Proteomic Study on the Developing High-lysine Rice Seeds

Leung, Hoi Ching

A Thesis Submitted in Partial Fulfillment

of the Requirements for the Degree of

Master of Philosophy

in

Biology

February 2007

Thesis/ Assessment Committee

Prof. L. W. Jiang (Chair)

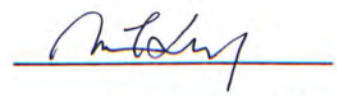
Prof. S. S. M. Sun (Thesis Supervisor)

Prof. S. M. Ngai (Committee Member)

Prof. N. S. Wong (External Examiner)

Statement from Author

All experimental work reported in this thesis was performed by the author, unless stated otherwise.



Leung Hoi Ching

Acknowledgements

First of all, I would like to express my sincere gratitude to my supervisor, Prof. S. S. M. Sun for offering me the opportunity to carry out this study. I have learned quite a lot from his extensive knowledge in plant molecular biology and many brilliant and creative ideas. I am deeply grateful to him for his continuous guidance, invaluable advises and concern throughout my study.

I would like to express my appreciation to my internal examiners, Prof. L.W. Jiang and Prof S. M. Ngai for their precious comments on my experiments and thesis. My thanks also go to my external examiner, Prof. N. S. Wong for his patience to review my thesis and available comments.

Besides, I would like to express my special thanks to Dr. Q. Q. Liu, who gave me the study materials and valuable information about my research. I also wish to thanks Helen Tsai for giving me technical support and helping me handling the MS machine, Freddie Kwok for teaching me various skills in Transmission Electron Microscopy and Mrs. Yeung Yeo Geok Yen in G50 for giving me assistance in data analysis of MS.

In addition, thanks must be given to all labmates in G90/91 as well (Marilyn Yu, Angela Yu, Steffi Long, Jennifer Wei, Gundam Wong, Juonjuon Lee, Sylvia lok, Yuk Wah Chan, Ringo Kwok, Ronald Yu, Dr. D.Y. Yuan, Smile Duan, Carmen Ho, Rebecca Chan, Dr Y.P. Bu, Raymond Cheung, Chiuyi Leung, Veronica Yu, Jim Yeung,

Melinda Liu, Siu Mei Chan, Tobi Kwok, Annual Tam, Eric Suen, Wendy Chen, Kit Tong, Stanley Cheung, Ruby Ng, William Ho and Pauline Yip), for their kind help, support, encouragement and technical advice during these years. Also, thanks for members of other labs (387, EG08 and AoE) and supporting teams in the Department of Biology, for their kind assistance.

I also like to thank my friends for their constant presence and support.

Finally, I express my warmest thanks to my parents and my sister for their moral support and encouragement in the course of my study.

Abstract

Rice (*Oryza sativa* L.) is the staple food of over half the world population. However, the nutritional quality of rice proteins is not optimal due to their deficiency in lysine, an essential amino acid cannot be synthesized by humans and must be obtained in the diet. An amino acid unbalanced cereal-based diet can affect the efficiency of converting plant to animal protein. Therefore it is necessary to enhance the lysine content of rice proteins in order to improve the nutritional quality of rice grains.

By genetic engineering, two types of high-lysine rice were successfully generated by integration of a lysine-rich protein (LRP) gene from winged bean and also by additional copies of lysine richer glutelin protein gene. The lysine content of rice seed protein was significantly enhanced (by up to 58% over the wild type). But the deposition and expression profiles of the recombinant proteins and their corresponding effects on other rice proteins remain unsolved. Before large-scale application of these approaches to improve the nutritional quality of rice seed proteins becomes possible, systematic analysis of the transgenic LRP rice in regard to the expression of the LRP gene and extra copies of glutelin gene is prerequisite.

In this study, two transgenic rice lines expressing the fusion of a heterologous LRP with glutelin and additional copies of glutelin gene (Gt1) were systematically

analyzed. The variations between the proteomes of developing rice seeds of wild-type and transgenic lines were first examined through proteomic approach, while the endosperm cells of the transgenic rice were observed and the depositions of the fusion protein as well as the storage proteins were studied via Transmission Electron Microscopy (TEM).

In total, 35 protein spots showing altered intensity in 2-D SDS-PAGE were identified by MS/MS. In the LRP/glutelin transgenic rice (FB rice), six extra spots of the LRP fusion protein with the same molecular weight but differed in isoelectric points were found. In the glutelin-added transgenic rice (GT rice), glutelin precursor and acidic subunits were found to be higher in abundance as compare with the wild type. In both types of rice, the chaperone binding protein (BiP) and protein disulphide isomerase (PDI) showed enhanced expression while the amounts of other storage proteins (prolamin, globulin and albumin) decreased significantly.

TEM study revealed the appearance of morphologically altered protein bodies (PBs) in the rice endosperm of the engineered rice seeds. In FB rice, the glutelin-containing PB (PB-II) remained morphologically unchanged but the prolamin-containing PB (PB-I) disappeared. A new type of PB harboring prolamin and LRP-fusion protein was readily observed. In GT rice, PB-I was fused with some smaller glutelin-containing PBs while PB-II was fused with smaller

prolamin-containing PBs.

Combining the proteomic analysis and electron microscopic observation, the results provide information at molecular and subcellular levels in the changes caused by introducing recombinant proteins into rice. This information suggests that the synthesis of foreign proteins in abundance may alter the process of protein body formation in transgenic rice. In addition, the expression of abundant recombinant proteins may induce ER stress so the chaperones BiP and PDI are actively expressed to cope with the unfolded proteins while the expression of other storage proteins is inhibited under the effect of unfolded protein response. As a result, the native storage-protein sorting and trafficking process is distorted and the protein body formation is affected.

摘要

稻米是世界半數以上人口的主要食物與營養的來源，然而稻米中的營養品質卻未能達致平衡，賴氨酸是稻米營養平衡的第一限制性因數。賴氨酸是一種必需氨基酸，即不能在人體中被製造，而需要從食物中吸收。這種氨基酸的不平衡性會影響植物蛋白向動物蛋白轉化的有效性，因此提高賴氨酸的含量能有效改進稻米的營養價值。

透過基因工程技術，兩種含有高賴氨酸蛋白的水稻已經成功產出，第一種方法是來將源於四梭豆的一個高賴氨酸蛋白質（Lysine-rich protein, LRP）基因轉化到水稻，另一方法則是把加強水稻中較富於賴氨酸的谷蛋白基因（Gt1）的表達，從而顯著提高水稻種子蛋白中的賴氨酸含量，藉此改良稻米的營養品質。然而，在這種高賴氨酸基因水稻能被廣泛生產及食用之前，對轉基因植物進行綜合的以及詳細的分析是必需的。

本研究利用兩種不同的高賴氨酸水稻作爲實驗研究材料，對表達高賴氨酸（LRP）與谷蛋白融合蛋白及表達高谷蛋白的轉基因植株進行了系統分析。首先我們利用蛋白質組學的方法，比較兩種高賴氨酸水稻與未轉化水稻的成熟中的水稻種子的蛋白質組，從而找出其中的差異。另外，我們亦透過穿透式電子顯微鏡（Transmission Electron Microscope, TEM）觀察轉基因水稻的種子內的胚乳細胞，找出異源蛋白和貯藏蛋白在胚乳細胞中的貯存情況。

透過雙向凝膠電泳（2D-PAGE）和串聯質譜（MS/MS）技術，共有三十五個蛋白被發現其表達模式有所改變。在表達高賴氨酸與谷蛋白融合蛋白的轉基因植株（FB）中，我們找出了六個代表高賴氨酸與谷蛋白融合蛋白的蛋白質點。它們擁有相同的分子量，但等電點是不同的。而在谷蛋白的轉基因植株（GT）中，谷蛋白的前驅蛋白和酸性亞基卻被發現其表達量有所提高。在這兩種高賴氨酸水稻中，屬於伴護蛋白（chaperone）的結合蛋白（binding protein, BiP）和二硫鍵異構酶（protein disulphide isomerase, PDI）的表達量都有所提高，而其他的貯藏蛋白（即醇溶蛋白、球蛋白和白蛋白）的表達量卻大幅下降。

利用穿透式電子顯微鏡觀察水稻種子的胚乳細胞，發現轉基因水稻種子的蛋白體的結構有所改變。在 FB 水稻中，儲存谷蛋白的蛋白體-II（protein body-II）沒有改變，而儲存醇溶蛋白的蛋白體-I（protein body-I）卻不見了。取而代之，一種新的蛋白體被發現，它同時貯存了醇溶蛋白和高賴氨酸蛋白。在 GT 水稻中，蛋白體-I 的表面被發現和一些儲存了谷蛋白的小型蛋白體相連，而蛋白體-II 的表面和一些儲存了醇溶蛋白的小型蛋白體相連。

本研究綜合蛋白質組學和穿透式電子顯微鏡的技術，在分子水平和細胞水平上找出異源蛋白在水稻體內表達時所產生的影響。研究結果顯示在轉基因水稻中，高表達量的異源蛋白有機會改變蛋白體的形成過程。這是因爲當大量異源蛋白在細胞中被表達時，未折疊的蛋白在內質網上增多，引起內質網應激（ER stress），從而引發未折疊蛋白反應（Unfolded Protein Response, UPR），伴護蛋白

BiP 和 PDI 的表達量因此被提高去處理那些未折疊的蛋白；同時，其他的貯藏蛋白的表達量則被減低。結果，在轉基因水稻的胚乳細胞中，原本的貯藏蛋白的分配和運輸途徑也受到影響，蛋白體的形成也因而改變。

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AA	Amino acid
Ab	Antibody
ACN	Acetonitrile
BAC	Bacterial artificial chromosome
BiP	Lumenal binding protein
BLAST	Basic local alignment search tool
BSA	Bovine serum albumin
CBB	Coomassie brilliant blue
CHAPS	3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate
CHCA	α-cyano-2-hydroxycinnamic acid
CCD	Charge Coupled Device
C-ER	Cisternal ER
Cys	Cysteine
Da	Dalton
DAF	Days after fertilization
DNA	Deoxyribonucleic acid
DV	Dense vesicle
DTT	Dithiothreitol
2D PAGE	Two-dimensional polyacrylamide gel electrophoresis
2D	Two-dimensional
EDTA	Ethylenediaminetetra-acetic acid
ER	Endoplasmic reticulum
FAO	Food and Agriculture Organization
FB	Transgenic rice containing LRP gene inserted into basic subunit of rice Gt1
g	Gram
GGPP	Geranylgernayl diphosphate
GluA	Glutelin A
IEF	Isoelectric focusing
GluB	Glutelin B

GT	Transgenic rice containing extra copies of Gt1 gene
Gt1	Rice glutelin 1 gene
HCTR	HC-toxin reductase
HPLC	High performance liquid chromatography
hr	Hours
HSP70	70 kDa Heat shock protein
HVR	Highly variable region
IAA	Iodoacetamide
IgE	Immunoglobulin E
IgG	Immunoglobulin G
pI	Isoelectric point
IPG	Immobilized pH gradients
IRRI	International Rice Research Institute
kDa	Kilodalton
LC	Liquid chromatography
LRP	Lysine-rich protein
LR while	White London Resin
M	Mole
MALDI-ToF	Matrix assisted laser desorption ionization-time of flight
Mb	Mega base pairs
mg	Milligram
min	Minutes
Met	Methionine
MS	Mass spectrometry
MSDB	Mass Spectrometry protein sequence DataBase
MS/MS	Tandem mass spectrometry
MW	Molecular weight
m/z	Mass/charge
N	Nitrogen
NaOH	Sodium hydroxide
NCBI	The National Center for Biotechnology Information
PAC	P1-derived artificial chromosome
PAGE	Polyacrylamide gel electrophoresis

PB-I	Protein body-I
PB-II	Protein body-II
PB-ER	Protein body-forming ER
PBS	Phosphate buffered saline
PBST	Phosphate buffered saline with Tween-20
PDI	Protein disulfide isomerase
PIR	Protein Information Resource
PMF	Peptide mass fingerprint
PSV	Protein storage vacuole
rpm	Revolutions per minute
SDS	Sodium dodecyl sulfate
SP	Signal peptide
TCA	Trichloroacetic acid
TEM	Transmission Electron Microcopy
TFA	Trifluoroacetic acid
TIGR	The Institute for Genomic Research
µg	Microgram
µl	Microliter
µm	Micrometer
UPR	Unfolded protein response
v	Volume
VAD	Vitamin A deficiency
vhr	Voltage hour
VPE	Vacuolar processing enzyme
VR	Variable region
WHO	World Health Organization

Chapter 1. General Introduction

Rice is the predominant staple food for over half of the world's population. It provides 700 calories/day/person for about 3,000 million people, most of whom live in developing countries (FAO, 2003). Since rice is a relatively cheap source of food, it is the major food in the developing world and thus serves as a sole source of energy and major dietary protein for the population.

However, while rice provides a substantial amount of dietary energy, it has an incomplete amino acid profile and contains deficient amounts of essential amino acids. Lysine is the first limiting amino acid in rice protein. The lysine values range from 3.8 to 4g/16g N in protein of rice, which is lower than the standard (5.5g/16g N) suggested by the World Health Organization (WHO). The low-income populations who can only effort rice-based diets may suffer from malnutrition.

Therefore, various approaches including traditional breeding and genetic engineering have been used to improve the lysine content in rice. In 2002, our lab has successfully produced the high-lysine rice by genetic engineering. Winged bean (*Psophocarpus tetragonolobus*), an edible bean widely consumed in Southeast Asia, contains a seed protein rich in lysine (10.8 mol %, 7.5 g/16 g N). The construct containing the gene encoding winged bean lysine-rich protein (LRP) inserted in the basic subunit of glutelin gene was expressed in rice (Liu, 2002) and the lysine content

was significantly enhanced (58% increased in total amino acid over the control rice). In 2004, transgenic high lysine rice over-expressing lysine-richer glutelin gene was also generated (30% increased in total amino acid) over the control rice (Liu, CUHK).

While lysine enhancement was achieved, initial results revealed that this genetic intervention resulted in the accumulation of larger amount of the 57-kDa glutelin precursor in both types of transgenic rice in comparison to wild type. Moreover, the transgenic rice showed a lower accumulation of other seed storage proteins including prolamin and globulin (Liu, CUHK). However, the molecular details and relationships between the expression of foreign proteins and native proteins as well as the subcellular deposition of the foreign proteins remain unclear. It has been reported earlier that high-lysine mutants of maize affected protein synthesis and protein body formation (Coleman *et al.*, 1997). However, it is not known whether such a correlation also exists in our high-lysine rice. Before the high-lysine rice can be released, comprehensive analyses including the expression of LRP/Gt1 fusion gene as well as extra copy of Gtl gene and their possible effects on other gene traits of the host plant are desirable. Recent advancements in proteomics provide opportunities to study protein expression of the transgenic rice in a systematic way.

This project thus focuses on systematic comparison of the composition of protein in high-lysine rice lines with the non-transgenic rice through proteomics approach

while the solubility and the subcellular localization of the transgenic glutelin-fusion LRP in transgenic rice grains will also be investigated. Transmission electron microscopic observation of rice endosperm can further elucidate the relationship between the expression of recombinant proteins and the subcellular changes in rice grain. This study will contribute to our understanding and effort in using molecular approach to improve the quality of rice seed proteins in the future.

Chapter 2. Literature Review

2.1 Nutritional quality of rice

Rice grain quality comprises grain appearance and milling, cooking, and nutritional quality. As rice provides a substantial source of dietary protein and energy for human consumption, the nutritional quality of rice is directly correlated to the quality of human diets. The nutritional value of rice mainly depends on the protein content and its amino acid makeup.

2.1.1 Classification of seed proteins

Protein is the second abundant component in rice grains. The seed protein content of rice ranges from 5 to 12% (Villareal and Juliano, 1978). Based on function, there are three main types of proteins: (1) storage proteins, (2) metabolic and structural proteins, and (3) protective proteins (Shewry and Casey, 1999). The metabolic and structural proteins are essential for the growth and structure of the seed. Many can also be called "housekeeping" proteins which are indispensable for all tissues of the plant while others are associated with the biochemical pathways of storage product biosynthesis. Protective proteins function against attack from pests and microbial pathogens. On the other hand, the seed storage proteins are non-enzymatic and have the major purpose of providing proteins (nitrogen and sulphur source) required during seed germination and

development of a new plant. Since metabolic and structural proteins and protective proteins only account for a very small amount of the total protein, storage proteins are thus the major component of seed proteins.

2.1.2 Amino acids composition of rice proteins

The nutritional quality of rice is mainly determined by the amino acid makeup of its protein. The protein in rice endosperm is of good quality compared with that of other cereal crops, especially the relatively higher essential amino acid lysine content (Table 1). However, the lysine content in rice is 3.8 g/16 g N, which is still far behind the WHO's recommended content (5.5 g/16 g of N) (WHO, 1973). Thus lysine is the first limiting amino acid in rice, which may result in nutritional deficiencies for humans who depend on rice protein alone as their entire protein requirement. Improving the protein content of rice will be an important effort.

Table 1. The protein and selected essential amino acids composition of eight whole-grain cereals (Juliano, 1985) and the standard from WHO (1973)

property	Brown rice	Wheat	Corn	Barley	Millet	Sorghum	Rye	Oat	WHO
Protein, %	7.3	10.6	9.8	11.0	11.5	8.3	8.7	9.3	
Lysine, g/16 g N	3.8	2.3	2.5	3.2	2.7	2.7	3.7	4.0	5.5
Threonine, g/16 g N	3.6	2.8	3.2	2.9	3.2	3.3	3.3	3.6	6.0
Methionine+ cysteine, g/16 g N	3.9	3.6	3.9	3.9	3.6	2.8	3.7	4.8	3.5
Tryptophan, g/16 g N	1.1	1.0	0.6	1.1	1.3	1.0	1.0	0.9	1.0

2.1.3 Other nutritional components of rice

Apart from seed protein, the nutritional quality of rice is also determined by fibers, lipids and micronutrients: minerals and vitamins. Crude fiber in brown and milled rice is 0.8 and 0.6%, respectively. Dietary fiber is usually defined as all plant food components that are not broken down by enzymes in the human digestive tract. Roth and Mehlmen (1978) reviewed the importance of dietary fiber in the maintenance of health and in the development of the specific disorders including obesity, diabetes, colonic-rectal cancer, and arteriosclerosis. The dietary fiber is highest in the outer bran layer of the grain and decreases toward the center.

Rice lipids are generally classified into nonstarch lipids found in the aleurone layer, embryo, and protein bodies of the endosperm and starch lipids, which are associated with starch granules (Juliano, 1983). The average lipids content is 0.65% and ranging from 0.19-2.73% (Juliano, 1972). Within the endosperm, lipids are unevenly distributed, with the highest in the outer layer while decreasing toward the center of the kernel (Julinao, 1983).

The mineral composition of the rice grain depends considerably on the availbility of soil nutrients during crop growth and on the diverse sampling and preparation (Juliano and Bechtel, 1985). Minerals content is higher in brown rice than in milled

rice. Among the microelements, iron and zinc are essential for human nutrition. But cadmium is a toxic pollutant which is higher in milled rice than brown rice.

The concentrations of vitamins are low in rice. Rice contains little or no vitamin A, C, or D. Most of the vitamins in rice belong to the water-soluble vitamin B group. Vitamin B1, B6 and E in brown rice are ten times higher than in milled rice, which are mainly distributed in aleurone layer, scutellum and embryo.

2.2 Rice seed storage proteins

2.2.1 Properties and classification of seed storage proteins

Despite wide variation in their detailed structures, all seed storage proteins have a number of common properties (Shewry *et al.*, 1995): (1) Their synthesis is at high levels in specific tissues and at certain developing stages and the process is easily influenced by nutritional conditions; (2) They are comprised of several kinds of polypeptides with different amino acid composition and physiochemical features. For instance, some of which are rich in sulphur amino acids and other of which are poor in them. The presence of these groups may allow the plant to maintain sufficient storage protein synthesis even in low sulphur availability; (3) All seed storage proteins are stored in discrete deposits called protein bodies (PB); (4) All storage protein fractions are heterogeneous, which consist of many polypeptides and controlled by multiple

gene families; (5) They have no enzymatic activity; and (6) Seed storage proteins serve to provide a store of amino acids, nitrogen and sulphur for use during germination and seedling growth.

Seed storage proteins can be divided into four classes according to their solubility properties. Albumins are water soluble; globulins are salt soluble; prolamins are alcohol soluble; and glutelins are soluble in weak alkali or weak acid (Osborn, 1924; Juliano, 1972). The major storage proteins found in rice are the glutelins, which accounts for 80% or more of the total seed protein (Juliano, 1972; Villareal and Juliano, 1978). The remaining 20 % is divided as follows: albumins and globulins, 4-10%; and prolamins, 5-10% (Juliano, 1972). However, later studies (Krishnan and Okita, 1986; Okita *et al.*, 1988) found that significant amounts of prolamin remained associated with the glutelin proteins in the glutelin fraction resulting in the underestimation of the total amount of prolamin while at the same time overestimating the amount glutelin proteins. More recent study showed that glutelins and prolamins comprise about 58% and 19% of the total protein in rice seeds (Li and Okita, 1993).

2.2.2 Composition and structure

2.2.2.1 Glutelin

Glutelin is the main form of storage protein in rice grain, accumulating greatly in the middle and late seed development (Yamagata *et al.*, 1982). Despite their general insolubility, the rice glutelins share similarity in amino acid sequence, protein structure and also synthesis pathway with the 11S globulins of legume (Zhao *et al.*, 1983; Wen and Luthe, 1985).

Glutelin is of three kinds of molecular weight in rice grain, the 57-kDa precursor, 37-39-kDa acidic subunit (large subunit) and 22-23-kDa alkaline subunit (small subunit), of which acidic and alkaline subunit are formed by precursor processing. Mature polypeptides all show heterogeneity. For example, the large subunit comprises at least 16 polypeptides and the small subunit consists of about 9 polypeptides (Wen and Luthe, 1985). Based on the amino acid sequence, the glutelins are classified into two subfamilies, designated as A and B subfamily (GluA and GluB). The GluA subfamily is composed of at least 4 members, A-1 (or Gt2), A-2 (or Gt1), A-3 (or Gt3) and A-4 (Okita *et al.*, 1989; Takaiwa and Oono, 1991) while the GluB subfamily consists of B-1, B-2, B-3 and B-4 (Masumura *et al.*, 1989). The homology in amino acid sequence of members within the same subfamilies is 80-88%, whereas that between GluA and GluB members is 60-65% (Takaiwa *et al.*, 1991). Analysis of the

deduced amino acid sequences from cDNAs showed that a typical signal peptide is composed of 24 AAs in the N-terminal of glutelin precursor, which contains a conserved leucine-rich hydrophobic core. This signal peptide (SP) plays an important role in leading glutelin precursor into the ER and further transporting to lumen of ER (Okita *et al.*, 1989; Takaiwa *et al.*, 1991). Seven cysteine-residuals exist in the precursor, of which the –122 and –315 positions are responsible for production of bisulfate bond between acidic subunit and alkaline subunit. There are 5 variable regions (VRs) in the different glutelin precursors; three of them are located in the C-terminal of signal peptide (SP), acidic subunit and alkaline subunit, respectively. The two others VRs present in the middle of acidic subunit. Highly variable region (HVR) is regarded as the optimal position to accept relatively large heterogeneous polypeptide sequence, and therefore it might act as the best location to modify this kind of protein and improve nutritional quality (Argos *et al.*, 1985).

2.2.2.2 Prolamin

In most of the cereal grains, prolamin accounts for a major proportion for the total grain proteins. However, prolamin is a smaller component of the storage protein in rice. Extraction of the prolamin fraction of rice has typically been performed in ethanol solutions up to 70%, and Juliano (1972) estimated the prolamin makes up about 5% of

the total protein of the rice grain by using this extraction method. However, more recent studies found that 55% propan-1-ol was much more effective in the extraction of rice prolamins than ethanol (Sugimoto *et al.*, 1986). In addition, Li and Okita (1993) estimated that the prolamin content in the line M201 was 18-20% of the total seed storage proteins using immunoblots probed with prolamin and glutelin antisera. Therefore, the prolamins comprise a much larger proportion of the total storage protein fraction than was suggested by earlier studies.

Three groups of prolamin can be classified on the basis of molecular weight: a major polypeptide species of about 13-kDa and two minor species of 10-kDa and 16-kDa (Ogawa *et al.*, 1987; Masumura *et al.*, 1989). The 16-kDa mature polypeptide is composed of 130-140 amino acids while its N-terminal of precursor contains a SP with 18-19 AAs (Kim and Okita, 1988a; 1988b). The 13-kDa prolamin has at least 7 polypeptide groups with different AA composition, whose SP in the N-terminal of precursor consists of 18-19 AAs (Kim and Okita, 1988a; 1988b; Sha *et al.*, 1996; Hibino *et al.*, 1989). The 10-kDa polypeptide is comprised of 110 AAs and its precursor SP is 24 AAs, longer than that of the two others (Masumara *et al.*, 1989). In comparison with the protein and DNA sequences, there is 70%-95% homology shared in different polypeptides inside the same molecular weight (MW), while the homology is low between different MW types. For instance, 47% homology is shared between

the 13 kDa and 16-kDa prolamin, whereas the 10-kDa polypeptide has no homology with the 13-kDa and 16-kDa (Sha *et al.*, 1996).

Cloning of rice prolamin cDNA and genomic clones revealed that the amino acid sequence of rice prolamin does not exhibit significant homology to the prolamins of other cereals (Kim and Okita, 1988a; 1988b). Two regions, however, show some conservation with other prolamins. Like the prolamins in maize, wheat, rye and barley, there is a glutamine-rich octapeptide sequence – QQQCCQQL in the middle of the sulphur-rich rice prolamins. Moreover, the SP has very strong similarity with the SP of the maize zeins (Masumura *et al.*, 1989; 1990). The limited sequence similarity and weak immunological cross-reactivity of the rice prolmins and the prolamins of other cereals indicates that they possess individual origin in biological evolution.

2.2.2.3 Albumin and globulin

Albumin in rice is heterogeneous and contains several polypeptides with molecular mass of 14-16 kDa. The precursor of albumin consists of 150-170 AAs and a SP with 26-27 AAs in the N-terminal. The homology of different polypeptides is 70-95% (Adachi *et al.*, 1993).

Reports on rice globulin are rare. Alpha-globulin is the major fraction in rice globulin and is controlled by a single gene. Its MW is about 26 kDa. The precursor is

composed of 186 AAs and has a SP with 22 AAs in the N-terminal (Shorrosh *et al.*, 1992).

Some salt-soluble albumin and globulin are considered as rice allergenic proteins due to the reactivity with immunoglobulin E (IgE) in sera from patients allergenic to rice (Shibasaki *et al.*, 1979; Matsuda *et al.*, 1991). They all belong to α–amylase/trypsin inhibitor family (Limas *et al.*, 1990; Izumi *et al.*, 1992; Adachi *et al.*, 1993; Alvarez *et al.*, 1995). Their apparent molecular weights are in the 12-16 kDa range.

2.2.3 Synthesis, assembly and deposition of rice seed storage proteins

Seeds act as strong amino acid sinks when storage proteins are synthesized during seed maturation. The storage proteins are deposited in specialized membrane-bound organelles called protein bodies, which can avoid interference with other cellular processes and protect them against premature breakdown. There are two types of protein bodies in rice endosperm. Prolamin mainly exists in the spherical protein body (PB-I) (Tanaka *et al.*, 1980). Glutelin and globulin are found in the irregular-shaped protein body (PB-II) (Krishnan *et al.*, 1986). Albumin is present in the aleurone protein bodies, which differ drastically in composition from the endosperm protein bodies (Tanaka *et al.*, 1973). The transport of proteins and their deposition into protein

bodies involves complicated processes, whose mechanisms are still incompletely understood. To date, it appears that two mechanisms of protein body formation occur, which may operate at different times or for different protein types.

