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
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 develops 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.
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.
Chapter 1: General Introduction

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 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 locale 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 locale by the tapetal cells, and the young microspore encased inside the callose wall will be released into the anther locale. 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 proplast.

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 through 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
Chapter 1: General Introduction

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, 1993). 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. 10
Chapter 1: General Introduction

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
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 transcriptomics (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. Figure 1.4 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 pI, 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 pl 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 (Bewick 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 incompatible 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
Chapter I: General Introduction

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).
Chapter 1: General Introduction

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:

\[
t = D \left(\frac{m/z}{2eEs}\right)^{1/2}, \quad \text{so that} \quad m/z = 2eEs \left(\frac{t}{D}\right)^2
\]

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 \( \alpha \)-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>
<thead>
<tr>
<th>Application</th>
<th>Program</th>
<th>Website or company name</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2-DE image analysis</strong></td>
<td>Melanie 4</td>
<td>GeneBio, Geneva, Switzerland</td>
</tr>
<tr>
<td></td>
<td>PDQuest</td>
<td>Bio_Rad, Hercules, CA, USA</td>
</tr>
<tr>
<td></td>
<td>Imagemaster®</td>
<td>Amersham Biosiences, Upsala</td>
</tr>
<tr>
<td></td>
<td>Progenesis</td>
<td>Nonlinear Dynamics Limited, UK</td>
</tr>
<tr>
<td></td>
<td>Z3</td>
<td>Compugen, Tel-Aviv, israel</td>
</tr>
<tr>
<td><strong>Peptide mass</strong></td>
<td>Mascot</td>
<td><a href="http://www.matrixscience.com">http://www.matrixscience.com</a></td>
</tr>
<tr>
<td>fingerprinting**</td>
<td>MS-Digest</td>
<td><a href="http://prospector.ucsf.edu/ucsfhtml4.0/msdigest.htm">http://prospector.ucsf.edu/ucsfhtml4.0/msdigest.htm</a></td>
</tr>
<tr>
<td></td>
<td>MS-Fit</td>
<td><a href="http://prospector.ucsf.edu/ucsfhtml4.0/msfit.htm">http://prospector.ucsf.edu/ucsfhtml4.0/msfit.htm</a></td>
</tr>
<tr>
<td></td>
<td>ProFound</td>
<td><a href="http://prowl.rockefeller.edu/cgi-bin/ProFound">http://prowl.rockefeller.edu/cgi-bin/ProFound</a></td>
</tr>
<tr>
<td></td>
<td>MOWSE</td>
<td><a href="http://www.hgmp.mrc.ac.uk/Bioinformatics/Webapp/mowse/">http://www.hgmp.mrc.ac.uk/Bioinformatics/Webapp/mowse/</a></td>
</tr>
<tr>
<td></td>
<td>ProteinLynx</td>
<td>Micromass, Manchester, UK</td>
</tr>
<tr>
<td></td>
<td>PepSea</td>
<td><a href="http://195.41.108.38/PA_PepSeaForm.html">http://195.41.108.38/PA_PepSeaForm.html</a></td>
</tr>
<tr>
<td><strong>MS/MS analysis</strong></td>
<td>Mascot</td>
<td><a href="http://www.matrixscience.com">http://www.matrixscience.com</a></td>
</tr>
<tr>
<td></td>
<td>MS-Seq</td>
<td><a href="http://prospector.ucsf.edu/ucsfhtml3.4msseq.htm">http://prospector.ucsf.edu/ucsfhtml3.4msseq.htm</a></td>
</tr>
<tr>
<td></td>
<td>MS-Tag</td>
<td><a href="http://prospector.ucsf.edu/ucsfhtml3.4mstagfd.htm">http://prospector.ucsf.edu/ucsfhtml3.4mstagfd.htm</a></td>
</tr>
<tr>
<td></td>
<td>Multident</td>
<td><a href="http://www.expasy.ch/tools/multident/">http://www.expasy.ch/tools/multident/</a></td>
</tr>
<tr>
<td></td>
<td>PepFrag</td>
<td><a href="http://prowl.rockefeller.edu/prowl/">http://prowl.rockefeller.edu/prowl/</a></td>
</tr>
<tr>
<td></td>
<td>TurboSEQUEST</td>
<td>Thermo Finnigan, San Jose, CA, USA</td>
</tr>
<tr>
<td></td>
<td>PepSea</td>
<td><a href="http://195.41.108.38/PA_PeptidePatternForm.html">http://195.41.108.38/PA_PeptidePatternForm.html</a></td>
</tr>
</tbody>
</table>
Chapter 1: General Introduction

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 (d0) and a heavy form (d8) 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 d0/d4
nicotinylolation, 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).
Chapter 1: General Introduction

(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, pI, 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.
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/m²) 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 μmol.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.
Chapter 2: Materials and Methods

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
Chapter 2: Materials and Methods

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 solubilisation

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:

<table>
<thead>
<tr>
<th>Component</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tris base</td>
<td>15.25 g</td>
</tr>
<tr>
<td>10 M HCl</td>
<td>10-12 mL</td>
</tr>
</tbody>
</table>

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:

<table>
<thead>
<tr>
<th>Component</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urea (8 M)</td>
<td>9.6 g</td>
</tr>
<tr>
<td>DTT (0.2%, w/v)</td>
<td>0.04 g</td>
</tr>
<tr>
<td>CHAPS (0.5%, w/v)</td>
<td>0.1 g</td>
</tr>
<tr>
<td>Bio-lyte ampholytes (0.52%, v/v)</td>
<td>260 µL</td>
</tr>
<tr>
<td>1% bromophenol blue</td>
<td>120 µL</td>
</tr>
</tbody>
</table>
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 × 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 × 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 × 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 × 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
Chapter 2: Materials and Methods

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)

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

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

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

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

2.3.4 Protein visualization

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

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

<table>
<thead>
<tr>
<th>Component</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ammonium sulphate (17%, w/v)</td>
<td>170 g</td>
</tr>
<tr>
<td>85% Phosphoric acid (3%, w/v)</td>
<td>36 mL</td>
</tr>
<tr>
<td>Coomassie G-250 (0.1%, w/v)</td>
<td>1 g</td>
</tr>
<tr>
<td>Methanol (34%, v/v)</td>
<td>340 mL</td>
</tr>
</tbody>
</table>

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 pI/s of spots was 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 scalpels. 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 pls of the matching sequences, excluding the signal peptide, were calculated using the on-line Compute pl/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
Chapter 2: Materials and Methods

of the 96 wells of a polypropylene microtitre plate (Medos, Langenselbold, Germany). Each well contained 10 μL of 10% (v/v) methanol which assisted 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 NH₄CO₃, 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 fingerprinting 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.
### Chapter 2: Materials and Methods

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