Understanding the Soybean Response to Salinity Stress: From the Viewpoint of Proteomics and Histone Modifications

WU, Tao

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Biology

The Chinese University of Hong Kong

September 2010

Thesis/ Assessment Committee

Professor JIANG Li Wen (Chair)

Professor NGAI Sai Ming (Thesis Supervisor)

Professor LAM Hon-Ming (Thesis Supervisor)

Professor WONG Yum Shing (Committee Member)

Professor WONG Ngok Shun (External Examiner)

Declaration

All experimental works reported in this thesis were performed by the author, unless stated the otherwise.

Wu Tao

Abstract of thesis entitled:

Understanding the soybean response to salinity stress: from the viewpoint of proteomics and histone modifications

Submitted by WU Tao

for the degree of Doctor of Philosophy

at the Chinese University of Hong Kong in September 2010

Soybean is an important economic crop and its production can be severely affected by salinity stress. At present, the soybean response to salinity stress is not clear. In my studies, I tried to understand this process from the perspective of proteomics and epigenetics, especially histone modifications.

Proteomics studies with 2-DE revealed that salt treatment may affect soybean photosynthesis and chloroplast formation. Comparison between the proteomic profiles of salt tolerant soybean variety (wild type) and salt sensitive soybean variety (cultivated, Union) indicated that protein levels in the detoxification and defense pathway as well as energy metabolism were higher in the wild type soybean, while the process of protein metabolism was less active. In addition, proteomic profiles of the cultivated soybean roots at different developmental stages were also compared to identify proteins related to soybean development. The expression of proteins which play critical roles in detoxification and defense pathways were higher at the seedling stage, especially the proteins which regulated the formation of ROS.

Histone modifications and histone variants are of importance in many biological processes. Whether they play some roles in regulating soybean salinity stress response is unknown. Previously, no study of histone modifications and histone variants in soybean were reported. In this study, I elucidated that in soybean leaves, mono-, di- and tri-methylation at Lysine (K) 4, 27 and 36, and acetylation at Lysine 14, 18 and 23 were present in histone H3. Moreover, H3K27 methylation and H3K36 methylation usually excluded each other. Although H3K79 methylation was

not reported in *Arabidopsis*, they were detected in soybean. In soybean histone H4, Lysine 8 and 12 were acetylated. In addition, the variants of histone H3 and H4 and their modifications were also determined. The variants of histone H3 were different at positions of $A^{31}F^{41}S^{87}S^{90}$ (histone variant H3.1) and $T^{31}Y^{41}H^{87}L^{90}$ (histone variant H3.2), respectively. Lysine 4 and 36 methylation were only detected in histone H3.2, suggesting that histone variant H3.2 might associate with actively transcribing genes. The two variants of histone H4 (H4.1 and H4.2) were different at amino acid 60. Moreover, I also found that the abundance of most of the histone modifications and histone variants did not change under the salinity stress except that H3K79 methylation would be up-regulated by the salinity stress.

In a parallel study, a PHD (plant homeodomain) finger domain containing protein, GmPHD1, was able to decipher the 'code' underlying H3K4 methylation. GmPHD1 was ubiquitously expressed in soybean and its expression increased upon salinity stress. GmPHD1 could bind to histone H3K4 methylation, with the preference to H3K4 dimethylation. It could then recruit several proteins, which were GmGNAT1, GmElongin A, and GmISWI. The interaction between GmPHD1 and GmGNAT1 was regulated by the self-acetylation of GmGNAT1. GmGNAT1 could also acetylate histone H3; GmElongin A was a transcription elongation factor; and GmISWI was a chromatin remodeling protein. Our data also indicated that the GmPHD1 located at the promoter of several soybean salt stress inducible genes. Therefore, the GmPHD1 recruited proteins to remodel the chromatin structure and facilitate the transcription of those salt stress inducible genes. Moreover, GmGNAT1 exhibited the preference to acetylate histone H3K14, therefore representing a kind of histone crosstalk between H3K4 methylation and H3K14 acetylation.