2.2.3.1 Storage protein folding and assembly in the ER

Seed storage protein (glutelin and prolamin) mRNAs are transcribed in the nucleus and then released into the cytosol and localized on the distinct ER membrane to synthesize protein. Previous researches found that there are two kinds of ERs exist in the endosperm. One is the cisternal ER (C-ER), consisting of a layer of pellicle and is distributed in the endosperm cells. Glutelin mRNAs are translated at this kind of ER. Another is the protein body-forming ER (PB-ER), which is the place where prolamin mRNA is localized (Li *et al.*, 1993a; 1993b).

After cotranslational cleavage of an N-terminal signal peptide, the nascent polypeptides are translocated into the ER lumen for processing. The polypeptides are subjected to chaperone-assisted folding, disulphide bond formation, glycosylation in specific cases and oligomerization (Li *et al.*, 1993b; Müntz, 1998).

Studies of other systems demonstrate that two types of ER luminal proteins may assist in the processes of protein folding and disulfide bond formation. Molecular chaperons of the HSP70/BiP family may facilitate folding by binding transiently to the

nascent polypeptides and may also prevent the formation of incorrect inter- or intramolecular interactions. In previous study, BiP was found to accumulate in higher than normal level in high-lysine maize mutants (Boston, 1991), possibly due to the presence of incorrectly folded zeins. The maize *floury 2* mutant is one of the well-studied mutant lines with twice the lysine content than the wild-type. The mutation is associated with elevated levels of BiP, and a novel 24-kDa polypeptide in the zein fraction. The 24-kDa polypeptide is a precursor of a 22-kDa α-zein protein with an uncleaved signal peptide. The presence of the signal peptide prevents normal folding of the protein. Much higher levels of BiP and malformed protein bodies were detected when these maize endosperms were compared with wild-type ones. The findings suggested that BiP is responsible for correct protein folding (Coleman *et al.*, 1995).

A second group of protein, protein disulfide isomerase (PDI) catalyzes disulfide bond formation in storage proteins. In rice *esp2* mutants, PDI level is suppressed in the accumulation of the glutelin precursor (Takemoto *et al.*, 2002). Electronic microscopic observation revealed that *esp2* mutants contained normal-appearing PB-II but lacks the normal PB-I. Instead, numerous small ER-derived PBs were observed. The results demonstrated that PDI plays an essential role in the segregation of glutelin precursor and prolamin polypeptides within the ER lumen.

2.2.3.2 Storage protein transport and protein body formation

Two routes of protein body formation appear to operate in developing rice endosperms, one is from the ER and the other is from protein storage vacuoles via the Golgi apparatus (Figure 1) (Krishnan *et al.*, 1986; Müntz, 1998). Glutelin and prolamin mRNAs are translated at different areas of the rough-ER, cis-ER and PB-ER, respectively. After disulfide linkage formation and trimerization in the ER lumen, 57 kDa glutelin precursor is transported to the Golgi apparatus. At the trans-Golgi cisternae glutelin precursor trimers are sorted into dense vesicles (DV). Glutelin precursor is processed into subunits by the vacuolar processing enzyme (VPE) and transformed into its deposition-compatible conformation. This occurs in a prevacuolar or vacuolar compartment. The latter is finally transformed into protein body II (PB-II).

On the other hand, transient binding of BiP to nascent prolamin polypeptides precedes sorting of prolamin into ER-dervied PB (PB-I). Li *et al.* (1993b) observed that many BiP are distributed within PB-I. BiP can bind to prolamin polypeptide to form protein complex, which is responsible for translocating prolamin from the outside of ER into its lumen, followed by folding in the lumen, assembling, and formation of PB-I. The restricted localization of abundance of BiP at the periphery of PB-I suggests that BiP functions to retain prolamins in the ER in a competent state

until it is deposited onto the protein body aggregate, at which time BiP is released (Li *et al.*, 1993b).

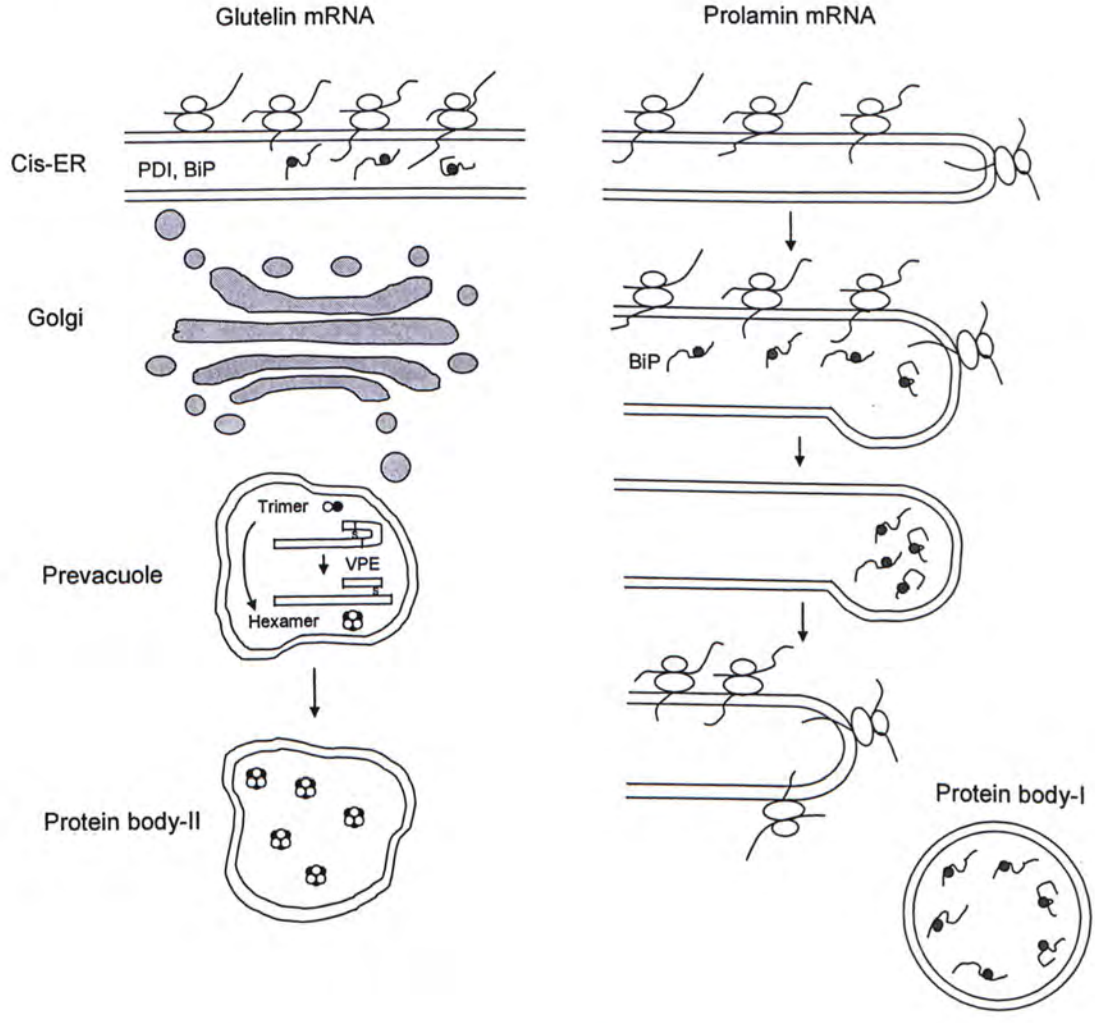


Figure 1. Two pathways of seed storage proteins deposition in developing rice grains.

PB-I contains prolamin while PB-II is composed of glutelin. The biosynthesis of these two PBs is different. PB-I is directly derived from ER membrane while PB-II comes from vacuole. Polypeptide trimers of glutelin precursor are secreted outside of ER and translocated into vacuole through Golgi apparatus. Inside the vacuole, the trimers are processed into small and large subunit, and then folded into hexamer, which are accumulated in the vacuole. Finally, these hexamers render the vacuole lyses into small pieces to form PB-II. On the other hand, prolamin polypeptides aggregate with BiP within the lumen of the ER. The prolamin-containing protein bodies (PB-I) are budding-off from ER directly. Hypothetical BiP retrieval is indicated.

2.2.3.3 Protein bodies and their distribution in endosperm

The two types of protein bodies in rice endosperm exhibit different morphologies (Figure 2). Under transmission electron microscopic observation, protein body I (PB-I) is spherical in shape with a clear margin; displays concentric rings; and is smaller in size (1-2 μm in diameter) and light in color. Protein body II (PB-II) shows irregular shape without a clear margin, and is large in size (2-3μm in size) and with high electron density (Bechtel and Pomeranz, 1978; Tanaka *et al.*, 1980; Yamagata *et al.*, 1982). In the endosperm, the number of protein bodies decreases with the distance from the surface of endosperm, and there are few protein bodies in the core of the rice kernel.

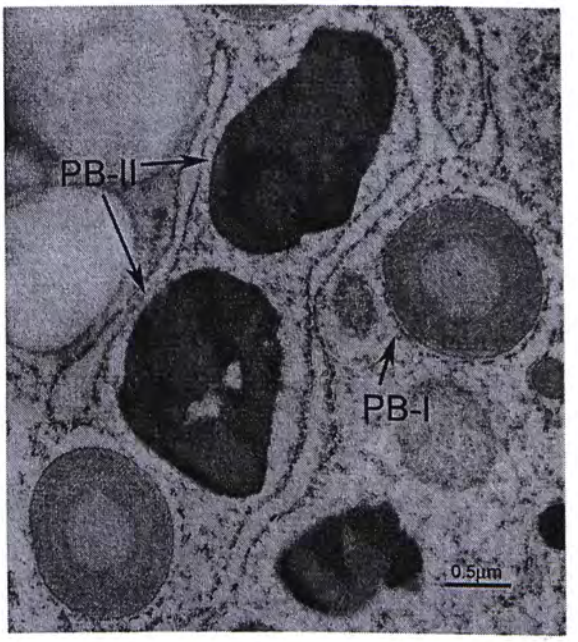


Figure 2. Two types of protein bodies in developing rice ensosperm cells Two morphologically distinct protein bodies (PBs), spherical PB-I with concentric strata and irregularly shaped electron-dense PB-II. PB-I contains prolamin and PB-II is composed of glutelin and globulin.

The rice prolamins are synthesized later than the glutelins during seed development (Yamagata *et al.*, 1982). By sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), storage proteins begin to accumulate in the Koshihikari rice endosperm by 5 DAF and these are the two subunits of glutelin (22-23 kDa and 37-39 kDa) and the 26 kDa globulin. The 10-16kDa prolamin synthesis is first detectable at approximately 8-10 days after flowering. Transmission electron microscopy and sucrose density gradient centrifugation of the starchy endosperm also showed that PB-II is formed earlier than PB-I. At 7 DAF, the protein bodies are mainly PB-II with some PB-I, and PB-I increases only gradually from 10 DAF onward (Yamagata *et al.*, 1982).

2.3 Transgenic approaches to improve the nutritional quality of rice seed proteins

2.3.1 General introduction

Seed storage proteins of cereal crops meet the major dietary protein requirement of over half of the world population. According to FAO estimate, plant sources provide 65% of the world supply of edible protein, of which 47% is from cereal grains, and the remaining 35% comes from animals. Cereals provide the cheapest sources of energy and protein to meet requirements for people in developing countries. However,

plant protein sources differ from animal sources in terms of the incomplete amino acid composition. Seed proteins in general are deficient in some essential amino acids, especially lysine and tryptophan in cereal proteins. As human beings cannot synthesize essential amino acids, complementing the essential amino acid profiles by consuming other sources of food is necessary. Non-staple foods such as fruits, vegetables, animal products and pulses are rich in micronutrients and proteins. However, such non-staple foods are simply too expensive to be eaten by the poor with low purchasing power.

Demand of non-staple foods in developing countries is growing faster than the demand for cereals, but supply is still considerable low and prices are still high. Increasing the productivity of non-staple foods through agricultural research is much more expensive than increasing the productivity of staple foods due to the large number of non-staple foods involved. Therefore, different efforts have been put to modify the cereal crops until its amino acid composition matches the balanced amino acid composition recommended by WHO for human diet.

2.3.2 Attempts to improve the nutritional quality of seed proteins

Because of the important values of seed storage protein and its AA composition, increasing their contents especially lysine is the key point in grain nutritional quality

improvement. Creation and selection of protein mutants is the major method to improve the grain nutritional quality by traditional breeding (Kumamaru *et al.*, 1988; Ogawa *et al.*, 1989). However, the limitations of this method include: low efficiency of mutant selection, the accompanying undesirable traits (Iida *et al.*, 1993; Schaeffer and Sharpe, 1990), and its minor effects on protein and lysine content enhancement.

In recent years, the rapid progress in molecular technologies provides more effective ways to improve the grain nutritional quality. Generally there are several approaches for improving the nutritional quality of seed proteins by recombinant DNA techniques, which include modification of protein sequence to elevate essential amino acid content, overexpression of elite homogeneous protein genes, transformation and overexpression of elite heterogeneous protein genes, synthesis of novel protein gene, and increase of free amino acid content (Beach and Ballo, 1992; Sun *et al.*, 1992; 1993; Mattews and Hughes, 1993; Habben and Larkins, 1995). After the first transgenic approach to improve methionine content of seed proteins by Altenbach *et al.* (1987; 1989), there has been a large flow of literature on these approaches. Various plant species such as corn (Wallace *et al.*, 1988; Ohtani *et al.*, 1991), potato (Tu *et al.*, 1998), soybean (Kho and Lumen, 1988; Saalbach *et al.* in 1994), wheat (Singh *et al.*, 1993), and tobacco (Shaul and Galili, 1992) have been reported in nutritional quality

improvement. There are also studies aiming at improving the rice grain nutritional quality.

2.3.3 Rice grain quality improvement by genetic engineering

2.3.3.1 Increase in the lysine content of rice endosperm

Compared with other cereals, rice is relatively low in protein content. At the same time, lysine is the first limiting essential amino acid in rice for human consumption. Of the major four storage proteins, glutelin contains relatively higher proportion of lysine but prolamin and albumin are low in lysine. PB-I containing prolamin is more difficult to be digested and absorbed by human than PB-II with glutelin. Furthermore, some members of the albumin and globulin fractions belonging to the α-amylase/trypsin inhibitor family are allergens to sensitive people (Alvarez *et al.*, 1995). Therefore, glutelin is considered to have higher nutritional value than other storage proteins (Tanaka *et al.*, 1975; Ogawa *et al.*, 1987).

To enhance the nutritional quality of rice grains, there are several feasible strategies: 1) to increase glutelin and decrease prolamin content; 2) to reduce albumin and globulin synthesis to diminish allergenic proteins in rice; 3) to enhance lysine content and to improve amino acid balance through regulation of key enzymes involve in lysine biosynthesis pathway and transformation with lysine-rich protein (Gao *et al.*,

2001; Liu, 2002).

Zheng *et al.* (1995) transformed rice with a gene encoding the seed storage protein, β-phaseolin, from the common bean (*Phaseolus vulgaris* L.) and found that β-phaseolin could be normally expressed and accumulated in the rice seeds under the control of the rice Gt1 promoter and was localized in PB-II. β-phaseolin has a relatively high lysine content of 6 mol% and the β-phaseolin protein was expressed as 4% of the total salt-soluble protein in the transgenic rice grains. Later, a lysine rich protein (LRP), containing a high percentage (over 10 mol%) of essential amino acid lysine, was cloned from winged bean (Sun *et al.,* 1998). This LRP gene was transformed into *Arabidopsis* and rice respectively and the lysine content of the transgenic seeds was significantly increased as a result of stable accumulation of this foreign protein (Cheng, 1999; Liu, 2002).

Foreign proteins could be highly expressed through fusion protein technique. Hoffman *et al.* (1988) first transformed a fusion gene, with a 45-bp sequence rich in methionine residues insertion into the β-phaseolin gene, into tobacco. Only 0.2% of the fusion protein at the level of normal protein was accumulated in the transgenic seeds. This low expression is probably due to the insertion sequence that may have destabilized the phaseolin trimers, resulting in the degradation of the modified phaseolin. Liu (CUHK, 2002) made use of the fusion protein approach by inserting the

LRP cDNA into the coding sequence of glutelin Gt1 structural gene, in frame, at three different regions, the acidic subunit, the basic subunit, and both subunits, resulting in three categories of chimeric fusion genes: Gt::LRP(A), Gt::LRP(B) and Gt::LRP(AB). These three transgenic rice lines are named FA, FB and FAB, respectively. The fusion proteins were highly expressed in transgenic seeds. The level of increase of lysine content in total amino acids could reach up as high as 58% in FB line when comparing with that of the wild type. Besides, several fusion polypeptides with abnormal sizes were detected in the seeds of transgenic plants of FA and FB lines, indicating that the fusion polypeptides might undergo unusual post-translational excision, while normal polypeptide of the fusion protein was observed in the FB line. Therefore FB line was selected for further study.

As glutelin is regard as more nutritious than prolamin in rice seeds because prolamin is low in lysine and is less digestible, another construct carrying the lysine-richer glutelin gene (Gt1) gene cloned from rice itself was transformed. Overexpressing glutelin in rice endosperm also improved the lysine concentration in rice protein (~30% increased in total AA) (Liu, CUHK).

Comparing the two methods in our lab, the increase in lysine content in Gt1/LRP fusion transgenic line (FB) is higher than that in Gt1 transgenic line (58% vs 30%). In both transgenic lines, extra copies of glutelin gene were introduced so that the lysine

content in both lines increase. In the FB line, LRP gene is inserted into the basic subunit of glutelin, this further increase the lysine content as the lysine content in LRP is even higher than that of glutelin.

2.3.3.2 Other examples of rice nutritional quality improvement

One of the examples is the production of β-carotene (provitamin A) in rice endosperm. Vitamin A deficiency (VAD) is a condition which afflicts millions of people in developing countries, especially children and pregnant women. Each year, it is estimated that VAD causes blindness in 250,000 to 500,000 malnourished children (WHO, 2004). Unlike photosynthetic tissues, rice endosperm contains neither β-carotene nor its carotenoid precursors (Burkhardt *et al.*, 1997). By genetic engineering, three β-carotene biosynthesis genes, phytoene synthase, phytoene desaturase, and lycopene cyclase were introduced into rice genomes (Ye *et al.*, 2000). The enzymes are expressed in rice endosperms so that geranylgernayl diphosphate (GGPP) could be converted into β-carotene and the rice grains became golden colour, with 1.6 mg/100 g endosperm of carotenoid in the transgenic seeds.

Besides, iron deficiency is considered the most serious micronutrient malnutrition problem in developing countries since approximately 2-5 billion people are affected at different levels (McPhail and Bothwell, 1992). Milled rice contains a

very low concentration of iron (0.2 to 2.8 mg/100 g endosperm). The iron content in rice might be enhanced by genetic engineering. It can be increased by expressing a protein called "ferritin" that stores iron in rice. Goto *et al.* (1999) transferred the ferritin gene isolated from soybean into rice and obtained a three-fold increase in iron concentration. The transgene was driven by the glutelin promoter GluB-1, and therefore it was exclusively expressed in the endosperm. In 2001, Lucca *et al.* were able to increase the iron content by twofold through expressing a ferritin gene from *Phaseolus vulgaris.* However, most of the iron was accumulated in aleurone cells which are lost during decortication, and therefore other strategies are necessary to enrich iron content in the endosperm of rice seeds.

2.3.4 Expression of recombinant protein in transgenic plants

Today, knowledge about subcellular localization of recombinant proteins in transgenic plants is still limited. Zheng *et al.* (1995) transformed a storage protein gene of β-Phaseolin from bean into rice using the rice Gt1 gene promoter. Results showed that β-Phaseolin was located in rice PB-II. Bagga *et al.* (1995) transferred a 15 kDa zein into tobacco and found that this protein was accumulated in a novel PBs in seeds. Recently, Yang *et al.* (2003) demonstrated that the human lysozyme coding sequence driven by rice glutelin promoters and signal peptides was expressed and the protein

accumulated in the rice PB-II. In addition, a transgenic line producing high lysozyme expression level exhibited morphologically different protein bodies with an unbalanced composition of lysozyme and native storage proteins. The study suggested that the high-level expression of recombinant protein distorted the trafficking and sorting of native storage proteins during rice endosperm development.

2.3.5 Effects of recombinant proteins on the high-lysine rice

After the production of high-lysine rice in our lab, initial studies were carried out. In general, the grain quality of transgenic rice remains similar to the wild type, except there was a notable increase in the chalkiness of the FB and GT rice. Besides, the two transgenic lines also share some other similar properties which are not present in wild type, which include glutelin precursor accumulation and repression of other seed storage proteins' expression. The fusion protein construct FB was found down regulate to most of the storage protein genes (including prolamin and globulin). Furthermore, the expression of chaperones BiP and PDI at both RNA and protein level was significantly enhanced in FB and GT transgenic plants. However, the more detailed effects of the transgenes on other genes, the localization of the recombinant proteins and their trafficking pathway have yet to be investigated.

2.4 Proteomics

2.4.1 General overview

Proteomics is playing an increasingly important role in scientific research. The best definition of proteomics is "any large-scale protein-based systematic analysis of the entire proteome or a defined sub-proteome from a cell, tissue, or entire organism" (Speicher, 2004). It originated in the mid-1990s due to two key enabling advances, availability of complete genome sequences and mass spectrometry advances that allowed high-sensitivity identification of proteins. Proteome analyses can be broadly categorized into three types of studies: quantitative protein profile comparisons, analysis of protein-protein interactions, and compositional analysis of simple proteomes or subproteomes such as organelles or large protein complexes.

There is a broad range of technologies used in proteomics, but the central paradigm is the use of 2-D gel electrophoresis (2D PAGE) followed by mass spectrometry (MS). In 2D PAGE, proteins are separated by isoelectric point (pI) in the first dimension and with molecular weight in the second dimension (O'Farrell, 1975). The individual proteins are subsequently removed from the gel and prepared, then analyzed by MS and searched of databases to determine their identities and characteristics. Fig 3 shows an overview of proteomics.

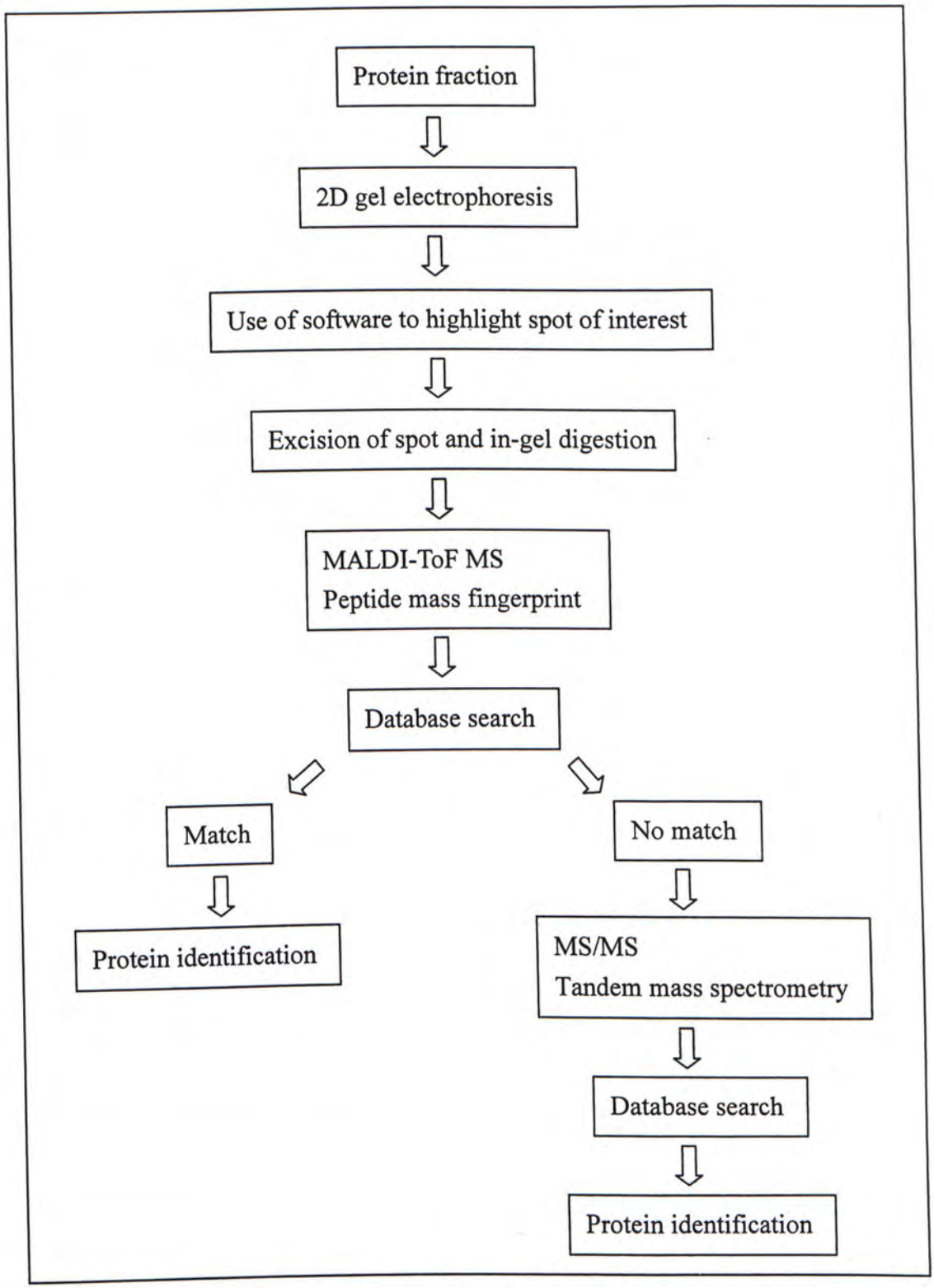


Figure 3. Overview for analysis of proteomics by mass spectrometry.
MALDI-ToF: Matrix-assisted laser desorption-ionization-time of flight

Zivy and de Vienne (2000) pointed out that proteomics play a linking role between genomics, genetics and physiology. In recent years, more and more researchers are interested in functional genomics research on sequenced genes, therefore proteomics is becoming a field in great demand with a large impact on plant biology. Proteomics is the next logical step after genome sequencing, but analysis of proteomes is much more complicated and challenging than sequencing a genome.

2.4.1.1 Two-dimensional polyacrylamide gel electrophoresis for proteome analysis

The first requirement for proteome analysis is the separation, visualization and analysis of the complex mixtures containing as many as several thousand proteins obtained from whole cells, tissues or organisms. Two-dimensional polyacrylamide gel electrophoresis (2D PAGE) is the core technology for separating complex protein mixtures. It was first developed by O'Farrell in 1975. In this technology, proteins are separated according to their isoelectric points (pI) in the first dimension and molecular weights in the second dimension.

The analysis begins with the solubilization of proteins from samples of interest using nonionic and zwitterionic detergents. This step converts the native sample into a suitable physicochemical state for isoelectric focusing (IEF) while preserving the

native charge and molecular weight of the constitute proteins. During solubilization procedure, it is most important to minimize protein modifications which might result in artifactual spots on the 2D gel. In addition, the impurities, such as lipids, salts, polysaccharides and nucleic acids that can interfere with the 2D PAGE separation should also be removed in this step. Therefore, solubilization of sample is one of the most critical factors for successful protein separation by 2D PAGE.

Sample preparation is followed by IEF using carrier ampholytes or immobilized pH gradients (IPG) to form the pH gradient. The technique of IPG is most commonly used as the first dimension. It offers high resolution and great reproducibility for IEF (Bjellqvist *et al.,* 1982). IPGs are based on the principle that the pH gradient is generated by a limited number (6-8) of well-defined chemicals (the 'Immobilines') which are co-polymerized with the acrylamide matrix.

Prior to the second dimension separation, the IPG strips are equilibrated in SDS equilibration buffer to allow the separated proteins to interact fully with SDS so that they can migrate properly during subsequent SDS PAGE. In addition, urea and glycerol are added to the equilibration buffer to reduce electroendosmotic effects which otherwise results in reduced protein transfer from the first to the second dimension. After an equilibration of the proteins, the resulting focused bands are separated according to size using traditional polyacrylamide gel electrophoresis

(PAGE) technique.

