Proteomic Study on the Starch Synthesis and Regulation in Developing Hybrid Rice Seeds

LONG Xiaohang

A Thesis Submitted in Partial Fulfillment

of the Requirement of the Degree of

Master of Philosophy

in

Biology

August 2006

Thesis/Assessment Committee

Professor H. M. LAM (Chair)

Professor S. S. M. SUN (Thesis Supervvisor)

Professor S. M. NGAI (Thesis Co-supervisor)

Professor L. W. JIANG (Committee Member)

Statement from Author

All experiments reported in this thesis were performed by the author unless specified otherwise in the text.

Signed Long Xiaohang

Long Xiaohang

Acknowledgements

First of all, I would like to express the deepest appreciation to my supervisor, Prof. Samuel S. M. Sun for offering me the most precious opportunity to do the research. During the past three years, he has kept giving me his great support and invaluable instruction on my research project. His magnificent personality, precise scientific thinking and dedicative spirit impress me deeply. It is a definite understatement for me to say that I am greatly grateful to have him as my supervisor.

Second, I would like to express my appreciation to my cosupervisor Prof. S. M. Ngai, and his patient guidance and valuable suggestions during my study and research period. I am thankful to my internal thesis committee Prof. H. M. Lam and Prof. L. W. Jiang and external committee Prof. N. S. Wong for their insightful advice and comments on my thesis.

My special gratitude goes to Marilyn Yu and Angela Yu, who have given me great assistance and support. Special thanks go to Ms. Helen Tsai for teaching me various skills in my projects, kindly explaining many questions and helping me do the MS.

Also, I would like to thanks Dr. J. M. Tu, Dr. Q. Q. Liu, Dr. B. X. Mo, who gave me great suggestions and valuable information about my research. Besides, I would like to give my special thanks to Dr. Abby Yu for her kind help, support, caring and encouragement. I would like to thank Dr. D. Y. Yuan and Miss M. J. Duan, who kindly provide me super hybrid rice seeds and gave me many invaluable suggestions of my research. I would like to thank Dr. M. Q. Zhong, who kindly taught me techniques and many suggestions of proteomics. Mr. Gundam Wong has kindly provided me the GBSS antibody, which has given me great help to make a further confirmation of this protein. Dr. Lucia Liu and Miss Jasmine Shum in AOE gave me great help during my study in AOE and Mrs. Yeung in G50 also gave me a lot of assistance in my data analysis.

Special thanks go to all labmates in G90, Rebecca, Jenifer, Eric, Ruby, Wendy, Juon Juon, Ivy, Gai, Zhang Junjun, William, Trista, Ronld, Yuk hua, Melineda, Ringo, Silva, Kit, KK,

Miss Yu, Pauline; and labmates in EG08, Mandy, Karen, Regina, Matthew, Christina for their friendliness and unselfish help during my study. I also thank all other fellow students in AoE lab and other labs. I am really grateful. I really enjoy the time I spent with them, which will be one of the most unforgettable recollections in my life.

Last but not the least, I would like to give heartfelt thanks to my parents, not only for their boundless love that gave me strength to overcome all the difficulties, but also for their understanding and support that enabled me to devote myself to my M.Phil study.

Abstract

Rice (*Oryza sativa L*) is the major dietary energy and nutrition source for about 50% of the world's population. Hybrid rice yields 25% higher than the best inbred lines and currently supplies half of the rice grains consumed in China. A super hybrid combination, LYP9, has been maintaining an average yield of 12 t/ha in large scale production for the last several years, however, its cooking and eating property, in general, requires further improvement.

Starch comprises about 90% of the total dry matter of polished rice, which in turn strongly influences the cooking and eating quality. In seeds, starch is synthesized in the amyloplasts as a long-term storage form of carbohydrates and is composed of amylose and amylopectin. The apparent amylose content (AAC) is recognized as one of the most important determinants of the eating and cooking quality of rice. According to China National Standard, amylose content of first grade *indica*-type rice should be 18-22%. However the amylose contents in hybrid rice parental varieties are generally higher or lower than this standard. Manipulating the amylose content to a suitable level will be a practicable way to improve the hybrid rice grain quality. Although researches on several key genes and enzymes involving in the starch synthesis and regulation have developed in recent years, it still remains a complicated pathway. Recent advancements in proteomics provide new opportunities to study this complicated pathway at the protein level.

In this study, we obtained important information on starch synthesis and regulation through the analysis of the proteome of amyloplasts during rice seed formation. We first developed a method specifically for the purification of rice amyloplasts. Then we used the seeds of hybrid line 9311 at 10DAF as a model material and characterized the expressed proteome of amyloplasts. As a result, we are the first group to analyze amyloplast proteins by 1D and 2D gel electrophoresis, and by PMF and MS/MS techniques. A total of 166 protein spots and bands were identified and their functions categorized. The most abundant category (48.8%) includes these proteins involved in metabolism. Through proteomic methods to profile the protein expression patterns, we further surveyed the proteome of super-hybrid rice, LYP9, and compared with those of its parental cultivars, 9311 and PA64S. The quantitative variations in proteomes of these three rice lines at 6, 10, 15 and 20DAF were

determined. New analysis methods were introduced to systematically analyze the proteomic data for first time. The 157 groups of amyloplast proteins during seed development were clustered by self-organized maps and compared by scatter plot. Through the systematical analysis, we found several trends of protein expression pattern in three rice lines and grouped them into three major expression patterns, increase, decrease and other pattern. The three expression patterns do not show great difference in the three rice lines. The decrease pattern is the most abundant pattern, mainly composed of metabolism proteins, indicating that the metabolism proteins are active at the early developing stages. We also found that proteins expression in hybrid seeds is quite similar to that of PA64S in several aspects, suggesting the maternal genes play a dominant role in the hybridization. Profiling the post translational modifications of expressed proteins during seed development was also carried out for the first time for the amyloplast proteins. We found that several starch synthesis proteins such as GBSS, SSS and SBE were glycosylated or phosphorylated and the glycosylation or phosphorylation levels of these proteins were examined. This study provides insight information towards our understanding of starch biosynthesis and regulation in rice, especially in hybrid rice seeds.

摘要

水稻是世界半數以上人口的主要食物與營養的來源。雜交水稻的產量比自交系水稻高出 25%，而在中國，雜交水稻提供了超過一半的稻米。儘管超級雜交水稻“兩優培九”是高產量水稻，平均畝產量 12 噸，但是它的蒸煮和食味品質并不理想，需要進一步提高。

澱粉占水稻乾重的 90%以上，對水稻的蒸煮和食味品質有非常大的影響。在種子中澱粉在澱粉質體中合成，由直鏈澱粉和支鏈澱粉兩種多糖組成，并以碳水化合物的形式長期儲存在澱粉質體中。水稻的蒸煮和食味品質主要決定因素之一就是直鏈澱粉的含量。根據中國國家標準，第一級的籼稻的直鏈澱粉的含量應該是 18-22%。然而超級雜交水稻的父母本的直鏈澱粉的含量或高或低於國家標準。因此調節直鏈澱粉含量到達一個合適的標準可以提高水稻的品質。儘管水稻澱粉合成和調控的一些相關基因和酶的研究已經取得了一些進展，澱粉合成和調控仍然是一個非常複雜的途徑。蛋白質組學的發展爲研究這一複雜的途徑在蛋白水平上提供了新的及有效的方法。本試驗通過對水稻種子澱粉質體的蛋白質組學的研究，從而獲得澱粉合成和調控過程的重要資訊，對改良稻米蒸煮和食味品質的目的有所貢獻。

在我們的研究中，我們首先選用雜交水稻 9311 開花後 10 天的種子作爲模型，主要針對這個時期的澱粉質體的蛋白質進行鑒定。我們改良了水稻澱粉質體的純化方法，第一個成功地運用單向和雙向電泳的方法分離水稻澱粉質體的蛋白，并用肽質量指紋圖譜和串聯質譜的方法鑒定出水稻澱粉質體的 166 個蛋白并對其功能進行了分類。48.8%的蛋白參與了澱粉合成的代謝過程。通過蛋白質組學的方法，我們進一步研究了超級雜交水稻“兩優培九”的澱粉質體的蛋白在水稻種子成熟過程中（開花後 6，10，15 和 20 天）的表達模式，并且和它的親本（9311 和 PA64S）進行了定量的比較。在分析過程中，我們首次引入了新的方法對蛋白質組的數據進行了系統的分析。運用了自我組織網絡映射的方法將 157 組澱粉質體的蛋白的表達模式進行歸類，并且應用分佈圖的方法對這 157 組蛋白的表達模式在這三個不同的水稻系

間進行了比較。通過這種系統的比較，我們發現了這三個水稻系澱粉質體的蛋白表達模式的一些規律。澱粉質體的蛋白表達模式主要有三種，降低型，升高型和其他型。這三種類型的含量在這三個水稻系中沒有明顯的差異。降低型主要由代謝相關的蛋白組成，這表明和代謝相關的蛋白主要在種子形成的早期表現活躍。我們同時還發現超級雜交水稻的蛋白表達模式和其母本 PA64S 更相似，這表明在雜交過程中，母性基因做出重要控制作用。我們第一次對澱粉質體的具有翻譯後修飾的蛋白的表達模式進行了研究。發現了一些和澱粉合成相關的蛋白，例如顆粒凝結型澱粉合成酶（GBSS），可溶性澱粉合成酶（SSS）和澱粉去分支酶（SBE）具有磷酸化或糖基化的翻譯後修飾，并對它們磷酸化和糖基化的程度進行了分析。通過我們的研究，尤其是對雜交水稻的澱粉合成和調控的機制提供了有效用的資訊。

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AA	amino acid
AAC	apparent amylose content
AC	amylose content
ACN	Acetonitrile
ADP-G	ADP-glucose
ADP-Glc	ADP-glucose
AGPase	ADP-glucose pyrophosphorylase
AGP-L	ADP-glucose pyrophosphorylase large subunit
AGP-S	ADP-glucose pyrophosphorylase small subunit
ATP	adenosine triphosphate
BiP	luminal binding protein
BLAST	basic local alignment search tool
bp	base pair
BSA	bovine serum albumin
CBB	coomassie brilliant blue
cDNA	complementary DNA
CHAPS	3-((3-cholamidopropyl)dimethylammonio)-1-propanesulfonate
CHCA	α-cyano-2-hydroxycinnamic acid
CID	collision-induced Dissociation
cm	Centimeter
CSIT	critical sterility inducing temperature
Cys	Cysteine
Cyt c	cytochrome c
2D	two-dimensional
2DE	two-dimensional electrophoresis
2DGE	two-dimensional electrophoresis
Da	Dalton
DAF	days after fertilization
DIGE	two dimensional differential gel electrophoresis
DNA	Deoxyribonucleic acid
dNTP	deoxyribonucleoside triphosphate
DTT	Dithiothreitol
EDTA	ethylenediaminetetra-acetic acid
ER	endoplasmic reticulum
ESI	electrospray ionisation
EST	expressed tag sequence
fmole	Femtomole
g	Gram
G-1-P	glucose-1-phosphate
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
GBSS	granule-bound starch synthase
GC	gel consistence
Glu	Glutamate
GPI	glycophosphatidlyinositol
GWD	glucan water dikinase
h	Hours
he	Hectare

HHRRC	Hunan Hybrid Rice Research Center
HNHRRDC	Hunan National Hybrid Rice Research and Developing Center
HPLC	high performance liquid chromatography
IEF	Isoelectric focusing
IAA	Iodoacetamide
ICAT	isotope-coded affinity tag
IgE	immunoglobulin E
IPG	immobilized pH gradient
JIPID	the Japan International Protein Information Database
kb	Kilobase
KD	Kilodalton
kDa	Kilodalton
kg	Kilogram
L	Liter
LC	liquid chromatography
LS	large subunit
LYP9	Liang you pei 9
m	Meter
m/z	mass/charge
M	Molar
MALDI	matrix assisted laser desorption ionization
MALDI-TOF	matrix assisted laser desorption ionization-time of flight
mg	milligram
min	Minute
MIPS	the Munich Information Center for Protein Sequences
ml	Milliliter
mm	Millimeter
mM	Millimole
$MgCl_2$	magnesium chloride
mRNA	Messenger RNA
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MSDB	Mass Spectrometry protein sequence DataBase
MudPIT	multi-dimensional protein identification technology
MW	molecular weight
NaCl	sodium chlolide
NaOH	sodium hydroxide
NADH	nicotinamide adenine dinucleotide
NADPH	nicotinamide adenine dinucleotide phosphate
NBRF	the National Biomedical Research Foundation
NCBI	the National Center for Biotechnology Information
ng	Nanogram
NH_4OAc	Ammonium acetate
OD	optical density
O-GlcNAc	β-*O*-linked *N*-acetylglucosamine
PA64S	Pei'ai 64S
PA64S/9311	Pei'ai 64S cross 9311
PAGE	polyacrylamide gel electrophoresis
PB I	protein body I
PB II	protein body II

PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDI	protein disulfide isomerase
3-PGA	3-phosphoglycerate
Pi	Phosphate
PIR	Protein Information Resource
PMF	peptide mass fingerprint
PMSF	phenylmethylsulfonyl fluoride
PPi	pyrophosphate
PSD	post-source decay
PTGMS	photo-and thermo-sensitive genic male sterility
PTM	post translational modification
RNA	ribonucleic acid
s	Second
SBE	starch branching enzyme
SDE	starch debranching enzyme
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
Ser	Serine
SP	signal peptide
SP	starch phosphorylase
SPS	sucrose-phosphate synthase
SSS	starch soluble synthase
TFA	trifluoroacetic acid
Thr	Threonine
Tris	tris(hydroxymethyl)amminomethane
TOF	time-of-flight
Tyr	Tyrosine
tRNA	transfer RNA
Tris-HCl	tris (hydroxymethyl) aminomethane hydrochloric acid
Trx	Thioredoxin
µg	Microgram
µl	Microliter
µM	Micromole
U	Unit
UPGMA	Unweighted Pair Group Method with Arithmetic Mean
UV	Ultraviolet
+ve	positive
v	Volume
V	Voltage
vac	Vacuum
Vh	voltage hour
w/v	weight by volume

Chapter 1 General Introduction and Literature Review

1.1 General introduction

Rice is the stable food for more than half of the world's population. While the available cropland is decreasing as a result of continuous industrialization and urbanization, the population of the world will increase from about 6.5 billion at the present to 8.5 billion in 2030, including 5 billion rice consumers. This severe situation greatly challenges rice production and supply. To meet the high demand of rice, agricultural economists estimated that the actual rice production has to be increased by 40% from the present 600 million tons to 850 million tons in 2030. For the increasing demand, rice has drawn crucial attention for its yield and quality improvements.

Hybrid rice supplies half of the rice grains consumed in China. The super hybrid rice yields 20%-30% higher than the best inbred lines (Yuan, 2004). And a super hybrid combination, LYP9 (Liang you pei 9), has been maintaining an average yield of 12 t/ha in large scale for the last several years. Due to its high yielding characteristic, super hybrid rice can be a very important solution to the potential food supply problem in China, even in the world. Because hybrid rice occupies 50% of the total rice cultivation area in China and its yield is 25% higher than regular rice in average, exploiting the potential of heterosis has become a key strategy for increasing productivity of crop plants (Xiao *et al.*, 1995). Although breeding of elite hybrid rice combinations has made great progress in China, the cooking and eating property of hybrid rice, however, requires further improvement.

Starch comprises about 90% of the total dry matter of polished rice, which in turn strongly influences the cooking and eating quality. The biosynthesis of starch is a tissue and subcellular compartment-specific process. In seeds, starch is synthesized in amyloplasts as a long-term storage form for carbohydrates and is composed of amylose and amylopectin. The apparent amylose content (AAC) is recognized as one of the most important determinants of the eating and cooking quality of rice. Amylose is synthesized by the

addition of one molecule of glucose at a time to a α-1,4-glucosyl chain, which is catalyzed by granule-bound starch synthase (GBSS). According to China National Standard, suitable amylose content level for *indica*-type should be 18-22%, whereas IRRI standard is 20-24%. But the amylose contents in super hybrid rice parental lines are generally either higher or lower than those standards, for example, parental lines PA64S is 5-8% higher than the standard level. Reducing or increasing the amylose content to a suitable level will be a practicable way to improve the hybrid rice grain quality. So understanding the formation and regulation of starch should benefit the improvement of grain quality.

In recent years, more and more scientists have been attempting to improve starch quality by means of plant genetic engineering (reviewed by Slattery *et al.*, 2000; Emes *et al.*, 2003; James *et al.*, 2003; Tetlow *et al.*, 2004a) and to further understand the starch biosynthesis pathway, the function and relationship between enzymes involved in this pathway, and to modify starch quantity and quality by regulation of enzymes expression (Shimada *et al.*, 1993; Terada *et al.*, 2000). But until now we only know several key enzymes involved in the starch synthesis, such as GBSS, SSS, SBE and APGase.

To fully illustrate the mechanism of starch biosynthesis, we need to understand the full complement and the functions of individual enzymes including the roles and interactions of multiple forms of these enzymes in starch synthesis. Proteomics can provide us with the unique opportunity to study the complicated metabolic pathway. Plant proteomic studies published to date have focused on global mapping of proteins such as maize leaves (Porubleva *et al.*, 2001), maize endosperm (Mechin *et al.*, 2004), wheat grain (Skylas *et al.*, 2005), rice leaf, root and seed (Koller *et al.*, 2002) or mapping the proteomes of various organs, tissue (Watson *et al.*, 2003), and cellular components (Peltier *et al.*, 2000; Andon *et al.*, 2002; Peltier *et al.*, 2002; Schubert *et al.*, 2002; Tanaka *et al.*, 2004; von Zychlinski *et al.*, 2005), or comparing protein differences between two or more samples (Vensel *et al.*, 2005). Considerable research effort has also been applied to the analysis of the rice proteome, and remarkable progress has been made in the systematic functional characterization of proteins expressed in the various tissues of rice (Komatsu *et al.*, 1993; Tsugita *et al.*, 1994; Hirano, 1997; Komatsu *et al.*, 1999; Imin *et al.*, 2001; Koller *et al.*, 2002; Shen *et al.*, 2002; Fukuda *et al.*, 2003). The analyses of cell organelle proteomes provide additional important information about protein localization and pathway compartmentalization. Most of the currently available plastid proteome information that

provides new insights into organelle specific metabolic functions has been reported from chloroplasts (Peltier *et al.*, 2000; Ferro *et al.*, 2002; Peltier *et al.*, 2002; Schubert *et al.*, 2002; Ferro *et al.*, 2003; Kleffmann *et al.*, 2004; Lonosky *et al.*, 2004) and rice etioplasts (von Zychlinski *et al.*, 2005).

Comparative proteomics is also a new insight of proteomics. Several studies have investigated temporal changes in plant proteomes involving up to four or more different time points (Wilson *et al.*, 2002; Gallardo *et al.*, 2003; Shen *et al.*, 2003; Watson *et al.*, 2003; Zhao *et al.*, 2005). 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). Another research (Tanaka *et al.*, 2005) using Unweighted Pair Group Method with Arithmetic Mean (UPGMA) clustering method to study the protein profiling in rice basal region and S-system to study the interactions of the clusters have also been reported.

Post translational modifications are another very important aspect in proteomics as they can potentially regulate metabolic pathway (Tetlow *et al.*, 2004b). However, proteomic study on rice amyloplast, especially in developing super hybrid rice lines in relation to starch synthesis is lagging.

In our study we systematically identified protein components in rice amyloplasts and try to elucidate their functions in the starch synthesis pathway. We focused our attention on the proteomes of three lines of hybrid rice (9311, PA64S and LYP9). LYP9 is a super hybrid rice which yields 20% to 30% more grains per hectare than other hybrid or high yielding non-hybrid higher-yield rice crops (Lu *et al.*, 2000). Its paternal cultivar, 9311, is an *indica* variety (*Oryza sativa L. subsp. indica*) and the major rice subspecies grown in China and many other Asian-Pacific regions. While the maternal cultivar, PA64S, has a major genetic background of *indica* and minor gene flows from *japonica* and *javanica*, two other major cultivated rice subspecies. To analyze the proteome of the hybrid (LYP9) which displays a significant yield advantage after hybridization, we profiled the proteome among hybrid and its parental lines. However, there is no existing systematic profiling information on the expressed proteins in the maturing rice seed organelle, especially in relation to rice nutritional quality and hybrid rice. Our aim is to profile the expression of proteins in relation to starch synthesis enzymes in amyloplasts of the parental and the hybrid

generations. The maturing seeds from 6, 10, 15 and 20 days after flowering (DAF) will be collected and studied. A complete functional understanding of the proteome requires, however, also the full characterization of the post translational modifications of proteins and the complex networks of protein-protein interactions. So in our study, the post translational modifications were carried out by the recent developed PTM staining techniques (Steinberg *et al.*, 2001; Steinberg *et al.*, 2003).

1.2 Literature review

1.2.1 Rice

1.2.1.1 Classification of rice

There are about 20 species of rice, of which *Oryza sativa* (the Asian cultivated species) and *Oryza galberrima* (the West African cultivated species) are the two cultivated species. According to morphological and ecological characteristics, *Oryza sativa* can be divided into two subspecies: *indica* and *japonica* (Kato, 1990). *Indica* has compact tillers and greater height with longer, slender grain and generally grows in the tropical and subtropical regions such as Southern China, India, Thailand and Vietnam. By contrast, *japonica* has cluster tillers, short, plump grain and is grown in temperate climates such as Northern China, Japan, Korea, Australia and California.

In the present study, PA64S, 9311 and their F1 hybrid (PA64S/9311) were used as experimental materials. PA64S, with *indica* and *javanica* genetic background (Yuan, 1997), is the maternal line with photo- or thermo-sensitive genic male sterility (PTGMS) for hybrid (PA64S/9311) production. The critical sterility inducing temperature (CSIT) of PA64S is 23.5℃, that is, PA64S will convert into male sterile if the temperature is above 23.5 ℃, at its sterility sensitive stage. Line 9311 is the paternal line for hybrid (PA64S/9311) production. It belongs to *indica* subspecies and is a major planting variety in South China. 9311 has high rice yield, good grain quality (first class standard) and multiple resistances to blast, blight and lodging (Dai *et al.*, 1997). PA64S/9311 is the super hybrid rice combination, commercially named Liang You Pei 9 (LYP9), which is the typical two-line intersubspecies hybrid with super high yield at 12 tons/ha, 15-20% higher than other hybrids. Furthermore, the rice grain quality of LYP9 reached second-class national standard and the morphology is excellent.

1.2.1.2 Rice grain quality

Quality of rice is not always easy to define as it depends on the consumer and the intended end use of the grain. All consumers want the best quality that they can afford. When countries reach self-sufficiency in rice production, the demand by the consumers for better quality rice will increase. Traditionally, plant breeders concentrated on breeding for high

yields and pest resistance. Recently the trend has changed to incorporate preferred quality characteristics that increase the total economic value of rice. Grain quality is not just dependent on the variety of rice, but also depends on crop production environment, harvesting, processing and milling systems and conditions.

Rice grain quality is assessed by its appearance quality, milling quality, cooking and eating quality and nutritional quality (Juliano, 1993). Cooking and eating quality is largely determined by the properties of the starch that makes up 90% of milled rice. Gelatinization temperature, amylose content and gel consistency are the important starch properties which influence cooking and eating quality.

On the whole, cooking and eating quality is largely affected by amylose content. The amylose and amylopectin ratio determined the starch property. Based on the amylose content, milled rice (white rice) can be classified into four classes: waxy, 1-2% amylose; low, 7-20%; intermediate, 20-25%; and high, >25% amylose (Juliano, 1979). Rice with high amylose content shows high volume expansion (not necessarily elongation) and high degree of flakiness. High amylose grains cook dry, are less tender, and become hard upon cooling. In contrast, low amylose rice cooks moist and sticky. Intermediate amylose rice is preferred in most of rice-growing areas of the world except where low amylose japonicas are grown. According to the rice amylose analysis performed by Hunan Hybrid Rice Research Center (HHRRC) in Changsha, Hunan, the amylose content of 9311 is 14.9%, which is the lowest among the three lines. The amylose content of F1 hybrid (21.4%) is 0.9% less than that of PA64S (22.3%) but 6.5% more than that of 9311 (14.9%).

Gelatinization temperature of starch which determines the time for cooking is a physical property of starch. Final gelatinization temperature ranges from 55 to 79 ℃ and can be classified as low (55 to 69 ℃), intermediate (70 to 74℃), and high (>74℃), where at least 90% of the starch granules swells irreversibly in hot water with loss of crystallinity and birefringence (Dela Cruz *et al.*, 2000). Environmental conditions such as temperature during grain development influence gelatinization temperature (Dela Cruz *et al.*, 1989). And gelatinization temperature is associated with amylose content (Jennings *et al.*, 1979).

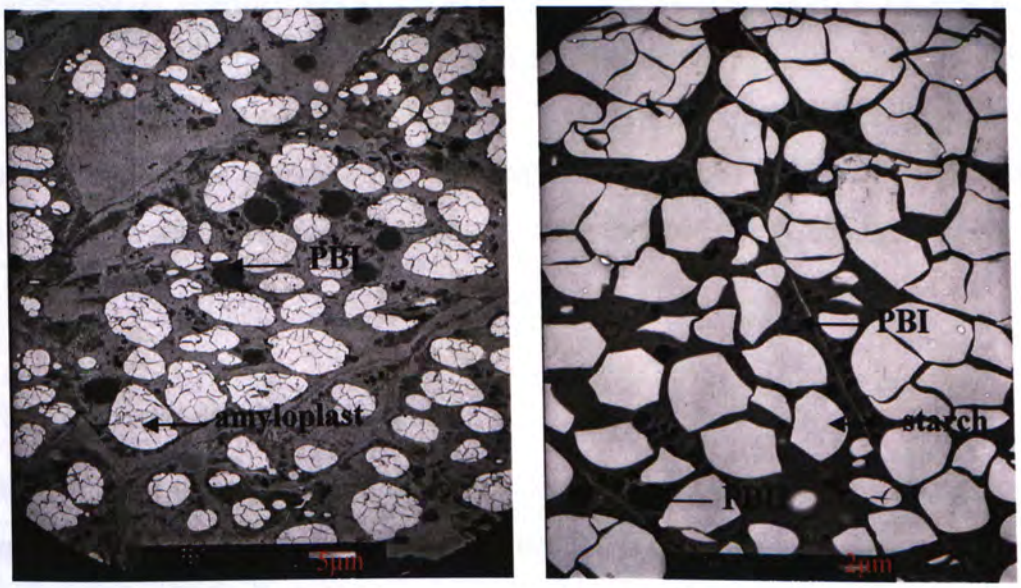
The gel consistency which measures the tendency of the cooked rice to harden on cooling may be divided into three categories: low (length of gel, more than 61 mm), medium (length of gel, 41 to 60), and high (length of gel, 40 or less).

1.2.2 Overview of current information on the starch biosynthesis and regulation during seed development

1.2.2.1 Starch property

1.2.2.1.1 Structure of rice starch granules

Starch is the major reserve in cereal crops. It is synthesized and accumulated in amyloplasts, where it exists as compound polyhedral granules 3-9 μm in size (Fig 1.1). In milled rice, starch makes up 90% of the dry weight. The cluster of starch granules within amyloplast is spherical, varying from 7-39 μm in diameter (Hayakawa *et al.*, 1980).



A: Magnified 1500× **B: Magnified 2500×**

Figure 1.1 Transmission Electron Microscope images of rice endosperm (Yuan, 2004)

1.2.2.1.2 Properties of rice starch

Starch is composed of two kinds of molecular components, amylose and amylopectin, which are high-molecular weight polysaccharides. Amylose is the minor fraction of starch granules with a linear molecular structure linked by α-1,4-glucosidic bonds. Amylopectin is a branched fraction with both α-1,4-glucosidic bonds and α-1,6-glucosidic linkage like

glycogen in animals and bacteria. Approximately every 20-25th glucose residue of the α-1, 4-glucosidic chains has a branch point which is linked by an α-1,6-glucosidic bond to an other α-1,4-glucose chain. The starch granule is a mixture of amylose and amylopectin, but how they associate together in a granule is still unclear, possibly through hydrogen bonding.

Amylose content is one of the most important factors affecting the cooking and processing behavior of rice. Amylose is synthesized by granule-bound starch synthase (GBSS) and the expression level seems to control the amylose content (Sano, 1984). High amylose rice shows high volume expansion, dryness and less tenderness; remains separate when cooked; and tends to retrograde rapidly upon cooling. Low amylose rice is moist and sticky when cooked and tends to retrograde slowly.

Starch granules also contain small amount of residual nitrogen, phosphorus and lipids. Most of the nitrogen is the *Wx* protein product (Sano, 1984) and perhaps starch synthase and its glucoprotein intermediates (Tandecarz *et al.*, 1975). Phosphorous appears mainly (93-98%) as 6-phosphoglucose phosphorous in the waxy starch granules but (75-89%) as phospholipid in the nonwaxy rice (Hizukuri *et al.*, 1983). Starch lipids probably are complexed with amylose and are either free fatty acids or monoacyl lipids. High amylose rice has slightly lower bound lipids than intermediate amylose rice (Choudhury *et al.*, 1980; Russell *et al.*, 1983).

1.2.2.2 Starch synthesis related proteins

Starch biosynthesis generates linear α-1, 4 glucosyl chains by adding an ADP-Glc to the nonreducing end of α-1,4 glucan primer. This reaction is carried out by the action of starch synthase (SS; EC 2.4.1.21), and the linear α-1,4 glucosyl chain products are used as the substrate for the next enzyme in the biosynthetic pathway, starch branching enzyme (SBE; EC 2.4.1.28). SBE introduces α-1,6 linkages between linear chains to form amylopectin. Further alterations in starch structure might occur because of the action of enzymes generally related to a degradative role in starch metabolism, the starch debranching enzymes (DBE; EC 2.4.1.41), phosphorylases, and glucanotransferases (D-enzyme).

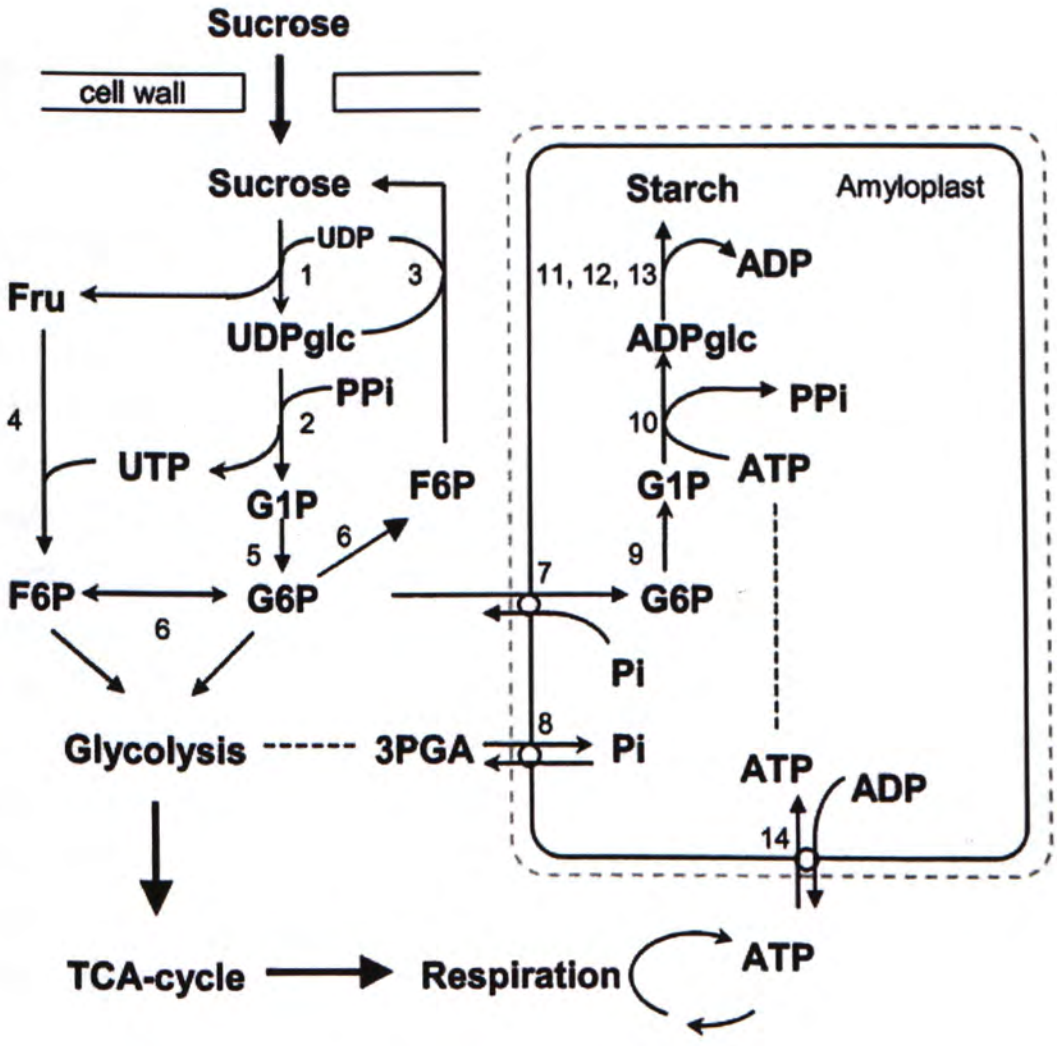


Figure 1.2 Pathway of sucrose to starch in potato tubers (Geigenberger, 2003)

(1) Sucrose synthase, (2) UDP-glucose pyrophosphorylase, (3) sucrose-phosphate synthase, (4) fructokinase, (5) cytosolic phosphoglucomutase, (6) phosphoglucoisomerase, (7) hexosephosphate translocater, (8) triose-phosphate translocator, (9) plastidial phosphoglucomutase, (10) ADP-glucose pyrophosphorylase, (11) soluble starch synthase, (12) granule-bound starch synthase, (13) branching enzyme, and (14) adenylate translocator.

1.2.2.2.1 The formation of ADP-glucose through AGPase

ADP-glucose pyrophosphorylase (AGPase, EC 2.7.7.27) catalyzes the first key regulatory step in starch biosynthesis (Fig 1.2). It forms the soluble precursor and substrate ADP-glucose for starch synthases. AGPase from higher plants is heterotetrameric enzyme composed of two large (AGP-L) subunits and two small (AGP-S) catalytic subunits encoded by at least two different genes (Preiss *et al.*, 1996). The large subunit plays more of a regulatory role while the small subunit plays more of a catalytic role. Based on the subcellular localization, AGPase can be divided into plastidial and cytosolic isoforms in several plants, such as maize (Denyer *et al.*, 1996), barley (Thorbjornsen *et al.*, 1996b), rice (Sikka *et al.*, 2001), and wheat (Tetlow *et al.*, 2003). In the developing endosperms of wheat, maize, barley, and rice, 65–95% of the total AGPase activity is the cytosolic isoform, implying that most of the storage starch biosynthesis in these tissues occurs through import of ADPGlc into amyloplasts. The differential expression of subunits in different tissues may produce AGPases with varying degrees of sensitivity to allosteric effectors, which are suited to the particular metabolic demands of a given plant tissue/organ. In cereal endosperms, the subcellular localization of AGPase isoforms is thought to be regulated by differential splicing of AGPase genes. Studies with barley (Thorbjornsen *et al.*, 1996a) indicate that the plastidial and cytosolic AGP-S subunit mRNAs are produced from a single gene by the use of two alternate first exons.

1.2.2.2.2 The synthesis of starch by starch synthases

The starch synthases catalyse the transfer of the glucosyl moiety of the soluble precursor ADP-glucose to the reducing end of a pre-existing α-1,4-linked glucan primer to synthesize the insoluble glucan polymers amylose and amylopectin. In higher plant, starch synthase can be separated into two groups, one is granule-bound starch synthase (GBSS) which is involved in amylose synthesis, and the other is the soluble starch synthase (SS) which is principally confined to amylopectin biosynthesis. Plants possess multiple isoforms of SSs, containing up to five isoforms that are categorized according to conserved sequence relationships. The isoforms within each of the major classes of SS genes are highly conserved in higher plants through the dicots and monocots (Ball *et al.*, 2003).

1.2.2.2.2.1 Amylose biosynthesis

Granule-bound starch synthase (GBSS), the major enzyme to elongate amylose, is tightly bound to the starch granules. It has two isoforms, GBSSI and GBSSII, which are encoded by two different genes. GBSSI is encoded by the *Waxy* locus (Nelson *et al.*, 1962; Shure *et al.*, 1983) in cereals and appears to be mostly confined to storage tissues. GBSSII is responsible for amylose synthesis in leaves and other non-storage tissues to accumulate transient starch (Fujita *et al.*, 1998; Nakamura *et al.*, 1998; Vrinten *et al.*, 2000). GBSSI consistes of a substrate-binding domain for ADP-glucose (Furukawa *et al.*, 1990), a typical transit peptide (Harn *et al.*, 1998) and putative motif KTGGL (Ainsworth *et al.*, 1993; Baba *et al.*, 1993; Cao *et al.*, 1999; Vrinten *et al.*, 2000). The mature GBSSI is a protein about 60 kDa.

The *waxy* mutants, which lack the GBSS, only have amylopectin, but still possess soluble starch synthases (Tsai, 1974), suggesting that different enzymes are involved in amylose and amylopectin synthesis. An interesting aspect of the control of polymer (amylose) elongation has been observed in the leaves of sweet potato (*Pomoea batatas*) where GBSSI transcript abundance and protein levels were shown to be under circadian control, in addition to being modulated by sucrose levels (Wang *et al.*, 2001).

1.2.2.2.2.2 Amylopectin biosynthesis

The second group of starch synthases, SS (SSI, SSII, SSIII, and SSIV) is exclusively involved in amylopectin biosynthesis, and their distribution within the plastid between the stroma and starch granules varies between species, tissue, and developmental stages (Harn *et al.*, 1998). The individual SS isoforms probably play unique roles in amylopectin biosynthesis. The study of SS mutants in a number of systems has been helpful in the assignment of *in vivo* functions/roles for the soluble and granule-associated SS isoforms in amylopectin synthesis. Recently, mutants of different isoforms of SSs have been identified. Mutations in isoforms from pea (*rug5*) (Craig *et al.*, 1998), maize (*dull1*) (Gao *et al.*, 1998) wheat (SGP-1) (Yamamori *et al.*, 2000), the green alga *Chlamydomonas reinhardtii* (*STA3*) (Fontaine *et al.*, 1993) all correspond with reduced starch synthase activity. Starch composition in all these mutants reveals a relative increase in amylose content, owing to a reduction in amylopectin synthesis. Overall, our present understanding indicates that loss of SSII or SSIIa results in reduced starch content, reduced amylopectin chain length

distribution, deformation of the starch granules, altered physicochemical properties of starch and perturbed crystallization.

1.2.2.2.3 Branching of the glucan chain by starch branching enzymes

Starch branching enzymes (SBE, EC 2.4.1.18) are also involved in amylopectin synthesis. They generates α-1,6-linkages by cleaving internal α-1,4 bonds and transferring the released reducing ends to C6 hydroxyls to form the branched structure of the amylopectin molecule. Two classes have been identified and classified on the basis of amino acid sequence homology (Burton *et al.*, 1995). The A class of SBEs consists of pea SBEI, maize SBEII and rice SBEIII, whereas maize SBEI, rice SBEI and pea SBEII contribute to the B class. SBEs differ in their enzymatic properties between classes. SBEII proteins transfer shorter chains and show a higher affinity towards amylopectin than their SBEI counterparts, which show higher rates of branching with amylose (Guan *et al.*, 1993; Takeda *et al.*, 1993). The construction of chimeric forms of maize SBEI and SBEII and analysis of their catalytic properties (Kuriki *et al.*, 1997) indicated that the N- and C-termini of these proteins play important roles in determining substrate preference, catalytic capacity, and chain length transfer. It has been shown from several studies that SBE isoforms are differentially and independently expressed during organ/tissue development and within the amyloplast (Mizuno *et al.*, 1992; Gao *et al.*, 1997; Sun *et al.*, 1998). SBEII isoforms are also partitioned between the plastid stroma and the starch granules. As with the granule-associated SS, the factors/mechanisms involved in partitioning the SBE proteins to the starch granules remain undetermined. The ability of proteins to become granule-associated may be a function of the relative affinities of their active sites for the glucan polymer. Recent study of starch branching enzyme (Tetlow *et al.*, 2004b) has revealed that SBEIIb may be activated by phosphorylation and form a protein complex with phosphorylated SBEI and starch phosphorylase (SP) to stimulate amylopectin biosynthesis. It has been speculated that the co-ordination of debranching, branching, and SS activities required for starch synthesis might be accomplished by physical association of the enzymes in a complex/complexes within the amyloplast (Ball *et al.*, 2003). The formation of complexes of starch metabolic enzymes via protein-protein interactions may directly alter the kinetic properties of individual components of the complex through conformational changes.

1.2.2.2.4 The role of debranching enzymes in polymer synthesis

Starch debranching enzymes (DBEs, EC 3.2.1.41, and EC 3.2.1.68) belong to the α-amylase super family (Jespersen *et al.*, 1993). They hydrolyze the α-1,6 glucan branches of amylopectin. Two groups of DBEs exist in plants; isoamylase-type and pullulanase-type which efficiently hydrolyse (debranch) α-1,6-linkages in amylopectin and pullulan (a fungal polymer of malto-triose residues). The difference is their substrate specificity in which pullulanases debranch pullulan and amylopectin but not glycogen, whereas isoamylase debranch both glycogen and amylopectin (Nakamura *et al.*, 1996). In cereals, isoamylase is a larger (400kDa) multimeric enzyme composed of one type of isoamylase subunit (Burton *et al.*, 1995). The precise roles for the isoamylase-type and pullulanase-type DBEs in starch biosynthesis are not yet known. Researcher (Ball *et al.*, 1996) has proposed a trimming model of SBE function. The starch granule is the result of 'trimming' by DBE of the highly branched glucans synthesized by SS and SBE.

1.2.2.2.5 Starch degradation in plastids

Starch degradation is part of the overall process of starch turnover which occurs in all starch-containing plastids to varying degrees. At the simplest level, the process of starch degradation requires an initial hydrolytic attack on the intact starch granule, followed by debranching (hydrolysis) of α-1,6-linkages to produce linear glucan chains, and finally, the degradation of the linear chains to glucosyl monomers. There is a range of plastidial enzymes with starch-degrading capabilities which may participate in the process of starch degradation and turnover, such as α-amylase, α-glucosidase, ß-amylase and starch phosphorylase (Zeeman *et al.*, 1998; Lao *et al.*, 1999).

1.2.2.2.6 Other enzymes involved in starch synthesis pathway

Starch phosphorylase (SP, EC 2.4.1.1) catalyses the reversible transfer of glucosyl units from glucose 1-phosphate (Glc1P) to the non-reducing end of α-1,4-linked glucan chains and may be driven in either a synthetic or a degradative direction by the relative concentrations of the soluble substrates. However, the role of SP in higher plant starch metabolism is unclear. Plastidial SP (referred to as Pho1 or the L-form) is characterized by higher affinity for amylopectin than glycogen. Studies with sweet potato roots have shown that the activity of the plastidial isozyme (L-form) of SP may be regulated by proteolysis of a 78-amino acid peptide (L78). Removal of L78 by an endogenous protease increased the

catalytic activity of SP in the phosphorolytic direction (Chen *et al.*, 2002a). It has been suggested that disproportionating enzymes may work in conjunction with SP, contributing to starch synthesis via the phosphorolytic SP reaction (Takaha *et al.*, 1998). Disproportionating enzyme (D-enzyme) disproportionate soluble oligosaccharides of at least three glucose residues or amylopectin into maltooligosaccharides (Peat *et al.*, 1953; Colleoni *et al.*, 1999). It has been shown that sucrose invertase, sucrose synthase and sucrose phosphate synthase which are crucial in sucrose methablosim and contribut to sink strength in divergent crop species may have the effect on the starch synthesis (Hajirezaei *et al.*, 2003; Roitsch *et al.*, 2003).

1.2.2.3 Starch biosynthesis regulation

Regulation of starch synthesis involves a flexible but complex network of highly interactive catabolic and anabolic reactions in the cytosol and plastids. Most interestingly, the plant cells must know when to use sucrose for starch synthesis and or for other metabolic processes. Also, it is well established that different genes and their isoforms diversely located in the plant cells are involved in starch metabolism. Thus, it is important that their activities are well coordinated for consistence with their roles. Overall, available evidence suggests that the entire regulatory network for starch synthesis relies on genetic, cellular metabolic status and environmental factors for its function. In the following sessions, major regulatory mechanisms of starch synthesis are reviewed.

1.2.2.3.1 Developmental regulation

The role of developmental regulation of starch synthesis genes has been established through spatial and temporal analysis studies in a number of crop species. It is governed by the interplay of genetic control, mitotic activity, histodifferentiation and cellular metabolic status (Borisjuk *et al.*, 2002). In barley, faba bean, wheat and maize, it has been shown that the pattern of storage starch accumulation correlates well with cell expansion but is spatially distinct from the pattern of mitotic activity. Detailed analysis has identified that metabolic status maintains a specific stage of differentiation and directs the next developmental program including starch biosynthesis (Wobus *et al.*, 1999). The expression profiles of the five major genes encoding the starch biosynthesis enzymes and their isoforms, a total of 15, during rice grain filling were also systematically carried out (Duan *et al.*, 2005). During grain filling, the two AGPase genes were active in the seeds at early

developmental stages (3DAF). The expression of the gene encoding GBSS I, the enzyme responsible for the synthesis of straight chain amylose in endosperm, was initiated at 3 DAF, and reached a peak at 6DAF. For the three major enzymes involving in the synthesis of amylopectin, namely SBE, SSS and SDE, the gene encoding the two members of SDE showed similar temporal expression patterns, with low level at 3DAF, peaking at 10DAF, and then declining to undetectable level at 20DAF. In other research, rice *sbeI* is immediately detectable post-pollination (Umemoto *et al.*, 1994), while *sbeIIb* in rice is detectable at 5 days after pollination (DAP) with maximal expression between 5 and 15 DAP (Mizuno *et al.*, 1992). Pea *sbeII* and *sbeI* genes are expressed early and late in embryo development, respectively (Smith, 1988). Similarly, in maize endosperm, it has been shown that *sbeI* is strongly expressed between 10 and 28 dap (Baba *et al.*, 1991; Gao *et al.*, 1996) whereas, *sbeIIb* is expressed throughout endosperm development (Gao *et al.*, 1996). In sorghum, the onset of *sbeIIb* and *sbeI* expression starts from 10 dap, attains maximum expression between 16 and 22 dap, and declines thereafter (Mutisya *et al.*, 2003). A similar pattern for *sbeI* in wheat has been reported (Morell *et al.*, 1997). In barley, *sbeIIa* and *sbeIIb* predominate during early stages of grain development, while *sbeI* is maximally expressed during late development (Sun *et al.*, 1998; Mutisya *et al.*, 2003). Taken together, these findings suggest that the expression profiles and expression levels of different isoforms are a well-coordinated process that involves the interactive effects of cellular mitotic changes, metabolic signals and genetic control, although the whole mechanism remains to be clearly understood.

