollen allergens, were identified from rice and characterised by proteome and immunochemical analyses. The *N*-terminal amino acid sequence profiles of three proteins on a 2-Dimensional electrophoresis (2-DE) gel of rice pollen proteins matched 100% to the protein sequences encoded by three rice ESTs. The deduced protein sequences from these ESTs share sequence identities of 41-43% with the protein sequences of the group II pollen allergens of different grasses, and sequence identity of 39% with the C-terminal portion of rice group I pollen allergens. Signal peptide sequences, which are similar to the leader peptides of other major pollen allergens, are also present in the deduced amino acid sequences. Polyclonal antibodies, produced in rabbits using Ory s 2 proteins purified by 2-DE, were used to investigate the developmental stage and tissue specific expression of Ory s 2 by immunochemical analyses. The results of immunochemical experiments demonstrated the late stage-specific expression patterns of Ory s 2 proteins, indicating the possible involvement of these proteins during the fertilization process. The Ory s 2 proteins do not cross-react with group II pollen allergens from some other common grasses.

5.1 INTRODUCTION

The grass family comprises about 9,000 species, and they are the most important group of plants both in ecological and economical terms. At the same time, they are also a main cause of allergic disease worldwide. Hayfever and seasonal allergic asthma resulting from the grass pollen allergies can affect a substantial proportion (20-25%) of the human population living in temperate regions (Wuthrich 1989; Griffith *et al.* 1991;).

Several groups of allergens from different grass pollens have already been isolated and characterised (for review see Andersson and Lidholm 2003; Ticha *et al.* 2002). The allergenic properties of most of the identified grass pollen allergens are well studied. However, the precise biological functions of most of the grass pollen allergens are largely unknown (Andersson and Lidholm 2003).

One group of well-characterised grass pollen allergens is the group II allergens, which are a group of acidic small proteins (pH 5.0-5.3) with a molecular weight of 10-12 kDa (Freidhoff *et al.* 1986; Ansari *et al.* 1989b). Lol p II from the perennial rye grass is one of the extensively investigated group II pollen allergens, towards which 45% of grass pollen allergic patients are sensitive (Freidhoff, Ehrlich-Kautzky *et al.* 1986; Ansari *et al.* 1987; Ansari *et al.* 1989a;). Group II allergens have also been identified and characterised from orchard grass (Roberts *et al.* 1993a), timothy grass (Dolecek *et al.* 1993), Bermuda grass, Kentucky blue grass and wheat (NCBI accession numbers are CAA10346, CAA10348 and CAA10349, respectively. Submitted by Sturaro et al).

Rice (Oryza sativa), a member of the grass subfamily Bambusoideae, is cultivated as a major cereal crop in most populated Asian regions. Airborne rice pollens originating from rice fields can be a trigger of seasonal hayfever for the rice farmers and people living near rice growing regions because protein extracts of rice pollen have shown cross-allergenicity with other grasses (Kimura *et al.* 1969; Smith *et al.* 1994a). It also has been reported that there is a significant correlation between airborne rice pollen counts and asthma episodes in children during the flowering season of rice plants in Japan (Matsumura *et al.* 1969). In Taiwan, skin prick tests of asthmatic patients using rice pollen extracts showed that 9.3% of the asthmatic patients had a positive reaction. In addition, immunoblotting studies using positive serum from asthma patients revealed three major allergens with molecular weights of 16, 26 and 32kDa, respectively (Tsai *et al.* 1990). Ory s 1, a gene coding for the group I allergen of rice pollen, has been cloned and characterised by molecular biological and immunochemical methods (Xu, Theerakulpisut *et al.* 1995a; Xu *et al.* 1999). Western blot analysis demonstrated that the Ory s 1 encodes a protein which shares high level of cross allergenicity with other grasses. These studies suggested that rice pollen allergens could be major triggers of hay fever and seasonal asthma. However, the rice pollen allergens were poorly characterised compared to their counterparts in other grasses. One reason for this situation could be the low prevalence of rice pollen as an aeroallergen in metropolitan

areas. Although group II pollen allergens have been identified from several common grasses, so far there has been no report about isolation and characterisation of group II pollen allergens from rice.

As presented in the results of Chapter 4, the *N*-terminal amino acid sequences of the two differentially displayed rice anther protein spots were shown to match three TC sequences in the database. The deduced protein sequences of the TC sequences were shown to share some level of sequence homology with grass group II pollen allergens. Because of the stage specific regulation and the high abundance levels in the late pollen developmental stages, these proteins were chosen for further characterization using immunochemical and computer analysis tools. In this chapter, the tissue and stage-specific expression of rice group II pollen allergens, and their cross-antigenicity with other grasses were studied using polyclonal antibodies raised against the gel-purified proteins. At the same time, an attempt was made to study the interaction of the group II pollen allergens with rice pollen and style proteins using immunoprecipitation in order to identify the protein interaction networks, in which the group II allergens are involved. The main aim of this chapter was to further characterise the spatial and temporal expression of rice group II pollen allergens, and to examine their use as potential protein markers of specific pollen developmental stages.

5.2 MATERIALS AND METHODS

5.2.1 Plant material

The Australian rice cultivar, Doongara, was grown in a glass house as described in Chapter 2. Mature pollen grains were isolated from fresh heading stage panicles according to Schrauwen *et al* (1990). with some modifications. In order to obtain a near synchronous pollen population, only the top three primary branches were used in the experiments. The panicles were homogenized for 1 min in iced 0.3 M mannitol using an electric blender (Quickie-Super-7, Matsushita Elec. Ltd, Osaka, Japan). The suspension was filtered through a nylon sieve (70 mm) and pollen grains were pelleted by centrifugation at low speed (250 g, 1 min). The pellet was washed twice in iced 0.3 M mannitol. After the final wash, the isolated pollen grains were kept at –80 ^{o}C until use. The anthers, palea, lemma, and stigmas were collected from the rice plants at heading stage by manual dissection of the spikelets using a pair of forceps and a scalpel. The top

flag leaves were also collected from the same plants for immunoblot analysis. All the plant materials were kept at –80 °C until use. The pollens of Bermuda grass (*Cynodon dactylon*), maize (*Zea mays*), Kentucky blue grass (*Poe pratensis*), rye grass (*Lolium perenne*), timothy grass (*Phleum pratense*) and wheat (*Triticum aestivum*) were obtained from Greer Laboratories (NC, USA.)

5.2.2 Two-dimensional gel electrophoresis

The protein extraction from rice pollens and the 2-DE analysis was carried out as described in Chapter 2.

5.2.3 Computer analysis

The conserved domain search of the predicted protein sequence was done using the Pfam 7.6 database available at http://pfam.wustl.edu/ and ScanProsite (Gattiker et al. 2002) available at http://www.expasy.ch/tools/scanprosite/. Multiple sequence alignments were performed using the ClustalX program (Thompson *et al.* 1997). The phylogenetic tree was constructed based on the distance matrix between sequence pairs using the MegAlign program of the DNASTAR software package (DNASTAR, Inc, Madison, USA)

5.2.4 Antibody production

Polyclonal antibody production using 2-DE purified proteins was done as previously described (Drenckhahn *et al.* 1993) with some modifications. Briefly, the pH 5 to 6 portions of multiple isoelectrically focused 18 cm IPG strips (pH 4 to 7), which contained the proteins of interest, were separated in the second dimension as described in section 2.3 of Chapter 2. Then, the proteins were electroblotted onto nitrocellulose membranes (Amershem Biosciences, Uppsala, Sweden), and the transferred protein spots were visualised by staining the membranes in 0.1% (w/v) Ponceau S in 1% (v/v) acetic acid for 1 min. The membrane bound protein spots were excised from the membranes, destained in deionised water and homogenised in liquid nitrogen until powdered. The homogenised membrane powder containing approximately 50 µg of gel purified Ory s 2 proteins was suspended in 2 mL of PBS (pH 7.6) and injected subcutaneously into New Zealand white rabbits. Two booster injections were administrated three and six weeks after the first injection. Blood was withdrawn from an

ear vein every two weeks and the antibody titre was checked by Western blot analysis. Prior to immunisation, pre-immune serum was taken as a control.

5.2.5 Immunoblotting

To study the tissue and stage specific expression of Ory s 2-A and Ory s 2-B/C, the plant materials including rice anthers collected at different stages, mature pollen, leaf, stigma, shoot, root and palea-lemma were homogenised in a liquid nitrogen-cooled mortar and pestle before adding SDS sample buffer (63 mM Tris-HCl, 10% Glycerol, 2% SDS, pH 6.8). The extracts were boiled, centrifuged (12000 x g) to remove insoluble materials, and the supernatant was collected. The protein concentration of the diluted supernatant was determined by Bradford dye-binding assay (Bio-Rad, Hercules, CA). The samples containing 60 μg of total proteins were fractionated using Novex precast 10-20% Tricine gels (Invitrogen, Carlsbad, CA). For 2-DE immunoblot of pollen proteins, the sample was fractionated using 11cm IPG strips (pH 4 to 7) and 12-14% ExcelGel-SDS gels. The proteins separated by SDS-PAGE and 2-DE were transferred onto PVDF membrane (Bio-Rad, Hercules, CA) using the same method as described in Chapter 2. After the protein transfer, non-specific binding sites on the membranes were blocked by incubating the membranes in 7% skim milk powder in TBS-T (20 mM Tris-HCl, pH 7.6, 150 mM NaCl, 0.05% (v/v) Tween 20) overnight at 4°C. The blots were incubated for 60 min at room temperature with antiserum (1:3000 dilution in TBS-T) raised against Ory s 2-A and Ory s 2-B/C, and were developed with an ECL-Plus western blot detection kit (Amersham Biosciences, Uppsala, Sweden) according to the manufacturer's instructions. Horseradish peroxidase conjugated anti-rabbit IgG (Amersham Biosciences, Uppsala, Sweden) was used as a secondary antibody.

5.2.6 Immunolocalisation

Rice anthers from heading stage were briefly degassed in a fixing solution containing 4% (v/v) formaldehyde in phosphate buffer (0.1 M NaPO4, pH 7.2) and incubated at room temperature for two hours. After three washes with phosphate buffer, the material was dehydrated through a graded ethanol series (60%, 70%, 90% and 100%, v/v). Then, the material was infiltrated and embedded in hydroxyethyl methacrylate resin in ultra thin PCR tubes using a Leica Historesin embedding kit

(Leica Instruments, Wetzlar, Germany) according to the manufacturer's instructions. Semi-thin sections (2-4 mm) were cut with a glass knife on a Reichert-Jung ultramicrotome (Leica Instruments, Germany) and were mounted on poly-L-lysine-coated glass slides (Sigma, St.Louis, MO). The immunological detection was carried out as described before (Cho and Kende 1998) with some modifications. Nonspecific binding sites of the sections were blocked by incubating the sections in 2% (w/v) cold fish gelatine in TBS-T for 30 min at room temperature. The sections were incubated with anti- Ory s 2-B/C antibody diluted to 1:50 with TBS-T containing 0.1% (w/v) bovine serum albumin overnight at 4°C in a moist chamber. After washing three times with TBS-T- BSA solution for 30 min, the sections were incubated with anti-rabbit IgG-alkaline phosphatase conjugate (1: 100 dilution, Boehringer Mannheim, Mannheim, Germany) in a TBS-T-BSA solution for 60 min in a moist chamber at room temperature. The sections were washed three times with TBS-T- BSA solution for 45 min and the colour reaction was performed using NBT/BCIP stock solution (Boehringer Mannheim) according to the manufacturer's instructions. Levamisole was added to the substrate at a final concentration of 1 mM to block the endogenous alkaline phosphatase activity of the anther sections. Control sections were treated identically except that pre-immune rabbit serum was substituted for Ory s 2-B/C antibody in the primary antibody incubation step. The slides were mounted with Permount (Fisher Scientific) and observed under Nikon Optiphot light microscope. Photographs were taken on Fujichrome 400 colour slide film using a Nikon FX-35 A camera.

5.2.7 Immunoprecipitation

Pooled plant material (200 mg of anthers and anthers plus style) was ground in a liquid nitrogen cooled mortar and pestle for 10 min and solubilised in 1 ml of protein extraction buffer (50mM Tris-HCl pH 7.6, 1 mM EDTA, 1mM DTT, 1x Protease inhibitor cocktail) on ice. The homogenate was centrifuged at 10000 *g* for 10 min at 4 °C and the supernatant was saved. The protein concentration of the supernatant was measured using Bradford assay. For each sample, 1 mg (250 µl in volume) of protein extract, 50 µl of anti Ory s-2 B/C antibody and 200 µl of immunoprecipitation buffer (50mM Tris-HCl pH 7.6, 100 mM NaCl, 0.5% (v/v) Nonidet-40, 1mM DTT, 1x Protease inhibitor cocktail) were mixed on ice. The mixture was incubated for 2 hours at 4°C on a rotating wheel. In the meantime, the protein A-agarose bead conjugate (Sigma,

St.Louis, MO) was pre-equilibrated by washing twice with immunoprecipitation buffer and suspended with twice the volume of immunoprecipitation buffer to make 50% protein A-agarose slurry. An aliquot of this slurry (100 µl) was added to the reaction mix on ice, and the reaction mixture was incubated for further 30 min at 4°C on a rotating wheel. The beads were collected by centrifugation at 3000 x g for 2 min at 4°C, and washed 3 times with the immunoprecipitation buffer to exclude the unbound proteins. Proteins were eluted from the beads by boiling for 1 min in 50 µl of SDS sample buffer (63 mM Tris-HCl, pH 6.8, 10% Glycerol, 2% SDS and 1mM DTT) and separated using Novex precast 10-20% Tricine gels. The proteins were visualized by colloidal coomassie staining.

5.3 RESULTS

5.3.1 2-DE analysis of pollen proteins

A proteome map of rice pollens (Figure 5.1), which were isolated from the heading stage panicles, was established by using exactly the same experimental conditions that were used for establishing the anther proteome maps (Chapter 4). The pollen proteome map was then compared to the anther proteome map at the heading stage (Figure 4.6, Chapter 4) to see if the protein spots t1-76 and t1-17 were anther or pollen specific. As shown in Figure 5.1, the protein spots t1-17 and t1-76 were present at the same gel locations where they were present on the anther proteome map at the heading stage (Figure 4.6, Chapter 4). Comparison of the spot volume showed that the volume of t1-76 and t1-17 on both proteome maps did not change significantly (less than 10%), indicating the pollen-specific expression of these proteins.

5.3.2 Identification of Ory s 2 isoforms by N-terminal sequencing and database searching

As presented in Chapter 4, the protein spots t1-17 and t1-76 were subjected to *N*-terminal microsequencing analysis. Unequivocal sequence data of 15 amino acid residues was obtained for protein spot t1-76 by *N*-terminal sequencing (Table 5.1). Clear sequence data was obtained for the first 15 amino acid residues for spot t1-17. However, a polymorphism was observed at residue 14 with serine and threonine both being detected in a 0.5 to 0.5 ratio. Figure 5.2 shows the chromatogram of the residue 14 obtained from the N-terminal sequencing of spot t1-17. Based on this data we

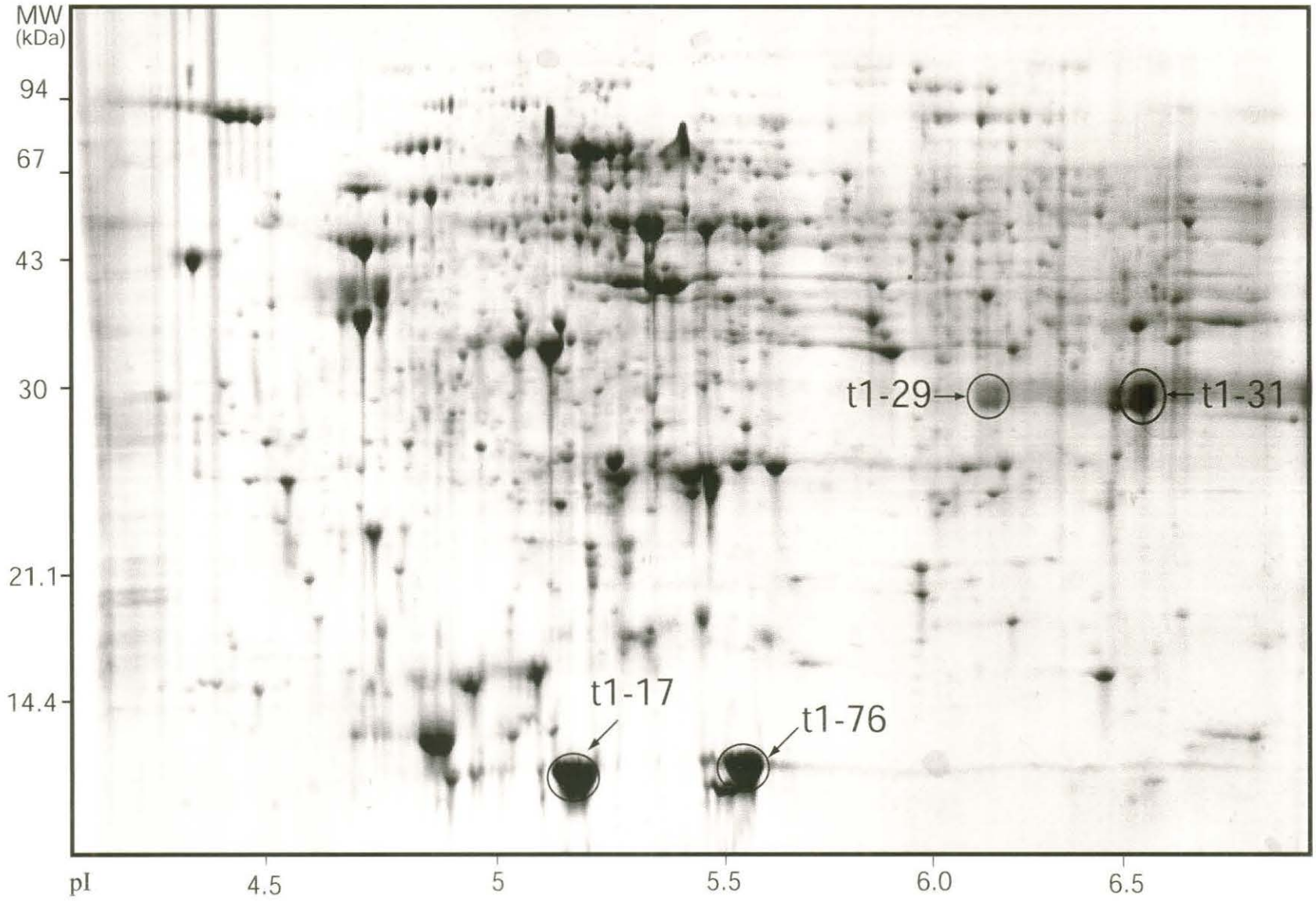
concluded that two distinct proteins with a very similar molecular weight and p*I*s are present at the position of protein spot t1-17. The separation of the anther proteins using 24 cm IPG strips also revealed the presence of two protein spots on the location of the spot t1-17 (data not shown). Therefore, two sets of N-terminal amino acid sequences as shown in Table 1 were used for the sequence homology searching for spot t1-17.