2.4.1.2 Protein visualization

After 2D PAGE, the separated proteins have to be visualized. There are various proteins staining methods include organic dyes (e.g. Coomassie blue), silver-staining, reverse staining with metal cations (e.g. zinc imidazole), fluorescence staining or labelling, and radioactive isotopes, using autoradiography, fluorography, or phosphor-imaging. Choosing of a specific protein stain emphasizes on high sensitivity (low detection limit), high linear dynamic range (for quantitative accuracy), reproducibility, and compatibility with protein analysis procedures, such as MS. Coomassie brilliant blue (CBB) staining methods have found widespread use for the detection of proteins on 2DE gels, because of its ease of use and compatibility with most subsequent protein analysis and characterization methods such as MS. Coomassie blue dye is capable of detecting as little as 0.2-0.5μg protein per spot, but this is considerably less sensitive than silver staining or fluorescence detection (Patton, 2000). Hence, typically no more than a few hundred protein spots can be visualized on a 2D gel by CBB staining, even if milligram amounts of protein have been loaded onto the gel. CBB in colloidal dispersions (Neuhoff *et al.*, 1988) and modifications (Candiano *et al.*, 2004) have been reported to be more sensitive than the classical CBB

stain, but are still less sensitive than the majority of chemical stains employed in 2D PAGE for proteomics. Silver staining methods (Oakley *et al.*, 1980; Merril *et al.*, 1984) are many times more sensitive than CBB. The detection limit is as low as 0.1 ng protein per spot. However, silver staining methods are less reproducible than CBB stains due to the subjective endpoint of the staining procedure which makes them less suitable for quantitative analysis. Silver staining method using aldehyde-based fixatives/sensitizers is the most sensitive one, but prevent subsequent protein analysis (e.g. by MS) due to protein cross-linkage. If aldehydes are omitted in the fixative and in the subsequent gel impregnating buffers (except in the developer), microchemical characterization by PMF is possible. Several silver staining protocols compatible with mass spectrometry have been published (Shevchenko *et al.*, 1996; Mortz *et al.*, 2001).

Due to the shortcomings of organic dyes for visualization and quantification of proteins, fluorescent detection of proteins has gained popularity for proteome analysis. Two major approaches for the fluorescent detection of proteins on 2D gels are currently practiced. These are: (i) pre-electrophoretic derivatization of proteins with fluorophores prior to the IEF, such as the cyanine-based dyes (Unlu *et al.*, 1997) which is commercially available as CyDyes (Amersham Biosciences), and (ii) post-electrophoretic labeling after SDS-PAGE, by intercalation of fluorophores into the SDS surrounding the proteins, such as SYPRO Ruby (Berggren *et al.*, 2002).

Fluorescent staining methods have a comparatively wide linear dynamic range (>10^3) and are relatively easy to perform. Furthermore, most fluorescent staining procedures are compatible with subsequent protein identification methods such as MS.

2.4.1.3 Computer-aided image analysis

One of the key objectives of proteomics is to identify the differential expression between control and experimental samples run on a series of 2-D gels. That is, the protein spots have changed in abundance (increased or decreased in volume). Once these gel features have been found, the proteins of interest can be identified using MS. This goal is usually accomplished with the help of computer-aided image analysis systems (Dowsey *et al.*, 2003). The typical sequence for image analysis is the following: (i) digital image acquisition, once the gel has been stained, the image can be digitized with flat-bed scanners, laser scanners, Charge Coupled Device (CCD) cameras or Fluor Imagers, depending on the protein visualization method used. Image normalization and background subtraction are also performed; (ii) spot detection and quantification; (iii) gel matching, *i.e.* an initial user guided pairing of a few spots between the reference and sample gels. The sample gel is then warped to align the landmarks and the rest of the spots were subjected to automatic pairing; (iv) identification of differentially expressed spots; (v) data presentation and interpretation;

and (vi) construction of 2D gel database. To analyze and document the separated and identified proteins, image analysis is essential. A number of software packages are available including the most widely used, such as ImageMaster 2D Platinum (Amersham Pharmacia Biotech, Sweden), Phoretix 2D Advanced (Nonlinear Dynamics Inc, USA), MasterScan (Scanalytics, USA), Melanie 3 (GeneBio, Switzerland; http://www.genebio.com/Melanie.html) and PDQuest (Bio-Rad Laboratories, USA).

2.4.1.4 Mass spectrometry-based methods for protein identification

Mass spectrometry (MS) techniques for proteome analysis are roughly divided into two major categories, peptide mass fingerprinting (PMF) or tandem mass spectrometry (MS/MS). Peptide mass fingerprints are the fastest method for identifying proteins recovered from 2D PAGE. The basic principle of this technique compares the measured peptide masses from MALDI-TOF spectra, and compares them against calculated peptide masses from database entries. Every protein results in a unique set of peptide masses after cleavage with a specific protease. Depending on the mass accuracy and mass resolution of the instrument, only a few peptide masses are sufficient for reliable protein identification. Analysis of protein by PMF is particularly successful with organisms containing simple genomes and few splice

variants. With more complex organisms, such as mammals, the probability of identifying a protein from PMF is lower and can be subjected to further analysis by tandem mass spectrometry. MS/MS data is obtained by analyzing daughter ions generated by peptide fragmentation using triple quadrupole, q-TOF, TOF-TOF, or ion trap instruments, with MALDI, capillary LC/MS, or nanospray interfaces.

2.4.1.5 Database search

For protein identification, search engines were used to search protein sequence databases using mass spectrometry data. Generally, there are two types of experimental data, one is peptide mass from the digestion of a protein by an enzyme of known cleavage specificity, and the other is using tandem mass spectrometry (MS/MS) data from one or more peptides. Algorithms are developed for protein identification based upon correlating experimental data with calculated peptide masses or fragment ion mass values, derived from proteins existing in sequence databases. A score (or ranking) is calculated to provide a measure of fit between the experimentally derived and calculated peptide masses. If the "unknown" protein is present in the sequence database, then the aim is to pull out the correct entry. If the sequence does not contain the unknown protein, then the aim is to identify those entries which exhibit the closest homology, often equivalent proteins from related species. By using an appropriate

scoring algorithm, the closest match can be identified.

Several search engines were capable for protein database searching, such as MASCOT (Perkins *et al.*, 1999) (http://www.matrixscience.com/), SEQUEST (Eng *et al.*, 1994) (http://fields.scripps.edu/sequest/), ProFound (Zhang and Chait, 2000) (http://prowl.rockefeller.edu/) and Protein Prospector (Clauser *et al.*, 1999) (http://prospector.ucsf.edu).

The most commonly used algorithms for mass spectrometry based protein identification are MASCOT and SEQUEST. MASCOT is based on the MOWSE algorithm (Pappin *et al.*, 1993) that uses the normalized distribution frequency of peptides in the sequence database. It supports all three types of search, including: PMF, sequence query, and MS/MS fragment ions search. SEQUEST can only use data from MS/MS fragmentation spectra of un-interpreted peptide. A cross-correlation function is calculated between the measured fragment mass spectrum and the proteins in the database. It is used to score the proteins in the database.

2.4.1.6 Protein sequence database

Protein sequence databases are also very crucial for protein identification since different databases contain different protein sequences. The commonly used protein sequence databases are listed below.

NCBInr

NCBInr is a comprehensive, non-identical protein and nucleic acid database maintained by NCBI (the National Center for Biotechnology Information) for use with their search tools BLAST and Entrez. The entries in the protein database, nr, have been compiled from GenBank CDS translations, PIR, SWISS-PROT, PRF, and PDB. NCBI has made strong efforts to cross-reference the sequences in these databases in order to avoid duplication.

MSDB

MSDB (Mass Spectrometry protein sequence DataBase) is a non-identical protein sequence database maintained by the Proteomics Department at the Hammersmith Campus of Imperial College London. MSDB is designed specifically for mass spectrometry applications.

Swiss-Prot

SWISS-PROT is a curated protein sequence database which strives to provide a high level of annotations (such as the description of the function of a protein, its domains structure, post-translational modifications, variants, etc), a minimal level of redundancy and high level of integration with other databases. It has been maintained collaboratively by the Department of Medical Biochemistry of the University of Geneva and the EMBL Data Library (now the EMBL Outstation of The European

Bioinformatics Institute - EBI).

PIR

The PIR (Protein Information Resource) database was initiated at the National Biomedical Research Foundation (NBRF) as a collection of sequences for the study of evolutionary relationships among proteins. The database is now an international collaboration of three data centers: the NBRF, the Munich Information Center for Protein Sequences (MIPS), and the Japan International Protein Information Database (JIPID). The three centers cooperate to produce and distribute a single database of 'wild-type' protein sequences.

The TIGR rice genome annotation database

The TIGR Rice Genome Annotation Database (http://www.tigr.org/tdb/e2k1/osa1/) (Yuan *et al.*, 2003) provides sequence and annotation data for the 12 rice chromosomes. This is a downloaded protein database which can be added into the MASCOT search engine to conduct the local protein database search. The pseudomolecules (virtual contigs) were constructed for each of the rice 12 chromosomes by resolving discrepancies between overlapping BAC/PAC clones, trimming the overlap regions at junction points in which the gene models are least disrupted, and linking the unique sequences to form a contiguous sequence.

2.4.2 Plant proteomics

Proteome analysis is becoming a powerful tool in the functional characterization of plants. In fact, two-dimensional electrophoresis techniques were established early on for barley (Görg *et al.*, 1988; Hurkman and Tanaka, 1988; Flengsrud and Kobro, 1989). Due to the development of sensitive techniques of mass spectrometric protein analysis and the acquisition of genome sequence information, proteome approaches open up new perspectives to analyze the complex functions of model plants and crop species at different levels. Within the past few years, a number of plant proteome studies have been published, some of them more descriptive, exploring global patterns of proteins present in a given tissue, organelle, or stage of development, including barley seed (Kristoffersen and Flengsrud, 2000), wheat grain endosperm (Skylas *et al.*, 2000), maize leaves (Porubleva *et al.*, 2001), Arabidopsis mitochondria (Millar *et al.*, 2001), rice leaf, root and seed (Koller *et al.*, 2002) and Arabidopsis chloroplast envelope membrane proteins (Ferro *et al.*, 2003). Apart from these, proteomic analyses of plant organs or tissues were also applied to investigate the responses to developmental changes or environmental factors on protein patterns and to compare lines with different genetic backgrounds. Wheat endosperm protein populations were compared at two developmental stages (Vensel *et al.*, 2005), which provided insight into biochemical events taking place during grain development and highlight the value

of proteomics in characterizing complex biochemical processes. Grain filling and seed maturation are also important processes that intensively studied in crop plants. Proteome analysis of the changes in proteins that occur during seed development of barley (Finnie *et al.*, 2002) identified some proteins which are involved and related them to the processes known to be taking place in the seed. Proteins in rice caryopses are essential to development and grain quality. A group of researchers (Lin *et al.*, 2005) not only profiled the protein expression patterns during rice caryopsis development, but also found some proteins associated with grain quality especially in a high temperature environment. These results could benefit the understanding of functions of rice caryopsis proteins and improvement of rice grain quality under temperature stress.

2.4.2.1 Rice proteomics

During the past couple of years, researchers have made great efforts to the study of rice proteome and remarkable progress has been made. Rice is an excellent model plant among the monocotyledonous cereal crops species as it has a relatively small genome (the rice genome is 440 Mb, the maize genome is 2500 Mb while that of barley is 4900 Mb), and it is suitable for efficient genetic analysis and transformation. The completion of the genome sequence of the rice two major cultivars, japonica type

and indica type (Goff *et al.*, 2002; Yu *et al.*, 2002), and subsequent prediction of the protein complement provide researchers with a unique opportunity to study globally expressed protein profiles in a given tissue. Several studies have dealt with mapping of proteomes for complex samples from rice, such as embryo and endosperm (Komatsu *et al.*, 1993); root (Zhong *et al.*, 1997); cultured suspension cells (Komatsu *et al.*, 1999); anther (Imin *et al.*, 2001); and leaf sheath (Shen *et al.*, 2002). Tsugita *et al.* (1994) have systematically analyzed rice proteins from 9 tissues and 1 organelle (leaf, stem, root, germ, dark germinated seedling, seed, bran, chaff, callus and chloroplast). A total of 4892 proteins was resolved, of which 137 protein spots were analyzed, and among them 56 proteins were sequenced. Besides, a systematic proteomic analysis of rice root, leaf, and seed tissues was obtained by application of both 2-D gel electrophoresis and HPLC-tandem mass spectrometry. Based on both methods, the identification of 2528 unique proteins (6296 peptides) was achieved (Koller *et al.*, 2002). The analysis of proteins from rice chloroplasts was the first organelle proteome work (Tsugita *et al.*, 1994). The proteomes of Golgi apparatus, mitochondria and other subcellular compartments have also been studied (Mikami *et al.*, 2001; Heazlewood *et al.*, 2003). Other rice proteome studies were performed to identify embryo proteins (Woo *et al.*, 2002) and to monitor the consequences of metal stress treatments (Hajduch *et al.*, 2001). The difference between the proteome of transgenic rice calli

over-expressing *YK1* gene, the homolog of maize HC-toxin reductase (HCTR) and that of wild type rice was compared (Takahashi *et al.*, 2005). Proteome analysis of cultured cells over-expressing *YK1* showed the up-regulation of several stress-related proteins, which enriched the knowledge to the genetic engineering of plants with a novel gene transfer.

The construction of rice proteome databases is also a great effort in rice proteome studies. The proteome reference maps of rice anthers were constructed according to the SWISS-2D PAGE standards, and are made available for public access by the authors at http://semele.anu.edu.au/2d/2d.html (Imin *et al.*, 2001). The extensive proteomic data generated by Koller *et al.*, 2002 on metabolic pathways in rice, are available as supplemental information on the PNAS website (www.pnas.org). The Rice Proteome Database (http://gene64.dna.affrc.go.jp/RPD/) is the first detailed proteome database to provide extensive information on at least 21 2D PAGE reference maps from rice tissues and subcellular compartment (Komatsu *et al.*, 2004). Rice etioplast protein identifications and related data were integrated into a database that is available at http://www.plprot.ethz.ch/ (von Zychlinski *et al.*, 2005).

2.4.2.2 Comparative proteomics

Comparative proteomics have been advanced in recent years with the hope to find

differences between two or more samples or treatments. Several studies have investigated changes in proteomes involving in green versus etiolated rice shoots (Komatsu *et al.*, 1999), rice leaf and leaf sheath following a jasmonic acid treatment (Rakwal and Komatsu, 2000), Arabidopsis seed germination and priming (Gallardo *et al.*, 2001; 2002), cell wall and extracellular matrix proteins from elicitor-treated Arabidopsis cell suspension cultures (Ndimba *et al.*, 2003), senescing white clover (*Trifolium repens*) (Wilson *et al.*, 2002), and rice after mechanical wounding of the leaf sheath (Shen *et al.*, 2003). However, reproducibility generally has been made cursorily in the studies. In addition, rigorous and quantitative clustering methods in comparative studies have not been evaluated and exploited. Until recently, using the de-etiolated (greening) of maize chloroplast as a model system, a general protocol that can be used to generate high quality, reproducible data set for comparative plant proteomics was developed (Lonosky *et al.*, 2004).

2.5 Hypothesis and objectives

The transgenic high lysine rice was successfully generated few years ago in our laboratory. On the basis of the previous studies, it becomes clear that there is an accumulation of glutelin precursor and chaperones BiP and PDI, while most of the storage proteins are suppressed. However, there is no study on the details of these changes and the correlations and possible impacts between the expression of recombinant proteins and the corresponding down regulated or enhanced host proteins. Based on the literature reviews and our previous study, the hypotheses of this study are:

(1) The expression of recombinant proteins promotes the expression level of some native proteins, such as BiP and PDI to cope with the increase in unfolded proteins.

(2) The expression of recombinant proteins down regulates the expression of other native proteins such as seed storage proteins possibly due to ER stress.

(3) The expression of recombinant proteins affects the deposition of storage proteins, leading to changes in the morphology of protein bodies in rice endosperms.

To test the hypotheses, the transgenic high lysine rice, FB and GT lines, will be used as study materials. Patterns of any differently expressed proteins will be studied by proteomic approach. The maturing seeds from 5, 10, 15 and 20 days after fertilization (DAF) will be collected and used in the study.

The morphology of protein bodies and the subcellular localization of the recombinant proteins, prolamin and glutelin will also be studied as a result of or in relation to the integration of the LRP and extra Gt1 genes into the rice plants.

This research will provide useful information to understand the molecular and cellular events involved in the maturation of high-lysine rice seeds, in our effort to enhance the grain quality.

Chapter 3. Materials and Methods

3.1 Materials

3.1.1 Chemicals and commercial kits

Most chemicals used in this project were of analytical grade and purchased from Sigma-Aldrich Chemical Co. (USA), Amersham Bioscience (Sweden), Roche Diagnostics Corporation (USA), Bio-Rad Laboratories, Inc (USA) or Invitrogen Corporation (New Zealand). The enzyme trypsin used in peptide digestion was from Promega Biotech Co., Ltd (USA). The 13 cm IPG strips of pH 3-10 and pH 6-12 and corresponding IPG buffer (pH 3-10 and pH 6-11) were purchased from Amersham Bioscience (Sweden). Western blot analysis was carried out using AURORA western blot chemiluminescent detection kit (USA). The glutaraldehyde, LR white resin, goat-anti-Rabbit IgG, goat-anti-Mouse IgG and uranyl acetate used for transmission electron microscope (TEM) observation were purchased from Electron Microscopy Sciences (USA).

3.1.2 Instruments

Protein gel electrophoresis was carried out using electrophoresis apparatus and power supply purchased from Bio-Rad Company (USA). The Ettan DALT*twelve* Large Format Vertical System for second-dimension SDS-PAGE was come from

Amersham Bioscience (Sweden). The freeze dry system [China scientific (HK) Ltd] was used to dry the deveoping rice seeds. The EttanTM IPGphor^{TM} Isoelectric Focusing System from Amersham Bioscience was used for first-dimension isoelectric focusing (IEF). MALDI-ToF/ToF mass spectrometry was carried out using the 4700 Proteomic Analyzer from Amersham Biosecience. Transmission electron microscope (JEM-1200 EXII, JEOL LTD, Japan) was used for TEM study.

3.1.3 Softwares

Mass spectrometry analysis was carried out using the 4700 Explorer™ Software while database searching was by the GPS Explorer™ software (Applied Biosystems, , USA). Image analysis was carried out using ImageMaster 2D Platinum, Version 5.00 (Amersham Biosciences, Sweden).

3.1.4 Plant materials

Two lines of high-lysine rice, named FB and GT lines were analyzed in this study (Fig. 4). They were constructed and developed previously by Dr. Q.Q. Liu (2002). Both of them are *japonica* rice variety wuxiangjing 9 (武香粳 9 號). The T_3 homozygous generations of transgenic plants were used for the analysis.

In FB line, extra copies of Gt1 glutelin gene with LRP insertion in the basic

subunit of glutelin gene were present. In GT line, extra copies of Gt1 glutelin gene were present (Figure 4). The lysine contents increased in total amino acid of the FB and GT lines were 58.35% and 29.74%, respectively (Liu, CUHK).

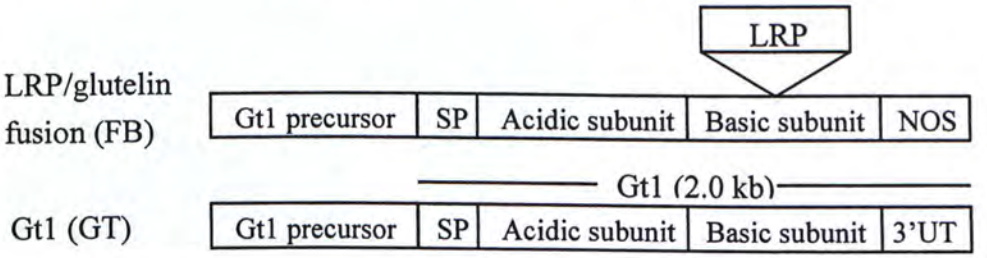


Figure 4. The transgenes of two high-lysine rice lines.
FB: The LRP gene was inserted into the basic subunit of Gt1 gene. GT: An extra copy of Gt1 gene was transformed. These genes were constructed and inserted into the super binary vector pSB130 for plant transformation (Liu, CUHK).

3.2 Methods

3.2.1 Collection of developing rice seeds

The rice plants (FB, GT and WT) were grown in the green house of the Department of Biology, CUHK. At the onset of flowering, the seed coats were marked with colour marker pens. Rice seeds were collected at 4 time points, 5, 10, 15 and 20 days after flowering (DAF) (Fig. 5 and 6). These time points are representative in the developmental process of rice seeds. All the harvested seeds were immediately wrapped in aluminium foil paper, frozen in liquid nitrogen and then stored into -80°C freezer until needed. Besides, some developing seeds at 11 to 13 DAF were collected and used in electron microscopic studies.

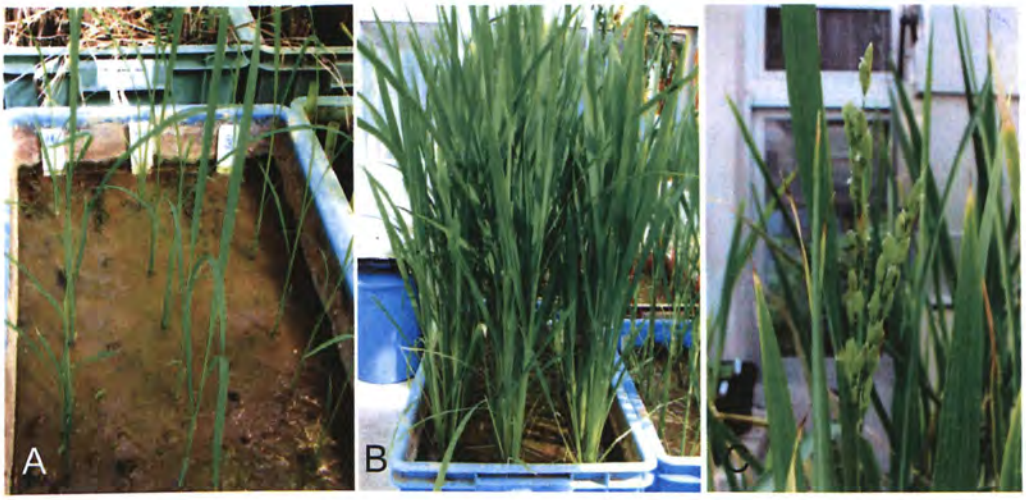


Figure 5. Planting rice materials and collecting developing rice seeds
Panels (A) Three lines of rice (FB, GT and WT) were grown in the same container; (B) The growing rice; and (C) The florets were labeled with colour marker pens.

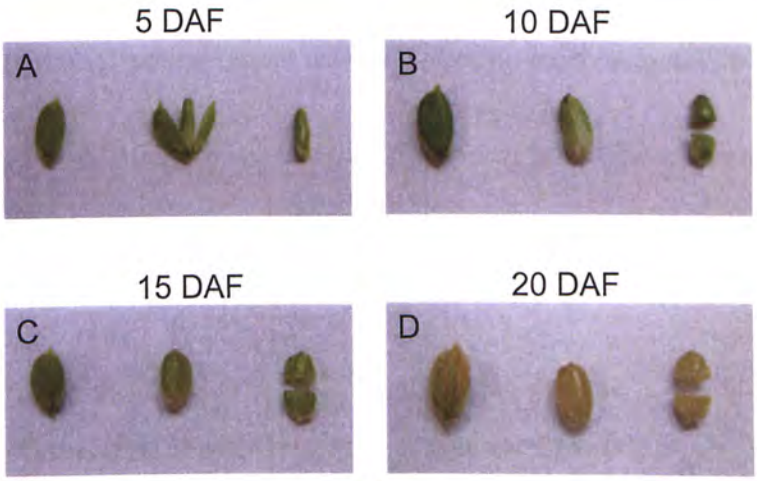


Figure 6. Developing rice seeds at 5, 10, 15 and 20 DAF.
Panels (A) Seed collected at 5 DAF (milky stage): the contents of the caryopsis are first watery but later turn milky in consistency; aleurone layer is developed; the embryo sac becomes filled with endosperm cells; and protein body appears; (B) Seed collected at 10 DAF (dough stage): cell division is essentially completed and the size of the starch granule increases most rapidly; and the milky portion of the grain turns into hard dough; (C) Seed collected at 15 DAF (yellow stage): the panicle colour changes from green to yellow; dry matter is accumulated; compound starch granule gains the maximum size; and potassium, calcium and manganese are accumulated rapidly; and (D) Seed collected at 20 DAF (mature stage): differentiation of embryo is fully completed and the grain is mature, fully developed, hard, and has turned yellow.

3.2.2 Extraction of rice seed proteins

3.2.2.1 Extraction of total protein

Extraction of total protein was carried out following the procedure of Yamagata (1982). The dried mature rice seeds of FB, GT and WT were grounded into fine powder with small mortars and pestles. For each sample, 0.05 g fine powder was used. Total protein extraction buffer (750 μl) containing 125 mM Tris-HCl, pH 6.8, 4 M urea, 4% SDS, and 5% β-mercaptoethanol was added and extracted by shaking at 1,400 rpm at 25°C for 1 hr. The extract was centrifuged at 18,000 x g in a microfuge at 4°C for 20 min. The supernatant was recovered as much as possible without disturbing the precipitate and transferred to a new Eppendorf tube. The total protein extracts were stored at -80°C until used.

3.2.2.2 Extraction of four fractions of rice seed proteins

The rice seeds of FB, GT and WT, respectively, were collected at four developing stages (5, 10, 15, 20 DAF) and freeze dried for 48 hr before protein extraction. Dry rice seeds were hand milled to remove the husks and were ground into fine powder using mortar and pestle. Proteins of different solubility were extracted from rice seeds according to Chrastil (1994). It consists of using four extraction buffers to extract different kinds of seed protein successively. This method could simplify the protein

mixture for better characterization in 2D PAGE. For each sample, 60 mg of seed power was used. Nine hundred microlitres of double-distilled water were added and water-soluble proteins were extracted with shaking at 1,400 rpm at 25°C for 1.5 hr. After centrifuged at 18,000 x g at 4°C for 10 min, the supernatant was recovered as much as possible without disturbing the precipitate. The precipitate was used for extraction of salt-soluble proteins, by mixing with 900 µl 0.05M NaCl and shaking at 1,400 rpm at 25°C for 1.5 hr. The mixture was centrifuged at the same conditions as described above. The upper phase was recovered into a new Eppendorf tube. Nine-hundred microlitres 60% n-propanol was added into the retained precipitate. This was followed by shaking at 1,400 rpm at 25°C for 1.5 hr and centrifuging at 4°C at 18,000 x g for 10 min. The supernatant was retained as alcohol-soluble protein fraction. Finally, the precipitate was shaked with 900 µl 0.02N NaOH to obtain the alkaline-soluble protein fraction. After extraction, the water-soluble proteins were further purified by dialysis using the Slide-A-Lyzer 2K Dialysis Cassette (Pierce Chemical Company). The samples were dialyzed for 2 h at room temperature with 600 ml of double distilled water. The ddH_2O was discarded and replaced with fresh ddH_2O for additional 2 h. The ddH_2O was replaced again and dialyzed overnight at 4°C. The dialyzed protein solution was transferred to a new Eppendorf tube and centrifuged at

18,000 x g at 4°C for 10 min. The supernatant was collected. All the protein samples were stored at -80°C until used.

3.2.3 2D gel electrophoresis

3.2.3.1 Protein precipitation and quantification

2D PAGE was performed according to the manufacturer's instructions (Amersham Biosciences, Sweden). The four fractions of protein extracts were concentrated by adding 4 volumes of precipitating solution [pure acetone, 0.3% dithiothreitol (DTT)] for 2 h at -20°C. After centrifuged at 18,000 x g at 4°C for 10 min, the supernatant was decanted and the white pellet was dried by vacuum. Finally, the pellet was resuspended in appropriate volume of rehydration solution containing 8 M urea, 4% (w/v) CHAPS, 0.5% (v/v) carrier ampholyte, 40 mM DTT, and 0.002% bromophenol blue. The concentrations of protein dissoived in the rehydration solution were estimated with the Bicinchoninic Acid Protein Assay Kit (Sigma, USA), using bovine serum albumin (BSA) as standard. Five microlitres of protein sample were 10-fold diluted, 1 ml of working color reagent (reagent A: reagent B, 50:1) was added and incubated for 30 min at 37 °C. At last, the absorbance of each sample and standard was read at 562 nm.