1.2.2.3.2. Diurnal regulation

Plant metabolic networks involve different gene expression-mediated pathways (Dijkwel *et al.*, 1997; Dunlap, 1999; Halford *et al.*, 1999). There are indications that starch metabolism and some of the genes involved are entrained to the circadian rhythm of the plant. Light-responsive elements such as G, GT1, and GATA have been identified in promoters of light regulated starch synthesis genes and convincingly shown to be essential for light responsive transcription (Millar *et al.*, 1996; Chattopadhyay *et al.*, 1998). *Cis*-acting elements that bind specific transcription factors (Terzaghi *et al.*, 1997; Chattopadhyay *et al.*, 1998) have also been identified in the *gbssI* promoter of *Snapdragon* (Merida *et al.*, 1999) and *sbe*2-1 in *Arabidopsis*. Besides, diurnal oscillation of AGPase in *Chlamydomonas* (Zabawinski *et al.*, 2001) and *gbssI* in *Snapdragon* (Merida *et al.*, 1999)

suggest the involvement of circadian mode of regulation. At the level of starch, it has been shown that plants transferred from a typical diurnal regimen to continuous light perturb starch synthesis during subjective night (Li *et al.*, 1992). Smith *et al.*, (2004) analysed the diurnal changes in the transcriptome of genes encoding the starch metabolizing enzyme in the leaves of the dicot Arabidopsis to reveal both transcriptional and post-transcriptional regulation of starch metabolism. However, their analyses were confined to leaves. More recently, work involving micro-array analysis has revealed that several *Arabidopsis* genes encoding starch synthesis/degrading enzymes (e.g. chloroplastic isoforms of α-amylase, β-amylase, glucan water dikinase (GWD), SBEII-1) are under the influence of circadian clock. In all cases, however, entrainment to the circadian mode of regulation may serve to prime the events in starch metabolism along with other key factors such as day length, temperature and nutrient availability. Whatever the significance, little is known about the underlying molecular mechanisms. This requires further studies.

1.2.2.3.3 3-PGA/Pi regulation

AGPase catalyzes a rate-limiting step of starch biosynthesis in photosynthetic tissues and in many developing sink organs. The control of carbon flux through this enzymatic step is through allosteric regulation where the catalytic rate is activated by 3-phosphoglyceric acid (3-PGA) and inhibited by phosphate (Pi) (Ghosh *et al.*, 1966; Dickinson *et al.*, 1969; Sowokinos *et al.*, 1982; Kleczkowski *et al.*, 1993b; Kleczkowski *et al.*, 1993a; Sikka *et al.*, 2001). Hence, the levels of these effector molecules in the plastid have a significant influence on the level of AGPase activity and, in turn, carbon flux towards starch biosynthesis. Leaf, potato tuber, and tomato fruit AGPases are sensitive to 3-PGA and phosphate, while endosperm AGPases, except for that of rice (Sikka *et al.*, 2001), are less sensitive to the regulators. Enzymatic analysis of rice seed extracts indicated that the major AGPase activity showed significant activation by 3-PGA (>40-fold) and inhibition by Pi at levels comparable to the leaf enzyme (Sikka *et al.*, 2001). The control of the AGPase by 3-PGA/Pi is likely to be essential to maintain proper carbon partitioning between sucrose and starch. The introduction of a catalytically active, allosterically insensitive enzyme directed into the cytoplasm or plastid may increase starch synthesis and, in turn, increase seed weights and grain yields. Efforts to engineer rice plants for elevated starch synthesis by expressing allosteric variant AGPases in the cytoplasm or amyloplast during seed development are underway.

1.2.2.3.4. Sugar signaling

Sugar signaling is a well-established phenomenon in a number of plants. Sugar signaling involves three critical steps i.e. sugar/signal sensing, signal transduction, and action at target genes. Sugar sensing might arise from changes in apoplast sugar concentrations, sugar flux through the plasma membrane, fluctuations in intracellular sugar levels, and, from glycolytic, plastidial and or vacuolar sugar species interconversions (Loreti *et al.*, 2001). Whatever the mode of sensing, it is established that sugar signals mediate many metabolic processes in plants including germination, seedling growth, leaf and root development, starch synthesis and senescence, in concert with expression of the relevant genes (Sheen, 1996; Smeekens, 2000). It has been shown that starch synthesis genes (Sun *et al.*, 2003) are activated by sucrose and the expression of α-amylase (Loreti *et al.*, 2000) is repressed by sucrose. Additionally, gene expression of AGPase and starch contents in rice cultured cells are cooperatively regulated by sucrose and ABA (Akihiro *et al.*, 2005). In spite of this evidence, it remains unclear how the sugar effects are sensed and or transduced to the target gene(s), although plant hexokinase's (HXK's) have been identified and implicated in the process (Kaplan *et al.*, 1997; Sherson *et al.*, 1999; Guglielminetti *et al.*, 2000; Pego *et al.*, 2000; Rolland *et al.*, 2002). Plants use HXK as a glucose sensor to interrelate nutrient, light, and hormone signaling networks for controlling growth and development in response to the changing environment and it has been shown that plant HXK's associates with chloroplast, amyloplast, mitochondrial membranes, or with the Golgi apparatus (Rolland *et al.*, 2002) providing further evidence for a role in starch synthesis. Recently, the existence of a sucrose sensor has been reported (Loreti *et al.*, 2000). It is speculated that the sensor functions either as an active sucrose carrier or by a metabolic-derived signal. Most interestingly, sucrose transporters display interactive specificity with the glucosyl hydroxls 3,4 and 6 of sucrose unravelling the significance of side group orientation or linkage position (Chandran *et al.*, 2003), and its binding with sucrose is dependent on the presence of a fructosyl moiety for a hydrophobic surface. Further downstream, sugar signals and or sensors, initiate complex signaling networks to target genes interlinked by plant Ser/Thr protein related kinases, *SnRKs* (Rolland *et al.*, 2002). Specific *cis*-elements e.g. sugar responsive elements SURE (Grierson *et al.*, 1994; Sun *et al.*, 2003), SP8 (Ishiguro *et al.*, 1994), TGGACGG (Maeo *et al.*, 2001), G-box (Giuliano *et al.*, 1988) and B-box (Grierson *et al.*, 1994; Zourelidou *et al.*, 2002) respond

to arriving signals and cause gene action in concert with transcription factors such as SUSIBA2 (Sun *et al.*, 2003).

1.2.2.3.5. Hormonal signaling

The interactive effects of sugar and hormonal signaling have long been known to modulate plant metabolic processes, including but not limited to seed maturation, dormancy and changes in gene expression (Sheen, 1996; Leon *et al.*, 2003). It has been shown that abscisic acid (ABA) enhances sugar induction of *Apl3* and *sbe2.2* gene expression in Arabidopsis (Rook *et al.*, 2001). In addition, it has been shown that some Arabidopsis ABA-insensitive mutants (*sis5*, sugar insensitive; *sun6*, sucrose uncoupled; *gin6,* glucose insensitive = *abi4*) and ABA synthesis mutants (*sis4*, *gin1* = *aba2*) are equally defective in sugar sensing, and vice verse (Arenas-Huertero *et al.*, 2000; Huijser *et al.*, 2000; Laby *et al.*, 2000). Additionally, ABA-insensitive mutants (*abi1*, *abi2* and *abi3*), unaltered in sugar sensitivity exist, indicating that a specific ABA-related transduction pathway mediates sugar sensing. Nonetheless, it is established that many of the sugar-sensing mutants are altered in their ability to synthesis or transduce ABA signals.

1.2.2.3.6 Post translational modification regulation

1.2.2.3.6.1 Post translational redox modulation

Post-translational modification of AGPase involving thioredoxin was suggested by Fu *et al.* (1998) following the observation of partial inactivation of the recombinant potato enzyme by the formation of intramolecular disulphide bonds between the N-termini of the AGPase small subunits. Redox control of AGPase through sucrose supply in potato tubers has recently been proposed, whereby reductive activation causes channeling of carbon to starch and away from respiratory/glycolytic metabolism (Tiessen *et al.*, 2002). In potato, post-translational redox modulation of AGPase occurs in a time frame of 30-60 min, and activation was shown to occur in response to factors directly or indirectly related to increased sucrose availability (Tiessen *et al.*, 2002). Working with a recombinant potato tuber AGPase, Fu *et al.* (1998) determined the regulatory site to be a cysteine at position 82 (Cys^{82}) on the AGPase small subunit. It is thought that under oxidizing conditions the Cys^{82} residues from AGPase small subunit form intramolecular disulphide bonds, resulting in an inactive dimer. The Cys^{82} is highly conserved amongst other forms of AGPase small subunit, with the notable exception of the cytosolic isoform of AGPase small subunit from

monocots, implying this form of regulation is restricted to plastidial AGPases. However, preliminary studies indicate that cytosolic AGPases in wheat and barley can dimerize under oxidizing conditions in vitro, suggesting that different cysteine residues may be involved in this case. Recent work has demonstrated that this phenomenon is relatively widespread, and includes photosynthetic, as well as non-photosynthetic tissues from a number of species (Hendriks *et al.*, 2003). Starch synthesis in leaves from a broad range of plants is controlled by redox modulation of AGPase activity in response to light and sugar levels. At night, leaf AGPase was converted to an inactive dimer (Hendriks *et al.*, 2003). The signaling components leading to redox modulation of AGPase are beginning to be understood, and are thought to involve sucrose and glucose acting via a SNF1-related protein kinase and hexokinase, respectively (Tiessen *et al.*, 2003). There is some evidence to suggest that the activities of certain members of the α-amylase 'super-family' may be redox modulated. Redox regulation of pullulanase-type DBEs has been proposed in a number of tissues, including spinach leaves and the endosperms of barley and maize (Beatty *et al.*, 1999; Schindler *et al.*, 2001). In vitro studies with β-amylase also showed reversible inactivation via disulphide interchanges (Spradlin *et al.*, 1970), and the amylolytic activity in spinach chloroplasts doubled upon the addition of the sulphydryl reagent dithiothreitol (Pongratz *et al.*, 1978), suggesting redox modulation may play a role in the diurnal regulation of amylolytic activities in leaves.

1.2.2.3.6.2 Protein phosphorylation

Until recently, virtually nothing was known about the role that protein phosphorylation plays in the processes of starch synthesis and degradation. Earlier protein phosphorylation studies with heterotrophic plastids suggested a likely role for this regulatory mechanism in non-photosynthetic plastid metabolism (Macherel *et al.*, 1986; Vescovi *et al.*, 1993; Lukaszewski *et al.*, 2001), but few phosphoproteins have been identified from the limited number of such studies, and none connected with starch metabolism. A recent study with isolated amyloplasts from wheat endosperm identified a number of phosphoproteins, including some involved in starch metabolism, indicating that some aspects of starch (amylopectin) biosynthesis may be controlled by protein phosphorylation (Tetlow *et al.*, 2004b). In this study, all the stromal isoforms of SBE and SP were shown to be phosphorylated at one or more serine residues by plastidial protein kinase(s) in amyloplasts following short incubations of intact organelles with (c-32P)-ATP. In vitro

dephosphorylation of the phosphoproteins reduced the activity of SBEIIa and SBEIIb in amyloplasts, and SBEIIa in chloroplasts, whilst having no measurable effect on the activity of SBEI. Phosphoproteins were also detected in the starch granules from isolated wheat endosperm amyloplasts, and among those identified were the granule-associated isoforms of SBEII, SSIIa, and an unidentified form of SS.

1.2.2.4 Rice grain quality improvement by genetic engineering

With the rapid progress of molecular biology and biotechnology, several approaches have been used for crops improvements, such as gene addition, gene subtraction, and regulation of biosynthesis pathway. However, only few have been reported in rice quality improvement.

1.2.2.4.1 Cooking and eating quality improvement

1.2.2.4.1.1 Manipulation of starch content

Starch is the major component of many crops. Many genes regulated its biosynthesis. Because antisense inhibition of AGPase in transgenic potatoes abates starch formation in tubers (Muller-Rober *et al.*, 1992), it seems reasonable that increasing AGPase activity might correlate with increased starch production. This was investigated by transforming potato var. Russett Burbank with a mutant *E. coli* AGPase gene, *glg16*, which encodes for a catalytically active enzyme that is only partially dependent on the activator and is refractive to the inhibitor (Stark *et al.*, 1992). The data showed a 30% increase in starch content than the non-transformed tubers. No significant increase in the starch content was observed when normal AGPase in *E.coli* was transformed.

1.2.2.4.1.2 Manipulation of amylose/ amylopectin ratio

GBSS is the only enzyme required for amylose synthesis. Waxy rice was obtained by transformation with antisense GBSS gene (Shimada *et al.*, 1993; Terada *et al.*, 2000; Liu *et al.*, 2003). The waxy mutants, which lack the GBSS, only have amylopectin, but still possess the soluble starch synthases (Tsai, 1974). Van der Leij *et al.*, (1991) and Itoh *et al.*, (2003) transformed GBSS gene with its promoter into amylose-free starch mutant and demonstrated or found amylose synthesis in potato and rice. There are two functional types of alleles in rice *Wx* locus, Wx^a and Wx^b (Sano, 1984). Expression level of Wx^a is 10-fold

higher than that of *Wx*^{b} at the RNA and protein levels. High expression of *Wx*^{a} results in high amylose content. Further studies indicated that all *Wx*^{a} alleles have a normal sequence of GT at the 5' splice junction of the first intron, on the other hand, *Wx*^{b} alleles have TT at the 5' splice junction. Terada *et al.* (2000) disrupted the target gene of Waxy by homologous recombination, resulting in glutinous rice. The same strategy can be used to decrease the amylose content by *Wx*^{b} substitution of *Wx*^{a}.

Manipulation of the levels of the isoforms of SBEs is a means of altering starch structure. For example, although SBE A is present at low levels in potato tubers (with SBE B being the major isoform), this isoform has a major impact on starch structure by increasing the amount of amylose (Jobling *et al.*, 1999).

As DBEs are suspected to play a role in amylopectin synthesis (Kubo *et al.*, 1999), abolition or enhancement of debranching activities might influence the structure and properties of the starch granules. Altering DBE levels can be accomplished using antisense RNA technology, through transgenic expression of an isoform behind a strong promoter, or by other means. One could selectively decrease or increase levels of debranching activity by targeting the pullulanase or isoamylase types of DBEs. Similar approaches that have been used or suggested for other enzymes, such as mutagenesis, heterologous expression or even DNA shuffling, perhaps between pullulanase and isoamylase types, might even be viable options for manipulating DBEs as well.

1.2.2.4.2 Other targets for manipulating starch quality and quantity

Although the targets such as AGPase, SS, SBE and DBE for manipulating the starch biosynthetic pathway seem obvious, other players can also influence starch quantity and quality.

Sucrose-phosphate synthase (SPS) is a key regulatory enzyme in the pathway of sucrose synthesis. Ono *et al.*, (1999), using maize SPS, have shown a positive correlation of SPS activity with respect to starch content and a positive correlation with sucrose to starch ratio. Transgenic plants with decreased levels of SPS activity showed diminished remobilization of N from flag leaves and suppressed the development of reproductive sinks. Contrary, the

results from Worrell *et al.*, (1991) indicated that when maize SPS gene was expressed in tomato, the starch content in leaves was decreased while the sucrose level doubled.

The export of sucrose from sucrose-synthesising tissues (source) and its uptake into sucrose utilizing tissues (sink) is thought to require sucrose transport proteins in the cell membrane (Upadhyaya *et al.*, 2000). The pattern of expression of sucrose transporter gene (*osSUT*) indicated its possible roles in mobilization of stem reserves and uptake of sucrose by the developing seed during grain filling (Hirose *et al.*, 1997). The results of antisense *osSUT* transgenic rice demonstrated its expression is essential for grain filling which in turn affects starch content.

Another means to generate more starch could be through the overexpression of BT-1, which is an adenylate transporter (Shannon *et al.*, 1998). Maize *bt1* mutants have a severe reduction in starch content and accumulate large quantities of ADP-Glc. These data are intriguing and over expression of BT-1 might have interesting effects.

1.2.3 Proteomics

1.2.3.1 General introduction

Nucleic acid-based analysis of biological systems can begin to provide data on the nature of individual genes and on the coordinate regulation among many genes. Experimental evidence, however, clearly shows a disparity between the relative expression levels of mRNA and their corresponding proteins (Anderson *et al.*, 1997; Gygi *et al.*, 1999). Furthermore, it has recently been proven mathematically that expression information from mRNA and proteins are required to understand a gene network (Hatzimanikatis *et al.*, 1999). The need and desire to understand total protein expression is motivating the field of 'Proteomics'. 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).

Proteomics is based on two-dimension gel electrophoresis (2DE) separation technique, mass spectrometry (MS) protein identification technique and bioinformatics to interpret the data. The basic 2DE-based proteomics methodology includes several steps: 1) solubilization of proteins from the sample (e.g. tissue); 2) separation of the proteins by 2DE; 3) computer-assisted analysis of protein spot patterns; 4) determination of specific attributes of the proteins of interest by MS; and searching of databases with these attributes to identify the proteins.

There are three major types of proteome studies, including: (1) proteome global mapping; (2) quantitative comparisons of proteins levels in multiple samples (protein profiling); (3) analysis of protein-protein interaction. Proteomics is the next logical step after genome sequencing, but analysis of proteomes is much more complicated and challenging than sequencing genome. Fig 1.3 is the overview of proteomics.

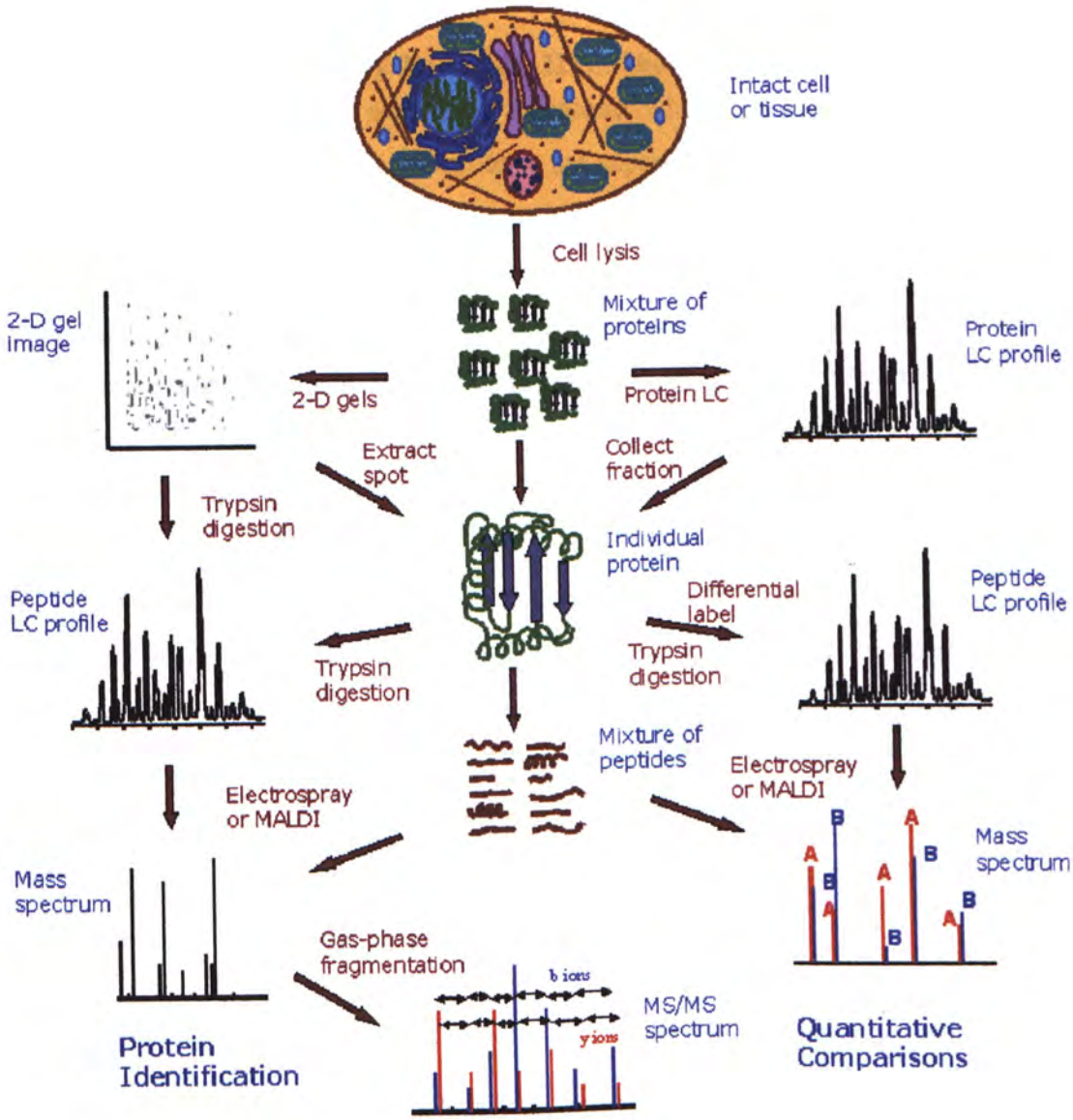


Figure 1.3 Overview of proteomics

(http://chemfacilities.chem.indiana.edu/facilities/proteomics/PRDFhomepage.htm)

1.2.3.2 Current technologies of proteomics

1.2.3.2.1 Protein separation by 2D or non-2D method

Currently, 2DE is a powerful separation technique, which allows simultaneous resolution of thousands of proteins. In 2D gel, proteins are separated by pI in the first dimension and with molecular weight in the second dimension (O'Farrell, 1975). Such an analysis begins with the solubilization of proteins from samples of interest using nonionic and zwitterionic detergents. Sample preparation is followed by isoelectric focusing (IEF) using carrier ampholytes or immobilized pH gradients to form the pH gradient. After an equilibration of the proteins in SDS and other reducing agents, the resulting focused bands are separated according to size using traditional polyacrylamide gel electrophoresis (PAGE).

The most important step in 2D gel electrophoresis (2DGE) is sample preparation. Sample preparation is to convert the native sample into a suitable physicochemical state for IEF while preserving the native charge and MW of the constitute proteins. So sample preparation should be as simple as possible to increase the reproducibility and protein modification must be minimized since they may result in artifactual spots on 2D gels. Prefractionation of complex mixtures significantly improves the 2DGE resolution, which will be discussed in the later session.

Although 2DGE is a widely used protein separation method, it has several weaknesses, including: 1) a limited capacity for the total amount of an extract that can be analyzed directly, which limits detection to the most low abundant proteins in the samples; 2) very large, very small, and membrane proteins are poorly recovered; 3) reproducibility is also a major problem for quantitative comparison; and 4) can not detect all the post translational modifications.

Due to the limitation of 2DGE, a number of different non 2D gel separation methods have been developed such as using two or more chromatographic separation modes which have been named by Yates and co-workers multi-dimensional protein identification technology (MudPIT). Using a high sensitivity nanocapillary ion exchange/reverse phase separation of tryptic peptides prior to MS/MS analysis, Yates demonstrated identification of ~1500 proteins from a yeast cell extract (Washburn *et al.*, 2001). Yates and co-workers have also systematical identified rice seed, leaf and root proteins by MudPIT, including 2528 unique

proteins (6296 peptides) (Koller *et al.*, 2002). The MudPIT method can also be used for qualitative analysis of relatively simple mixtures such as macromolecular complexes or complete proteomes. Quantitative profile comparisons by MudPIT is by introduction of stable isotope labeling into one of two samples to be compared using either metabolic labeling or chemical modification (Washburn *et al.*, 2002; Kubota *et al.*, 2003; Peng *et al.*, 2003). Despite some limitations, 2DGE technology remains as the gold standard in the proteome analysis.

1.2.3.2.2 Protein visualization

After 2DE, the separated proteins have to be visualized. The most important properties of protein visualization methods are high sensitivity (low detection limit), high linear dynamic range (for quantitative accuracy), reproducibility, and compatibility with protein identification procedures, such as MS. Proteins staining methods include anionic dyes (e.g. Coomassie blue), negative staining with metal cations (e.g. zinc imidazole), silver staining, fluorescence staining or labelling, and radioactive isotopes, using autoradiography, fluorography, or Phosphor-imaging.

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. However, in terms of the requirements for proteome analysis, the principal limitation of CBB stains lies in their insufficient sensitivity, which does not permit the detection of low abundance proteins-the detection limit of CBB stains is in the range of 0.2-0.5μg protein per spot. Hence, typically no more than a few hundred protein spots can be visualized on a 2D gel, 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 2DE for proteomics.

Silver staining methods (Oakley *et al.*, 1980; Merril *et al.*, 1981) are up to several times more sensitive than CBB. The detection limit is as low as 0.1 ng protein or spot. They provide a linear response with over a 10- to 40-fold range in protein concentration, which is slightly worse than with CBB staining. However, silver staining methods are far from

stoichiometric, and are much less reproducible than CBB stains due to the subjective end-point 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 increasingly gained popularity for proteome analysis. Two major approaches for the fluorescent detection of proteins on 2DE gels are currently practiced. These are: (i) covalent derivatization of proteins with fluorophores prior to IEF, such as the cyanine-based dyes (Unlu *et al.*, 1997) that react with cysteinyl residues and lysyl residues, which is commercially available as CyDyes (Amersham Biosciences), and (ii) post-electrophoretic protein staining by intercalation of fluorophores into the SDS micelles coating 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 use. Furthermore, most fluorescent staining procedures are compatible with subsequent protein identification methods such as MS.

1.2.3.2.3 Computer-assisted 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 computerized image analysis systems (Dowsey *et al.*, 2003). The traditional workflow for a 2DE software package is (i) preprocessing of the gel images, *i.e.* image normalization, cropping and background subtraction; (ii) spot segmentation, detection and expression quantification; (iii) landmarking, *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; (iv) matching, *i.e.* automatic pairing of the rest of the spots; (v)

identification of differentially expressed spots; (vi) data presentation and interpretation; and (vii) creation of 2D gel databases (Dowsey *et al.*, 2003).

To analyze and document the separated and identified proteins, image analysis is essential. A number of software packages is available including the most widely used, such as Image Master 2D Elite (Amersham Pharmacia Biotech, Uppsala, Sweden), PDQuest (Bio-Rad Laboratories, Hercules, CA, USA), Melanie 3 (GeneBio, Geneva, Switzerland; http://www.genebio.com/Melanie.html/), and MasterScan (Scanalytics, Billerica, MA, USA).

1.2.3.2.4 Protein identification by mass spectrometry

Two MS approaches are commonly used to identify proteins by in gel digestion-peptide mass fingerprinting (PMF) or tandem mass spectrometry (MS/MS). Peptide mass fingerprints are the fastest method for identifying proteins recovered from 2DE. The basic principle of this technique is the comparison of the measured peptide masses with 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 required for reliable protein identification. An alternative method for identifying proteins and peptides is the fragmentation of isolated peptide ions by post source decay (PSD) (Spengler *et al.*, 1992) or collision induced dissociation (CID) (Shevchenko *et al.*, 1996) after Matrix-assisted Laser Desorption Ionization (MALDI) or Electrospray Ionization (ESI). Both techniques lead to quite statistical fragmentations at the amide bonds of a peptide mainly depending on the ionization and fragmentation methods and the peptide sequence. The generated ions are called b-ions when the N-terminus is included and y-ions when the C-terminus is included in the fragment ions (Biemann, 1992). The mass difference of two consecutive b- or y-ions corresponds to an amino acid mass and, therefore, a PSD or CID fragment ion spectrum contains the sequence information of a peptide, which can be called de novo sequencing. Peptide fragmentation can be generated by using triple quadrupole, q-TOF, TOF-TOF, or ion trap instruments, with MALDI, capillary LC/MS, or nanospray interfaces.

1.2.3.2.5 Database search

1.2.3.2.5.1 Database searching software

For protein identification, it needs to search a protein sequence database 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, and the other is using tandem mass spectrometry (MS/MS) data from one or more peptides. The experimental data are compared with calculated peptide mass or fragment ion mass values, obtained by applying appropriate cleavage rules to the entries in a sequence database. Corresponding mass values are counted or scored in a way that allows the peptides or protein which best matches the data to be identified. 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 algorithms and programs have been available for protein database searching, such as Protein Prospector (http://prospector.ucsf.edu), ProFound (Zhang *et al.*, 2000) (http://prowl.rockefeller.edu/), PeptideSearch (http://www.narrador.embl-heidelberg.de), Mascot (Perkins *et al.*, 1999) (http://www.matrixscience.com/), and SEQUEST (Griffin *et al.*, 1995; Yates *et al.*, 1995a; Yates *et al.*, 1995b; MacCoss *et al.*, 2002; Tabb *et al.*, 2002; Tabb *et al.*, 2003b; Tabb *et al.*, 2003a).

The most widely used protein database searching programs are Mascot and SEQUEST. Mascot is based on the MOWSE algorithm (Pappin *et al.*, 1993) which incorporates probability-based scoring. It supports all three types of search, including: PMF, sequence query, and MS/MS 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. The cross-correlation function is used to score the proteins in the database.

1.2.3.2.5.2 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

NCBI (the National Center for Biotechnology Information) maintains composite, non-identical protein and nucleic acid databases. The entries in the protein database, nr, have been compiled from GenBank CDS translations, PIR, SWISS-PROT, PRF, and PDB.

MSDB

MSDB (Mass Spectrometry protein sequence DataBase) is a comprehensive, 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.

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

This is a downloaded protein database which can be added into the Mascot search engine to conduct the local protein database search. We have used TIGR Rice Genome Pseudomolecules version 2 (http://www.tigr.org/tdb/e2k1/osa1/) (Yuan *et al.*, 2003). 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 contigous sequence.

1.2.3.2.5.3 Evaluating database hits

Problems are commonly encountered in evaluating the significance of protein identification by database searching. There are several criteria for protein database

searching, especially in PMF results, such as spectrum quality, protein score, protein coverage, and peptide account. PMF results, especially of false positive candidates can be verified by using MS/MS results. Also MS/MS spectra should be examined to determine whether observed ions are consistent with the amino acid composition and chemistry of the candidate sequence.

1.2.3.2.6 Bioinformatics involved in proteomics

Proteomics technologies are under continuous improvements and new technologies are frequently introduced. Bioinformatics already plays a very important role in proteome analysis and will significantly increase the speed and the value of proteome analysis. Its important role in the analysis of proteome presents at several levels.

The first of these levels is in the 2D image analysis. 2D image analysis programs have been continuously improved and enhanced over the years in terms of faster matching algorithms with lesser manual intervention and with focus on automation and better integration of data from various sources. New 2D software packages have also emerged which offer completely new approaches to image analysis and novel algorithms for more reliable spot detection and matching. Several programs include options such as control of a spot cutting robot, automated import of protein identification results from mass spectrometry, superior annotation flexibility or multi-channel image merging of up to three different images to independent colour channels for fast image comparison.

The second level is in the development of proteomics tools, especially in MS data interpretation. Protein analysis using mass spectrometry aims at deriving information about the primary structure of peptides and proteins in order to identify proteins and their modifications. The role of bioinformatics is fundamental for the elaboration of mass spectrometry data. Typical tasks for MS interpretation software in proteomics are MS data pre-processing, peptide mass fingerprinting, peptide fragmentation fingerprinting and *de novo* sequencing. To process data efficiently, new software packages and algorithms are continuously being developed to improve protein identification and characterization in terms of high-throughput and statistical accuracy. The programs used to do the database search have already listed previously. Some of these are available via the internet with links to many through the ExPASy proteomics server (http://www.expasy.org/tools/). The

development of proteomics tools allows not only identification of proteins but further characterization ranging from calculation of basic physicochemical properties to prediction of potential post-translational modifications (http://www.cbs.dtu.dk/services/) and three-dimensional structures.

The third level is the creation of 2DE protein expression databases. An increasing number of 2DE protein expression databases are being established. The SWISS-2DPAGE (http://www.expasy.org/ch2d/2dindex.html) was a pioneering venture in this field and continues to be one of the most comprehensive 2-DE databases containing protein maps for human, mammals, mouse, yeast, plants, bacteria, viruses and specific cell lines.

Nowadays high throughput acquisition of proteome data is possible. The young and rapidly emerging field of bioinformatics is introducing new algorithms to handle large and heterogeneous data sets and to improve the knowledge discovery process in proteomics.

1.2.3.2.7 Post translational modification

Post translational modifications (PTMs) occur almost in all proteins analyzed to date. The function of a modified protein is often strongly affected by these modifications. Understanding the PTMs of a target protein may increase our understanding of the function of this protein, the molecular processes in which it takes place or the interaction among other proteins. So with the increased knowledge about PTMs and the development of technology in proteomics, PTMs have already become a new challenge of proteomics. Here we mainly discuss two major types of PTMs, glycosylation and phosphorylation.

1.2.3.2.7.1 Glycosylation

The most prevalent protein modification is glycosylation and 50% of all proteins are thought to be glycosylated based on bioinformatics analysis of genomic information (Hansen *et al.*, 1995; Apweiler *et al.*, 1999). Glycosylation not only determine the protein structure, have dramatic effect on protein properties such as *in vivo* protein stability, thermal stability, proteolytic susceptibility and solubility (Van den Steen *et al.*, 1998) but also affects protein localization and trafficking, antigenicity, biological activity and half-life, as well as cell-cell interactions (Stanley, 1992; Varki, 1993; Hounsell *et al.*, 1996). There are four types of glycosylation, N-linked glycosylation, O-linked glycosylation, C-

mannosylation and glycophosphatidlyinositol (GPI) anchor attachments, with O- and N-linked glycosylation as the major ones.

1.2.3.2.7.1.1 N-linked glycosylation

N-linked glycans are attached to the proteins by an amide bond to an asparagine residue. The process occurs in the endoplasmic reticulum-Golgi complex during translation and is known to influence protein folding. Several glycosylation processing enzymes exist in the ER. The sequence motif Asn-Xaa-Ser/Thr (Xaa is any amino acid except Pro) has been defined as a prerequisite for N-glycosylation (Gavel *et al.*, 1990).

1.2.3.2.7.1.2 O-linked glycosylation

O-linked glycosylation reactions may happen at two cellular locations in the cell. Complex O-glycosyaltion takes place in the Golgi. Recently, O-glycosylation has also been shown to occur in the nucleus and cytoplasm of cells by adding an GlcNAc to a serine or threonine residue (Hart, 1997; Snow *et al.*, 1998). This simple modification is dynamic and has been found at sites that are identical to those used by serine/threonine kinases. The O-GlcNAc modification appears to play a pivotal role in signaling (Wells *et al.*, 2001).There is no acceptor motif defined for O-linked glycosylation. The only common characteristic among most O-glycosylation sites is that glycosylation occurs on serine and threonine residues in close proximity to proline residues, and that the acceptor site is usually in a beta-conformation.

1.2.3.2.7.2 Phosphorylation

Protein phosphorylation is the primary means of switching the activity of a cellular protein rapidly from one state to another. Thus, phosphorylation is considered as being a key event in many signal transduction pathways of biological systems. Phosphorylation of substrate sites at serine, threonine or tyrosine residues is performed by members of the protein kinase family. Protein kinase cascades play essential roles in the regulation of cellular processes like metabolism, proliferation, differentiation and apoptosis in animals and yeast (Johnson *et al.*, 1996; Pinna *et al.*, 1996; Graves *et al.*, 1997; Kolibaba *et al.*, 1997; Hunter, 1998; Johnson *et al.*, 1998). In plants, there is evidence that phosphorylation plays an

important role in signaling pathways triggered by abiotic stress, pathogen invasion, and plant hormones (Knetsch *et al.*, 1996; Sheen, 1996; Zhang *et al.*, 1998).

1.2.3.2.7.3 Strategies for studying PTMs

Direct analysis of modifications and their positions within the protein sequence is required because genomic research is not able to predict post-translational modification of proteins for sure. Generally, PTM studies use several methods to separate modified peptides, and then compare the difference between the unmodified peptides with mass spectrometry or directly detect the modification by mass spectrometry. PTM analysis by mass spectrometry, however, poses some particular challenges. In contrast to glycosylation, phosphorylation seems easier to analyze due to their simple molecular structure, whereas glycans consist of a variety of complex substructures. For detail, please refer to some good reviews on PTMs are available (Blom *et al.*, 2004; Reinders *et al.*, 2004). With the technology development, more and more methods have been employed to study PTMs. For example, the newly developed Pro-Q Diamond phosphoprotein dye (Steinberg *et al.*, 2003) and Pro-Q Emerald 300 dye (Steinberg *et al.*, 2001) technology are suitable for the fluorescent detection of phosphoprotein and glycoproteins directly in SDS gels to the nanogram level. A new commercial western blot detection kit for O-GlcNAc modification is available with high specificity (Comer *et al.*, 2000; Comer *et al.*, 2001).

Bioinformatics tools are also widely employed in the PTM protein sequence analysis and mass spectrometry data interpretation. Several PTM prediction softwares are available (Table 1.2) that use prior knowledge to predict probable modification sites. The softwares are useful for rational design of experimental approaches aiming at determination of the actual utilizable modification sites. Prediction methods may work well for some proteins, but they often tends to generate many candidates. Thus, the putative sites should be confirmed by experimental data. For example, using GlycoMod (Cooper *et al.*, 2001a) (http://www.expasy.ch/tools/glycomod/), which is a software to compare the experimental mass with the predicted possible oligosaccharide structures that occur on proteins, we can find which peptides having the glycosylation and use such as MS/MS method to further confirm the results. There are also several databases which collect information on PTMs of known proteins, such as RESID and GlycoSuite, which are listed in Table 1.1.

PTM analysis is in its infancy, but the research field is developing very quickly. Integration of various technologies for PTM protein enrichment, detection, quantification and sequencing is needed. These methods have to be coupled to novel bioinformatics tools for analysis and visualization of the protein data that is obtained. Methods are continuously being refined at all levels, most notably in sample preparation methods, tandem mass spectrometry techniques and bioinformatics software. These developments will enable comprehensive and quantitative analysis of post translational modification events in cells.

Table 1.1 Post translational modification databases

RESID
RESID (Garavelli, 2003) (http://pir.georgetown.edu/pirwww/dbinfo/resid.html) is a database of protein post-translational modifications with descriptive, chemical, structural and bibliographic information.
GlycoSuite
GlycoSuite (Cooper *et al.*, 2001b; Cooper *et al.*, 2003) (www.glycosuite.com) is a database which is connected to GlycoMod and comprised over 8100 entries of known glycosylation and glycan structure.

Table 1.2 Post translational modification predication software

Phosphorylation site prediction	
ScanSite	http://scansite.mit.edu
NetPhos (Generic phosphorylation sites in eukaryotic proteins)	http://www.cbs.dtu.dk/services/NetPhos/
NetPhosK (Kinase specific phosphorylation sites in eukaryotic proteins)	http://www.cbs.dtu.dk/services/NetPhosK/
Prosite	http://www.expasy.org/prosite/
Glycosylation site predication	
NetNGlyc (N-glycosylation sites in human proteins)	http://www.cbs.dtu.dk/services/NetNGlyc/
NetAcet (N-terminal acetylation in eukaryotic proteins)	http://www.cbs.dtu.dk/services/NetAcet/
NetOGlyc (O-GalNAc (mucin type) glycosylation sites in mammalian proteins)	http://www.cbs.dtu.dk/services/NetOGlyc/
DictyOGlyc (GlcNAc O-glycosylation sites in *Dictyostelium discoideum*)	http://www.cbs.dtu.dk/services/DictyOGlyc/
YinOYang (O-(beta)-GlcNAc glycosylation and *Yin-Yang* sites (intracellular/nuclear proteins))	http://www.cbs.dtu.dk/services/YinOYang/

1.2.3.2.8 Other aspects of proteomics

Proteomics is an exciting new approach to biological and biochemical research. In this session, other newly developed techniques in proteomics will be reviewed.

1.2.3.2.8.1 2D DIGE

Two dimensional differential gel electrophoresis (2D DIGE) is a relatively new approach to quantitative proteome analysis using 2D gels (Unlu *et al.*, 1997). It involves separately labeling two or three different protein extracts with distinct fluorophores (cyanine dyes) that have similar charge, size, and reactivity toward lysine residues in the proteins. Because different cyanine dyes have distinguishable spectra, the individual fluorescently labeled proteins samples can be combined prior to gel electrophoresis. The proteins from each sample are visualized separately by imaging with appropriate excitation and emission wavelengths for the different dyes to obtain quantitative data. The ability to combine two or three labeled samples on a single 2D gel eliminates problems associated with gel-to-gel variation. Tonge *et al.*, (2001) have determined the quantitative variation in the 2-D DIGE process and established statistically valid thresholds for assigning quantitative changes between samples.

1.2.3.2.8.2 LC/LC-MS/MS

Liquid chromatography (LC) is an emerging alternative to 2DE for protein separation, which can overcome some persistent limitations of 2DE. Two of the most common LC/LC-MS/MS approaches are the multi-dimensional protein identification technology (MudPIT) and isotope-coded affinity tag (ICAT) methods.

1.2.3.2.8.2.1 MudPIT

The interesting multidimensional protein identification technology (MudPIT) approach was introduced by Yates laboratory (Washburn *et al.*, 2001). MudPIT is a technique for the separation and identification of complex protein and peptide mixtures. Rather than using traditional 2D gel electrophoresis, MudPIT separates peptides by 2D LC. In this way, the separation can be interfaced directly with the ion source of a mass spectrometer (Previous session for details).

1.2.3.2.8.2.2 ICAT

A recently developed, attractive method for quantitative comparison of two proteomes is the isotope-coded affinity tag (ICAT) method (Gygi *et al.*, 1999). Two forms of a cysteine-specific labeling reagent coupled to biotin affinity tag by a linker, which are known as heavy and light ICAT, are used. Heavy ICAT has eight deuterium atoms in the reagent's linker region, while light ICAT contains hydrogen instead. This creates 8 Da mass difference between two ICAT forms. To compare two samples, each sample is labeled with a different form of the ICAT reagent under conditions where all cysteines are derivatized (end-point labeling), the samples are combined, subjected to trypsin digestion, cysteine containing peptides are affinity purified using the biotin tag, and these peptides are analyzed by LC-MS/MS. The relative abundance of proteins identified by MS/MS analysis of cysteine-containing peptides is estimated from the relative amounts of the light and heavy forms of each peptide detected.

Use of ICAT also permits quantification and discrimination on the mass spectrometer itself. Several studies verifying the advantages of using ICAT for quantitative protein profiling have been recently published (Smolka *et al.*, 2002; Griffin *et al.*, 2003; Hansen *et al.*, 2003; Ranish *et al.*, 2003). In contrast to 2DE, the amount of protein that can be processed is theoretically unlimited. This is an important strength of this method which can improve the detection of low abundant proteins. The process of ICAT is easier to automate, and specific classes of protein such as very acidic, very basic, and membrane proteins, which are difficult to detect by 2DE may be readily detected. Conversely many protein changes that can often be readily detected on 2D gels may be difficult or impossible to detect using the ICAT method, including: proteolytic processing, changes in the spliced form of a protein that is expressed, and changes in PTMs.

1.2.3.3 Plant proteomics

Proteome analysis is becoming a powerful tool in the functional characterization of plants. Due to the availability of vast nucleotide sequence information and based on the progress achieved in sensitive and rapid protein identification by mass spectrometry, proteome approaches open up new perspectives to analyze the complex functions of model plants and crop species at different levels. Proteomics analysis of various tissues and organelles has revealed diverse functional categories of proteins.

Along with general limitations of the current available technologies, plant proteome approaches face specific challenges. Sample preparation is often more difficult due to the rigidity of plant cell walls or can be compromised by the accumulation of large quantities of secondary compounds in the central vacuole, which upon tissue disruption can lead to protein precipitation. Due to the availability of complete genomic sequence information on rice and Arabidopsis and of large mutant collections, a number of recent proteome studies have been carried out.

1.2.3.3.1 Proteome analysis of plant tissues and organs

Plant proteomic studies published to date have focused on global mapping of proteins, such as maize leaves (Porubleva *et al.*, 2001), maize endosperm (Mechin *et al.*, 2004), wheat grain (Skylas *et al.*, 2005), rice leaf, root and seed (Koller *et al.*, 2002). Mapping the proteomes of various organs, tissue (Watson *et al.*, 2003) and subcellular components (Peltier *et al.*, 2000; Andon *et al.*, 2002; Peltier *et al.*, 2002; Schubert *et al.*, 2002; Tanaka *et al.*, 2004; von Zychlinski *et al.*, 2005) have also been systematically studied.

Proteomic analyses of plant organs or tissues were also applied to monitor developmental changes or the influence of environmental stimuli on protein patterns and to compare lines with different genetic backgrounds. Developmental changes in the metabolic protein profiles of wheat endosperm have been studied at two time points (Vensel *et al.*, 2005), which provided insight into biochemical events taking place during wheat grain development and highlight the value of proteomics in characterizing complex biochemical processes. Rice caryopses proteins are very important to development and grain quality. A group of researchers (Lin *et al.*, 2005) not only profiled the developmental changes of rice caryopses proteins, but also found several proteins related to grain quality, and identified their high temperature response. These results should benefit the understanding of functions of rice caryopsis proteins and improvement of rice grain quality under temperature stress. Using proteomic analysis, an investigation aimed at a better understanding of the molecular adaptation mechanisms of cold stress was carried out in rice and six proteins were identified as cold responses proteins (Cui *et al.*, 2005). Grain filling and seed maturation are other important processes intensively studied in crop plants. Proteome analysis of barley seed during maturation (Finnie *et al.*, 2002) found changes of

some proteins which are related to the processes known to be taking place in the seed and cultivar-specific spot variations.

The complete genomic sequence of rice (Goff *et al.*, 2002), one of the most important crop plants, holds great promise for functional genomics including proteome approach as a tool to detect novel traits for breeding. Details of rice proteomics will be discussed in the later session.

1.2.3.3.2 Plant organelle proteomics

Although 2D gels were invented in 1975, the technology is still tedious and difficult to apply. The crucial question is how to relate the changes in expression level of proteins on 2D gels to the biology of a system when we can see only a minor fraction of all the proteins present.

Subcellular fractionation allows the characterization of individual organelle proteomes, and the protein location in relation to their function. Subcellular fractionation consists of two major steps, disruption of the cellular organization and fractionation of the homogenate to separate the different organelles. Such a homogenate can then be resolved by differential centrifugation into several fractions containing mainly 1) nuclei, heavy mitochondria, cytoskeletal networks, and plasma membrane; 2) light mitochondria, lysosomes, and peroxisomes; 3) Golgi, endosomes and microsomes, and endoplasmic reticulum (ER); and 4) cytosol. Each oganelle is characterized by size, density, charge, and other properties on which the separation relies (Pasquali *et al.*, 1999). Most of the currently available plastid proteome information that provides new insights into organelle specific metabolic functions has been reported from chloroplasts. Systematical analyses of protein fractions enriched for lumenal and/or peripheral proteins of the thylakoids from pea, spinach and *A. thaliana*, were carried out (Kieselbach *et al.*, 1998; Peltier *et al.*, 2000; Peltier *et al.*, 2002; Schubert *et al.*, 2002; Friso *et al.*, 2004).

Several recent papers describe efforts to experimentally identify the inner and/or outer envelope proteome. In a series of excellent papers, the proteome of highly purified envelope membranes of chloroplasts from spinach and *A. thaliana* was systematically analyzed, using extraction with alkaline or saline treatment, chloroform/methanol of

different ratios, 1DE and off-line and on-line (ESI-MS/MS) analysis (Seigneurin-Berny *et al.*, 1999; Ferro *et al.*, 2002; Ferro *et al.*, 2003). The fractionation strategy, using salt washes as well organic solvent extraction, proved important to extend the number of identified proteins. Proteins of highly purified outer envelope membranes of pea chloroplasts were also analyzed by ESIMS/MS (Schleiff *et al.*, 2003).