FASTA searches of the SWISS-PROT non-redundant protein database using these three sets of experimental *N*-terminal amino acid sequences did not yield any significant matches. BLAST searches of the GenBank protein database for short exact matches revealed that the two N-terminal sequences derived from spot t1-17 exhibited 100% sequence identities in 15 amino acid to two hypothetical rice proteins, OSJNBa0050F15.10 (CAD40510) and OSJNBa0050F15.9 (CAD40509), predicted from genomic sequencing (Table 1). Further tBLASTn searching of TIGR rice gene index revealed that each of the three *N*-terminal sequences displayed 100% identity over 15 amino acid residues to the ORF products of three TC sequences (Table 5.1). Each of the matching ORFs encodes for a polypeptide of 117 amino acids with a molecular weight of about 12 kDa. SignalP analysis predicted 23 amino acid long signal peptide sequences in all three deduced protein sequences (Figure 5.3). Furthermore, the cleavage sites of the signal peptides are predicted to be located immediately before the start of the experimentally determined *N*-terminal sequences. The calculated molecular weights and p*I*s of the deduced mature proteins are in close agreement with those experimentally observed on 2-DE gels. The homology searching of the NCBI protein database using the translated protein sequences revealed that all three sequences share some degree of sequence identities with the group II pollen allergens from various grasses. Therefore, they were designated as Ory s 2-A (t1-76) and Ory s 2-B/C (t1-17).

5.3.3 Sequence analysis of the matching ORFs

Alignment of the deduced protein sequences using the ClustalX program showed that these three sequences share 86.3% (between Ory s 2-A and Ory s 2-C), 90.6% (between Ory s 2-A and Ory s 2-B), and 93.2% (between Ory s 2-B and Ory s 2-C) amino acid sequence identities with each other (Figure 5.3). The results of conserved domain searching show that all three predicted protein sequences contain a conserved pollen allergen domain (PFAM accession PF01357), which matches to amino acid residues that starts from 26 to 102 in the deduced protein sequences. They also contain

Figure 5.1. 2-DE map of proteins extracted from the mature pollen of rice cultivar Doongara. For the first dimension, 1 mg of extracted pollen proteins was loaded on an IPG strip (linear pH 4-7, 18 cm). For the second dimension, a 12-14% SDS-PAGE gel was used. Proteins were visualised by colloidal Coomassie staining. The molecular weight markers are denoted on the left and p*I* is indicated at the bottom. The protein spots t1-17 and t1-76 were subjected to N-terminal micro sequencing analysis and the protein spots t1-29 and t1-31 were identified as isoforms of rice group I pollen allergens (β-expansins) in Chapter 4.


MW
(kDa)
94
67
43
30
21.1
14.4
t1-29
t1-31
t1-17
t1-76
pI
4.5
5
5.5
6.0
6.5

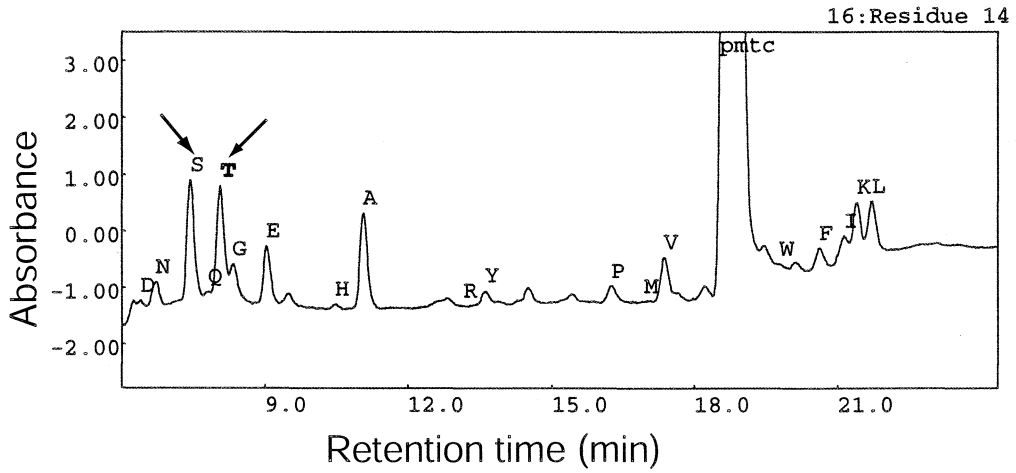


Figure 5.2. Chromatogram of the residue 14 obtained from the N-terminal micro sequencing of spot t1-17. The single letter amino acid abbreviations above the peaks indicated by arrows are corresponding to two amino acids, Serine and Threonine, obtained for residue 14 of spot t1-17.

Table 5.1. Proteins analysed by *N*-terminal microsequencing and database searching

Spot I.D.	Mr/pI[a)]	*N*-terminal Sequence	Homology (% identity)	NCBI accession	MW/pI[b)]	Signal peptide	TIGR accession
t1-76	11/5.6	TEVTFKVGEGSSGKS	100%		10.0 / 5.45	Yes	TC114263
t1-17	11/5.2	TELTFKVAEGSSASS	100%	CAD40510	10.1 / 5.12	Yes	TC125017
		TELTFKVAEGSSATS	100%	CAD40509	10.1 / 5.13	Yes	TC129660

[a)] The observed molecular mass and pI of the protein spots.

[b)] The predicted molecular mass and isoelectric point of the matched peptides calculated from the ORFs after excluding the predicted signal peptide sequences.

```
          ***:*****:**.*****:******:****.****..**************:********
Orys2-A:  MASLSSFRLSVAVAALLVVGSCATEVTFKVGEGSSGKSLELVTNVAISEVEIKEKGGKDW 60
Orys2-B:  MASLSSFRLAVAVAALLVVGSCATELTFKVAEGSSASSLELVTNVAISEVEIKEKGGKDW 60
Orys2-C:  MASMSSFRLAVAAAALLVIGSCATELTFKVAEGSSATSLELVTNVAISEVEVKEKGGKDW 60

          *.****.:***::****.************************************::
Orys2-A:  VALKESSTNTWSLKSEAALKGPFSVRFLVKNGGYRVVDDVIPESFTAGSEYKSGINV    117
Orys2-B:  VALKESSSNTWTIKSEAPLKGPFSVRFLVKNGGYRVVDDVIPESFTAGSEYKSGIQL    117
Orys2-C:  VGLKESGSNTWTLKSEAPLKGPFSVRFLVKNGGYRVVDDVIPESFTAGSEYKSGIQL    117
```

Figure 5.3. Multiple alignment of the deduced amino acid sequences of three isoforms of Ory s 2 protein. The deduced amino acid sequences of Ory s 2-A (TC114263), Ory s 2-B (TC125017) and Ory s 2-C (TC129660) are aligned using ClustalX program (Thompson). The "*" symbol indicates that the residues in that column are identical in all sequences in the alignment and the ":" and "." indicate that the residues in that column are conserved and semi-conserved, respectively. The experimentally determined N-terminal sequences of Ory s 2 proteins are underlined in red. The blue bar on the top of the columns shows the location of the pollen allergen domain (PFAM accession PF01357) in the alignment.

an expansin cellulose-binding domain (EXPANSIN_CBD, PROSITE accession PS50843), which matches to amino acid residues 47 to 113. Both of the domains are the characteristic of group I and group II pollen allergens.

Figure 5.4 demonstrates the alignment of the deduced protein sequences with amino acid sequences of known group II pollen allergens in the database. The Ory s 2 isoforms share comparable amino acid sequence identities (41.7% to 43.5%) with group II pollen allergen proteins of orchard grass, Bermuda grass, timothy grass, wheat, velvet grass and Kentucky blue grass. Phylogenetic analysis, which was based on the multiple sequence alignment, yielded a phylogenetic tree as shown in Figure 5.5. On this phylogenetic tree, the isoforms of Ory s 2 form a distinct subgroup, well separated from the other group II pollen allergens. Sequence alignment analysis also indicated 39%, 41% and 38% sequence identities between the deduced protein sequences of Ory s 2-A, Ory s 2-B and Ory 2-C, and the C-terminal portion (starts from amino acid residue 168 to 267) of rice group I pollen allergen (Ory s 1), respectively (sequence alignment not shown).

5.3.4 Immunoblot analysis

Immunological approaches were employed to see if there was any significant immunological cross-reactivity among the two identified protein spots. Western immunoblot analyses of 2-DE separated pollen proteins (in pH range of 4 to 7) were carried out using the polyclonal antibodies produced against the protein spots Ory s 2-A and Ory s 2-B/C. As expected, the Ory s 2-A antibody detected both protein spots t1-76 and t1-17 (Figure 5.6). However, the Ory s 2-A antibody also cross-reacted to a neighbouring low molecular weight protein spot which was located just below the t1-76, indicating the contamination of the t1-76 protein material used for the injection. A very low detection signal was also present in the gel region corresponding to the rice group I pollen allergens (β-expansins). The antibody produced against Ory s 2-B/C detected both protein spots t1-17 and t1-76, but it did not cross-react with the low molecular weight protein next to t1-76 (Figure 5.7). The control experiments using the pre-immune serum did not detect the same spots in the replica blot. The Ory s 2-B/C antibody was used for the subsequent immunoblot and tissue localisation experiments.

5.3.5 Stage- and tissue-specific expression of Ory s 2

Expressions of Ory s 2 during different pollen developmental stages and in different floral tissues were studied by Western blot immuno-detection analysis. Western blots of rice proteins extracted from rice anthers at four different developmental stages and from different parts of the rice plant, were probed with the anti Ory s 2-B/C polyclonal antibody (Figure 5.8). The results showed that Ory s 2-B/C was not detected in anthers at the uninucleate microspore stage. Reactivity was first detected at a low level in anthers of early binucleate microspore stage and the expression gradually increased to a high level in anthers at the mature pollen stage. A similar level of expression was observed in anther samples of mature pollen stage and pollen that was isolated from the panicle of the same stage. The same immunoblot analysis also revealed that the Orys 2-B/C is only expressed in mature pollen and late stage anther but not in any other part of inflorescence, leaves, shoots or roots of rice.

5.3.6 Cross-reactivity of Ory s 2 to other grasses

The alignment of deduced amino acid sequences of Ory s 2 to group II pollen allergens of other grasses indicated that they share considerable sequence identity. The anti Ory s 2-B/C polyclonal antibody was used to investigate whether there was any significant immunological cross-reactivity between Ory s 2 and group II pollen allergens of other grasses. The Western immunoblot results showed that the polyclonal antibody did not bind to any protein components, within molecular weight range of 10-12 kDa (group II allergens), in protein extracts of six common grasses (Figure 5.9). However, it bound to a band with a molecular weight of approximately 32 kDa in maize pollen extracts. Although there is a reasonably high level of sequence similarity between Ory s 2 and group II pollen allergens from other grasses as demonstrated by our sequence alignment analysis, no cross-reactivity of Ory s 2 to these allergens was detected.

5.3.7 Immunolocalization of Ory s 2 in rice anthers

The 2 DE and immunoblot analysis results show that Ory s 2 proteins are expressed only in late stage pollen and anthers. To confirm pollen-specific expression of Ory s 2, tissue localization of Ory s 2-B/C in cross sections of rice anthers was performed using immuno-light microscopy. The Orys 2-B/C is exclusively localized to the cytoplasm

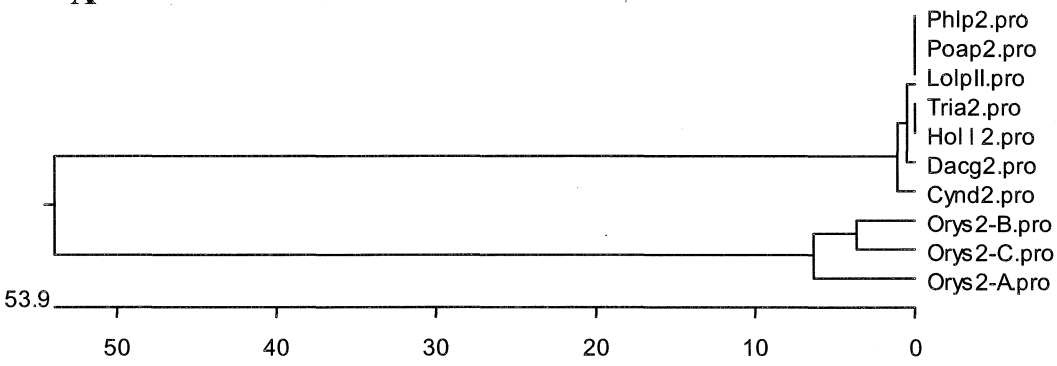
Figure. 5.4 Amino acid sequence alignment of Ory s 2-A, B and C with group II pollen allergens of other common grasses. Multiple sequence alignment was performed using ClustalX program (Thompson, 1997). The accession numbers and sources of the proteins are: Hol l 2 (CAA10347) of velvet grass; Lol p II (CAA10350) of rye grass; Poa p 2 (CAA10348) of Kentucky bluegrass; Tri a 2 (CAA10349) of wheat; Phl p 2 (P43214) of timothy grass; Dac g 2 (CAA10345.1) of orchard grass and Cyn d 2 (CAA10346) of Bermuda grass. The "*" symbol indicates that the residues in that column are identical in all sequences in the alignment and the ":" and "." indicate that the residues in that column are conserved and semi-conserved, respectively.

```
          *** **   *::*. * *. *: *..::**.* :**. . * ::.:     ::***::*:
Orys2-A --MASLSS-FRLSVAVAALLVVGS-CATEVTFKVGEGSSGKSLELVTNVA---ISEVEIKEK    55
Orys2-B --MASLSS-FRLAVAVAALLVVGS-CATELTFKVAEGSSASSLELVTNVA---ISEVEIKEK    55
Orys2-C --MASMSS-FRLAVAAAALLVIGS-CATELTFKVAEGSSATSLELVTNVA---ISEVEVKEK    55
    Hol MSMASSSSSSLLAMAVLAALFAGAWCVPKVTFTVEKGSNEKHLAVLVKYEGDTMAEVELREH    62
 LolpII MSMASSSSSSLLAMAVLAALFAGAWCVPKVTFTVEKGSNEKHLAVLVKYEGDTMAEVELREH    62
  Poap2 MSMASSSSSSLLAMAVLAALFAGAWCVPKVTFTVEKGSNEKHLAVLVKYEGDTMAEVELREH    62
  Tria2 MSMASSSSSSLLAMAVLAALFAGAWCVPKVTFTVEKGSNEKHLAVLVKYEGDTMAEVELREH    62
 PhlpII MSMASSSSSSLLAMAVLAALFAGAWCVPKVTFTVEKGSNEKHLAVLVKYEGDTMAEVELREH    62
  Dacg2 MSMASSSSSGLLAMAVLAALFAGAWCVPKVTFTVEKGSNEKHLAVLVKYEGDTMAEVELREH    62
  Cynd2 MSMASSSSSGLLAMAVLAALFAGAWCVPKVTFTVEKGSNEKHLAVLVKYEGDTMAEVELREH    62

         *..:**.:.:.  ..*::.** .*:***..***.::*  .*.***:**.:* *: * .
Orys2-A GGKDWVALKESSTNTWSLKSEAALKGPFSVRFLVKNGGYRVVDDVIPESFTAGSEYKSGINV   117
Orys2-B GGKDWVALKESSSNTWTIKSEAPLKGPFSVRFLVKNGGYRVVDDVIPESFTAGSEYKSGIQL   117
Orys2-C GGKDWVGLKESGSNTWTLKSEAPLKGPFSVRFLVKNGGYRVVDDVIPESFTAGSEYKSGIQL   117
    Hol GSDEWVAMTKGEGGVWTFDSEEPLQGPFNFRFLTEKGMKNVFDDVVPEKYTIGATYAPEE--   122
 LolpII GSDEWVAMTKGEGGVWTFDSEEPLQGPFNFRFLTEKGMKNVFDDVVPEKYTIGATYAPEE--   122
  Poap2 GSDEWVAMTKGEGGVWTFDSEEPLQGPFNFRFLTEKGMKNVFDDVVPEKYTIGATYAPEE--   122
  Tria2 GSDEWVAMTKGEGGVWTFDSEEPLQGPFNFRFLTEKGMKNVFDDVVPEKYTIGATYAPEE--   122
 PhlpII GSDEWVAMTKGEGGVWTFDSEEPLQGPFNFRFLTEKGMKNVFDDVVPEKYTIGATYAPEE--   122
  Dacg2 GSDEWVAMTKGEGGVWTFDSEEPLQGPFNFRFLTEKGMKNVFDDVVPEKYTIGATYAPEE--   122
  Cynd2 GSDEWVAMTKGEGGVWTFDSEEPLKGPFNFRFLTEKGMKNVFDDVVPEKYTIGATYAPEE--   122
```

Figure 5.5. **A**- A phylogenetic tree representing the alignment of Ory s 2 proteins with group 2 pollen allergens from other common grasses. The phylogenetic tree was constructed based on the distance matrix between sequence pairs using the MegAlign program of the DNASTAR software package (DNASTAR, Inc, Madison, USA). The scale at the bottom represents the substitution events between the sequence pairs. **B-** The taxonomic classification of 20 common allergenic grass genera according to Andersson *et al* (2003).

A


Phlp2.pro
Poap2.pro
LolpII.pro
Tria2.pro
Hol I 2.pro
Dacg2.pro
Cynd2.pro
Orys2-B.pro
Orys2-C.pro
Orys2-A.pro
53.9
50
40
30
20
10
0

B

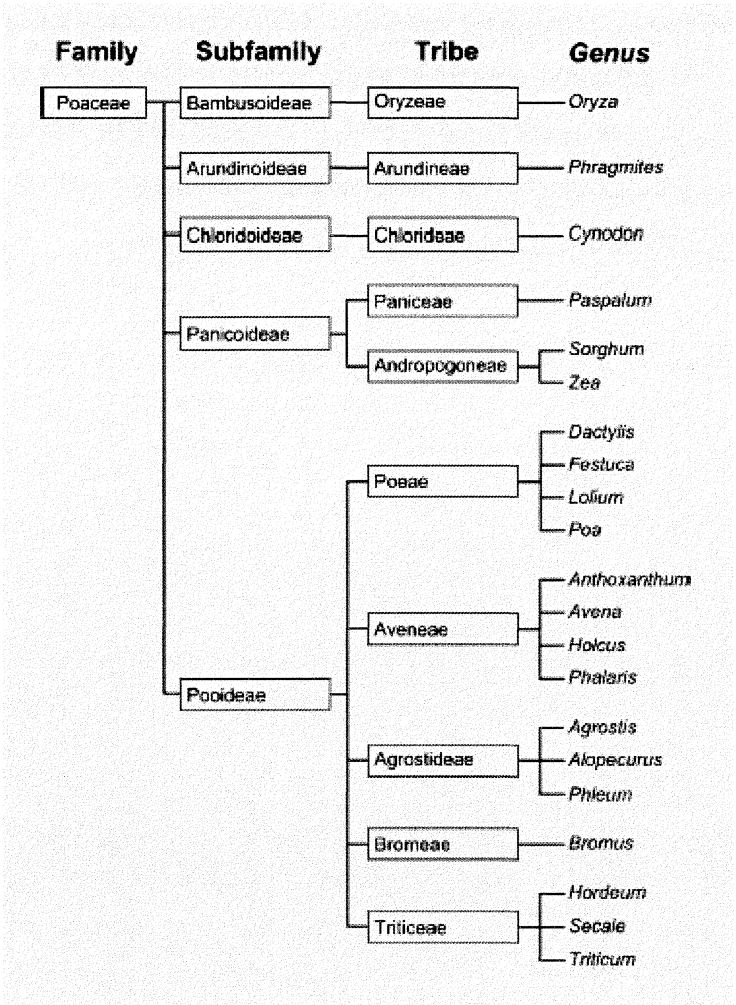

Family
Subfamily
Tribe
Genus
Poaceae
Bambusoideae
Oryzeae
Oryza
Arundinoideae
Arundineae
Phragmites
Chloridoideae
Chlorideae
Cynodon
Panicoideae
Paniceae
Paspalum
Andropogoneae
Sorghum
Zea
Pooideae
Poeae
Dactylis
Festuca
Lolium
Poa
Aveneae
Anthoxanthum
Avena
Holcus
Phalaris
Agrostideae
Agrostis
Alopecurus
Phleum
Bromeae
Bromus
Triticeae
Hordeum
Secale
Triticum

Figure 5.6. 2-DE immunoblot analysis using anti Ory s 2-A serum. 500 μg of pollen protein extract was separated using a 11 cm IPG strip (pH 4-7) in the first dimension and SDS-PAGE in the second dimension. **A.** The separated proteins were electroblotted onto PVDF membrane and probed with anti Ory s 2-B/C serum using ECL-Plus western blot detection kit. The detection signals marked as Ory s 2-A and Ory s 2-B, C are corresponding to the protein spots t1-76 and t1-17 respectively in Figure 5.6-B. **B.** A replica gel of Figure 5.6-A was stained using colloidal coomassie and the protein spots corresponding to the detection signals in Figure 5.6-A are marked.