中文摘要

大豆是非常重要的經濟作物，但鹽脅迫對其產量有極大的影響。目前我們對大豆鹽脅迫的反應機制還不是特別清楚。本課題將從蛋白質組學和表觀遺傳學（尤其是組蛋白修飾）的角度對此進行研究。

蛋白質組學的研究表明鹽脅迫會抑制大豆的光合作用。對野生型（耐鹽品種）和栽培型（敏鹽品種）大豆的蛋白質組的比較結果顯示野生型大豆中解毒途徑和能量代謝途徑更活躍，但栽培型大豆中蛋白質的新陳代謝則更迅速。在不同的生長發育階段，大豆根蛋白質也呈現一種動態變化。許多參與解毒途徑的蛋白質在苗期的表達量比在芽期的表達量明顯增高，如過氧化酶。

組蛋白修飾和異構體是表觀遺傳學的重要組成部分，他們參與生物體內眾多的生理生化過程。但他們如何參與植物鹽脅迫反應尚無研究。之前尚無大豆組蛋白修飾和異構體的研究報導。此研究發現在大豆葉片中，組蛋白 H3K4，H3K27 和 H3K36 可被單、雙和三甲基化，H3K14，H3K18，H3K23， H4K8 和 H4K12 可被乙醯化。此外被甲基化修飾的 H3K27 和 H3K36 一般不共存於同一組蛋白 H3 上。雖然 H3K79 在擬南芥中沒有甲基化，但此修飾卻存在於大豆中。

此研究還鑒定了大豆組蛋白 H3 和 H4 的異構體及他們的修飾。兩個組蛋白 H3 異構體在 4 個氨基酸位點有不同，分別是 $A^{31}F^{41}S^{87}S^{90}$ (異構體 H3.1) 和 $T^{31}Y^{41}H^{87}L^{90}$ (異構體 H3.2)。甲基化修飾的 H3K4 和 H3K36 主要存在於異構體 H3.2 中，暗示異構體 H3.2 和活性轉錄的基因相關聯。兩個組蛋白 H4 異構體在第 60 位氨基酸不同。鹽脅迫下，H3K79 的甲基化會上升，但其他大多數的組

蛋白修飾和異構體沒有明顯的變化。

許多研究表明甲基化修飾的 H3K4 是活躍轉錄的基因的標誌。本研究表明核蛋白 GmPHD1 可以“解碼” 甲基化修飾的 H3K4。GmPHD1 在大豆中廣泛表達而且鹽脅迫下表達上升。GmPHD1 通過其 PHD 指結構識別甲基化修飾的 H3K4，並且對雙甲基化修飾的 H3K4 具有最大親和力。ChIP 的結果顯示 GmPHD1 分佈在某些大豆鹽脅迫誘導基因的啓動子區。GmPHD1 可以和 GmGNAT1, GmElongin A 及 GmISWI 相互作用。GmGNAT1 是組蛋白乙醯轉移酶；GmElongin A 參與基因轉錄的延伸過程；GmISWI 是染色質重塑蛋白。GmGNAT1 也可對自身進行乙醯化進而調節其與 GmPHD1 的相互作用。GmPHD1 因而通過這些機制調節大豆的基因轉錄和表達。GmGNAT1 主要促進組蛋白 H3K14 乙醯化，表明植物的組蛋白修飾之間也存在“交流”。

Acknowledgements

I would like to express my sincere thanks to my supervisors, Prof. NGAI Sai-Ming and Prof. LAM Hon-Ming, for giving me the chance to study these interesting projects. During my Ph.D study, Prof. NGAI and Prof. LAM gave me not only a good training in the techniques, but also the freedom and the way of learning and thinking independently. I am indebted to Prof. NGAI and Prof. LAM, not only for their patient guidance and continuous encouragement throughout my study, but also for their invaluable suggestions and critical comments on the preparation of this thesis.

I would like to express my faithful gratitude to the members of thesis committee, Prof. Jiang Liwen, Prof. Wong Yum Shing, and Prof. Wong Ngok Shun in Hong Kong Baptist University. Their suggestions and comments on my studies and this thesis are really constructive. I am also grateful to Prof. Cheung Chi Keung's help and kindness in these years.