3.2.3.2 Isoelectric focusing (IEF)

Proteins were mixed with 250 µl rehydration solution including 8 M urea, 0.5% CHAPS, 0.2% DTT, 0.5% IPG buffer (pH3-10), 0.002% bromophenol blue and shaked at 1,400 rpm at 25°C for 1.5 hr and centrifuge at 4°C at 18,000 x g for 10 min. The supernatant was applied into the strip holder and the IPG strip was positioned in the strip holder. The IPG cover fluid was overlaid on the IPG strip to minimize evaporation. The first-dimension isoelectric focusing was performed by rehydration for 8 hr, followed by holding at 30 volt for 6 hr, 500 volt for 30 min, 1000 volt for 30 min, and then 3000 vhr, 4000 vhr and finally IEF at 6000 vhr. After IEF, the IPG strips were stored in individual glass test tubes at -80°C until used.

3.2.3.3 IPG strips equilibration

Before carrying out the SDS-PAGE, the IPG strips in individual tubes were equilibrated for 20 min with 7.5 ml equilibration buffer containing 50 mM Tris-HCl, pH 8.8, 6 M urea, 30%(v/v) glycerol, 2%(w/v) SDS, 0.002%(w/v) bromophenol blue and 75 mg DTT (freshly prepared) and followed by 20 min in 7.5 ml equilibration buffer with 187.5 mg iodoacetamide (IAA) instead of 75 mg DTT. These treated IPG strips were ready for the second-dimension SDS PAGE.

3.2.3.4 Second-dimension SDS PAGE

The second-dimension SDS PAGE was run on a 12% polyacrylamide gel using the PROTEAN II XL Vertical Electrophoresis Cells (Bio-rad). The equilibrated IPG strips were inserted down to contact the gel slab. This was followed by sealing the IPG strips using agarose sealing solution [SDS electrophoresis buffer (25mM Tris-base, 192 mM glycine, 0.1% SDS), 0.5% agarose, 0.002% bromophenol blue]. The second-dimension SDS PAGE was performed under 80 W at room temperature until the dye front was approximately 1 mm from the bottom of the gel.

3.2.3.5 Silver staining of 2D gel

The gels were silver stained according to the following procedures. All steps were performed with gentle shaking of the staining tray. The gels were first fixed with fixation solution (40% ethanol, 10% acetic acid and 50% distilled water) for 60 min. Then the gels were immersed in sensitizing solution (30% ethanol, 6.8% sodium acetate, 0.2% sodium thiosulfate) for 60 min. This was followed by washing the gels with distilled water three times, for 20 min every time and silver reaction in a 0.25% silver nitrate solution for 30 min. After rinsed them two times (1 min each time) the gels were developed in developing solution (2.5% sodium carbonate, 0.148% formaldehyde) until appropriate protein spots were found. The gels was then

transferred to the stop solution (1.5% Na_2-EDTA) for 30 min and washed with distilled water two times, 10 min every time. Finally, the gels were preserved in preservation solution including 8.7% glycerol.

3.2.3.6 Image and data analysis

The stained 2D gels were scanned by 2D ImageScanner (Amersham Biosciences, Sweden). The spots were automatically detected by the software ImageMaster 2D Platinum, Version 5.00 (Amersham Biosciences, Sweden) with manual check. To compare the spots in different gels, the corresponding spots were automatically matched in all gels by ImageMaster with manual check. The spot quantity was expressed as its volume and automatically normalized as a percentage of the total volume of all the spots present in a gel by ImageMaster. Scatter plot analysis was done by the ImageMaster.

3.2.4 MALDI-ToF mass spectrometry (Matrix Assisted Laser Desorption Ionization-time of Flight)

3.2.4.1 Sample destaining

The available spots were excised from the silver staining gels with sterilized knife. Each protein spot was cut into small pieces and put into a new microtube, immersed in

40 μl mixture solution (30 mM potassium ferricyanide: 100 mM sodium thiosulphate, 1:1). Once the dark stain had been removed, the spots were washed with distilled water. The protein spots were then equilibrated in 200 mM ammonium bicarbonate 10 min for 2 times.

3.2.4.2 In-gel digestion with trypsin

The destained protein spots were dehydrated with 25 μl acetonitrile (ACN) 3 times, for 10 min each time and dried in speed vacuum (Gene Co. Ltd.) for 5 min. These spots were rehydrated with 10 μl trypsin in buffer (40ng/ul trypsin in 50 mM ammonium bicarbonate) and incubated on ice for 30 min. More 50 mM ammonium bicarbonate was added to cover the hydrated spots and digested at 30°C for overnight. In the second day, 20 μl 50 mM ammonium bicarbonate was added into the digested spots and sonicated for 10 min with ultrasonic cleaner [China Scientific (HK) Ltd.]. Then 20 μl of the second extraction buffer [ACN: 5% trifluoroacetic acid (TFA),1:1] was added and sonicated for 10 min again. After transferring the supernatant into a new 1 ml microtube, extraction was repeated for 2 more times with the second extraction buffer and 1 time with 10 μl acetonitrile. The supernatant extracts were combined and finally dried with speed vacuum.

3.2.4.3 Desalination of the digested sample with Zip-Tip™

The vacuum–dried sample was resuspended in 10 µl 0.1% TFA. The Zip Tip C18 (Millipore) was wetted by aspirated 10 µl wetting solution (50% ACN) into the tip and dispensed to waste for twice. The wetted Zip Tip was equilibrated by aspirating 10 µl equilibration solution (0.1% TFA) into the tip and dispensed to waste twice. After binding the sample into Zip Tip by aspirating and dispensing 10 cycles, the Zip Tip was washed with 10 µl washing solution (0.1% TFA) by aspirating and dispensing into waste for 5 times. At last, the digested peptides were eluted with 2 µl elution solution (50% ACN, 0.1 % TFA) by aspirating and dispensing for 5 times in a clean 1 ml microtube.

3.2.4.4 Protein identification by mass spectrometry and database searching

After desalting with Zip Tip, 0.5µl of the peptide solution were spotted on the AB 4700 proteomics analyzer plate and incubated at room temperature until air dried. Then 0.5 µl CHCA (α-cyano-hydroxycinnamic acid) solution was overlaid on the dried sample as a matrix. After matrix was air dried the target plate was loaded into the AB 4700 Proteomics Analyzer. Protein identification was done with ToF/ToF Optics. The mass spectrometry was used in positive ion reflector mode to generate a mass spectrum of the peptides in each sample. The software was used to process the mass

spectra and identify proteins by searching against the integrated NCBI non-redundant database and TIGR rice database (http://www.tigr.org/tdb/e2k1/osa1/). The following parameters were used for database searching with the MALDI-ToF peptide mass and MS/MS data: monoisotopic mass accuracy; max 0.2 Da mass error; one missed cleavage using trypsin; iodoacetamide modified Cys and oxidation of Met as partial modifications. MS/MS was performed on all samples and the spectra were interpreted with the aid of GPS Explorer™. Only identification results with an expectation score below 0.05 (>95% confidence) were considered as positive identifications. MS/MS was performed on all samples and the spectra were interpreted with the aid of GPS Explorer.

3.2.5 Detection of LRP fusion protein in 2D PAGE

3.2.5.1 2D gel electrophoresis

The appearance of LRP-fusion protein on 2D gel could be confirmed by Western blotting using anti-LRP antibody. The total protein extracts of FB and WT rice were precipitated by TCA/acetone precipitation method. Four volumes of precipitating solution (10 % TCA, acetone,and 0.3% DTT) was added and kept for 2 hr at -20°C. After centrifuged at 18,000 x g at 4°C for 10 min, the supernatant was decanted and the white pellet was washed by adding 200 µl of pure acetone with 0.3% DTT for 1 hr at

-20°C. After centrifuged, the supernatant was decanted and the white pellet was dried by vacuum and then was resuspended in rehydration solution. The subsequent steps were the same as those mentioned in part 3.2.3 except the IPG strips used this time were 7 cm with pH 3-10. Thus the steps in isoelectric focusing was shortened: rehydration for 8 hr, followed by holding at 30 voltages for 6 hr, 500 voltages for 30 min, 1000 voltages for 30 min and then 3000 vhr and finally 4000 vhr.

The proteins in the IPG strips were separated by 12% polyacrylamide separating gel and a 4% polyacrylamide stacking gel. Electrophoresis was run with 40 voltages for stacking gel and 100 voltages for separating gel. After about 4 hr electrophoresis, the SDS-PAGE gels were removed from gel tank and prepared for Western blot analysis or stained with silver staining solution.

3.2.5.2 Western blotting using anti-LRP antibody

After separated by SDS PAGE, the proteins were transferred onto nitrocellulose memberane using BioRad Trans-blot electrophoretic transfer cell. Western blot analysis was performed according to AURORA western blot chemiluminescent detection system. Nitrocellulose membrane with the binding proteins was washed for 5 min with PBST (1x PBS with Tween 20) and blocked for 90 min with blocking buffer containing 1x PBS and 0.2% AURORA blocking reagent. After blocking, the

membrane was incubated with LRP-specific antiserum (LRP antibody : blocking buffer, 1:5000) (provided by H.W. Wong, CUHK) for 1 hr with shaking. The membrane was washed by PBST for 2 times, 15 min for each time. This was followed by reaction with 1:5000 (v:v) anti-rabbit alkaline phosphate conjugate (GIBOCOL) blocking for 1 h and washing with PBST for 2 times, 15 min for each time. Finally, LRP was detected using AURORA chemiluminescent system.

3.2.6 Antiserum production

3.2.6.1 Purification of glutelin and prolamin proteins

Glutelins and prolamins from 2 g rice flour were extracted according to previous method. The protein from the supernatants was precipitated separately in 4 volumes of pure acetone with 0.3% DTT for 2 hr at -20°C. After centrifugation at 18,000 x g at 4°C for 10 min, the supernatant was removed and the pellet dissolved in 10 ml solvent buffer (125 mM Tris-HCl, pH 6.8, 4 M urea, 4% SDS, and 5% β-mercaptoethanol). One millilitre of concentrated proteins was fractionated on 12% SDS-PAGE gel. Three duplicate gels were performed to obtain more fractionated proteins. Gel slices containing 57-kDa glutelin precursors, and the 13-kDa prolamins were excised and stored at -20°C until they were macerated and used as antigens.

3.2.6.2 Immunization of rabbits and mice

The excised gels of prolamin or glutelin were mashed with minimal amount of 10 mM Tris-HCl, pH 7.5, in small mortar and pestle. The purity and quantity of the purified proteins were checked by SDS-PAGE, using BSA as standard. Approximately the concentration of prolamin polypeptides and glutelin prolepetides were 2 μg/μl and 0.8 μg/μl, respectively. Antibodies to the gel-purified prolamins and glutelin precursors were raised in rabbit and Balb/C mice, respectively. One rabbit and five mice were prepared by the Laboratory Animal Service Center (CUHK). For the initial immunization, Freund's complete adjuvant (Sigma, USA) was mixed with the proteins and the contents were injected into the animals. One hundred micrograms of prolamin polypeptides for rabbit and 10 μg of glutelin precursors for mice were used in each injection. Three more booster injections were given with the antigens and incomplete Freund's adjuvant (Sigma, USA) at two-week intervals. The animals were terminated and blood samples were collected one week after the third booster injection. The blood samples were allowed to clot for 1 hr at room temperature, and were centrifuged at 3000 x g at 4°C to remove the blood cells. The recovered sera were stored at –20°C.

3.2.6.3 Testing of antibody specificity

Western blotting was performed to determine the specificity of the anti-serum.

Ten microlitres wild type total protein extracts (~10 μg) was mixed with equal volume of 2X sample loading buffer (40% sucrose, 0.125M Tris-HCl, pH 6.8, 2% SDS, 2 mM EDTA, 0.01% bromophenol blue and 1% β-mercaptoethanol) and treated at 99°C for 10 min. The treated samples were spinned down and loaded into the wells of 1.0 mm slab SDS-PAGE gel. After separated by SDS PAGE, Western blot analysis was performed using AURORA chemiluminescent system. The glutelin-specific antiserum and prolamin-specific antiserum were used as the primary antibody while the corresponding anti-mouse and anti-rabbit alkaline phosphate conjugates were used as the secondary antibody. The specificity of anti-glutelin and anti-prolamin was thus detected.

3.2.7 Transmission electron microscopy (TEM)

3.2.7.1 Sample fixation and section preparation

Immature seeds (11-13 DAF) of FB, GT and WT rice were cut into half with a razor blade and the portion without the embryo was immediately fixed overnight at 4°C with 0.1% gluteraldehyde and 4% paraformaldehyde buffered at pH 7.0 with 0.1 M sodium phosphate. The fixed samples were rinsed three times in 0.1 M sodium phosphate buffer, 15 min each and then dehydrated at room temperature by a series of

ethanol concentrations as follows: incubated 30%, 50% and 70% ethanol, respectively, for 10 min. This was followed by 85%, 95% and twice of 100% ethanol for 20 min incubation. Tissues were then incubated in a gradually increasing concentration of white London Resin (LR white), starting with 1:3, 1:1 and 3:1 (v:v) LR white to ethanol for 1 hr each time, and with 100% LR white overnight at 4°C. Incubation in pure LR white was repeated once for 4 hr, after which the tissues were put in gelatin capsules for 24 hr at 60 °C to polymerize. Ultrathin sections were cut by Reichert Ultracuts microtome and mounted on grids.

3.2.7.2 TEM observation

The grids were floated on a drop of uranyl acetate stain for 30 min. The grids were then rinsed with distilled water 3 times, 1 min for each wash, and allowed to air dry. Microscopical observations were carried out with a transmission electron microscope (JEM-1200 EXII).

3.2.7.3 Immunocytochemical observation

The samples were fixed, dehydrated, embedded and cut as mentioned previously. Rice sections mounted on grids were incubated in 60 µl TBS blocking buffer with 3% BSA for 30 min at room temperature, followed by hybridization with 60 µl primary

antibody (1:30 LRP antiserum in 1x TBS with 1% BSA and 0.1% Tween 20) for 1 h at room temperature. Nonspecifically bound antibodies were removed by washing the section three times using 1x TBS buffer, 5 min each time. Then the secondary antibody, goat anti-rabbit polyclonal serum conjugated to 15 nm of gold particle, in 1:30 dilution, was incubated with the sections for 45 min at room temperature. Finally, the grids were washed 3 times, 5 min for each wash, in 1X TBS and 2 min in distilled water and then allowed to air dry. The grid was stained with uranyl acetate stain for 30 min. The grids were than rinsed with distilled water 3 times, 1 min for each wash, and allowed to air dry before viewing under the TEM.

In double labeling, the procedures were unchanged except 60 µl of the reaction buffer (1x TBS with 1% BSA and 0.1% Tween 20) containing the primary antibody of anti-glutelin from mouse and anti-prolamin from rabbit was added to the section and co-incbated for 1 h. Thirty microlitres of the reaction buffer containing 1:30 dilution of secondary antibodies conjucated with gold particles (10 nm for anti-rabbit serum and 15nm or 25 nm for anti-mouse serum) was added to the sections and these were incubated for another 1 hr.

Chapter 4. Results

4.1 Proteomic analysis of high-lysine rice

In the present study, two types of high-lysine rice, FB and GT lines, and wild type rice were used as experimental materials for the profiling study. To assess changes in the transgenic rice proteome during seed development, we performed two-dimensional SDS-PAGE on proteins of developing rice seeds from four time points, 5, 10, 15 and 20 DAF. These time points are representative in the developmental process of rice endosperm (Duan, 2003).

4.1.1 Extraction of proteins

Seed storage proteins were initially classified into four categories according to their solubility properties by Osborne (1924). The four classes are albumins (water-soluble), globulins (saline-soluble), prolamins (alcohol-soluble), and glutelins (alkaline/acid-soluble). In this study, the four classes of proteins in every developmental stage of rice seeds under study were extracted according to their solubility properties. The extraction procedure not only differentiated various categories of proteins occurring in seeds but also simplified the protein mixture for characterization.

4.1.2 The proteomic profiles of different storage proteins in developing high-lysine rice seeds

After sequential protein extractions, the four fractions of proteins from the four developmental stages of rice seeds were used for 2D PAGE. All four fractions of proteins were resolved in pH 3-10 gels. Additionally, the proteins extracted by weak alkaline were resolved in pH 6-11 gels. The 2D gels of 15 DAF were used as the reference maps for comparison. A total of sixty 2D SDS-PAGE gels was established for the three rice lines at four developing stages to study the effects of recombinant proteins on rice proteome. For each gel, the proteins expressed differently from the wild type rice were identified by MS/MS. Fig. 7-10 show the representative images of 2D gels of rice proteins. In order to obtain a reliable result from 2D images for quantitative analysis, the rice samples collected from each developing stage were extracted at least twice for protein and three replicate 2D gels were run for each sample. Silver-stained spots from each developing stage were statistically counted using ImageMaster 2D Platinum, Version 5.00. To discard experimental variations in 2-D gels between different rice lines and the different stages, the volume of each spot was automatically normalized to the percentage of total volume of all the spots present in a gel by ImageMaster 2D Platinum, Version 5.00.

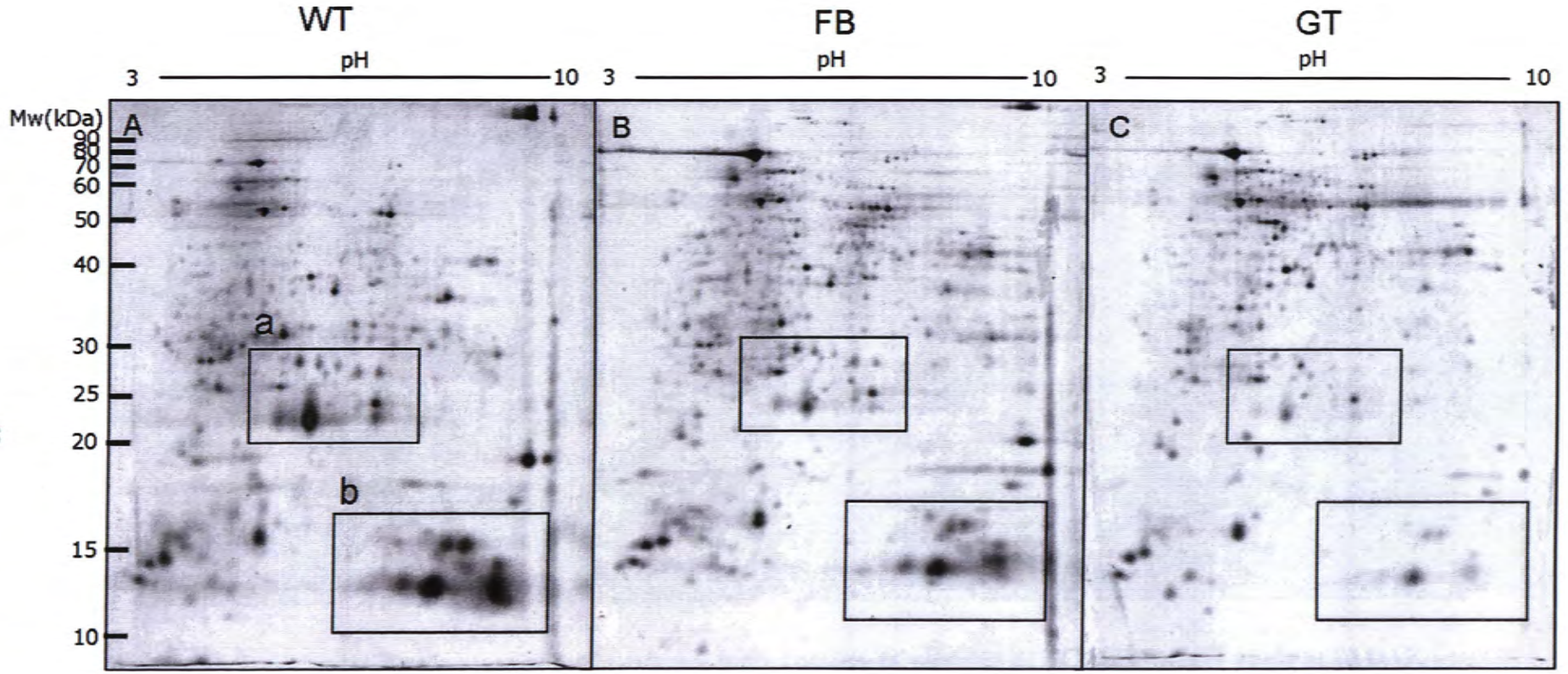


Figure 7. Two-dimensional separation of salt soluble fraction of proteins of WT, FB and GT seeds at 15 DAF. Panels (A) WT; (B) FB; and (C) GT. The boxs (a) and (b) indicate the regions of proteins expressed differently between the wild types and transgenic rices. Remarks: Sample loading; 100 μg of protein; first dimension: pH 3-10 13cm IPG strips; second dimension: 12 % SDS-PAGE; stain method: silver staining.

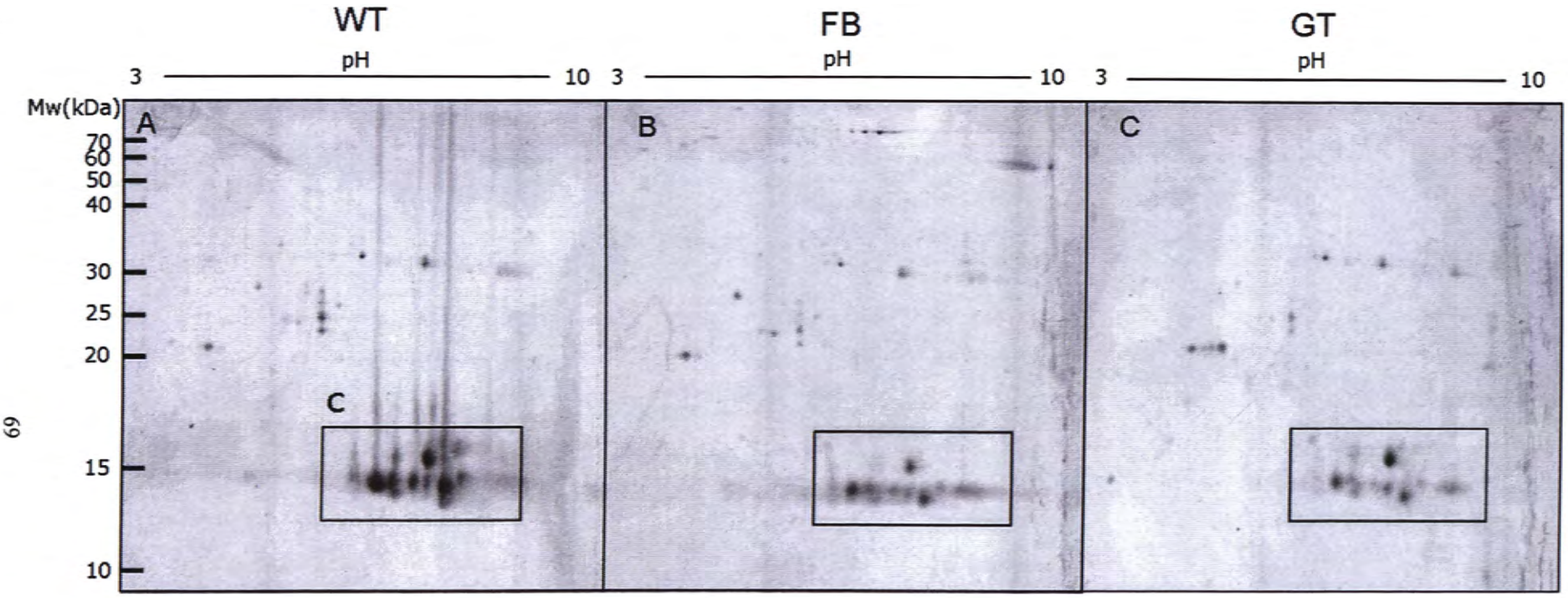


Figure 8. Two-dimensional separation of alcohol soluble fraction of proteins of WT, FB and GT seeds at 15 DAF. Panels (A) WT; (B) FB; and (C) GT. Proteins expressed differently in the wild type seeds and transgenic seeds were indicated by a box (c). Remarks: Sample loading; 50 µg of protein; first dimension: pH 3-10 13cm IPG strips; second dimension: 15 % SDS-PAGE; stain method: silver staining.

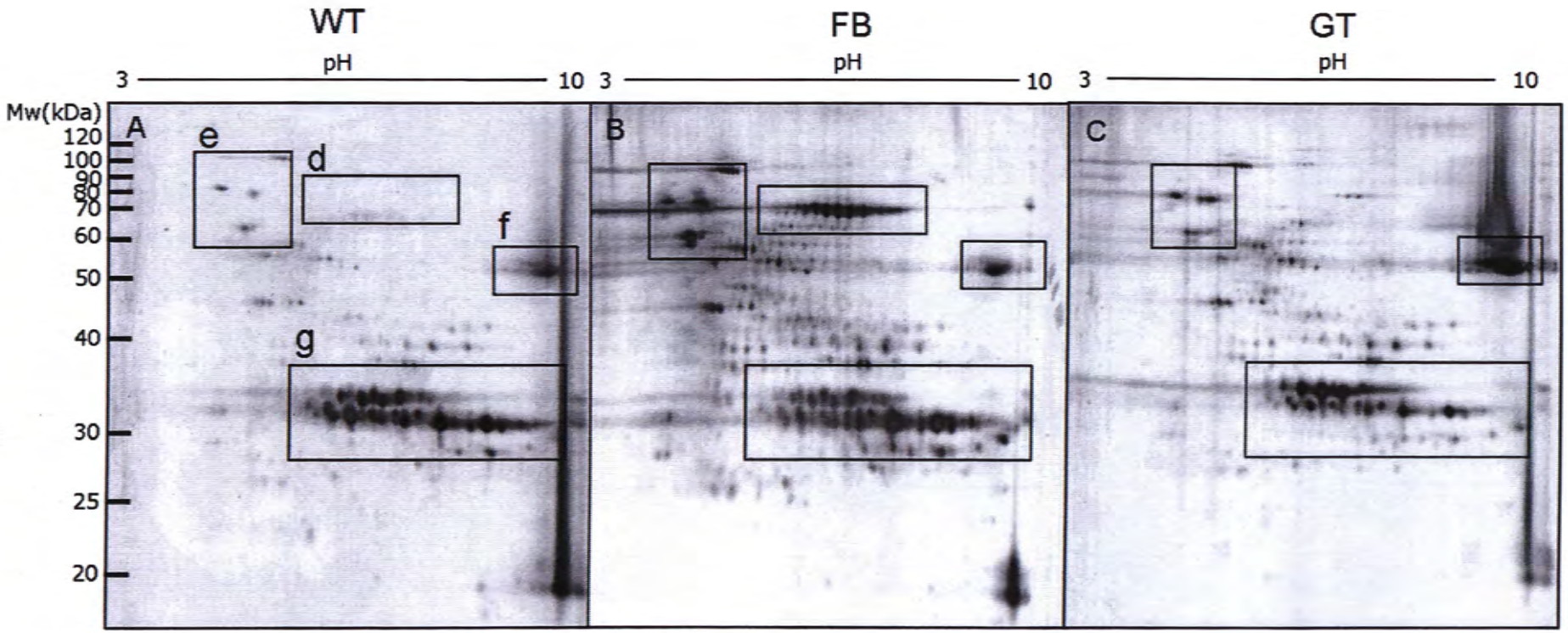


Figure 9. Two-dimensional separation of alkaline soluble fraction of proteins of WT, FB and GT seeds at 15 DAF. Panels (A) WT; (B) FB; and (C) GT. The boxes (d), (e), (f), & (g) show the regions of proteins expressed differently the between wild types and transgenic rices. Remarks: Sample loading; 100 µg of protein; first dimension: pH 3-10 13cm IPG strips; second dimension: 12 % SDS-PAGE; stain method: silver staining.