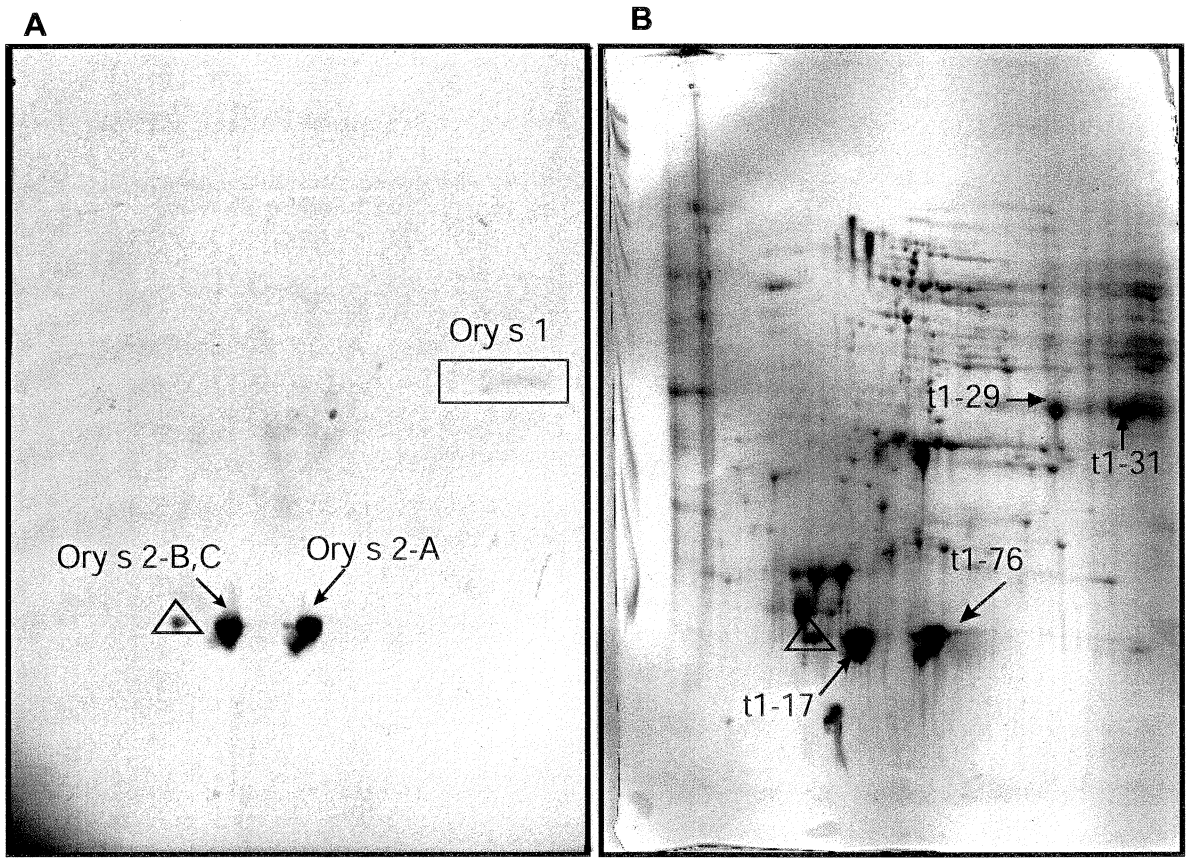

A
Ory s 1
Ory s 2-B,C
Ory s 2-A
B
t1-29
t1-31
t1-76
t1-17

Figure 5.7. 2-DE immunoblot analysis using anti Ory s 2-B/C serum. 500 μg of pollen protein extract was separated using a 11 cm IPG strip in first dimension and SDS-PAGE in second dimension. **A.** The separated proteins were electroblotted onto PVDF membrane and probed with anti Ory s 2-B/C serum using ECL-Plus western blot detection kit. The detection signals marked as Ory s 2-A and Ory s 2-B/C are corresponding to the protein spots t1-76 and t1-17 in Figure 5.7-B respectively. **B.** A replica gel of Figure 5.7-A was stained using colloidal coomassie stain and the protein spots corresponding to the detection signals in Figure 5.7-A are marked.

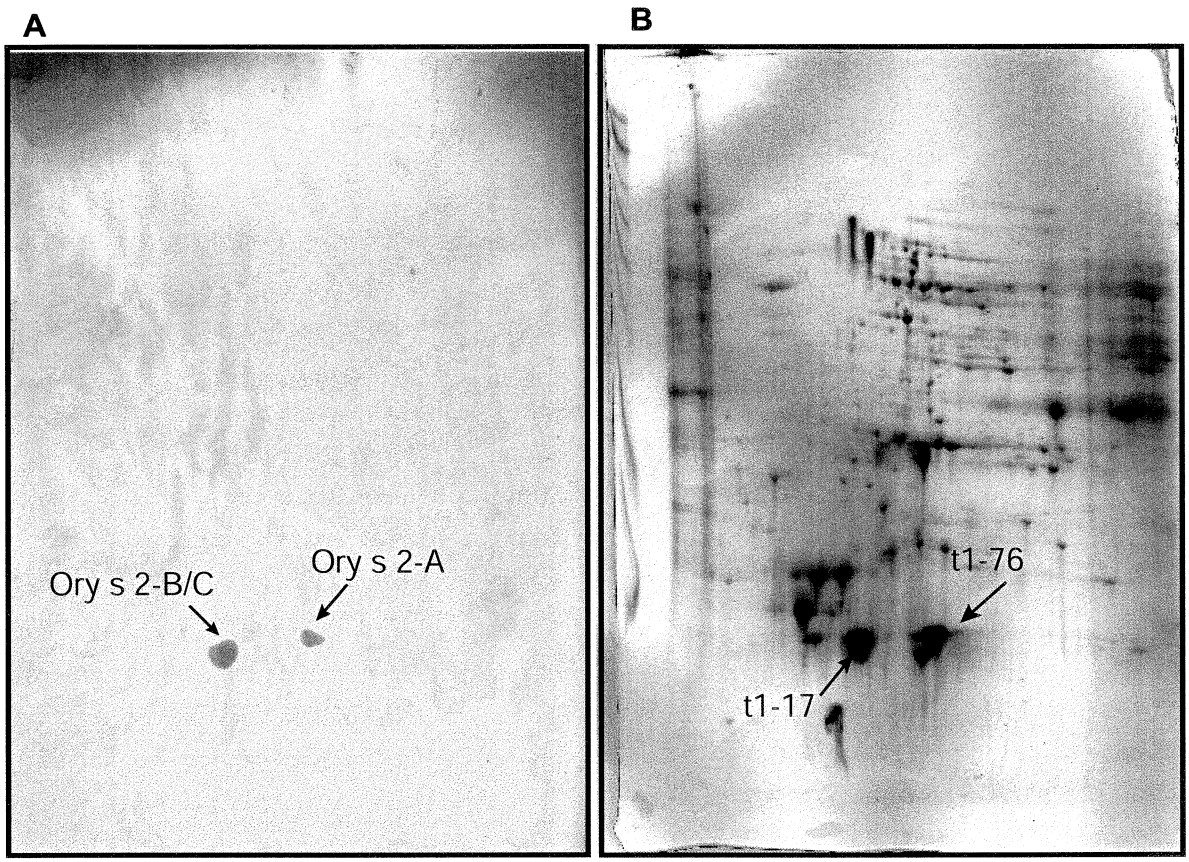

A
Ory s 2-B/C
Ory s 2-A
B
t1-76
t1-17

Figure 5.8. Stage and tissue specific expression of Ory s 2 revealed by western immunoblot analysis. Total soluble proteins, which were extracted from rice pollen and anthers at different developmental stages and from different tissues of rice plant, were fractionated by SDS-PAGE and electroblotted onto PVDF membrane. The blot was probed using anti Ory s 2-B/C antibody. Lane 1, anthers at 8 days before heading (DBH); lane 2, anthers at 5DBH; lane 3,anthers at 3DBH stage; lane 4, anthers at heading stage; lane 5, pollen at heading stage; lane 6, leaf; lane 7, stigma; lane 8, shoot; lane 9, root and lane 10, palea and lemma. Lane C, heading stage pollen protein was probed using pre-immune serum as a negative control.

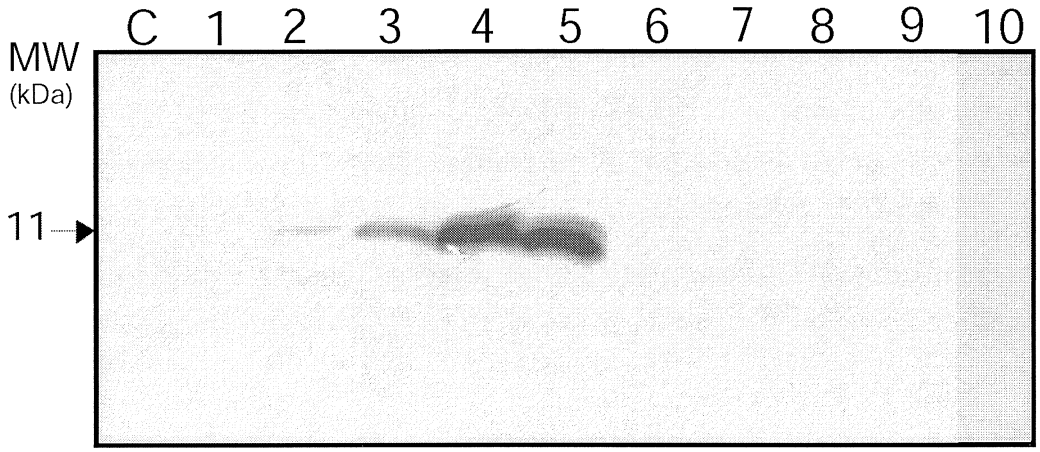
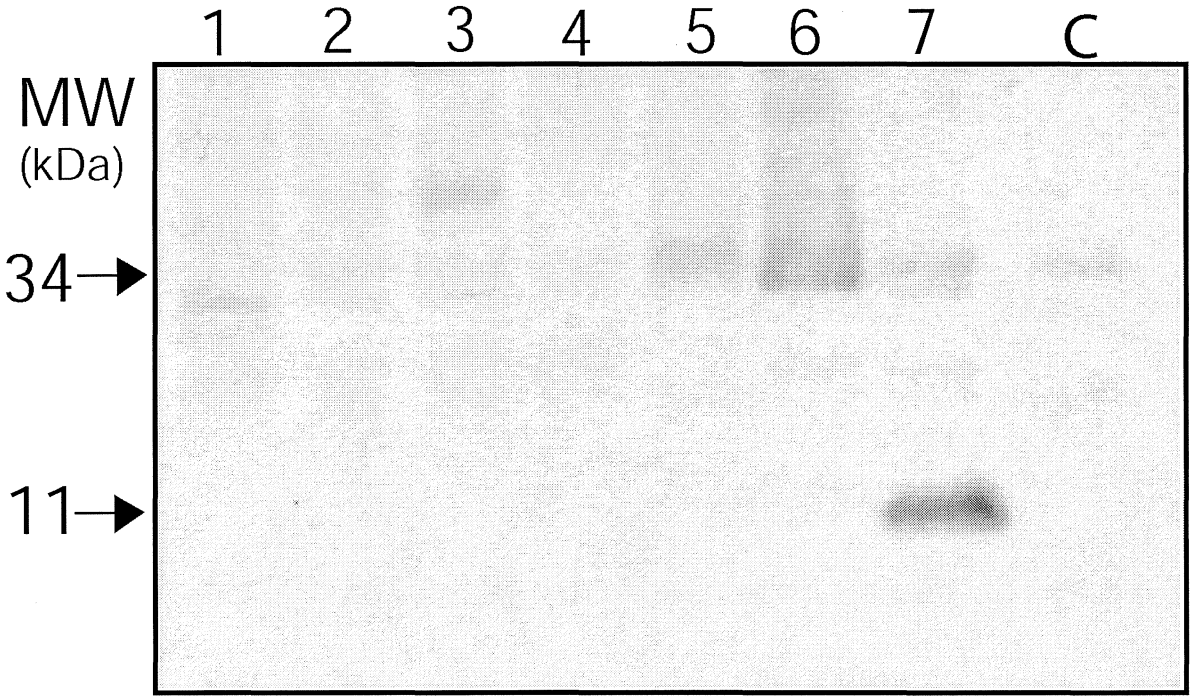

C 1 2 3 4 5 6 7 8 9 10
MW
(kDa)
11

Figure 5.9. Cross-reactivity of Ory s 2-A antibody with group II allergens of other grasses. Soluble pollen proteins from Bermuda grass (lane 1), Kentucky blue grass (lane 2), timothy grass, perennial rye grass (lane 4), wheat (lane 5), maize (lane 6) and rice (lane 7) were separated by SDS-PAGE and blotted onto PVDF membrane. The blot was probed with anti Ory s 2-B/C antibody and detected with ECL-Plus detection kit. Lane C, heading stage pollen protein was probed using pre-immune serum as a negative control.


1
2
3
4
5
6
7
C
MW
(kDa)
34
11

and cell wall of mature pollen (Figure 5.10), but no color reaction product was detected in the nuclei region of pollen and anther wall tissues. In control experiments using the pre-immune serum, the anther sections showed a low background color.

5.3.8 Immunoprecipitation

In order to examine if the Ory s 2 proteins interact with proteins of style origin in the rice flower, an attempt was made to co-immunoprecipitate Ory s 2 and other proteins (both anther and style origin) which can form a protein complex with Ory s 2 under native conditions. Figure 5.11 shows a coomassie stained gel of the immunoprecipitate of soluble protein fractions from the anthers at the heading stage and from the anthers, including style tissues, at the late heading stage using the Ory s 2-B/C antibody. Comparison of the protein profiles of immunoprecipitates of anthers and the anthers including styles did not reveal any difference. A band with a molecular weight of 11 kDa, which is corresponding to the Ory s 2 proteins, was present in both the immunoprecipitates that were performed with the Ory s 2-B/C antibody but not in the control experiments which were performed with pre-immune serum. As shown in Figure 5.11, there are a number of high molecular weight protein bands present in increased amount in the immunoprecipitates compared to the control experiment. But this result was not reproducible in a repeated experiment under the more stringent immunoprecipitation conditions.

5.4 DISCUSSION

Type I allergy is a prevalent global medical problem. Among the most potent elicitors of type I allergy are group II/III allergens of grass pollen. The group II allergens have been identified and characterised from a number of common grasses. However, so far there is no report on identification and characterisation of group II/III pollen allergens of rice, a widely cultivated crop plant and a member of the grass family. The accumulated rice EST and genome sequences in the public databases already contain partial or complete sequences of genes coding for some rice allergens. These databases have made it easier to identify these allergens using proteome analysis techniques without using immunological screening and cloning procedures. As described in Chapter 4, during our proteome analysis of rice male gametophyte development, the peptide mass fingerprinting and *N*-terminal amino acid sequence

profiles of two late stage specific proteins were matched to the translation products of three unannotated EST sequences in public databases. Further investigations were carried out using immunochemical and bioinformatic research tools to further characterize the proteins.

Based on the sequence homology among deduced protein sequences of matching ORFs, it can be concluded that Ory s 2-A, Ory s 2-B and Ory s 2-C are homologous proteins. Cross-reactivity of the polyclonal antibodies to Ory s 2-A and Ory s 2-B/C protein spots on a 2-DE blot also supported this conclusion. The results of further bioinformatic analyses of matching TC sequences in combination with immunochemical analysis provided several lines of evidence that the three proteins are rice homologues of group II pollen allergens of the grass family. First, the deduced amino acid sequences of the matching ESTs have 23 amino acid long signal peptide sequences. Furthermore, these signal peptide sequences share around 45% of sequence identities with signal peptides predicted for other well-characterised group II allergens of other grasses. This means that Ory s 2 proteins are likely to be secreted, which is one of the common characteristics of pollen allergens. Second, the experimental and calculated molecular weights and pIs of the protein spots on 2-DE gels are in close agreement with molecular weights and pIs of group II pollen allergens from the grass family. They exist as immunologically indistinguishable multiple isoforms in the pI range of 5.0-5.4 as it was reported for Lol pII (Ansari *et al.* 1989b; Dolecek *et al.* 1993). Third, they share comparable sequence identities with other group II pollen allergens as demonstrated by sequence alignment analysis. At the same time, conserved pollen allergen domains and some other sequence motifs that are specific to grass pollen allergens are also present in Ory s 2 proteins. Fourth, the Ory s 2 sequences have 39% of amino acid sequence identity with the C-terminal portion of Ory s 1 (group I pollen allergen of rice). It was reported in rye grass that Lol pII also shares a similar level of sequence identity (33%) with the C-terminal part of Lol pI (Griffith *et al.* 1991). Finally, the temporal and spatial expression patterns of Ory s 2 proteins demonstrated by immunochemical experiments are consistent with the expression pattern of Ory s 1 and other group II allergens (Dolecek *et al.* 1993; Xu *et al.* 1995a). The evidence presented above indicates that these three proteins are the rice homologues of grass group II pollen allergens. Therefore, we designated them as Ory s 2-A, Ory s 2-B (OSJNBa0050F15.10) and Ory s 2-C (OSJNBa0050F15.9) according to the

Figure 5.10. Immunolight microscopic localisation of Ory s 2 on cross sections of heading stage rice anthers. **A.** The semi-thin anther cross sections were probed with anti Ory s 2-A antibody and detected with NBT/BCIP colour reaction substrate. The insert shows one pollen grain at higher magnification. The colour reaction product was detected in cytoplasm and cell wall of pollen grains. **B.** The anther cross sections were treated with pre-immune serum and detected in parallel with **A.** (Bars = 100 μm)