Special thanks go to Ms. Helen TSAI in Prof. NGAI's lab. With her expert technical assistance, I could carry out my projects smoothly. Also I would express my thanks to Miss Wong Fuk Ling in Prof. LAM's lab for her help in my preparing experimental materials.

Special thanks also go to all labmates in Prof. NGAI's lab and Prof. LAM's lab for their friendliness and unselfish help during my study. All other fellow students in AoE lab and technician and staff in the Department of Biology are also nice to me and provide me with lots of help, I am really grateful. With their direct or indirect support and encouragement, my study and life in Hong Kong were easy and happy!

Last but not the least; I would like to give heartfelt thanks to my family, not only for their boundless love that gave me strength to overcome all the difficulties, but also for their support that enabled me to devote myself to my Ph.D study.

Thank you all!

June, 2010

Wu Tao

Table of contents

General abbreviations

2DE	Two-dimensional gel electrophoresis
AFLP	Amplified fragment length polymorphism
APX	Ascorbate peroxidase
bp	Base pair
BPTF	Bromodomain PHD finger transcription factor
cDNA	Complementary DNA
ChIP	Chromatin immunoprecipitation
CID	Collision induced dissociation
cm	Centimeter
Da	Dalton
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleoside triphosphate
Dot1	Disruptor of telomeric silencing
g	Microgram
GNAT	GCN5-related N-acetyltransferase
GST	Glutathione *S*-transferase
HDAC	Histone deacetylase
HAT	Histone acetyltransferase
ICAT	Isotope coded affinity tag
IEF	Isoelectric focusing
ING	*In*hibitor of growth
IPG	Immobilized pH gradient
iTRAQ	Multiplexed isobaric tagging technology
K	Lysine
LC	Liquid chromatography
MALDI	Matrix assisted laser desorption ionization
MBP	Maltose binding protein

mg	Milligram
MH^+	Monoisotopic peptide mass
MLS	Wild above ground tissues germination stage
MLG	Wild above ground tissues germination stage
ml	Milliliter
mm	Millimeter
mM	Millimole
MRG	Wild root germination stage
MRS	Wild root seedling stage
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MW	Molecular weight
MudPIT	Multidimensional protein identification technology
OD	Optical density
PCR	Polymerase chain reaction
PEV	Position effect variegation
PHD	Plant homeodomain
PMF	Peptide mass finger printing
PPP	Pentose phosphate pathway
PTM	Posttranslational modifications
PR-HPLC	Reversed-phase high performance liquid chromatography
R	Arginine
RAPD	Random amplification of polymorphic DNA
ROS	Reactive oxygen species
S/N	Signal to noise
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
TCA cycle	Tricarboxylic acid cycle
TOF	Time-of-flight
ULS	Union above ground tissues seedling stage

ULG	Union above ground tissues germination stage
URS	Union root seedling stage
URG	Union root germination stage
UV	Ultraviolet
v	Volume
V	Voltage
Vh	Voltage hour
w	Weight
µl	Microliter
µM	Micromole

Abbreviations of chemicals

ACN	Acetonitrile
CHAPS	3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate
dNTP	Deoxyribonucleoside triphosphate
DTT	Dithiothreitol
EDTA	2,2',2",2'"-(ethane-1,2-diyldinitrilo) tetraacetic acid
HCl	Hydrochloric acid
IAA	Iodoacetamide
KCl	Potassium chloride
KLH	Keyhole Limpet Hemocyanin
$MgCl_2$	Magnesium chloride
NaCl	Sodium chloride
$NaHCO_3$	Sodium bicarbonate
Na_2CO_3	Sodium carbonate
PBS	Phosphate buffered saline
PMSF	Phenylmethylsulfonyl fluoride
PVDF	Polyvinylidene difluoride
TCA	Trichloroacetic acid
TFA	Trifluoroacetic acid