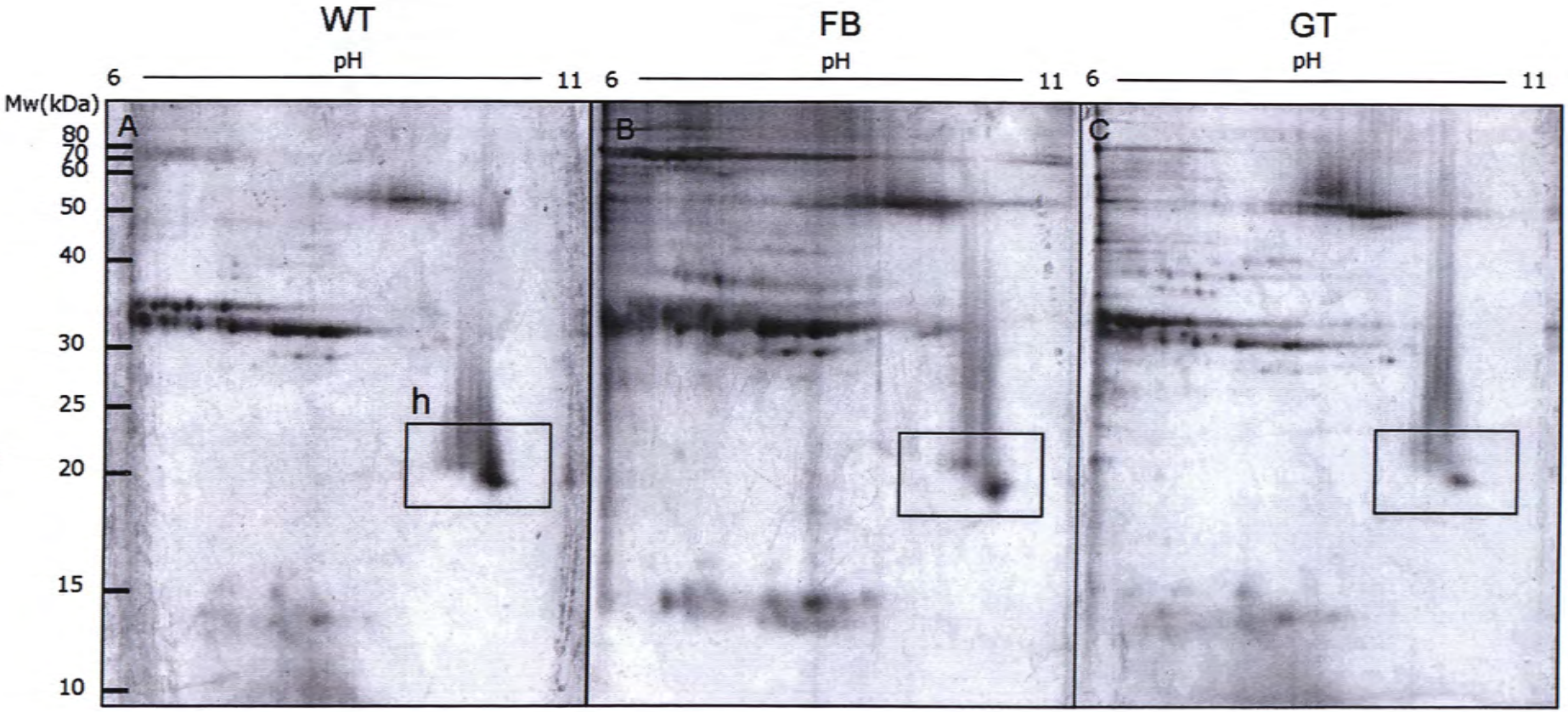


Figure 10. Two-dimensional separation of alkaline soluble fraction of proteins of WT, FB and GT seeds at 15 DAF. Panels (A) WT; (B) FB; and (C) GT. The square (h) shows the region of protein whose abundance differed in wild types and transgenic rices. Remarks: Sample loading; 100 µg of protein; first dimension: pH 6-11 13cm IPG strips; second dimension: 12 % SDS-PAGE; stain method: silver stain.

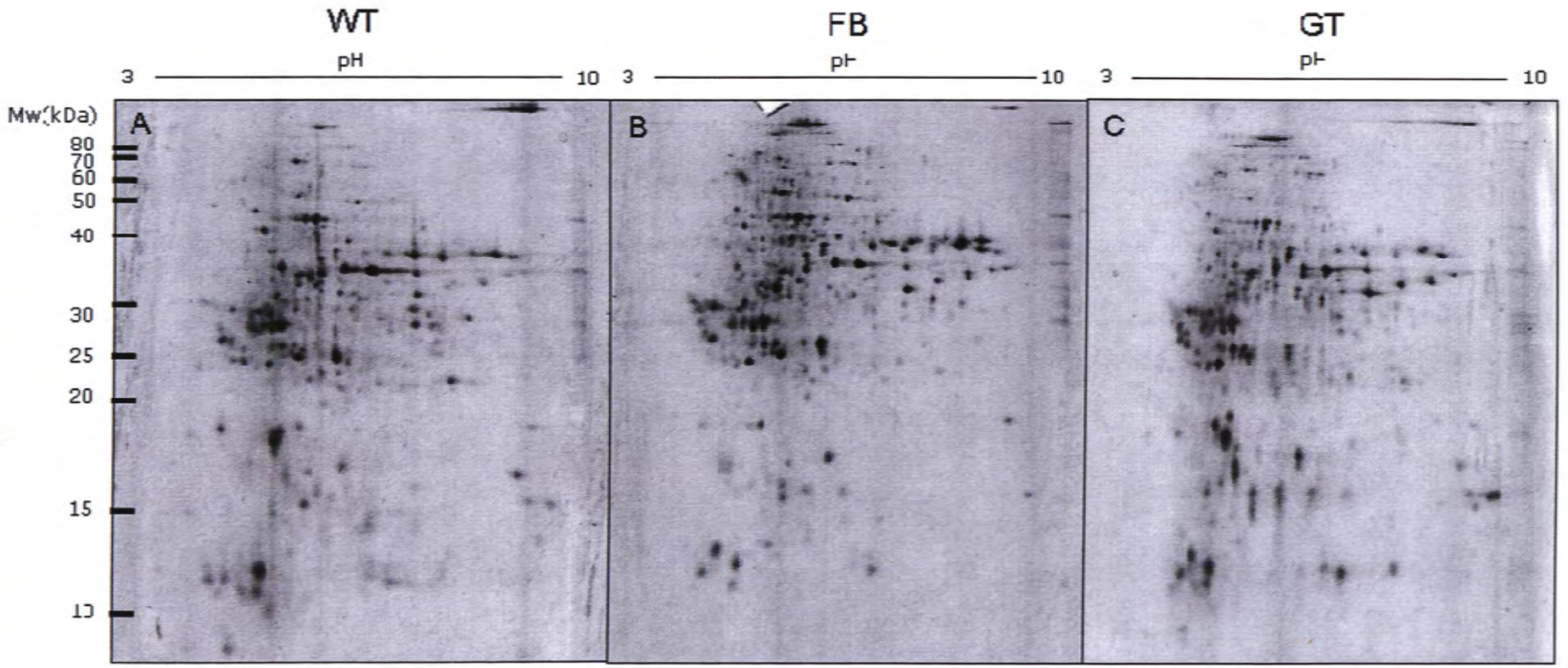


Figure 11. Two-dimensional separation of water soluble fraction of proteins of WT, FB and GT seeds at 15 DAF. Panels (A) WT; (B) FB; and (C) GT. The gels show that there are no detectable changes in the expression levels of proteins between the wild types and transgenic rices. Remarks: Sample loading; 100 µg of protein; first dimension: pH 3-10 13cm IPG strips; second dimension: 12 % SDS-PAGE; stain method: silver stain.

The proteins whose abundance varied in high-lysine rice were identified and summarized in Table 2. The proteins were identified first by PMF and then confirmed by MS/MS. There were several criteria for protein identification. At least four different predicted peptide masses need to match the observed masses for an identification to be considered valid. The mass accuracy should be less than 0.2 Da, and protein coverage should be higher than 10%. And only identification results with expectation scores above 54 (>95% confidence) were considered as positive. The last criterion is that in addition to technical repeats, at least two biological repeats were carried out. Combining the PMF and MS/MS results could lead to more reliable protein identification. For each identified protein, we have provided the spot ID, accession number, protein name, calculated pI/MW, experimental pI/MW, protein coverage, protein score and functional classification of the protein.

Table 2. Summary of proteins identified by 2D PAGE

Spot ID	Extracted fraction	Protein Name	Accession number (NCBI)	Cal. MW (kDa)	Cal. pI	Exp. MW (kDa)	Exp. pI	Protein score	Protein coverage	Functional classification
1	Salt soluble	19 kDa globulin precursor	20159	21.04	7.51	22.8	5.6	91	23%	Storage protein
2	Salt soluble	19 kDa globulin precursor	20159	21.04	7.51	22.7	5.8	67	20%	Storage protein
3	Salt soluble	Allergen RA5B precursor	2118429	17.28	8.36	15.7	7.5	107	35%	Allergenic protein
4	Salt soluble	Allergen RA5B precursor	34900132	17.27	8.36	15.8	7.6	165	16%	Allergenic protein
5	Salt soluble	Trypsin inhibitor (Bowman-Birk)	476550	15.06	7.31	14.6	7.5	69	40%	Allergenic protein
6	Salt soluble	Hypothetical protein	31431093	14.54	12.13	13.6	6.7	46	20%	unknown
7	Salt soluble	Hypothetical protein	50939899	11.26	10.68	13.4	7.3	36	22%	unknown
8	Alcohol soluble	13 kDa prolamin	34900246	16.93	8.82	15.8	7.5	154	12%	Storage protein
9	Alcohol soluble	13 kDa prolamin	34900250	17.00	8.79	15.4	7.4	149	15%	Storage protein
10	Alcohol soluble	13 kDa prolamin	34900246	16.93	8.82	14.0	6.7	147	11%	Storage protein
11	Alcohol soluble	13 kDa prolamin	34900246	16.92	8.82	14.0	6.9	134	11%	Storage protein
12	Alcohol soluble	13 kDa prolamin	34900248	16.98	7.90	13.3	7.0	178	33%	Storage protein
13	Alcohol soluble	13 kDa prolamin	34900246	16.92	8.82	14.0	7.2	146	12%	Storage protein
14	Alcohol soluble	Prolamin 7	55168130	16.86	8.79	13.8	7.3	139	38%	Storage protein
15	Alcohol soluble	Prolamin 7	55168130	16.86	8.79	13.2	7.5	177	38%	Storage protein
16	Alkaline soluble	Glutelin	225710	56.16	8.93	70.5	6.2	96	17%	Storage protein
17	Alkaline soluble	Glutelin	225710	56.16	8.93	70.5	6.3	82	11%	Storage protein
18	Alkaline soluble	Glutelin II precursor	72300	56.29	8.93	70.5	6.5	143	14%	Storage protein
19	Alkaline soluble	Glutelin II precursor	72300	56.29	8.93	70.5	6.7	209	21%	Storage protein
20	Alkaline soluble	Glutelin II precursor	72300	56.29	8.93	70.5	6.8	97	13%	Storage protein

Spot ID	Extracted fraction	Protein Name	Accession number (NCBI)	Cal. MW (kDa)	Cal. pI	Exp. MW (kDa)	Exp. pI	Protein score	Protein coverage	Functional classification
21	Alkaline soluble	Glutelin	226588	56.30	8.96	70.5	7.0	205	18%	Storage protein
22	Alkaline soluble	Endosperm lumenal binding protein	2267006	73.50	5.30	74.3	5.3	430	34%	Protein synthesis
23	Alkaline soluble	Protein disulfide isomerase	7209794	33.38	4.81	63.5	5.1	145	16%	Protein synthesis
24	Alkaline soluble	Glutelin II precursor	31431938	56.27	8.93	53.1	8.3	224	22%	Storage protein
25	Alkaline soluble	Glutelin	20210	56.64	9.33	34.3	5.8	179	12%	Storage protein
26	Alkaline soluble	Glutelin	20212	56.19	9.32	34.0	6.0	104	11%	Storage protein
27	Alkaline soluble	preprolglutelin (AA -24 to 476) [Oryza sativa (japonica cultivar-group)]	20227	56.78	9.00	34.2	6.1	135	11%	Storage protein
28	Alkaline soluble	Glutelin	225959	56.16	8.56	33.8	6.2	159	14%	Storage protein
29	Alkaline soluble	preprolglutelin (AA -24 to 476) [Oryza sativa (japonica cultivar-group)]	20227	56.78	9.00	34.1	6.4	160	13%	Storage protein
30	Alkaline soluble	Glutelin	556403	56.21	9.09	33.7	6.5	136	14%	Storage protein
31	Alkaline soluble	preprolglutelin (AA -24 to 476) [Oryza sativa (japonica cultivar-group)]	20227	56.78	9.00	34.1	6.8	314	17%	Storage protein
32	Alkaline soluble	Glutelin	20208	56.04	8.81	32.6	6.1	118	14%	Storage protein
33	Alkaline soluble	Glutelin	20212	56.19	9.32	31.7	7.4	302	15%	Storage protein
34	Alkaline soluble	Glutelin	20212	56.19	9.32	31.5	7.6	220	15%	Storage protein
35	Alkaline soluble	Glutelin	20212	56.19	9.32	19.8	9.9	270	17%	Storage protein

4.1.3 Quantitative analysis of protein spots

Quantitative analysis of the protein spots varied in high-lysine rice was also carried out. For each seed developmental stage, the 2-D PAGE gels were analyzed with Imagemaster 2D Platinum, Version 5.00. The spot quantity was expressed as its volume and automatically normalized as a percentage of the total volume of all the spots present in a gel by ImageMaster. The expression of seed proteins were summarized in Table 3. The symbols "+, ++, +++, ++++ & +++++" correspond to the relative expression level ranging from, 0.1-2, 2.1-5, 5.1-10, 10.1-15 to 15.1 or above (% volume), respectively. The symbol "–" means that a protein spot was not detected. At each developmental stage, the gels from three independent samples were used to quantitatively measure the proteins.

Table 3. The relative expression level of individual protein during seed development.

Spot no.	Protein name	Line	Developmental stage			
			5 DAF	10 DAF	15 DAF	20 DAF
1	19 kDa globulin precursor	WT	-	+	++++	+++++
		FB	-	-	++	++
		GT	-	+	++	++
2	19 kDa globulin precursor	WT	+	+++	+++++	+++++
		FB	+	++	+++	+++
		GT	+	++	+++	+++
3	Allergen RA5B precursor	WT	-	+	+++	+++
		FB	-	-	+	+
		GT	-	+	+	+
4	Allergen RA5B precursor	WT	-	+	+++	+++
		FB	-	-	+	+
		GT	-	-	+	+
5	Trypsin inhibitor (Bowman-Birk)	WT	-	+	++	++
		FB	-	-	+	+
		GT	-	-	+	+
6	Hypothetical protein	WT	-	+++	++++	+++++
		FB	-	++	+++	++++
		GT	-	++	++	++
7	Hypothetical protein	WT	-	+++	++++	+++++
		FB	-	++	+++	++++
		GT	-	++	++	++
8	13 kda prolamin	WT	-	+	++	++++
		FB	-	-	+	+
		GT	-	-	+	+
9	13 kda prolamin	WT	++	+++	++++	+++++
		FB	+	+	++	+++
		GT	+	++	+++	+++
10	13 kda prolamin	WT	++	+++	++++	++++
		FB	+	+	++	++
		GT	+	+	++	++
11	13 kda prolamin	WT	+	+	+++	+++
		FB	-	+	+	+
		GT	-	+	+	+
12	13 kda prolamin	WT	+	+	+++	+++
		FB	-	+	+	+
		GT	-	+	+	+
13	13 kda prolamin	WT	+	+	+++	+++
		FB	-	+	+	+
		GT	-	+	+	+
14	Prolamin 7	WT	+	+	+++	+++
		FB	-	+	++	++
		GT	+	+	+	+
15	Prolamin 7	WT	++	+++	++++	++++
		FB	+	++	++	+++
		GT	+	++	++	++
16	Glutelin	FB	-	+	++	++
17	Glutelin	FB	-	++	++	++
18	Glutelin II precursor	FB	+	++	++++	+++
19	Glutelin II precursor	FB	+	++	++++	+++
20	Glutelin II precursor	FB	+	++	++++	+++
21	Glutelin	FB	+	+	+++	++

Spot no.	Protein name	Line	Developmental stage			
			5 DAF	10 DAF	15 DAF	20 DAF
22	Endosperm lumenal binding protein	WT	-	-	++	++
		FB	+	++	++++	++++
		GT	-	+	++++	++++
23	Protein disulfide isomerase	WT	+	+	+	++
		FB	+	++	++++	++++
		GT	+	+	++	++++
24	Glutelin	WT	-	-	+++	+++
		FB	-	-	+++	+++
		GT	-	+++	+++++	+++++
25	Glutelin	WT	-	-	+	+
		FB	-	-	+	+
		GT	+	++	++	++
26	Glutelin	WT	-	+	++	++
		FB	-	-	+	+
		GT	++	+++	++++	++++
27	preprolglutelin (AA -24 to 476) [Oryza sativa (japonica cultivar-group)]	WT	-	+	+	+
		FB	-	+	++	++
		GT	-	-	-	-
28	Glutelin	WT	+	++	+++	+++
		FB	+	+	+++	+++
		GT	++	+++	++++	++++
29	preprolglutelin (AA -24 to 476) [Oryza sativa (japonica cultivar-group)]	WT	+	++	++	++
		FB	+	++	++	++
		GT	-	+	+	+
30	Glutelin II precursor	WT	+	+++	+++	+++
		FB	+	+	++	+++
		GT	+	+++	++++	++++
31	preprolglutelin (AA -24 to 476) [Oryza sativa (japonica cultivar-group)]	WT	+	++	+++	+++
		FB	+	++	+++	+++
		GT	+	+	+	+
32	Glutelin	WT	+	+	++++	++++
		FB	+	+	+++	+++
		GT	+	+	++	++
33	Glutelin	WT	+	+++	++++	++++
		FB	+	+++	++++	++++
		GT	+	++	++	++
34	Glutelin	WT	+	+++	++++	++++
		FB	+	+++	++++	++++
		GT	+	++	++	++
35	Glutelin	WT	+	+++	++++	+++++
		FB	-	++	+++	++++
		GT	+	+	++	++

The symbols "+, ++, +++, ++++ & +++++" correspond to the relative expression level ranging from, 0.1-2, 2.1-5, 5.1-10, 10.1-15 to 15.1 or above (% volume), respectively. The data was obtained from computer analysis. The symbol "–" means that protein spot was absent.

4.1.4 Proteomic analysis of salt-soluble proteins

Salt-soluble proteins were extracted with 0.5 N NaCl. From 2D PAGE, seven proteins out of more than 248 proteins are clearly with lower abundance in the transgenic lines as compared with the wild type line (Fig. 7 regions a & b; Fig. 12A & B). These proteins are designated as proteins spot 1-7. By MS/MS, spot 1 and 2 were identified as the 21-kDa precursor of 19-kDa globulin (Table 2). In 2D gels, they show almost the same molecular weight, 22.8 kDa, but with different pI points, namely 5.55 (protein 1) and 5.75 (protein 2). In wild type seeds, they were accumulated relatively slowly from 5 to 10 DAF; but increased rapidly thereafter. In FB and GT seeds, they were expressed at a low level throughout their maturation and their expression levels were quite similar.

Proteins 3 and 4 were identified as allergen RA5B precursor while protein 5 was trypsin inhibitor (Table 2). All of them are allergen-related proteins in rice and are rice albumins even though they were present in salt-soluble fraction. Their molecular weights are 15.7, 15.8 and 14.6 kDa, respectively, and their pIs ranged from 7.5 to 7.6. The expression of the allergenic proteins was first detected at 10 DAF and reached the peak level at 15 DAF and remained high until maturity in wild type seeds. For FB and GT during seed development, they were expressed in a much lower level, especially in GT seeds. The protein spots 6 and 7 are hypothetical proteins whose functions are still

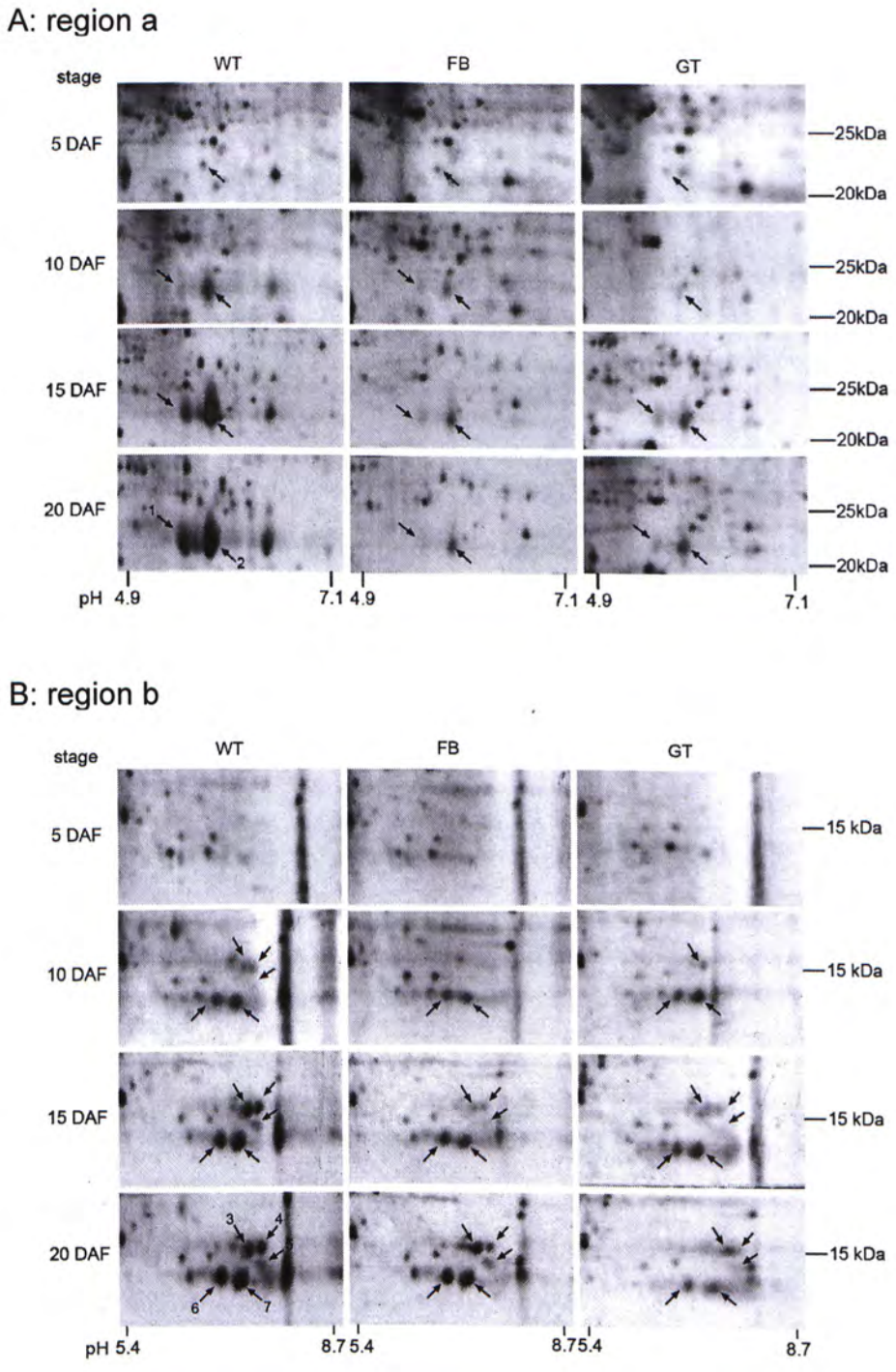


Figure 12. Comparisons of protein spots of salt-soluble fraction in 2D gels after silver-staining.

Panels (A) Spots 1 and 2 are 19 kDa globulin precursor with lower accumulation level in the FB and GT rice seeds than in the WT seeds; and (B) Spots 3, 4 and 5 are allergenic proteins. They show reduced expression level in the FB and GT rice seeds than in the WT seeds. Spots 6 and 7, with reduced expression level, are of unknown function.

unknown (Table 2). They are abundance in WT seeds but less in FB and GT seeds.

4.1.5 Proteomic analysis of alcohol-soluble proteins

Alcohol-soluble proteins were extracted with 60% n-propanol. In rice seeds, only prolamins are soluble in alcohol-water mixtures so only around 20 protein spots were observed on the gel in this fraction. When compared the transgenic lines with the wild type by 2D PAGE, we found eight protein spots, designated as 8-15, all showed lower level of accumulation in both FB and GT lines (Fig. 8 region c and Fig. 13). All of the polypeptides were identified as the members of prolamin family by MS/MS (Table 2). Their pIs range from 6.7 to 7.5. From the 2D gels, spots 8 and 9 have molecular weight of about 15 kDa while spots 10 to 15 with molecular weights of 13.2 to 14 kDa. In both FB and GT seeds, the expression of prolamins 9, 10 and 15, was detected at 5 DAF while 13, 14 at 10 DAF and 8, 11 and 12 at 15 DAF. Their expression levels were quite low. In wild type seeds, prolamins 9, 10, 13, 14 and 15 were accumulated at 5 DAF while prolamins 8, 11 and 12 at 10 DAF. They showed a common expression pattern of increasing level throughout the maturation and reached the peak level at 20 DAF. The accumulation of prolamins in WT seeds was much higher than that in FB and GT seeds.

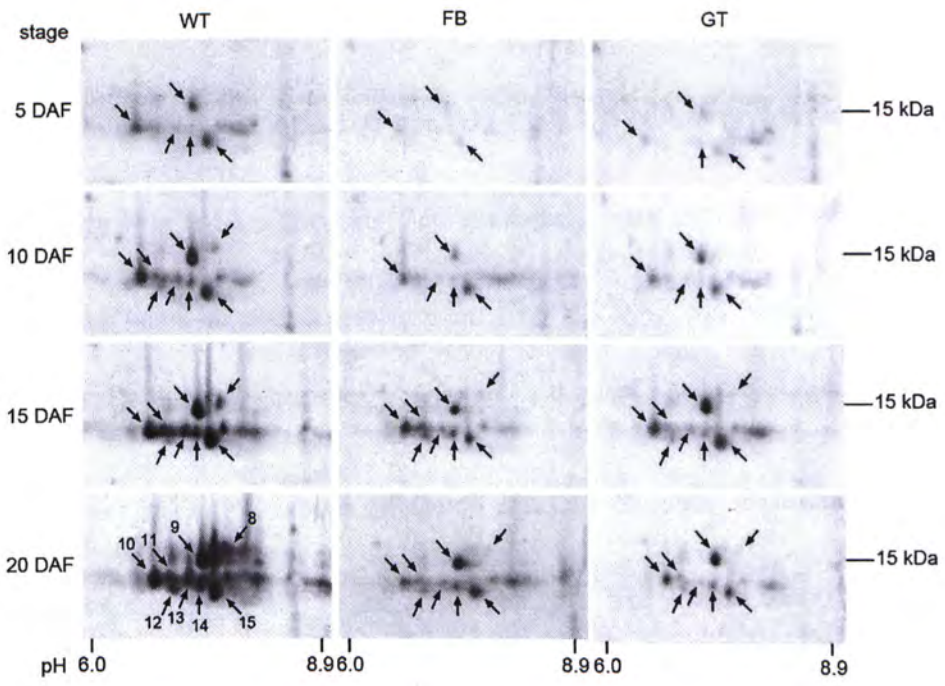


Figure 13. Comparisons of protein spots of alcohol-soluble fraction in 2D gels after silver-staining.
Spots 8, 9, 10, 11, 12, 13, 14 and 15 are prolamin polypeptides. When compared to the WT seeds, all of them show a decreasing accumulation in the FB and GT rice seeds during seed development.