Recently, an extensive proteomics study was published in which chloroplast proteins were separated with different fractionation techniques, followed by LC-ESI-MS/MS, resulting in the identification of 687 proteins (Kleffmann *et al.*, 2004). So far, comparative proteomics has only rarely been applied to tackle these important issues in chloroplasts. Lonosky *et al.* (2004) addressed the statistical analysis of comparative proteomics using 2D gels, with the plastid proteome of greening maize leaves as example.

For mitochondria proteome, large-scale analysis of mitochondrial proteins was carried out in plants with high resolution (Kruft *et al.*, 2001; Millar *et al.*, 2001; Bardel *et al.*, 2002; Bykova *et al.*, 2003a; Heazlewood *et al.*, 2003; Hochholdinger *et al.*, 2004). PTM studies have also been published for the mitochondrial proteome (Bykova *et al.*, 2003a; Bykova *et al.*, 2003b; Balmer *et al.*, 2004; Kristensen *et al.*, 2004). Integrating these predicted and experimental datasets is now required for an in-depth understanding of the mitochondrial proteome and for its PTMs to further understand the diversity of mitochondrial function in plants.

A few pioneering efforts resulting in the first catalogues of proteins are noted for cell wall (Robertson *et al.*, 1997), plasma membrane (Santoni *et al.*, 1998), endoplasmic reticulum, and Golgi apparatus (Prime *et al.*, 2000; Maltman *et al.*, 2002) and peroxisome (Fukao *et al.*, 2002), and amyloplast proteome in wheat (Andon *et al.*, 2002).

Subcellular fractionation and purification of organelles has always been a challenge. Most fractionation protocols take advantage of the physical property of proteins, e.g., density or charge. However, differential subcellular compartments may share similar physical properties leading to co-fractionation at least to some extent in conventional gradients and difficulty in separation by centrifugation. Many assays can be used to assess the results of subcellular fractionation. For evaluation of the mechanical and functional integrity of organelles, several methods may be used. First, quantitative western blotting to follow the

distribution of specific organelle-marker proteins can be used (Fialka *et al.*, 1997; Pasquali *et al.*, 1999). Second, for morphological analysis of all fractions, standard electron microscopy procedures can be applied. It is essential to point out that complete purification is, with few exceptions, hardly possible. However, it is still very powerful for proteome studies.

Nevertheless, subcellular fractionation and proteomics are an ideal combination. Subcellular fractionation allows access to intracellular organelles and multi-protein complexes. Low abundant proteins and signaling complexes can be enriched, and at the same time complexity of the sample can be reduced. Analyzing subcellular fractions and organelles allows also tracking proteins that shuttle between different compartments, eg, between the cytoplasm and nucleus. Importantly, subcellular fractionation is a flexible and adjustable approach that may be efficiently combined not only with 2DGE but also with gel-independent techniques.

1.2.3.3.3 Post translational modifications in plant

Proteins in plants and other organisms undergo numerous PTMs, which help to regulate protein function and can alter protein localization. Although information of PTMs in plant proteomics is very limited, our understanding of the regulation of plant metabolism by PTM has increased substantially over the past few years. Protein phosphorylation and thiol/disulfide redox modulation are important modifications in plants, and it is likely that O-glycosylation and S-nitrosylation will also emerge as important mechanisms. It is known that several thylakoid proteins are reversible phosphorylated in response to environmental changes. Vener and colleagues used MS to analyze the reversible phosphorylation of surface-exposed hydrophylic loops of the more abundant thylakoid proteins under different physiological conditions (Vener *et al.*, 2000). The combination of 2DE with Triton X-114-based two-phase separation (Sherrier *et al.*, 1999) showed that glycosylphosphatidylinositol-anchored proteins are a relatively abundant class of proteins at the plant plasma membrane and extracellular matrix. One of these was shown to be an arabinogalactan protein, a class of proteins known to be associated with cellular differentiation. Redox control of AGPase through sucrose supply in potato tubers has recently been proposed, whereby reductive activation causes channeling of carbon to starch and away from respiratory/glycolytic metabolism (Tiessen *et al.*, 2002). A recent

study (Balmer *et al.*, 2006), which applied proteomic and immunological approaches, found that amyloplasts isolated from wheat starchy endosperm resemble chloroplasts in containing a complete ferredoxin (a Fe-S protein)-thioredoxin (a regulatory disulfide protein) system composed of ferredoxin, ferredoxin/thioredoxin reductase (FTR), and thioredoxin (m-type). The finding of a complete ferredoxin/thioredoxin system opens the door to a role for redox in reversibly regulating enzymes of amyloplasts as established for chloroplasts.

1.2.3.4 Recent progress in rice proteomics

1.2.3.4.1 General introduction of 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. 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); green and etiolated shoot (Komatsu *et al.*, 1999); cultured suspension cells (Komatsu *et al.*, 1999); anther (Imin *et al.*, 2001); and leaf sheath (Shen *et al.*, 2002). 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). Tsugita *et al.* (1994) have systematically analyzed and identified 4892 proteins from nine tissues and one organelle of rice (leaf, stem, root, germ, dark germinated seedling, seed, bran, chaff, callus and chloroplast). Recently, a survey of the proteome complement of rice root, leaf, and seed tissues was obtained by application of both 2-D gel electrophoresis and LC-based separation methods for complex peptide mixtures after digestion of protein extracts. Based on both methods, the identification of 2528 unique proteins (6296 peptides) was achieved (Koller *et al.*, 2002).

The proteomes of Golgi, mitochondria and other subcellular compartments have also been studied. The analysis of proteins from rice chloroplasts was the first organelle proteome work (Tsugita *et al.*, 1994). Golgi complex is a very important multifunctional organelle in rice particularly for biosynthesis of complex cell surface polysaccharides and the processing and modification of glycoproteins. This work was done by Mikami *et al.* (Mikami *et al.*, 2001). Mitochondria play a pivotal role in energy metabolism in eukaryotic cells. A comprehensive report on rice mitochondrial proteome had been presented (Heazlewood *et al.*, 2003), which used blue native SDS-PAGE and reverse phase HPLC to

separate the hydrophobic membrane proteins and peptides. Using proteome analysis von Zychlinski *et al.* (2005) examined the global state of protein expression in rice etioplasts to establish comprehensive information on complex metabolic and regulatory networks that function in a heterotrophic plastid. The authors developed a novel protocol for isolation of highly purified rice etioplasts (100% purity) and subsequent systematic identified 240 unique proteins of rice etioplast proteome which provide new insights into heterotrophic plant metabolism and control of gene expression. However, systematic analysis of the proteome of rice amyloplast, especially for hybrid rice seeds during development was not been reported.

1.2.3.4.2 Rice proteome database construction

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-2DPAGE 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 recently constructed rice proteome database website (http://gene64.dna.affrc.go.jp/RPD/) provides extensive information on the progress of rice proteome research (Komatsu *et al.*, 2004). Rice etioplast protein identifications and related data were integrated into a database that is available at https://www.pb.ipw.biol.ethz.ch/plprot (von Zychlinski *et al.*, 2005).

1.2.3.4.3 Comparative proteomics

Comparative proteomics have been raised in recent years with the hope to find differences between two or more samples or treatments. Several studies have investigated temporal changes in plant proteomes involving two or more different time points (Wilson *et al.*, 2002; Gallardo *et al.*, 2003; Shen *et al.*, 2003; Watson *et al.*, 2003; Vensel *et al.*, 2005; Zhao *et al.*, 2005). However, quantitative measures of reproducibility were not reported (only qualitative) nor rigorous quantitative analyses were conducted to group proteins into expression classes (e.g. clustering analyses). 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). While for rice, in 1999 a comparative 2DE proteomic study was performed on green and etiolated shoots of rice, and 85 proteins were subjected to analysis by gas-phase sequencer (Komatsu *et al.*, 1999). Proteomic changes in rice leaves during development of

field-grown rice plants were also studied (Zhao *et al.*, 2005). Another research (Tanaka *et al.*, 2005) using Unweighted Pair Group Method with Arithmetic Mean (UPGMA) clustering method to study the protein profiles of the rice basal region and S-system so as to analyse the interactions of the clusters had also been reported.

1.2.3.4.4 Post translational modification study of rice proteome

A complete functional understanding of the proteome requires, however, full characterization of the PTMs of proteins and the complex networks of protein-protein interactions. However, few studies have been reported about PTMs in plants, especially in starch synthesis related proteins.

Protein kinase cascades play essential roles in diverse intracellular signaling processes in animals and yeast. In plants, there is evidence that protein phosphorylation plays an important role in signaling pathways triggered by abiotic stress, pathogen invasion, and plant hormones (Knetsch *et al.*, 1996; Sheen, 1996; Zhang *et al.*, 1998). A recent study with isolated amyloplasts from wheat endosperm identified a number of phosphoproteins including some involved in starch metabolism, indicating that some aspects of starch (amylopectin) biosynthesis may be controlled by protein phosphorylation (Tetlow *et al.*, 2004b). This is the first report that establishing a starch synthesis protein complex model which may have been regulated by protein phosphorylation. The results also revealed the importance of phosphorylation to the starch synthesis.

The function of protein glycosylation was related to the storage protein stability during accumulation (Lamport, 1980; Shewry *et al.*, 1995b; Kishimoto *et al.*, 1999) and to the prevention of the degradation of the storage proteins during seed germination and in early seedling stage (Faye *et al.*, 1989). In rice caryopsis proteins, both glutelin (rice storage protein) and GBSS have been identified to be glycosylated (Lin *et al.*, 2005). Storage proteins have been reported glycosylated earlier (Kishimoto *et al.*, 1999), however the function of glycosylation of GBSS, the enzyme to elongate the amylose, remains unclear. Thus further characterization of PTMs is necessary to understand its function of PTMs in rice proteome and starch synthesis.

Chapter 2 Materials and methods

2.1 Materials

2.1.1 Plant materials

The super hybrid rice (*Oryza sativa L. ssp. Indica*) parental line 9311 was planted in the growth facility at the Chinese University of Hong Kong. For amyloplast preparation, developing seeds were labeled on the days of flowering and harvested at 10 DAF and used immediately.

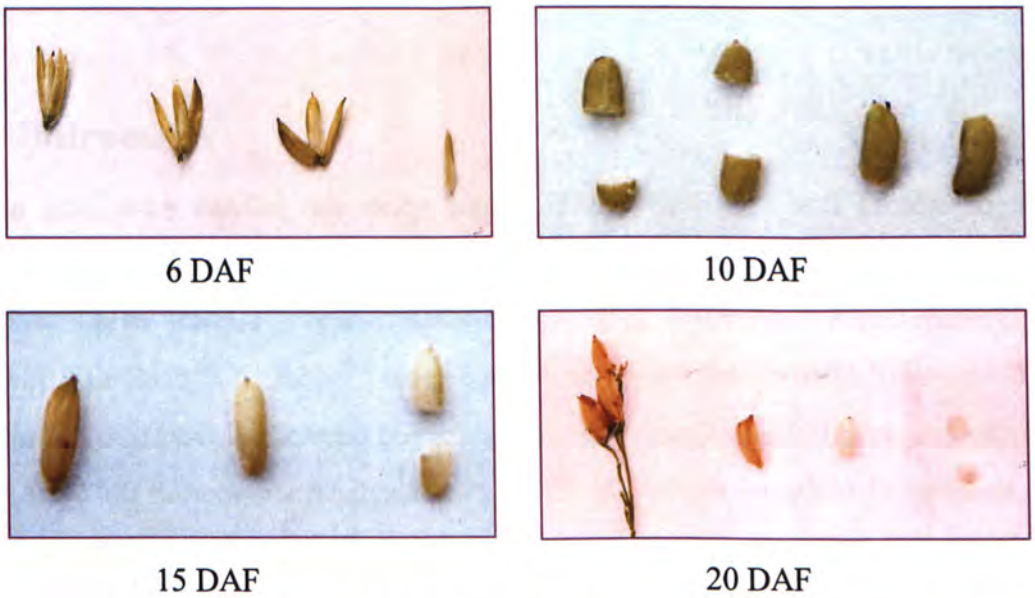


Figure 2.1 F1 seeds of super hybrid rice (9311/PA64S) at different developmental stages (Duan, 2003)

For protein profiling, the seeds of super hybrid rice parental lines 9311 and PA64S and F1 hybrid were provided by Hunan National Hybrid Rice Research and Developing Center (HNHRRDC). Because the female parental hybrid rice line (PA64S) is sterile at higher temperature (>23°C) and fertile at temperature below 23°C, the self-crossing seeds of PA64S were collected in the experimental field treated with cold water (19-21°C average). This cold treatment induces and converts male sterility into fertility at the fertility sensitive

stage, i.e. during pollen mother cell formation. Male parental hybrid rice line 9311 is relatively easy to set self-crossed seeds at different stages due to its fertility, bigger stigma, and vigorous pollen activity. The F1 hybrid seeds were acquired by artificial pollination. Seeds were labeled on the days of flowering and harvested at 6, 10, 15, 20 DAF (Fig 2.1).

2.1.2 Chemical reagents and commercial kits

All chemicals of analytical grade were from Sigma (St. Louis, MO, USA). The Plusone™ 2D Quant kit, IPG strips including pH3-10, 4-7, 3-11NL from 7cm, 18cm to 24cm, IPG buffer (pH3-10, 4-7 and 3-11NL), precast polyacrylamide gels (DALT Gel 12.5, 255×196×1 mm) and Plusone™ silver staining kit were obtained from Amersham Bioscience (Uppsala, Sweden). Modified trypsin (sequencing grade) was purchased from Promega (Madison, WI, USA). Cytochrome c oxidase assay kit was obtained from Sigma (St. Louis, MO, USA).

2.1.3 Instruments

Protein gels were carried out using Mini-PROTEAN 3 Cell and Protean II xi Cell electrophoresis apparatus purchased from Bio-Rad Company (Hercules, USA), and Ettan DALT*six* Large Format Vertical System was from Amersham Bioscience (Uppsala, Sweden). The Ettan™ IPGphor™ Isoelectric Focusing System for first-dimension IEF was also from Amersham Bioscience (Uppsala, Sweden). Mass spectrometry was carried out using AB 4700 Proteomics Analyzer with TOF/TOF Optics (Applied Biosystems, Foster City, CA, USA).

2.1.4 Software

Listed below are the softwares used in this study:

Image analysis was carried out using ImageMaster 2D Platinum, Version 5.00 (Amersham Biosciences, Uppsala, Sweden).

Mass spectrometry analysis was carried out using the 4700 Explorer™ Software while database searching was by the GPS Explorer™ software (Applied Biosystems, Foster City, CA, USA).

Protein transit peptide cleavage site was predicted by the ChloroP algorithm (http://www.cbs.dtu.dk/services/ChloroP/).

Protein signal peptide cleavage site was predicted by the SignalP algorithm (http://www.cbs.dtu.dk/services/SignalP/).

SOM (Self Organizing Map) was performed on normalized medians using version 1.0 of Gene Cluster (http://www-genome.wi.mit.edu/cancer/software/software.html).

2.2 Methods

2.2.1 Fractionation of amyloplast and amyloplast membrane proteins

Purification of rice amyloplast and amyloplast membrane proteins was based on the procedure of wheat amyloplast isolation with some modifications (Andon *et al.*, 2002). Rice seeds were harvested on ice and use immediately. After dehusked, rice endosperm was gently squeezed out and collected in the collection buffer (0.5M sorbitol, 50mM HEPES pH7.5) on ice. The collection buffer was then removed and replaced with chopping buffer (0.8M sorbitol, 50mM HEPES pH7.5, 1mM EDTA, 1mM KCl, and 2mM $MgCl_2$). After 1h at 4°C, endosperm tissue was finely chopped with a razor blade. About 6ml crude homogenate was gently pipetted off and filtered through four layers of Miracloth (Calbiochem, Bioshops Stortford, Herts, UK) onto a 6ml 2% Nycodenz layer (Sigma, St. Louis, MO, USA) which was dissolved in chopping buffer. Amyloplasts were separated from the endosperm extract by centrifugation through the Nycodenz cushion at 30g for 10min. The amyloplasts pellet was resuspended into chopping buffer and centrifuged through a 2% Nycodenz layer again. The amyloplast pellet was washed 4 times with 1ml SDS buffer (62.5mM Tris-HCl, pH6.8; 10% glycerol; 5% 2-mercaptoethanol; 2.8% SDS) and 3 times with acetone before store at -80°C. For the membrane fraction purification, the amyloplast pellet was resuspended in 1 mL of rupturing buffer (100mM Tricine-NaOH pH7.8, 1mM EDTA, 1mM DTT, 1mM PMSF, 100μM each of leupeptin, bestatin, pepstatin, 3,4-dichloroiso coumarin, chymostatin, 1,10-phenanthroline, phosphoramidon, and pefabloc (Boehringer Mannheim, Indianapolis, IN, USA), and ruptured by three cycles of freezing in liquid nitrogen and thawing on ice. Ruptured amyloplasts were centrifuged at 10,000g for 10min at 4°C to pellet the starch and debris. The supernatant was removed and membranes were pelleted by centrifugation at 100,000g for 20min at 4°C. The membrane pellet was stored at -80°C before use.

2.2.2 Marker enzyme assays

Amyloplast intactness and yield were analyzed by comparing the activities of cytosolic and plastid marker enzymes bwtween lysed and intact amyloplast preparations. Amyloplasts were ruptured by addition of 0.1% (v/v) Triton X-100 followed by vortexing. Marker enzymes used were alcohol dehydrogenase (ADH; EC 1.1.1.1), cytochrome C oxidase (EC 1.9.3.1) and alkaline pyrophosphatase (APPase; EC 3.6.1.1). Both enzymes were assayed at 25°C according to established methods for APPase (Gross *et al.*, 1986), ADH (MacDonald *et al.*, 1983) and cytochrome C oxidase (technical bulletin of Sigma). The intactness of amyloplasts was measured by the latency of marker enzymes. Latency (L) is the difference in enzyme activities in ruptured (R) and intact (I) amyloplasts, relative to that of the ruptured (R) and expressed as a percentage: L= ((R-1/R))*100.

2.2.3 2D gel electrophoresis

2DGE was performed according to the manufacturer's instructions (Amersham Biosciences, Uppsala, Sweden). The rehydration buffer (8M urea; 4% CHAPS; 19mM DTT; 2% IPG Buffer, 0.002% bromophenol blue) was directly added to the amyloplast pellet. This mixture was extensively vortexed for 1h and centrifuged (14,000g, 10min) to pellet insoluble materials. The protein concentration in the supernatant was measured with PlusOne™ 2D Quant Kit. IPG strips with pH4-7 and pH3-11 NL 18cm were used for protein identification, and pH3-10 24cm IPG strips were used for protein profiling. To improve the entry of high molecular weight proteins into IPG strips, low voltage was applied during rehydration. IEF was performed at 20°C on an IPGphore™ for rehydration for 8h, 30V for 6h, 500 V for 0.5h, 1000 V for 0.5h, 3000V for 1h, 4000 V for 1h, and 8000 V for 4 h, for a total of 39, 930 Vh. Prior to second dimension SDS-PAGE, the IPG strips were first equilibrated for 15 min in 6M Urea, 30% glycerol, 2% SDS, 50mM Tris-HCl pH 8.8, and 1% DTT and secondly for 15 min in 6M urea, 30% glycerol, 2% SDS, 50mM Tris-HCl pH 8.8, and 2.5% iodoacetamide. 18 cm IPG strips were performed on 12.5% self-cast SDS-PAGE gels and 80V constant voltage per gel in a Protean II xi Cell. 24cm IPG strips were performed on 12.5% precast gels (DALT Gel 12.5) and 5W constant Watt per gel in the Ettan DALT*Six* Large Format Vertical System. The marker proteins were applied on the gels via a small piece of filter paper for calibration of the molecular weights. The gels were stained by Plusone™ silver staining kit.

2.2.4 Silver staining of 2D gel

The silver staining was performed according to the instruction of Plusone^{TM} silver staining kit. The gels were fixed in fixing solution (40% ethanol, 10% acetic acid and 50% distilled water) for 30min each time. Then sensitized the gels by sensitizing solution without glutardialdehyde (30% ethanol, 0.2% sodium thiosulfate, and 6.8%(w/v) sodium acetate) for 30min. This was followed by washing the gels with distilled water three times, for 5min each time and silver reaction in a 0.2% silver nitrate solution without formaldehye for 20min. After rinsed them twice (1min each time) the gels were developed in developing solution (2.5% sodium carbonate and 0.148‰ formaldehyde) until appropriate protein spots were found. The gels were then transferred to the stop solution (1.46% EDTA-Na_2) for 20min and washed with distilled water twice, 20min each time. For mass spectrometry analysis, the gels were preserved in the distilled water in the dark. For long time preservation, the gels can be stored in preservation solution including 8.7% glycerol.

2.2.5 In-gel digestion of protein spots

Selected spots were carefully excised with a needle. Silver staining spots were destained in sodium thiosulfate (100mM): potassium ferricyanide (30mM) (1:1 v/v) and equilibrium in 200mM ammonium bicarbonate 10 min twice. For Coomassie blue staining, the spots were destained in ammonium bicarbonate (50mM): methanol (1:1 v/v). After dehydrating with 100% ACN (acetonitrile) three times and dried in speed vacuum (Gene Co. Ltd.) for 5min, a total of 10μl trypsin solution (40ng/μl trypsin in 50mM ammonium bicarbonate) was added to the dried gel pieces. More 50mM ammonium bicarbonate was added to cover the hydrated spots and incubated at 30°C overnight. Resulting peptides were extracted using, 20μl of 50% ACN and 0.1% TFA (trifluoracetic acid) and sonicated for 10min with ultrasonic cleaner (China Scientific (HK) Ltd.). Then 20μl of the second extraction buffer (ACN: 5% TFA 1:1) was added and sonicated for 10min again. After removing the supernatant into a new microtube extraction was repeated for 3 times with the second extraction buffer and the supernatant extracts were combined. Finally, dried the supernatant extracts with speed vacuum.

2.2.6 Desalination of the digested sample with ZipTip

The vacuum dried sample was resuspended in 10μl 0.1% TFA. The ZipTip μC18 (Millipore, Billerica, MA, USA) was wetted by aspirated 10μl wetting solution (50% ACN) into the tip and dispensed to waste twice. The wetted ZipTip was equilibrated by aspirating 10μl equilibration solution (0.1% TFA) into the tip and dispensed to waste twice. After binding the sample into ZipTip by aspirating and dispensing 10 cycles in the tube the ZipTip was washed with 10μl washing solution (0.1% TFA) by aspirating and dispensing into waste for 5 times. At last, the digested peptides were eluted with 2μl elution solution by aspirating and dispensing for 5 times in a clean 0.5 ml microtube. After desalting with ZipTip, 0.5μl of the peptide mixture was spotted onto AB 4700 proteomics analyzer plate using CHCA (α-cyano-hydroxycinnamic acid) as a matrix.

2.2.7 Protein identification by mass spectrometry and database searching

Protein identification was done by using AB 4700 Proteomics Analyzer 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.

2.2.8 Image and data analysis

2D gels were scanned by 2D ImageScanner (Amersham Biosciences, Uppsala, Sweden) The spots were automatically detected by the software ImageMaster 2D Platinum, Version 5.00 (Amersham Biosciences, Uppsala, 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. The quantity of the spot which did not appear on 2-D gels was assumed as 0. A repeated measurement was given as the mean ± SEM. Scatter plot analysis was done by the ImageMaster. SOM was performed on normalized medians using version 1.0 of Gene Cluster.

2.2.9 Extraction of starch granule associated proteins

The method of extraction of starch granule associated proteins was according to the previous method (Peng *et al.*, 1999). Aliquots of 50mg of endosperm powders were washed 4 times with 1ml SDS buffer (62.5mM Tris-HCl, pH6.8; 10% glycerol; 5% 2-mercaptoethanol; 2.8% SDS) and 3 times with acetone. The granule associated proteins were released by boiling in 0.5ml SDS buffer for 10min under constant agitation. The resulting starch paste was frozen at -20°C for 1h, and thawed on ice for 20min before centrifugation at 25,000g for 30 min. The starch granule associated proteins were in the supernatant.

2.2.10 Western blot analysis

After separated by SDS-PAGE, the proteins were transferred into Nitrocellulose memberane using BioRad Trans-blot electrophoretic transfer cell. Western blot analysis was performed according to AURORA western blot chemiluminescent detection system (Vancouver, B.B., Canada). Nitrocellulose with the bound proteins was washed for 5min with PBST (1x phosphate-buffered saline with 0.1% Tween 20) for 3 times and blocked for 90 min with blocking buffer containing 1x PBS and 0.2% AURORA blocking reagent. After Nitrocellulose was blocked by blocking buffer with GBSS-specific antiserum (G. Wong, the Chinese University of Hong Kong) for 60min, it was washed by PBST for 3 times, 10 min for each time. Followed by blocking for 45 min with 1:5000 (v/v) anti-rabbit alkaline phosphatase conjugate (GIBOCOL) and washed by PBST for 3 times, 10min for each time. Finally, GBSS signal was detected using AURORA chemiluminescent substrate.

2.2.11 Sample preparation for N terminal sequencing

When N-terminal sequencing was performed, the SDS-PAGE gels were blotted onto PVDF membranes (Bio-rad, Hercules, USA) using the Bio-Rad TransBlot Cell according to the manufacturer's instructions. After transfer at 44V for 2h, membranes were washed

twice for 5min with deionized water and then stained for 5min with the Coomassie blue staining solution containing: 0.1% (w/v) Coomassie Blue R-250, 10% (v/v) acetic acid and 40% (v/v) methanol. The membranes were destained with a solution of 10% (v/v) acetic acid and 40% (v/v) methanol. Proteins were excised and sequenced on an ABI 492 Procise sequencer (Applied Biosystems, Foster City, CA, USA).

2.2.12 Phosphorylation and glycosylation assays

After SDS-PAGE, selected gels were subjected to phosphorylation and glycosylation assays. The phosphorylation (Pro-Q Diamond Phosphoprotein Gel Stain Kit P-33300; Molecular Probes, Eugene, OR, USA) and glycosylation (Pro-Q Emerald 300 Glycoprotein Gel and Blot Stain Kit P-1857; Molecular Probes, Eugene, OR, USA) assays were performed according to the manufacturer's instructions. Stained gels were visualized with a UV transilluminator at 300 nm.

In brief, for phosphoprotein, after SDS-PAGE, the gel was immersed in the fixing solution (50% methanol and 10% acetic acid) and incubated at room temperature with gentle agitation for 30min. After washing in dH_2O for 10min for 4 times, the gel was incubated in Pro-Q Diamond phosphoprotein staining solution for 75-120min in the dark. The gel was then incubated in the dark in the destaining solution (4% ACN and 50mM sodium acetate, pH4.0) at room temperature for a total of about 2h with two changes of destaining solution.

For glycoprotein gel, immerse the gel in the fixing solution (50% methanol and 5% acetic acid) and incubate at room temperature with gentle agitation for 45min. After fixing, incubate the gel in the washing solution (3% glacial acetic acid) for 10min twice. Oxidize the gel in the oxidizing solution (250ml of 3% acetic acid to the bottle containing the periodic acid) for 30min and then wash in the washing solution for 3 times each for 10-20min. The Pro-Q Emerald 300 stock solution (adding 6ml *N'N*-Dimethylformamide (DMF) to the vial containing the Pro-Q Emerald 300 reagent) was diluted 50 folds with Pro-Q Emerald 300 staining buffer before incubating the gel in the dark for 90-120min. The gel was incubated in the washing solution at room temperature for 3 times 15-20min each time.

Chapter 3 Results

3.1 Protein identification by 1D and 2D PAGE

3.1.1 Isolation and purification of amyloplasts from rice seeds

We have developed a series of fractionation processes using Nycodenz density gradient centrifugation which allows the isolation of intact and highly pure amyloplasts from rice seeds. In this protocol, a 2% Nycodenz layer centrifugation is used twice. The second 2% Nycodenz layer can further purify the amyloplasts, by largely removing the contaminants from other organelles, such as cytosol and mitochondria (Tetlow *et al.*, 1993).

The purified amyloplasts fraction was evaluated by light microscopy and marker enzymes assay. Light microscopy analysis with iodine staining (Fig 3.1) showed that the amyloplasts by this protocol contained very little extraneous materials from surrounding tissues.

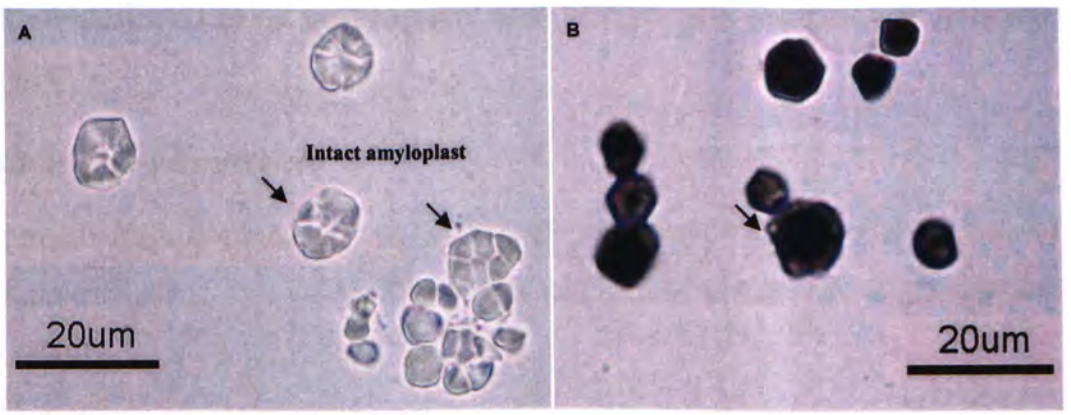


Figure 3.1 Light microscope observation of amyloplast pellets

A. Rice amyloplasts; B. Rice amyloplasts stained with iodine solution. The arrow points to an intact amyloplast.

The integrity and function of the purified amyloplasts were assayed to ensure that proteins were not being lost by amyloplast rupture during isolation and that the key functions were maintained. Table 3.1 shows the results of marker enzyme assay and intactness of amyloplasts. The amyloplast samples were found to be 62% intact based on the latency of APPase activity. The intactness at 62% is relatively high in comparison to previous similar research (Tetlow *et al.*, 1993).

Table 3.1 Marker enzyme assay and intactness for rice amyloplast

Marker enzyme	Cell compartment	Specific activity (μmol/ml min μg)	
		amyloplast pellet	Crude
ADH	Cytosol	0	0.0245
Cyt C Oxidase	Mitochondria	0	$7.38x10^{-3}$
APPase	Amyloplast	$2.65x10^{-4}$	
		$4.26x10^{-4}$ (ruptered)	
		62% (intactness)	

3.1.2 Identification of amyloplast and amyloplast membrane proteins by MS/MS

3.1.2.1 Sample preparation

For protein identification, we choose 10DAF 9311 seeds as a model because at this developmental stage, cell differentiation was taken place and the starch granules increase most rapidly.

For sample preparation, because several proteins are intrinsically associated with starch granules, such as GBSS, it is quite difficult to extract proteins from amyloplasts. In order to extract proteins, the starch granules thus need to be solubilized. We directly used the 2D rehydration buffer which contains 8M urea to solubilize starch granule (Fig 3.2A). Although TCA/acetone preparation method is widely used in protein enrichment, we did not employ this method because starch and polysaccharides will co-precipitate with TCA/acetone and the proteins will be dramatically lost. Fig 3.2B shows the 2D image of amyloplast proteins using total protein extraction buffer (62.5mM Tris-HCl, pH6.8; 10%

glycerol; 5% 2-Mercaptoethanol; 2.8% SDS) at 99°C followed with TCA/acetone precipitation method in comparison to the 2D rehydration buffer method (Fig 3.2A).

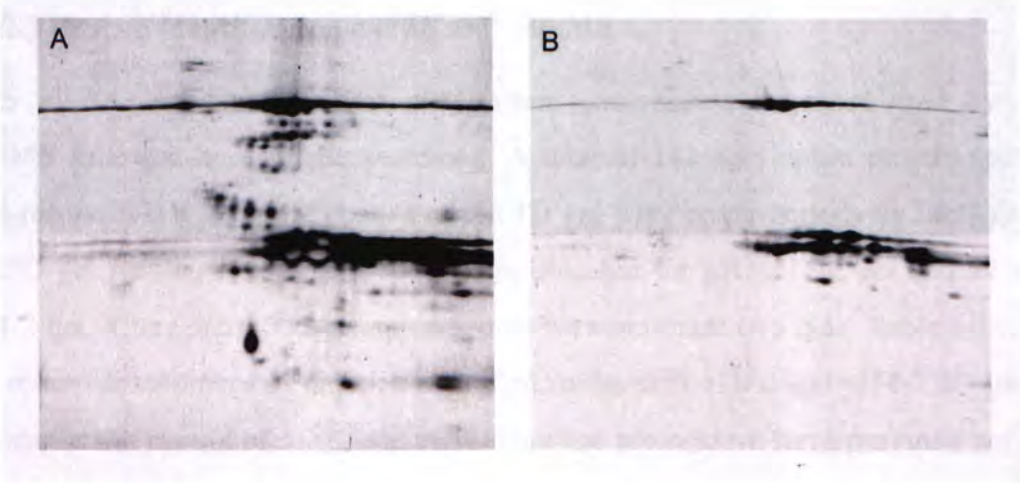


Figure 3.2 Comparison of different extraction methods for amyloplast proteins.

A. 2D image of amyloplast proteins using 2D rehydration buffer method; B. 2D image of amyloplast proteins using total protein extraction buffer followed with TCA/acetone precipitation method. Sample load: 5 μg of proteins; first dimension: pH 3-10, 7cm IPG strip; Second dimension: SDS-PAGE (12.5%); staining method: silver stain.

3.1.2.2 2D and 1D gel electrophoresis

The rice amyloplast proteins were first focused in an 18cm, pH3-11 NL IPG strip then followed by a separation on 12.5% polyacrymide gel. In order to separate the protein completely, we further employed an 18cm narrow pH4-7 strip. Fig 3.3A shows the image of 2D gel of rice amyloplast proteins with pH3-11 NL and Fig 3.3B shows the image of 2D gel of rice amyloplast proteins separated by pH4-7 IPG strip. Although low voltage was applied during rehydration to improve the entry of high molecular weight proteins into the IPG strip, the high molecular weight proteins such as starch branching enzyme were quite difficult to entry into the IPG strip. So 1DE was carried out. A great difference of sample preparation between 1DE and 2DE is the use of detergent. Since the very strong SDS detergent is not allowed in IEF preparation, some proteins may not be extracted without the presence of SDS during sample preparation, such as some hydrophobic proteins. Also, 1DE can accommodate high protein sample loading which in turn allows the detection of more low-abundance proteins. Fig 3.4A shows a 1DE of total protein of

rice endosperm and Fig 3.4B shows a 1DE of rice amyloplast proteins. We have also examined the amyloplast membrane proteins in 1DE, as shown in Fig 3.4C.

3.1.2.3 Protein identification by MS and MS/MS

After electrophoresis separation, the protein components were analyzed by MS and MS/MS followed by database searching. A total of 141 amyloplast protein spots of the well-resolved 2D gel and 25 bands of the 1D gel were comprehensively identified. From the 2D gel results, 104 protein spots were obtained for pH3-11NL gel and 91 spots for pH4-7 gel. There were 54 overlapping spots between these two gels. Table 3.2 and Table 3.3 show the summary of proteins identified in the pH3-11NL and pH4-7 2D gels. Table 3.4 shows the results of 1DE. For each identified protein, we have provided not only the spot ID, accession number, protein name, calculated pI/Mw, experimental pI/Mw and functional classification of the protein, but also the information on protein identification, such as protein coverage, protein score and peptide account. Additionally, we predicted the transit and signal peptide for each protein using the ChloroP and SignalP online software. The amyloplast proteome we reported covers a wide variety of enzymes and provides a comprehensive view of their metabolic pathways in the amyloplast.

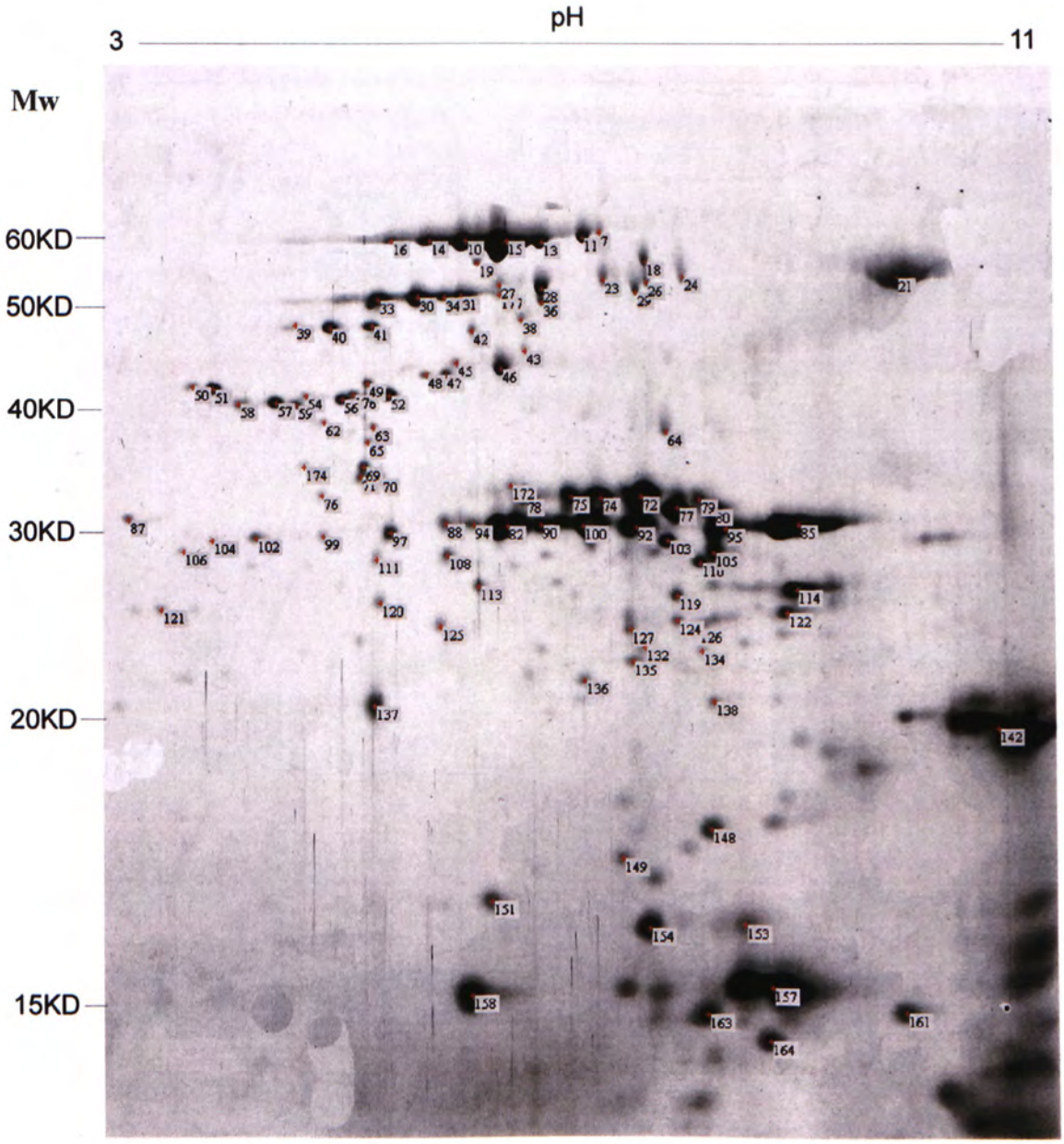


Figure 3.3A 2D SDS-PAGE of rice (9311) seed amyloplast proteins at pH3-11NL range

Amyloplast proteins were extracted from rice seeds of line 9311 at 10DAF. Sample load: 50 μg of proteins; first dimension: pH3-11NL, 18cm IPG strip; Second dimension: SDS-PAGE (12.5%); staining method: silver stain. The number on each spot was automatically labeled by the ImageMaster 2D Platinum, Version 5.00 (Amersham Biosciences, Uppsala, Sweden).

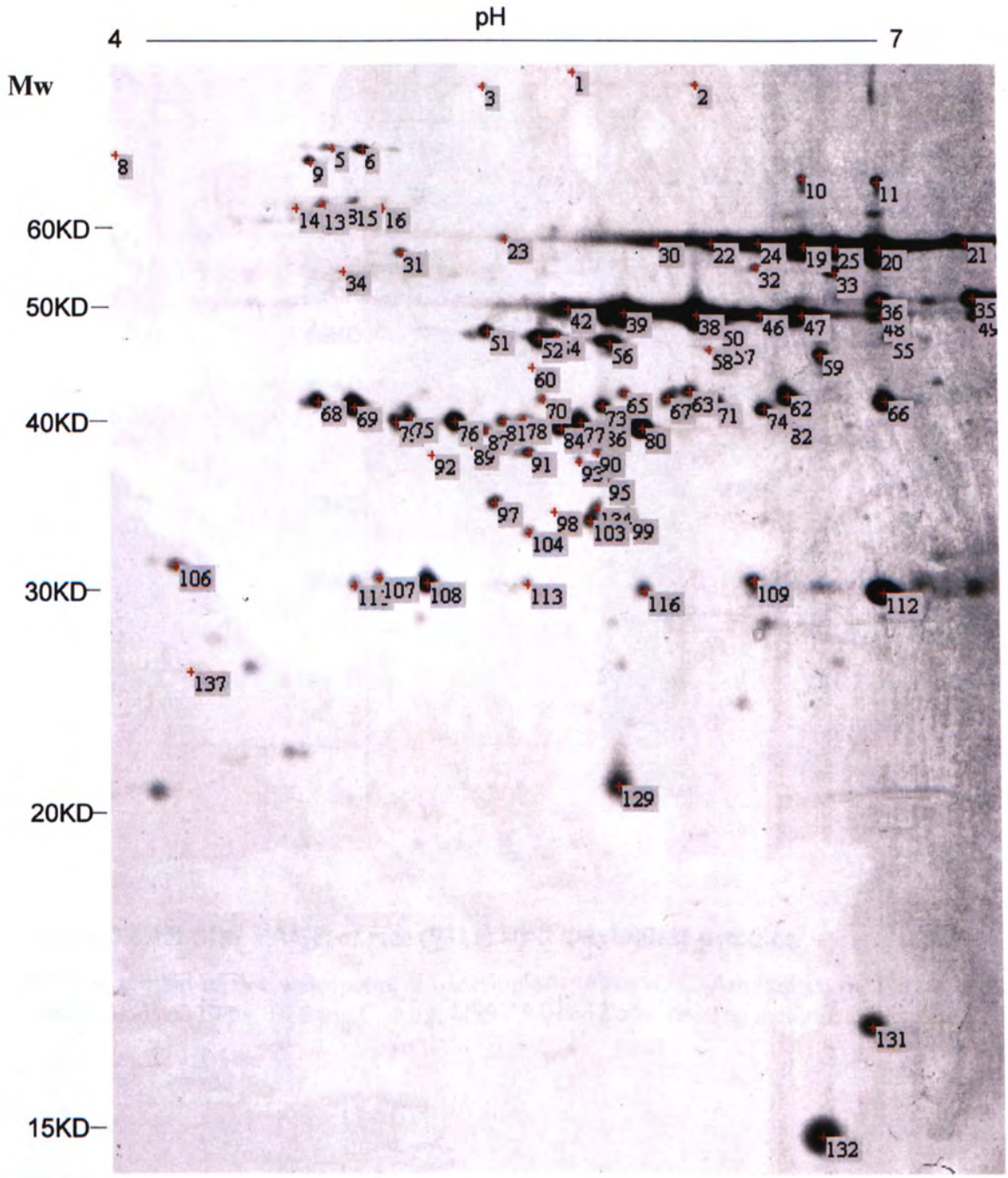


Figure 3.3B 2D SDS-PAGE of rice (9311) seed amyloplast proteins at pH4-7 range

Amyloplast proteins were extracted from rice seeds of line 9311 at 10DAF. Sample load: 50 μg of proteins; first dimension: pH4-7, 18cm IPG strip; Second dimension: SDS-PAGE (12.5%); staining method: silver stain. The number on each spot was automatically labeled by the ImageMaster 2D Platinum, Version 5.00 (Amersham Biosciences, Uppsala, Sweden).

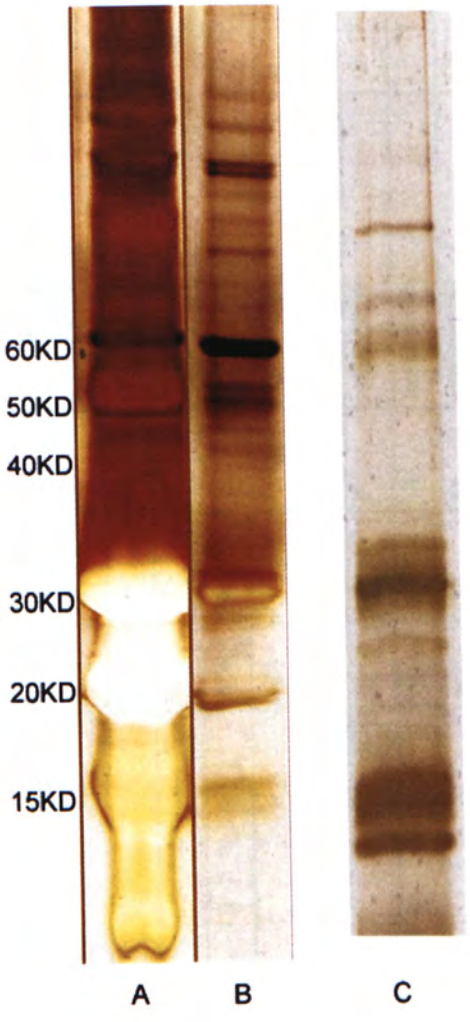


Figure 3.4 1D SDS-PAGE of rice (9311) seed amyloplast proteins

A. Total protein of rice endosperm; B. Amyloplasts proteins; C. Amyloplast membrane proteins. Sample load: A. 10 μg; B. 5 μg; C. 1 μg; SDS-PAGE: 12.5%; staining method: silver stain.