List of tables

List of figures

Chapter 1 Introduction

1.1 Salinity stress and plants' response to salinity stress

Salinity stress, which is caused by the accumulation of excessive amount of salts (mainly sodium chloride) in the soil, is one of the most severe abiotic stresses that constraint not only crop plant growth but also crop productivity in many parts of the world. Both natural processes and human activities results in salinization of soil. Natural environmental factors such as backflow of seawater into the seashore region could elevate salt contents in the soil (Jain and Selvaraj, 1997). However, prolonged irrigation is the major human activity causing salinization in agricultural lands, particularly in arid and semi-arid region (Ashraf, 1994; Kozlowski, 2000). Today, nearly 340 million hectares of irrigated land, approximately one-third of the world's irrigated land, are salt-affected and unsuitable for crop cultivation (Kozlowski, 2000; Owens, 2001). As for China, more than 7 million hectares of land are classified as saline (Sun, 1987). It has been predicted that salinization of agricultural land will affect 30% of cultivated land within the next 25 years, and this will increase to 50% by the year 2050 (Wang *et al.*, 2003).

Salinity imposes three kinds of stresses on plants: (1) water deficiency that results from the physiological water loss because of the relatively high salt concentrations in the soil; (2) ion toxicity and nutritional imbalance resulting from accumulation of high concentration of Na^+ and Cl^- in the cytosol and following impairment in nutrient acquisition, such as the Ca^{2+}, K^+ input (Blumwald *et al.*, 2000); and (3) oxidative stress caused by the reactive oxygen species (ROS), including hydroperoxyl radical (HO_2^-), superoxide ($O^{\cdot 2-}$), hydrogen peroxide (H_2O_2) and hydroxyl radical ($OH^{\cdot}$), +which can react with and damage proteins, membrane lipids, and DNA (Dat *et al.*, 2000).

Despite of the severe stress that the salinity caused, plants have evolved several mechanisms to cope with it, including salt exclusion, selective ion uptake, ion compartmentation and ion retranslocation (Blumwald *et al.*, 2000; Shi *et al.*, 2000; Zhang *et al.*, 2001; Xiong and Zhu, 2002). For example, high concentrations of Na^+

and Cl^- ion can be sequestered into vacuoles by ion transport (Hamada *et al.*, 2001). Many ion transport proteins that involved in ion uptake and transport and ion homeostasis play important roles in these processes. In some plants, some specialized tissues have been developed to deal with high salt in the soil. In the mangrove *Atriplex spp.*, excess salts in the cell can be stored in salt glands or salt bladders and then be secreted out the leaves (Hamada *et al.*, 2001). Another strategy that many plants adopte to cope with the stress is the accumulation of compatible osmolytes which are low molecular weight hydrophilic compounds carrying no net charge at physiological pH, such as proline and glycine-betaine. Usually, they will not affect normal metabolic reactions and can facilitate water absorption by lowering cellular osmotic potential. Accumulation of these compatible osmolytes have been observed in many plants under abiotic stresses (Yeo, 1998; Xiong and Zhu, 2002).

1.2 Soybean

Soybean (*Glycine max (L.) Merrill*) is one of the most important economical legume crops in the world because of its richness of nutritional compositions including protein, oil, sugar and fiber in the seed. A 60-pound of soybeans yields about 48 pounds of protein-rich meal and 11 pounds of oil. It is estimated that soybean contributes to 30% of edible vegetable oil and 69% of high-protein feed supplements worldwide (http://www.soystats.com/2008/Default-frames.htm). Moreover, other fractions and derivatives of the soybean seed have substantial economic importance in a wide range of industrial, food, pharmaceutical, and agricultural products. At present, China is the fourth largest soybean production country (http://www.soystats.com/2008/Default-frames.htm).

Previously, depending on their threshold salinity level (maximum salinity without yield loss), crop species have been categorized into four groups: sensitive, moderately sensitive, moderately tolerant and tolerant (Maas and Hoffman, 1977). Soybean has many varieties all over the world, exhibiting variability in their level of injury under the salt stress. Soybean yield of sensitive cultivars is decreased dramatically under salt stress. Soybean yield was 80% at 4.0 dS/m and 44% at 6.7

dS/m versus 100% at 0.8 dS/m (Lee *et al.* 2009).