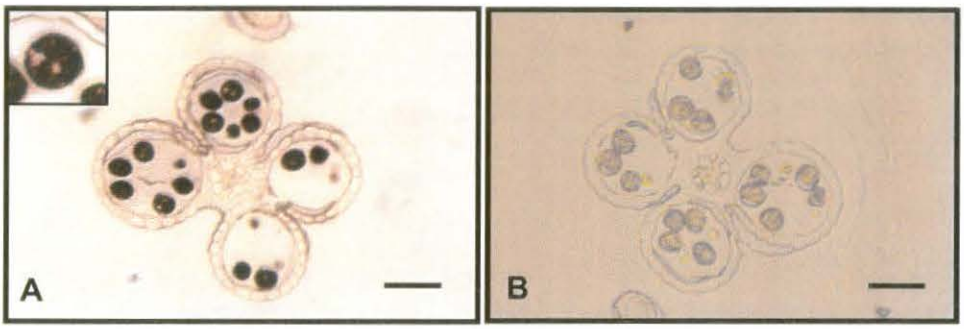
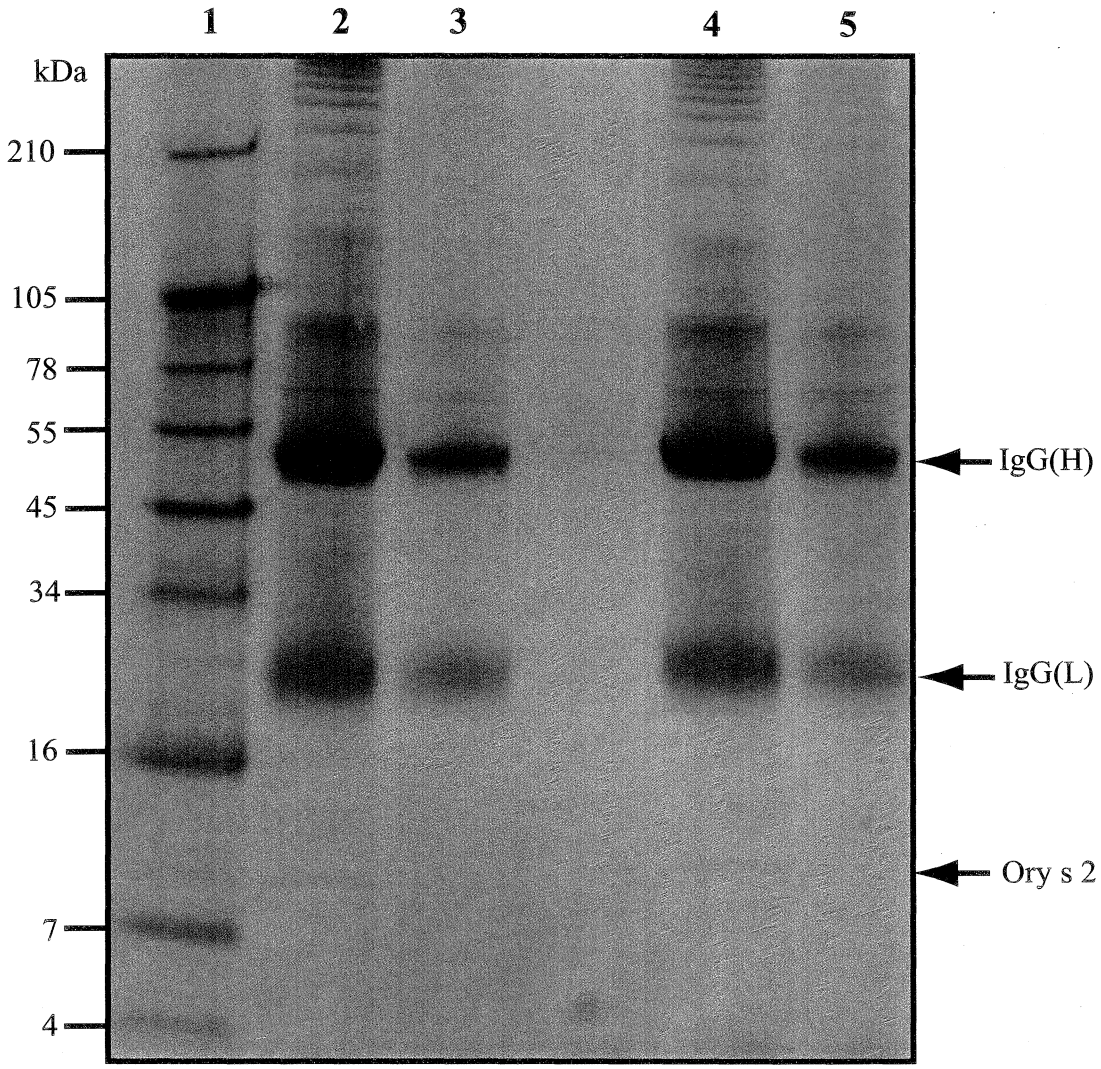

A
B

Figure 5.11. Immunoprecipitation of anther and style proteins using anti Ory s 2-B/C antibody. Proteins were extracted under native conditions from anthers plus styles (lane 2 and lane 3) and from anthers alone (lane 4 and lane 5) at the heading stage. Immunoprecipitation experiments were carried out using anti Ory s 2-B/C serum (lane 2 and lane 4). The control experiments were also performed in parallel using the pre-immune serum (lane 3 and lane 5). The precipitated proteins were separated using Novex precast 10-20% Tricine gels and visualised by colloidal coomassie staining. Protein standards were co-migrated in lane 1. Arrows indicate the heavy and light chains of IgG in the immunoprecipitant. A band corresponding to Ory s 2 is also indicated by an arrow.


1
2
3
4
5
kDa
210
105
78
55
45
34
16
7
4
IgG(H)
IgG(L)
Ory s 2

international allergen nomenclature system (King *et al.* 1994). The *N*-terminal amino acid sequence of Ory s 2-A protein was submitted to the SWISS-PROT protein database with accession number P83466, and the submission of *N*-terminal sequences of Ory s 2-B and Ory s 2-C is pending. The unavailability of serum from patients who are allergic to rice pollen prevented further investigations into the allergenic properties of this group of protein using immunochemical techniques.

Early studies show that grass pollen allergens that belong to the same allergen group but from different grass origins have immunological cross-reactivity (Smith *et al.* 1994a; Smith *et al.* 1994b). Furthermore, cross-reactivity of Ory s 1 with group I allergens from timothy, rye and orchard grasses has been demonstrated (Xu *et al.* 1995b). In this report it is evident from the sequence alignment analysis that Ory s 2 proteins share a considerable level of sequence similarity with group II allergens from other grasses. Therefore, we expected some degree of cross-binding of the anti-Ory s 2 antibody to group II allergens of other common grasses. However, our immunoblot analysis showed that polyclonal the anti-Ory s 2 antibody of rice plant, which is a member of grass subfamily Bambusoideae, does not cross-bind to the counterparts from other three grass subfamilies including Pooideae (rye grass, Kentucky blue grass, timothy grass and wheat), Chloridoideae (Bermuda grass) and Panicoideae (maize). A possible explanation for this is that the allergenic epitopes of group II allergens are probably not very conserved among the grass subfamilies. The close conformation between the phylogenetic relationship among the group II allergens and the taxonomic classification of 20 common grass genera also supports this explanation (Figure 5.5). Another factor likely to contribute to a lack of cross-reactivity between group II allergens is the molecular size of group II allergens. The small molecular size of group II allergens may reduce the probability of conservation of common allergenic epitopes among group II allergens from different grass origins. Therefore, a lower level of cross-reactivity among the group II allergens than that of among the group I allergens can be expected. However, the Ory s 2 antibody cross bound to a ~32 kDa protein band in maize pollen extracts. These proteins presumably are the group I allergens of maize and they have some conserved cross-reactive epitopes with Ory s 2. Zea mI, the group-I pollen allergens of maize, have been identified and characterized using molecular and immunochemical tools and displayed cross-reactivity to group I allergens from other grasses (Broadwater *et al.* 1993). The cross-reactivity between Ory s 2 and other grass

group II pollen allergens needs to be further studied using IgE from rice pollen allergic patients. Further study on immunological features of rice group II pollen allergens may lead to the application of these allergens for the immunotherapy of rice pollen allergy.

Although the allergic properties of grass pollen allergens have been studied intensively by immunologists, their native biological functions in plant reproductive developmental process are not very well understood. In terms of biological function, the group I allergens are the best-studied group among the eleven groups of grass pollen allergens. Cosgrove *et al.* reported that group I pollen allergens share low level of sequence homology with expansins and they also demonstrated *in-vitro* cell expansion activities in grass cell walls. Therefore the group I pollen allergens were classified as β-expansin family (Cosgrove *et al.* 1997). However, the biochemical mechanism underlying expansin activities of group I allergen remains highly controversial as different biochemical activities were proposed to group I allergens (Grobe *et al.* 1999; Grobe, Poppelmann *et al.* 2002; Li and Cosgrove 2001). The considerable level of sequence similarity and existence of some common functional domains between Ory s 2 and Ory s 1 raise the question of whether Ory s 2 also has expansin activity. Until now, there is no report in scientific journals about the physiological function of group II pollen allergens.

The developmental stage specific expressions and the high abundance of Ory s 2 proteins in mature pollen revealed by our investigation indicates that these proteins should have important functions in pollen development and pollen stigma interactions. In order to gain some knowledge about the biological function of group II allergens, attempts was made to immunoprecipitate and subsequently identify proteins of pollen and style tissue origins which bind to Ory s 2 in native biological conditions. No conclusive differences were observed between the SDS-PAGE profiles of test and control experiments. Optimization of the immunprecipitation conditions through more repeated experiments was not pursued further because of the limited availability of rice anthers and styles.

Elucidation of the exact biological function of group II allergens using immunochemical and biochemical assays and molecular biology tools in the future will contribute to the better understanding of molecular mechanisms underlying the male

gametophyte developmental processes and also to the elucidation of the precise biological functions of other grass pollen allergens in plant development.

CHAPTER 6 GENERAL DISCUSSION

6.1 INTRODUCTION

In the post-genome sequencing era, proteomics has become an important functional genomics tool providing biological information at the protein level, and has been widely applied to different aspects of the life sciences. The ability of proteomics to deal with large-scale determination of gene and cellular function at the protein level means that it can be used to comprehensively characterize the complex molecular networks underlying various biological processes. This study was directed at investigating the usefulness of 2-DE-based comparative proteomics to study the changes in global protein expression during the male gametophyte development in rice plants, and demonstrates how the potential of this technique can be realized when applied to real biological problems.

6.2 THE ACHIEVEMENTS OF THIS THESIS

Chapter 3:

- The male gametophyte developmental process in rice cultivar Doongara was studied to identify unique cytological "markers" for the major cellular events which take place inside the developing anther.
    - The allometric relationships between different cytological events and vegetative growth measurements such as anther length, auricle distance and the days before heading were used to define discrete developmental stages in male gametophyte development in rice cultivar Doongara. This enabled the collection of near homogenous anther populations for each developmental stage.

Chapter 4:

- Anther proteome maps were established within the pH ranges of 4-7 and 6-11 for six microspore developmental stages. More than 3,500 protein spots were detected by silver staining in the combined pH range of 4-11.
- Comparison of the proteome maps revealed that the expression levels of 150 protein spots changed at different stages during the development. Putative identities

were assigned to 49 protein spots by peptide mass fingerprinting analysis. A further eight low molecular weight protein spots were identified by N-terminal sequencing.

- The integration of the protein expression and identification data demonstrated the active participation of three different signaling and metabolic pathways in the later stages of pollen development.

Chapter 5:

- Three isoforms of developmentally expressed rice homologues of grass group II pollen allergens (Ory s 2) were identified and characterized using bioinformatics. The N-terminal sequence of one isoform has been deposited in the public protein database
- Polyclonal antibodies were produced against Ory s 2 proteins using gel-separated and nitrocellulose-bound proteins as the antigen. The antibodies were used for the further characterization of these proteins.
  - Rice group 2 pollen allergens displayed low levels of sequence similarity and cross reactivity with their counterparts from other common grasses.

6.3 RICE AS A MODEL SYSTEM FOR MALE GAMETOPHYTE RESEARCH

One of the achievements of this thesis is its contribution to the use of rice plants as a model system to study the male gametophyte development.in crops. As outlined in section 1.2.4.3, several features of rice plants make them an ideal model system for plant research (Goff 1999; Izawa and Shimamoto 1996; Salse *et al.* 2002). These include the availability of the complete draft rice genome sequence and the large number of rice EST sequences, both of which are critical to the use of peptide mass fingerprinting for high-throughput identification of proteins from 2-DE gels. The only impediment to the use of rice as a model for the study of male gametophyte development is the relative difficulty of collecting sufficient amounts of homogenous anther materials at the different developmental stages. This bottleneck is caused by the small size of rice anthers and the heterogeneous maturation pattern of rice flowers on the rice panicle. The cold susceptible Australian rice cultivar Doongara was chosen for this project with the expectation that the experimental results of this project can be

integrated with other ongoing research projects of our group investigating the cold-induced male sterility in rice using the same cultivar. As presented in Chapter 3, the extensive cytological and growth analyses were conducted in this cultivar to define the allometric relationship between growth measurement and microspore developmental stages. These studies provided the basis for the selective collection of a homogenous anther population for each predefined microspore developmental stage. In combination with continuous plant growth throughout the year under a controlled environmental facility, the micro dissection of rice spikelets provided sufficient anther materials for proteomic profiling of male gametophyte development.

The research on this model system resulted in the identification of new male gametophyte-specific genes and also in the detection of molecular networks which are active in the developing rice anthers. Further exploitation of this approach promises to generate further insight into plant reproductive development.

6.4 PROTEOMICS AS A RESEARCH TOOL TO STUDY MALE GAMETOPHYTE DEVELOMENTAL PROCESS

In the post genomic era, proteomics has emerged as a valuable analytical tool to deal with the analysis of whole genome at the protein level. The unparalleled capacity of 2-DE to simultaneously display thousands of proteins has been utilized in plant research to differentially display the protein contents of various plant tissues under particular physiological conditions, such as biotic and abiotic stresses and plant microbe interactions (Imin, Kerim *et al.* 2001; Natera, Guerreiro *et al.* 2000; Rossignol 2001; van Wijk 2001). In order to expand the application of proteomics in plant science, this thesis set out to systematically investigate one of the fundamental reproductive developmental processes in plants using 2-DE based comparative proteomics. The proteome maps of six developmental stages of anthers were established, new proteins were identified and some of the metabolic and signaling pathways involved in rice anther development were detected. Overall, the primary findings of this thesis demonstrated the potential of applying proteomics to other aspects of plant developmental biology.

However, it also became apparent that the current 2-DE based proteomics technology is not free from problems. Protein solubility and the ability to detect low

abundance proteins are major limitations, which results in the incomplete representation of proteomes on 2-DE gels. Hydrophobic proteins, which represents up to 30% of the genome coding capacity of eukaryotic cells (Wallin and von Heijne 1998), are generally underrepresented in proteome maps due to their low solubility in currently available detergents which are compatible with 2-DE. The total number of predicted genes from draft sequences of *japonica* rice genome range from 32,000 to 61,000 (Delseny 2003; Goff, Ricke *et al.* 2002). In this study, the 3,500 rice anther proteins displayed on 2-DE gels only accounts for 10% of the minimum coding capacity of the rice genome, assuming one gene product per estimated gene. This figure will decrease further if the number of protein isoforms detected in this study is taken into the account. Furthermore, the transcriptional analyses of pollen gene expression in monocot plants have estimated that about 20,000 and 24,000 different mRNAs are present in the mature pollen grains of *Tradescantia* and maize, respectively (Mascarenhas 1990). If it is assumed that the similar numbers of genes are expressed in rice pollen, the proteins detected on the anther 2-DE maps could not fully reflect the true complexity of eukaryotic gene expression.

Another major limitation lies in the area of post-separation protein identification techniques. As shown in Chapter 4, the success rate with protein identification by PMF analysis was lower than expected in this study. The protein identification by *N*-terminal analysis and peptide mass fingerprinting of eight protein spots using the same database resulted in the totally different success rates. The eight protein spots, which had not been identified by PMF analysis, were successfully identified by *N*-terminal sequencing analysis. This indicated that the under representation of the matching database entries was not the main reason contributing to the failure of protein identification of these eight protein spots by PMF analysis. Furthermore, the searches of different databases using the same PMF data also resulted in different success rates (section 4.3.3). These examples highlighted the imperfectness of PMF analysis and the importance of database quality in the success rate of protein identification. The real capacity of proteomics can be fully utilized if only the complete or near complete proteome of a tissue or cell is successfully displayed and every protein is identified.

The current experimental limitations in 2-DE based proteomics can be reduced in a number of ways. Solubilisation of hydrophobic proteins can be improved by the application of new zwitterionic detergents (Chevallet *et al.* 1998) and by sequential

extraction of the sample with reagents of increasing solubilizing power (Molloy *et al.* 1999). Detection and identification of highly hydrophobic proteins can also be achieved by MudPIT technology which based on chromatographic separation of peptides generated from a complex protein mixture (Koller *et al.* 2002) One dimensional SDS PAGE, which is compatible with solubilisation of membrane proteins using SDS, can also used to the separation of highly hydrophobic proteins, even though the resolution of this technique is much lower than 2-DE analysis (Andon *et al.* 2002) Separation of protein samples using narrow range (1 pH unit) IPG strips to generate overlapping proteome contig maps (Cordwell *et al.* 2000) and the enrichment of low-abundance proteins by preparative isoelectric pre-fractionation step using a multi-compartment electrolyser (MCE) can contribute to the visualization of low-abundance proteins (Herbert *et al.* 2001).

6.5 GLOBAL CHANGES IN PROTEIN EXPRESSION DURING THE RICE MALE GAMETOPHYTE DEVELOPMENT

The establishment of 2-DE based proteomics as a viable differential display tool has enabled the investigation of global changes in gene expression of a biological system at the protein level. As presented in Chapter 4, while the global expression pattern of rice anther proteins was largely unaltered, the comparisons of rice anther proteome maps showed that the expression level of 150 proteins changed more than two fold during the course of anther development. These 150 proteins account for less than 5% of the total proteins that were visualized on 2-DE gels. Most significant changes in protein expression levels were detected between the proteome maps of the pollen mother cell stage and the heading stage. More than 50 new protein spots were newly expressed at the heading stage compared to the pollen mother cell stage.