Table 3.2 Summary of protein identified by 2D pH3-11NL SDS-PAGE

Spot ID	Cal Mw (kDa)	Cal pI (pH)	Exp Mw (kDa)	Exp pI (pH)	Accession number	Protein Name	Cover age	Protein Score	Peptide Count	Functional classification	Remarks
7	66.424	8.34	58	6.74	gi|19911776	GBSS	22%	120	7	metabolism	TP: 77
10	66.281	8.34	58	5.73	gi|19911776	GBSS	15%	87	5	metabolism	TP: 77
11	66.358	8.34	58	6.60	gi|19911776	GBSS	26%	184	11	metabolism	TP: 77
13	66.281	8.34	58	6.28	gi|19911776	GBSS	16%	140	7	metabolism	TP: 77
14	66.424	8.34	58	5.40	gi|19911776	GBSS	23%	260	12	metabolism	TP: 77
15	66.424	8.34	58	6.10	gi|19911776	GBSS	44%	108	18	metabolism	TP: 77
16	66.424	8.34	58	5.05	gi|19911776	GBSS	30%	132	15	metabolism	TP: 77
18	66.424	8.34	57	7.10	gi|19911776	GBSS	30%	150	15	metabolism	TP: 77
19	48.920	6.46	57	5.84	gi|29367525	fructose-6-phosphate 1-phosphotransferase	15%	97	12	metabolism	TP: 33
21	56.247	9.09	56	9.10	gi|50902034	glutelin type I precursor	15%	99	5	protein destination	SP: 24
23	66.389	8.16	56	6.77	gi|19911776	GBSS	26%	74	11	metabolism	TP: 77
24	116.382	6.73	56	7.40	gi|988270	sucrose phosphate synthase	15%	89	4	metabolism	SP: 36
26	66.281	8.34	55	7.11	gi|19911776	GBSS	18%	140	6	metabolism	TP: 77
27	66.434	8.34	55	6.03	gi|19911776	GBSS	28%	130	14	metabolism	TP: 77
28	66.348	8.36	55	6.27	gi|19911776	GBSS	26%	120	13	metabolism	TP: 77
29	66.358	8.34	54	7.03	gi|19911776	GBSS	15%	109	8	metabolism	TP: 77
30	66.389	8.16	54	5.30	gi|19911776	GBSS	29%	248	12	metabolism	TP: 77
31	66.358	8.34	54	5.69	gi|19911776	GBSS	15%	133	8	metabolism	TP: 77
33	66.358	8.34	53	4.93	gi|19911776	GBSS	20%	82	7	metabolism	TP: 77
34	66.424	8.34	54	5.54	gi|19911776	GBSS	18%	94	5	metabolism	TP: 77
36	66.348	8.36	53	6.28	gi|19911776	GBSS	27%	134	14	metabolism	TP: 77
38	56.254	9.48	51	6.12	9629.m02468	Similar to probable thiamin pyrophosphokinase At2g44750 (imported) - Arabidopsis thaliana	21%	97	13	energy	

Spot ID	Cal Mw (kDa)	Cal pI (pH)	Exp Mw	Exp pI (pH)	Accession number	Protein Name	Cover age	Protein Score	Peptide Count	Functional classification	Remarks
39	207.109	9.27	51	4.28	gi|34909624	Putative glucan synthase	16%	89	20	metabolism	TP: 8
40	66.348	8.36	50	4.55	gi|19911776	GBSS	29%	126	11	metabolism	TP: 77
41	66.281	8.34	50	4.88	gi|19911776	GBSS	20%	325	8	metabolism	TP: 77
42	45.009	7.98	50	5.79	9634.m04530	galactasyl transferase, putative	15%	90	4	metabolism	
45	66.424	8.34	47	5.65	gi|19911776	GBSS	30%	111	12	metabolism	TP: 77
46	66.358	8.34	46	6.05	gi|19911776	GBSS	24%	125	14	metabolism	TP: 77
47	66.348	8.36	45	5.55	gi|19911776	GBSS	28%	134	15	metabolism	TP: 77
48	66.348	8.36	45	5.37	gi|19911776	GBSS	21%	90	11	metabolism	TP: 77
49	116.382	6.73	44	4.84	gi|988270	sucrose phosphate synthase	15%	90	4	metabolism	TP: 36
50	66.348	8.36	44	3.49	gi|19911776	GBSS	15%	106	8	metabolism	TP: 77
51	66.348	8.36	44	3.65	gi|19911776	GBSS	22%	132	12	metabolism	TP: 77
52	66.348	8.36	43	5.03	gi|19911776	GBSS	31%	100	17	metabolism	TP: 77
54	60.096	7.52	43	4.37	gi|13129489	putative sugar transporter	15%	99	8	transport facilitation	SP: 21
56	66.424	8.34	43	4.65	gi|19911776	GBSS	24%	140	14	metabolism	TP: 77
57	66.348	8.36	43	4.13	gi|19911776	GBSS	24%	130	13	metabolism	TP: 77
58	62.867	10.21	42	3.85	gi|21672079	Putative ribosomal protein S23 (S12)	23%	79	14	protein synthesis	TP: 58
59	153.845	8.66	42	4.29	gi|50920201	putative gag-pol precursor	14%	43	9	retrotransposons and plasmid proteins	TP: 62
62	68.627	9.35	41	4.50	gi|33242909	transmembrane kinase	14%	53	11	cellular communication	TP: 8
63	74.510	5.58	40	4.89	gi|34901134	putative probable transcription repressor HOTR	15%	97	4	transcription	
64	43.317	8.99	40	7.26	gi|49388082	putative glyceraldehydes 3-phosphate	15%	120	4	energy	TP: 57
65	170.335	8.4	39	4.84	gi|37536652	putative gag-pol precursor	14%	67	9	retrotransposons and plasmid proteins	TP: 62
69	19.135	6.75	37	4.81	9629.m02036	dTDP-glucose-4,6-deyhdratase like protein	15%	86	4	metabolism	
70	66.348	8.36	36	4.96	gi|19911776	GBSS	20%	98	8	metabolism	TP: 77

Spot ID	Cal Mw (kDa)	Cal pI (pH)	Exp Mw (kDa)	Exp pI (pH)	Accession number	Protein Name	Cover age	Protein Score	Peptide Count	Functional classification	Remarks
71	58.417	8.64	36	4.79	gi|2625086	ADP-glucose pyrophosphorylase large subunit (Cucumis melo]	14%	98	4	metabolism	TP: 33
72	56.190	9.32	35	7.07	gi|20212	glutelin	15%	88	4	protein destination	acidic subunit
74	55.820	8.81	35	6.75	gi|100680	glutelin	24%	89	8	protein destination	acidic subunit
75	56.640	9.33	35	6.50	gi|20210	glutelin	16%	101	4	protein destination	acidic subunit
76	174.723	4.99	35	4.49	gi|21429218	putative soluble starch synthase III-2	15%	104	4	metabolism	TP: 18
77	56.190	9.32	34	7.35	gi|20212	glutelin	26%	89	9	protein destination	acidic subunit
78	56.157	8.93	35	6.14	gi|225710	glutelin	15%	70	5	protein destination	acidic subunit
79	49.769	8.74	35	7.53	gi|100681	glutelin	15%	70	4	protein destination	acidic subunit
80	57.084	8.96	34	7.64	gi|50907885	glutelin	24%	75	8	protein destination	acidic subunit
82	55.820	8.81	33	6.00	gi|100680	glutelin	20%	133	8	protein destination	acidic subunit
85	56.190	9.32	33	8.31	gi|20212	glutelin	24%	139	14	protein destination	acidic subunit
87	66.348	8.36	33	3.00	gi|19911776	GBSS	15%	108	8	metabolism	TP: 77
88	56.157	8.93	33	5.54	gi|225710	glutelin	15%	88	4	protein destination	acidic subunit
90	56.640	9.33	33	6.26	gi|20210	glutelin	22%	104	8	protein destination	acidic subunit
92	56.190	9.32	33	7.02	gi|20212	glutelin	16%	135	4	protein destination	acidic subunit
94	56.190	9.32	33	5.79	gi|20212	glutelin	15%	88	4	protein destination	acidic subunit
95	56.640	9.33	33	7.73	gi|20210	glutelin	16%	101	4	protein destination	acidic subunit
97	26.465	11.81	32	5.04	gi|50933531	Hypothetical protein	30%	87	6	Hypothetical protein	TP: 85
99	68.704	6.40	32	4.50	9634.m02599	1,4-alpha-glucan branching enzyme,putative	18%	99	8	metabolism	
100	56.190	9.32	33	6.60	gi|20212	glutelin	20%	121	8	protein destination	acidic subunit
102	66.348	8.36	32	3.98	gi|19911776	GBSS	24%	156	8	metabolism	TP: 77
103	54.825	9.09	32	7.28	9630.m02963	Hypothetical protein	15%	90	10	Hypothetical protein	
104	66.348	8.36	32	3.64	gi|19911776	GBSS	19%	179	11	metabolism	TP: 77

Spot ID	Cal Mw (kDa)	Cal pI (pH)	Exp Mw	Exp pI (pH)	Accession number	Protein Name	Cover age	Protein Score	Peptide Count	Functional classification	Remarks
105	56.217	9.17	31	7.64	gi|225737	glutelin	18%	121	8	protein destination	acidic subunit
106	68.627	9.35	31	3.41	gi|34901134	putative probable transcription repressor HOTR	26%	58	15	transcription	
108	40.043	11.08	31	5.56	gi|34895398	P0445D12.6	19%	58	7	Hypothetical protein	TP: 14
110	56.640	9.33	30	7.53	gi|20210	glutelin	22%	104	8	protein destination	acidic subunit
111	33.152	11.14	30	4.91	gi|37532874	Hypothetical protein	16%	89	5	Hypothetical protein	TP: 54
113	38.666	11.85	28	5.84	gi|34914578	P0455H0310.22	16%	80	5	Hypothetical protein	TP: 58
114	56.217	9.17	28	8.29	gi|225737	glutelin	18%	90	8	protein destination	acidic subunit
119	57.571	5.54	28	7.35	gi|13540812	ADP-glucose pyrophosphorylase large subunit isoform	20%	90	4	metabolism	TP: 18
120	14.775	9.94	27	4.94	gi|52077170	ubiquitin / ribosomal protein CEP52	30%	53	4	protein synthesis	TP: 14
121	21.989	5.59	27	3.25	gi|553108	UDP-glucose starch glycosyl transferase	22%	97	4	metabolism	TP: 23
122	49.769	8.74	27	8.21	gi|100681	glutelin	17%	75	4	protein destination	acidic subunit
124	33.785	11.39	26	7.35	gi|50917087	hypothetical protein	15%	54	6	Hypothetical protein	SP: 32
125	20.102	5.35	26	5.49	gi|18149173	starch branching enzyme (Phaseolus vulgaris)	27%	141	4	metabolism	TP: 47
126	31.279	10.21	25	7.51	gi|34906126	P0020E09.23	15%	47	8	Hypothetical protein	TP: 62
127	66.358	8.34	25	6.97	gi|19911776	GBSS	18%	87	7	metabolism	TP: 77
132	66.281	8.34	24	7.09	gi|19911776	GBSS	18%	83	7	metabolism	TP: 77
134	49.769	8.74	24	7.55	gi|100681	glutelin 3 precursor - rice	17%	75	4	protein destination	basic subunit
135	66.358	8.34	23	7.00	gi|19911776	GBSS	25%	84	14	metabolism	TP: 77
136	206.780	9.02	22	6.61	gi|7270678	putative glucan synthase component (Arabidopsis thaliana)	17%	88	9	metabolism	TP: 16
137	21.041	7.48	21	4.89	gi|51038053	globulin	30%	104	4	protein destination	SP: 22
138	57.084	8.96	21	7.64	gi|50907885	glutelin	18%	74	5	protein destination	basic subunit
142	56.190	9.32	21	9.89	gi|20212	glutelin	29%	174	10	protein destination	basic subunit
148	20.351	6.84	19	7.61	gi|34912758	P0407B12.16	15%	89	7	energy	SP: 19

Spot ID	Cal Mw (kDa)	Cal pI (pH)	Exp Mw (kDa)	Exp pI (pH)	Accession number	Protein Name	Cover age	Protein Score	Peptide Count	Functional classification	Remarks
149	174.833	4.99	18	6.91	gi|21429218	putative soluble starch synthase III-2	29%	78	7	metabolism	TP: 18
151	19.064	9.03	17	5.95	gi|347451	ribulose 1,5-bisphosphate carboxylase	23%	120	5	energy	TP: 40
153	17.267	8.36	17	7.87	gi|34900132	allergenic RA5B precursor	35%	150	4	Hypothetical protein	TP: 55
154	43.227	11.16	17	7.12	gi|20171	alpha-amylase	22%	70	10	metabolism	SP: 26
157	16.859	8.79	16	8.08	gi|6180118	prolamin	48%	366	4	protein destination	SP: 19
158	16.985	7.90	16	5.76	gi|34900248	prolamin	20%	230	4	protein destination	SP: 19
161	16.762	8.79	16	9.13	gi|82505	prolamin	20%	208	4	protein destination	SP: 19
163	16.918	9.17	16	7.57	gi|20304	prolamin	36%	225	4	protein destination	SP: 19
164	16.762	8.79	15	8.06	gi|82505	prolamin	48%	141	4	protein destination	SP: 19
172			36	6.03	9x20-36	-KANCLKVDNK/-PAHCKNK				Hypothetical protein	
174	207.109	9.27	37	4.35	gi|34909624	putative glucan synthase	15%	88	4	metabolism	TP: 8
176	66.389	8.16	43	4.73	gi|19911776	GBSS	22%	90	12	metabolism	TP: 77
177	66.358	8.34	54	6.04	gi|19911776	GBSS	32%	85	17	metabolism	TP: 77

Remarks: TP = transit peptide; SP = signal peptide

Table 3.3 Summary of protein identified by 2D pH4-7 SDS-PAGE

Spot ID	Cal Mw (kDa)	Cal pI (pH)	Exp Mw (kDa)	Exp pI (pH)	Accession number	Protein Name	Cover age	Protein score	Peptide count	Functional classification	Remarks
1	91.434	7.36	87	5.19	gi|12025466	a-1,4-glucan phophorylase H	16%	100	4	metabolism	TP: 25
2	100.655	6.95	84	5.56	gi|34907810	P0703B11.26	15%	99	4	hypothetical protein	TP: 17
3	102.736	5.98	84	4.92	gi|7434944	PPDK precursor	16%	109	5	energy	TP: 34
5	73.495	5.3	73	4.47	gi|2267006	endosperm lumenal binding protein	22%	202	12	transport facilitation	TP: 24 SP: 25
6	69.352	5.57	73	4.56	gi|15384987	soluble starch synthase 2	23%	168	7	metabolism	TP: 15
8	98.317	5.92	72	3.82	gi|25008209	Phosphoenolpyruvate carboxylase (PEPCase) (PEPC)	15%	101	8	metabolism	
9	91.490	5.59	71	4.40	gi|18461259	putative beta-galactosidase	17%	89	13	metabolism	SP: 24
10	92.071	5.96	68	5.87	gi|20366	sucrose synthase	15%	106	9	metabolism	TP: 7
11	77.768	6.18	68	6.10	gi|34896390	putative ABC transporter	15%	89	9	transport facilitation	TP: 41
13	56.196	9.09	63	4.44	gi|225174	glutelin precursor	24%	106	11	protein destination	SP: 24
14	130.854	9.13	64	4.36	gi|8096424	Similar to Arabidopsis thaliana chromosome2 BAC F13B15; putative non-LTR retroelement	30%	87	23	retrotransposons and plasmid proteins	
15	51.888	5.84	64	4.53	gi|22535568	putative glycosyltransferase	20%	98	4	metabolism	SP: 37
16	75.098	5.26	64	4.62	gi|37531722	hypothetical protein	15%	68	8	hypothetical protein	TP: 8
23	66.358	8.34	60	4.98	gi|12025466	GBSS	30%	134	15	metabolism	TP: 77
25	66.358	8.34	58	5.97	gi|19911776	GBSS	40%	220	17	metabolism	TP: 77
31	94.411	6.82	58	4.68	gi|20805185	alpha 1,4-glucan phosphorylase H isozyme	15%	99	8	metabolism	TP: 10
32	55.856	8.96	56	5.73	9631.m00680	sucrose/H+symporter	22%	120	9	transport facilitation	
34	54.845	6.23	56	4.50	gi|50725517	AGPase ss	16%	90	8	metabolism	TP: 50
42	66.358	8.34	53	5.17	gi|19911776	GBSS	20%	108	12	metabolism	TP: 77
50	57.021	11.7	51	5.63	gi|34906494	P0432C03.4	15%	90	15	hypothetical protein	TP: 7
54	66.348	8.36	50	5.14	gi|19911776	GBSS	29%	130	14	metabolism	TP: 77

Spot ID	Cal Mw (kDa)	Cal pI (pH)	Exp Mw (kDa)	Exp pI (pH)	Accession number	Protein Name	Cover age	Protein score	Peptide count	Functional classification	Remarks
57	38.335	11.96	50	5.66	gi|34913538	OJ1029_F04.20	28%	79	12	hypothetical protein	TP: 72
58	46.122	9.74	49	5.60	gi|34909646	putative beta-amylase	15%	99	8	metabolism	TP: 55
60	47.103	6.95	48	5.07	gi|14140154	DnaJ-like protein	20%	99	15	protein destination	TP: 20
63	37.527	11.39	46	5.54	gi|50939039	putative glucanase	15%	87	12	metabolism	TP: 49
65	102.736	5.98	46	5.34	gi|7434944	pyruvate, phosphate dikinase (EC 2.7.9.1)	15%	102	7	energy	TP: 34
67	51.638	5.46	45	5.47	gi|7417426	UDP-glucose pyrophosphorylase	25%	167	8	metabolism	TP: 20
70	21.989	5.59	45	5.10	gi|553108	UDP glucosyltransferase	45%	143	7	metabolism	TP: 23
75	43.986	10.06	44	4.70	9639.m02229	transposon protein, putative, mariner sub-class	30%	105	7	retrotransposons and plasmid proteins	
78	63.268	8.69	44	5.04	gi|50918167	chloroplast protein-translocon-like protein	15%	89	8	transport facilitation	TP: 21; SP:
82	45.102	8.43	44	5.83	gi|476752	rubisco ls	21%	97	10	energy	TP: 8
86	43.437	11.52	43	5.27	gi|34907220	P0504E02.4	14%	68	9	hypothetical protein	TP: 14
89	45.476	6.27	42	4.88	gi|18642697	Unknown protein	15%	98	4	hypothetical protein	TP: 16
92	41.056	9.20	41	4.76	gi|5091602	10A19I.6	15%	90	4	hypothetical protein	TP: 17
93	66.348	8.36	41	5.21	gi|19911776	GBSS	22%	109	12	metabolism	TP: 77
98	34.909	5.09	38	5.13	gi|4097944	beta-1,3-glucanse precursor	15%	78	7	metabolism	SP: 25
138	66.791	8.92	64	4.44	gi|37537160	unknown protein	25%	89	4	hypothetical protein	

Remarks: TP = transit peptide; SP = signal peptide

Table 3.4 Summary of protein identified by 1D SDS-PAGE

Spot ID	Cal Mw (kDa)	Cal pI (pH)	Exp Mw	Exp pI (pH)	Accession number	Protein Name	Cover age	Protein score	Peptide count	Functional classification	Remarks
04b	93.236	6.35	98		gi|4704818	starch-branching enzyme I	20%	150	9	metabolism	TP: 57
05b	38.565	7.95	37.5		gi|34909322	putative diphosphonucleotide phosphatase	15%	80	6	metabolism	TP: 16
07a	64.840	8.52	64		gi|41469080	putative glucose-6-phosphate dehydrogenase	25%	79	10	metabolism	TP: 46
13a	51.670	5.46	53		gi|7417426	UDP-glucose pyrophosphorylase	21%	89	8	metabolism	TP: 20
13b	15.294	9.84	17		gi|34896292	putative beta-glucosidase-aggregating factor	17%	78	6	metabolism	TP: 71
15a	59.146	6.31	60		gi|37536926	putative chloroplast-targeted beta-amylase	27%	75	10	metabolism	TP: 45
17a	52.932	9.16	51.5		gi|34909216	putative beta-glucosidase	18%	94	8	metabolism	SP: 22
01b	88.375	5.28	81		gi|51535402	Soluble starch synthase II	28%	120	15	metabolism	
02b	92.757	5.69	85		gi|50909141	Branching enzyme-3	29%	156	12	metabolism	TP: 65
05a-1	143.496	6.12	64		gi|20303616	putative reverse transcriptase	14%	78	16	retrotransposons and plasmid proteins	
08a	167.126	8.99	61		9630.m04799	retrotransposon protein, putative, Ty1-copia sub-class	14%	89	18	retrotransposons and plasmid proteins	
04a	11.642	11.74	13		gi|50943101	hypothetical protein	22%	94	6	hypothetical protein	
07a	21.144	9.26	21		gi|29027773	OJ1710_H11.19	23%	90	4	hypothetical protein	TP: 30
17a	29.715	11.27	34		gi|21902024	B1097D05.25	23%	82	10	hypothetical protein	TP: 50
31b	56.565	9.26	59		gi|20232	unnamed protein product	19%	96	6	hypothetical protein	TP: 60; SP: 24
32b	56.565	9.26	58.5		gi|20232	unnamed protein product	19%	99	7	hypothetical protein	TP: 60; SP: 24
31b	56.565	9.26	58		gi|20232	unnamed protein product	19%	80	5	hypothetical protein	TP: 60; SP: 24
04b	66.749	9.17	66		gi|34912792	P0501G01.5	20%	78	8	hypothetical protein	TP: 57
124	25.999	9.63	27		9639.m03827	ribosomal protein	24%	90	8	protein synthesis	
295	25.999	9.63	26		9639.m03827	ribosomal protein	25%	95	9	protein synthesis	
16	136.879	8.8	120		gi|34902102	chloroplast import-associated channel protein	27%	95	13	transport facilitation	TP: 31

Spot ID	Cal Mw (kDa)	Cal pI (pH)	Exp Mw	Exp pI (pH)	Accession number	Protein Name	Cover age	Protein score	Peptide count	Functional classification	Remarks
m19	73.126	8.21	77		9636.m02962	Leucine Rich Repeat, putative	15%	102	14	protein destination	
m1	56.449	9.36	59		gi|19033566	putative potassium channel regulatory factor	18%	80	10	transport facilitation	TP: 62
m2	39.444	6.56	43		gi|34897162	putative membrane protein	15%	91	9	transport facilitation	TP: 33
17a	39.153	5.88	35		gi|15451567	Putative phosphoenolpyruvate kinase	15%	97	12	energy	

Remarks: TP = transit peptide; SP = signal peptide

3.1.3 Functional classification of identified proteins

To understand the function of the amyloplast proteins, the identified proteins were sorted into 9 different categories according to Koller *et al.* (2002) as shown in Fig 3.5. The categories cover: hypothetical protein/unknown (14.5%), protein destination (19.3%), retrotransposon and plastid protein (3.6%), protein synthesis (2.4%), cellular communication (0.6%), metabolism (48.8%), energy (4.8%), transport facilitation (4.8%) and transcription (1.2%). Of the 166 proteins identified by MS/MS, the most abundant category was classified as being involved in metabolism (48.8%). Storage proteins such as glutelin and prolamin were also identified in the amyloplast fraction. They can be considered as the proteins strongly specifically associated with amyloplasts. Proteins involved in other cellular activities were also detected in the amyloplast proteome.

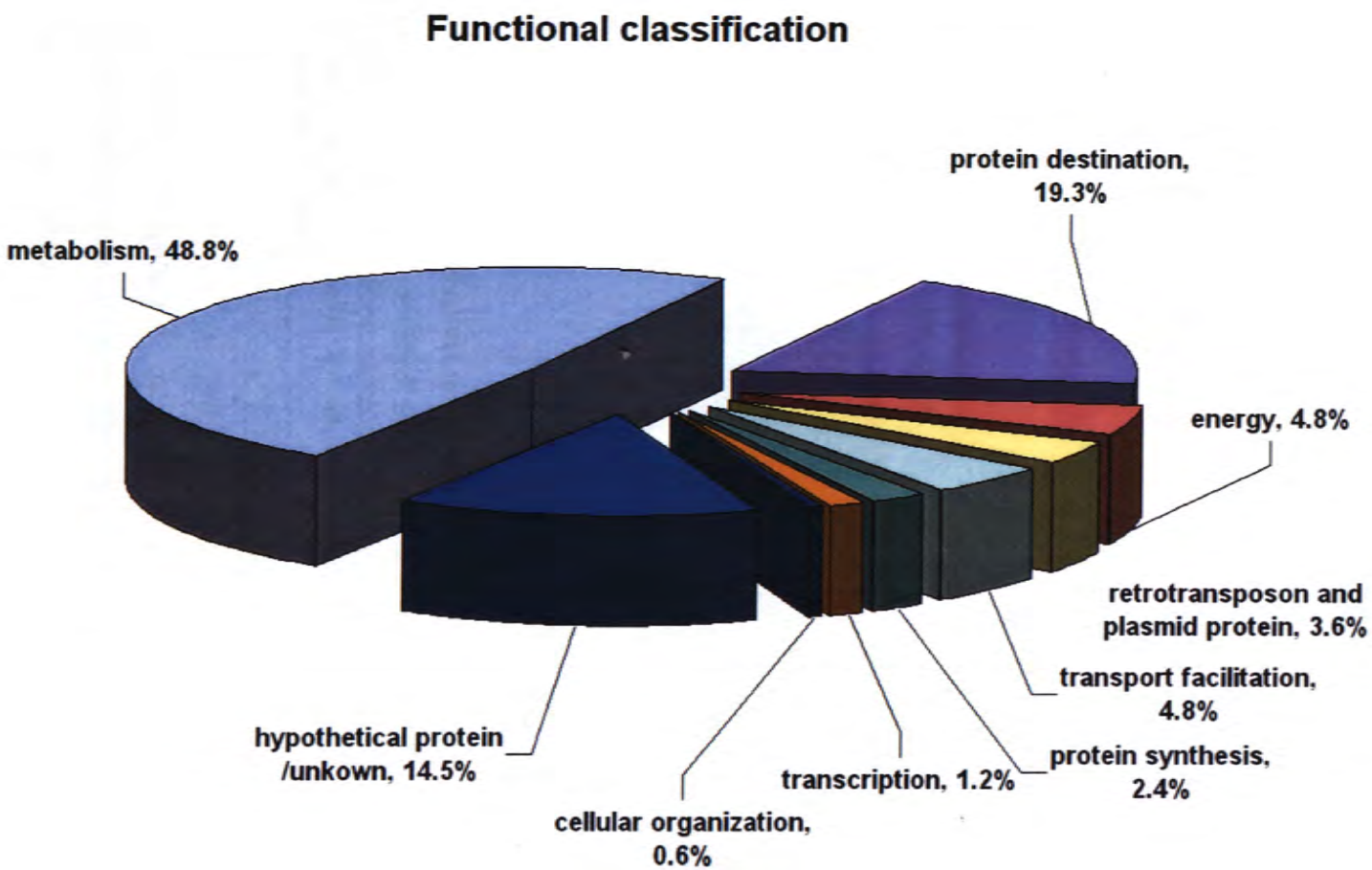


Figure 3.5 Functional classifications of rice amyloplasts proteins

The 166 of rice seed amyloplast proteins of line 9311 were grouped into 9 categories, including hypothetical protein/unknown (14.5%), protein destination (19.3%), retrotransposon and plastid protein (3.6%), protein synthesis (2.4%), cellular communication (0.6%), metabolism (48.8%), energy (4.8%), transport facilitation (4.8%) and transcription (1.2%).

3.1.3.1 Metabolism proteins

As indicated in Fig 3.5, a variety of proteins associated with metabolism were identified. Since starch is synthesized in the rice endosperm amyloplasts, the proteome of amyloplasts should be highly correlated with starch synthesis. Several other function proteins indicate that they may also involve in facilitation of starch synthesis. Indeed 79% of the metabolism proteins are starch synthesis related proteins (Fig 3.6).

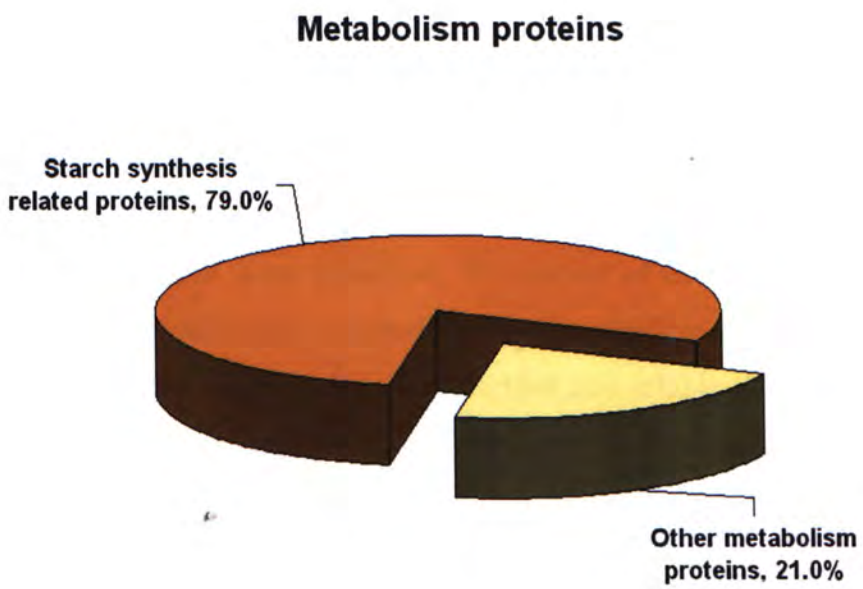


Figure 3.6 Rice amyloplast metabolism proteins

Metabolism proteins identified from amyloplasts were further divided into two categories according to their functions: starch synthesis related proteins and other metabolism proteins. 79% of the metabolism proteins are starch synthesis related proteins and 21% are related to other metabolism functions.

For starch biosynthesis, most attention has been focused on the ADP-glucose pyrophosphorylase (AGPase) since the first step in starch biosynthetic pathway is the conversion of glucose-1-phosphate into ADP-glucose, through the action of this enzyme. The ADP-glucose then serves as a glucosyl donor for glucan synthesis by starch synthases, starch branching and debranching enzymes (Slattery *et al.*, 2000; Beckles *et al.*, 2001) AGPase is a homotetramer in bacteria; while in plants, it is a heterotetramer of two different, yet evolutionary related subunits (reviewed by (Sivak *et al.*, 1993). The two small subunits are responsible for catalytic property and the two large subunits for

regulatory property (Smith-White *et al.*, 1992; Fu *et al.*, 1998). The AGPase was detected in our analysis, with two protein spots for the large subunit (spots 119 and 71 in the pH3-11NL gel) and one spot of the small subunit (spot 34 in the pH4-7 gel). Relative apparent molecular masses of the large subunits on the 2D gel (spot 119, 28 kDa and spot 71, 36kDa) are not consistent with the theoretical values (57–58 kDa for the large subunit). These differences may be caused by the fragmentation of the large subunit. The use of subunit-specific antibodies should help us to confirm the presence and relatedness of the endosperm subunit isoforms.

Starch synthases (SSs) utilize ADPG in its elongation of the linear chains by α-1,4 linkages formation. Cereal endosperm contains at least five SS isoforms that are categorized according to their conserved sequence relationships. Four isoforms, termed SSI, SSIIa, SSIIb, and SSIII, are believed to have unique functions in amylopectin synthesis, although their precise roles have not been identified. Elucidation of the roles of the soluble SS isoforms at the protein level thus is essential to our understanding of amylopectin synthesis. Among the starch synthesizing enzymes, we identified one soluble starch synthase II (spot 6 in the pH4-7 gel) and two isoforms (spots 76 and 149 in the pH3-11 NL gel) of soluble starch synthase III-2.

Starch branching enzymes (SBEs) generate α-1,6 linkages by cleaving internal α-1,4 bonds and transferring the released reducing ends to the C6 hydroxyls. There are two classes of branching enzyme (BEI and BEII) that differ in the lengths of the transferred chains *in vitro*, with BEII transferring shorter chains than BEI (Guan *et al.*, 1993; Takeda *et al.*, 1993). In cereals, there are two closely related forms of BEII (BEIIa and BEIIb) (Mizuno *et al.*, 2001; Rahman *et al.*, 2001). We identified 2 spots (spots 99 and 125 in the pH4-7 gel) of starch branching enzyme and 2 bands in 1D gel. Our results are consistent with the results that multiple forms of starch branching enzyme exist and they generate different lengths of branched chains or branch points at different frequencies.

Two proteins of starch phosphorylase enzyme were identified among the excised spots (spots 1 and 31 in pH4-7 gel). Enzymes for the starch degradation pathway include debranching enzymes, disproportionating enzymes, isoamylases, and α-glucosidases (Kossmann *et al.*, 2000). But we only identified the α-amylases and β-amylases. The presence of several starch biosynthetic enzymes and the absence of most catabolic

enzymes among the identified protein spots might simply be explained by the fact that starch is just beginning to be synthesized at 10DAF stage.

During protein identification, we observed several isoforms of starch synthesis related enzymes which may play different functional roles. In previous studies (Zhu *et al.*, 2003), researchers found that most genes in the starch biosynthetic pathway show multiple distinct spatial and temporal expression patterns, suggesting that different isoforms of a given enzyme may be expressed in different tissues and at different developmental stages.

3.1.3.2 Non starch synthesis metabolism proteins

In the amyloplast fraction, we observed several proteins involved in the sucrose to starch synthesis pathway with high coordination, such as sucrose synthase, UDP-glucose pyrophosphorylase, and sucrose phosphate synthase. For sucrose synthase and sucrose phosphate synthase, we observed two different isoforms of each. We have also observed one putative sugar transporter (spot 54 in the pH3-11 NL gel) and one sucrose/H^+ symporter (spots 32 in the pH4-7 gel). For sugar signal mediated starch synthesis, our results may provide some insights into the mechanism of sugar signaling in plant.

3.1.3.3 Protein destination

Storage proteins are one of the major storage compounds in rice endosperm. On the basis of their extraction and solubility, storage proteins in rice seeds can be classified into four types including: diluted acidic- or alkaline-soluble glutelin, salt-soluble globulin, alcohol-soluble prolamin and water-soluble albumin (Juliano, 1982). Rice seed storage proteins possess some characteristics as those of other cereals. First, they are expressed at high level in specific tissues and at certain developmental stages and regulated by nutrients. A second common property of them is that in mature seeds they all are deposited discretely in specialized membrane-bounded storage organelles called protein bodies. Finally, each storage protein fraction is a mixture of many polypeptides which exhibit polymorphisms arising from multiple gene families and, in some cases, proteolytic processing and glycosylation (Shewry *et al.*, 1995).

We tried washing the starch granules with washing buffer eight times, but still detected the storage proteins especially glutelin in all the amyloplast preparations, indicating that a small proportion of storage proteins is tightly bound to the starch granules. This finding agrees well with the notion that storage proteins are usually tightly but peripherally associated with the starch granules (Mu-Forster *et al.*, 1998). Similar results were also obtained for GBSS protein in the barley starch proteome (Boren *et al.*, 2004) and in the wheat amyloplast proteome (Andon *et al.*, 2002). Although the presence of storage proteins in amyloplast fraction may be caused by artifacts of subcellular fractionation procedure, their association might also be biologically significant. But in our analysis, we are most interested in those proteins that fall outside this category.

3.1.3.4 Proteins with other functions

Several proteins were identified as unknown functions although their MS/MS spectrums were quite good. For example, allergenic protein showed significant homology to the α-amylase/trypsin inhibitor family from wheat and barley and it was isolated with reactivity for IgE of rice-allergic patients (Tada *et al.*, 1996). For these proteins with unknown functions they may also have the potential to facilitate starch synthesis.

3.1.4 Cross-correlation of experimental and calculated Mw of proteins

When searching the database with PMF or MS/MS data, it is possible to use the pI and Mw information of the proteins from the 2D gels to reduce the number of candidates searched in the database. We found that the experimental Mw of several proteins did not match with the calculated Mw when searching the database. We cross-correlated the calculated and experimental values for Mw and found that the experimental Mw of most proteins were lower than or correspond to the calculated Mw (Fig. 3.7).

We use the ChloroP and SignalP software to predict each protein to see whether it contains any transit peptide or signal peptide (Table 3.2, 3.3 and 3.4). Of the 166 proteins, 106 (63.9%) were predicted to have transit peptides, 44 (26.5%) with signal peptides and 21 (12.6%) without transit peptides and signal peptides. Five proteins were predicted to contain both transit peptides and signal peptides. This may partially explained why the experimental Mw of amyloplast proteins is lower than the calculated precursor proteins. Of

course an incorrect prediction of plastid proteins can be caused by the presence of non-canonical transit sequences (reviewed by Schleiff *et al.*, 2001; Miras *et al.*, 2002). Dual targeting of proteins to plastids and other sub-cellular compartments has been reported for several proteins, further compounding the problem in predicting plastid targeting (Silva-Filho, 2003; Kiessling *et al.*, 2004; Wall *et al.*, 2004); review by (Peeters *et al.*, 2001).

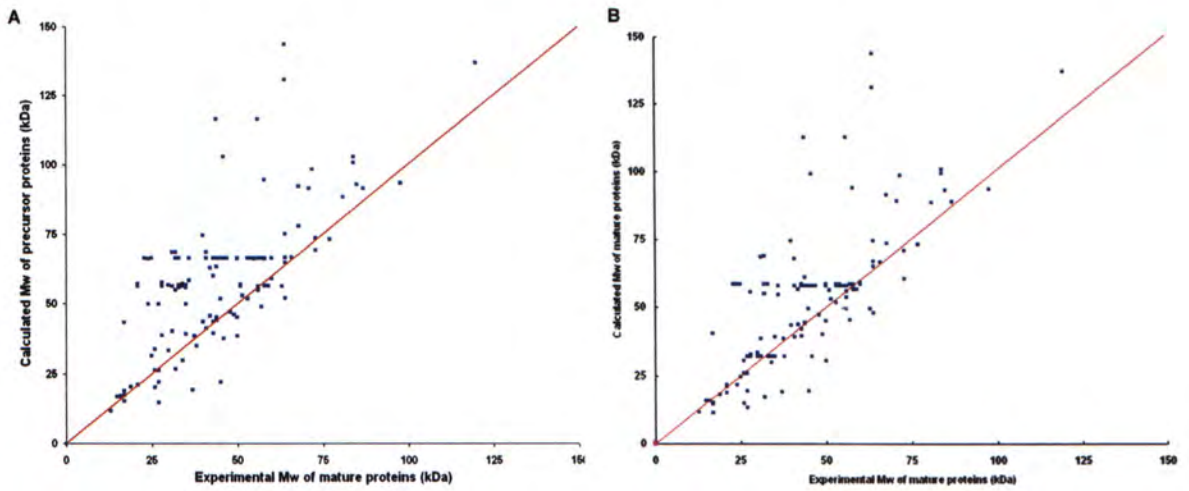


Figure 3.7 Cross-Correlation between experimental and calculated Mw of precursors and mature proteins in rice amyloplast proteome

Cross-correlations between experimental Mw of mature proteins and calculated Mw of precursor proteins (A) of rice seed amyloplasts and experimental Mw of mature proteins and calculated Mw of mature proteins (B) are compared. The red lines indicate the regression lines.

3.1.5 Granule bound starch synthase (GBSS)

After protein identification, it is quite interesting to find that there are multiple fragments (a total of 43 spots) of GBSS with different molecular weights from 60kDa to 28kDa and with different pI values. Western blot results further confirmed the MS/MS results (Fig 3.8).

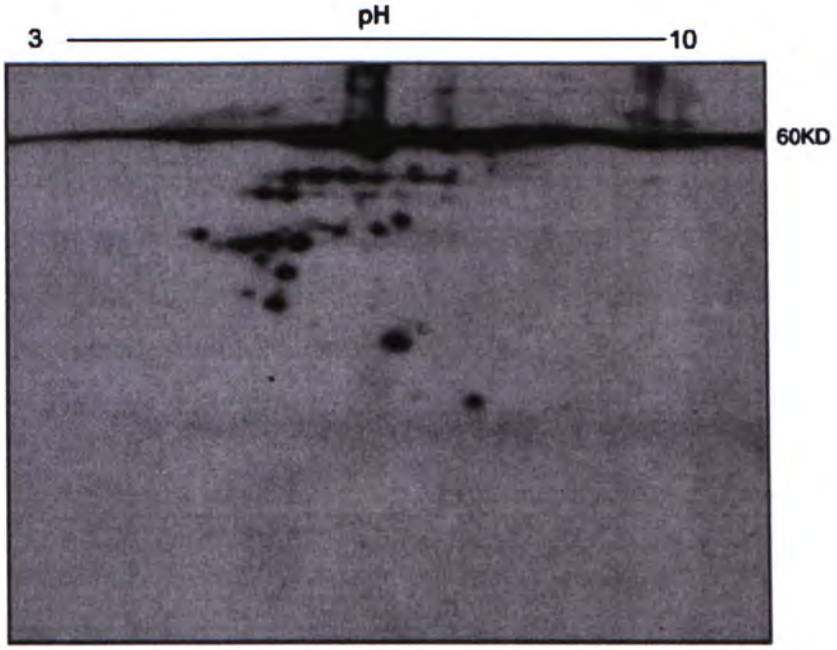


Figure 3.8 Western blot of GBSS

Ten μg of starch granule associated proteins was loaded onto the 7cm pH3-10 IPG strip. After 2D SDS-PAGE, the proteins were transferred onto the nitrocellulose membrane. GBSS proteins were detected by the anti-GBSS antibody (1:5000 dilution) as the primary antibody.

Since GBSS is encoded only by the *Waxy (Wx)* gene for the synthesis of amylose in rice, we thus do not regard the various molecular weight GBSS spots as isoforms, but GBSS fragments. Previous studies had revealed that amylose content in rice was correlated to the ability of intron 1 being excised from the leader sequence of the *Wx* transcripts (Wang *et al.*, 1995). We are interested to know the relationship between the expression of each fragments and the amylose content and whether these fragments have functions or not. Question may be raised that these fragments could be caused by degradation during sample preparation. We thus performed experiments to test such possibility. Parallel experiments of amyloplasts proteins extraction were carried out in the presence or absence of the protease inhibitor cocktail (35μg/ml PMSF; 0.3mg/ml EDTA; 0.7μg/ml pepstatin A; 0.5μg/ml leupeptin). Results were showed in Fig 3.9. If proteolysis is a primary cause, degradation would be expected to increase in the sample without the addition of the protease inhibitor cocktail. However, the 2DE images were similar regardless of adding protease inhibitor or not. Since multiple spots were highly reproducible in number and relative abundance, it is unlikely that the multiple spots observed here were artifacts of protein damage during sample preparation.

Furthermore, the formation of GBSS protein fragmentations had been observed in 1D gel (Mu *et al.*, 1998) and in 2D gel (Boren *et al.*, 2004; Lin *et al.*, 2005), and the origin of these products was unknown. The need to classify these granule associated protein fragments prompted us to investigate the identity of the granule intrinsic proteins and the nature of their association with the starch granules. We thus purified GBSS protein and its fragments from the starch granules for further analysis.

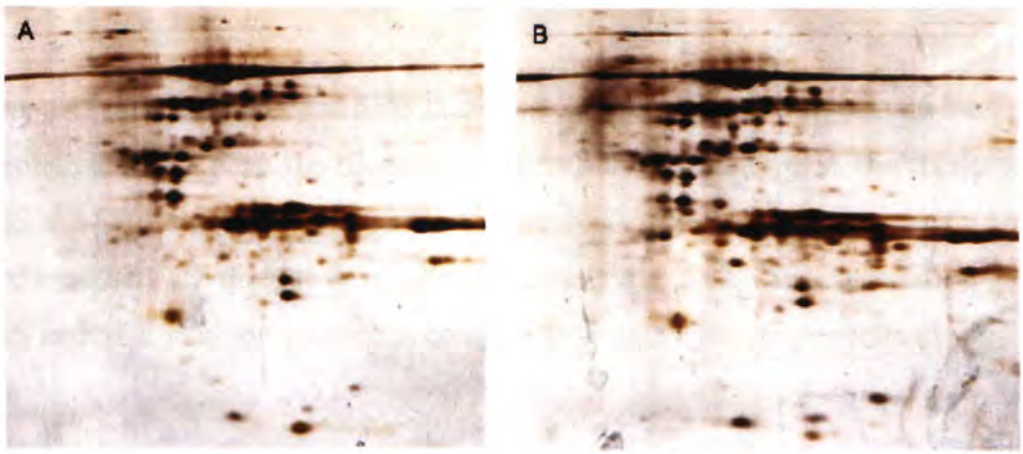


Figure 3.9 Amloplast proteins and protease inhibitors

Amyloplast purification was as described in the previous section. During the purification, the buffers were without adding the protease inhibitor cocktail （A） or adding the protease inhibitor cocktail (B). After purifying the amyloplasts, 10 μg proteins were load onto the pH 3-10, 7cm IPG strip and the SDS-PAGE gels were visualized by silver staining.

3.1.5 N-terminal sequencing

Using the ChloroP prediction algorithm (Emanuelsson *et al.*, 1999) at the ChloroP www server (http://www.cbs.dtu.dk/services/ChloroP/), we predicted the transit peptide of GBSS to be 76 amino acids long, yielding a processed GBSS protein of around 60 kDa. The estimated length of the GBSS transit peptide is in good agreement with our N-terminal sequencing results which confirm that the transit peptide cleavage site is at the 77th amino acid. The amino acid motif KTGGL (Ainsworth *et al.*, 1993; Baba *et al.*, 1993), the putative binding site for ADPGlc (Furukawa *et al.*, 1990), is located 19 amino acids from the N-terminus of mature GBSS (Nakamura *et al.*, 1998).

We sought to identify these GBSS fragments with an aim to determine whether they have a function in starch synthesis. Our approach was first to purify starch granule associated proteins (Fig 3.10) and then to identify them by analyzing their amino acid sequence. So Fig 3.11 shows the summary of our N-terminal sequencing results. During the course of this study, we observed that N-termini of the 55kDa, 50kDa and 28kDa polypeptides exhibited complete identity to the N-termini of the 60kDa GBSS protein. For polypeptides around 45kDa and 35kDa, their N-termini indicate that they do not contain the putative binding site for ADPGlc, thus the 45 and 35kDa GBSS fragments may have lost their functions.

We have also identified SSS and SBE in our starch granule associated protein fraction, but their protein amounts were significantly lower than that of the GBSS. The results showed that SSS and SBE were loosely bound to the starch granules and such association had been reported by others (Darlington *et al.*, 2000; Boren *et al.*, 2004). This association indicates that SSS and SBE may form a protein complex to elongate the amylopectin (Tetlow *et al.*, 2004b).

Figure 3.10 1D SDS-PAGE of starch granule associated proteins

Ten μg of the starch granule associated proteins were loaded onto the 1D SDS-PAGE gel and visualized by the silver staining. The Mw of the 6 selected bands to do the N-terminal sequencing were 60 kDa, 55 kDa, 50 kDa, 45kDa, 35 kDa and 28 kDa.

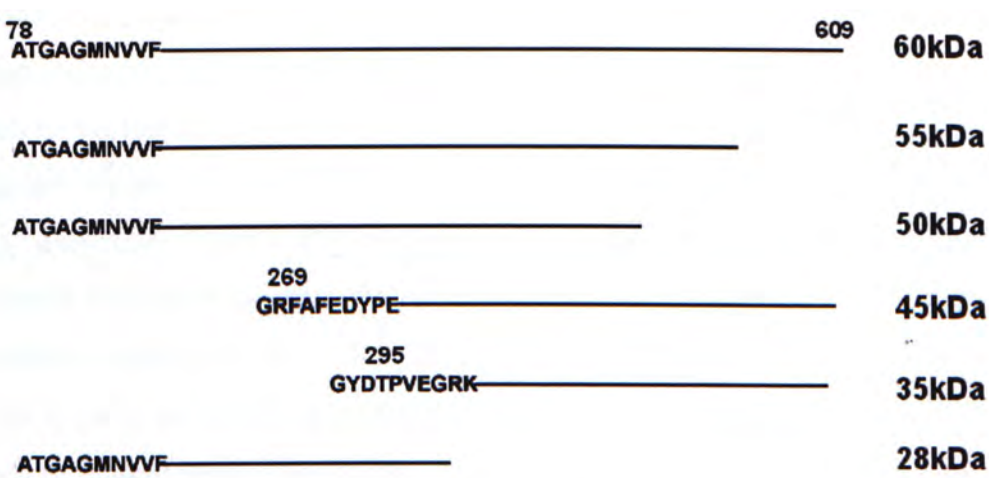


Figure 3.11 Summary of N-terminal sequencing results

The N-termini of the 60kDa, 55kDa, 50kDa and 28kDa GBSS are at the 78th amino acid. The N-termini of the 45kDa GBSS is at the 269th amino acid. The N-termini of the 35kDa GBSS is at the 295th amino acid.