1.3 Proteomics

1.3.1 Introduction to proteomics

Proteomics is the study of the proteome, a term to describe "the analysis of the entire PROTEin complement expressed by a genOME, or by a cell or tissue type" (Wilkins *et al.*, 1996).

Recently, great progresses have been achieved in the field of proteomics. On one side, large numbers of genome have been sequenced, which provide the 'blueprint' of the possible gene products and made the possibility of identification of the proteins by searching against these database. On the other side, highly advanced technology have been developed, which greatly increase the sensitivity and accuracy of the proteomcis (Kavallaris and Marshall, 2005).

Two-dimensional gel electrophoresis (2DE) coupled with protein identification by mass spectrometry (MS) have always been the most widely used tools in the proteomic studies since its appearance in the late 70's (O'Farrell, 1975) because of its capacity in separating and visualizing a large number of proteins at one time. In the first dimension of isoelectric focusing (IEF), a mixture of proteins is separated on the basis of their corresponding isoelectric points (pI) by ampholyte gradients. While in the second dimension, the separated proteins in the first dimension will be further separated according to their molecular weights (MW) (Figure 1.1). Nowadays, the immobilized pH gradients (IPGs)-based 2DE technology has greatly enhanced the reproducibility, handling, resolution, and separation of very acidic and/or basic proteins in the proteomics.

Mass spectrometry (MS) is able to identify the proteins with extremely high sensitivity and accuracy according to the mass to charge ratios (*m/z*) of peptides. During the MS analysis, the sample must first be ionized to generate ions by the ionization sources, for example Electrospray Ionization (ESI) (Fenn *et al.*, 1989) or Matrix-Assisted Laser Desorption Ionization (MALDI) (Tanaka *et al.*, 1988), which allow the transfer of large, polar, thermally labile biomolecules into the gaseous

phase for mass analysis. Then these ions will be separated by mass analyzers such as time-of-flight (TOF) according to their *m/z* ratios and their signals are detected by ion detectors (Lane, 2005). By determining a series of accurate masses of peptides resulting from a digested unknown protein, a peptide mass fingerprint (PMF) of the unknown protein can be obtained. With the reference of PMF, data can be compared to *in silico* fingerprints obtained by theoretical cleavage of protein sequences stored in databases, and the top-scoring proteins are retrieved as possible candidate proteins (Lane, 2005). Nevertheless, protein identification can be further confirmed by tandem MS/MS analysis, in which some peptides detected in MS analysis are selected for further fragmentation and their *m/z* ratios will also be recorded. The potential amino acid sequence of the peptide could be deduced from the generated fragment ions. Similarly, the posttranslational modifications on the peptide can be identified based on the mass shift of the peptides by comparing their detected mass and theoretical mass. For example, if the mass of a peptide shift about +14Da from its theoretical mass, then it is likely that a methyl group ($-CH_3$) is present in this peptide. With the tandem MS/MS analysis, we can even pinpoint where this group locates. Therefore mass spectrometry can also be applied to globally identify the protein posttranslational modifications, such as methylation, acetylation and phosphorylation etc.

In addition to the classical 2D-PAGE, other proteomic methods which also be able to supply the quantitative information of the differentially expression proteins have been developed. Some of them are based on the stable isotope labeling of peptides, such as isotope-coded affinity tag (ICAT), in which peptides originating from two different samples are labeled either with a heavy or a light ICAT at cysteine residues before they are mixed. Similarly, iTRAQ, which stands for isotopic tagging for relative and absolute quantification, would targets primary amines and thus labels all peptides in a mixture. Subsequently, the peptides with different isotopic tag can be distinguished in following mass spectrometric analyses by a characteristic mass shift and their relative quantities in the different samples can also be calculated by comparing their intensity. These approaches deliver quantitative data in temporal and

spatial resolution, reveal functional interactions between proteins and protein complexes, and depict global proteome dynamics and proteome performances (Baginsky, 2009).