Transcriptional profiling studies of male gametophyte development showed that in pollen grains of *Tradescantia* and maize, the male gametophyte specific genes could be classified into "early" and "late" genes (Mascarenhas 1990). According to those studies, most of the "early" genes were transcriptionally active just after pollen meiosis. The "late" genes become active soon after first pollen mitosis and are thought to play major roles during the later part of pollen maturation and during pollen germination. The transcripts of some "late" genes are stored untranslated in pollen grains until pollen germination (Stinson, Eissenberg *et al.* 1987). The late stage-specific proteins that were

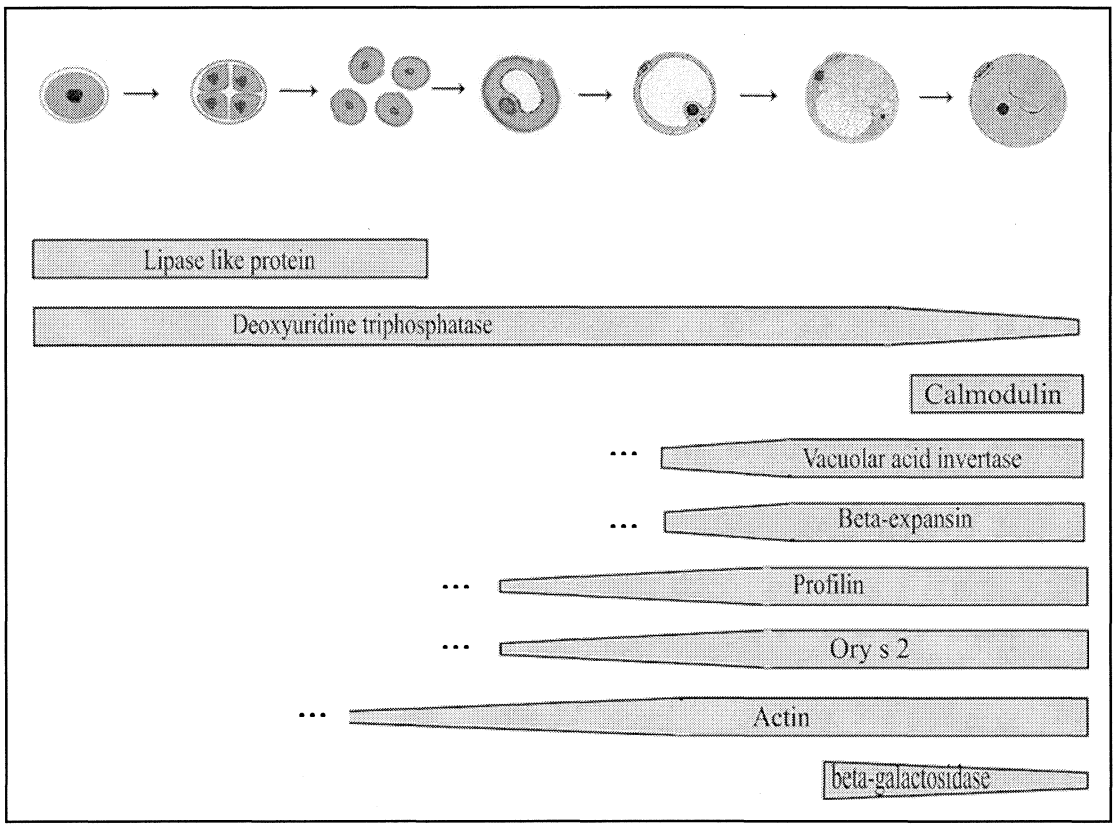


Figure 6.1. Temporal expression patterns of some differentially displayed protein identified in this study.

detected in this study are presumably the products of the "late" genes, because the expression of most of these proteins starts at around the first pollen mitosis (Table 4.1, Chapter 4), which conforms to the above-mentioned temporal expression pattern of "late" genes. Furthermore, the comparison of proteome maps also revealed a small number of proteins representing the "early" class gene products. However, the number of the detected early stage-specific proteins is much less than those of late stage-specific proteins in this study. A possible explanation for this is that the abundance level of "early" gene products is probably below the limit of detection of the silver staining used in this study. It is also possible that some transcripts of the "early" genes remain untranslated until the onset of certain later stages. The examples of "early" and "late" gene products, which were identified in this study, are shown in Figure 6.1. These results of this study clearly demonstrate the potential power of proteomics to test and complement the transcriptional profiling studies at the protein level.

6.6 METABOLIC AND REGULATORY PATHWAYS INVOLVED IN POLLEN DEVELOPMENT

Out of the150 differentially displayed protein spots, 49 proteins were tentatively identified by PMF analysis and eight proteins by *N*-terminal sequencing and homology searches. In order to better understand the molecular networks in which these proteins are involved, putative functional classifications were assigned to the identified proteins according to the functional classifications in *A. thaliana* and *M.trancatula* proteome projects (Kaul, Koo *et al.* 2000; Mathesius, Keijzers *et al.* 2001). A statistical summary of the functional classification is shown Figure 6.1. Largest group of proteins (26%) are involved in metabolism and the next largest group matched to hypothetical or unknown proteins. Unlike other studies, the proteins involved in cell wall and cytoskeleton formation constituted the third largest group (18%), perhaps indicating the importance of these processes in pollen development. Several functional protein classes such as cell signaling and protein synthesis that have been identified in other projects are underrepresented, probably because of the low number of identified proteins in this study.

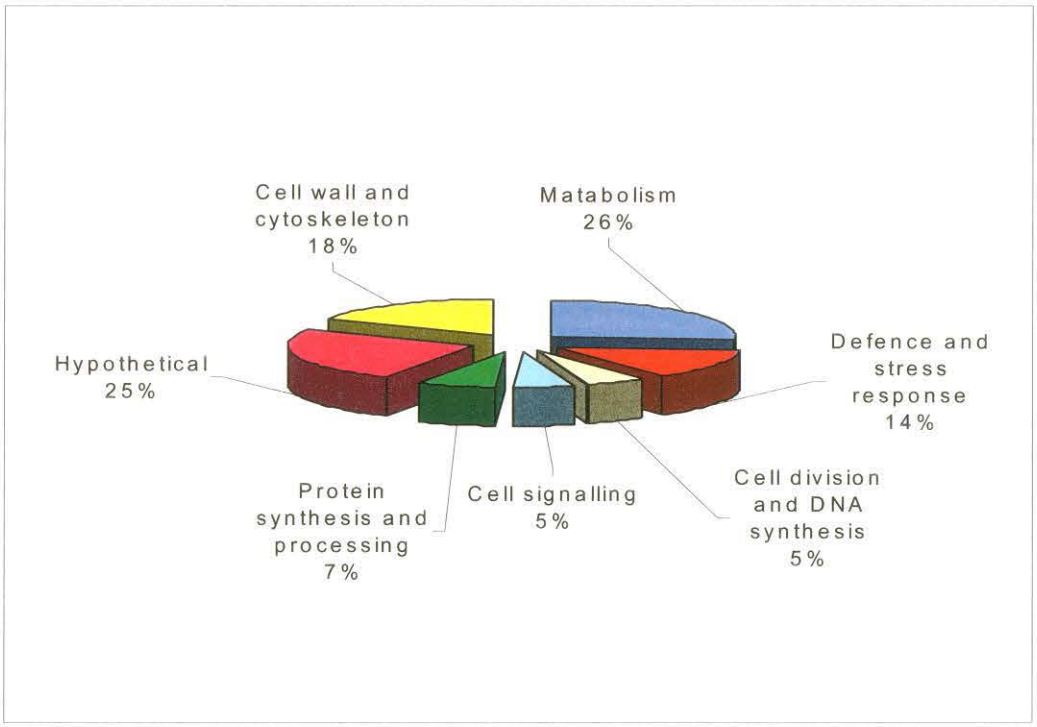


Figure 6.2. Summary of the functional classification of the identified rice anther proteins.

Based on the predicted identities of the limited number of differentially displayed anther proteins and their temporal regulation patterns, three developmentally regulated metabolic or regulatory pathways were identified that are active in developing rice anthers. Table 6.1 lists the predicted identities of three groups of proteins and their proposed biological functions in cellular processes in which they are involved. The ubiquitin/26 proteosome pathway, of which three components were identified, is involved in the selective proteolysis of proteins in plants (discussed in Chapter 4). Recent analysis of *Arabidopsis thaliana* genome revealed that a sizable fraction (~5%) of the plant proteome is devoted to encoding the components of this pathway and the individual components are connected to almost all aspects of plant biology (Fu *et al.* 1999; Vierstra 2003). The up-regulation of two components indicates the possible role of COP9 signalosome-mediated light signal sensing in anther dehiscence at the flowering stage. The up-regulation of three proteins involved in starch and sucrose metabolism at later stages of pollen development coincides with the increased starch synthesis activity in pollen grains. Late stage-specific expression of B-expansin, profilin and actin conforms to their proposed function in cell wall expansion and pollen stigma interaction.

Considerable portion of the identified proteins in this study are primary metabolic enzymes which are involved in basic "housekeeping" metabolic processes in plants. However, it is possible that these housekeeping proteins have other function(s) during

Table 6.1. Summary of the proteins involved in metabolic pathways

Pathway	Identity	Isoform number	Proposed Function
Ubiquitin / 26 proteosome	Jab 1/CSN5	1	repression of photomorphogenesis
	Ubiquitin	1	protein degradation
	20S proteosome	1	protein degradation
Starch /sucrose metabolism	Vacuolar acid invertase	2	hydrolysis of sucrose
	Fructokinase II	2	fructose phosphorylation
	Beta-galactosidase	2	hydrolysis of O-glycosides
Cell wall and cytoskeleton	Actin	1	cytoskeleton formation
	Profilin	3	actin binding
	Beta-expansin	3	cell wall loosening
	Group II allergens	3	cell wall loosening ??

the male gametophyte development in anthers in addition to their main role in primary metabolism. Recent studies showed that some proteins could "moonlight", or have more than one distinct function in an organism. Therefore, they are called moonlighting proteins (Jeffery 1999). The function of moonlighting proteins can vary as a consequence of changes in cellular localization, cell type, oligomeric state, or the concentration of a ligand, substrate, cofactor or product. One of the identified proteins in this study, quinone oxidoreductase, has been reported to have moonlight functions in animals (Piatigorsky 1998). Recent results by the Johnston *et al.* show that the regulatory proteins of 26S proteosome complex, of which a subunit was identified in this study, has a role in nucleotide excision repair and transcription elongation (Gonzalez *et al.* 2002; Jeffery 2003). This moonlighting mechanism of proteins adds another dimension to the cellular complexity of eukaryotic gene expression.

Regarding the significant cellular changes that take place inside anthers during the development, the differential protein expression profiles that were obtained in this study are probably an under representation of protein changes in developing anthers. Therefore, it is still inadequate to provide a full insight into the molecular mechanism of male gametophyte developmental process. Complete achievement of such a goal in this study was impeded mainly by the current limitations in 2-DE and protein identification techniques (discussed in section 1.3.5).

6.7 FUNCTION OF IDENTIFIED PROTEINS IN COLD INDUCED MALE STERILITY

One of the aims of this study is to assist the research projects aimed at understanding the molecular mechanism of cold-induced male sterility in rice (abnormal development) by providing information about proteins involved in the normal development of rice anthers. Two of the identified proteins in this study, β-expansin and nucleoside diphosphate kinase, have been found to be involved in cold-induced male sterility in rice anthers by proteomic analysis (Imin 2002). An increase in the partial degradation or cleavage of these proteins was detected in trinucleate anthers which had been treated with cold shock at the young microspore stage. The involvement of β-expansin in the process of cold-induced male sterility suggests that the sterility resulting from cold treatment is possibly partially attributed to incomplete pollen tube growth. In

the same study, the β-6-subunit of 20S proteosome was identified and observed to be up-regulated by two fold in response to cold treatment at the young microspore stage. The identification of another subunit of the 20S proteosome and other constitutes of ubiquitin-proteosome pathway in our study provided further information about the cellular processes in which this cold responsive protein (β-6-subunit of 20S proteosome) is involved (section 4.4.1). Future studies in cold-induced male sterility can be directed at the stage-specific anther proteins and the relevant biological pathways identified in this project. This will contribute to the understanding of molecular mechanism of the stress-response process in plants.

6.8 BIOLOGICAL FUNCTION OF POLLEN ALLERGENS IN PLANT REPRODUCTION

Grass pollen allergens have been studied extensively since the 1960s, and to date, eleven different groups of grass pollen allergens have been identified and characterised from one or more species (Andersson and Lidholm 2003; Suphioglu 2000). The pollen allergens have been poorly characterized in terms of their biological function in plants, although their immunological features have been well studied. Out of the eleven groups of pollen allergens characterised in grasses, biological functions have been proposed for only a few groups based on the sequence homology to other known proteins or based on the results of *in vitro* biochemical assays (Andersson and Lidholm 2003). The group I allergens are thought to have an expansin activity and to play important roles in pollen tube growth during fertilization (Cosgrove 1997; Cosgrove, Bedinger *et al.* 1997). The group 5 allergen of timothy grass pollen has been proposed to have ribonuclease activity (Bufe *et al.* 1995). The profilins are involved in signalling pathways that regulate the pollen cytoskeleton remodelling during the pollen engorgement and pollen tube growth (Vidali and Hepler 2001).

In this study, multiple isoforms of three groups of rice pollen allergens have been identified. These allergens included three isoforms of group I allergen (β-expansin), three isoforms of group 12 (profilin) and three isoforms of group II allergens. In mature pollen, these three groups of allergens are the most abundant proteins. The comparison of proteome maps from different developmental stages demonstrated a very similar temporal expression pattern of these allergens during the anther development. The

expression of β-expansin and profilin was first detected at the early or late binucleate stages and steadily increased to high levels at the pollen maturity. The expression of these proteins coincides with the formation of exine and intine wall layers and enlargement of pollen grains, indicating the involvement of these proteins in pollen cell wall structure and cell enlargement (discussed in Chapter 4).

Multiple isoforms of group II pollen allergens were identified in this study using a combination of proteomics and bioinformatics. As revealed by immunoblot analysis, their expression is specific to pollen, and they also displayed a temporal regulation pattern which is similar to β-expansins, implying their possible involvement in pollen development and fertilization processes. Until very recently, there was no report in scientific journals about the physiological function of group II pollen allergens. Li *et al.* very recently reported that the group 2 pollen allergens from ryegrass and timothy grass exhibited plant cell wall-loosening activity which is similar to β-expansins (Li *et al.* 2003). They concluded that at least part of the characteristic cell wall-loosening action of β-expansin is attributable to its C-terminal domain, which shares a high level of sequence homology with the corresponding group 2 allergens. Even if the cell expansion activity of the group II pollen allergens is confirmed, further research still needs to be done before the biological significance and exact molecular mechanism of redundant cell expansion activities of the group I and group II pollen allergens in plant reproductive development are elucidated.

6.9 PRODUCTION OF POLYCLONAL ANTIBODIES USING A COMBINATION OF 2-DE AND NITROCELLULOSE BOUND ANTIGEN

During the past few decades, immunocytochemical methods have become powerful and indispensable investigation tools for different aspects of plant science. The application of immunochemical methods in plant research often involves the production of monoclonal and polyclonal antibodies directed against some minor proteins, or against nonsoluble or membrane proteins, which are difficult to purify in native form by liquid-based fractionation approaches. In addition to its role as a differential display tool, the 2-DE has proved to be a quick and efficient method of obtaining pure antigenic material for immunization from crude or partially purified protein extracts (Diano and

Bevic 1997). Once proteins are separated on 2-DE gels, the protein spots of interest can be excised from the lightly stained gels, ground and injected into animals with or without adjuvant. The generally poor results derived from this approach may be due to the toxicity of SDS and acrylamide in the antigen which cause intensive tissue and cell damage, resulting in the early death of the experimental animals (personal communication with Dr Nijat Imin, Australian National University, Canberra). As an alternative approach, the 2-DE separated proteins can be transferred to nitrocellulose membrane and the protein-bearing nitrocellulose is then solublized by physical or chemical means before being administrated to animals (Knudsen 1985). As demonstrated in Chapter 5, the antibody production to the 2-DE-purified Ory s 2 proteins was successful by this method using as little as 50 μg proteins. The antibody production elicited to the low molecular weight proteins, without conjugation to carrier proteins, was probably partially attributable to the intrinsic allergenic nature of the Ory s 2 proteins. Another factor that contributed to the good immune response might be the adjuvant role of the nitrocellulose acting to slowly release the bound antigens, and thereby elicit a good immune response in animals.

The antibody production to 2-DE separated proteins in this study has demonstrated the parallel application of 2-DE based proteomics as a micropreparative tool in addition to its main application as a large-scale analytical approach. However, it should be kept in mind that the injection of a distinct protein spot does not necessarily result in the production of a mono-specific antibody as one protein spot on a gel can contain more than one protein population because some proteins with the similar p*I* and molecular weight can co-migrate to the same gel positions. As presented in Chapter 4 (Table 4.2), this is the case in two out of eight protein spots (25%) which were subjected to *N*-terminal sequencing. Therefore, special care should be taken while producing antibodies to 2-DE proteins identified by PMF analysis, because it is very difficult to confirm the homogeneity of the analyzed spots from the generated PMF data. Application of narrow range IPG strips in the micro- preparative 2-DE should increase the chance of obtaining homogenous antigens for antibody production.

6.10 CONCLUSION AND FUTURE PERSPECTIVES

This thesis has demonstrated the usefulness of 2-DE based proteomics for investigating the changing pattern of gene expression at the protein level during

developmental processes in plants. In order to successfully follow a developmental sequence it is vital to have access to homogenous cell populations representing particular developmental stages. Otherwise, the real changes in protein expression levels will remain diluted or swamped by the overlapping or more abundant protein population. In this project, a considerable effort was made collect near homogenous anther populations by excluding the non-anther parts of the rice flower. This effort certainly resulted in the partial enrichment of some anther-specific and low abundant proteins.

An alternative method to collecting homogenous anther samples from rice flowers at different developmental stages is to subject the mixed microspore population to fractionation in a percoll or sucrose gradient (Fan, Armstrong *et al.* 1988). Although highly uniform microspore populations representing different developmental stages can be obtained in sufficient amounts, this method is not suitable for the isolation of pre-meiotic and early young microspores which lack of rigid pollen wall. As presented in Chapter 5, the isolation of mature pollen grains and the analysis of pollen proteins by 2-DE have proven successful in our study. Unfortunately, a similar research approach has not been extended to the study of microspores from other stages due to the time limitation of the PhD project.

The development of methods for the isolation of generative cells, sperm cells, embryo sacs and egg cells from angiosperm plants has opened up new possibilities for plant reproductive research, utilizing proteomic and transcriptional profiling of generative or sperm cells. Viable sperm or generative cells can be isolated from the pollen grains or growing pollen tubes by enzymatic digestion, osmotic bursting or squashing, filtration and centrifugation (Matthysrochon *et al.* 1987; Theunis *et al.* 1991). In several plant species, the isolated sperm and generative cell has been subjected to transcriptional profiling, resulting in the identification of generative cell specific-genes and in the isolation of the male gametic cell-specific promoters (Singh *et al.* 2003; Xu, Swoboda *et al.* 1999b). McCormick's laboratory recently initiated a large-scale project to construct and sequence cDNA libraries from isolated sperms, eggs, and embryo sacs of maize (for details visit http://www.pgec.usda.gov/McCormick/McCormick/mclab.html). Similar studies have also been conducted at the protein level in *Plumbago zeylanica L* (Geltz and Russell 1988). The proteins extracted from sperm, cytoplasmic particulate and water-soluble fractions of mature pollen have been

separated and compared. However, so far there is no report about the large-scale profiling of sperm or generative cell proteins in crop plants. Large-scale proteome profiling of isolated generative or sperm cells will provide a deeper insight into the protein composition of male gametophytes.

In parallel to proteomic profiling, it is also important to study the dynamic changes of gene expression during the male gametophyte development at the transcriptome level using large scale and high-throughput mRNA profiling techniques such as SAGE (serial analysis of gene expression) and DNA microarray analysis. This may enable the identification of some genes whose protein products are not displayed by 2-DE analysis. At the same time, the large-scale biochemical profiling approach, which is also known as metabolomics, can be applied to profile the changes of chemical constituents of male gamete cells at different developmental stages. This approach can complement the proteomic and transcriptomic analysis with functional data at the metabolome level.

It is hoped that in the new era of Systems Biology, technological improvements in proteomics and other global expression profiling tools will ultimately lead to the more complete unraveling of the molecular and biochemical mechanisms underlying male gametophyte developmental processes in crop plants. The knowledge derived from such studies can readily be exploited to resolve the practical problems such as cold- induced male sterility and self-incompatibility in plants, and will eventually contribute to the improvements of crop plants to the benefit of ever increasing world population.

REFERENCES

Aebersold R, Mann M (2003) Mass spectrometry-based proteomics. *Nature* **422**, 198-207.