3.2 Protein profiling

3.2.1 Seed collection and stages chosen

In the present study, PA64S, 9311 and their F1 hybrid (PA64S/9311) were used as experimental materials for profiling study. PA64S is a female parental line of hybrid rice with photo- and thermo-sensitive genic male sterility (PTGMS). It has a typical morphology of super hybrid rice and a high combining ability with different restorer lines. Line 9311 is the male parental line and a good restorer line with a better rice quality. PA64S/9311 is the super hybrid rice combination, named LYP9, which is the typical two-line inter-subspecies hybrid. Furthermore, the rice grain quality of hybrid reached second-class national standard and the morphology is excellent. Hence, research on protein expression and comparison between these three lines of rice and for this combination during grain development is necessary and significant in improving its grain quality and hopefully modified cultivars with super high yield and excellent grain quality can be developed in the future.

To assess changes in the rice amyloplast proteome during seed development, we performed two-dimensional SDS-PAGE on proteins isolated from plastid-enriched fractions from four time points, 6, 10, 15 and 20 DAF. These time points are representative in the developmental process of amyloplasts (Duan, 2003). At 5-7 DAF (termed milky stage), the plumule and coleoptile are present and the embryo attains the capacity to germinate; aleurone layer is fully developed; the embryo sac becomes filled with endosperm cells; crystalline protein body appears. At 9-10 DAF (dough stage), cell division is essentially completed and the size of the starch granule increases most rapidly. At 14-15 DAF (yellow stage), caryopsis reaches the maximum width; dry matter is actually accumulated; compound starch granule gains the maximum size; potassium, calcium and manganese are accumulated rapidly in the hull. At 20-21 DAF (mature stage), the differentiation of embryo is fully completed; the caryopsis attains the maximum thickness and the pericarp reaches its mature state; aleurone layer matured; and starch granules are formed.

3.2.2 The proteomic profiles of rice amyloplasts at different developing stages

Our earlier research, as described in previous session, had built 2D reference maps for amyloplast proteins of 9311 at 10DAF using pH3-11NL and pH4-7. The 2D reference map was composed of 141 spots while a 1D reference map contained 25 bands. Here we established twelve 2D SDS-PAGE gels for the three rice lines each at four developing stages by employing pH3-10 24cm IPG strip and 12.5% precast gel to study the amyloplast proteome. For each gel, proteins were identified by PMF and MS/MS.

During profiling, 50μg of protein was loaded on each gel and stained by silver nitrate. Fig 3.12 shows the representative images of 2D gels of rice amyloplast proteins. In order to obtain a reliable result from 2D images for quantitative analysis, the rice amyloplast samples collected from each developing stage were extracted twice for protein. Two 2D gels were run in parallel 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.

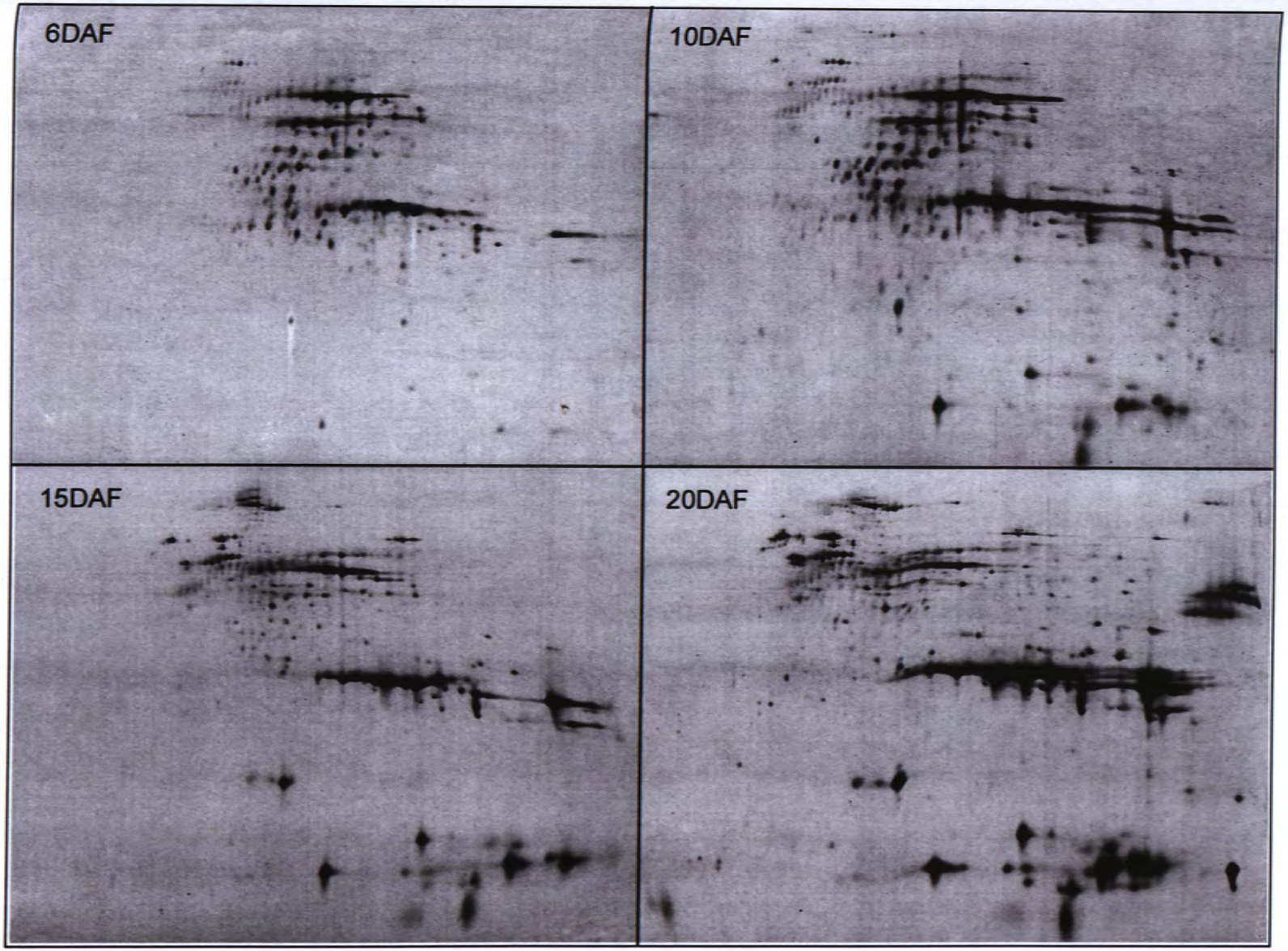


Figure 3.12A Profiling the rice amyloplast proteins of 9311 seeds during development

The amyloplast proteins of rice line 9311 were examined by 2D SDS-PAGE at four developing stages, 6 DAF, 10DAF, 15DAF, and 20DAF. Sample load: 50 μg of protein; first dimension: pH 3-10 24cm IPG strips; second dimension: 12.5% precast gels; stain method: silver stain.

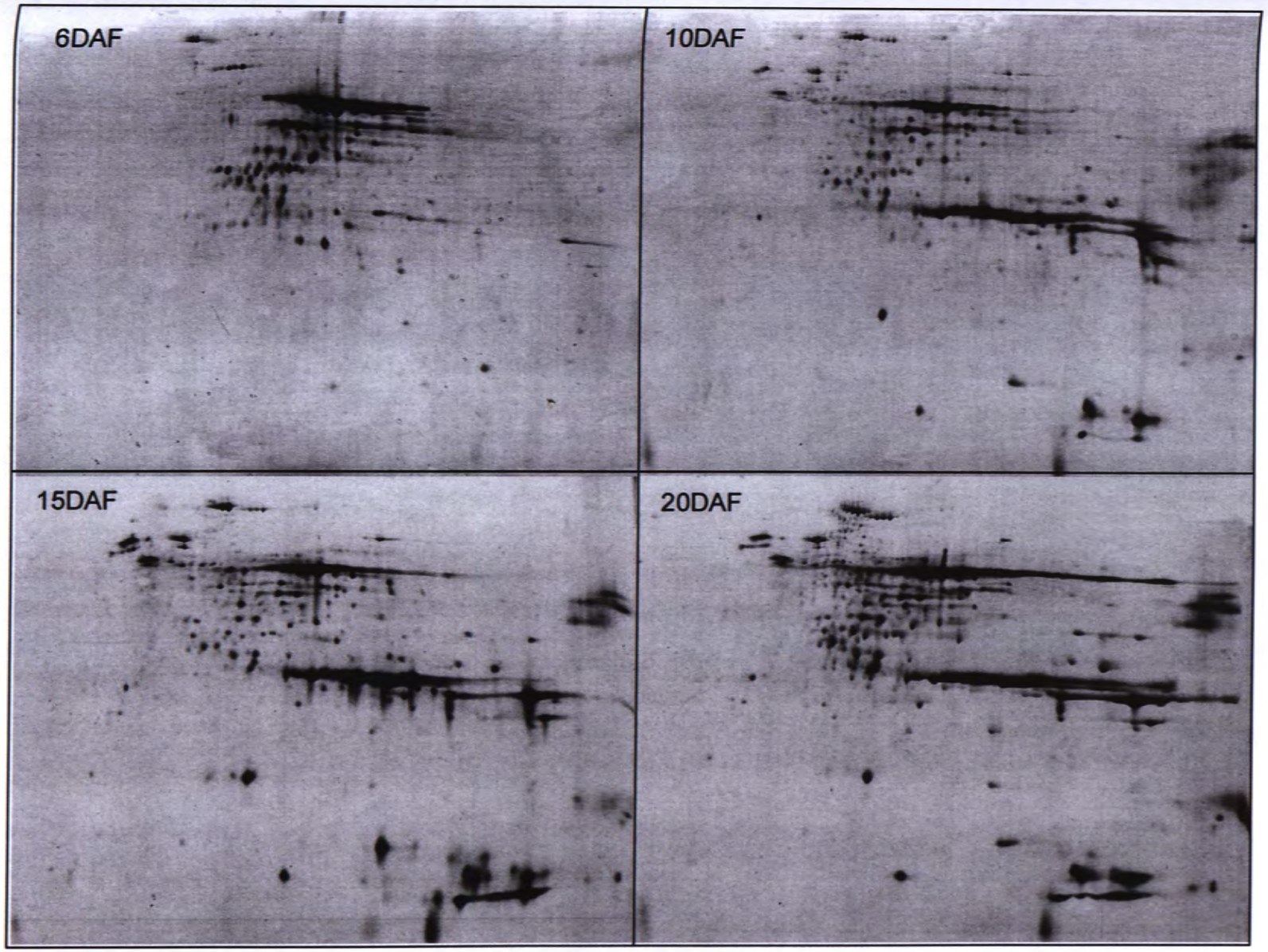


Figure 3.12B Profiling the rice amyloplast proteins of the hybrid seeds during development.

The amyloplast proteins of rice line Hybrid were examined by 2D SDS-PAGE at four developing stages, 6 DAF, 10DAF, 15DAF, and 20DAF. Sample load: 50 μg of protein; first dimension: pH 3-10 24cm IPG strips; second dimension: 12.5% precast gels; stain method: silver stain.

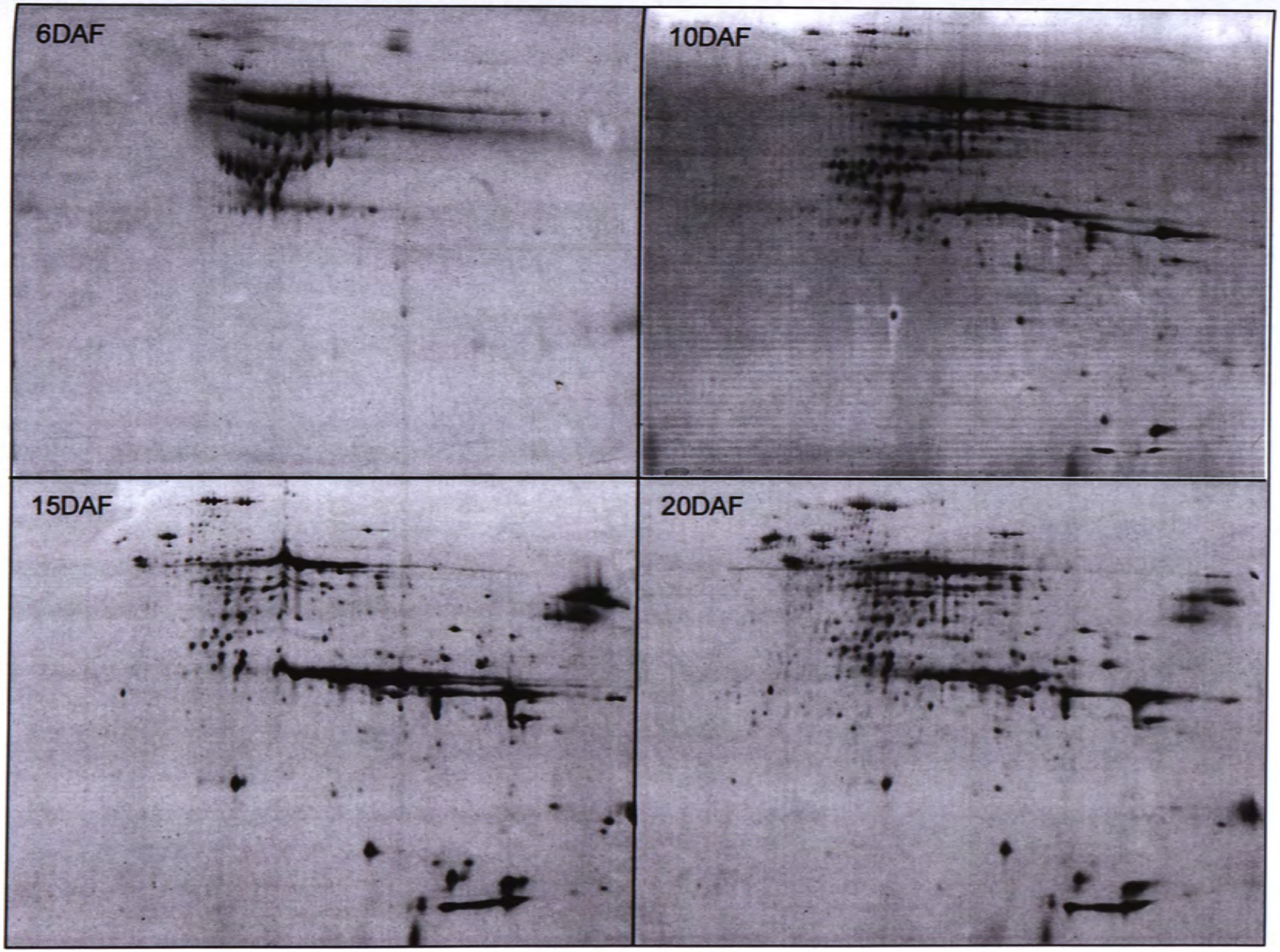


Figure 3.12C Profiling the rice amyloplast proteins of PA64S seeds during development

The amyloplast proteins of rice line PA64S were examined by 2D SDS-PAGE at four developing stages, 6 DAF, 10DAF, 15DAF, and 20DAF. Sample load: 50 μg of protein; first dimension: pH 3-10 24cm IPG strips; second dimension: 12.5% precast gels; stain method: silver stain.

3.2.4 Comparing the proteome of three rice lines

3.2.4.1 Spot number analysis

Fig 3.13 revealed that the total number of amyloplast proteins on 2D gels gradually increased in all three rice lines during development. The number of protein spots in 9311 (109±5) is relatively higher than the hybrid (98±3) and PA64S (90±6) at early developmental stage (6DAF). This is because the proteins related to protein destination in hybrid and PA64S appeared (at 10DAF) later than in 9311 (6DAF). The increase of proteins with unknown function, not identified proteins and proteins with function of destination led to the increase of the total number of spots.

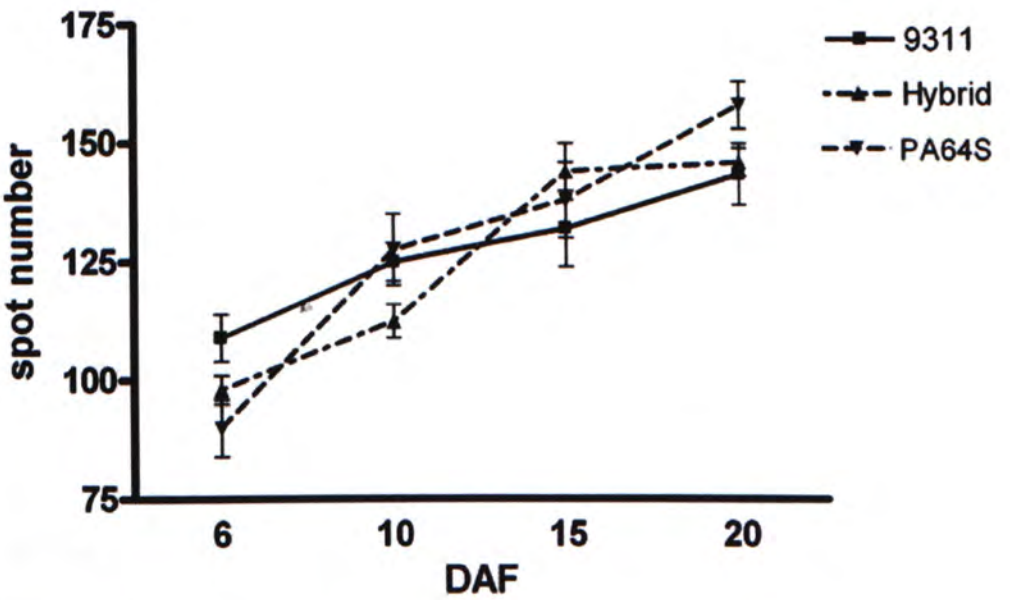


Figure 3.13 Analysis of amyloplast proteins spot number

The spot numbers of each rice line at four developing stages were statistically counted using ImageMaster 2D Platinum, Version 5.00 based on the 2D SDS-PAGE gels in Fig 3.12.

3.2.4.2 Functional distribution analysis

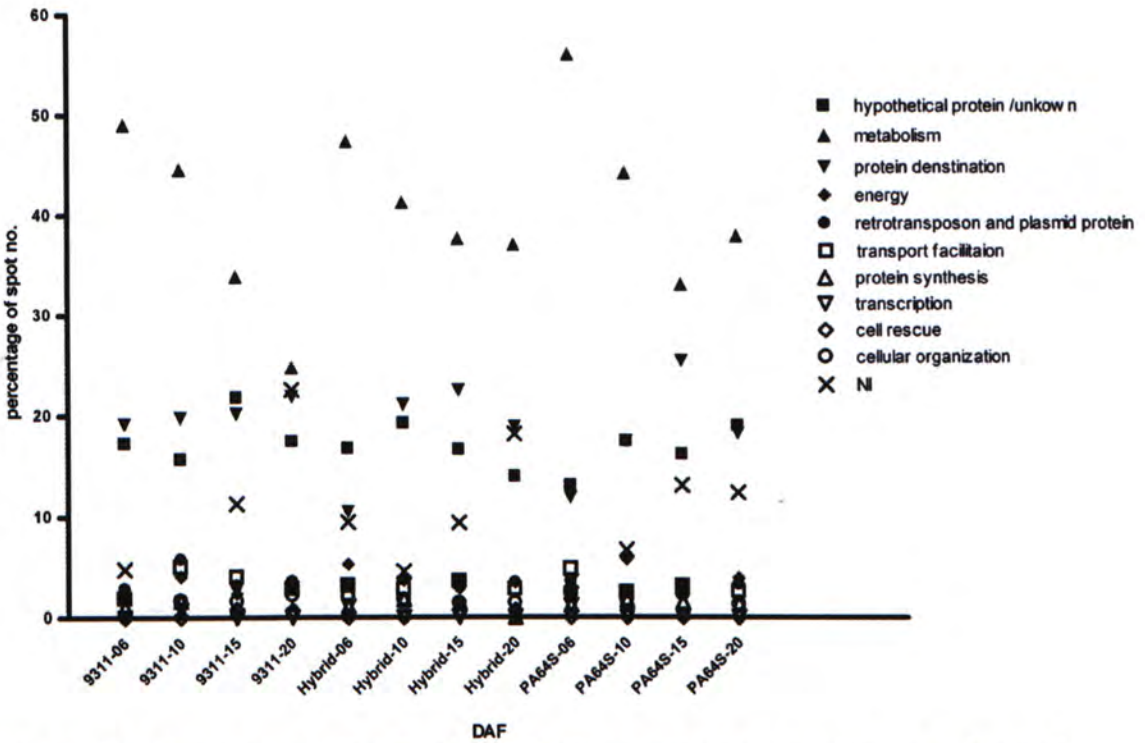


Figure 3.14 Functional distributions of the amyloplast proteins of three rice lines at four developing stages

According to the putative functions, rice amyloplast proteins were classified into 11 categories. For each category, the protein spot numbers (%) of the three rice lines at four developing stages were compared.

The fundamental goal of this study is to define the differentially expressed proteins throughout the process of rice seed development. To gain insight into the biochemical and molecular processes occurring in the amyloplasts, the identified proteins were classified according to putative functions (Koller *et al.*, 2002). Fig 3.14 shows the functional distributions of amyloplast proteins of the three rice lines at four developing stages. Metabolism proteins are the most abundant category. The spot number (%) of metabolism proteins is the highest, more than any other functions at each stage but gradually decreased with time during development. However, their profiles are somewhat different. For 9311, the spot number (%) declined dramatically and linearly at later stages during development; for PA64S, the spot number (56%) is the highest at 6DAF (9311 49.0%; hybrid 47.4%), decreased rapidly through 10 DAF (44.2%) and 15 DAF (33.1%), but, increased somewhat at 20 DAF (37.9%); and for

the hybrid the spot number (%) at the early stage, 6DAF (47.4%), was lower than those of the two parents but decreased slowly during seed development and maintained at relatively high number at 37.1%. As shown in Fig 3.6, 79% of metabolism proteins (9311 at 10DAF) were involved in starch synthesis. Although at early stage (6DAF), the development of rice endosperm just begin and the starch granules increase rapidly at 10DAF, our results showed that most of the proteins related to metabolism appear at the very beginning to have their function and gradually disappear at later stages.

The protein destination proteins increased during development. For 9311 at 6DAF, the spot number (%) is 8.7% and 7.3% higher than those of the hybrid and PA64S, suggesting that at 6 DAF, 9311 is active in synthesis of proteins related to protein destination, more than the two other lines. Our results are also consistent with previous researches that glutelin gene is expressed early at 5 DAF (Kim *et al.*, 1993) and the synthesis and accumulation of prolamin proteins is greatly increased at latter stages (Li *et al.*, 1993). The distribution of proteins with other functions did not change greatly during development and among the three rice lines.

3.2.4.3 Protein amount analysis

The frequency distribution of protein amount of 9311 at 10 DAF was analyzed and shown in Fig 3.15. We separated the proteins into two categories according to their amounts: the high abundant spots (normalized spot volumes ranging from 0.75-8.75%) and the low abundant spots (normalized spot volumes ranging from 0-0.75%). Among these spots, 17.8% belongs to the first category and 82.2% belong to the other category. As expected, the most abundant proteins correspond to the metabolism proteins (43.8%) and storage proteins (50.0%), while the hypothetical proteins are only 6.2%. The low abundant proteins including hypothetical protein/unknown (25%), metabolism (37.2%), protein destination (15.5%), retrotransposon and plastid protein (4.1%), protein synthesis (3.4%), cellular communication (0.7%), energy (5.4%), transport facilitation (4.7%) and transcription (0.7%), however, could play an important role in protein or starch deposition, in transport facilitation, in maintaining cell structure during reserve deposition, or in the transition from a highly active to a quiescent state during seed development.

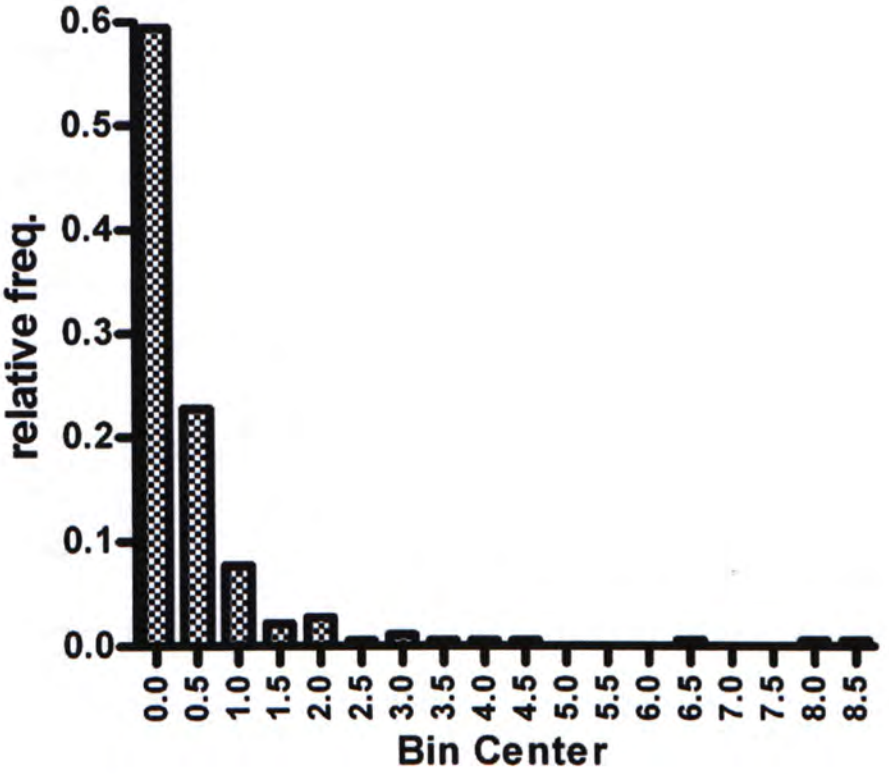


Figure 3.15 Frequency distributions of protein amount of 9311 at 10DAF

The protein amount was the value of the spot volume counted by ImageMaster 2D Platinum, Version 5.00. To discard experimental variations, 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.

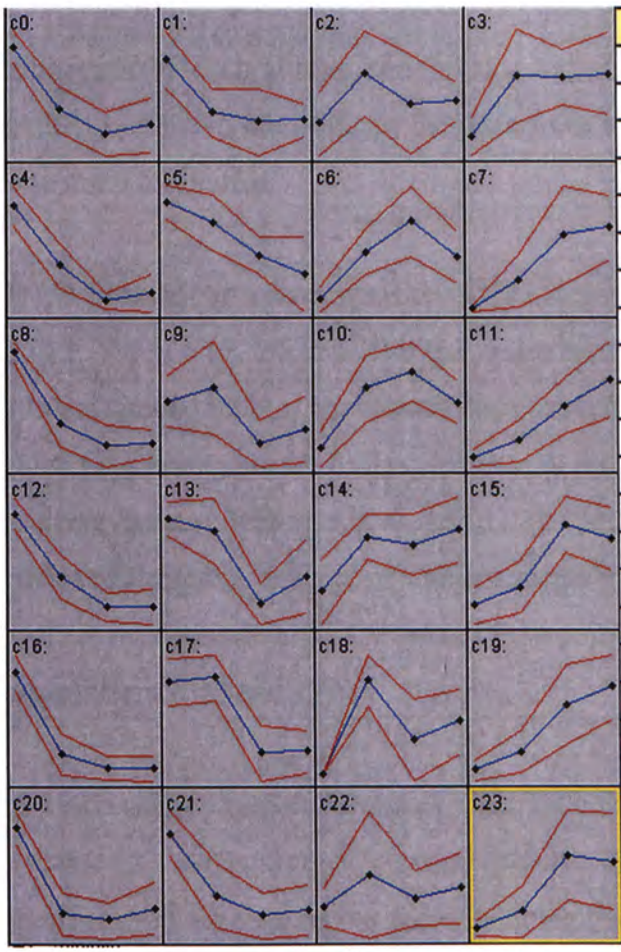
3.2.5 Comparison of expression pattern: cluster analysis (SOM)

3.2.5.1 Cluster analysis of rice amyloplast proteome

Although a growing number of comparative proteomics studies have been reported in plant systems, the grouping of proteins into expression classes has generally been qualitative, and rigorous quantitative measures have been lacking. Recently, using the de-etiolated maize chloroplasts 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). Another research (Tanaka *et al.*, 2005) using Unweighted Pair Group Method with Arithmetic Mean (UPGMA) clustering method to study the protein profiles of rice basal region and S-system and the interactions of the clusters had also been reported.

In our present research, we used nonhierarchical neural network clustering method, self-organized mapping (SOM), to determine the patterns of change in protein expression. SOM has been used previously for microarray data (Maleck *et al.*, 2000; Chen *et al.*, 2002b) but not for proteomics data. SOM clustering works well for large data sets because neural networks are less influenced by noise and the shape of the data distribution (Dopazo *et al.*, 2001). The SOM algorithm maps high-dimensional data onto an ordered two-dimensional space, resulting in an ordered grid where each cell represents a model pattern for the corresponding set of data points. Using SOM clustering method to study changes in protein expression pattern have been evaluated (Lonosky *et al.*, 2004) and demonstrated that it can provide an accurate reflection of the actual patterns of change that occur in individual proteins.

Employing the SOM method, we selected the same 157 proteins in each rice line. For each protein, we assigned spots in each line at the four developing stages as the same group. A total of 471 (157x3) groups of spots was analyzed by the SOM method. The quantity of the protein which did not appear on the gel was assumed as 0. We predefined 24 clusters (0-23) and Fig 3.16 shows the results of SOM analysis of our data. The pattern inside each cell represents the expression pattern over time for that cluster. The data in our research provide a reliable method to assess patterns of change in the plastid proteome during development. Using our methodology, we were able to obtain reproducible data for replicated gels and to classify most of the visible spots on these gels as high quality, facilitating the estimation of spot quantities.



Cluster	9311	Hybrid	PA64S
0	3	1	4
1	0	3	2
2	1	2	1
3	2	1	1
4	2	5	6
5	2	0	1
6	5	5	3
7	3	3	3
8	1	3	6
9	3	2	7
10	1	5	4
11	6	1	2
12	8	12	4
13	3	6	8
14	5	8	6
15	5	5	5
16	10	6	8
17	16	0	9
18	2	5	10
19	14	5	15
20	4	22	3
21	16	13	8
22	23	20	25
23	22	24	16

Figure 3.16 Expression patterns by SOM analysis and number of spots within each cluster

A total of 471 groups of spots in three rice line was selected to do the SOM analysis. We predefined 24 clusters. The pattern inside each cell represents the expression pattern over time for that cluster. The number in each cell represents the number of the proteins for each line and each cluster.

3.2.5.2 Three major categories of rice amyloplast proteome expression patterns

Although for each cluster, the developmental expression pattern is different, they show similar trends. So according to the patterns in spot volumes (%), we divided the 24 clusters into three categories:

(1) Decreasing protein expression: This category includes 12 clusters (clusters 0, 1, 4, 5, 8, 9, 12, 13, 16, 17, 20 and 21), the major portion of the spots. These 12 clusters can be further divided into (a) continuous decrease (cluster 5); (b) continuous decrease for the first three time points but no further decrease at the latest time point (clusters 8 and 12) or with a slight increase (clusters 0, 4 and 21); (c) decrease for the first time point but reaching a plateau (clusters 1 and 16) or a slight increase at the later three time points (cluster 20); and (d) no decrease until the second time point and increase from the third time point and on (clusters 9, 13 and 17);

(2) Increasing protein expression: This includes 9 clusters (clusters 3, 6, 7, 10, 11, 14, 15, 19, and 23); (a) continuous increase (clusters 7, 11 and 19); (b) continuous increase during the first third time points but decrease at the fourth time point (clusters 6, 10, 15 and 23); (c) increase from the first time point and reaching a plateau at the later time points (cluster 3); and (d) a slight decrease from the second to third time points during increase (cluster 14);

(3) Other patterns of protein expression: Alternating increase and decrease at successive time points, such as cluster 2, 18 and 22 (3 clusters).

Several trends emerge from the data. One is that each expression category is nearly identical in all three rice lines. The decrease pattern is the most abundant category in all three lines. The decrease pattern in 9311 (46.5%) is the same as hybrid (46.5%) but a little greater than PA64S (42.0%). The increase pattern in 9311 (37.6%) is a little higher than hybrid (36.3%) and PA64S (35.7%).

We further analyzed the components of each expression pattern (Fig 3.17) and found that the composition of each expression category is also similar in these three rice lines. For example, metabolism proteins are the major component in decrease pattern (28% in 9311; 28% in hybrid; 28.7% in PA64S), while, only 7.6% in increase pattern (in 9311, hybrid and

PA64S), and 5.1% (9311), 5.1% (hybrid) and 4.5% (PA64S) in the other patterns. It seems that a given functional class of protein is generally coordinately regulated in expression, at least during part of the plastid development. The enzymes of starch synthesis generally have the highest expression level at the early developing stage before gradually decreasing, such as GBSS, soluble starch synthase, and ADP-glucose pyrophosphorylase large subunit. There are proteins of other functions with the decrease pattern, such as ribosomal protein and putative sugar transporter. The increase pattern of α-amylase is different since during cereal seed germination, α-amylase plays an important role in hydrolyzing the endosperm starch into metabolizable sugars, providing the energy for the growth of roots and shoots (Akazawa *et al.*, 1985; Beck *et al.*, 1989).

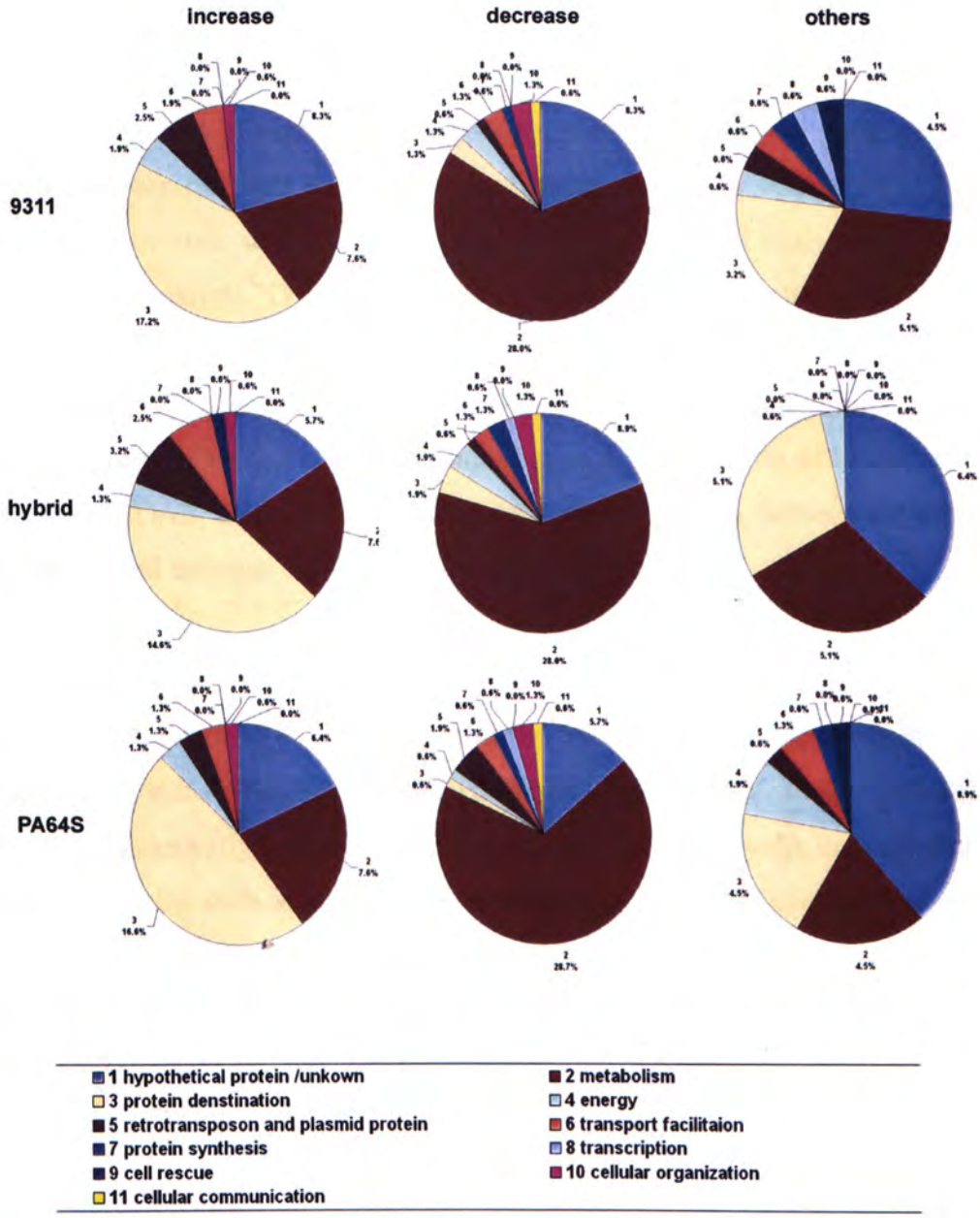


Figure 3.17 Functional distributions of amyloplast proteins of the three rice lines within three expression patterns

After SOM analysis, protein expression patterns were mainly divided into three categories: increase, decrease and others. According to the proteins' function, we further divided them into 11 classes in each expression pattern of the three rice lines.

3.2.6 Scatter plots analysis

Scatter plots are often used to reflect the similarity between two different gels. Scatter plots among the three lines at each developing stages were compared and we hope to find the difference between different lines at the same stages (Fig 3.18). A total of 157 groups of spots in three rice lines, which were also selected to do the SOM analysis, were used to do the scatter plots analysis. The correlation coefficiency of each scatter plot was calculated. The correlation coefficiencies between 6DAF 9311 and 6DAF PA64S (0.4875) and between 6DAF 9311 and 6DAF hybrid (0.5041) are dramatically lower than others, indicating a great difference in the proteomes between 9311 and PA64S and between 9311 and hybrid at 6DAF, indicating that the great difference exists between 6DAF 9311 and 6DAF PA64S and between 6DAF 9311 and 6DAF hybrid.

Further analysis of the proteins in the three gels, we found that the expression levels of metabolism proteins in 9311 are lower than the hybrid and PA64S, while the expression levels of hybrid and PA64S are similar. The expression levels of storage proteins-glutelin in 9311 are dramatically higher than in hybrid and PA64S. Results indicate that most of metabolism proteins such as GBSS are accumulated at the early developing stage (6DAF). For 9311, the storage proteins are accumulated at 6DAF but for the other two lines, they are later than 9311, at 10DAF.

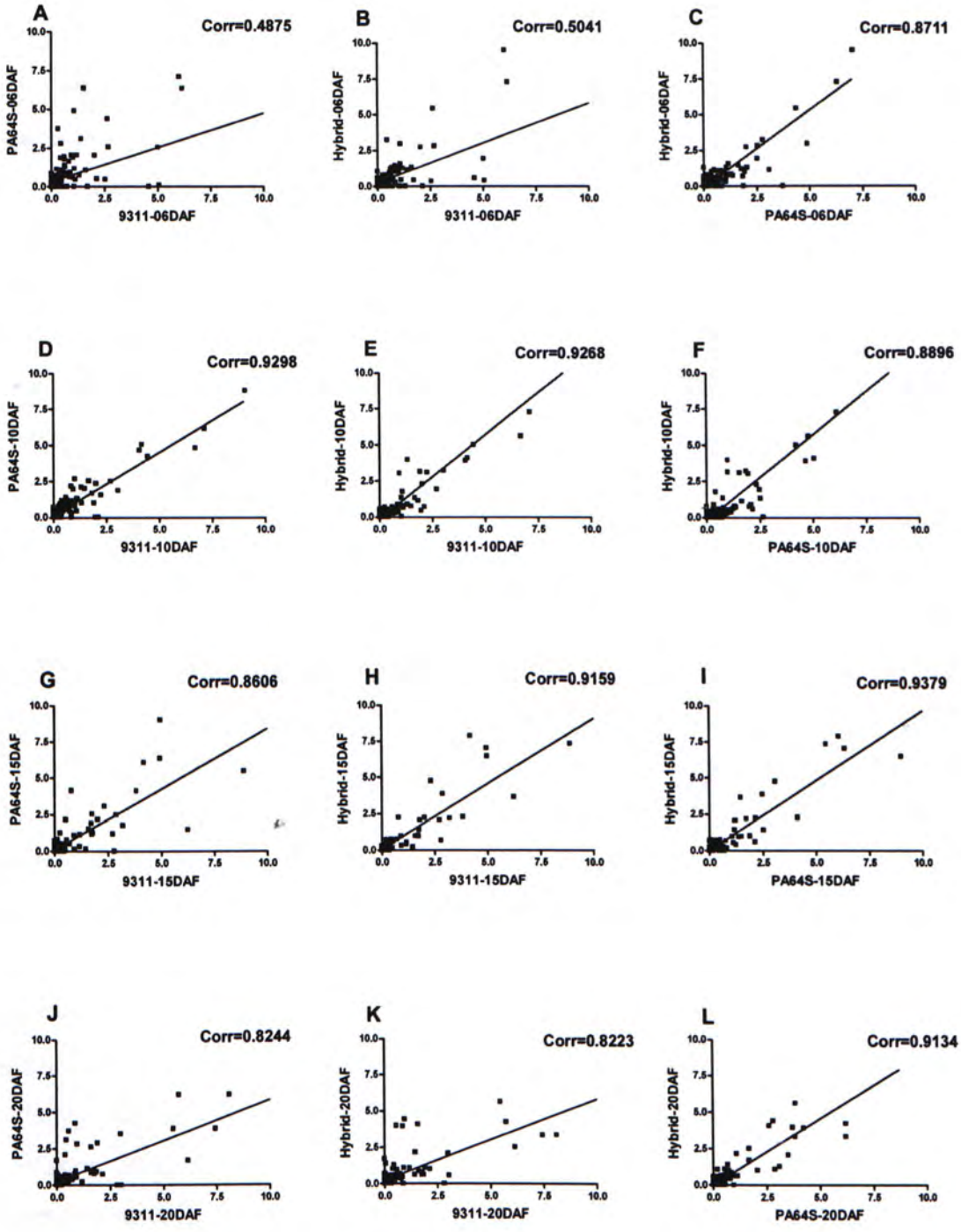


Figure 3.18 Scatter plot analysis of rice amyloplast proteins between two different rice lines at the same developing stage

The scatter plot analysis used the spot volume (%) of the selected 157 proteins in each rice line at the four developing stages. The spot volume (%) of each line at each developing stage was directly plotted onto the plot. The 'Corr' is the correlation coefficiency of each plot.

3.2.7 Comparison of changes in proteins related to starch synthesis

Rice grain development was divided into the following stages in this study: early stage (5-7 DAF) when total dry matter in seeds starts to increase and endosperm starch begins to accumulate; middle stage (9-10 DAF) when endosperm starch and grain dry weight rapidly increase; late stage (14-15 DAF) when both endosperm starch and seed weight continue to increase, but approaching maximum values and becoming constant; and mature stage (20-21 DAF), when differentiation of embryo is fully completed, the caryopsis attains the maximum thickness, the pericarp reaches its mature state, aleurone layer matured, and starch granules are formed. The expression profiles of AGPase, GBSS, SSS, SBE and SP throughout seed development were followed (Fig 3.20).

3.2.7.1 GBSS

GBSS is one of the most important proteins in starch synthesis pathway for amylose synthesis (Nelson *et al.*, 1962; Shure *et al.*, 1983). As described in previous session, we found that the GBSS protein has multiple spots with different molecular weights and different pI points in 9311 at 10DAF and the cause of this fragmentation may be by heat or high concentration urea (Mu *et al.*, 1998). We found that not only 9311 at 10DAF, but all the three lines at four developing stages showed similar patterns. We have extensively studied the multiple spots of GBSS. Fig 3.19 shows the changes in spot number (%) and total expression level of GBSS. During seed development of all the three lines, both the spot number (%) and total expression level decreased, indicating that the function of this protein in the synthesis of amylose begins at the early developmental stage and its activity gradually declines with time. Comparison of the changes of spot number (%) and expression level among the three rice lines revealed that both spot number (%) and expression level in PA64S are the highest, while 9311 the lowest and significantly lower than PA64S, and the hybrid is slightly lower than PA64S. This result is consistent with mRNA expression level and amylose content (Duan *et al.*, 2005). Previous studies revealed that amylose content in rice is correlated with plants ability to excise intron 1 from the leader sequence of the *Wx* transcript (Wang *et al.*, 1995). Because of the unefficient excision of intron 1 in 9311, the level of *Wx* mature mRNA was low, resulting in lower GBSS protein expression level in 9311 than the other two lines. According to the rice amylose contents performed by the HHRRC in Changsha, Hunan, the amylose content of

9311 is 14.9%, the lowest among the three lines. The amylose content of F1 hybrid (21.4%) is 0.9% less than that of PA64S (22.3%) but 6.5% more than that of 9311 (14.9%). Our results are coincided with the low level of *Wx* mature mRNA and amylose content in 9311.

We have analyzed the expression pattern of each GBSS spot in detail using cluster analysis. Although among spots, their expression patterns may show some difference, 83.3% (9311), 85.7% (hybrid) and 83.3% (PA64S), GBSS spots neverthless share similar decreasing pattern, which is consistent with the total GBSS expression pattern.