4.1.6 Proteomic analysis of alkaline-soluble proteins

Alkaline-soluble proteins including glutelin were extracted with 0.02N NaOH. Among the more than 156 protein spots on the gels we found some variations in 20 of them are individually designated as proteins 16-35. Proteins 16-21 were only detected in FB rice seeds (Fig. 9 region d; Fig. 14A). By MS/MS, they were identified as glutelin (Table 2) and considered as the lysine-rich protein (LRP) fusion polypeptides. They had an apparent molecular mass of 70.5 kDa, which is consistent with the

theoretical value of LRP fusion protein (74 kDa). This fusion protein contains the LRP (17 kDa) inserted into the basic subunit of glutelin precursor (57 kDa). Immunoblotting was also performed to confirm the LRP fusion protein identity by anti-LRP antibody (Fig. 14A). The LRP fusion proteins were first detected at 5 DAF and the expression level increased greatly from 10 to 15 DAF and then slowed down from 15 to 20 DAF.

In Fig. 9 region e, two spots exhibited marked increase in intensity in the transgenic rices when compared to the wild type rice. They were identified as BiP (spot 22) and PDI (spot 23), both involving in protein folding/sorting (Table 2). The experimental molecular weight of BiP is 74.3 kDa with a pI of 5.3 which are coincided with the calculated ones. The experimental molecular weight of PDI is 63.5 kDa but the calculated MW is 33.38 kDa. This is because PDI is always in dimmer form, linking by intermolecular disulphide bond. In the wild type seeds, PDI appeared at 5 DAF and was expressed stably until maturation while BiP appeared at 10 DAF and was accumulated until maturation (Fig. 14B). The accumulation of BiP and PDI is greater in FB and GT seeds when compared to WT seeds. In FB and GT seeds, BiP and PDI were expressed from 5 DAF and the amount of these two proteins increased drastically after 10 DAF. In FB seeds, the expression of BiP and PDI declined

somewhat during 15 to 20DAF. Temporally, their expression patterns are similar to that of the LRP fusion proteins, spot no. 16-21 (Fig. 14A).

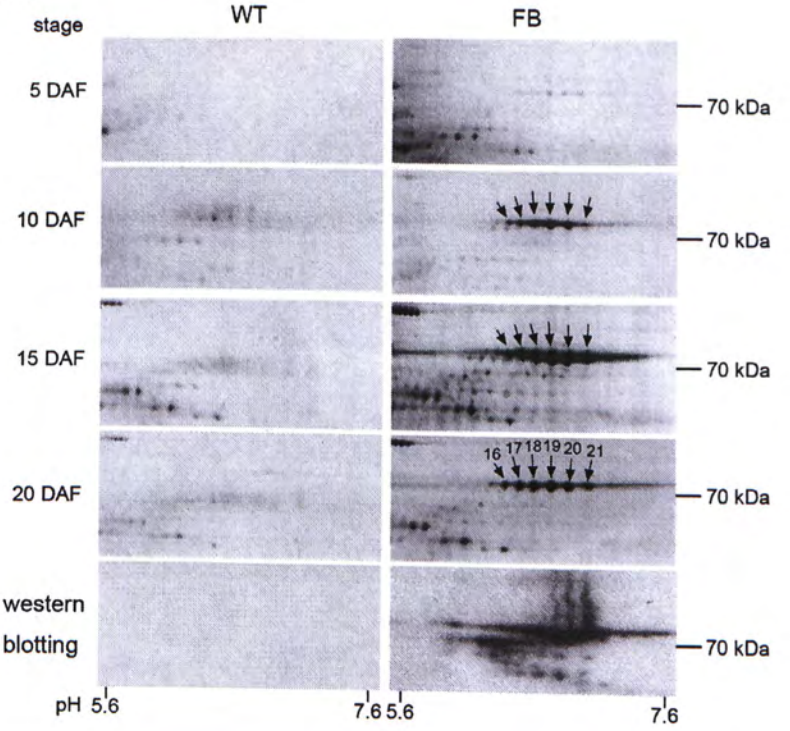
In Fig. 9 region f, a spot (# 24) showed extremely high intensity in the 2D-gel of GT rice seeds. It was identified as the glutelin precursor (Table 2). Its size is 53.1 kDa with pI of 8.3. Judging by the spot intensity, the amount of glutelin precursor is much greater in GT seeds than in FB and WT seeds (Fig. 14C). In GT rice, the glutelin precursor appeared at 10 DAF while it was detected at 15 DAF in WT and FB seeds.

In the 2D-PAGE analysis, the acidic subunits can be separated into 23 spots with molecular weights ranging from 31 to 34 kDa and pI ranging from 5.8 to 7.6. Among the 23 spots(Fig. 9 region g), 10 spots changed in their expression in transgenic lines. They are designated as spots 25 to 34 (Table 2; Fig. 14D). The polypeptide of spot 25 was expressed and accumulated at a higher level in GT seeds but at a much lower level in WT and FB rice. Likewise, the polypeptides of spots 26, 28 and 30 were of high amount in GT rice. They were first expressed at 5 DAF and their accumulation level rose to the highest level at 15 DAF and remained in the later developing stage. From the results, polypeptides 25, 26, 28 and 30 should be encoded by Gt1 gene so that their expression was greatly enhanced in GT rice. In the WT and FB samples, spots 26, 28 and 30 were expressed at a much gentle pattern during maturation. On the other hand, the accumulation of some acid subunit polypeptides was inhibited in GT seeds. Spots

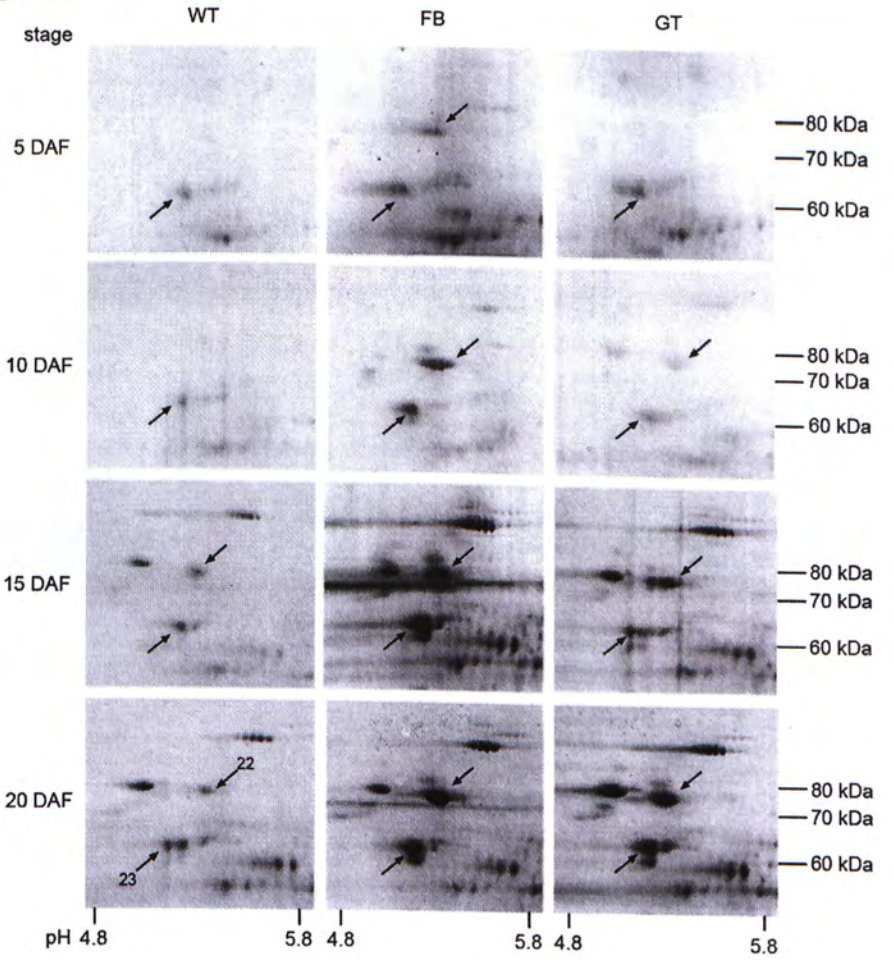
27 and 29 were hardly found in the 2D gels while spots 31, 32, 33 and 34 were expressed at a lower level.

The 2D PAGE was also carried out using pH 6-11 IPG strips. It is because the basic subunits of glutelin could not be separated in pH 3-10 IPG strips. In figure 10 region h, a spot was found with a weaker expression in GT rice when compared to wild type rice. By MS/MS, it was identified as glutelin. The apparent molecular weight of it on the 2D gels is 19.8 with pI 9.9. It is thus considered as the basic subunit of glutelin. The expression level of glutelin basic subunit was quite similar in WT and FB rice while that of GT rice was shown to be down-regulated.

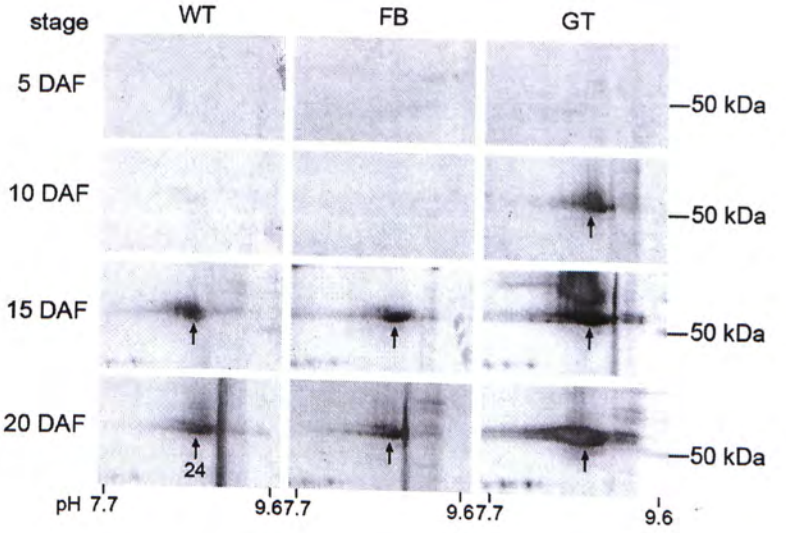
A: region d



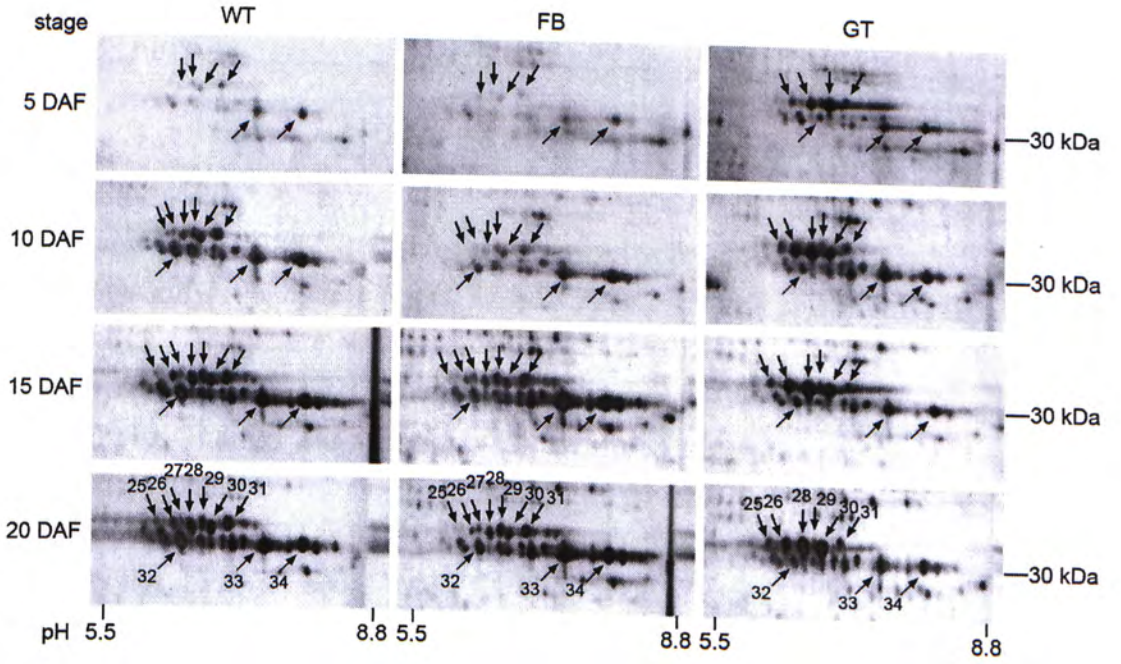
B: region e



C: region f



D: region g



E: region h

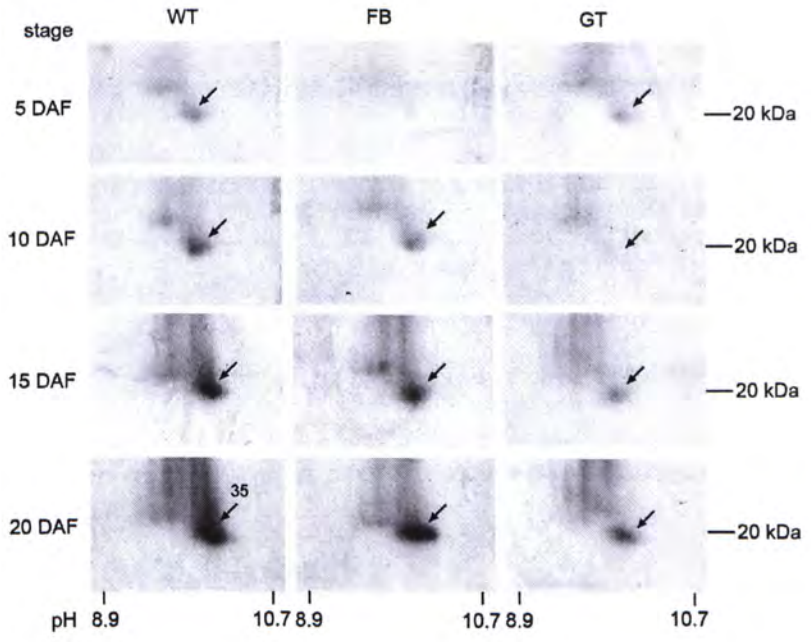


Figure 14. Comparisons of protein spots of alkaline-soluble fraction in 2D gels after silver-staining.

Panels (A) Protein spots 16-21, appeared in FB rice seeds only, are identified as glutelin by MS/MS sequencing. From their molecular weights and western blot analysis using anti-LRP antibody, the protein spots are regarded as the LRP-fusion protein; (B) Spots 22 and 23 are endoplasmic lumenal binding protein (BiP) and protein disulfide isomerase (PDI), respectively. Both of the spots are accumulated at higher levels in FB and GT lines corresponding to the WT line; (C) In GT rice seeds, spot 24, identified as glutelin precursor, shows an obvious increase in its expression during seed development; (D) In GT rice seeds, four spots, 25, 26, 28 and 30, increase significantly in abundance with respsect to the WT and FB rice seeds, while the expression of other proteins 27, 29, 31, 32 and 33 are down-regulated in GT sample; and (E) Spot 35, the glutelin basic subunit, shows lower expression level in GT rice seeds when compared to WT and FB rice.

4.1.7 Proteomic analysis of water-soluble proteins

Water-soluble proteins were extracted with distilled water. Albumins belong to this group of proteins. From the gels, there are no significant differences in the accumulation levels of water soluble proteins in both FB and GT transgenic lines and the wild type (Fig. 11).

4.1.8 Comparison of changes in protein expression patterns of specific proteins in the high-lysine rice

After 2-D gel electrophoresis, quantitative analysis of the protein spots was performed, as mentioned in 4.1.3. The expressions of proteins were also compared by graphical method according to the results of quantitative analysis to elucidate their changes after the insertion of foreign genes (Fig. 15 & 16). The graphs will be discussed in the discussion part.

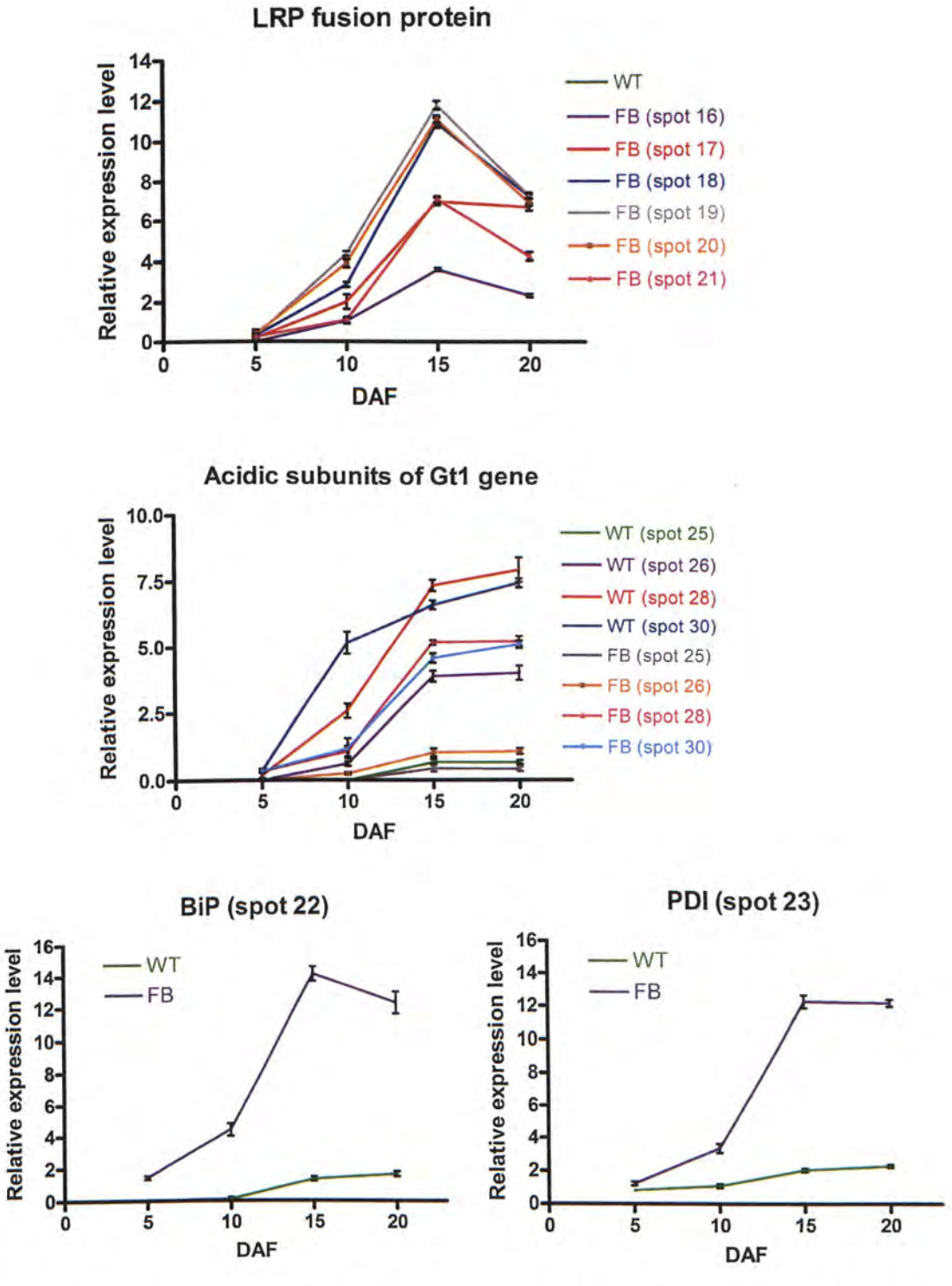


Figure 15. Protein profiles of four proteins, LRP fusion, glutelin, BiP and PDI, during FB rice seed development.

The four proteins exhibit similar expression patterns during seed development, including LRP fusion protein, glutelin acidic subunits encoded by Gt1 gene, BiP and PDI.

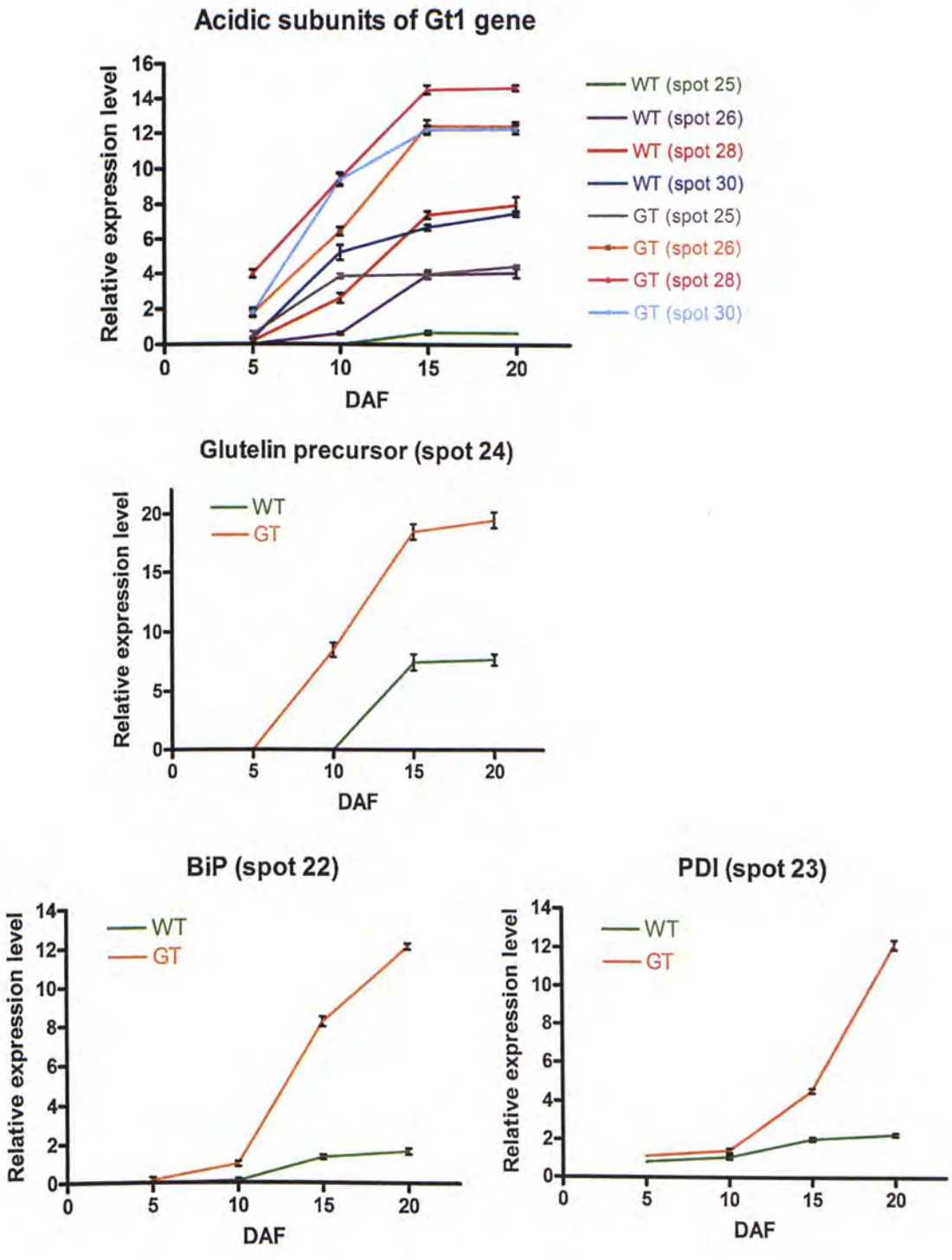


Figure 16. Protein profiles of glutelin proteins, BiP and PDI during GT rice seed development.

The four proteins exhibit similar expression patterns during seed development, including glutelin acidic subunits encoded by Gt1 gene, glutelin precursor, BiP and PDI.

4.2 Antibody production

4.2.1 The production of anti-prolamin and anti-glutelin antibodies

Anti-prolamin and anti-glutelin antibodies were produced for finding out the distribution of prolamin and glutelin in developing rice protein bodies. Total prolamin and glutein fractions were extracted from mature wild type rice seeds and were separated by SDS-PAGE. The bands of 13 kDa prolamin and 57 kDa glutelin precursor (Figure 17A) were excised from the surrounding gel and macerated. The purity of the gel-eluted prolamin and glutelin was confirmed by SDS-PAGE (Figure 17B) and used in rabbit and mice immunization.

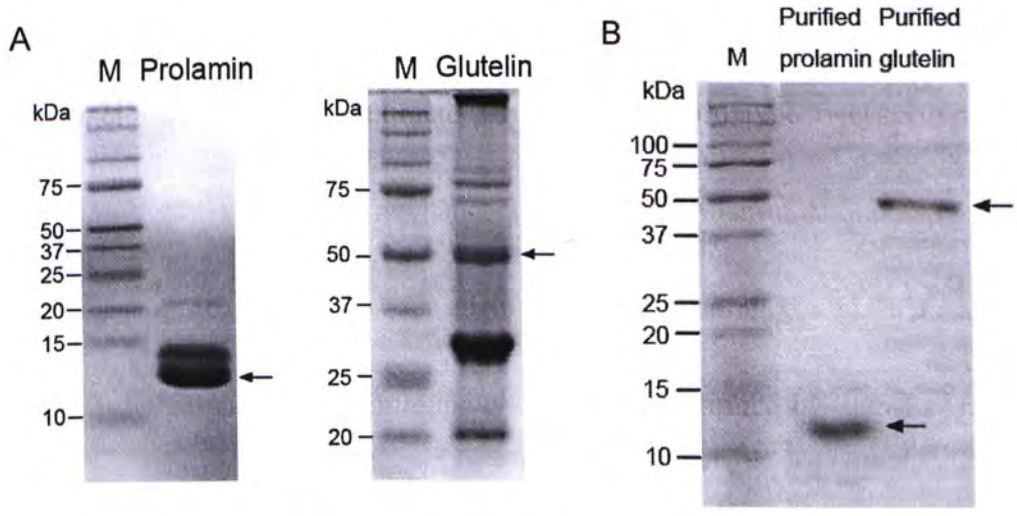


Figure 17. Gel purification of prolamin and glutelin precursor from rice seeds.

Panels (A) The proteins of prolamin fraction and glutelin fraction were resolved in 12% SDS-PAGE and stained in Coomassie brilliant blue; and (B) The purified prolamin and glutelin precursor were separated in SDS-PAGE and stained in Coomassie brilliant blue. The positions of 13 kDa prolamin and 57 kDa glutelin precursor are indicated by arrows. Key: M: Precision Plus Protein Dual Color Standards (Bio-Rad).

4.2.2 The specificity of anti-prolamin and anti-glutelin antibodies

One rabbit and five mice were immunized with prolamin and glutelin as the antigens, respectively. Western blotting using the anti-sera was carried out to test for their specificity. The antisera of anti-prolamin and anti-glutelin were diluted to 1:1000 and 1:5000, respectively.

The immunoblot results (Fig. 18) reveal that the rabbit antiserum against 13 kDa prolamin can react with both 10 kDa and 13 kDa-prolamin of the total proteins of mature rice seeds. For the anti-glutelin from mouse #1, it appears to reacted with the 57 kDa glutelin precursor, 37-38 kDa acidic subunits and 23-25 kDa basic subunits. The anti-glutelin antisera from mice #2 to #4 also gave the same reaction pattern as mouse #1.

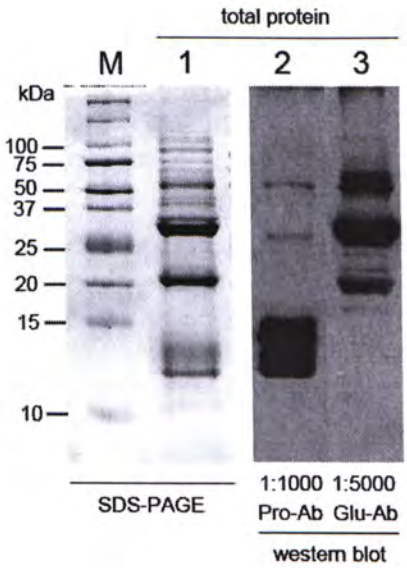


Figure 18. Specificity of anti-prolamin and anti-glutelin sera.
Total rice seed protein (20 μg) was resolved in 12% SDS-PAGE and blotted onto nitrocellulose membrane. Blots were detected using anti-serum of prolamin or glutelin as primary antibody and corresponding alkaline phosphatase conjugated anti-rabbit IgG or anti-mouse IgG as secondary antibody. Key: M: Precision Plus Protein Dual Color Standards (Bio-Rad).