Aebersold RH, Leavitt J, Saavedra RA, Hood LE, Kent SB (1987) Internal amino acid sequence analysis of proteins separated by one- or two-dimensional gel electrophoresis after in situ protease digestion on nitrocellulose. *Proc Natl Acad Sci U S A* **84**, 6970-6974.

Agrawal GK, Rakwal R, Yonekura M, Kubo A, Saji H (2002) Proteome analysis of differentially displayed proteins as a tool for investigating ozone stress in rice (Oryza sativa L.) seedlings. *Proteomics* **2**, 947-959.

Albani D, Robert LS, Donaldson PA, Altosaar I, Arnison PG, Fabijanski SF (1990) Characterization of a pollen-specific gene family from Brassica napus which is activated during early microspore development. *Plant Mol Biol* **15**, 605-622.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389-3402.

Andersen JS, Mann M (2000) Functional genomics by mass spectrometry. *FEBS Lett* **480**, 25-31.

Anderson L, Seilhamer J (1997) A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* **18**, 533-537.

Anderson NG, Anderson L (1982) The Human Protein Index. *Clin Chem* **28**, 739-748.

Anderson NL, Anderson NG (1998) Proteome and proteomics: new technologies, new concepts, and new words. *Electrophoresis* **19**, 1853-1861.

Andersson K, Lidholm J (2003) Characteristics and immunobiology of grass pollen allergens. *Int Arch Allergy Immunol* **130**, 87-107.

Andon NL, Eckert D, Yates JR, 3rd, Haynes PA (2003) High-throughput functional affinity purification of mannose binding proteins from Oryza sativa. *Proteomics* **3**, 1270-1278.

Andon NL, Hollingworth S, Koller A, Greenland AJ,Yates JR., Haynes PA. (2002). Proteomic characterization of wheat amyloplasts using identification of proteins by tandem mass spectrometry. *Proteomics* **2**, 1156-1168

Ansari AA, Freidhoff LR, Marsh DG (1989a) Molecular genetics of human immune responsiveness to Lolium perenne (rye) allergen, Lol p III. *Int Arch Allergy Appl Immunol* **88**, 164-169.

Ansari AA, Kihara TK, Marsh DG (1987) Immunochemical studies of Lolium perenne (rye grass) pollen allergens, Lol p I, II, and III. *J Immunol* **139**, 4034-4041.

Ansari AA, Shenbagamurthi P, Marsh DG (1989b) Complete primary structure of a Lolium perenne (perennial rye grass) pollen allergen, Lol p III: comparison with known Lol p I and II sequences. *Biochemistry* **28**, 8665-8670.

Ashcroft AE (2003) Protein and peptide identification: the role of mass spectrometry in proteomics. *Nat Prod Rep* **20**, 202-215.

Asturias JA, Arilla MC, Gomezbayon N, Martinez J, Martinez A, Palacios R (1997) Cloning and high level expression of cynodon dactylon (bermuda grass) pollen profilin (cyn d 12) in escherichia coli - purification and characterization of the allergen. *Clinical & Experimental Allergy* **27**, 1307-1313.

Bae S, Cho J, Choi Y, Park K (2003). Analysis of the Arabidopsis nuclear proteome and its response to cold stress. *Plant J* **36**(5): 652-663.

Bahrman N, Plomion C, Petit RJ, Kremer A (1997) CONTRIBUTION OF TWO-DIMENSIONAL ELECTROPHORESIS OF PROTEINS TO MARITIME PINE GENETICS [Review]. *Annales des Sciences Forestieres* **54**, 225-236.

Baluska F, Salaj J, Mathur J, Braun M, Jasper F, Samaj J, Chua NH, Barlow PW, Volkmann D (2000) Root hair formation: F-actin-dependent tip growth is initiated by local assembly of profilin-supported F-actin meshworks accumulated within expansin-enriched bulges. *Dev Biol* **227**, 618-632.

Bedinger P (1992) The remarkable biology of pollen. *Plant Cell* **4**, 879-887.

Bedinger PA, Edgerton MD (1990) Developmental Staging of Maize Microspores Reveals a Transition in Developing Microspore Proteins. *Plant Physiology* **92**, 474-479.

Bevan M, Murphy G (1999) The small, the large and the wild: the value of comparison in plant genomics. *Trends Genet* **15**, 211-214.

Bjellqvist B, Ek K, Righetti PG, Gianazza E, Gorg A, Westermeier R, Postel W (1982) Isoelectric focusing in immobilized pH gradients: principle, methodology and some applications. *J Biochem Biophys Methods* **6**, 317-339.

Blomstedt CK, Knox RB, Singh MB (1996) Generative cells of lilium longiflorum possess translatable mrna and functional protein synthesis machinery. *Plant Molecular Biology* **31**, 1083-1086.

Blomstedt CK, Taylor PE, Singh MB, Knox RB (1997) The identification of an anther specific antigen in brassica species using a heterologous monoclonal antibody. *Annals of Botany* **80**, 197-204.

Bork P (2000) Powers and pitfalls in sequence analysis: the 70% hurdle. *Genome Res* **10**, 398-400.

Bork P, Dandekar T, Diaz-Lazcoz Y, Eisenhaber F, Huynen M, Yuan Y (1998) Predicting function: from genes to genomes and back. *J Mol Biol* **283**, 707-725.

Broadwater AH, RubinsteinAL, Chay CH, Klapper DG, Bedinger PA (1993) Zea mI, the maize homolog of the allergen-encoding Lol pI gene of rye grass. *Gene* **131**(2): 227-230.

Bufe A, Schramm G, Keown MB, Schlaak M, Becker WM (1995) Major allergen Phl p Vb in timothy grass is a novel pollen RNase. *FEBS Lett* **363**, 6-12.

Chamovitz DA, Segal D (2001) JAB1/CSN5 and the COP9 signalosome. A complex situation. *EMBO Rep* **2**, 96-101.

Chasan R, Walbot V (1993) Mechanisms of Plant Reproduction: Questions and Approaches. *Plant Cell* **5**, 1139-1146.

Chaurand P, Luetzenkirchen F, Spengler B (1999) Peptide and protein identification by matrix-assisted laser desorption ionization (MALDI) and MALDI-post-source decay time-of-flight mass spectrometry. *J Am Soc Mass Spectrom* **10**, 91-103.

Chevallet M, Santoni V, Poinas A, Rouquie D, Fuchs A, Kieffer S, Rossignol M, Lunardi J, Garin J, Rabilloud T (1998) New zwitterionic detergents improve the analysis of membrane proteins by two-dimensional electrophoresis. *Electrophoresis* **19**, 1901-1909.

Chivasa S, Ndimba K, Simon W J, Robertson D, Yu XL, Knox JP, Bolwell P, Slabas AR (2002). Proteomic analysis of the Arabidopsis thaliana cell wall. *Electrophoresis* **23**(11): 1754-65.

Cho HT, Kende H (1998) Tissue localization of expansins in deepwater rice. *Plant Journal* **15**, 805-812.

Chung YY, Kim SR, Kang HG, Noh YS, Park MC, Finkel D, An GH (1995) Characterization of two rice mads box genes homologous to globosa. *Plant Science* **109**, 45-56.

Clarke SR, Staiger CJ, Gibbon BC, Franklin-Tong VE (1998) A potential signaling role for profilin in pollen of Papaver rhoeas. *Plant Cell* **10**, 967-979.

Cordewener J, Bergervoet J, Liu CM (2000) Changes in protein synthesis and phosphorylation during microspore embryogenesis in Brassica napus. *Journal of Plant Physiology* **156**, 156-163.

Cordewener JHG, Busink R, Traas JA, Custers JBM, Dons HJM, Campagne MMV (1994) Induction of microspore embryogenesis in brassica napus l is accompanied by specific changes in protein synthesis. *Planta* **195**, 50-56.

Cordwell SJ, Nouwens AS, Verrills NM, Basseal DJ, Walsh BJ (2000) Subproteomics based upon protein cellular location and relative solubilities in conjunction with composite two-dimensional electrophoresis gels. *Electrophoresis* **21**, 1094-1103.

Cosgrove DJ (1997) Creeping walls, softening fruit, and penetrating pollen tubes - the growing roles of expansins. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 5504-5505.

Cosgrove DJ (2000) Loosening of plant cell walls by expansins [Review]. *Nature* **407**, 321-326.

Cosgrove DJ, Bedinger P, Durachko DM (1997) Group i allergens of grass pollen as cell wall-loosening agents. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 6559-6564.

Cozzone AJ (1998) Post-translational modification of proteins by reversible phosphorylation in prokaryotes. *Biochimie* **80**, 43-48.

Damerval C, Ziv M, Granier F, de Vienne D (1988) Two dimensional electrophoresis in plant biology. In 'Advances in Electrophoresis'. (Eds A Chrambach, MJ Dunn and BJ Radola) pp. 263-340. (VCH: Weinheim, Germany)

de Saizieu A, Certa U, Warrington J, Gray C, Keck W, Mous J (1998) Bacterial transcript imaging by hybridization of total RNA to oligonucleotide arrays. *Nat Biotechnol* **16**, 45-48.

Delseny M (2003) Towards an accurate sequence of the rice genome. *Curr Opin Plant Biol* **6**, 101-105.

Deng XW, Dubiel W, Wei N, Hofmann K, Mundt K (2000) Unified nomenclature for the COP9 signalosome and its subunits: an essential regulator of development. *Trends Genet* **16**, 289.

Diano M, Bevic A (1997) Production of Highly Specific Polyclonal Antibodies Using a Combination of 2D Electrophoresis and Nitrocellulose-Bound Antigen. In 'The Protein Protocols Handbook'. (Ed. JM Walker) pp. 703-710. (Humana Press: Totowa, NJ)

Dolecek C, Vrtala S, Laffer S, Steinberger P, Kraft D, Scheiner O, Valenta R (1993) Molecular characterization of Phl p II, a major timothy grass (Phleum pratense) pollen allergen. *FEBS Lett* **335**, 299-304.

Drenckhahn D, Jons T, Schmitz F (1993) Production of polyclonal antibodies against proteins and peptides. *Methods Cell Biol* **37**, 7-56.

Dubey H, Grover A (2001) Current initiatives in proteomics research: The plant perspective. *Current Science* **80**, 262-269.

Edman P (1949) A method for the determination of the amino acid sequence of peptides. *Arch Biochem Biophys* **22**, 475-483.

Edman P, Begg G (1967) A protein sequenator. *Eur J Biochem* **1**, 80-91.

Eisen MB, Brown PO (1999) DNA arrays for analysis of gene expression. *Methods Enzymol* **303**, 179-205.

Eisenstein E, Gilliland GL, Herzberg O, Moult J, Orban J, Poljak RJ, Banerjei L, Richardson D, Howard AJ (2000) Biological function made crystal clear - annotation of hypothetical proteins via structural genomics [Review]. *Current Opinion in Biotechnology* **11**, 25-30.

Erickson RO (1948) Cytological and growth correlations in the flower bud and anther of Lilium longiflorum. *American Journal of Botany* **35**, 729-739.

Fan Z, Armstrong KC, Keller WA (1988) Development of Microspores Invivo and Invitro in Brassica-Napus L. *Protoplasma* **147**, 191-199.

Faure JE, Digonnet C, Dumas C (1994) An in vitro system for adhesion and fusion of maize gametes. *Science* **263**, 1598-1600.

Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**, 64-71.

Ferrari BC, Attfield PV, Veal DA, Bell PJ (2003) Application of the novel fluorescent dye Beljian red to the differentiation of Giardia cysts. *J Microbiol Methods* **52**, 133-135.

Freidhoff LR, Ehrlich-Kautzky E, Grant JH, Meyers DA, Marsh DG (1986) A study of the human immune response to Lolium perenne (rye) pollen and its components, Lol p I and Lol p II (rye I and rye II). I. Prevalence of reactivity to the allergens and correlations among skin test, IgE antibody, and IgG antibody data. *J Allergy Clin Immunol* **78**, 1190-1201.

Fu H, Girod PA, Doelling JH, van Nocker S, Hochstrasser M, Finley D, Vierstra RD (1999) Structure and functional analysis of the 26S proteasome subunits from plants. *Mol Biol Rep* **26**, 137-146.

Gale MD, Devos KM (1998) Comparative genetics in the grasses. *Proc Natl Acad Sci U S A* **95**, 1971-1974.

Garrels JI, McLaughlin CS, Warner JR, Futcher B, Latter GI, Kobayashi R, Schwender B, Volpe T, Anderson DS, Mesquita-Fuentes R, Payne WE (1997) Proteome studies of Saccharomyces cerevisiae: identification and characterization of abundant proteins. *Electrophoresis* **18**, 1347-1360.

Gasser CS (1991) Molecular studies on the differentiation of floral organs. *Annual Review of Plant Physiology and Plant Molecular Biology* **42**, 621-649.

Geltz NR, Russell SD (1988) Two-Dimensional Electrophoretic Studies of the Proteins and Polypeptides in Mature Pollen Grains and the Male Germ Unit of Plumbago-Zeylanica. *Plant Physiology* **88**, 764-769.

Gevaert K, Vandekerckhove J (2000) Protein identification methods in proteomics. *Electrophoresis* **21**, 1145-1154.

Gharahdaghi F, Weinberg CR, Meagher DA, Imai BS, Mische SM (1999) Mass spectrometric identification of proteins from silver-stained polyacrylamide gel: a method for the removal of silver ions to enhance sensitivity. *Electrophoresis* **20**, 601-605.

Goff SA (1999) Rice as a model for cereal genomics. *Curr Opin Plant Biol* **2**, 86-89.

Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, Hadley D, Hutchison D, Martin C, Katagiri F, Lange BM, Moughamer T, Xia Y, Budworth P, Zhong J, Miguel T, Paszkowski U, Zhang S, Colbert M, Sun WL, Chen L, Cooper B, Park S, Wood TC, Mao L, Quail P, Wing R, Dean R, Yu Y, Zharkikh A, Shen R, Sahasrabudhe S, Thomas A, Cannings R, Gutin A, Pruss D, Reid J, Tavtigian S, Mitchell J, Eldredge G, Scholl T, Miller RM, Bhatnagar S, Adey N, Rubano T, Tusneem N, Robinson R, Feldhaus J, Macalma T, Oliphant A, Briggs S (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). *Science* **296**, 92-100.

Goldberg RB (1988) Plants: novel developmental processes. *Science* **240**, 1460-1467.

Gonzalez F, Delahodde A, Kodadek T, Johnston SA (2002) Recruitment of a 19S proteasome subcomplex to an activated promoter. *Science* **296**, 548-550.

Gorg A, Boguth G, Obermaier C, Posch A, Weiss W (1995) Two-dimensional polyacrylamide gel electrophoresis with immobilized pH gradients in the first dimension (IPG-Dalt): the state of the art and the controversy of vertical versus horizontal systems. *Electrophoresis* **16**, 1079-1086.

Gorg A, Obermaier C, Boguth G, Harder A, Scheibe B, Wildgruber R, Weiss W (2000) The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* **21**, 1037-1053.

Gorg A, Postel W, Gunther S (1988) The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* **9**, 531-546.

Gould KS, Lord EM (1988) Growth of Anthers in Lilium-Longiflorum - a Kinematic Analysis. *Planta* **173**, 161-171.

Griffith IJ, Smith PM, Pollock J, Theerakulpisut P, Avjioglu A, Davies S, Hough T, Singh MB, Simpson RJ, Ward LD, et al. (1991) Cloning and sequencing of Lol pI, the major allergenic protein of rye- grass pollen. *FEBS Lett* **279**, 210-215.

Grobe K, Becker WM, Schlaak M, Petersen A (1999) Grass group I allergens (beta-expansins) are novel, papain-related proteinases. *Eur J Biochem* **263**, 33-40.

Grobe K, Poppelmann M, Becker WM, Petersen A (2002) Properties of group I allergens from grass pollen and their relation to cathepsin B, a member of the C1 family of cysteine proteinases. *Eur J Biochem* **269**, 2083-2092.

Guerreiro N, Redmond JW, Rolfe BG, Djordjevic MA (1997) New Rhizobium leguminosarum flavonoid-induced proteins revealed by proteome analysis of differentially displayed proteins. *Mol Plant Microbe Interact* **10**, 506-516.

Gygi SP, Aebersold R (2000) Mass spectrometry and proteomics [In Process Citation]. *Curr Opin Chem Biol* **4**, 489-494.

Gygi SP, Corthals GL, Zhang Y, Rochon Y, Aebersold R (2000) Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc Natl Acad Sci U S A* **97**, 9390-9395.

Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R (1999a) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* **17**, 994-999.

Gygi SP, Rochon Y, Franza BR, Aebersold R (1999b) Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol* **19**, 1720-1730.

Hajduch M, Rakwal R, Agrawal GK, Yonekura M, Pretova A (2001) High-resolution two-dimensional electrophoresis separation of proteins from metal-stressed rice (Oryza sativa L.) leaves: drastic reductions/fragmentation of ribulose-1,5-bisphosphate carboxylase/oxygenase and induction of stress-related proteins. *Electrophoresis* **22**, 2824-2831.

Hanson DD, Hamilton DA, Travis JL, Bashe DM, Mascarenhas JP (1989) Characterization of a pollen-specific cDNA clone from Zea mays and its expression. *Plant Cell* **1**, 173-179.

Harry JL, Wilkins MR, Herbert BR, Packer NH, Gooley AA, Williams KL (2000) Proteomics: capacity versus utility. *Electrophoresis* **21**, 1071-1081.

Haynes PA, Gygi SP, Figeys D, Aebersold R (1998) Proteome analysis: biological assay or data archive? *Electrophoresis* **19**, 1862-1871.

Haynes PA, Yates JR, 3rd (2000) Proteome profiling-pitfalls and progress. *Yeast* **17**, 81-87.

Heazlewood JL, Howell KA, Whelan J, Millar AH (2003) Towards an analysis of the rice mitochondrial proteome. *Plant Physiol* **132**, 230-242.

Hellmann H, Estelle M (2002) Plant development: regulation by protein degradation. *Science* **297**, 793-797.

Henzel WJ, Billeci TM, Stults JT, Wong SC, Grimley C, Watanabe C (1993) Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proc Natl Acad Sci U S A* **90**, 5011-5015.

Herbert B (1999) Advances in protein solubilisation for two-dimensional electrophoresis. *Electrophoresis* **20**, 660-663.

Herbert BR, Harry JL, Packer NH, Gooley AA, Pedersen SK, Williams KL (2001) What place for polyacrylamide in proteomics? *Trends Biotechnol* **19**, S3-9.

Herskovits TT, Jaillet H, Gadegbeku B (1970) On the structural stability and solvent denaturation of proteins. II. Denaturation by the ureas. *J Biol Chem* **245**, 4544-4550.

Hewick RM, Hunkapiller MW, Hood LE, Dreyer WJ (1981) A gas-liquid solid phase peptide and protein sequenator. *J Biol Chem* **256**, 7990-7997.

Hieter P, Boguski M (1997) Functional genomics: it's all how you read it. *Science* **278**, 601-602.

Hihara Y, Hara C, Uchimiya H (1996) Isolation and characterization of two cdna clones for mrnas that are abundantly expressed in immature anthers of rice (oryza sativa l.). *Plant Molecular Biology* **30**, 1181-1193.

Hird DL, Worrall D, Hodge R, Smartt S, Paul W, Scott R (1993) The anther-specific protein encoded by the Brassica napus and Arabidopsis thaliana A6 gene displays similarity to beta-1,3-glucanases. *Plant J* **4**, 1023-1033.