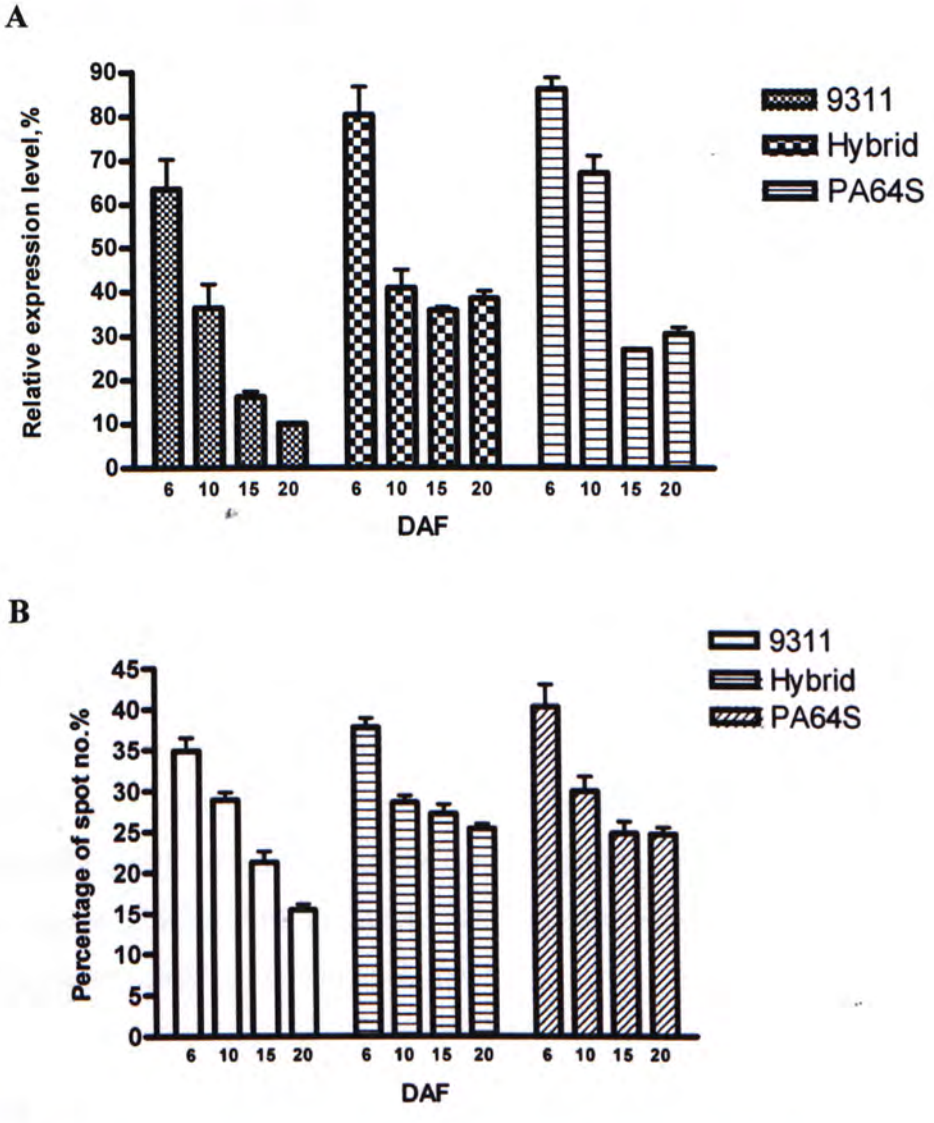


Figure 3.19 Comparison of GBSS expression level and spot number

Relative expression level （A） and Spot number (%) (B) of the three rice lines (9311, Hybrid and PA64S) at four developing stages (6DAF, 10DAF, 15DAF and 20DAF) were compared.

3.2.7.2 AGPase

In a previous study (Nakamura, 2002), it was shown that rice endosperm contains at least two AGPase large subunits and one small subunit. In our experiment, we detected two isoforms of the AGPase large subunit and one AGPase small subunit. Both of the large subunits and small subunit were highly expressed at the early stage (6DAF) but decreased in expression during development, suggesting that the AGPase small subunit combines with AGPase large subunits to form the plastidial AGPase, an important event before or/and at the early stage of endosperm development. It is interesting to note that the expression level of one large subunit was higher than the other.

3.2.7.3 SSS

SSS is another very important protein in amylopectin synthesis (Harn *et al.*, 1998; Edwards *et al.*, 1999; Commuri *et al.*, 2001). We only identified two isoforms of soluble starch synthase, SSSII and SSSIII-2 in our study. The expression pattern is similar with GBSS, but the protein amount is dramatically lower than GBSS. A possible reason is that GBSS is tightly bound to the starch granule and retains stability.

3.2.7.4 SBE

Plants contain two BEs, BEI and BEII. Rice contains a single gene for BEI and two genes for BEII (BEIIa and BEIIb). It has been reported that BEIIb is specifically expressed in the endosperm whereas both BEI and BEIIa are expressed in all tissues examined (Yamanouchi *et al.*, 1992; Mizuno *et al.*, 1993). In our samples, we identified one SBE. The expression profile of SBE was of the decrease pattern, but with a little increase during the later stages (15-20DAF) in hybrid. Our results were consistent with its mRNA expression pattern observed (Duan *et al.*, 2005).

3.2.7.5 SP

We detected two isoforms of starch phosphorylase (SP) and their expression patterns were quite different with other starch synthesis related proteins. Both SPs at stage 6 DAF were quite low in expression level, while increased rapidly from 6 DAF to 15 DAF and decreased somewhat at the later stage. SP is generally considered to be involved in starch

degradation, but recent studies suggest that SP may play a role in starch biosynthesis, although the precise mechanism is unclear (Colleoni *et al.*, 1999; Ball *et al.*, 2003; Tetlow *et al.*, 2004b). The association of SP with SBEI and SBEIIb in a putative complex in amyloplasts, as proposed by Tetlow *et al* (2004b), suggests that SP is involved in starch synthesis. There are two alternative means by which SP can perform such a biosynthetic role. Firstly, SBEs could utilize the products of SP activity, i.e. adding branch points to the glucan chains elongated by SP (operating in a glucan forming direction). A second, more likely scenario, is that SP operates after SBE activity by modifying the structure of amylopectin via phosphorolysis. Many biological multi-protein complexes contain one or more large proteins that are believed to be essential components, both functionally and structurally (Nakayama *et al.*, 2002). SP normally functions as a large multimeric protein, and may also act as a scaffold in addition to other functions within the protein complex. The expression of SP was later than SBE during seed development, which may suggest that the protein complex may not form at the very early stage. The protein expression pattern of SP was quite different with the pattern at the transcript level. From Ohdan's results (Ohdan *et al.*, 2005), the SP transcripts were numerous at the onset of seed development, rapidly increased to peak at 5 DAF, and diminished to a low but significant level until 15 DAF. This may suggest that although the transcripts level of SP is decreased during development, the protein is still accumulated and not degraded, and also suggested that SP may not only involve in the starch synthesis, but also involve in the starch degradation.

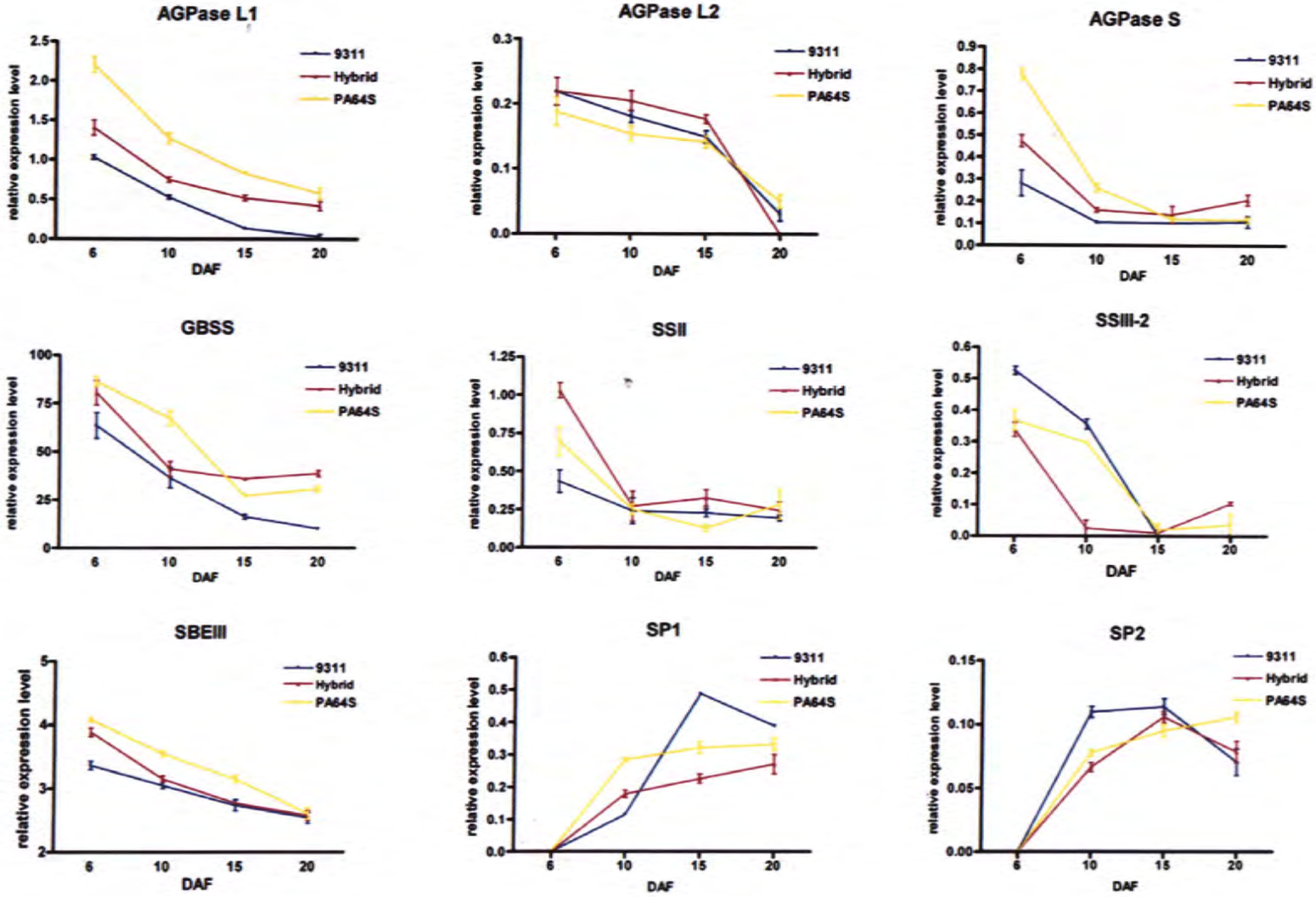


Figure 3.20A Protein expression patterns of starch synthesis related proteins during seed development

Expression patterns of nine starch synthesis related proteins of three rice lines 9311 (blue line), hybrid (pink line) and PA64S (yellow line) were profiled, including AGPase L1, AGPase L2, AGPase S, GBSS, SSII, SSIII-2, SBEIII, SP1 and SP2.

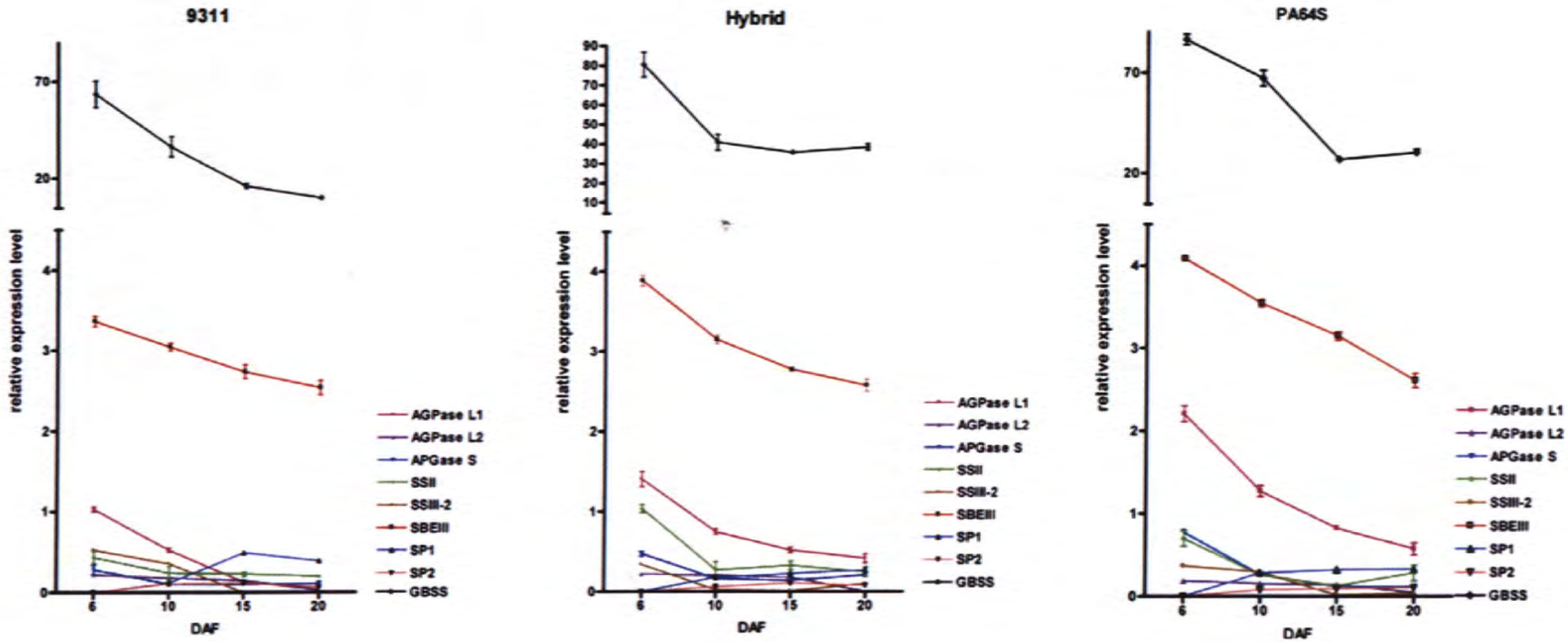


Figure 3.20B Protein expression patterns of starch synthesis related proteins during seed development

Nine starch synthesis related proteins were profiled in each rice lines (9311, PA64S and hybrid), including AGPase L1, AGPase L2, AGPase S, GBSS, SSII, SSIII-2, SBEIII, SP1 and SP2.

3.3 Study on protein post translational modifications

3.3.1 Post translational modifications that potentially regulate starch synthesis

PTM is another very important reason to cause multiple spots of the same protein. We have employed recent technology to detect the most common PTMs, glycosylation (Steinberg *et al.*, 2001) and phosphorylation (Steinberg *et al.*, 2003). The results in Fig 3.21 showed that GBSS and its fragments were phosphorylated and glycosylated. Protein kinase cascades play essential roles in diverse intracellular signaling processes in animals and yeast. In plants, there is evidence that protein phosphorylation plays an important role in signaling pathways triggered by abiotic stress, pathogen invasion, and plant hormones (Knetsch *et al.*, 1996; Sheen, 1996; Hart, 1997; Zhang *et al.*, 1998). Despite the importance and widespread occurrence of protein phosphorylation in plants, little is known about its possible role in regulating starch biosynthesis. A recent study with isolated amyloplasts from wheat endosperm identified a number of phosphoproteins, including some involved in starch metabolism, indicating that some aspects of starch (amylopectin) biosynthesis may be controlled by protein phosphorylation (Tetlow *et al.*, 2004b). This result revealed the importance of phosphorylation in starch synthesis and its potential in regulation of the starch biosynthesis. Although we have detected phosphorylation on GBSS fragments, however, whether they are functional, need further investigation. Besides GBSS, other proteins were also detected with PTMs. SSS and SBE are phosphorylated but not glycosylated (Fig 3.21). Some measure of coordination and regulation of starch biosynthetic enzymes is increasingly thought to occur through specific protein-protein interaction. There should be a complex relationship between the rate of amylose synthesis and the overall rate of starch synthesis in storage organ.

It is known that glycosylation is related to storage protein stability during accumulation (Lamport, 1980; Shewry *et al.*, 1995a; Kishimoto *et al.*, 1999), by preventing the degradation of storage proteins during seed germination and in early seedlings (Faye *et al.*, 1989). For GBSS, it was only recently reported on its glycosylation (Lin *et al.*, 2005). As for the function of glycosylation of GBSS, the sugar side chain may interact with starch granules or to prevent it degradation during seed development.

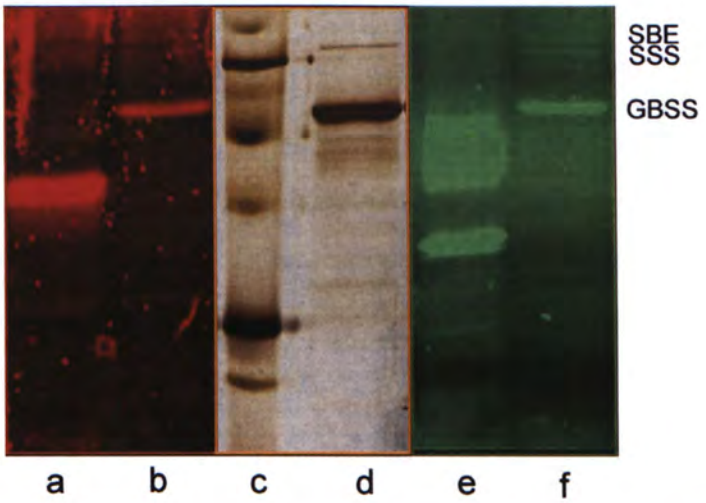


Figure 3.21A Glycosylation and phosphorylation staining of rice seed starch associated proteins in 1D SDS-PAGE

Five μg of rice seed starch associated proteins were separated by 1D SDS-PAGE and followed by glycosylation staining (a, b), silver staining (c, d) and phosphorylation staining (e, f). Glycoprotein standard (a), protein Mw marker (c) and phosphoprotein standard (e) were parallelly run with starch associated proteins (b, d and f). GBSS was detected with glycosylation and phosphorylation. SBE and SSS were detected with phosphorylation.

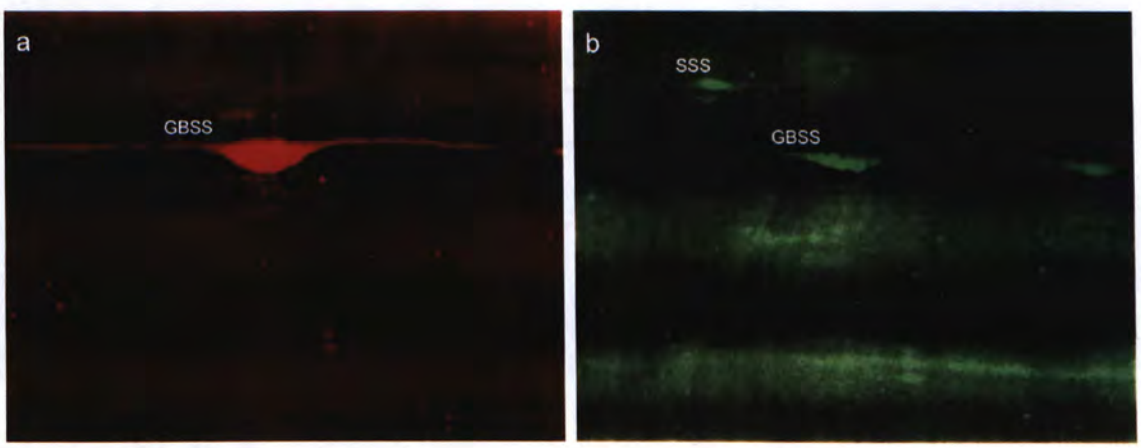


Figure 3.21B Glycosylation and phosphorylation staining of rice seed starch associated proteins in 2D SDS-PAGE.

Ten μg of rice seed starch associated proteins was separated by 2D SDS-PAGE and followed by glycosylation staining (a), and phosphorylation staining (b). GBSS was detected with glycosylation and phosphorylation. SSS were detected with phosphorylation.

3.3.2 Post translational modifications at different developing stages

3.3.2.1 Profiling of post translational modifications of rice amyloplast proteome

We found that several proteins including GBSS, SSS and SBE showed PTMs, such as phosphorylation or glycosylation. PTM study in starch synthesis related proteins have also been reported recently (Tetlow *et al.*, 2004b; Lin *et al.*, 2005). But our study represents the first time to profile the amyloplast proteins with PTMs by using the phosphoprotein and glycoprotein staining methods. Fig 3.22 shows the 2D images of profiling results.

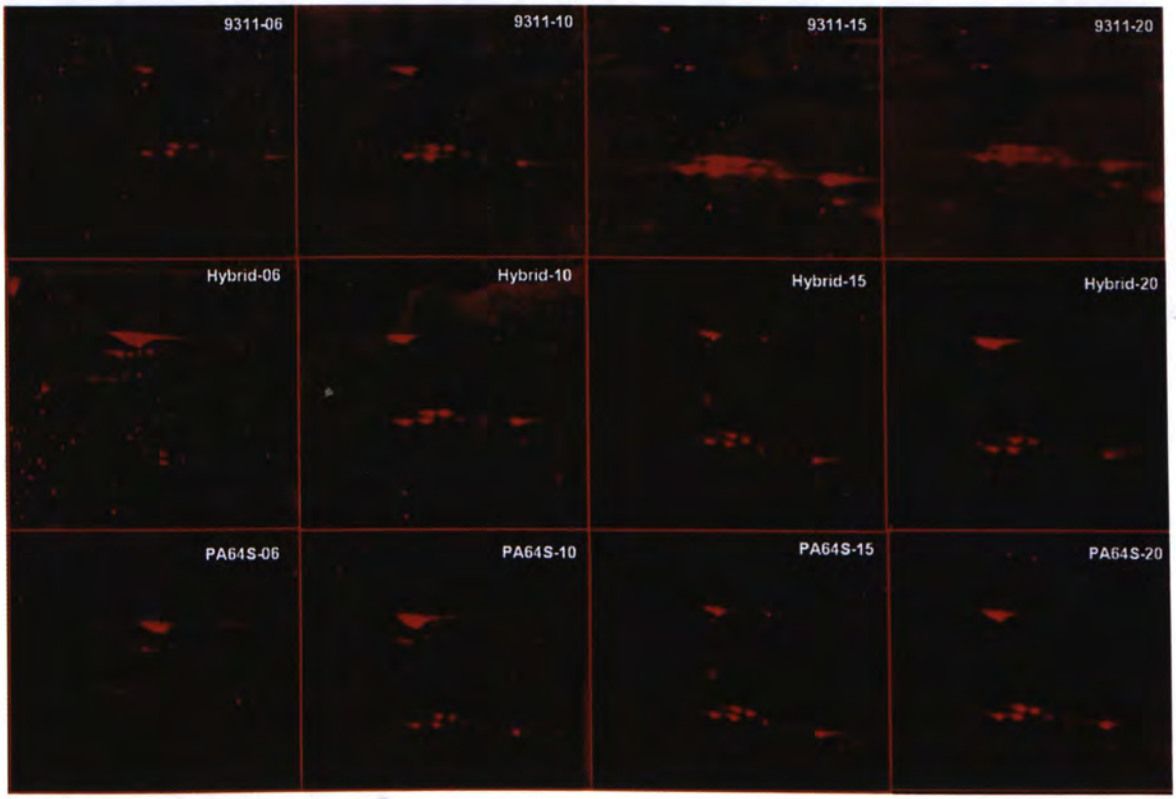


Figure 3.22A 2D image of glycoprotein staining of amyloplast proteins in the three rice lines at four developing stages

Ten μg of rice seed amyloplast proteins was separated by 2D SDS-PAGE and followed by glycosylation staining.

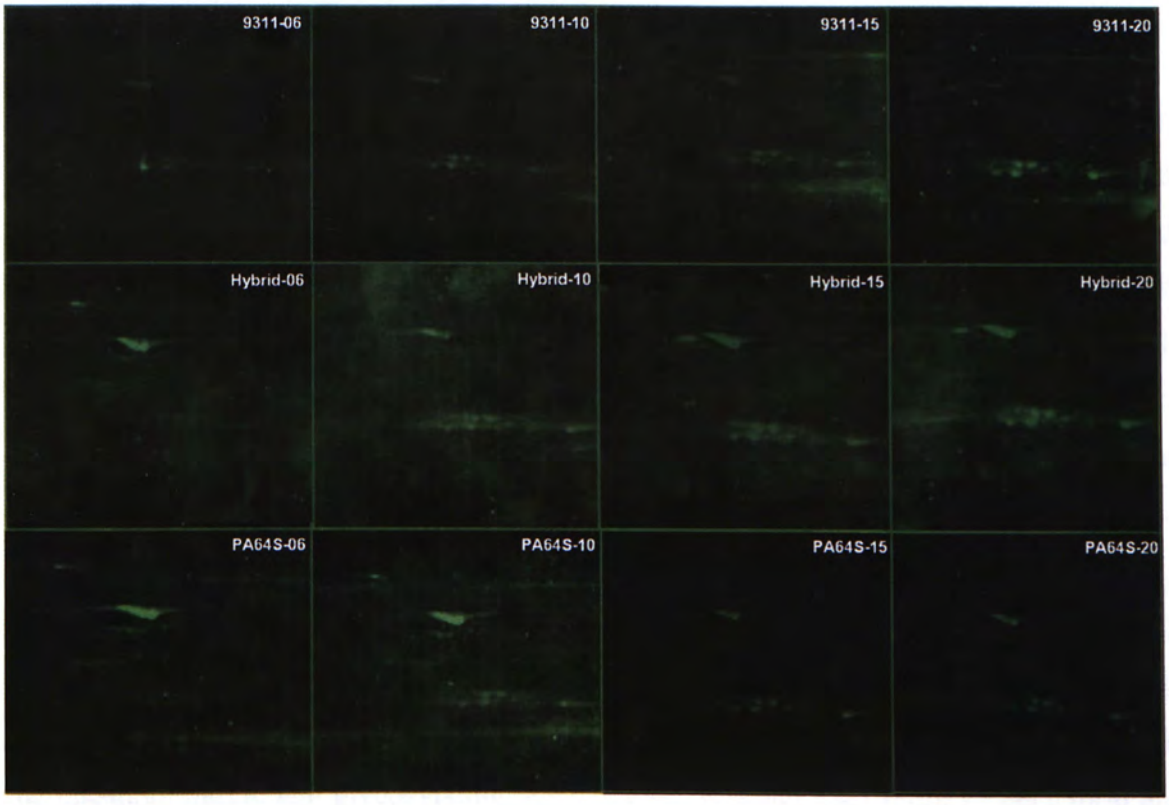


Figure 3.22B 2D images of phosphoprotein staining of amyloplast proteins in the three rice lines at four developing stages

Ten μg of rice seed amyloplast proteins was separated by 2D SDS-PAGE and followed by phosphorylation staining.

3.3.2.2 Starch synthesis related proteins

3.3.2.2.1 GBSS

We focused on the post translational modifications of GBSS. Fig 3.22 shows the profiling by 2-D images of GBSS protein using phosphoprotein and glycoprotein staining method. We observed that the major 60kDa GBSS was phosphorylated and glycosylated in all developing stages. The phosphorylated and glycosylated GBSS fragments were only observed in the early stages, however from the silver staining gels, the GBSS fragments were always detected during seed development, indicating that the GBSS fragments may be only functional at early stages. The expression level is higher in PA64S and hybrid than in 9311. Since the expression level of GBSS is different in the three lines, we have analyzed the relative phosphorylation level and glycosylation level of GBSS (Fig 3.24). Although both the GBSS expression level and the phosphorylated and glycosylated GBSS decreased, their phosphorylation level and glycosylation level also decreased. The patterns of phosphorylation and glycosylation levels in 9311, hybrid and PA64S are similar, while the phosphorylation and glycosylation level in 9311 is lower than those in the hybrid and PA64S.

It have been reported that glycosylation was related to the storage protein stability during accumulation (Lamport, 1980; Shewry *et al.*, 1995b; Kishimoto *et al.*, 1999) by preventing their degradation during seed germination and in the early seedling stage. We thus assume that glycosylation may also have the function to stabilize GBSS by preventing its degradation. At early seed developing stage, since GBSS needs to have high activity to synthesis amylose, the glycosylation level would be higher than in later stage (Fig 3.24).

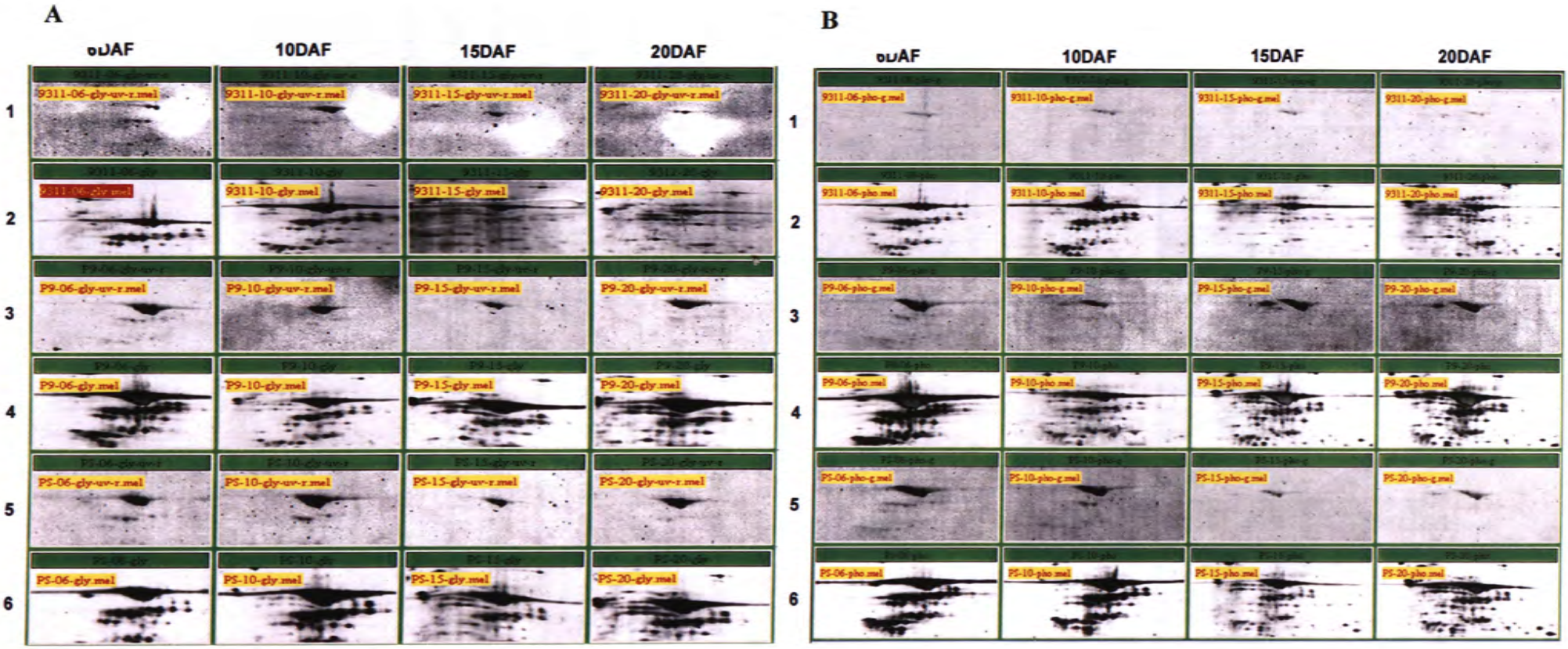


Figure 3.23 Post translational modifications of GBSS

Ten µg of rice seed amyloplast proteins was separated by 2D SDS-PAGE. The gels were first stained by glycosylation staining or phosphorylation staining, and then stained by the silver staining. Glycosylation (A) and phosphorylation staining (B) of GBSS in the three rice line, 9311 (1), hybrid (3) and PA64S (5) at four developing stages (6DAF, 10DAF, 15DAF and 20DAF) were shown in the panels. The rows 2, 4 and 6 are the 2D images of GBSS by silver staining.

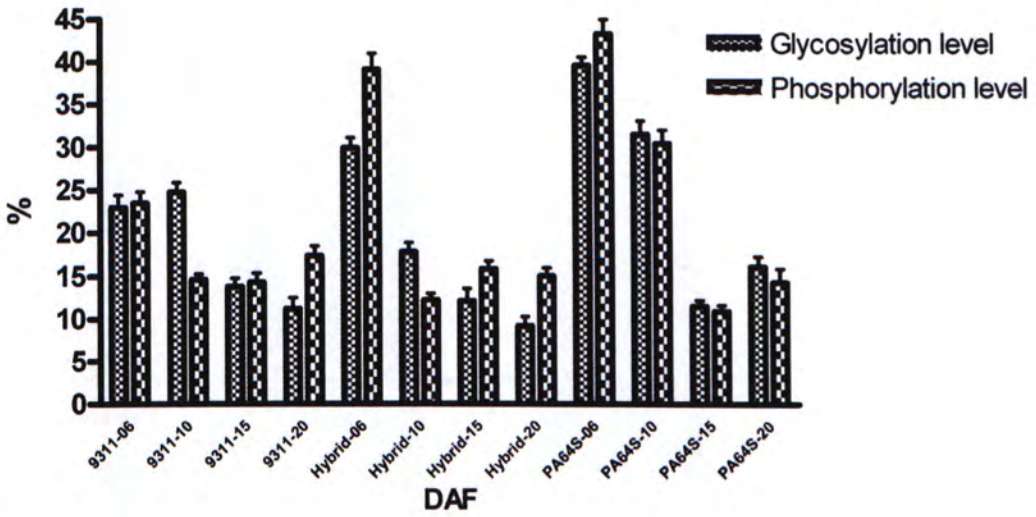


Figure 3.24 Post translational modification analysis of GBSS

The columns represent the glycosylation level and phosphorylation level of GBSS in the three rice lines at four developing stages.

3.3.2.2.2 SSS

Fig 3.25 shows the glycosylation and phosphorylation staining of soluble starch synthase. The results are quite interesting. We found that only phosphorylation can be detected in SSS in lines PA64S and hybrid, but not 9311. Fig 3.26 shows the relative SSS expression level and the phosphorylated SSS expression level. The expression level in 9311 is not much lower than the other two lines at the same stage, while no signal of phosphorylated SSS was detected in 9311. This result indicates that the SSS in line 9311 is not being phosphorylated. This is surprising and interesting finding, but at this time we do not know why SSS in 9311 is not phosphorylated. For the other two lines, the phosphorylation level of SSS shows a decreasing pattern which is similar in GBSS.

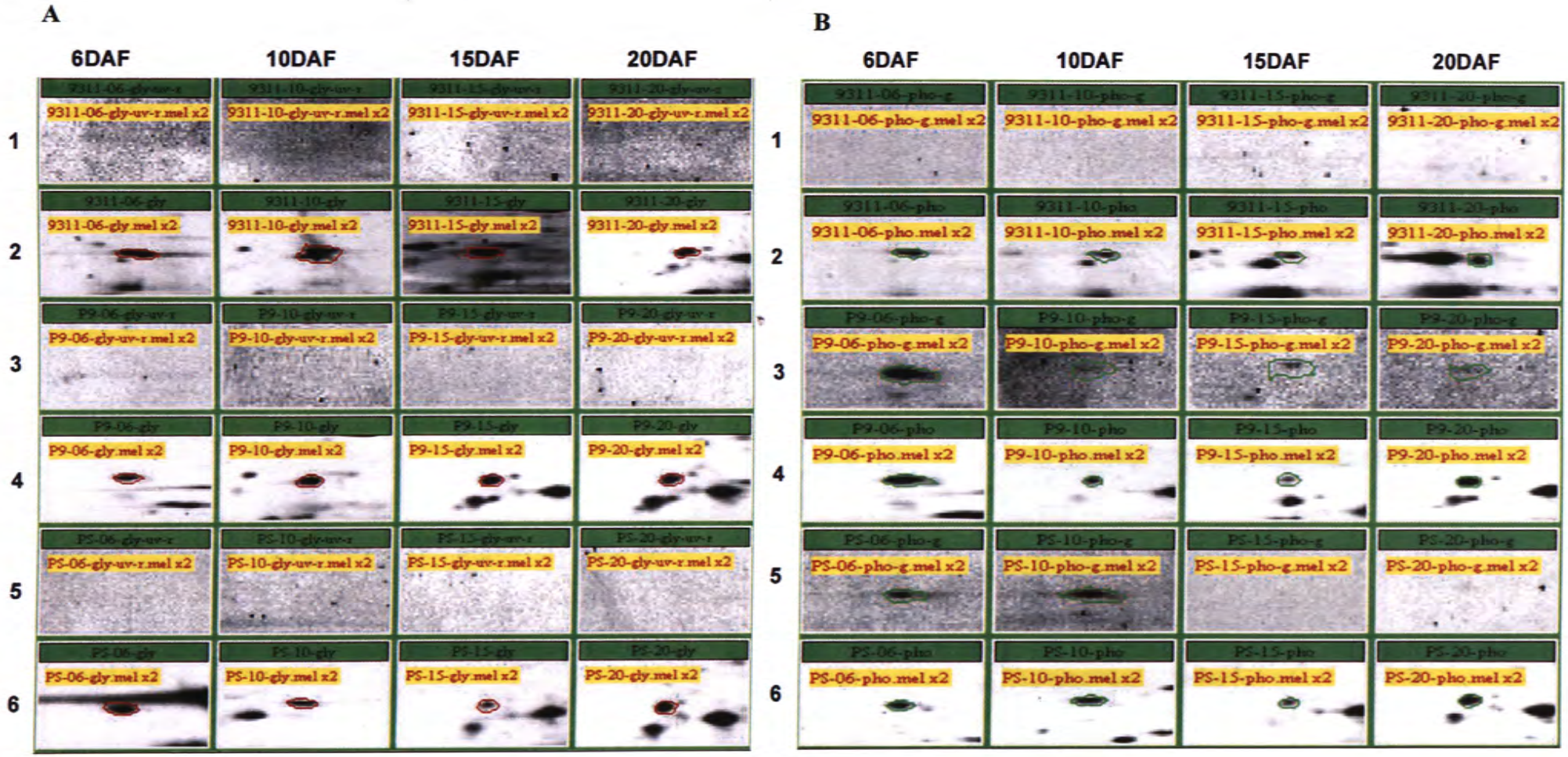


Figure 3.25 Post translational modifications of SSS

Ten μg of rice seed amyloplast proteins was separated by 2D SDS-PAGE. The gels were first stained by glycosylation staining or phosphorylation staining, and then stained by the silver staining. Glycosylation (A) and phosphorylation staining (B) of SSS in the three rice line, 9311 (1), hybrid (3) and PA64S (5) at four developing stages (6DAF, 10DAF, 15DAF and 20DAF) were shown in the panels. The rows 2, 4 and 6 are the 2D images of SSS by silver staining.

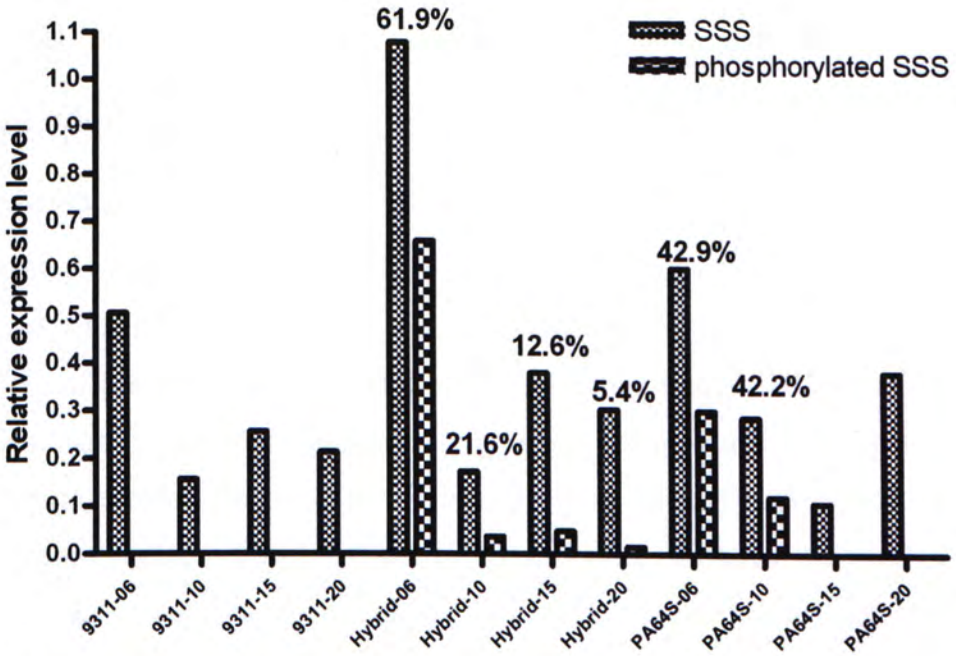


Figure 3.26 Phosphorylation level analysis of SSS

The columns represent the relative expression level of SSS and phosphorylated SSS in the three rice lines at four developing stages. The percentages represent the phosphorylation levels of SSS, using the relative expression level of phosphorylated SSS divide the SSS.

Chapter 4 Discussion

4.1 Methodology

4.1.1 Amyloplast isolation

The integrity of a subcellular proteome is largely dependent on the purification of the target compartment away from other cellular contaminants. To reduce ambiguities in the assignment of proteins to the amyloplast proteome, we first tested different established plastid isolation methods for the purification of rice amyloplast. Sucrose density gradient centrifugation exposed plastids to strong osmotic stress, due to the extremely fragile nature of the rice amyloplasts (Tetlow *et al.*, 1993), this method makes the isolation of intact plastids quite difficult. Percoll density gradient is often used in purifying amyloplasts in maize or wheat, while in rice this method yielded plastids contaminated with other cell organelles and particles, especially protein bodies. Because rice protein bodies I and II are tightly associated with rice amyloplasts (Adoration *et al.*, 1993), which make isolation of high-purity amyloplasts from rice is more difficult than from maize or wheat. In this study we developed a series of fractionation process using Nycodenz density gradient which allowed the isolation of intact and highly pure amyloplast from rice seeds with low speed centrifugation. Nycodenz is a novel density gradient medium. Several researchers used Nycodenz gradient to obtain the high purity plastids. For example, rice etioplasts were highly purified from dark-grown leaves by a novel protocol using Nycodenz density gradient (von Zychlinski *et al.*, 2005). In our method, we use 2% Nycodenz layer twice. The second 2% Nycodenz layer can further purify the amyloplasts and largely remove the contaminants from other organelles, such as cytosol and mitochondria (Tetlow *et al.*, 1993).

4.1.2 Protein extraction from amyloplasts

For protein sample preparation, the starchy characteristics of amyloplasts make it difficult to extract proteins from amyloplasts. Several proteins are intrinsically associated with starch granules, such as granule bound starch synthase. In order to extract these proteins, we need to solubilize the starch granules, either by adding 8M urea extraction buffer or heating the sample at 99°C for 10min. We directly used the 2D rehydration buffer which

contains 8M urea to solubilize starch granules. Although TCA/acetone method is widely used in protein enrichment, we did not employ it because starch and polysaccharides will co-precipitate with TCA/acetone and the proteins will be trapped and dramatically lost.

4.1.3 Protein identification by PMF and MS/MS

The proteins were identified by standard PMF and MS/MS technologies. For identification, we set several criteria for database establishment. All the results were first obtained by PMF and then confirmed by MS/MS. 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 15%. And only identification results with expectation scores below 0.05 (>95% confidence) were considered as positive. Apparent molecular weights and isoelectric points should be close to the ones predicted from the deduced primary sequences of the proteins. However, several identified proteins were accepted with significant differences between theoretical and observed pI. The last criterion is that in addition to technical repeats, two biological repeats were carried out.

Combining the PMF and MS/MS results should lead to more reliable protein identification than MS/MS alone. If only by the de novo results, an identified sequence should have very high quality with which the sequence can be predicted correctly by the software. If the sequence quality is not high enough, it will increase the possibility of mismatching the wrong sequence in the database. Also, the peptide selected to sequence should really belong to the protein not the contaminants. Combination of these two methods can correct the false identifications that may be created by any one of these identification methods.

Since database searching has been reported as one of the greatest stumbling blocks in proteomic research, especially in plants, we combined the NCBInr database and TIGR rice database in our data searching. TIGR rice database is a downloaded protein database that we added into our Mascot search engine to do the local protein database search. We used TIGR Rice Genome Pseudomolecules version 2 (http://www.tigr.org/tdb/e2k1/osa1/) (Yuan *et al.*, 2003). Several researches have used the TIGR rice database in database search (von Zychlinski *et al.*, 2005).

4.1.4 Methods used to study protein expression patterns

Comparative proteomics have raised the hope in recent years to find the differences between two or more samples or treatments. Several studies have investigated temporal changes in plant proteomes involving two or more different time points (Wilson *et al.*, 2002; Gallardo *et al.*, 2003; Shen *et al.*, 2003; Watson *et al.*, 2003; Vensel *et al.*, 2005; Zhao *et al.*, 2005). However, quantitative measures with reproducibility were not reported, only qualitative, nor were there rigorous quantitative analyses conducted to group proteins into expression classes (e.g. clustering analyses). Recently, more researchers focused on large scale comparative proteomics, which requires analysis of large amounts of data. Traditional method by plotting the expression pattern manually is laborious and time consuming and it is quite difficult to summarize and compare expression patterns. Two previous researches which employed new analysis methods were adopted as models in our data analysis of large-scale protein expression patterns. One is the adaptive resonance theory 2, a nonhierarchical, neural clustering techniques used to analyze the expression patterns of the de-etiolated (greening) of maize chloroplasts (Lonosky *et al.*, 2004). The other research is the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) clustering method used in protein profiling of the rice basal region and S-system to study the interactions of the clusters (Tanaka *et al.*, 2005).

To explore the aspect of possible changes in starch biosynthesis in the seeds of hybrid (LYP9), we have systematically studied the expression patterns of amyloplasts proteins, in the hybrid and its paternal cultivar 9311 (*Oryza sativa* L. ssp. *Indica*) and its maternal cultivar PA64S (mixed background of *indica*, *japonica* and *javanica*). In our present research, we used nonhierarchical neural network clustering method, the self-organized mapping (SOM), to determine the patterns of change in protein expression. SOM has been used previously for microarray data (Maleck *et al.*, 2000; Chen *et al.*, 2002b) but not for proteomics data. The SOM clustering method works well for large data sets because neural networks are less influenced by noise and the shape of the data distribution (Dopazo *et al.*, 2001). The SOM algorithm maps high-dimensional data onto an ordered two-dimensional space, resulting in an ordered grid where each cell represents a model pattern for the corresponding set of data points. Also, the SOM can automatically deal with large scale data, so that we can easily summarize the expression patterns. And there is no need for us to plot the expression pattern for each spot manually. Using SOM clustering method to study changes in protein expression pattern had been evaluated (Lonosky *et al.*, 2004)

which indicated that SOM can provide an accurate reflection of the actual patterns of change that occur in individual proteins. Using SOM, we have generated high-quality reproducible data sets for comparative proteomics. Of the three lines, the protein spots in rice amyloplasts, either in spot number, or in spot distribution pattern, or protein expression pattern, the hybrid was quite similar to those of the PA64S, suggesting the dominance of maternal genes in the hybridization, at least in the proteome of developing rice amyloplasts.

Cluster analysis can be used to compare the expression patterns among the three rice lines, while scatter plots can be used to reflect the similarity between two different gels. Scatter plots are often used in microarray data analysis to reflect the similarity between two different samples. Here we also employed this method in our proteomic data analysis which can help us to find the difference between different lines at the same stages (Fig 3.18). We found the great difference between 9311 and hybrid at 6DAF and between 9311 and PA64S at 6DAF. Comparing with the 2D spot patterns of the amyloplast proteins in more detail revealed that the expression of metabolism proteins especially GBSS in PA64S and hybrid were earlier than in 9311, suggesting the metabolism activities such as starch synthesis in PA64S and hybrid were earlier than in 9311. Also, the patterns of expression profile of the hybrid proteins are quite similar to those of the PA64S (e.g. GBSS and glutelin), especially at the early developing stages, again suggesting that the maternal genes are dominant in the hybridization.

4.1.5 New methods introduced to study post translational modifications

Both the Pro-Q Diamond phosphoprotein gel stain and Pro-Q Emerald 300 glycoprotein gel stain are breakthrough technology that provides a simple method for selectively staining phosphoproteins and glycoproteins directly in polyacrylamide gels. There is no need to label proteins with radioisotopes, transfer the proteins to blotting membrane or use specific antibodies. The procedure is thus very simple and fast.

Pro-Q Diamond phosphoprotein gel stain detects phosphate groups attached to tyrosine, serine or threonine residues. The sensitivity limit ranges from ~1-16 ng of phosphoprotein. Pro-Q Emerald stain detects as little as 1 ng/band of glycoprotein and is ~50-fold more sensitive than the standard periodic acid-Schiff base method using acidic fuchsin dye and

at least as sensitive as using biotin hydrazide with streptavidin-horseradish peroxidase and ECL detection. In addition, Pro-Q Emerald stain detects glycoproteins more easily and reliably than mobility-shift assays, which only detect glycoproteins susceptible to specific deglycosylating enzymes and which can only be used in 1-D gels. The detection of phosphoprotein and glycoprotein with these stains can be used by UV-light sources. Directly superimposing phosphorylation and glycosylation images with the total-protein stain image allows unambiguous match of protein bands or spots. So quantitate changes in expression and post-translational modification per protein can be observed.