4.3 Transmission electron microscopy observation of rice protein bodies

4.3.1 Morphology of protein bodies in high-lysine rice

Prolamin and glutelin are accumulated in rice as major storage proteins. They are stored in two separate compartments, PB-I (prolamin) and PB-II (glutelin), in the rice endosperm. Prolamin-containing PB-I is spherical, with a diameter of about 1-2 μm, and exhibits concentric rings of varying electron density. Glutelin is deposited in PB-II which is irregularly shaped, about 2-3 μm in diameter and of highly uniform electron density. Yang *et al.* (2003) demonstrated that high expression of human lysozyme gene fused with rice glutelin led to formation of morphologically altered protein bodies. In this project, electron microscopic observation of developing rice endosperm revealed that both transgenic lines exhibited morphologically-changed protein bodies (Fig. 19). In the wild type rice, two types of PBs, PB-I and PB-II, can be readily distinguished (Fig. 19A). Fig. 19B shows the types of PB observed in FB rice. Normal-appearing PB-II, but not PB-I can be seen. Instead, a new type of PB appears. This new PB has an irregularly shaped structure that with size varied from 0.5 μm to 2 μm in diameter. It has higher electron density at the outer part while lower at the inner part and the whole PB is full of cracks (Fig. 19C). Many ribosomes are observed on the outside of new type of PB. On the other hand, in GT rice, both PB-I and PB-II

change in their morphology (Fig. 17D). For the spherical PB-I, the lamellar structure surrounding the PB becomes uneven. Several small inclusion bodies (0.1 μm to 0.5 μm in diameter), which have the same electronic density as PB-II, are bound to the peripheral of PB-I. Besides, the irregularly shaped PB-II appears normal, but the peripheral also bound with some small inclusion bodies that have the same electron density as PB-I.

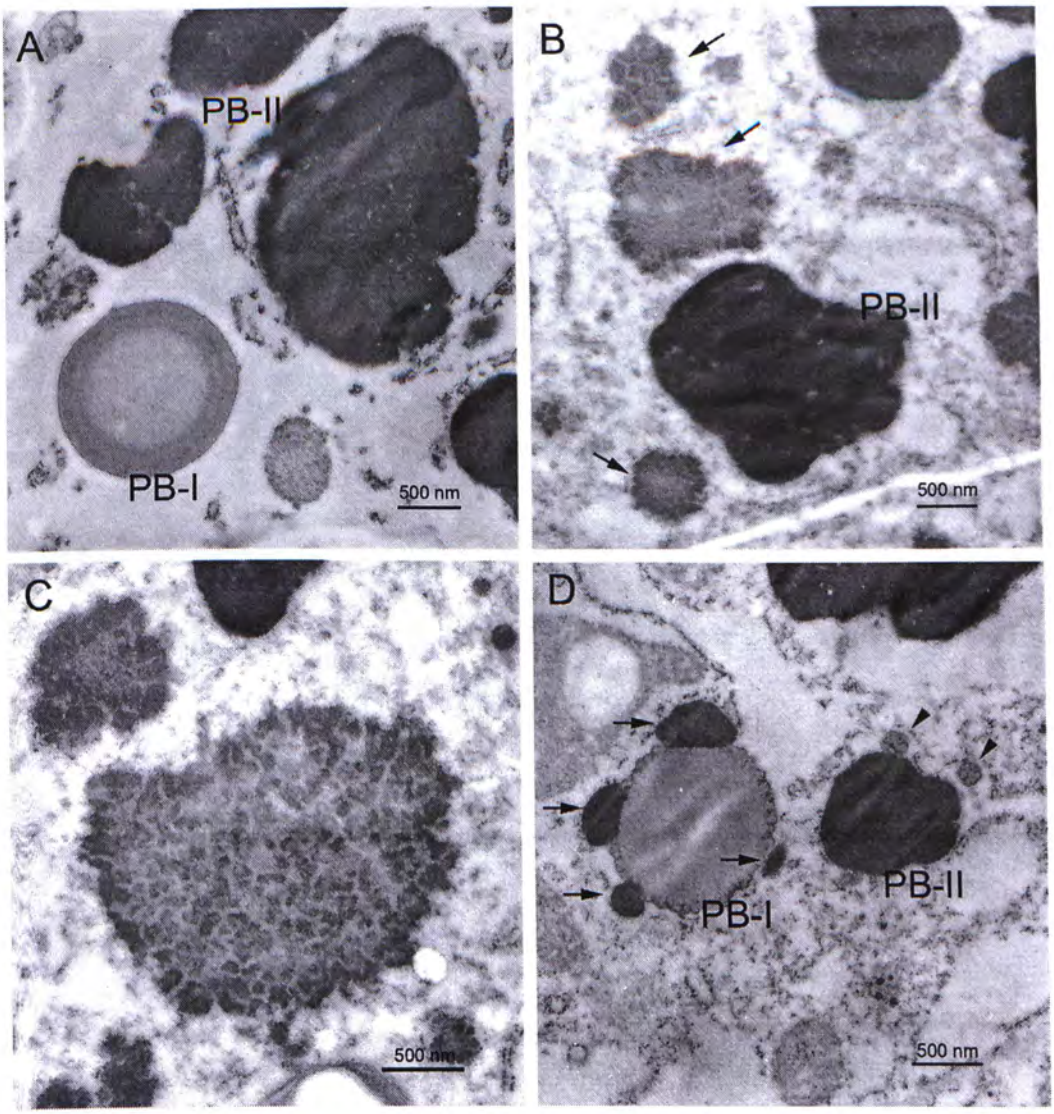


Figure 19. Electron microscopic observation of developing endosperm at 12 DAF from non-transgenic and high lysine rice plants.
Panels (A) Two types of normal protein body (PB-I and PB-II) in the endosperm cells of a wild-type plant; (B) A new type of protein body full of cracks (arrows) is found instead of PB-I in the endosperm cells of the FB rice while normal PB-II is observed; (C) The new type of protein body is full of cracks. It has higher electron density at the outer part while lower at the inner part; and (D) Portion of an endosperm cell of the GT rice, showing some small inclusion bodies with high electron density (arrows) are fused to PB-I while some small inclusion bodies with low electron density (arrowheads) are fused to PB-II. Bar = 500 nm

4.3.2 Subcellular localization of storage proteins and LRP

To obtain direct evidence for the localization of glutelin and prolamin polypeptides in the high lysine rice protein bodies, immunocytochemical studies were conducted using the antibodies for glutelin and prolamin. In all three types of rice, anti-glutelin antibody was co-incubated with anti-prolamin antibody in the immuno-reaction. Futhermore, LRP is a salt soluble protein which is localized in the cytosol of winged bean (Liu, CUHK). Where does LRP go when it is fused with glutelin and expressed in rice? In order to investigate the deposition of LRP fusion protein in FB transgenic seeds, Anti-LRP was also used. Results show that in wild-type rice, the prolamin is localized in circular shaped protein bodies (PB-I) while the glutelin in irregular shaped protein bodies (PB-II) (Fig. 20A). In FB rice, prolamin is localized in the new type of protein body (Fig. 20B) only. The signal for anti-glutelin antibody could be observed in PB-II and also in the new type of protein body. Figure 20C showed that LRP is distributed throughout all protein-body types. This means that the LRP fusion protein is targeted to both PB-II and the new PB. In GT rice, prolamin is deposited in PB-I and smaller inclusion bodies with low electron density. Glutelin is deposited in PB-II and in smaller inclusion bodies with high electron density (Fig. 18D).

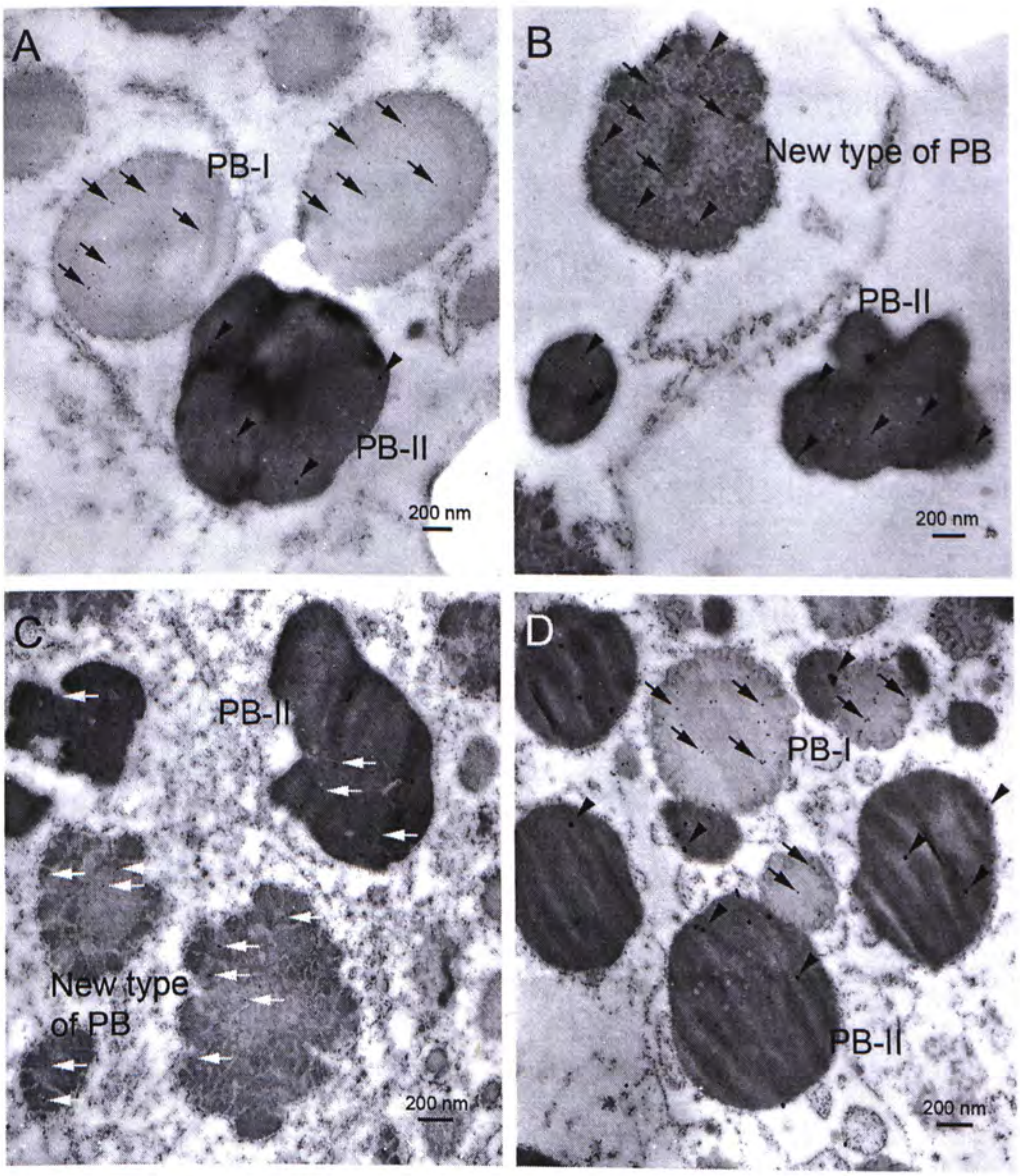


Figure 20. Immunoelectron microscopic observation of developing rice endosperm at 12 DAF from non-transgenic and high lysine rice plants.
Panels (A) In WT rice, prolamin is localized in PB-I (10 nm gold, arrows). Glutelin is localized in PB-II (25 nm gold, arrowheads); (B) In FB rice, prolamin was localized in the new type of PB (10 nm gold, arrows). Glutelin is localized in PB-II and the new type of PB (15 nm gold, arrowheads); (C) Both PB-II and the new type of PB are labeled by the anti-LRP antibodies (10 nm gold, white arrows) in FB rice; and (D) In GT rice, prolamin is deposited in PB-I and small inclusion bodies with low electron density (10 nm gold, arrows). Glutelin is deposited in PB-II and small inclusion bodies with high electron density (25 nm gold, arrowheads). Bar = 200 nm

Chapter 5. Discussion

5.1 Protein profiling of LRP fusion protein and its effects on the expression of other proteins

In the present study rice endosperm proteins of transgenic rice were profiled by 2D PAGE, and differentially expressed proteins were analyzed by tandem mass spectrometry. In FB line, LRP fusion protein is produced. In GT line, more glutelin precursor and acidic subunits are synthesized. Generally, after the insertion of LRP-fusion protein or extra copies of Gt1 gene, the expression of other storage proteins including albumin, globulin and prolamin are suppressed while that of BiP and PDI are enhanced.

In FB rice, the LRP fusion protein consists of multiple spots with slightly different isoelectric points (pI) on 2D gels, but with the same apparent molecular weight. The pattern of multiple identifications of a single protein occurred from a series of spots with the same molecular weight but slightly different pI could be the result of post-translational modification, by phosphorylation (Cash *et al.*, 1999) or deamidation (Sarioglu *et al.*, 2000). Another possible cause is that these patterns could be artefacts of the acetone precipitation process (Antelmann *et al.*, 2001). Further

detailed analyses are required to reveal the subtle changes in the charge of the polypeptide.

The deposit of LRP fusion protein was initiated at 5 DAF, peaked at 15 DAF, and declined at 20 DAF in FB rice. This protein profile is similar to that of the Gt1 gene (spots 25, 26, 28 and 30) (Fig. 15) which appears that the accumulation patterns of the introduced fusion protein and the native Gt1 protein under the same Gt1 regulatory sequences remain similar. From the expression profile, there is a slightly decrease in the accumulation of LRP fusion protein during 15 to 20 DAF. The possible reason is that the protein pattern of the LRP fusion protein on the 2D gels is affected by the accumulation of other proteins. Since equal protein was loaded for 2D PAGE, when the other alkaline-soluble proteins accumulate significantly at the latest stage, the proportion of LRP fusion protein in that fraction would decrease so that the amount of LRP fusion protein appears declining from 15 to 20 DAF.

The accumulation of BiP and PDI in rice endosperm also changes in the FB rice (Fig. 15). Their expression profiles are up-regulated when compared with the WT rice while their patterns are similar to that of the LRP fusion proteins. This suggests that the production of LRP fusion protein in rice endosperm may be correlated with the expression of chaperones BiP and PDI.

On the other hand, the accumulation of several storage proteins, including two

proteins which belong to globulins, three 14-16 kDa allergenic proteins and eight polypeptides of prolamin are down-regulated during grain development in FB rice.

5.2 Over-expression of glutelin and its effects on the expression of other proteins

In GT rice, Gt1 gene was over-expressed. The results of 2D PAGE of the alkaline-soluble fraction show that the accumulation of the four polypeptides (spot 25, 26, 28 and 30) of glutelin acidic subunits are significantly enhanced during grain development when compared with the WT rice (Fig. 16). At the same time, the six polypeptides of acid subunits are repressed in GT rice development. In addition, the accumulation of glutelin precursor is dramatically increased in GT rice over wild-type. In wild-type rice endosperm, glutelin precursor is proteolytically cleaved to form acidic and basic subunits, resulting in low amounts of the precursor form. In GT rice, as extra copies of Gt1 genes are inserted to the rice genome, the amount of glutelin precursor is perhaps too high to be fully processed into the mature forms so that high intensity of protein spot of glutelin precursor appeared in the 2D gels. Moreover, the amounts of BiP and PDI are also up-regulated when compared with WT rice and their accumulation patterns are the same with that of the polypeptides of Gt1 gene. This indicates that the production of extra glutelin protein in rice endosperms is correlated

with the accumulation of chaperones BiP and PDI.

Same as the FB rice, prolamin, globulin and albumin fractions are down-regulated in GT rice. The results of protein profiling of the high-lysine rice support the notion that expression of recombinant proteins especially storage proteins in abundant amounts will in turn suppress the expression of other storage proteins.

5.3 Formation of malformed protein bodies and deposition of storage proteins

The induction of recombinant proteins in the transgenic rice also accompanies changes in protein body morphology. In normal endosperm, prolamin and glutelin are synthesized on the ER, translocated into the ER lumen and then deposited in two morphologically distinct compartments, PB-I and PB-II, respectively. Prolamin in the ER lumen is packaged to form PB-I, whereas the glutelin precursor is transported first to the Golgi and then to a protein storage vacuole, where glutelin precursor is proteolytically cleaved to form acidic and basic subunits. The protein storage vacuole eventually forms PB-II.

From the EM study, in contrast to the wild-type protein bodies, the circular PB-Is were not observed in FB rice at all. Instead, a large number of novel PBs which were irregular in shaped, full of cracks and of varied size were observed in addition to

normal PB-II. Many ribosomes were observed on the outside of the new type of PB, suggesting that these PBs were derived from ER. By immunocytochemical study, anti-prolamin antibody labeled the new type of PB while anti-glutelin antibody and anti-LRP antibody were labeled both PB-II and the new type of PB in the transgenic rice seeds. On the basis of these, the new type of PB thus contains a mixture of proteins, including prolamin, LRP fusion protein and possibly the glutelin. Theoretically, since LRP-fusion protein gene includes the glutelin promoter, signal sequence and coding sequence, the protein should be localized to PB-II only. But the observation that the LRP-fusion protein was also located in other type of PB is contrary to our expectation. The results demonstrate that the expression of LRP-fusion protein affects the protein trafficking pathway.

In GT rice, both PB-I and PB-II could be observed but also with morphological changes. For the spherical PB-I, the lamellar structure surrounding the PB became uneven. Several small inclusion bodies, with high electronic density, bound to the peripheral was of PB-I. Besides, though the irregularly shaped PB-II appeared normal, its peripheral also bound with some small inclusion bodies low in electron density. Through immunocytochemical study, prolamin was found to deposit in PB-I as well as the small inclusion bodies with low electron density while glutelin in PB-II and the small inclusion bodies with high electron density.

5.4 Relationship between changes in protein expression and the Unfolded Protein Response

The present study shows that the morphology of protein bodies alters in both types of transgenic rice. The prolamin and glutelin precursor polypeptides within the ER lumen may not be separated normally in the high-lysine rice endosperm. One possible mechanism, the Unfolded Protein Response (UPR) may have accounted for the changes. UPR is a cellular response to the ER stress. Under stress conditions, proteins in the ER may misfold and accumulate into aggregates. To prevent the accumulation of unfolded proteins, the ER initiates the UPR system, which is a unique signal transduction pathway that activates the transcription of BiP chaperones. Increased amount of BiP is synthesized to bind to the misfolded proteins and assists proper folding. UPR also mediates the up-regulation of other genes that encode ER-resident proteins including other chaperones and disulphide-bond forming enzymes such as PDI. The subsequent reactions down-regulate the expression of storage proteins which can alleviate the ER stress (Bertolotti *et al.*, 2000; Schroder and Kaufman, 2005).

In previous studies, the existence of UPR was proven in *esp2* rice mutant. Large amount of glutelin precursor in *esp2* endosperm associated with the increased amounts of BiP and reduced levels of prolamin (Kumamaru *et al.*, 1988), symptoms of UPR in

esp2 endosperm. Such a condition has also been observed in the maize *floury-2* mutant in which the accumulation of abnormally processed zein proteins mediated the elevated levels of BiP and PDI (Boston *et al.*, 1991; Coleman *et al.*, 1995; Li and Larkins, 1996).

The results of the present study support that UPR occurred in the high-lysine rice endosperm. In FB rice, LRP-fusion protein is expressed while in GT rice, glutelin is over-expressed. In both cases, the increased amounts of protein entering ER lead to stress condition in ER which triggers the UPR. This then enhances the expression of protein folding enzymes and affects the genes in the secretary and ER-associated protein degradation pathways, leading to the down-regulation of the storage proteins transporting and processing through the secretory pathway. In this project, the accumulation of BiP and PDI was found greatly increased in the FB and GT rice. Furthermore, the reduction in storage proteins including prolamins, globulins and albumins accumulation was also observed. Both findings support the occurrence of UPR in the FB and GT transgenic rice.

The observed changes in protein body morphology in two types of high-lysine rice also imply that ER stress occurs when the ER is overloaded with recombinant protein. Yang *et al.* (2002) reported that the morphological changes in protein bodies in rice endosperm are due to the high lysozyme expression. The high-level expression

of recombinant protein distorts the trafficking and sorting of native storage proteins in rice endosperm and affects the normal protein-body formation.

The FB rice has a pronounced effect on PB-I formation and no apparent effect on PB-II. One explanation for this biased LRP-mediated effect on only PB-I is that the glutelin and prolamin have overlapping but distinct profiles of gene expression. Glutelin is expressed at an earlier stage of seed development, in contrast to the prolamin, which is preferentially expressed during latter stages (Yamagata *et al.*, 1986). Hence, substantial amounts of glutelin and LRP-fusion protein have already been synthesized and packaged into PB-II before the onset of prolamin accumulation. When the synthesis of prolamin starts, ER stress is caused as the ER is overloaded with glutelin, LRP-fusion protein as well as prolamin. Although the LRP fusion protein is driven by the Gt1 promoter/ GT signal peptide and the glutelin protein with its sorting signal, it was found to localize not only in PB-II but also in the new type of PB. It is possible that the LRP-fusion protein aggregates with prolamin in the ER, distorting PB-I formation and causing cracks to appear in the new type of PB in FB rice.

For GT rice, it seems that PB-I and PB-II bind together at their early developing stage. However, how do the novel types of protein bodies originate and form remains to be studied.

5.5 Effect of transgenes on rice grain quality

Rice grain quality is integrational characteristics including milling quality, appearance quality, cooking and eating quality, and nutritional quality. Each of them has its own parameters. According to previous studies, the T_3 transgenic rice lines of FB and GT that we used in the present study showed a high degree of chalkiness (Liu, CUHK). A similar case was observed in *floury-2* maize. The 24 kDa α-zein gene of *floury-2* maize contained a site mutation so it could not be properly processed to form a 22 kDa α-zein. Subsequently, the 24 kDa α-zein was anchored to the ER membrane and remained on the surface of the protein body, resulting in the formation of asymmetrical and malformed PB and starchy endosperm (Coleman *et al.*, 1995; Coleman *et al.*, 1997). As LRP/glutelin fusion construct (FB) was strongly expressed in the transgenic rice seeds, it is possible that the abundant LRP/glutelin fusion protein could not be processed and folded properly in ER lumen, disrupting the normal development of the protein bodies and leading to higher chalkiness endosperm. As chalkiness lowers the appearance, cooking and eating quality of rice, it is necessary to make further studies of the chalkiness of the FB endosperm. And understanding the mechanism causing increased chalkiness in the high-lysine rice may allow its reduction and removal. The quality protein maize (QPM) gives us a good example to restore a vitreous kernel for the high lysine opaque2 (*o2*) maize through breeding

(Prasanna *et al.*, 2001; Gibbon *et al.*, 2003). Currently our lab is crossing the FB rice with hybrid rice varieties, aiming at selection of high yielding and high-lysine hybrid rice lines, with low chalkiness. With further understanding of the molecular events involving in over expression of LRP-fusion protein and glutelin, such as UPR and their effect in protein trafficking and protein body formation, we may be able to tackle the problem of chalkiness with biotechnological approach as well.

5.6 Allergenic effects of transgenic rice

The first reported allergens in rice were 14-16 kDa proteins which were detected using sera from patients allergic to rice (Matsuda *et al.*, 1991). A 16 kDa protein was later recognized as a major rice allergen. This protein has significant amino acid homology to barley trypsin inhibitor and wheat alpha amylase inhibitor (Izumi *et al.*, 1992). Symptoms reported in rice–allergic individuals include abdominal cramping and similar pain, nausea, vomiting, rhinitis, rhinoconjunctivitis, asthma, contact urticaria, angioedema and anaphylaxis. There have been several attempts to produce hypoallergenic rice. Rice products of reduced allergenicity have been developed by specifically hydrolyzing or reducing allergenic proteins using protease, alkali, and ultra-high pressure treatment (Yamazaki & Sasagawa, 1997). Some rice products of reduced allergenicity were proven to be effective for individuals hypersensitive to rice

and with atopic dermatitis (Watanabe *et al.,* 1990). Furthermore, transgenic rice lines with reduced expression levels of the 14-16 kDa allergens are under development. From our proteomic results, the amount of allergen proteins at around 14 to 16 kDa were found dramatically decreased in the transgenic FB and GT rice when compared with that of wild type. This suggested that the high-lysine rice not only contains higher nutritional quality but can also reduce the allergenic effects on the patients.

5.7 Future perspectives

The effects of inserting LRP-fusion protein construct or extra copies of glutelin gene into rice genome were studied in the research. The next step is trying to alleviate the ER stress so that the protein sorting and trafficking process will not be affected or less affected by the expression of foreign proteins. Leborgne-Castel *et al.* (1999) over-expressed BiP in tobacco and alleviated ER stress caused by environmental factors. It is significant that the endosperm cell can adapt to aggregate and place the recombinant protein in storage form. Therefore, it will be of great interest to elucidate the pathway of recombinant protein trafficking in rice endosperm (wild type and transgenic), and to determine whether a membrane delimits the small protein bodies, and if such a membrane is derived from the ER, the Golgi apparatus or transport vesicles. We do not know if the ER stress in our case could be alleviated with

increased chaperone expression, but further studies to our understand how recombinant protein expression affects native protein trafficking during endosperm development should contribute to our effort of using genetic intervention for crop improvement.

Chapter 6. Conclusions

This study attempts to comprehensively evaluate the protein profiles of the transgenic rice plants expressing the LRP/Gt1 fusion genes and Gt1 genes. From the results of proteomic analysis, the two transgenic lines share some similar properties which are not present in wild type:

1. In salt-soluble fraction, the accumulation of two globulin proteins are down-regulated during grain development in FB as well as GT rice. Three allergenic proteins of albumin nature are also reduced significantly in accumulation in both FB and GT rice with reference to WT rice.
2. In alcohol-soluble fraction, the levels of eight polypeptides of prolamin are down-regulated in both transgenic lines when compared to WT line.
3. In alkaline-soluble fraction, the levels of chaperone proteins including BiP and PDI are increased in FB and GT lines with respect to WT line.

There are also some findings that are specific to only one transgenic line:

1. In FB rice, the LRP fusion protein are appears at 5 DAF, peaks at 15 DAF, and declines at 20 DAF, as revealed by 2D SDS-PAGE.
2. In GT rice, for the alkaline-soluble protein fraction, glutelin precursor and the four polypeptides of glutelin acidic subunits increase in protein level during grain

development, when compared with WT rice. At the same time, six polypeptides of acid subunits are also down regulated in GT rice during development.

TEM study showed that the FB rice contains normal-appearing glutelin-containing PB-II, but lacks the prolamin-containing PB-I. Instead, a new type of protein body irregular in shape and full of cracks can be observed. Both the LRP-fusion protein and prolamin are deposited in the novel protein body. In addition, changes in morphology of PB-I and PB-II occur in the GT rice. Small protein bodies can be seen to fuse with the surface of normal protein bodies. While PB-I fuses with smaller glutelin-containing PB, PB-II fuses with smaller prolamin-containing PB.

On the basis of the results, we have shown that the expression of recombinant proteins alters the native protein expression level. The LRP fusion protein or extra glutelin in ER could trigger the Unfolded Protein Response in transgenic rice endosperm. The elevated levels of chaperones, BiP and PDI demonstrate the presence of the UPR. More BiP and PDI are expressed to deal with the accumulated proteins in ER. The UPR also suppresses the expression of other storage proteins to alleviate the ER stress, which agrees with our results that the expression of other storage proteins including albumin, globulin and prolamin were down-regulated. Furthermore, the expression of recombinant proteins affects the deposition of storage proteins, changing the morphology of protein bodies in rice endosperms.

References

Adachi, T., Izumi, H., Yamada, T., Tanaka, K., Takeuchi, S., Nakamura, R. and Matsuda, T. (1993) Gene structure and expression of rice seed allergenic proteins belonging to the a-amylase/trypsin inhibitor family. *Plant Molecular Biology* 21: 239-248.