Holtorf H, Guitton MC, Reski R (2002) Plant functional genomics. *Naturwissenschaften* **89**, 235-249.

Hood L (2003) Systems biology: integrating technology, biology, and computation. *Mech Ageing Dev* **124**, 9-16.

Horgan GW, Glasbey CA (1995) Uses of digital image analysis in electrophoresis. *Electrophoresis* **16**, 298-305.

Huber LA (2003) Is proteomics heading in the wrong direction? *Nat Rev Mol Cell Biol* **4**, 74-80.

Imin N (2002) Molecular mechanisms causing cold induced male sterility in rice: A proteomic investigation. PhD thesis, Australian national University.

Imin N, Kerim T, Weinman JJ, Rolfe BG (2001) Characterisation of rice anther proteins expressed at the young microspore stage. *Proteomics* **1**, 1149-1161.

Izawa T, Shimamoto K (1996) Becoming a model plant - the importance of rice to plant science. *Trends in Plant Science* **1**, 95-99.

Jacobs DI, van der Heijden R, Verpoorte R (2000) Proteomics in plant biotechnology and secondary metabolism research. *Phytochemical Analysis* **11**, 277-287.

Jeffery CJ (1999) Moonlighting proteins. *Trends in Biochemical Sciences* **24**, 8-11.

Jeffery CJ (2003) Moonlighting proteins: old proteins learning new tricks. *Trends in Genetics* **19**, 415-417.

Jensen ON, Wilm M, Shevchenko A, Mann M (1999) Peptide sequencing of 2-DE gel-isolated proteins by nanoelectrospray tandem mass spectrometry. *Methods Mol Biol* **112**, 571-588.

Jeon JS, Chung YY, Lee S, Yi GH, Oh BG, An GH (1999) Isolation and characterization of an anther-specific gene, RA8, from rice (Oryza sativa L.). *Plant Molecular Biology* **39**, 35-44.

Jong K (1997) 'Laboratory manual of plant cytological techniques.' (Royal Botanic Garden Edinburgh: Edinburgh)

Kalinowski A, Winiarczyk K, Radlowski M (2002) Pollen coat proteins after two-dimensional gel electrophoresis and pollen wall ultrastructure of Secale cereale and *Festuca pratensis*. *Sex Plant Reprod* **15**, 75-83.

Kalinowski A, Winiarczyk K, Wojciechowska B (2001) Pollen proteins after two-dimensional gel electrophoresis and pollen morphology of the amphiploids Aegilops kotschyi and Ae. variabilis with Secale cereale. *Sexual Plant Reproduction* **14**, 153-161.

Kamalay JC, Goldberg RB (1980) Regulation of structural gene expression in tobacco. *Cell* **19**, 935-946.

Kamalay JC, Goldberg RB (1984) Organ-specific nuclear RNAs in tobacco. *Proc Natl Acad Sci U S A* **81**, 2801-2805.

Kamo M, Kawakami T, Miyatake N, Tsugita A (1995) Separation and characterization of Arabidopsis thaliana proteins by two- dimensional gel electrophoresis. *Electrophoresis* **16**, 423-430.

Kang HG, Jang S, Chung JE, Cho YG, An G (1997) Characterization of two rice MADS box genes that control flowering time. *Mol Cells* **7**, 559-566.

Karas M, Hillenkamp F (1988) Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* **60**, 2299-2301.

Karniol B, Chamovitz DA (2000) The COP9 signalosome: from light signaling to general developmental regulation and back. *Curr Opin Plant Biol* **3**, 387-393.

Kaul S, Koo HL, Jenkins J, Rizzo M, Rooney T, Tallon LJ, Feldblyum T, Nierman W, Benito MI, Lin XY, Town CD, Venter JC, Fraser CM, Tabata S, Nakamura Y, Kaneko T, Sato S, Asamizu E, Kato T, Kotani H, Sasamoto S, Ecker JR, Theologis A, Federspiel NA, Palm CJ, et al. (2000) Analysis of the

genome sequence of the flowering plant Arabidopsis thaliana [Review]. *Item Corporate Author Arabidopsis Genome Initiative* **408**, 796-815.

Kersten B, Burkle L, Kuhn EJ, Giavalisco P, Konthur Z, Lueking A, Walter G, Eickhoff H, Schneider U (2002) Large-scale plant proteomics. *Plant Mol Biol* **48**, 133-141.

Kikuchi S, Satoh K, Nagata T, Kawagashira N, Doi K, Kishimoto N, Yazaki J, Ishikawa M, Yamada H, Ooka H, Hotta I, Kojima K, Namiki T, Ohneda E, Yahagi W, Suzuki K, Li CJ, Ohtsuki K, Shishiki T, Otomo Y, Murakami K, Iida Y, Sugano S, Fujimura T, Suzuki Y, Tsunoda Y, Kurosaki T, Kodama T, Masuda H, Kobayashi M, Xie Q, Lu M, Narikawa R, Sugiyama A, Mizuno K, Yokomizo S, Niikura J, Ikeda R, Ishibiki J, Kawamata M, Yoshimura A, Miura J, Kusumegi T, Oka M, Ryu R, Ueda M, Matsubara K, Kawai J, Carninci P, Adachi J, Aizawa K, Arakawa T, Fukuda S, Hara A, Hashidume W, Hayatsu N, Imotani K, Ishii Y, Itoh M, Kagawa I, Kondo S, Konno H, Miyazaki A, Osato N, Ota Y, Saito R, Sasaki D, Sato K, Shibata K, Shinagawa A, Shiraki T, Yoshino M, Hayashizaki Y (2003) Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice. *Science* **301**, 376-379.

Kimura T, Todokoro M, Kuroume T, Tatemo K, Matsumura T (1969) Rice pollen asthma. II. Cross antigenicity between rice pollen and other grass pollens. *Arerugi* **18**, 1005-1016.

King TP, Hoffman D, Lowenstein H, Marsh DG, Platts-Mills TA, Thomas W (1994) Allergen nomenclature. WHO/IUIS Allergen Nomenclature Subcommittee. *Int Arch Allergy Immunol* **105**, 224-233.

Klose J (1975) Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. *Humangenetik* **26**, 231-243.

Klose J, Kobalz U (1995) Two-dimensional electrophoresis of proteins: an updated protocol and implications for a functional analysis of the genome. *Electrophoresis* **16**, 1034-1059.

Knox RB, Zee SY, Blomstedt C, Singh MB (1993) MALE GAMETES AND FERTILIZATION IN ANGIOSPERMS [Review]. *New Phytologist* **125**, 679-694.

Knudsen KA (1985) Proteins transferred to nitrocellulose for use as immunogens. *Anal Biochem* **147**, 285-288.

Kobayashi T, Kobayashi E, Sato S, Hotta Y, Miyajima N, Tanaka A, Tabata S (1994) Characterization of cDNAs induced in meiotic prophase in lily microsporocytes. *DNA Res* **1**, 15-26.

Koller A, Washburn MP, Lange BM, Andon NL, Deciu C, Haynes PA, Hays L, Schieltz D, Ulaszek R, Wei J, Wolters D, Yates JR, 3rd (2002) From the

Cover: Proteomic survey of metabolic pathways in rice. *Proc Natl Acad Sci U S A* **99**, 11969-11974.

Koltunow AM (1990) Different Temporal anad Spatial Gene expression Patterns occur during Anther Development. *The Plant Cell* **2**, 1201-1224.

Komatsu S, Kajiwara H, Hirano H (1993) A rice protein library - a data-file of rice proteins separated by 2-dimensional electrophoresis. *Theoretical & Applied Genetics* **86**, 935-942.

Kovar DR, Drobak BK, Staiger CJ (2000) Maize profilin isoforms are functionally distinct. *Plant Cell* **12**, 583-598.

Kowyama Y, Saba T, Tsuji T, Kawase T (1994) Specific developmental stages of gametogenesis for radiosensitivity and mutagenesis in rice. *Euphytica* **80**, 27-38.

Kranz E, Dresselhaus T (1996) IN VITRO FERTILIZATION WITH ISOLATED HIGHER PLANT GAMETES [Review]. *Trends in Plant Science* **1**, 82-89.

Kranz E, Lorz H (1993) In Vitro Fertilization with Isolated, Single Gametes Results in Zygotic Embryogenesis and Fertile Maize Plants. *Plant Cell* **5**, 739-746.

Kurland CG (1991) Codon bias and gene expression. *FEBS Lett* **285**, 165-169.

Lay Jr. JO, Holland RD (2000) Rapid identification of bacteria based on spectral patterns using MALDI-TOF MS. In 'Methods in Molecular Biology'. (Ed. JR Chapman) pp. 461-487. (Humana Press: Totowa, New Jersey)

Lemkin PF (1997) Comparing two-dimensional electrophoretic gel images across the Internet. *Electrophoresis* **18**, 461-470.

Li LC, Cosgrove DJ (2001) Grass group I pollen allergens (beta-expansins) lack proteinase activity and do not cause wall loosening via proteolysis. *Eur J Biochem* **268**, 4217-4226.

Li L-C, Shieh MW, Cosgrove DJ (2003) Group 2 and group 3 allergens of grass pollen have plant cell wall-loosening activity characteristic of b-expansins. In 'American Society of Plant Biologists'. Honolulu, Hawaii USA

Liebler CD (2002) 'Introduction to proteomics: Tools for the new biology.' (Humana Press: Totowa, NJ)

Lifschitz E (1988) Molecular markers for the floral program. *Flowering Newsl* **6**, 16-20.

Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, Garvik BM, Yates JR, 3rd (1999) Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol* **17**, 676-682.

Lockhart DJ, Winzeler EA (2000) Genomics, gene expression and DNA arrays. *Nature* **405**, 827-836.

Lopez MF, Berggren K, Chernokalskaya E, Lazarev A, Robinson M, Patton WF (2000) A comparison of silver stain and SYPRO Ruby Protein Gel Stain with respect to protein detection in two-dimensional gels and identification by peptide mass profiling. *Electrophoresis* **21**, 3673-3683.

Mann M, Hendrickson RC, Pandey A (2001) Analysis of proteins and proteomes by mass spectrometry. *Annu Rev Biochem* **70**, 437-473.

Mann M, Jensen ON (2003) Proteomic analysis of post-translational modifications. *Nat Biotechnol* **21**, 255-261.

Mann M, Ong SE, Gronborg M, Steen H, Jensen ON, Pandey A (2002) Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome. *Trends Biotechnol* **20**, 261-268.

Mann M, Pandey A (2001) Use of mass spectrometry-derived data to annotate nucleotide and protein sequence databases. *Trends Biochem Sci* **26**, 54-61.

Mann M, Wilm M (1994) Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal Chem* **66**, 4390-4399.

Marknell DeWitt A, Niederberger V, Lehtonen P, Spitzauer S, Sperr WR, Valent P, Valenta R, Lidholm J (2002) Molecular and immunological characterization of a novel timothy grass (Phleum pratense) pollen allergen, Phl p 11. *Clin Exp Allergy* **32**, 1329-1340.

Martin SE, Shabanowitz J, Hunt DF, Marto JA (2000) Subfemtomole MS and MS/MS peptide sequence analysis using nano-HPLC micro-ESI fourier transform ion cyclotron resonance mass spectrometry. *Anal Chem* **72**, 4266-4274.

Mascarenhas JP (1990) Gene activity during pollen development. . *Annual review of Plant Physiology and Plant Molecular biology* **41**, 317-338.

Mathesius U, Imin N, Chen H, Djordjevic MA, Weinman JJ, Natera SH, Morris AC, Kerim T, Paul S, Menzel C, Weiller GF, Rolfe BG (2002) Evaluation of proteome reference maps for cross-species identification of proteins by peptide mass fingerprinting. *Proteomics* **2**, 1288-1303.

Mathesius U, Keijzers G, Natera SH, Weinman JJ, Djordjevic MA, Rolfe BG (2001) Establishment of a root proteome reference map for the model legume Medicago truncatula using the expressed sequence tag database for peptide mass fingerprinting. *Proteomics* **1**, 1424-1440.

Matsudaira P (1987) Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. *J Biol Chem* **262**, 10035-10038.

Matsumura T, Kimura T, Tatemo K, Kato E, Todokoro M, Nakajima S, Kuroume T (1969) Rice pollen asthma. *Journal of Asthma Research* **7**, 7-16.

Matsuo T, Hoshikawa K. (1993) 'Science of the Rice Plant.' (Food and Agriculture Policy Recearch Center, Tokyo)

Matsushima V, Manaka T (1956) Developmental process of the young panicle of rice and its diagnosis in rice plants, developmental process of young panicles in all stems, its standard and characteristics of each developmental stages. *Assoc Agric. Technol., Tokyo,*, 1-58.

Matthysrochon E, Vergne P, Detchepare S, Dumas C (1987) Male Germ Unit Isolation from 3 Tricellular Pollen Species - Brassica-Oleracea, Zea-Mays, and Triticum-Aestivum. *Plant Physiology* **83**, 464-466.

Mayfield JA, Fiebig A, Johnstone SE, Preuss D (2001) Gene families from the Arabidopsis thaliana pollen coat proteome. *Science* **292**, 2482-2485.

McCormack AL, Schieltz DM, Goode B, Yang S, Barnes G, Drubin D, Yates JR, 3rd (1997) Direct analysis and identification of proteins in mixtures by LC/MS/MS and database searching at the low-femtomole level. *Anal Chem* **69**, 767-776.

McCormick S (1993) Male Gametophyte Development. *Plant Cell* **5**, 1265-1275.

McCurdy DW, Kovar DR, Staiger CJ (2001) Actin and actin-binding proteins in higher plants. *Protoplasma* **215**, 89-104.

Medzihradszky KF, Campbell JM, Baldwin MA, Falick AM, Juhasz P, Vestal ML, Burlingame AL (2000) The characteristics of peptide collision-induced dissociation using a high-performance MALDI-TOF/TOF tandem mass spectrometer. *Anal Chem* **72**, 552-558.

Molloy MP, Herbert BR, Williams KL, Gooley AA (1999) Extraction of Escherichia coli proteins with organic solvents prior to two-dimensional electrophoresis. *Electrophoresis* **20**, 701-704.

Moons A, Prinsen E, Bauw G, Van Montagu M (1997) Antagonistic effects of abscisic acid and jasmonates on salt stress-inducible transcripts in rice roots. *Plant Cell* **9**, 2243-2259.

Mori T, Tanaka I (2000) Isolation of the ftsZ gene from plastid-deficient generative cells of Lilium longiflorum. *Protoplasma* **214**, 57-64.

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

Munchbach M, Quadroni M, Miotto G, James P (2000) Quantitation and facilitated de novo sequencing of proteins by isotopic N-terminal labeling of peptides with a fragmentation-directing moiety. *Anal Chem* **72**, 4047-4057.

Natera SH, Guerreiro N, Djordjevic MA (2000) Proteome analysis of differentially displayed proteins as a tool for the investigation of symbiosis. *Mol Plant Microbe Interact* **13**, 995-1009.

Nelson DL, Cox MM (2000) 'Princples of Biochemistry.' (Worth Publishers: New York)

Neuhoff V, Arold N, Taube D, Ehrhardt W (1988) Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. *Electrophoresis* **9**, 255-262.

Nguyen-Quoc B, Foyer CH (2001) A role for 'futile cycles' involving invertase and sucrose synthase in sucrose metabolism of tomato fruit. *J Exp Bot* **52**, 881-889.

Nielsen H, Engelbrecht J, Brunak S, von Heijne G (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng* **10**, 1-6.

Nishiyama I (1982) Male sterility caused by cooling treatment at the young microspore stage in rice plants. *Japanese Journal of Crop Science* **51**, 462-469.

O'Farrell PH (1975) High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* **250**, 4007-4021.

Oldenhof MT, de Groot PF, Visser JH, Schrauwen JA, Wullems GJ (1996) Isolation and characterization of a microspore-specific gene from tobacco. *Plant Mol Biol* **31**, 213-225.

Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **1**, 376-386.

Ong SE, Pandey A (2001) An evaluation of the use of two-dimensional gel electrophoresis in proteomics. *Biomol Eng* **18**, 195-205.

Pandey A, Lewitter F (1999) Nucleotide sequence databases: a gold mine for biologists. *Trends Biochem Sci* **24**, 276-280.

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

Patton WF (2002) Detection technologies in proteome analysis. *J Chromatogr B Analyt Technol Biomed Life Sci* **771**, 3-31.

Pearson WR, Lipman DJ (1988) Improved tools for biological sequence comparison. *Proc Natl Acad Sci U S A* **85**, 2444-2448.

Piatigorsky J (1998) Multifunctional lens crystallins and corneal enzymes. More than meets the eye. *Ann N Y Acad Sci* **842**, 7-15.

Porubleva L, Vander Velden K, Kothari S, Oliver DJ, Chitnis PR (2001) The proteome of maize leaves: use of gene sequences and expressed sequence tag data for identification of proteins with peptide mass fingerprints. *Electrophoresis* **22**, 1724-1738.

Quadroni M, James P (1999) Proteomics and automation. *Electrophoresis* **20**, 664-677.

Rabilloud T (1999) Solubilization of proteins in 2-D electrophoresis. In '2-D Proteome Analysis Protocols'. (Ed. AJ Link) pp. 9-19. (Humana press: Totowa)

Rabilloud T, Vuillard L, Gilly C, Lawrence JJ (1994) Silver-staining of proteins in polyacrylamide gels: a general overview. *Cell Mol Biol (Noisy-le-grand)* **40**, 57-75.

Raghavan V (1988) Anther and Pollen Development in Rice (Oryza Sativa). *American Journal of Botany* **75**, 183-196.

Raikhel NV, Coruzzi GM (2003) Plant systems biology. *Plant Physiol* **132**, 403.

Rakwal R, Agrawal GK, Kubo A, Yonekura M, Tamogami S, Saji H, Iwahashi H (2003) Defense/stress responses elicited in rice seedlings exposed to the gaseous air pollutant sulfur dioxide. *Environmental and Experimental Botany* **49**, 223-235.

Rakwal R, Agrawal GK, Yonekura M (1999) Separation of proteins from stressed rice (Oryza sativa L.) leaf tissues by two-dimensional polyacrylamide gel electrophoresis: induction of pathogenesis-related and cellular protectant proteins by jasmonic acid, UV irradiation and copper chloride. *Electrophoresis* **20**, 3472-3478.

Rakwal R, Komatsu S (2000) Role of jasmonate in the rice (Oryza sativa L.) self-defense mechanism using proteome analysis [In Process Citation]. *Electrophoresis* **21**, 2492-2500.

Ramachandran S, Christensen HE, Ishimaru Y, Dong CH, Chao-Ming W, Cleary AL, Chua NH (2000) Profilin plays a role in cell elongation, cell shape maintenance, and flowering in arabidopsis. *Plant Physiol* **124**, 1637-1647.

Raven PH, Evert RF, Eichhorn SE (1992) 'Biology of Plants.' (Worth Publishers: New York)

Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Seraphin B (1999) A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol* **17**, 1030-1032.

Righetti PG, Macelloni C (1982) New polyacrylamide matrices for drift-free isoelectric focusing. *J Biochem Biophys Methods* **6**, 1-15.

Roberts AM, Bevan LJ, Flora PS, Jepson I, Walker MR (1993a) Nucleotide sequence of cDNA encoding the group II allergen of cocksfoot/orchard grass (Dactylis glomerata), Dac g II. *Allergy* **48**, 615-623.