Using these two staining methods, we found that some of the amyloplasts proteins are phosphoproteins or glycoproteins, such as GBSS, SSS and SBE. Also after comparing the expression level of phosphoproteins or glycoproteins with the expression level of this protein, we found that changes occur at the post translational level during seed development.

4.2 Characteristics of rice amyloplast proteins

Rice grain filling is a highly coordinated process. The synthesis and transport of carbohydrates are not only coordinated at the transcriptional level, but also at the post translational level. The proteomic methods used in this study, along with the complete genome information, provide a powerful tool for studying and understanding functions and interactions between proteins.

Of the 166 amyloplast proteins, the most abundant category was classified as being involved in metabolism (48.8%). A variety of enzymes associated with metabolism were identified. Since starch is synthesized in endosperm amyloplasts, the proteome of amyloplasts obviously should relate to starch synthesis. Indeed, 79% of the amyloplast metabolism proteins are starch synthesis related. We also found that in amyloplast, 17.8% belongs to the high abundant proteins while 82.2% belongs to the low abundant proteins. The most abundant proteins correspond to metabolism proteins (43.8%) and storage proteins (50.0%), while the hypothetical proteins are only 6.2%. The low abundant proteins which including hypothetical protein/unknown (25%), metabolism (37.2%), protein destination (15.5%), retrotransposon and plastid protein (4.1%), protein synthesis (3.4%), cellular communication (0.7%), energy (5.4%), transport facilitation (4.7%) and

transcription (0.7%) could play an important role in protein or starch deposition, in transport facilitation, in maintaining cell structure during reserve deposition, or in the transition from a highly active to a quiescent state during seed development.

4.2.1 Amyloplast proteins associated with starch granules

During protein extraction, we found that some of the amyloplast proteins are tightly associated with starch granules, while some are loosely attached. Several studies indicated that amyloplast proteins contain these two types of proteins (Mu-Forster *et al.*, 1996; Mu-Forster *et al.*, 1998; Darlington *et al.*, 2000; Baldwin, 2001). We have tried using SDS buffer to wash the starch granules for eight times, but tightly associated proteins are still attached to the starch granules. In order to obtain the whole amyloplast proteins, 8M urea buffer was used to solubilize the starch granules in 2DE while heating was used in 1DE.

4.2.2 Most proteins in amyloplast proteome contain the transit peptide

When searching the database with PMF or MS/MS data, it is possible to use the pI and Mw information of the proteins from the 2D gels to reduce the number of candidates searched in the database. We found that the experimental Mw of several proteins did not match the calculated Mw when searching the database. We cross-correlated the calculated and experimental values for Mw and found that the experimental Mw of most proteins were lower or correspond to the calculated Mw (Fig 3.7). Using this method, we found that of the 166 proteins identified, 63.9% of the amyloplast proteins contain transit peptides as predicted by ChloroP software, 26.5% with signal peptides and 12.6% without transit peptides and signal peptides. 3.0% of the proteins were predicted to contain both transit peptides and signal peptides.

This may be due to several reasons. One is that the proteins may possess the transit peptide targeting to the amyloplasts. So the calculated Mw in the database is the Mw of precursor proteins, while the experimental Mw in the 2D gels is the Mw of mature proteins whose transit peptides have been cleaved. For example, GBSS contains a 77-amino acid transit peptide. The calculated Mw of GBSS is 66kDa. After subtracting the Mw of the transit peptide, the calculated Mw of mature GBSS is 58kDa, which matches the experimental mature protein (58kDa). Cross-correlation between the experimental and calculated Mw and pI generally shows good agreement between experimentation and theory after removal

of the presequences in our samples. Of course an incorrect prediction of plastid proteins can be caused by the presence of non-canonical transit sequences (reviewed by Schleiff *et al.*, 2001; Miras *et al.*, 2002). Dual targeting of proteins to plastids and other subcellular compartments has been reported for several proteins, further compounding the problem in predicting plastid targeting (Silva-Filho, 2003; Kiessling *et al.*, 2004; Wall *et al.*, 2004; review by Peeters *et al.*, 2001).

The second reason is that the protein spots on 2D gels are the subunits or fragments of some proteins, such as AGPase large subunit, glutelin acidic subunit and glutelin basic subunit.

A third reason is that if there is a significant mistake in gene assignment, or if there is a post-translational modification or processing event, these data points will lie outside the correlation and highlight such modifications or annotation errors. For example, in the research of Peltier *et al.* (2002), they found a gene annotation error. After reannotation of this gene by using cross-correlation plots, they found a very large difference between the predicted and the experimental pI value. Cross-correlation plots can help to identify annotation errors that affect pI or Mw as well as unexpected splicing events.

As amyloplasts have a double membrane envelope and a complex of internal membranes, transit peptides are required to target proteins to the amyloplasts. Our results support our identified proteins as authentic amyloplast proteins.

4.2.3 Multiple isoforms of starch synthesis related proteins

During the protein identification, we observed that there are several isoforms of starch synthesis related enzymes and these isoforms may have different functions. Zhu *et al.* (2003) found that most genes in the starch biosynthetic pathway show multiple distinct spatial and temporal expression patterns, suggesting that different isoforms of a given enzyme may be expressed in different tissues and at different developmental stages. In our study by examining protein expression in the amyloplast proteome, we confirmed the presence of different protein isoforms of the same enzyme in the starch synthesis pathway during grain development. The different expression of various isoforms in this pathway suggests their functional differences, including the enzymes GBSS, soluble starch synthase,

AGPase, sucrose synthase, sucrose phosphate synthase. Our studies also showed that the multiple forms of starch biosynthesis enzymes are functioning in hybrid rice and their parental lines. During seed formation, their expression patterns vary at time and level, exhibiting a complex network of protein actions and interactions.

4.2.3.1 Multiple spots of GBSS

GBSS is one of the most important proteins in starch synthesis pathway by its synthesis of amylose (Nelson *et al.*, 1962; Shure *et al.*, 1983). Previous studies have showed that GBSS protein is of multiple spots in both 1D and 2D gels (Boren *et al.*, 2004; Lin *et al.*, 2005). The origin of these products was unknown. In our study we found that GBSS protein has multiple spots with different molecular weights and different pI points not only in 9311 developing seeds at 10DAF, but all the three lines at four developing stages with similar patterns. GBSS is encoded only by the *Waxy (Wx)* gene. The question may be raised that the fragments may be caused by degradation during sample preparation. We have performed experiments to test this possibility. Amyloplasts proteins were extracted by adding protease inhibitor cocktail or not. Results (Fig 3.9) showed that the 2D images were similar regardless of adding protease inhibitors or not. The multiple GBSS spots were highly reproducible in number and relative abundance, rendering it unlikely that the multiple spots seen here are artifacts of protein damage during sample preparation. Previous research (Mu *et al.*, 1998) had mentioned that heat-induced fragmentation may exist in the maize waxy protein during protein extraction from starch granules. Due to the specific characteristics of amyloplasts, GBSS is tightly associated with the starch granules. So it requires heating the starch granules to release the GBSS. Although the 2D extraction method does not heat the starch granules, 8M urea in the 2D extraction buffer also leads to the solubilization of starch granules. If the GBSS binds to the starch at multiple domains, granule expansion could place added physical stress on this polypeptide during granule gelatinization, leading to fragmentation at multiple sites. Since multiple spots are highly reproducible in number and relative abundance, the interaction between GBSS and starch may be at specific sites. So it could be speculated that the GBSS has more starch-binding domains and greater binding affinity for starch than either the SBE or SSS proteins. It will, therefore, be important to determine whether the mechanism by which the GBSS protein interacts with the starch matrix is unique relative to SBE and SSS. Identification of the starch-binding domains for each of these granule associated polypeptides should provide

information to better understand the binding interactions between starch and proteins and the effects of such interactions on starch functionality.

The need to classify these granule associated GBSS protein spots prompted us to investigate the identity of these granule intrinsic polypeptides and the nature of their association with the starch granules. Our approach was first to purify starch granule associated proteins and then to identify them by analyzing their amino acid sequence. Using the ChloroP prediction algorithm, we predicted the transit peptide cleavage site of GBSS to be 76 amino acids long, yielding a processed GBSS protein of around 60 kDa. The estimated length of the GBSS transit peptide is in good agreement with our N-terminal sequencing results indicating that the transit peptide cleavage site is at the 77th amino acid. The amino acid motif KTGGL (Ainsworth *et al.*, 1993; Baba *et al.*, 1993), the putative binding site for ADPGlc (Furukawa *et al.*, 1990), is located 19 amino acids from the N-terminus of mature GBSS (Nakamura *et al.*, 1998). We observed that the N-termini of the 55kDa, 50 kDa and 28 kDa polypeptides are of complete identity to the full length N-terminus of the 60KD GBSS protein. For polypeptides around 45 kDa and 35 kDa, their N-termini are different and they do not contain the putative binding site for ADPGlc, suggesting that the GBSS fragments around 45 kDa and 35 kDa may have lost their function in starch synthesis.

We had further extensively studied the multiple spots of GBSS in detail. During seed development in all the three lines, both the GBSS spot number and total expression level decreased, indicating that the function of this protein to synthesis amylose began at the early stage and its activity gradually declined. Compared the changes in spot number and expression level among the three lines (Fig 3.19), we observed that both spot number and expression level in PA64S are the highest, while 9311 is the lowest and significantly lower than the PA64S, and the hybrid is in between, slightly lower than the PA64S. This result is consistent with the mRNA expression level and amylose content (Duan *et al.*, 2005). Previous researches revealed that amylose content in rice is correlated with its ability to excise intron 1 from the leader sequence of the *Wx* transcripts (Wang *et al.*, 1995). Because of a point mutation, excision of intron 1 in 9311 is unefficient, leading to low level of mature *Wx* mRNA, GBSS protein expression, and amylose in 9311 than the other two lines. According to the analysis performed by HHRRC, the amylose content of 9311 is 14.9%, which is the lowest among the three lines, while the amylose content of hybrid is 21.4%

and PA64S is 22.3%. Our proteome results thus are coincided with the low level of mature *Wx* mRNA and amylose content in 9311.

We have also analyzed the expression pattern for each GBSS spot in detail using cluster analysis (data not shown). Although for spots, the expression patterns are a little different, however, 83.3% (9311), 85.7% (hybrid) and 83.3% (PA64S) of GBSS spots among the three lines share similar decreased pattern during seed development, which is also consistent with the total GBSS expression pattern.

PTM is also a very important factor in causing multiple spots of GBSS. We have profiled the expression of GBSS by 2DE with PTM staining methods. From the results (Fig 3.23) we found that GBSS not only have been glycosylated and phosphorylated, but at different developing stages, the glycosylation and phosphorylation levels are different, which indicate that PTMs may also regulate the GBSS expression and its function in amylose synthesis.

In our study, we have analyzed the relative phosphorylation and glycosylation levels of GBSS (Fig 3.24). Both the phosphorylation and glycosylation levels decrease, and the patterns of phosphorylated and glycosylated levels in 9311, hybrid and PA64S are similar to the protein expression level. Phosphorylation in GBSS takes place mainly at early seed developmental stages, indicating that active forms of GBSS may appear at early stages with the highest activity, while the activity decreases during maturation. These results are consistent with our findings in the changes of spot number and expression level of GBSS during development. The difference in phosphorylation level among the 9311, hybrid and PA64S are also consistent with the difference among their protein expression levels.

4.2.4 Expression patterns of amyloplast proteome

Super hybrid rice combination LYP9 (PA64S/9311) was developed in 1998 by Hunan Hybrid Rice Research Center (HHRRC) and Jiangsu Agricultural Academy in China. This rice variety has a yield of average 12 t/ha in a large-scale production but its grain quality only reaches second-grade or even third-grade in certain traits. Hence, research on protein expression during grain development for this combination is necessary and significant in improving its grain quality and hopefully modified cultivars with super high yield and

excellent grain quality can be developed in the future.

We have profiled the amyloplast proteome during seed development in LYP9 and its parental lines 9311 and PA64S (Fig 3.12). The following trends can be observed.

1. The total number of amyloplast proteins on 2-D gels gradually increases in all three hybrid rice lines during seed development (Fig 3.13). The functional classification results (Fig 3.14) at each developing stages in each rice lines showed that metabolism proteins are the most abundant category. The spot number (%) in metabolism proteins is the highest, more than the proteins with other functions at each stage, but gradually decreases through the four stages. The patterns in all three lines are similar, while the spot number (%) in 9311 is generally lower than those in PA64S and the hybrid, suggesting a relatively lower concentration of proteins/enzymes and metabolic activities in 9311.

2. According to the cluster analysis of protein expression patterns, we observed several trends of protein expression pattern in the three rice lines, including three major expression patterns, i.e. the increase, decrease and other pattern (Fig 3.16).

a. The three expression patterns, however, do not differentiate among three rice lines. The decrease pattern in 9311 (46.5%) is the same as the hybrid (46.5%), although a little higher than PA64S (42.0%). The increase pattern in 9311 (37.6%) is a little higher than the hybrid (36.3%) and PA64S (35.7%). The other patterns in 9311 is 15.9%, in PA64S, 22.3% and in hybrid, 17.2%.

b. Most of the metabolism proteins in the three rice lines are of the decrease pattern. The decrease pattern of PA64S is of 72.6%, a little higher than the 9311 (68.3%) and hybrid (69.4%).

c. The decrease pattern covers the most abundant protein category in all the three lines. For instance, the enzymes of starch synthesis generally show the highest expression level at early developing stage but gradually decrease, such as the GBSS, soluble starch synthase and ADP-glucose pyrophosphorylase large subunit. There are other function proteins with the decrease pattern, such as the ribosomal proteins, putative sugar transporter. The increase pattern observed for α-amylase during development is different, since during

cereal seed germination, α-amylase plays an important role in hydrolyzing the endosperm starch into metabolizable sugars, providing the energy for the growth of roots and shoots (Akazawa *et al.*, 1985; Beck *et al.*, 1989).

From the clustering analysis, most of the proteins with the similar functions commonly share similar expression patterns; thus identified proteins with unknown functions can be related to other proteins whose functions have been determined.

3. Scatter plot analysis (Fig 3.18) between two different gels at the same developing stage revealed that the patterns of expression profile of hybrid proteins are quite similar to there of the PA64S (e.g. GBSS and glutelin), suggesting the dominance of the maternal genes in the hybrids.

4.2.5 Post translational modifications potentially regulate starch synthesis

Protein kinase cascades play essential roles in diverse intracellular signaling processes in animals and yeast. In plants, there is evidence that protein phosphorylation plays an important role in signaling pathways triggered by abiotic stress, pathogen invasion, and plant hormones (Knetsch *et al.*, 1996; Sheen, 1996; Zhang *et al.*, 1998). Despite the importance and widespread occurrence of protein phosphorylation in plants, little is known about the factors that regulate starch biosynthesis. A recent study with isolated amyloplasts from wheat endosperm identified a number of phosphoproteins, including some involved in starch metabolism, indicating that some aspects of starch (amylopectin) biosynthesis may be controlled by protein phosphorylation (Tetlow *et al.*, 2004b). In this study, we detected the phosphorylation on some of the starch synthesis related proteins, such as GBSS (and its fragments), SSS and SBE (Fig 3.21), but how phosphorylation regulates these proteins needs further investigation. Nevertheless, in this study, we observed wide occurrence phosphorylation of starch synthesis related proteins than before.

Combination of the 2DE and PTM staining methods can help us to identify which protein has the PTMs, and in turn allows us to study their expression profiles. Although we did not know the exact function of phosphorylation (and glycosylation) on GBSS, it is the first time to profile proteins with PTM and to study their PTM level. A few PTM studies in

starch synthesis related proteins had only been reported recently (Tetlow *et al.*, 2004b; Lin *et al.*, 2005).

Most of the glycoproteins will go through the secretory pathway for their glycosylation, but some cytosolic or nucleus proteins can be glycosylated (see literature review). Both the GBSS and SSS are nuclear-encoded proteins and contain transit peptides, while GBSS is a glycoprotein and SSS is not. This result indicates that the synthesis of GBSS and SSS may go through different pathways. The SSS may be directly targeted to the amyloplast. For GBSS, the situation is more complicated than SSS since we do not know the glycosylation type of GBSS. If GBSS is O-GlcNAc glycosylation, it could be glycosylated on the way to the amyloplast, maybe in the cytosol. If not, it may go through the secretory pathway. How does the GBSS go through the ER-Golgi complex without a signal peptide and how does the GBSS target to the amyloplast after going through the ER-Golgi complex? Research (Sulli *et al.*, 1999) found that the Euglena chloroplast protein precursors are transported as integral membrane proteins from the endoplasmic reticulum (ER) to the Golgi apparatus prior to chloroplast localization. All Euglena chloroplast protein precursors have functionally similar bipartite presequences composed of an N-terminal signal peptide domain and a stromal targeting domain containing a hydrophobic region approximately 60 amino acids from the predicted signal peptidase cleavage site. However, GBSS only contains a transit peptide and no signal peptide as predicted by SignalP. Further study on the glycosylation of GBSS is thus of interest. The first step is to determine the glycosylation type of GBSS, for example, by the new commercial western kit to detect O-GlcNAc glycosylation, and then to determine the glycosylation sites by MS.

4.3 Other characteristic aspects of amyloplast proteome

4.3.1 Comparison between the rice and wheat amyloplast proteomes

Andon *et al.*, (2002) has previously studied the wheat amyloplast proteome by 2D and 1D SDS-PAGE, followed by LC-MS/MS. In the research, they identified 171 proteins, of which 108 were from whole amyloplasts and 63 from purified amyloplast membranes. The identified proteins were classified into seven categories. In the whole amyloplast, 37% belongs to unknown/hypothetical, 6% to ribosomal/transcription regulation, 23% to energy metabolism, 0% to transporter proteins, 3% to receporter/signaling, 25% to protein destination/storage, and 6% to disease/defense. In the amyloplast membrane proteins, 25%

belongs to unknown/hypothetical, 24% to ribosomal/transcription regulation, 21% to energy metabolism, 5% to transporter proteins, 6% to receptor/signaling, 19% to protein destination/storage, and 0% to disease/defense. Our results show some differences from those of the wheat amyloplasts. In our study the most abundant identified proteins are metabolism proteins (48.8%), while in wheat amyloplasts the most abundant category is hypothetical proteins (37%). One of the reasons causing this difference could be the database. Since wheat lacks genomic information, the authors combined the available genomic sequences of wheat and other cereal sequences to do the database searching. The author directly used the de novo sequence to search the database; but some of the identified proteins in wheat are different from other cereal proteins, although the sequences may be partially the same.

We identified several isoforms of proteins in our samples, such as multiple spots of GBSS, while in wheat amyloplasts, only one band of GBSS was identified in 1D gel. But from the reported 2D SDS-PAGE of wheat amyloplasts, we observed similar patterns of GBSS around 60kDa, as in our rice amyloplasts, however, the authors did not report their identification results of the proteins in this region. We found that GBSS is the predominant protein in this region.

4.3.2 Proteomic comparisons among the three rice lines

Although the morphology of hybrid seeds is more similar to the 9311, several aspects from this study reveal that the hybrid protein expression pattern is quite similar to that of PA64S, suggesting maternal genes are notably dominant in the hybrid. For example, the scatter plots reveal that there is more similarity between the hybrid and PA64S than between the hybrid and 9311 in terms of expression pattern, especially at early developing stages. The expression level of GBSS in the hybrid line during seed development is also more similar to that of the PA64S than to the 9311. Although the patterns of phosphorylated and glycosylated levels in 9311, hybrid and PA64S are similar, the relative phosphorylation and glycosylation levels in 9311 are lower than those in the hybrid and PA64S.

4.3.3 Comparison of starch synthesis enzymes at protein and transcript levels

The expression profiles of the five major genes encoding the starch biosynthesis enzymes and their isoforms, including AGPase, GBSS, SSS, SBE and DBE, a total of 15, during rice grain filling had been also systematically carried out at the transcript level (Duan, 2003; Duan *et al.*, 2005). During grain filling, the two AGPase genes were active in the seeds at early developmental stage (3DAF). The expression of the genes encoding GBSS, the enzyme responsible for the synthesis of straight chain amylose in endosperm, was initiated at 3DAF and reached a peak at 6DAF. For the three major enzymes involving in the synthesis of amylopectin, namely SBE, SSS and SDE, the gene encoding the two members of SDE showed similar temporal expression patterns, with low level at 3DAF, peaking at 10DAF, and then declining to undetectable level at 20DAF. Similar studies were also carried out by another research group (Ohdan *et al.*, 2005) in which comprehensive analysis of transcript level of 27 rice genes encoding six classes of starch synthesis enzymes was carried out by quantitative real-time RT-PCR. They profiled not only AGPase, GBSS, SSS, SBE and DBE, but also SP and DPE which are considered as playing partial roles in starch biosynthesis. All of these genes were active at the early developing stages and gradually decreased during development. From our results, the expression profiles of the proteins of AGPase, GBSS, SSS and SBE at the protein level are highly consistent with the expression profiles at the transcripts level. At the transcripts level, the amount of GBSS transcript was markedly higher than any of the SS genes, with its maximum level being more than 3-fold higher than that of the most vigorously expressed SS gene (SSIIa) and approximately 55-fold over the least-expressed SS gene (SSIVa) (Ohdan *et al.*, 2005). In our results, at the protein level, the maximum level of GBSS is about 145-fold higher than that of the SSS. The expression difference between GBSS and SSS at the protein level is thus dramatically greater than at the transcript level. This may suggest that GBSS continuously accumulates and plays major role in the synthesis amylose during the whole development. Also tightly bounding to the starch granules may stabilize and prevent the degradation of GBSS. The protein expression pattern of SP, however, is quite different with the pattern at the transcript level. From Ohdan's results the SP transcripts were numerous at the onset of seed development, rapidly increased to peak at 5 DAF, and diminished to a low but significant level until 15 DAF (Ohdan *et al.*, 2005). This may suggest that although the transcripts level of SP is decreasing during development, the protein is still accumulated and not degraded. It also suggests that SP may not only involve in starch synthesis, but also in starch degradation.

4.3.4 Comparison of the starch synthesis related proteins among the three rice lines

In our study, we have not only systematically profiled the protein expression patterns of starch synthesis related proteins, but also made a comparison among the three different rice lines, 9311, PA64S and the hybrid. We found that only the GBSS expression level shows great difference among these three rice lines. The GBSS expression level in 9311 is lower than those of the PA64S and the hybrid. Our observations are highly consistent with previous studies (Duan, 2003; Duan *et al.*, 2005). In their studies, they systematically studied the expression profiles of 15 genes in starch synthesis, including five major genes encoding the AGPase, GBSS, SSS, SBE, and DBE among the three rice lines 9311, PA64S and their hybrid. In their results, except for the GBSS, the transcript levels of the other four major genes were similar and no great difference was found among the three rice lines. For the GBSS, its transcript levels in PA64S and the hybrid were significantly higher than that of the 9311 through the whole seed development. Thus at both protein and transcript levels, their patterns and amounts revealed that the great difference in starch quality among the three rice lines is mainly caused by the expression of GBSS. The amylose content of 9311 has been determined as 14.9%, hybrid, 21.4% and PA64S, 22.3% (Duan, 2003). Unefficient excision of intron 1 in 9311(Wang *et al.*, 1995) has been shown leading to the low level of mature *Wx* mRNA, which in turn results in lower amylose content in 9311 than in the other two lines. In our research, we observed that both of the spot number and expression level of GBSS in PA64S are the highest, while 9311 is the lowest and significantly lower than that of the PA64S. The GBSS protein level in the hybrid, on the other hand, is slightly lower than that of the PA64S. Our results thus reveal that the protein expression level of GBSS is correlated with its mRNA expression level and in turn the amylose content.

4.4 Limitations of proteomic approach in directly answering the question on how to improve eating and cooling quality

The ultimate goal of our project is to improve the eating and cooking property of super hybrid rice and its further yield, which are closely related to starch property and quality. Starch synthesis is through a very complicated pathway. The present results collectively provide clear and concrete evidence that the expressions of the numerous proteins involved in starch metabolism are highly co-ordinated. However, the following aspects cannot

be/have not been able to adequately addressed in our proteomic research. For example, the flow of metabolites from outside into the amyloplast; the partition of carbon between sucrose and starch; the catabolism of starch; the influence of post translational process other than glycosylation and phosphorylation; and the effect due to uneven distribution of GBSS in starch granules.

Proteomics is an exciting new approach to biological and biomedical research, and recent technology developments have allowed us to undertake comprehensive analyses of proteomes. The massive amount of data generated in these experimental studies requires much effort to translate into biological knowledge. The greatest limitations in proteomic research thus are translating the massive amount of data into the biological knowledge and linking it with the metabolic pathway. How to use the information obtained in our proteomic research to answer the question to improve eating and cooking quality certainly need further investigation.

Chapter 5 Conclusion

To fully elucidate the mechanism and regulation of starch biosynthesis, we need to understand the function of individual enzymes in the pathway as well as the interactions of these enzymes and their multiple forms in starch synthesis. Proteomics provides us a unique opportunity to study this complicated metabolic pathway. Our study adds a new set of data crucial in our understanding of the molecular mechanisms involved in the starch synthesis regulatory networks of cultivated and possible in the heterosis of hybrid rice.

In our study we systematically determined the protein components in rice amyloplasts where storage starch is synthesized and try to elucidate their roles in the starch synthesis pathway. We selected hybrid rice line 9311 at 10DAF as the model material to study the amyloplast proteome. Of the 166 proteins identified by MS/MS, the most abundant category (48.8%) was classified as being involved in metabolism. A variety of enzymes associated with metabolism were identified. In agreement with the fact that starch is synthesized in endosperm amyloplasts, 79% of the metabolism proteins were identified as starch synthesis related proteins. We also found that in the amyloplast proteins, 17.8% belongs to the high abundant proteins and 82.2% belongs to the low abundant proteins. As expected, the most abundant proteins are largely (43.8%) in the metabolism proteins category. By identifying the proteins expressed in the amyloplast proteome, we confirmed the presence of different isoforms of the same enzyme in the starch synthesis pathway during grain development. The differences in expression of the various isoforms in this pathway during seed development suggest their functional differences, including GBSS, soluble starch synthase, AGPase, sucrose synthase, and sucrose phosphate synthase.

The hybrid (LYP9) of parental lines 9311 and PA64S displays significant yield advantage. To explore the possible refection of such advantage in the proteome of amyloplasts where the most abundant rice grain component, starch, is synthesized and stored, we focused our attention on the hybrid and its parental lines. We profiled their amyloplast proteomes during seed formation. The aim is to analyze the expression of proteins related to starch synthesis in the amyloplasts of the parental and the hybrid generations. The maturing seeds of 6, 10, 15 and 20 days after flowering (DAF) were collected and their proteomes during

seed development were studied. The total numbers of amyloplast proteins on 2-D gels are found to gradually increase in all three lines of rice during seed development. The functional classification results of each rice line at different developing stages showed that metabolism proteins are the most abundant category and gradually decreased during four stages. The patterns of metabolism proteins in all three lines are similar, while in 9311 it is normally lower than in PA64S and hybrid. Employed the SOM method, we have selected 157 groups of spots in three rice lines with four developing stages to do the cluster analysis. We have found several trends of protein expression pattern in three rice lines. There are three major expression patterns in three rice lines, increase, decrease and other pattern. The three expression patterns do not have great difference in three rice lines. The decrease pattern is the most abundant category in all three lines. Most of metabolism proteins in each line are decrease pattern. Although the morphology of hybrid seed is more similar with 9311, several aspects reveal that hybrid proteins expression is quite similar to that of PA64S, suggesting the maternal genes involved in determining the amyloplast proteome are dominant during hybrid seed development.

A complete functional understanding of the proteome requires, however, full characterization of the post translational modifications of proteins and the complex networks of protein-protein interactions. In our study, we were the first to carry out the expression profiling of proteins with post translational modifications during seed development by the recently developed PTM staining techniques in combination with 2D SDS-PAGE. We detected phosphorylation on some of the starch synthesis related proteins, such as GBSS (and its fragments), SSS and SBE, suggesting phosphorylation is likely to play a role in regulation starch synthesis, but the mechanism of the regulation needs further investigation. Glycosylation was also detected in GBSS. These post translational modifications thus may be a very important cause of the multiple spots of GBSS that we observed. Further more, we found that GBSS not only was glycosylated and phosphorylated, but at different developing stages the glycosylation and phosphorylation levels were different, indicating that post translational modifications may regulate GBSS expression and its function in amylose synthesis.

Chapter 6 Future perspectives

The ultimate goal of super hybrid rice project in our lab is to improve the eating and cooking property of super hybrid rice and its further yield, which are closely related to starch property and quality. Proteomics is an exciting new approach to biological and biomedical research, and has rapidly grown into a major research and commercial enterprise with strong prospects of dramatically advancing our knowledge of basic biological processes. However, before we can manipulate the expression of selected candidate proteins by molecular biotechnology to achieve these goals, further information and studies are required for efficient genetic engineering.

First of all, we have set up 2D reference maps of rice amyloplasts in three rice lines at four developing stages by 2-D gel electrophoresis followed by mass spectrometry, which is a systematic platform to identify organelle proteome and analyze quantitative proteomics. Although we have identified many new proteins in amyloplasts, the number is far from covering the expressed proteins in amyloplasts. For example, brittle-1 protein which is an adenylate translocator facilitating transfer of extraplastidial synthesized ADP-Glucose into amyloplasts, and ferredoxin, ferredoxin-Trx reductase, and Trx which are the components of the ferredoxin/Trx system originally described in chloroplast were not found in our collections. To increase the capacity, more proteomic platform should be set up such as LC/MSMS, which can give great helps to find more components involved in the starch synthesis.

Secondly, improving eating and cooking quality involved fine-tuning of metabolic pathway. It is very likely that the metabolic flow is a key player in this case. There may not be any single magic regulator that can optimize the ratio of amylose: amylopectin for eating and cooking quality improvement. Further investigation is thus to set up a metabolic pathway network which can link the single proteins together and find relationships among each proteins. With the help of profiling protein expressions patterns, we found that normally proteins with similar functions also share similar expression patterns, which gives us an insight into set up the metabolic network and more information on the regulatory mechanisms.

Finally, post translational modification of amyloplast proteins is a very important regulatory mechanism in starch synthesis regulation. During our study, we found several proteins have the phosphorylation and glycosylation. How they employ these regulatory mechanisms in the starch synthesis is still unknown. More interesting is that we have identified one of our amyloplast proteins GBSS has O-GlcNAc (data not shown). Most of the glycoproteins will go through the secretory pathway for their glycosylation, but proteins with O-GlcNAc are cytosolic or nucleus proteins. Since GBSS is O-GlcNAc glycosylation, it could be glycosylated on the way to the amyloplast or in the amyloplast. How and where GBSS is glycosylated is of great interest. If we can detect the glycosylation sites and elucidate the GBSS transporting pathway, it will provide further and important information on the regulation of starch synthesis.

References

Adoration, P.R., Li, X.L., Okita, T.W., and Juliano, B.O. (1993). Characterization of poorly digested protein of cooked rice protein bodies. *Cereal Chem* 70: 101-105.

Ainsworth, C., Clark, J., and Balsdon, J. (1993). Expression, organisation and structure of the genes encoding the waxy protein (granule-bound starch synthase) in wheat. *Plant Mol Biol* 22: 67-82.

Akazawa, T., and Hara-Mishimura, I. (1985). Topographic aspects of biosynthesis, extracellular section and intracellular storage of proteins in plant cells. *Annu Rev Plant Physiol* 70: 441-472.

Akihiro, T., Mizuno, K., and Fujimura, T. (2005). Gene expression of ADP-glucose pyrophosphorylase and starch contents in rice cultured cells are cooperatively regulated by sucrose and ABA. *Plant Cell Physiol* 46: 937-946.

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

Andon, N.L., Hollingworth, S., Koller, A., Greenland, A.J., Yates, J.R., 3rd, and Haynes, P.A. (2002). Proteomic characterization of wheat amyloplasts using identification of proteins by tandem mass spectrometry. *Proteomics* 2: 1156-1168.

Apweiler, R., Hermjakob, H., and Sharon, N. (1999). On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim Biophys Acta* 1473: 4-8.

Arenas-Huertero, F., Arroyo, A., Zhou, L., Sheen, J., and Leon, P. (2000). Analysis of Arabidopsis glucose insensitive mutants, gin5 and gin6, reveals a central role of the plant hormone ABA in the regulation of plant vegetative development by sugar. *Genes Dev* 14: 2085-2096.

Baba, T., Kimura, K., Mizuno, K., Etoh, H., Ishida, Y., Shida, O., and Arai, Y. (1991). Sequence conservation of the catalytic regions of amylolytic enzymes in maize branching enzyme-I. *Biochem Biophys Res Commun* 181: 87-94.

Baba, T., Nishihara, M., Mizuno, K., Kawasaki, T., Shimada, H., Kobayashi, E., Ohnishi, S., Tanaka, K., and Arai, Y. (1993). Identification, cDNA cloning, and gene expression of soluble starch synthase in rice (Oryza sativa L.) immature seeds. *Plant Physiol* 103: 565-573.

Baldwin, P.M. (2001). Starch Granule-Associated Proteins and Polypeptides: A Review. *Starch* 53: 475-503.

Ball, S., Guan, H.P., James, M., Myers, A., Keeling, P., Mouille, G., Buleon, A., Colonna, P., and Preiss, J. (1996). From glycogen to amylopectin: a model for the biogenesis of the plant starch granule. *Cell* 86: 349-352.

Ball, S.G., and Morell, M.K. (2003). From bacterial glycogen to starch: understanding the biogenesis of the plant starch granule. *Annu Rev Plant Biol* 54: 207-233.

Balmer, Y., Vensel, W.H., Cai, N., Manieri, W., Schurmann, P., Hurkman, W.J., and Buchanan, B.B. (2006). A complete ferredoxin/thioredoxin system regulates fundamental processes in amyloplasts. *Proc Natl Acad Sci U S A.*

Balmer, Y., Vensel, W.H., Tanaka, C.K., Hurkman, W.J., Gelhaye, E., Rouhier, N., Jacquot, J.P., Manieri, W., Schurmann, P., Droux, M., and Buchanan, B.B. (2004). Thioredoxin links redox to the regulation of fundamental processes of plant mitochondria. *Proc Natl Acad Sci U S A* 101: 2642-2647.

Bardel, J., Louwagie, M., Jaquinod, M., Jourdain, A., Luche, S., Rabilloud, T., Macherel, D., Garin, J., and Bourguignon, J. (2002). A survey of the plant mitochondrial proteome in relation to development. *Proteomics* 2: 880-898.

Beatty, M.K., Rahman, A., Cao, H., Woodman, W., Lee, M., Myers, A.M., and James, M.G. (1999). Purification and molecular genetic characterization of ZPU1, a pullulanase-type starch-debranching enzyme from maize. *Plant Physiol* 119: 255-266.

Beck, E., and Ziegler, P. (1989). Biosynthesis and Degradation of Starch in Higher Plants. *Annual Review of Plant Physiology and Plant Molecular Biology* 40: 95-117.

Beckles, D.M., Smith, A.M., and ap Rees, T. (2001). A Cytosolic ADP-Glucose Pyrophosphorylase Is a Feature of Graminaceous Endosperms, But Not of Other Starch-Storing Organs. *Plant Physiol.* 125: 818-827.

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.

Biemann, K. (1992). Mass Spectrometry of Peptides and Proteins. *Annual Review of Biochemistry* 61: 977-1010.

Blom, N., Sicheritz-Ponten, T., Gupta, R., Gammeltoft, S., and Brunak, S. (2004). Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* 4: 1633-1649.

Boren, M., Larsson, H., Falk, A., and Jansson, C. (2004). The barley starch granule proteome - internalized granule polypeptides of the mature endosperm. *Plant Science* 166: 617-626.

Borisjuk, L., Walenta, S., Rolletschek, H., Mueller-Klieser, W., Wobus, U., and Weber, H. (2002). Spatial analysis of plant metabolism: sucrose imaging within Vicia faba cotyledons reveals specific developmental patterns. *Plant J* 29: 521-530.

Burton, R.A., Bewley, J.D., Smith, A.M., Bhattacharyya, M.K., Tatge, H., Ring, S., Bull, V., Hamilton, W.D.O., and Martin, C. (1995). Starch branching enzymes belonging to distinct enzyme families are differentially expressed during pea embryo development. *Plant Journal* 7: 3-15.

Bykova, N.V., Egsgaard, H., and Moller, I.M. (2003a). Identification of 14 new phosphoproteins involved in important plant mitochondrial processes. *FEBS Lett* 540: 141-146.

Bykova, N.V., Stensballe, A., Egsgaard, H., Jensen, O.N., and Moller, I.M. (2003b). Phosphorylation of formate dehydrogenase in potato tuber mitochondria. *J Biol Chem* 278: 26021-26030.

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.

Cao, H., Imparl-Radosevich, J., Guan, H., Keeling, P.L., James, M.G., and Myers, A.M. (1999). Identification of the soluble starch synthase activities of maize endosperm. *Plant Physiol* 120: 205-216.

Chandran, D., Reinders, A., and Ward, J.M. (2003). Substrate specificity of the Arabidopsis thaliana sucrose transporter AtSUC2. *J Biol Chem* 278: 44320-44325.

Chattopadhyay, S., Ang, L.H., Puente, P., Deng, X.W., and Wei, N. (1998). Arabidopsis bZIP protein HY5 directly interacts with light-responsive promoters in mediating light control of gene expression. *Plant Cell* 10: 673-683.

Chen, H.M., Chang, S.C., Wu, C.C., Cuo, T.S., Wu, J.S., and Juang, R.H. (2002a). Regulation of the catalytic behaviour of L-form starch phosphorylase from sweet potato roots by proteolysis. *Physiol Plant* 114: 506-515.

Chen, W., Provart, N.J., Glazebrook, J., Katagiri, F., Chang, H.-S., Eulgem, T., Mauch, F., Luan, S., Zou, G., Whitham, S.A., Budworth, P.R., Tao, Y., Xie, Z., Chen, X., Lam, S., Kreps, J.A., Harper, J.F., Si-Ammour, A., Mauch-Mani, B., Heinlein, M., Kobayashi, K., Hohn, T., Dangl, J.L., Wang, X., and Zhu, T. (2002b). Expression Profile Matrix of Arabidopsis Transcription Factor Genes Suggests Their Putative Functions in Response to Environmental Stresses. *Plant Cell* 14: 559-574.

Choudhury, N.H., and Juliano, B.O. (1980). Lipids in developing and mature rice grain. *Phytochemistry* 19: 1063-1069.

Colleoni, C., Dauvill e, D., Mouille, G., Bul on, A., Gallant, D., Bouchet, B., Morell, M., Samuel, M., Delrue, B., d'Hulst, C., Bliard, C., Nuzillard, J.M., and Ball, S. (1999). Genetic and biochemical evidence for the involvement of alpha-1,4 glucanotransferases in amylopectin synthesis. *Plant Physiol* 120: 993-1004.

Comer, F.I., and Hart, G.W. (2000). O-Glycosylation of nuclear and cytosolic proteins. Dynamic interplay between O-GlcNAc and O-phosphate. *J Biol Chem* 275: 29179-29182.

Comer, F.I., Vosseller, K., Wells, L., Accavitti, M.A., and Hart, G.W. (2001). Characterization of a mouse monoclonal antibody specific for O-linked N-acetylglucosamine. *Anal Biochem* 293: 169-177.

Commuri, P.D., and Keeling, P.L. (2001). Chain-length specificities of maize starch synthase I enzyme: studies of glucan affinity and catalytic properties. *Plant J* 25: 475-486.

Cooper, C.A., Gasteiger, E., and Packer, N.H. (2001a). GlycoMod--a software tool for determining glycosylation compositions from mass spectrometric data. *Proteomics* 1: 340-349.

Cooper, C.A., Harrison, M.J., Wilkins, M.R., and Packer, N.H. (2001b). GlycoSuiteDB: a new curated relational database of glycoprotein glycan structures and their biological sources. *Nucleic Acids Res* 29: 332-335.

Cooper, C.A., Joshi, H.J., Harrison, M.J., Wilkins, M.R., and Packer, N.H. (2003). GlycoSuiteDB: a curated relational database of glycoprotein glycan structures and their biological sources. 2003 update. *Nucleic Acids Res* 31: 511-513.

Craig, J., Lloyd, J.R., Tomlinson, K., Barber, L., Edwards, A., Wang, T.L., Martin, C., Hedley, C.L., and Smith, A.M. (1998). Mutations in the Gene Encoding Starch Synthase II Profoundly Alter Amylopectin Structure in Pea Embr yos. *Plant Cell* 10: 413-426.

Cui, S., Huang, F., Wang, J., Ma, X., Cheng, Y., and Liu, J. (2005). A proteomic analysis of cold stress responses in rice seedlings. *Proteomics* 5: 3162-3172.

Dai, Z.Y., Zhao, B.H., Liu, X.J., Xia, G.H., Tan, C.L., Zhang, B.Q., and Zhang, H.X. (1997). 優質高産多抗中秈稻新品種揚稻 6 號. *江蘇農業科學* 4: 13-15.

Darlington, H.F., Tecsi, L., Harris, N., Griggs, D.L., Cantrell, I.C., and Shewry, P.R. (2000). Starch Granule Associated Proteins in Barley and Wheat *Journal of Cereal Science* 32: 21-29.

Dela Cruz, N., and Khush, G.S. (2000). Rice grain quality evaluation procedures. In Aromatic rices, R.K. Singh, U.S. Singh, and G.S. Khush, eds (Oxford & IBH publishing Co. Pvt. Ltd. New Delhi.), pp. 15-28.

Dela Cruz, N., Kumar, I., Kaushik, R.P., and Khush, G.S. (1989). Effect of temperature during grain development on stability of cooking quality component in rice. *Japanese J. Breed.* 39: 299-306.

Denyer, K., Dunlap, F., Thorbjornsen, T., Keeling, P., and Smith, A.M. (1996). The Major Form of ADP-Glucose Pyrophosphorylase in Maize Endosperm Is Extra-Plastidial. *Plant Physiol.* 112: 779-785.

Dickinson, D.B., and Preiss, J. (1969). ADP glucose pyrophosphorylase from maize endosperm. *Arch Biochem Biophys* 130: 119-128.

Dijkwel, P.P., Huijser, C., Weisbeek, P.J., Chua, N.H., and Smeekens, S.C. (1997). Sucrose control of phytochrome A signaling in Arabidopsis. *Plant Cell* 9: 583-595.

Dopazo, J., Zanders, E., Dragoni, I., Amphlett, G., and Falciani, F. (2001). Methods and approaches in the analysis of gene expression data. *Journal of Immunological Methods* 250: 93-112.

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., and Sun, S.S. (2005). Profiling the expression of genes controlling rice grain quality. *Plant Mol Biol* 59: 165-178.

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.

Dunlap, J.C. (1999). Molecular bases for circadian clocks. *Cell* 96: 271-290.

Edwards, A., Fulton, D.C., Hylton, C.M., Jobling, S.A., Gidley, M., Rossner, U., Martin, C., and Smith, A.M. (1999). A combined reduction in activity of starch synthases II and III of potato has novel effects on the starch of tubers. *The Plant Journal* 17: 251-261.

Emanuelsson, O., Nielsen, H., and von Heijne, G. (1999). ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci* 8: 978-984.

Emes, M.J., Bowsher, C.G., Hedley, C., Burrell, M.M., Scrase-Field, E.S.F., and Tetlow, I.J. (2003). Starch synthesis and carbon partitioning in developing endosperm. *J.Exp.Bot.* 54: 569-575.

Faye, L., and Chrispeels, M.J. (1989). Apparent Inhibition of Beta-Fructosidase Secretion by Tunicamycin May Be Explained by Breakdown of the Unglycosylated Protein During Secretion. *Plant Physiol.* 89: 845-851.

Ferro, M., Salvi, D., Riviere-Rolland, H., Vermat, T., Seigneurin-Berny, D., Grunwald, D., Garin, J., Joyard, J., and Rolland, N. (2002). Integral membrane proteins of the chloroplast envelope: identification and subcellular localization of new transporters. *Proc Natl Acad Sci* 99: 11487-11492.

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. *Mol Cell Proteomics* 2: 325-345.

Fialka, I., Pasquali, C., Lottspeich, F., Ahorn, H., and Huber, L.A. (1997). Subcellular fractionation of polarized epithelial cells and identification of organelle-specific proteins by two-dimensional gel electrophoresis. *Electrophoresis* 18: 2582-2590.

Finnie, C., Melchior, S., Roepstorff, P., and Svensson, B. (2002). Proteome Analysis of Grain Filling and Seed Maturation in Barley. *Plant Physiol.* 129: 1308-1319.

Fontaine, T., D'Hulst, C., Maddelein, M.L., Routier, F., Pepin, T.M., Decq, A., Wieruszeski, J.M., Delrue, B., Van den Koornhuyse, N., and Bossu et, a. (1993). Toward an understanding of the biogenesis of the starch granule. Evidence that Chlamydomonas soluble starch synthase II controls the synthesis of intermediate size glucans of amylopectin. *The Journal Of Biological Chemistry* 268: 16223-16230.

Friso, G., Giacomelli, L., Ytterberg, A.J., Peltier, J.B., Rudella, A., Sun, Q., and Wijk, K.J. (2004). In-depth analysis of the thylakoid membrane proteome of Arabidopsis thaliana

chloroplasts: new proteins, new functions, and a plastid proteome database. *Plant Cell* 16: 478-499.

Fu, Y., Ballicora, M.A., Leykam, J.F., and Preiss, J. (1998). Mechanism of reductive activation of potato tuber ADP-glucose pyrophosphorylase. *J Biol Chem* 273: 25045-25052.

Fujita, N., and Taira, T. (1998). A 56-kDa protein is a novel granule-bound starch synthase existing in the pericarps, aleurone layers, and embryos of immature seed in diploid wheat (Triticum monococcum L.). *Planta* 207: 125-132.

Fukao, Y., Hayashi, M., and Nishimura, M. (2002). Proteomic analysis of leaf peroxisomal proteins in greening cotyledons of Arabidopsis thaliana. *Plant Cell Physiol* 43: 689-696.

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

Furukawa, K., Tagaya, M., Inouye, M., Preiss, J., and Fukui, T. (1990). Identification of lysine 15 at the active site in Escherichia coli glycogen synthase. Conservation of Lys-X-Gly-Gly sequence in the bacterial and mammalian enzymes. *J Biol Chem* 265: 2086-2090.

Gallardo, K., Le Signor, C., Vandekerckhove, J., Thompson, R.D., and Burstin, J. (2003). Proteomics of Medicago truncatula Seed Development Establishes the Time Frame of Diverse Metabolic Processes Related to Reserve Accumulation. *Plant Physiol.* 133: 664-682.

Gao, M., Fisher, D.K., Kim, K.N., Shannon, J.C., and Guiltinan, M.J. (1996). Evolutionary conservation and expression patterns of maize starch branching enzyme I and IIb genes suggests isoform specialization. *Plant Mol Biol* 30: 1223-1232.