Altenbach, S.B., Pearson, K.W., Leung, F.W. and Sun, S.S.M. (1987) Cloning and sequence analysis of a cDNA encoding a Brazil nut protein exceptionally rich in methionine. *Plant Molecular Biology* 8: 239-250.

Altenbach, S.B., Pearson, K.W., Meeker, Staracy, L. and Sun, S.S.M. (1989) Enhancement of the methionine content of seed protein by the expression of a chimeric gene encoding a methionine-rich protein in transgenic plants. *Plant Molecular Biology* 13: 513-522.

Alvarez, A.M., Adachi, T., Nakase, M., Aoki, N., Nakamura, R. and Matsuda, T. (1995) Classification of rice allergenic protein cDNAs belonging to the alpha-amylase/trypsin inhibitor gene family. *Biochimica et biophysica acta* 1251: 201-204.

Argos, P., Narayana, S.V. and Nielsen, N.C. (1985) Structural similarity between legumin and vicilin storage proteins from legumes. *The EMBO journal* 4: 1111-1117.

Bagga, S., Adams, H., Kemp, J.D. and Sengupta-Gopalan, C. (1995) Accumulation of 15-Kilodalton Zein in Novel Protein Bodies in Transgenic Tobacco. *Plant Physiology* 107: 13-23.

Beach, L.R. and Ballo, B. (1992) Enhancing the nutritional value of seed crops. In *Biosynthesis and Molecular Regulation of Amino Acids in Plants*. Edited by Singh, B.K., Flore, H.E. and Shannon, J.C. pp. 229-236. American Society of Plant Physiology, Rockville, Maryland.

Bechtel, D.B. and Pomeranz, Y. (1978) Ultrastrure of the mature ungerminated rice (Oryza sativa) caryopsis. *American Journal of Botany* 65: 684-685.

Berggren, K.N., Schulenberg, B., Lopez, M.F., Steinberg, T.H., Bogdanova, A., Smejkal, G., Wang, A. and Patton, W.F. (2002) An improved formulation of SYPRO Ruby protein gel stain: comparison with the original formulation and with a ruthenium II tris (bathophenanthroline disulfonate) formulation. *Proteomics* 2: 486-498.

Bertolotti, A., Zhang, Y., Hendershot, L.M., Harding, H.P. and Ron, D. (2000) Dynamic interaction of BiP and ER stress transducers in the unfolded-protein response. *Nature Cell Biology* 2: 326-332.

Bjellqvist, B., Ek, K., Righetti, P.G., Gianazza, E., Gorg, A., Westermeier, R. and Postel, W. (1982) Isoelectric focusing in immobilized pH gradients: principle, methodology and some applications. *Journal of Biochemical and biophysical methods* 6: 317-339.

Boston, R.S., Fontes, E.B., Shank, B.B. and Wrobel, R.L. (1991) Increased expression of the maize immunoglobulin binding protein homolog b-70 in three zein regulatory mutants. *The Plant Cell* 3: 497-505.

Burkhardt, P.K., Beyer, P., Wunn, J., Kloti, A., Armstrong, G.A., Schledz, M., von Lintig, J. and Potrykus, I. (1997) Transgenic rice (Oryza sativa) endosperm expressing daffodil (Narcissus pseudonarcissus) phytoene synthase accumulates phytoene, a key intermediate of provitamin A biosynthesis. *Plant Journal* 11: 1071-1078.

Candiano, G., Bruschi, M., Musante, L., Santucci, L., Ghiggeri, G.M., Carnemolla, B., Orecchia, P., Zardi, L. and Righetti, P.G. (2004) Blue silver: a very sensitive colloidal Coomassie G-250 staining for proteome analysis. *Electrophoresis* 25: 1327-1333.

Cheng, S.M.K. (1999) Expression of LRP in Arabidopsis Plants. M. Phil. thesis. The Chinese University of Hong Kong, Hong Kong.

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

Coleman, C.E., Clore, A.M., Ranch, J.P., Higgins, R., Lopes, M.A. and Larkins, B.A. (1997) Expression of a mutant alpha-zein creates the floury2 phenotype in transgenic maize. *Proceedings of the National Academy of Sciences of the United States of America* 94: 7094-7097.

Coleman, C.E., Lopes, M.A., Gillikin, J.W., Boston, R.S. and Larkins, B.A. (1995) A defective signal peptide in the maize high-lysine mutant floury 2. *Proceedings of the National Academy of Sciences of the United States of America* 92: 6828-6831.

Dowsey, A.W., Dunn, M.J. and Yang, G.Z. (2003) The role of bioinformatics in two-dimensional gel electrophoresis. *Proteomics* 3: 1567-1596.

Duan, M.J. (2003) Profiling the expression of grain quality releated genes in developing hybrid rice seeds. M. Phil thesis. The Chinese University of Hong Kong., Hong Kong.

FAO (2003) Rice is life: Aspect of rice-based system. In *Concept paper of The International Year of Rice 2004* pp. 3-9, FAO, Rome.

Ferro, M., Salvi, D., Brugiere, S., Miras, S., Kowalski, S., Louwagie, M., Garin, J., Joyard, J. and Rolland, N. (2003) Proteomics of the chloroplast envelope membranes from Arabidopsis thaliana. *Molecular and Cell Proteomics* 2: 325-345.

Finnie, C., Melchior, S., Roepstorff, P. and Svensson, B. (2002) Proteome analysis of grain filling and seed maturation in barley. *Plant Physiology* 129: 1308-1319.

Flengsrud, R. and Kobro, G. (1989) A method for two-dimensional electrophoresis of proteins from green plant tissues. *Analytical Biochemistry* 177: 33-36.

Gallardo, K., Job, C., Groot, S.P., Puype, M., Demol, H., Vandekerckhove, J. and Job, D. (2001) Proteomic analysis of arabidopsis seed germination and priming. *Plant Physiology* 126: 835-848.

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

Gao, Y.F., Jing, Y.X., Shen, S.H., Tian, S.P., Kuang, T.Y. and Sun, S.S.M. (2001) Transfer of Lysine-rich protein gene into rice and production of fertile transgenic plants. *Acta Botamica Sinica* 43: 505-511.

Gibbon, B.C., X. Wang, B.A. Larkins. (2003) Altered starch structure is associated with endosperm modification in Quality Protein Maize. *Proceeding of the National Academy of Sciences* 100(26): 15329–15334.

Goff, S.A., Ricke, D., Lan, T.H., Presting, G., Wang, R., et al. (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). *Science* 296: 92-100.

Gorg, A., Postel, W., Domscheit, A. and Gunther, S. (1988) Two-dimensional electrophoresis with immobilized pH gradients of leaf proteins from barley (*Hordeum vulgare*): method, reproducibility and genetic aspects. *Electrophoresis* 9: 681-692.

Goto, F., Y.T., Shigemoto, N., Toki, S. and Takaiwa, F. (1999) Iron fortification of rice seed by the soybean ferritin gene. *Nature Biotechnology* 17: 282-286.

Habben, J.E. and Larkins, B.A. (1995) Genetic modification of seed proteins. *Current Opinion in Biotechnology* 6: 171-174.

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

Heazlewood, J.L., Howell, K.A., Whelan, J. and Millar, A.H. (2003) Towards an analysis of the rice mitochondrial proteome. *Plant Physiology* 132: 230-242.

Hibino, T., Kidzu, K., Masumura, T., Ohtsuki, K.,Tanaka, K., Kawabata, M. and Fujii., S. (1989) Amino acid composition of rice prolamin polypeptide. *Agricultural and biological chemistry* 53: 512-518.

Hoffman, L.M., Donaldson, D.D. and Herman, E.M. (1988) A modified storage protein is synthesized, processed, and degraded in the seeds of transgenic plants. *Plant Molecular Biology* 11: 717-729.

Hurkman, W.J. and Tanaka, C.K. (1988) Polypeptide changes induced by salt stress, water deficit, and osmotic stress in barley roots: a comparison using two-dimensional gel electrophoresis. *Electrophoresis* 9: 781-787.

Iida, S., Amano, E. and Nishio, T. (1993) A rice (Oryza sativa L.) mutant having a low content of glutelin and a high content of prolamine. *Theoretical and Applied Genetics* 87: 374-378.

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

Izumi, H., Adachi, T., Fujii, N., Matsuda, T., Nakamura, R., Tanaka, K., Urisu, A. and Kurosawa, Y. (1992) Nucleotide sequence of a cDNA clone encoding a major allergenic protein in rice seeds. Homology of the deduced amino acid sequence with members of alpha-amylase/trypsin inhibitor family. *FEBS Letters* 302: 213-216.

Juliano, B.O. (1972) The rice caryopsis and its composition. In *Rice: Chemistry and Technology.* Edited by Houston, D.F. pp. 16-74. American Association of Cereal Chemistry

Juliano, B.O. (1983) Lipids in rice and rice processing. In *Lipids in cereal technology.* Edited by Barnes, P.J. pp. 305-330. Academic Press, London.

Juliano, B.O. (1985) *Rice: Chemistry and Technology.* St. Paul, Minnesota.

Juliano, B.O. and Bechtel, D.B. (1985) The rice grain and its gross composition. In *Rice: Chemistry and Technology.* Edited by Juliano, B.O. pp. 17–57, St. Paul, Minnesota.

Kho, C.J. and de Lumen, B.O. (1988) Identification and isolation of methionine-cysteine rich proteins in soybean seed. *Plant Foods for Human Nutrition* 38: 287-296.

Kim, W.T. and Okita, T.W. (1988a) Nucleotide and primary sequence of a major rice prolamine. *FEBS letters* 231: 308-310.

Kim, W.T. and Okita, T.W. (1988b) Structure, Expression, and Heterogeneity of the Rice Seed Prolamines. *Plant Physiology* 88: 649-655.

Koller, A., Washburn, M.P., Lange, B.M., Andon, N.L., Deciu, C., et al. (2002) Proteomic survey of metabolic pathways in rice. *Proceedings of the National Academy of Sciences of the United States of America* 99: 11969-11974.

Komatsu, S., Kajiwara, H., and Hirano, H. (1993) A rice protein library: a data-file of rice proteins separated by two-dimensional electrophoresis. *Theoretical and Applied Genetics* 86: 953-942.

Komatsu, S., Kojima, K., Suzuki, K., Ozaki, K., and Higo, K. (2004) Rice Proteome Database based on two-dimensional polyacrylamide gel electrophoresis: its status in 2003. *Nucleic Acids Research* 32: 388-392.

Komatsu, S., Muhammad, A. and Rakwal, R. (1999) Separation and characterization of proteins from green and etiolated shoots of rice (Oryza sativa L.): towards a rice proteome. *Electrophoresis* 20: 630-636.

Krishnan, H.B., Franceschi, V.R. and Okita, T.W. (1986) Immunochemical studies on the role of the Golgi complex in protein body formation in rice seeds. *Planta* 169: 471-480.

Krishnan, H.B. and Okita, T.W. (1986) Structural relationship among the rice glutelin polypeptides. *Plant Physiology* 81:748-753.

Kristoffersen, H.E. and Flengsrud, R. (2000) Separation and characterization of basic barley seed proteins. *Electrophoresis* 21: 3693-3700.

Kumamaru, T., Satoh, H., Iwata, N., Omura, T., Ogawa, M. and Tanaka, K. (1988) Mutants for rice storage proteins 1: screening of mutants for rice storage proteins of protein bodies in the starchy endosperm. *Theoretical and Applied Genetics* 76: 11-16.

Leborgne-Castel, N., Jelitto-Van Dooren, E.P., Crofts, A.J. and Denecke, J. (1999) Overexpression of BiP in tobacco alleviates endoplasmic reticulum stress. *The Plant Cell* 11: 459-470.

Li, C.P. and Larkins, B.A. (1996) Expression of protein disulfide isomerase is elevated in the endosperm of the maize floury-2 mutant. *Plant Molecular Biolology* 30: 873-882.

Li, X., Franceschi, V.R. and Okita, T.W. (1993a) Segregation of storage protein mRNAs on the rough endoplasmic reticulum membranes of rice endosperm cells. *Cell* 72: 869-879.

Li, X. and Okita, T.W. (1993) Accumulation of prolamins and glutelins during rice development:a quantitative evaluation. *Plant Cell Physiology* 34: 385 - 390.

Li, X., Wu, Y., Zhang, D.Z., Gillikin, J.W., Boston, R.S., Franceschi, V.R. and Okita, T.W. (1993b) Rice prolamine protein body biogenesis: a BiP-mediated process. *Science* 262: 1054-1056.

Limas, G.G., Salinas, M., Moneo, I., Fischer, S., Wittman-Liebold, B. and Hendez, E. (1990) Purification and characterization of ten new rice NaCl-soluble proteins: identification of four protein-synthesis inhibitors and two immunoglubulin-binding proteins. *Planta* 181: 1-9.

Lin, S.K., Chang, M.C., Tsai, Y.G. and Lur, H.S. (2005) Proteomic analysis of the expression of proteins related to rice quality during caryopsis development and the effect of high temperature on expression. *Proteomics* 5: 2140-2156.

Liu, Q.Q. (2002) Genetically Engineering rice for increased lysine. Doctoral Dissertation. Yanzhou University, Yangzhou, China.

Lonosky, P.M., Zhang, X., Honavar, V.G., Dobbs, D.L., Fu, A. and Rodermel, S.R. (2004) A proteomic analysis of maize chloroplast biogenesis. *Plant Physiology* 134: 560-574.

Lucca, P., Hurrell, R. and Potrykus, I. (2001) Genetic engineering approached to improve the bioavailability and the level of iron in rice grains. *Theoretical and Applied Genetics* 102: 392-397.

Masumura, T., Hibino, T., Kidzu, K., Mitsukawa, N., Tanaka, K. and Fujii, S. (1990) Cloning and characterization of a cDNA encoding a rice 13 kDa prolamin. *Molecular & General Genetics* 221: 1-7.

Masumura, T., Shibata, D., Hibino, T., Kato, T., Kawabe, K., Takeba G., Tanaka, K. and Fujii, S. (1989) cDNA cloning of an mRNA encoding a sulfur-rich 10 kDa prolamin polypeptide in rice seeds. *Plant Molecular Biology* 12: 123-130.

Matsuda, T., Nomura, R., Sugiyama, M. and Nakamura, R. (1991) Immunochemical studies on rice allergenic proteins *Agricultural and Biological Chemistry* 55: 509-513.

Mattews, B.F. and Hughes, C.A. (1993) Nutritional improvement of the aspartate family of amino acid in edible crop plants. *Amino Acids* 4: 21-34.

McPhail, A.P. and Bothwell, T.H. (1989) Fortification of the diet as strategy for prevention of iron deficiency. *Acta Paediatr* 361: 114-124.

Merril, C.R., Goldman, D. and Van Keuren, M.L. (1984) Gel protein stains: silver stain. *Methods in Enzymology* 104: 441-447.

Mikami, S., Hori, H., and Mitsui, T. (2001) Separation of distinct components of rice Golgi complex by sucrose density gradient centrifugation. *Plant Science* 161: 665-675.

Millar, A.H., Sweetlove, L.J., Giege, P. and Leaver, C.J. (2001) Analysis of the Arabidopsis mitochondrial proteome. *Plant Physiology* 127: 1711-1727.

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

Müntz, K. (1998) Deposition of storage proteins. *Plant Molecular Biology* 38: 77-99.

Ndimba, B.K., Chivasa, S., Hamilton, J.M., Simon, W.J. and Slabas, A.R. (2003) Proteomic analysis of changes in the extracellular matrix of Arabidopsis cell suspension cultures induced by fungal elicitors. *Proteomics* 3: 1047-1059.

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

O'Farrell, P.H. (1975) High resolution two-dimensional electrophoresis of proteins. *The Journal of Biological Chemistry* 250: 4007-4021.

Oakley, B.R., Kirsch, D.R. and Morris, N.R. (1980) A simplified ultrasensitive silver stain for detecting proteins in polyacrylamide gels. *Analytical Biochemistry* 105: 361-363.

Ogawa, M., Kumamaru, T., Satoh, H., Iwata, N., Omura, T., Kasai, Z. and Tanaka, K. (1987) Purification of protein body-I of rice seed and its polypeptide composition. *Plant Cell Physiology* 28: 1517-1527.

Okita, T.W., Krishnan, H.B. and Kim, W.T. (1988) Immunological relationships among the major seed proteins of cereals. *Plant Science* 57:103-111.

Ogawa, M., Kumamaru, T., Satoh, H., Omura, T., Park, T., Shintaku, K. and Baba, K. (1989) Mutants for rice storage proteins: isolation and characterization of protein bodies from rice mutants. *Theoretical and Applied Genetics* 78: 305–310.

Ohtani, T., Galili, G., Wallace, J.C., Thompson, G.A. and Larkins, B.A. (1991) Normal and lysine-containing zeins are unstable in transgenic tobacco seeds. *Plant Molecular Biology* 16: 117-128.

Okita, T.W., Hwang, Y.S., Hnilo, J., Kim, W.T., Aryan, A.P., Larson, R. and Krishnan, H.B. (1989) Structure and expression of the rice glutelin multigene family. *The Journal of biological chemistry* 264: 12573-12581.

Osborn, T.B. (1924) *The Vegetable Proteins*. Longmans, London.

Pappin, D.J., Hojrup, P. and Bleasby, A.J. (1993) Rapid identification of proteins by peptide-mass fingerprinting. *Current Biology* 3: 327-332.

Patton, W.F. (2000) A thousand points of light: the application of fluorescence detection technologies to two-dimensional gel electrophoresis and proteomics. *Electrophoresis* 21: 1123-1144.

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

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

Prasanna, B.M., Vasal, S.K., Kassahun, B. and Singh, N.N. (2001) Quality protein maize. *Current Science* 81: 1308–1319.

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

Roth, H.P. and Mehlman, M.A. (1978) Symposium on role of dietary fiber in health. *American Journal of Clinical Nutrition* 31: S1-S291.

Saalbach, I., Pickardt, T., Machemehl, F., Saalbach, G., Schieder, O. and Muntz, K. (1994) A chimeric gene encoding the methionine-rich 2S albumin of the Brazil nut (Bertholletia excelsa H.B.K.) is stably expressed and inherited in transgenic grain legumes. *Molecular & General Genetics* 242: 226-236.

Schaeffer, G.W. and Sharpe, F.T. (1990) Modification of amino acid composition of endosperm proteins from in-vitro-selected high lysine mutants in rice. *Theoretical and Applied Genetics* 80: 841-846.

Schroder, M. and Kaufman, R.J. (2005) ER stress and the unfolded protein response. *Mutation Research* 569: 29-63.

Sha, S., Sugiyama, Y., Mitsukawa, N., Masumura, T. and Tanaka, K. (1996) Cloning and sequencing of a rice gene encoding the 13-kDa prolamin polypeptide. *Bioscience, Biotechnology, and Biochemistry* 60: 335-337.

Shaul, O. and Galili, G. (1992) Increased lysine synthesis in tobacco plants that express high levels of bacterial dihydrodipicolinate synthase in their chloroplasts. *Plant Journal* 2: 203-209.

Shen, S., Jing, Y. and Kuang, T. (2003) Proteomics approach to identify wound-response related proteins from rice leaf sheath. *Proteomics* 3: 527-535.

Shen, S., Matsubae, M., Takao, T., Tanaka, N. and Komatsu, S. (2002) A proteomic analysis of leaf sheaths from rice. *Journal of Biochemistry* 132: 613-620.

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

Shewry, P.R. and Casey, R. (1999) Seed proteins. In *Seed proteins*. Edited by Shewry, P.R. and Casey, R. pp. 1-10. Kluwer Academic Publishers, Dordrecht.

Shewry, P.R., Napier, J.A. and Tatham, A.S. (1995) Seed storage proteins: structures and biosynthesis. *The Plant Cell* 7: 945-956.

Shibasaki, M., Suzuki, S., Nemoto, H. and Kuroume, T. (1979) Allergenicity and lymphocyte-stimulating property of rice protein. *The Journal of Allergy and Clinical Immunology* 64: 259-265.

Shorrosh, B.S., Wen, L., Zen, K.C., Huang, J.K., Pan, J.S., Hermodson, M.A., Tanaka, K., Muthukrishnan, S. and Reeck, G.R. (1992) A novel cereal storage protein: molecular genetics of the 19 kDa globulin of rice. *Plant Molecular Biology* 18: 151-154.

Singh, N.K., Donovan, G.R., Carpenter, H.C., Skerritt, J.H. and Langridge, P. (1993) Isolation and characterization of wheat triticin cDNA revealing a unique lysine-rich repetitive domain. *Plant Molecular Biology* 22: 227-237.

Skylas, D.J., Mackintosh, J.A., Cordwell, S.J., Basseal, D.J., Walsh, B.J., Harry, J., Blumenthal, C., Copeland, L., Wrigley, C.W. and Rathmell, W. (2000) Proteome Approach to the Characterisation of Protein Composition in the Developing and Mature Wheat-grain Endosperm. *Journal of Cereal Science* 32: 169-188.

Speicher, D.W. (2004) Overview of proteome analysis. In *Proteome Analysis: Interpreting the Genome*. Edited by Speicher, D.W. pp. 1-18. Elsevier, Netherlands.

Sugimoto, T., K. Tanaka, and Z. Kasai. (1986) Improved extraction of rice prolamin. *Agricultural and Biological Chemistry* 50: 2409-2411.

Sun, S.S.M., Xing, L.W., Jing, X.Y. and Liu, B.L. (1998) Lysine rich protein from winged bean. U.S. patent application no.08/964,722.

Sun, S.S.M., Zuo, W.N. and Tu, H.M. (1992) Molecular approaches for enhancing amino acid quality of plant proteins. In *Biosynthesis and Molecular Regulation of Amino Acids in Plants*. Edited by Singh, B.K., Flore, H.E., and Shannon, J.C. pp. 208-126. American Society of Plant Physiology.

Sun, S.S.M. and Larkins, B.A. (1993) Transgenic plants for improving seed storage proteins. In *Transgenic Plants vol.1*. Edited by Kung, S.D., and Wu, R. pp. 339-371. Academic Press.

Takahashi, H., Hotta, Y., Hayashi, M., Kawai-Yamada, M., Komatsu, S. and Uchimiya H. (2005) High throughput metabolome and proteome analysis of transgenic rice plants (Oryza sativa L.). *Plant Biotechnology* 22: 47–50.

Takaiwa, F. and Oono, K. (1991) Genomic DNA sequences of two new genes for new storage protein glutelin in rice. *Idengaku zasshi* 66: 161-171.

Takaiwa, F., Oono, K., Wing, D. and Kato, A. (1991) Sequence of three members and expression of a new major subfamily of glutelin genes from rice. *Plant Molecular Biology* 17: 875-885.

Takemoto, Y., Coughlan, S.J., Okita, T.W., Satoh, H., Ogawa, M. and Kumamaru, T. (2002) The rice mutant esp2 greatly accumulates the glutelin precursor and deletes the protein disulfide isomerase. *Plant Physiology* 128: 1212-1222.

Tanaka, K., Sugimoto, T., Ogawa, M. and Kasai, Z. (1980) Isolation and characterization of two types of protein bodies in the rice endosperm. *Agricultural and Biological Chemistry* 44: 1633-1639.

Tanaka, K., Yoshida, T., Asada, K. and Kasai, Z. (1973) Subcellular particles isolated from aleurone layer of rice seeds. *Archives of Biochemistry and Biophysics* 155: 136-143.

Tanaka, Y., Hayashida, S. and Hongo, M. (1975) The relationship of the feces protein particles to rice protein bodies. *Agricultural and Biological Chemistry* 39: 515-518.

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

Tu, H.M., Godfrey, L.W. and Sun, S.S. (1998) Expression of the Brazil nut methionine-rich protein and mutants with increased methionine in transgenic potato. *Plant Molecular Biology* 37: 829-838.

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

Vensel, W.H., Tanaka, C.K., Cai, N., Wong, J.H., Buchanan, B.B. and Hurkman, W.J. (2005) Developmental changes in the metabolic protein profiles of wheat endosperm. *Proteomics* 5: 1594-1611.

Villareal, R.M. and Juliano, B.O. (1978) Properties of glutelin from mature and developing rice grain. *Photochemistry* 17: 177-182.

von Zychlinski, A., Kleffmann, T., Krishnamurthy, N., Sjolander, K., Baginsky, S., and Gruissem, W. (2005) Proteome analysis of the rice etioplast: metabolic and regulatory networks and novel protein functions. *Molecular and Cell Proteomics* 4: 1072-1984.

Wallace, J.C., Galili, G., Kawata, E.E., Cuellar, R.E., Shotwell, M.A. and Larkins, B.A. (1988) Aggregation of lysine-containing zeins into protein bodies in Xenopus oocytes. *Science* 240: 662-664.

Watanabe, M., Miyakawa, J., Ikezawa, Z., Suzuki, Y., Hirao, T., Yoshizawa, T. and Arai, S. (1990) Production of hypoallergenic rice by enzymatic decomposition of constituent proteins. *Journal of Food Science* 55 781-783.

Wen, T.N. and Luthe, D.S. (1985) Biochemical characterization of rice glutelin. *Plant Physiology* 78: 172-177.

WHO (1973) Energy and protein requirements. In *WHO Technical Report Series 522*. World Health Organization, Geneva.

WHO (2004) Vitamin. In *Vitamin and mineral requirements in human nutrition* pp. 17-37. World Health Organization and Food and Agricultural Organization of the United Nations, Geneva.

Wilson, K.A., McManus, M.T., Gordon, M.E. and Jordan, T.W. (2002) The proteomics of senescence in leaves of white clover, Trifolium repens (L.). *Proteomics* 2: 1114-1122.

Woo, S.H., Fukuda, M., Islam, N., Takaoka, M., Kawasaki, H. and Hirano, H. (2002) Efficient peptide mapping and its application to identify embryo proteins in rice proteome analysis. *Electrophoresis* 23: 647-654.

Yamagata, H., Sugimoto, T., Tanaka, K. and Kasai, Z. (1982) Biosynthesis of Storage Proteins in Developing Rice Seeds. *Plant Physiology* 70: 1094-1100.

Yamagata, H. and Tanaka, K. (1986) The site of synthesis and accumulation of storage proteins. *Plant Cell Physiology* 27: 135-145.

Yamazaki, A. and Sasagawa, A. (1997) Story of Development of High-Pressure Processed Foods Part 3: A method of development of low allergenic cooked rice and bread. The *Review of High Pressure Science Technology* 6: 236-241.

Yang, D., Guo, F., Liu, B., Huang, N. and Watkins, S.C. (2003) Expression and localization of human lysozyme in the endosperm of transgenic rice. *Planta* 216: 597-603.

Ye, X., Al-Babili, S., Kloti, A., Zhang, J., Lucca, P., Beyer, P. and Potrykus, I. (2000) Engineering the provitamin A (beta-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. *Science* 287: 303-305.

Yu, J., Hu, S., Wang, J., Wong, G.K., Li, S., et al. (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. indica). *Science* 296: 79-92.

Yuan, Q., Ouyang, S., Liu, J., Suh, B., Cheung, F., Sultana, R., Lee, D., Quackenbush, J. and Buell, C.R. (2003) The TIGR rice genome annotation resource: annotating the rice genome and creating resources for plant biologists. *Nucleic Acids Sesearch* 31: 229-233.

Zhang, W. and Chait, B.T. (2000) ProFound: an expert system for protein identification using mass spectrometric peptide mapping information. *Analytical Chemistry* 72: 2482-2489.

Zhao, C., Wang, J., Cao, M., Zhao, K., Shao, J., Lei, T., Yin, J., Hill, G.G., Xu, N. and Liu, S. (2005) Proteomic changes in rice leaves during development of field-grown rice plants. *Proteomics* 5: 961-972.

Zhao, W.M., Gatehouse, J.A. and Boulter, D. (1983) The purification and partial amino-acid sequence of a polypeptide from the glutelin fraction of rice grains; homology to pea legumin. *FEBS Letters* 162:96-102.

Zheng, Z., Sumi, K., Tanaka, K. and Murai, N. (1995) The bean seed storage protein β-phaseolin is synthesized, processed, and accumulated in the vacuolar type-II protein bodies of transgenic rice endosperm. *Plant Physiology* 109: 777-786.

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

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