Roberts JK (2002) Proteomics and a future generation of plant molecular biologists. *Plant Mol Biol* **48**, 143-154.

Roberts MR, Foster GD, Blundell RP, Robinson SW, Kumar A, Draper J, Scott R (1993b) Gametophytic and sporophytic expression of an anther-specific Arabidopsis thaliana gene. *Plant J* **3**, 111-120.

Rogers HJ, Maund SL, Johnson LH (2001) A beta-galactosidase-like gene is expressed during tobacco pollen development. *J Exp Bot* **52**, 67-75.

Rossignol M (2001) Analysis of the plant proteome. *Curr Opin Biotechnol* **12**, 131-134.

Salekdeh GH, Siopongco J, Wade LJ, Ghareyazie B, Bennett J (2002a) Proteomic analysis of rice leaves during drought stress and recovery. *Proteomics* **2**, 1131-1145.

Salekdeh GH, Siopongco J, Wade LJ, Ghareyazie B, Bennett J (2002b) A proteomic approach to analyzing drought- and salt-responsiveness in rice. *Field Crops Research* **76**, 199-219.

Salse J, Piegu B, Cooke R, Delseny M (2002) Synteny between Arabidopsis thaliana and rice at the genome level: a tool to identify conservation in the ongoing rice genome sequencing project. *Nucleic Acids Res* **30**, 2316-2328.

Santoni V, Molloy M, Rabilloud T (2000) Membrane proteins and proteomics: un amour impossible? *Electrophoresis* **21**, 1054-1070.

Sasaki T, Burr B (2000) International Rice Genome Sequencing Project: the effort to completely sequence the rice genome. *Curr Opin Plant Biol* **3**, 138-141.

Satake T, Hayase H (1970) *Proc.CropSci.Soc. Jpn.* **39**, 468-473.

Schrauwen J, de Groot PF, van Herpen M, van der Lee T, Reynen WH, Weterings K, Wullems G (1990) Stage-related expression of mRNAs during pollen development in lily and tobacco. *Planta* **182**, 298-304.

Scott R, Dagless E, Hodge R, Paul W, Soufleri I, Draper J (1991) Patterns of gene expression in developing anthers of Brassica napus. *Plant Mol Biol* **17**, 195-207.

Scott R, Hodge, R., Paul, D. and Draper, J. (1991) The molecular biology of anther differentiation. *Plant Science* **80**, 167-191.

Shaw MM, Riederer BM (2003) Sample preparation for two-dimensional gel electrophoresis. *Proteomics* **3**, 1408-1417.

Shen S, Matsubae M, Takao T, Tanaka N, Komatsu S (2002) A proteomic analysis of leaf sheaths from rice. *J Biochem (Tokyo)* **132**, 613-620.

Shen S, Sharma A, Komatsu S (2003) Characterization of Proteins Responsive to Gibberellin in the Leaf- Sheath of Rice (Oryza sativa L.) Seedling Using Proteome Analysis. *Biol Pharm Bull* **26**, 129-136.

Shevchenko A, Sunyaev S, Loboda A, Bork P, Ens W, Standing KG (2001) Charting the proteomes of organisms with unsequenced genomes by MALDI-quadrupole time-of-flight mass spectrometry and BLAST homology searching. *Anal Chem* **73**, 1917-1926.

Shevchenko A, Wilm M, Vorm O, Mann M (1996) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal Chem* **68**, 850-858.

Sidoli A, Tamborini E, Giuntini I, Levi S, Volonte G, Paini C, De Lalla C, Siccardi AG, Baralle FE, Galliani S, et al. (1993) Cloning, expression, and immunological characterization of recombinant Lolium perenne allergen Lol p II. *J Biol Chem* **268**, 21819-21825.

Simoens CR, Peleman J, Valvekens D, Van Montagu M, Inze D (1988) Isolation of genes expressed in specific tissues of Arabidopsis thaliana by differential screening of a genomic library. *Gene* **67**, 1-11.

Singh M, Bhalla PL, Xu H, Singh MB (2003) Isolation and characterization of a flowering plant male gametic cell-specific promoter. *FEBS Lett* **542**, 47-52.

Singh MB, Oneill PM, Knox RB (1985) Initiation of Postmeiotic Beta-Galactosidase Synthesis During Microsporogenesis in Oilseed Rape. *Plant Physiology* **77**, 225-228.

Skylas DJ, Mackintosh JA, Cordwell SJ, Basseal DJ, Walsh BJ, Harry J, Blumenthal C, Copeland L, Wrigley CW, Rathmell W (2000) Proteome approach to the characterisation of protein composition in the developing and mature wheat-grain endosperm. *Journal of Cereal Science* **32**, 169-188.

Smith AG, Gasser CS, Budelier KA, Fraley RT (1990) Identification and characterization of stamen- and tapetum-specific genes from tomato. *Mol Gen Genet* **222**, 9-16.

Smith DL, Gross KC (2000) A family of at least seven beta-galactosidase genes is expressed during tomato fruit development. *Plant Physiol* **123**, 1173-1183.

Smith PM, Avjioglu A, Ward LR, Simpson RJ, Knox RB, Singh MB (1994a) Isolation and characterization of group-I isoallergens from Bermuda grass pollen. *Int Arch Allergy Immunol* **104**, 57-64.

Smith PM, Ong EK, Knox RB, Singh MB (1994b) Immunological relationships among group I and group V allergens from grass pollen. *Mol Immunol* **31**, 491-498.

Stinson JR, Eissenberg AJ, Willing RP, Pe M, E., Hanson DD, Mascarenhas JP (1987) Genes expressed in the male gametophyte flowering plants and their isolation. *Plant Physiol* **83**, 442-447.

Sturm A (1999) Invertases. Primary structures, functions, and roles in plant development and sucrose partitioning. *Plant Physiology* **121**, 1-7.

Suphioglu C (2000) What are the important allergens in grass pollen that are linked to human allergic disease? *Clin Exp Allergy* **30**, 1335-1341.

Tebbutt SJ, Rogers HJ, Lonsdale DM (1994) Characterization of a tobacco gene encoding a pollen-specific polygalacturonase. *Plant Mol Biol* **25**, 283-297.

Thangavelu M, Belostotsky D, Bevan MW, Flavell RB, Rogers HJ, Lonsdale DM (1993) Partial characterization of the Nicotiana tabacum actin gene family: evidence for pollen-specific expression of one of the gene family members. *Mol Gen Genet* **240**, 290-295.

Theerakulpisut P, Xu H, Singh MB, Pettitt JM, Knox RB (1991) Isolation and developmental expression of Bcp1, an anther-specific cDNA clone in Brassica campestris. *Plant Cell* **3**, 1073-1084.

Theunis CH, Pierson ES, Cresti M (1991) Isolation of Male and Female Gametes in Higher-Plants. *Sexual Plant Reproduction* **4**, 145-154.

Thiellement H, Bahrman N, Damerval C, Plomion C, Rossignol M, Santoni V, de Vienne D, Zivy M (1999) Proteomics for genetic and physiological studies in plants. *Electrophoresis* **20**, 2013-2026.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**, 4876-4882.

Ticha M, Pacakova V, Stulik K (2002) Proteomics of allergens. *J Chromatogr B Analyt Technol Biomed Life Sci* **771**, 343- 353.

Toriyama K, Hanaoka K, Okada T, Watanabe M (1998) Molecular cloning of a cdna encoding a pollen extracellular protein as a potential source of a pollen allergen in brassica rapa. *FEBS Letters* **424**, 234-238.

Towbin H, Staehelin T, Gordon J (1992) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. 1979. *Biotechnology* **24**, 145-149.

Treacy BK, Hattori J, Prud'homme I, Barbour E, Boutilier K, Baszczynski CL, Huang B, Johnson DA, Miki BL (1997) Bnm1, a Brassica pollen-specific gene. *Plant Mol Biol* **34**, 603-611.

Triantafillidou D, Georgatsos JG (2001) Barley beta-galactosidase: structure, function, heterogeneity, and gene origin. *J Protein Chem* **20**, 551-562.

Tsai YT, Chen SH, Lin KL, Hsieh KH (1990) Rice pollen allergy in Taiwan. *Ann Allergy* **65**, 459-462.

Tsuchiya T, Toriyama K, Ejiri S, Hinata K (1994) Molecular characterization of rice genes specifically expressed in the anther tapetum. *Plant Mol Biol* **26**, 1737-1746.

Tsuchiya T, Toriyama K, Nasrallah ME, Ejiri S (1992) Isolation of genes abundantly expressed in rice anthers at the microspore stage. *Plant Mol Biol* **20**, 1189-1193.

Tsugita A, Kamo M, Kawakami T, Ohki Y (1996) Two-dimensional electrophoresis of plant proteins and standardization of gel patterns. *Electrophoresis* **17**, 855-865.

Tsugita A, Kawakami T, Uchiyama Y, Kamo M, Miyatake N, Nozu Y (1994) Separation and characterization of rice proteins. *Electrophoresis* **15**, 708-720.

Turcich MP, Hamilton DA, Mascarenhas JP (1993) Isolation and characterization of pollen-specific maize genes with sequence homology to ragweed allergens and pectate lyases. *Plant Mol Biol* **23**, 1061-1065.

Twell D (1994) The diversity and regulation of gene expression in th epathway of male gametophyte development. In 'Molecular and cellular aspects of plant reproduction'. (Eds RJ Scott and AD Stead). (Cambridge University Press)

Twell D, Wing R, Yamaguchi J, McCormick S (1989) Isolation and expression of an anther-specific gene from tomato. *Mol Gen Genet* **217**, 240-245.

Tyers M, Mann M (2003) From genomics to proteomics. *Nature* **422**, 193-197.

Unlu M, Morgan ME, Minden JS (1997) Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* **18**, 2071-2077.

Ursin VM, Yamaguchi J, McCormick S (1989) Gametophytic and sporophytic expression of anther-specific genes in developing tomato anthers. *Plant Cell* **1**, 727-736.

van Wijk KJ (2001) Challenges and prospects of plant proteomics. *Plant Physiol* **126**, 501-508.

Vergne P, Delvallee I, Dumas C (1987) Rapid assessment of microspore and pollen development stage in wheat and maize using DAPI and membrane permeabilization. *Stain Technol* **62**, 299-304.

Vidali L, Hepler PK (1997) Characterization and localization of profilin in pollen grains and tubes of Lilium longiflorum. *Cell Motil Cytoskeleton* **36**, 323-338.

Vidali L, Hepler PK (2001) Actin and pollen tube growth. *Protoplasma* **215**, 64-76.

Vierstra RD (2003) The ubiquitin/26S proteasome pathway, the complex last chapter in the life of many plant proteins. *Trends Plant Sci* **8**, 135-142.

Vihinen M (2001) Bioinformatics in proteomics. *Biomol Eng* **18**, 241-248.

Wallin E, von Heijne G (1998) Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci* **7**, 1029-1038.

Wang CS, Walling LL, Eckard KJ, Lord EM (1992a) Immunological Characterization of a Tapetal Protein in Developing Anthers of Lilium-Longiflorum. *Plant Physiology* **99**, 822-829.

Wang CS, Walling LL, Eckard KJ, Lord EM (1992b) Patterns of Protein Accumulation in Developing Anthers of Lilium-Longiflorum Correlate with Histological Events. *American Journal of Botany* **79**, 118-127.

Wang CS, Walling LL, Eckard KJ, Lord EM (1993) Characterization of an anther-specific glycoprotein in lilium-longiflorum. *American Journal of Botany* **80**, 1155-1161.

Wasinger VC, Cordwell SJ, Cerpapoljak A, Yan JX, Gooley AA, Wilkins MR, Duncan MW, Harris R, Williams KL, Humpherysmith I (1995) Progress with gene-product mapping of the mollicutes - mycoplasma genitalium. *Electrophoresis* **16**, 1090-1094.

Wasinger VC, Corthals GL (2002) Proteomic tools for biomedicine. *J Chromatogr B Analyt Technol Biomed Life Sci* **771**, 33-48.

Wasinger VC, Pollack JD, Humphery-Smith I (2000) The proteome of Mycoplasma genitalium. Chaps-soluble component. *Eur J Biochem* **267**, 1571-1582.

Wilkins MR, Pasquali C, Appel RD, Ou K, Golaz O, Sanchez JC, Yan JX, Gooley AA, Hughes G, Humphery-Smith I, Williams KL, Hochstrasser DF (1996a) From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Biotechnology (N Y)* **14**, 61-65.

Wilkins MR, Sanchez JC, Gooley AA, Appel RD, Humphery-Smith I, Hochstrasser DF, Williams KL (1996b) Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnol Genet Eng Rev* **13**, 19-50.

Willing RP, Bashe D, Mascarenhas JP (1988) An Analysis of the Quantity and Diversity of Messenger-Rnas from Pollen and Shoots of Zea-Mays. *Theoretical and Applied Genetics* **75**, 751-753.

Willing RP, Mascarenhas JP (1984) Analysis of the Complexity and Diversity of Messenger-Rnas from Pollen and Shoots of Tradescantia. *Plant Physiology* **75**, 865-868.

Wilm M, Mann M (1996) Analytical properties of the nanoelectrospray ion source. *Anal Chem* **68**, 1-8.

Wolters DA, Washburn MP, Yates JR, 3rd (2001) An automated multidimensional protein identification technology for shotgun proteomics. *Anal Chem* **73**, 5683-5690.

Wolters-Arts M, Lush WM, Mariani C (1998) Lipids are required for directional pollen-tube growth. *Nature* **392**, 818-821.

Wuthrich B (1989) Epidemiology of the allergic diseases: are they really on the increase? *Int Arch Allergy Appl Immunol* **90**, 3-10.

Xu H, Swoboda I, Bhalla PL, Singh MB (1999a) Male gametic cell-specific expression of H2A and H3 histone genes. *Plant Mol Biol* **39**, 607-614.

Xu H, Swoboda I, Bhalla PL, Singh MB (1999b) Male gametic cell-specific gene expression in flowering plants. *Proc Natl Acad Sci U S A* **96**, 2554-2558.

Xu H, Theerakulpisut P, Goulding N, Suphioglu C, Singh MB, Bhalla PL (1995a) Cloning, expression and immunological characterization of Ory s 1, the major allergen of rice pollen. *Gene* **164**, 255-259.

Xu H, Theerakulpisut P, Taylor PE, Knox RB, Singh MB, Bhalla PL (1995b) Isolation of a gene preferentially expressed in mature anthers of rice (oryza sativa l.). *Protoplasma* **187**, 127-131.

Xu HL, Goulding N, Zhang Y, Swoboda I, Singh MB, Bhalla PL (1999) Promoter region of Ory s 1, the major rice pollen allergen gene. *Sexual Plant Reproduction* **12**, 125-126.

Yanagida M (2002) Functional proteomics; current achievements. *J Chromatogr B Analyt Technol Biomed Life Sci* **771**, 89-106.

Yang WC, Sundaresan V (2000) Genetics of gametophyte biogenesis in Arabidopsis. *Curr Opin Plant Biol* **3**, 53-57.

Yates JR, 3rd (1998) Mass spectrometry and the age of the proteome. *J Mass Spectrom* **33**, 1-19.

Yates JR, 3rd (2000) Mass spectrometry. From genomics to proteomics. *Trends Genet* **16**, 5-8.

Yates JR, 3rd, Eng JK, McCormack AL (1995) Mining genomes: correlating tandem mass spectra of modified and unmodified peptides to sequences in nucleotide databases. *Anal Chem* **67**, 3202-3210.

Yee A, Pardee K, Christendat D, Savchenko A, Edwards AM, Arrowsmith CH (2003) Structural proteomics: toward high-throughput structural biology as a tool in functional genomics. *Acc Chem Res* **36**, 183-189.

Ylstra B, Garrido D, Busscher J, Vantunen AJ (1998) Hexose transport in growing petunia pollen tubes and characterization of a pollen-specific, putative monosaccharide transporter. *Plant Physiology* **118**, 297-304.

Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, Cao M, Liu J, Sun J, Tang J, Chen Y, Huang X, Lin W, Ye C, Tong W, Cong L, Geng J, Han Y, Li L, Li W, Hu G, Li J, Liu Z, Qi Q, Li T, Wang X, Lu H, Wu T, Zhu M, Ni P, Han H, Dong W, Ren X, Feng X, Cui P, Li X, Wang H, Xu X, Zhai W, Xu Z, Zhang J, He S, Xu J, Zhang K, Zheng X, Dong J, Zeng W, Tao L, Ye J, Tan J, Chen X, He J, Liu D, Tian W, Tian C, Xia H, Bao Q, Li G, Gao H, Cao T, Zhao W, Li P, Chen W, Zhang Y, Hu J, Liu S, Yang J, Zhang G, Xiong Y, Li Z, Mao L, Zhou C, Zhu Z, Chen R, Hao B, Zheng W, Chen S, Guo W, Tao M, Zhu L, Yuan L, Yang H (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. indica). *Science* **296**, 79-92.

Zapata FJ, Khush GS, Crill JP, Neu MH, Romero RO, Torrizo LB, Alejar M (1983) Rice anther culture at IRRI. In 'Cell and tissue culture techniques for cereal crop improvement' pp. 27-146. (Science Press: Beijing)

Zhang W, Chait BT (2000) ProFound: an expert system for protein identification using mass spectrometric peptide mapping information. *Anal Chem* **72**, 2482-2489.

Zheng HH, Qu LJ, Liu MH, Zhang Y, Shen YP, Wei JM, Pan NS, Gu HY, Chen ZL (2000) An anther-specific chalcone synthase-like gene D5 related to rice pollen development. *Chinese Science Bulletin* **45**, 1921-1926.

Zhong BX, Karibe H, Komatsu S, Ichimura H, Nagamura Y, Sasaki T, Hirano H (1997) Screening of rice genes from a cdna catalog based on the sequence data-file of proteins separated by two-dimensional electrophoresis. *Breeding Science* **47**, 245-251.

Zivy M, de Vienne D (2000) Proteomics: a link between genomics, genetics and physiology. *Plant Mol Biol* **44**, 575-580.

Zou JT, Zhan XY, Wu HM, Wang H, Cheung AY (1994) Characterization of a rice pollen-specific gene and its expression. *American Journal of Botany* **81**, 552-561.

APPENDIX

Refereed Publications

1. **Kerim, T.,** Imin, N., Weinman, J. J. and Rolfe, B. G (2003). Proteome analysis reveals developmentally expressed rice homologues of grass group II pollen allergens, *Functional Plant biology*, 30 (8): 843-852.

2. **Kerim, T.**, Imin, N., Weinman, J. J. and Rolfe, B. G. (2003). Proteome analysis of male gametophyte development in rice anthers, *Proteomics* **3**(5): 738-51.

3. Imin, N., **Kerim, T.**, Weinman, J. J. and Rolfe, B. G. (2001). Characterisation of rice anther proteins expressed at the young microspore stage, *Proteomics* **1**(9): 1149-61.

4. Mathesius, U., Imin, N., Chen, H., Djordjevic, M. A., Weinman, J. J., Natera, S. H., Morris, A. C., **Kerim, T.**, Paul, S., Menzel, C., Weiller, G. F. and Rolfe, B. G. (2002). Evaluation of proteome reference maps for cross-species identification of proteins by peptide mass fingerprinting, *Proteomics* **2**(9): 1288-303.

5. Imin N, **Kerim T**, Weinman JJ, Rolfe BG. Effect of Early Cold Stress on the Maturation of Rice Anthers *Proteomics* (in press).