Gao, M., Fisher, D.K., Kim, K.N., Shannon, J.C., and Guiltinan, M.J. (1997). Independent genetic control of maize starch-branching enzymes IIa and IIb. Isolation and characterization of a Sbe2a cDNA. *Plant Physiol* 114: 69-78.

Gao, M., Wanat, J., Stinard, P.S., James, M.G., and Myers, A.M. (1998). Characterization of dull1, a Maize Gene Coding for a Novel Starch Synthase. *Plant Cell* 10: 399-412.

Garavelli, J.S. (2003). The RESID Database of Protein Modifications: 2003 developments. *Nucleic Acids Res* 31: 499-501.

Gavel, Y., and von Heijne, G. (1990). Sequence differences between glycosylated and non-glycosylated Asn-X-Thr/Ser acceptor sites: implications for protein engineering. *Protein Eng* 3: 433-442.

Geigenberger, P. (2003). Regulation of sucrose to starch conversion in growing potato tubers. *J.Exp.Bot.* 54: 457-465.

Ghosh, H.P., and Preiss, J. (1966). Adenosine diphosphate glucose pyrophosphorylase. A regulatory enzyme in the biosynthesis of starch in spinach leaf chloroplasts. *J Biol Chem* 241: 4491-4504.

Giuliano, G., Hoffman, N.E., Ko, K., Scolnik, P.A., and Cashmore, A.R. (1988). A light-entrained circadian clock controls transcription of several plant genes. *EMBO J* 7: 3635-3642.

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

Graves, L.M., Bornfeldt, K.E., and Krebs, E.G. (1997). Historical perspectives and new insights involving the MAP kinase cascades. *Adv Second Messenger Phosphoprotein Res* 31: 49-62.

Grierson, C., Du, J.S., de Torres Zabala, M., Beggs, K., Smith, C., Holdsworth, M., and Bevan, M. (1994). Separate cis sequences and trans factors direct metabolic and developmental regulation of a potato tuber storage protein gene. *Plant J* 5: 815-826.

Griffin, P.R., MacCoss, M.J., Eng, J.K., Blevins, R.A., Aaronson, J.S., and Yates, J.R., 3rd. (1995). Direct database searching with MALDI-PSD spectra of peptides. *Rapid Commun Mass Spectrom* 9: 1546-1551.

Griffin, T.J., Lock, C.M., Li, X.J., Patel, A., Chervetsova, I., Lee, H., Wright, M.E., Ranish, J.A., Chen, S.S., and Aebersold, R. (2003). Abundance ratio-dependent proteomic analysis by mass spectrometry. *Anal Chem* 75: 867-874.

Gross, P., and ap Rees, T. (1986). Alkaline inorganic pyrophosphatase and starch synthesis in amyloplasts. *Planta* 167: 140-145.

Guan, H.P., and Preiss, J. (1993). Differentiation of the Properties of the Branching Isozymes from Maize (Zea mays). *Plant Physiol* 102: 1269-1273.

Guglielminetti, L., Perata, P., Morita, A., Loreti, E., Yamaguchi, J., and Alpi, A. (2000). Characterization of isoforms of hexose kinases in rice embryo. *Phytochemistry* 53: 195-200.

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

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.

Hajirezaei, M.R., Bornke, F., Peisker, M., Takahata, Y., Lerchl, J., Kirakosyan, A., and Sonnewald, U. (2003). Decreased sucrose content triggers starch breakdown and respiration in stored potato tubers (Solanum tuberosum). *J Exp Bot* 54: 477-488.

Halford, N.G., Purcell, P.C., and Hardie, D.G. (1999). Is hexokinase really a sugar sensor in plants? *Trends Plant* Sci 4: 117-120.

Hansen, J.E., Lund, O., Engelbrecht, J., Bohr, H., and Nielsen, J.O. (1995). Prediction of O-glycosylation of mammalian proteins: specificity patterns of UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase. *Biochem J* 308 (Pt 3): 801-813.

Hansen, K.C., Schmitt-Ulms, G., Chalkley, R.J., Hirsch, J., Baldwin, M.A., and Burlingame, A.L. (2003). Mass Spectrometric Analysis of Protein Mixtures at Low Levels Using Cleavable 13C-Isotope-coded Affinity Tag and Multidimensional Chromatography. *Mol Cell Proteomics* 2: 299-314.

Harn, C., Knight, M., Ramakrishnan, A., Guan, H., Keeling, P.L., and Wasserman, B.P. (1998). Isolation and characterization of the zSSIIa and zSSIIb starch synthase cDNA clones from maize endosperm. *Plant Mol Biol* 37: 639-649.

Hart, G.W. (1997). Dynamic O-linked glycosylation of nuclear and cytoskeletal proteins. *Annu Rev Biochem* 66: 315-335.

Hatzimanikatis, V., and Lee, K.H. (1999). Dynamical analysis of gene networks requires both mRNA and protein expression information. *Metab Eng* 1: 275-281.

Hayakawa, T., Seo, S.W., and Igaaue, I. (1980). Electron microscopic observation of rice grain. Morphology of rice starch. *J. Japan. Soc.Starch Sc.* 27: 173-327.

Heazlewood, J.L., Howell, K.A., Whelan, J., and Millar, A.H. (2003). Towards an Analysis of the Rice Mitochondrial Proteome. *Plant Physiol.* 132: 230-242.

Hendriks, J.H., Kolbe, A., Gibon, Y., Stitt, M., and Geigenberger, P. (2003). ADP-glucose pyrophosphorylase is activated by posttranslational redox-modification in response to light and to sugars in leaves of Arabidopsis and other plant species. *Plant Physiol* 133: 838-849.

Hirano, H. (1997). Screening of rice genes from the cDNA catalog using the data obtained by protein sequencing. *J Protein Chem* 16: 533-536.

Hirose, T., Imaizumi, N., Scofield, G.N., Furbank, R.T., and Ohsugi, R. (1997). cDNA cloning and tissue specific expression of a gene for sucrose transporter from rice (Oryza sativa L.). *Plant Cell Physiol* 38: 1389-1396.

Hizukuri, S., Shurasaka, K., and Juliano, B.O. (1983). Phosphorous and amylose branching in rice starch granules. *Starch* 35: 348-350.

Hochholdinger, F., Guo, L., and Schnable, P.S. (2004). Cytoplasmic regulation of the accumulation of nuclear-encoded proteins in the mitochondrial proteome of maize. *Plant J* 37: 199-208.

Hounsell, E.F., Davies, M.J., and Renouf, D.V. (1996). O-linked protein glycosylation structure and function. *Glycoconj J* 13: 19-26.

Huijser, C., Kortstee, A., Pego, J., Weisbeek, P., Wisman, E., and Smeekens, S. (2000). The Arabidopsis SUCROSE UNCOUPLED-6 gene is identical to ABSCISIC ACID INSENSITIVE-4: involvement of abscisic acid in sugar responses. *Plant J* 23: 577-585.

Hunter, T. (1998). The Croonian Lecture 1997. The phosphorylation of proteins on tyrosine: its role in cell growth and disease. *Philos Trans R Soc Lond B Biol Sci* 353: 583-605.

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.

Ishiguro, S., and Nakamura, K. (1994). Characterization of a cDNA encoding a novel DNA-binding protein, SPF1, that recognizes SP8 sequences in the 5' upstream regions of genes coding for sporamin and beta-amylase from sweet potato. *Mol Gen Genet* 244: 563-571.

Itoh, K., Ozaki, H., Okada, K., Hori, H., Takeda, Y., and Mitsui, T. (2003). Introduction of Wx transgene into rice wx mutants leads to both high- and low-amylose rice. *Plant Cell* Physiol 44: 473-480.

James, M.G., Denyer, K., and Myers, A.M. (2003). Starch synthesis in the cereal endosperm. *Curr Opin Plant Biol* 6: 215-222.

Jennings, P.R., Coffman, W.R., and Kauffman, H.E. (1979). Grain quality. In Rice improvement (Intl. Rice Res. Inst., Los Banos, Laguna, Philippines), pp. 101-120.

Jespersen, H.M., MacGregor, E.A., Henrissat, B., Sierks, M.R., and Svensson, B. (1993). Starch- and glycogen-debranching and branching enzymes: prediction of structural features of the catalytic (beta/alpha)8-barrel domain and evolutionary relationship to other amylolytic enzymes. *J Protein Chem* 12: 791-805.

Jobling, S.A., Schwall, G.P., Westcott, R.J., Sidebottom, C.M., Debet, M., Gidley, M.J., Jeffcoat, R., and Safford, R. (1999). A minor form of starch branching enzyme in potato (Solanum tuberosum L.) tubers has a major effect on starch structure: cloning and characterisation of multiple forms of SBE A. *The Plant Journal: For Cell And Molecular Biology* 18: 163-171.

Johnson, L.N., Noble, M.E., and Owen, D.J. (1996). Active and inactive protein kinases: structural basis for regulation. *Cell* 85: 149-158.

Johnson, L.N., Lowe, E.D., Noble, M.E., and Owen, D.J. (1998). The Eleventh Datta Lecture. The structural basis for substrate recognition and control by protein kinases. *FEBS Lett* 430: 1-11.

Juliano, B.O. (1979). The chemical basis of rice grain quality. In Proc.Workshop on Chemical aspects of rice grain quality (Int.Rice Res. Inst., Los Banos, Laguna, Phillippines.), pp. 69-90.

Juliano, B.O. (1982). Rice: Chemistry and Technology.

Juliano, B.O. (1993). Rice in human nutrition. (Rome: Food and Agriculture Organization of the United Nations).

Kaplan, C.P., Tugal, H.B., and Baker, A. (1997). Isolation of a cDNA encoding an Arabidopsis galactokinase by functional expression in yeast. *Plant Mol Biol* 34: 497-506.

Kato, H. (1990). Temperature response and inheritance of a thermosensitive genic male sterility in rice. *Japan J. Breed.* 40(suppl. 1): 352-369.

Kieselbach, T., Hagman, Andersson, B., and Schroder, W.P. (1998). The thylakoid lumen of chloroplasts. Isolation and characterization. *J Biol Chem* 273: 6710-6716.

Kiessling, J., Martin, A., Gremillon, L., Rensing, S.A., Nick, P., Sarnighausen, E., Decker, E.L., and Reski, R. (2004b). Dual targeting of plastid division protein FtsZ to chloroplasts and the cytoplasm. *EMBO Rep* 5: 889-894.

Kim, W.T., Li, X., and Okita, T.W. (1993). Expression of storage protein multigene families in developing rice. *Plant Cell Physiology* 34: 595-603.

Kishimoto, T., Watanabe, M., Mitsui, T., and Hori, H. (1999). Glutelin Basic Subunits Have a Mammalian Mucin-Type O-Linked Disaccharide Side Chain. *Archives of Biochemistry and Biophysics* 370: 271-277.

Kleczkowski, L.A., Villand, P., Preiss, J., and Olsen, O.A. (1993a). Kinetic mechanism and regulation of ADP-glucose pyrophosphorylase from barley (Hordeum vulgare) leaves. *J Biol Chem* 268: 6228-6233.

Kleczkowski, L.A., Villand, P., Luthi, E., Olsen, O.A., and Preiss, J. (1993b). Insensitivity of barley endosperm ADP-glucose pyrophosphorylase to 3-phosphoglycerate and orthophosphate regulation. *Plant Physiol* 101: 179-186.

Kleffmann, T., Russenberger, D., von Zychlinski, A., Christopher, W., Sjolander, K., Gruissem, W., and Baginsky, S. (2004). The Arabidopsis thaliana chloroplast proteome reveals pathway abundance and novel protein functions. *Curr Biol* 14: 354-362.

Knetsch, M., Wang, M., Snaar-Jagalska, B.E., and Heimovaara-Dijkstra, S. (1996). Abscisic Acid Induces Mitogen-Activated Protein Kinase Activation in Barley Aleurone Protoplasts. *Plant Cell* 8: 1061-1067.

Kolibaba, K.S., and Druker, B.J. (1997). Protein tyrosine kinases and cancer. *Biochim Biophys Acta* 1333: F217-248.

Koller, A., Washburn, M.P., Lange, B.M., Andon, N.L., Deciu, C., Haynes, P.A., Hays, L., Schieltz, D., Ulaszek, R., Wei, J., Wolters, D., and Yates, J.R., 3rd. (2002). Proteomic survey of metabolic pathways in rice. *Proc Natl Acad Sci U S A* 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. *Theor. Appl. Genet.* 86: 953-942.

Komatsu, S., Rakwal, R., and Li, Z. (1999). Separation and characterization of proteins in rice suspension cultured cells. *Plant Cell Tissue Organ Culture* 55: 183-192.

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 Res.* 32: D388-D392.

Kossmann, J., and Lloyd, J. (2000). Understanding and influencing starch biochemistry. *Crit Rev Biochem Mol Biol* 35: 141-196.

Kristensen, B.K., Askerlund, P., Bykova, N.V., Egsgaard, H., and Moller, I.M. (2004). Identification of oxidised proteins in the matrix of rice leaf mitochondria by immunoprecipitation and two-dimensional liquid chromatography-tandem mass spectrometry. *Phytochemistry* 65: 1839-1851.

Kruft, V., Eubel, H., Jansch, L., Werhahn, W., and Braun, H.P. (2001). Proteomic approach to identify novel mitochondrial proteins in Arabidopsis. *Plant Physiol* 127: 1694-1710.

Kubo, A., Fujita, N., Harada, K., Matsuda, T., Satoh, H., and Nakamura, Y. (1999). The starch-debranching enzymes isoamylase and pullulanase are both involved in amylopectin biosynthesis in rice endosperm. *Plant Physiol.* 121: 399-409.

Kubota, K., Wakabayashi, K., and Matsuoka, T. (2003). Proteome analysis of secreted proteins during osteoclast differentiation using two different methods: two-dimensional electrophoresis and isotope-coded affinity tags analysis with two-dimensional chromatography. *Proteomics* 3: 616-626.

Kuriki, T., Stewart, D.C., and Preiss, J. (1997). Construction of Chimeric Enzymes out of Maize Endosperm Branching Enzymes I and II:. ACTIVITY AND PROPERTIES. *J. Biol. Chem.* 272: 28999-29004.

Laby, R.J., Kincaid, M.S., Kim, D., and Gibson, S.I. (2000). The Arabidopsis sugar-insensitive mutants sis4 and sis5 are defective in abscisic acid synthesis and response. *Plant J* 23: 587-596.

Lamport, D.T.A. (1980). The Biochemistry of Plants. (Academic Press, New York.).

Lao, N.T., Schoneveld, O., Mould, R.M., Hibberd, J.M., Gray, J.C., and Kavanagh, T.A. (1999). An Arabidopsis gene encoding a chloroplast-targeted beta-amylase. *Plant J* 20: 519-527.

Leon, P., and Sheen, J. (2003). Sugar and hormone connections. *Trends Plant Sci* 8: 110-116.

Li, B., Geiger, D.R., and Shieh, W.J. (1992). Evidence for circadian regulation of starch and sucrose synthesis in sugar beet leaves. *Plant Physiol.* 99: 1393-1399.

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

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., Wang, Z., Chen, X., Cai, X., Tang, S., Yu, H., Zhang, J., Hong, M., and Gu, M. (2003). Stable Inheritance of the Antisense Waxy Gene in Transgenic Rice with Reduced Amylose Level and Improved Quality. *Transgenic Research* 12: 71-82.

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 Physiol* 134: 560-574.

Loreti, E., Alpi, A., and Perata, P. (2000). Glucose and disaccharide-sensing mechanisms modulate the expression of alpha-amylase in barley embryos. *Plant Physiol* 123: 939-948.

Loreti, E., Bellis, L.D., Alpi, A., and Perata, P. (2001). Why and how do plant cells sense sugars? *Annals of Botany* 88: 803-812.

Lu, C., and Zhou, J. (2000). Breeding and utilization of two-line interspecific hybrid rice Liangyou Peijiu. *Hybrid Rice* 15: 4-5.

Lukaszewski, K.M., Bowsher, C.G., Savory, P.J., and Emes, M.J. (2001). Protein phosphorylation in pea root plastids. *Plant Cell Physiol* 42: 642-649.

MacCoss, M.J., Wu, C.C., and Yates, J.R., 3rd. (2002). Probability-based validation of protein identifications using a modified SEQUEST algorithm. *Anal Chem* 74: 5593-5599.

MacDonald, F.D., and ap Rees, T. (1983). Enzymic properties of. amyloplasts from suspension cultures of soybean. *Biochem. Biophys. Acta* 755: 81-89.

Macherel, D., Viale, A., and Akazawa, T. (1986). Protein-Phosphorylation in Amyloplasts Isolated from Suspension-Cultured Cells of Sycamore (Acer-Pseudoplatanus L). *Plant Physiol.* 80: 1041-1044.

Maeo, K., Tomiya, T., Hayashi, K., Akaike, M., Morikami, A., Ishiguro, S., and Nakamura, K. (2001). Sugar-responsible elements in the promoter of a gene for beta-amylase of sweet potato. *Plant Mol Biol* 46: 627-637.

Maleck, K., Levine, A., Eulgem, T., Morgan, A., Schmid, J., Lawton, K.A., Dangl, J.L., and Dietrich, R.A. (2000). The transcriptome of Arabidopsis thaliana during systemic acquired resistance. *Nat Genet* 26: 403-410.

Maltman, D.J., Simon, W.J., Wheeler, C.H., Dunn, M.J., Wait, R., and Slabas, A.R. (2002). Proteomic analysis of the endoplasmic reticulum from developing and germinating seed of castor (Ricinus communis). *Electrophoresis* 23: 626-639.

Mechin, V., Balliau, T., Chateau-Joubert, S., Davanture, M., Langella, O., Negroni, L., Prioul, J.L., Thevenot, C., Zivy, M., and Damerval, C. (2004). A two-dimensional proteome map of maize endosperm. *Phytochemistry* 65: 1609-1618.

Merida, A., Rodriguez-Galan, J.M., Vincent, C., and Romero, J.M. (1999). Expression of the granule-bound starch synthase I (Waxy) gene from snapdragon is developmentally and circadian clock regulated. *Plant Physiol* 120: 401-410.

Merril, C.R., Goldman, D., Sedman, S.A., and Ebert, M.H. (1981). Ultrasensitive stain for proteins in polyacrylamide gels shows regional variation in cerebrospinal fluid proteins. *Science* 211: 1437-1438.

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 Physiol* 127: 1711-1727.

Millar, A.J., and Kay, S.A. (1996). Integration of circadian and phototransduction pathways in the network controlling CAB gene transcription in Arabidopsis. *Proc Natl Acad Sci U S A* 93: 15491-15496.

Miras, S., Salvi, D., Ferro, M., Grunwald, D., Garin, J., Joyard, J., and Rolland, N. (2002). Non-canonical transit peptide for import into the chloroplast. *J Biol Chem* 277: 47770-47778.

Mizuno, K., Kimura, K., Arai, Y., Kawasaki, T., Shimada, H., and Baba, T. (1992). Starch branching enzymes from immature rice seeds. *J Biochem (Tokyo)* 112: 643-651.

Mizuno, K., Kawasaki, T., Shimada, H., Satoh, H., Kobayashi, E., Okumura, S., Arai, Y., and Baba, T. (1993). Alteration of the structural properties of starch components by the lack of an isoform of starch branching enzyme in rice seeds. *J Biol Chem* 268: 19084-19091.

Mizuno, K., Kobayashi, E., Tachibana, M., Kawasaki, T., Fujimura, T., Funane, K., Kobayashi, M., and Baba, T. (2001). Characterization of an isoform of rice starch branching enzyme, RBE4, in developing seeds. *Plant Cell Physiol* 42: 349-357.

Morell, M.K., Blennow, A., Kosar-Hashemi, B., and Samuel, M.S. (1997). Differential expression and properties of starch branching enzyme isoforms in developing wheat endosperm. *Plant Physiol* 113: 201-208.

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.

Mu, H.H., Mu-Forster, C., Bohonko, M., and Wasserman, B.P. (1998). Heat-Induced Fragmentation of the Maize Waxy Protein During Protein Extraction from Starch Granules. *Cereal Chem.* 75: 480-483.

Mu-Forster, C., and Wasserman, B.P. (1998). Surface localization of zein storage proteins in starch granules from maize endosperm. Proteolytic removal by thermolysin and in vitro cross-linking of granule-associated polypeptides. *Plant Physiol* 116: 1563-1571.

Mu-Forster, C., Huang, R., Powers, J.R., Harriman, R.W., Knight, M., Singletary, G.W., Keeling, P.L., and Wasserman, B.P. (1996). Physical Association of Starch Biosynthetic Enzymes with Starch Granules of Maize Endosperm (Granule-Associated Forms of Starch Synthase I and Starch Branching Enzyme II). *Plant Physiol.* 111: 821-829.

Muller-Rober, B., Sonnewald, U., and Willmitzer, L. (1992). Inhibition of the ADP-glucose pyrophosphorylase in transgenic potatoes leads to sugar-storing tubers and influences tuber formation and expression of tuber storage protein genes. *The EMBO Journal* 11: 1229-1238.

Mutisya, J., Sathish, P., Sun, C., Andersson, L., Ahlandsberg, S., Baguma, Y., Palmqvist, S., Odhiambo, B., Aman, P., and Jansson, C. (2003). Starch branching enzymes in sorghum (Sorghum bicolor) and barley (Hordeum vulgare): comparative analyses of enzyme structure and gene expression. *J Plant Physiol* 160: 921-930.

Nakamura, T., Vrinten, P., Hayakawa, K., and Ikeda, J. (1998). Characterization of a Granule-Bound Starch Synthase Isoform Found in the Pericarp of Wheat. *Plant Physiol.* 118: 451-459.

Nakamura, Y. (2002). Towards a Better Understanding of the Metabolic System for Amylopectin Biosynthesis in Plants: Rice Endosperm as a Model Tissue. *Plant Cell Physiol.* 43: 718-725.

Nakamura, Y., Umemoto, T., Ogata, N., Kuboki, Y., Yano, M., and Sasaki, T. (1996). Starch debranching enzyme (R-enzyme or pullulanase) from developing rice endosperm: purification, cDNA and chromosomal localization of the gene. *Planta* 199: 209-218.

Nakayama, M., Kikuno, R., and Ohara, O. (2002). Protein-protein interactions between large proteins: two-hybrid screening using a functionally classified library composed of long cDNAs. *Genome Res* 12: 1773-1784.

Nelson, O.E., and Rines, H.W. (1962). The enzymatic deficiency in the waxy mutant of maize. *BiochemBiophys Res Commun* 9: 297-300.

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.

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

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

Ohdan, T., Francisco, P.B., Jr., Sawada, T., Hirose, T., Terao, T., Satoh, H., and Nakamura, Y. (2005). Expression profiling of genes involved in starch synthesis in sink and source organs of rice. *J Exp Bot* 56: 3229-3244.

Ono, K., Ishimaru, K., Aoki, N., Takahashi, S., Ozawa, K., Ohkawa, Y., and Ohsugi, R. (1999). Characterization of a maize sucrose-phosphate synthase protein and its effect on carbon partitioning in transgenic rice plant. *Plant Prod. Sci.* 2: 172?77.

P. R.Adoration, X.L.L., T. W.Okita, B. O.Juliano. (1993). Characterization of poorly digested protein of cooked rice protein bodies. *Cereal Chem* 70: 101-105.

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

Pasquali, C., Fialka, I., and Huber, L.A. (1999). Subcellular fractionation, electromigration analysis and mapping of organelles. *Journal of Chromatography B: Biomedical Sciences and Applications* 722: 89-102.

Peat, S., Whelan, W.J., and Rees, W.R. (1953). D-enzyme: a disproportionating enzyme in potato juice. *Nature* 172: 158.

Peeters, N., and Small, I. (2001). Dual targeting to mitochondria and chloroplasts. *Biochim Biophys Acta* 1541: 54-63.

Pego, J.V., and Smeekens, S.C. (2000). Plant fructokinases: a sweet family get-together. *Trends Plant Sci* 5: 531-536.

Peltier, J.B., Friso, G., Kalume, D.E., Roepstorff, P., Nilsson, F., Adamska, I., and van Wijk, K.J. (2000). Proteomics of the Chloroplast: Systematic Identification and Targeting Analysis of Lumenal and Peripheral Thylakoid Proteins. *Plant Cell* 12: 319-342.

Peltier, J.B., Emanuelsson, O., Kalume, D.E., Ytterberg, J., Friso, G., Rudella, A., Liberles, D.A., Soderberg, L., Roepstorff, P., von Heijne, G., and van Wijk, K.J. (2002). Central functions of the lumenal and peripheral thylakoid proteome of Arabidopsis determined by experimentation and genome-wide prediction. *Plant Cell* 14: 211-236.

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

Peng, M., Gao, M., Abdel-Aal, E.-S.M., Hucl, P., and Chibbar, R.N. (1999). Separation and Characterization of A- and B-Type Starch Granules in Wheat Endosperm. *Cereal Chem.* 76: 375-379.

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.

Pinna, L.A., and Ruzzene, M. (1996). How do protein kinases recognize their substrates? *Biochim Biophys Acta* 1314: 191-225.

Pongratz, P., and Beck, E. (1978). Diurnal Oscillation of Amylolytic Activity in Spinach Chloroplasts. *Plant Physiol* 62: 687-689.

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.

Preiss, J., and Sivak, M. (1996). Starch synthesis in sinks and sources. (New York: Marcel Dekker).

Prime, T.A., Sherrier, D.J., Mahon, P., Packman, L.C., and Dupree, P. (2000). A proteomic analysis of organelles from Arabidopsis thaliana. *Electrophoresis* 21: 3488-3499.

Rahman, S., Regina, A., Li, Z., Mukai, Y., Yamamoto, M., Kosar-Hashemi, B., Abrahams, S., and Morell, M.K. (2001). Comparison of starch-branching enzyme genes reveals evolutionary relationships among isoforms. Characterization of a gene for starch-branching enzyme IIa from the wheat genome donor Aegilops tauschii. *Plant Physiol* 125: 1314-1324.

Ranish, J.A., Yi, E.C., Leslie, D.M., Purvine, S.O., Goodlett, D.R., Eng, J., and Aebersold, R. (2003). The study of macromolecular complexes by quantitative proteomics. *Nat Genet* 33: 349-355.

Reinders, J., Lewandrowski, U., Moebius, J., Wagner, Y., and Sickmann, A. (2004). Challenges in mass spectrometry-based proteomics. *Proteomics* 4: 3686-3703.

Robertson, D., Mitchell, G.P., Gilroy, J.S., Gerrish, C., Bolwell, G.P., and Slabas, A.R. (1997). Differential extraction and protein sequencing reveals major differences in patterns of primary cell wall proteins from plants. *J Biol Chem* 272: 15841-15848.

Roitsch, T., Balibrea, M.E., Hofmann, M., Proels, R., and Sinha, A.K. (2003). Extracellular invertase: key metabolic enzyme and PR protein. *J Exp Bot* 54: 513-524.

Rolland, F., Moore, B., and Sheen, J. (2002). Sugar sensing and signaling in plants. *Plant Cell* 14 Suppl: S185-205.

Rook, F., Corke, F., Card, R., Munz, G., Smith, C., and Bevan, M.W. (2001). Impaired sucrose-induction mutants reveal the modulation of sugar-induced starch biosynthetic gene expression by abscisic acid signalling. *Plant J* 26: 421-433.

Russell, P.L., and Juliano, B.O. (1983). Differential scanning calorimetry of rice starches. *Starch* 35: 382-386.

Sano, Y. (1984). Differential regulation of waxy gene expression in rice endosperm. *Theor Appl Genet* 68: 467-473.

Santoni, V., Rouquie, D., Doumas, P., Mansion, M., Boutry, M., Degand, H., Dupree, P., Packman, L., Sherrier, J., Prime, T., Bauw, G., Posada, E., Rouze, P., Dehais, P., Sahnoun, I., Barlier, I., and Rossignol, M. (1998). Use of a proteome strategy for tagging proteins present at the plasma membrane. *Plant J* 16: 633-641.

Schindler, I., Renz, A., Schmid, F.X., and Beck, E. (2001). Activation of spinach pullulanase by reduction results in a decrease in the number of isomeric forms. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* 1548: 175-186.

Schleiff, E., and Klosgen, R.B. (2001). Without a little help from 'my' friends: direct insertion of proteins into chloroplast membranes? *Biochim Biophys Acta* 1541: 22-33.

Schleiff, E., Eichacker, L.A., Eckart, K., Becker, T., Mirus, O., Stahl, T., and Soll, J. (2003). Prediction of the plant beta-barrel proteome: a case study of the chloroplast outer envelope. *Protein Sci* 12: 748-759.

Schubert, M., Petersson, U.A., Haas, B.J., Funk, C., Schroder, W.P., and Kieselbach, T. (2002). Proteome map of the chloroplast lumen of Arabidopsis thaliana. *J Biol Chem* 277: 8354-8365.

Seigneurin-Berny, D., Rolland, N., Garin, J., and Joyard, J. (1999). Technical Advance: Differential extraction of hydrophobic proteins from chloroplast envelope membranes: a subcellular-specific proteomic approach to identify rare intrinsic membrane proteins. *Plant J* 19: 217-228.

Shannon, J.C., Pien, F.M., Cao, H., and Liu, K.C. (1998). Brittle-1, an Adenylate Translocator, Facilitates Transfer of Extraplastidial Synthesized ADP-Glucose into Amyloplasts of Maize Endosperms. *Plant Physiol.* 117: 1235-1252.

Sheen, J. (1996). Ca2+-dependent protein kinases and stress signal transduction in plants. *Science* 274: 1900-1902.

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. *J Biochem (Tokyo)* 132: 613-620.

Sherrier, D.J., Prime, T.A., and Dupree, P. (1999). Glycosylphosphatidylinositol-anchored cell-surface proteins from Arabidopsis. *Electrophoresis* 20: 2027-2035.

Sherson, S., Gy, I., Medd, J., Schmidt, R., Dean, C., Kreis, M., Lecharny, A., and Cobbett, C. (1999). The arabinose kinase, ARA1, gene of Arabidopsis is a novel member of the galactose kinase gene family. *Plant Mol Biol* 39: 1003-1012.

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

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

Shimada, H., Tada, Y., Kawasaki, T., and Fujimura, T. (1993). Antisense Regulation of the Rice Waxy Gene-Expression Using a Pcr-Amplified Fragment of the Rice Genome Reduces the Amylose Content in Grain Starch. *Theoretical and Applied Genetics* 86: 665-672.

Shure, M., Wessler, S., and Fedoroff, N. (1983). Molecular identification and isolation of the Waxy locus in maize. *Cell* 35: 225-233.

Sikka, V.K., Choi, S.-B., Kavakli, I.H., Sakulsingharoj, C., Gupta, S., Ito, H., and Okita, T.W. (2001). Subcellular compartmentation and allosteric regulation of the rice endosperm ADPglucose pyrophosphorylase. *Plant Science* 161: 461-468.

Silva-Filho, M.C. (2003). One ticket for multiple destinations: dual targeting of proteins to distinct subcellular locations. *Curr Opin Plant Biol* 6: 589-595.

Sivak, M.N., Wagner, M., and Preiss, J. (1993). Biochemical Evidence for the Role of the Waxy Protein from Pea (Pisum sativum L.) as a Granule-Bound Starch Synthase. *Plant Physiol* 103: 1355-1359.

Skylas, D.J., Van Dyk, D., and Wrigley, C.W. (2005). Proteomics of wheat grain. *New Approaches in Cereal Science* 41: 165-179.

Slattery, C.J., Kavakli, I.H., and Okita, T.W. (2000). Engineering starch for increased quantity and quality. *Trends Plant Sci* 5: 291-298.

Smeekens, S. (2000). Sugar-Induced Signal Transduction in Plants. *Annu Rev Plant Physiol Plant Mol Biol* 51: 49-81.

Smith, A.M. (1988). Major differences in isoforms of starch branching enzyme between developing embryos of round- and wrinkled-seeded peas (Pisum sativum L.). *Planta* 175: 270-279.

Smith, S.M., Fulton, D.C., Chia, T., Thorneycroft, D., Chapple, A., Dunstan, H., Hylton, C., Zeeman, S.C., and Smith, A.M. (2004). Diurnal changes in the transcriptome encoding enzymes of starch metabolism provide evidence for both transcriptional and posttranscriptional regulation of starch metabolism in Arabidopsis leaves. *Plant Physiol* 136: 2687-2699.

Smith-White, B.J., and Preiss, J. (1992). Comparison of proteins of ADP-glucose pyrophosphorylase from diverse sources. *J Mol Evol* 34: 449-464.

Smolka, M., Zhou, H., and Aebersold, R. (2002). Quantitative protein profiling using two-dimensional gel electrophoresis, isotope-coded affinity tag labeling, and mass spectrometry. *Mol Cell Proteomics* 1: 19-29.

Snow, D.M., and Hart, G.W. (1998). Nuclear and cytoplasmic glycosylation. *Int Rev Cytol* 181: 43-74.

Sowokinos, J., and Preiss, J. (1982). Pyrophosphorylases in Solanum tuberosum.III. Purification, physical, and catalytic properties of ADPglucose pyrophosphorylase in potatoes. *Plant Physiol* 69: 1459-1466.

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

Spengler, B., Kirsch, D., Kaufmann, R., and Jaeger, E. (1992). Peptide sequencing by matrix-assisted laser-desorption mass spectrometry. *Rapid Commun Mass Spectrom* 6: 105-108.

Spradlin, J., and Thoma, J.A. (1970). Beta-amylase thiol groups. Possible regulator sites. *J Biol Chem* 245: 117-127.

Stanley, P. (1992). Glycosylation engineering. *Glycobiology* 2: 99-107.

Stark, D.M., Timmerman, K.P., Barry, G.F., Preiss, J., and Kishore, G.M. (1992). Regulation of the amount of starch in plant tissues by ADP glucose pyrophosphorylase. *SCIENCE* 258: 287-292.

Steinberg, T.H., Pretty On Top, K., Berggren, K.N., Kemper, C., Jones, L., Diwu, Z., Haugland, R.P., and Patton, W.F. (2001). Rapid and simple single nanogram detection of glycoproteins in polyacrylamide gels and on electroblots. *Proteomics* 1: 841-855.

Steinberg, T.H., Agnew, B.J., Gee, K.R., Leung, W.Y., Goodman, T., Schulenberg, B., Hendrickson, J., Beechem, J.M., Haugland, R.P., and Patton, W.F. (2003). Global quantitative phosphoprotein analysis using Multiplexed Proteomics technology. *Proteomics* 3: 1128-1144.

Sulli, C., Fang, Z., Muchhal, U., and Schwartzbach, S.D. (1999). Topology of Euglena Chloroplast Protein Precursors within Endoplasmic Reticulum to Golgi to Chloroplast Transport Vesicles. *J. Biol. Chem.* 274: 457-463.

Sun, C., Sathish, P., Ahlandsberg, S., and Jansson, C. (1998). The two genes encoding starch-branching enzymes IIa and IIb are differentially expressed in barley. *Plant Physiol* 118: 37-49.

Sun, C., Palmqvist, S., Olsson, H., Boren, M., Ahlandsberg, S., and Jansson, C. (2003). A novel WRKY transcription factor, SUSIBA2, participates in sugar signaling in barley by binding to the sugar-responsive elements of the isol promoter. *Plant Cell* 15: 2076-2092.

Tabb, D.L., McDonald, W.H., and Yates, J.R., 3rd. (2002). DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *J Proteome Res* 1: 21-26.

Tabb, D.L., Saraf, A., and Yates, J.R., 3rd. (2003a). GutenTag: high-throughput sequence tagging via an empirically derived fragmentation model. *Anal Chem* 75: 6415-6421.

Tabb, D.L., MacCoss, M.J., Wu, C.C., Anderson, S.D., and Yates, J.R., 3rd. (2003b). Similarity among tandem mass spectra from proteomic experiments: detection, significance, and utility. *Anal Chem* 75: 2470-2477.

Tada, Y., Nakase, M., Adachi, T., Nakamura, R., Shimada, H., Takahashi, M., Fujimura, T., and Matsuda, T. (1996). Reduction of 14 to 16 kDa allergenic proteins in transgenic rice plants by antisense gene. *FEBS Lett.* 391: 341-345.

Takaha, T., Yanase, M., Takata, H., Okada, S., and Smith, S.M. (1998). Cyclic glucans produced by the intramolecular transglycosylation activity of potato D-enzyme on amylopectin. *Biochem Biophys Res Commun* 247: 493-497.

Takeda, Y., Guan, H.P., and Preiss, J. (1993). Branching of amylose by the branching isoenzymes of maize endosperm. *Carbohydr Res* 240: 253-263.

Tanaka, N., Mitsui, S., Nobori, H., Yanagi, K., and Komatsu, S. (2005). Expression and Function of Proteins during Development of the Basal Region in Rice Seedlings. *Mol Cell Proteomics* 4: 796-808.

Tanaka, N., Fujita, M., Handa, H., Murayama, S., Uemura, M., Kawamura, Y., Mitsui, T., Mikami, S., Tozawa, Y., Yoshinaga, T., and Komatsu, S. (2004). *Proteomics of the rice cell: systematic identification of the protein populations in subcellular compartments* 271: 566-576.

Tandecarz, J., Lavintman, N., and Cardini, C.E. (1975). Biosynthesis of starch. Formation of a glucoproteic acceptor by a potato non-sedimentable preparation. *Biochim Biophys Acta* 399: 345-355.

Terada, R., Nakajima, M., Isshiki, M., Okagaki, R.J., Wessler, S.R., and Shimamoto, K. (2000). Antisense Waxy Genes with Highly Active Promoters Effectively Suppress Waxy Gene Expression in Transgenic Rice. *Plant Cell Physiol.* 41: 881-888.

Terzaghi, W.B., Bertekap, R.L., Jr., and Cashmore, A.R. (1997). Intracellular localization of GBF proteins and blue light-induced import of GBF2 fusion proteins into the nucleus of cultured Arabidopsis and soybean cells. *Plant J* 11: 967-982.

Tetlow, I.J., Blissett, K.J., and Emes, M.J. (1993). A rapid method for the isolation of purified amyloplasts from wheat endosperm. *Planta* 189: 597-600.

Tetlow, I.J., Morell, M.K., and Emes, M.J. (2004a). Recent developments in understanding the regulation of starch metabolism in higher plants. *J.Exp.Bot.* 55: 2131-2145.

Tetlow, I.J., Davies, E.J., Vardy, K.A., Bowsher, C.G., Burrell, M.M., and Emes, M.J. (2003). Subcellular localization of ADPglucose pyrophosphorylase in developing wheat endosperm and analysis of the properties of a plastidial isoform. *J.Exp.Bot.* 54: 715-725.

Tetlow, I.J., Wait, R., Lu, Z., Akkasaeng, R., Bowsher, C.G., Esposito, S., Kosar-Hashemi, B., Morell, M.K., and Emes, M.J. (2004b). Protein Phosphorylation in Amyloplasts Regulates Starch Branching Enzyme Activity and Protein-Protein Interactions. *Plant Cell* 16: 694-708.

Thorbjornsen, T., Villand, P., Kleczkowski, L.A., and Olsen, O.A. (1996a). A single gene encodes two different transcripts for the ADP-glucose pyrophosphorylase small subunit from barley (Hordeum vulgare). *Biochem J* 313: 149-154.

Thorbjornsen, T., Villand, P., Denyer, K., Olsen, O.-A., and Smith, A.M. (1996b). Distinct isoforms of ADPglucose pyrophosphorylase occur inside and outside the amyloplasts in barley endosperm. *The Plant Journal* 10: 243-250.

Tiessen, A., Hendriks, J.H., Stitt, M., Branscheid, A., Gibon, Y., Farre, E.M., and Geigenberger, P. (2002). Starch synthesis in potato tubers is regulated by post-translational redox modification of ADP-glucose pyrophosphorylase: a novel regulatory mechanism linking starch synthesis to the sucrose supply. *Plant Cell* 14: 2191-2213.

Tiessen, A., Prescha, K., Branscheid, A., Palacios, N., McKibbin, R., Halford, N.G., and Geigenberger, P. (2003). Evidence that SNF1-related kinase and hexokinase are involved in separate sugar-signalling pathways modulating post-translational redox activation of ADP-glucose pyrophosphorylase in potato tubers. *Plant J* 35: 490-500.

Tonge, R., Shaw, J., Middleton, B., Rowlinson, R., Rayner, S., Young, J., Pognan, F., Hawkins, E., Currie, I., and Davison, M. (2001). Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology. *Proteomics* 1: 377-396.

Tsai, C.-Y. (1974). The function of the Waxy locus in starch synthesis in maize endosperm. *Biochemical Genetics (Historical Archive)* 11: 83-96.

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

Umemoto, N., Nakamura, K., and Ishikura, N. (1994). Effect of grain location on the panicle on activities involved in starch synthesis in rice endosperm. *Photochemistry* 3: 843-847.

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.

Upadhyaya, N.M., Zhou, X.R., Wu, L., Ramm, K., and Dennis, E.S. (2000). The tms2 gene works as a negative selection marker in rice. *Plant Mol. Biol. Rep.* 18: 227-233.

Van den Steen, P., Rudd, P.M., Dwek, R.A., and Opdenakker, G. (1998). Concepts and principles of O-linked glycosylation. *Crit Rev Biochem Mol Biol* 33: 151-208.

van der Leij, F.R., Visser, R.G., Ponstein, A.S., Jacobsen, E., and Feenstra, W.J. (1991). Sequence of the structural gene for granule-bound starch synthase of potato (Solanum tuberosum L.) and evidence for a single point deletion in the amf allele. *Mol Gen Genet* 228: 240-248.

Varki, A. (1993). Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* 3: 97-130.

Vener, A.V., Rokka, A., Fulgosi, H., Andersson, B., and Herrmann, R.G. (2000). A cyclophilin-regulated PP2A-like protein phosphatase in thylakoid membranes of plant chloroplasts. *Biochemistry* 39: 2130.

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.

Vescovi, E.G., and Viale, A.M. (1993). Protein-Phosphorylation in Plastids from Ripening Tomato Fruits (Lycopersicon-Esculentum). *Plant and Cell Physiology* 34: 905-911.

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. *Mol Cell Proteomics*: M500018-MCP500200.

Vrinten, P.L., and Nakamura, T. (2000). Wheat granule-bound starch synthase I and II are encoded by separate genes that are expressed in different tissues. *Plant Physiol* 122: 255-264.

Wall, M.K., Mitchenall, L.A., and Maxwell, A. (2004). Arabidopsis thaliana DNA gyrase is targeted to chloroplasts and mitochondria. *Proc Natl Acad Sci U S A* 101: 7821-7826.

Wang, S.-J., Yeh, K.-W., and Tsai, C.-Y. (2001). Regulation of starch granule-bound starch synthase I gene expression by circadian clock and sucrose in the source tissue of sweet potato. *Plant Science* 161: 635-644.

Wang, Z.Y., Zheng, F.Q., Shen, G.Z., Gao, J.P., Snustad, D.P., Li, M.G., Zhang, J.L., and Hong, M.M. (1995). The amylose content in rice endosperm is related to the post-transcriptional regulation of the waxy gene. *The Plant Journal* 7: 613-622.

Washburn, M.P., Wolters, D., and Yates, J.R., 3rd. (2001). Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* 19: 242-247.

Washburn, M.P., Ulaszek, R., Deciu, C., Schieltz, D.M., and Yates, J.R., 3rd. (2002). Analysis of quantitative proteomic data generated via multidimensional protein identification technology. *Anal Chem* 74: 1650-1657.

Watson, B.S., Asirvatham, V.S., Wang, L., and Sumner, L.W. (2003). Mapping the proteome of barrel medic (Medicago truncatula). *Plant Physiol* 131: 1104-1123.

Wells, L., Vosseller, K., and Hart, G.W. (2001). Glycosylation of nucleocytoplasmic proteins: signal transduction and O-GlcNAc. *Science* 291: 2376-2378.

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.

Wobus, U., and Weber, H. (1999). Sugars as signal molecules in plant seed development. *Biol Chem* 380: 937-944.

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.

Worrell, A.C., Bruneau, J.M., Summerfelt, K., Boersig, M., and Voelker, T.A. (1991). Expression of a maize sucrose phosphate synthase in tomato alters leaf carbohydrate partitioning. *Plant Cell* 3: 1121-1130.

Xiao, J., Li, J., Yuan, L., and Tanksley, S.D. (1995). Dominance is the major genetic basis of heterosis in rice as revealed by QTL analysis using molecular markers. *Genetics* 140: 745-754.

Yamamori, M., Fujita, S., Hayakawa, K., Matsuki, J., and Yasui, T. (2000). Genetic elimination of a starch granule protein, SGP-1, of wheat generates an altered starch with apparent high amylose. *TAG Theoretical and Applied Genetics* 101: 21-29.

Yamanouchi, H., and Nakamura, Y. (1992). Organ specificity of isoforms of starch branching enzyme (Q-enzyme) in rice. *Plant and Cell Physiology* 33: 985?91.

Yates, J.R., 3rd, Eng, J.K., and McCormack, A.L. (1995a). Mining genomes: correlating tandem mass spectra of modified and unmodified peptides to sequences in nucleotide databases. *Anal Chem* 67: 3202-3210.

Yates, J.R., 3rd, Eng, J.K., McCormack, A.L., and Schieltz, D. (1995b). Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. *Anal Chem* 67: 1426-1436.

Yuan, D.Y. (2004). Biochemical and molecular characterization of transgenic rice expressing a lysine-rich protein from winged bean. Ph.D Thesis, The Chinese University of Hong Kong.

Yuan, L.P. (1997). Exploiting crop heterosis b y two-line system hybrids: Current status and future prospects. (Changsha, Hunan, China,September).

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. *Nucl. Acids Res.* 31: 229-233.

Zabawinski, C., Van Den Koornhuyse, N., D'Hulst, C., Schlichting, R., Giersch, C., Delrue, B., Lacroix, J.M., Preiss, J., and Ball, S. (2001). Starchless mutants of Chlamydomonas reinhardtii lack the small subunit of a heterotetrameric ADP-glucose pyrophosphorylase. *J Bacteriol* 183: 1069-1077.

Zeeman, S.C., Northrop, F., Smith, A.M., and Rees, T.a. (1998). A starch-accumulating mutant ofArabidopsis thalianadeficient in a chloroplastic starch-hydrolysing enzyme. *The Plant Journal* 15: 357-365.

Zhang, S., Du, H., and Klessig, D.F. (1998). Activation of the tobacco SIP kinase by both a cell wall-derived carbohydrate elicitor and purified proteinaceous elicitins from Phytophthora spp. *Plant Cell* 10: 435-450.

Zhang, W., and Chait, B.T. (2000). ProFound: an expert system for protein identification using mass spectrometric peptide mapping information. *Anal Chem* 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.

Zhong, B., Karibe, H., Komatsu, S., Ichimura, H., Nagamura, Y., Sasaki, T., and Hirano, H. (1997). Screening of rice genes from a cDNA based on the sequence data-file of proteins separated by two-dimensional electrophoresis. *Breeding Sci.* 47: 255-251.

Zhu, T., Budworth, P., Chen, W., Provart, N., Chang, H.-S., Guimil, S., Su, W., Estes, B., Zou, G., and Wang, X. (2003). Transcriptional control of nutrient partitioning during rice grain filling. *Plant Biotechnology Journal* 1: 59-70.

Zourelidou, M., de Torres-Zabala, M., Smith, C., and Bevan, M.W. (2002). Storekeeper defines a new class of plant-specific DNA-binding proteins and is a putative regulator of patatin expression. *Plant J* 30: 489-497.