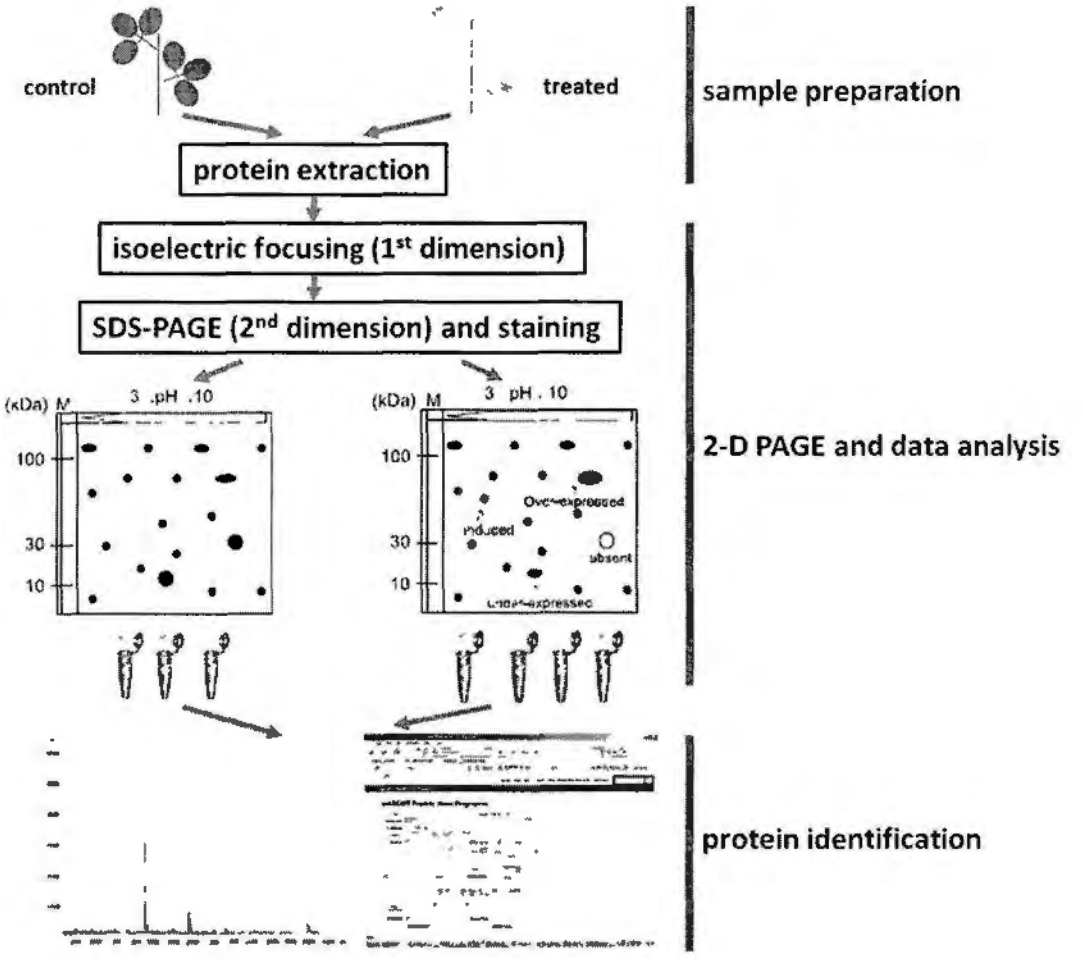


Figure 1.1: Demonstration of the processes of the classical 2D-PAGE (modified from Qureshi *et al.*, 2007). The proteins are extracted from two experimental samples. After separation in IEF and SDS-PAGE gel, the differentially expressed proteins between these two samples were selected out and identified by mass spectrometry.

1.3.2 Why proteomics?

Several high-throughput RNA measurement tools such as differential display and cDNA microarrays have been developed for the analysis of the transcriptome. Studies in *Arabidopsis* using cDNA microarray have revealed that the plants

regulated gene expression to increase their tolerance to salinity stress (Seki *et al.*, 2002). Some genes have good correlation at the mRNA and protein levels, however, in some case, the amount of mRNA do not offer insight into the quantity and quality of their final gene products, namely the proteins, such as rice SALT and tobacco osmotin (Jiang *et al.*, 2007; Qureshi *et al.*, 2007). Large differences in mRNA and protein turnover, for example some mRNAs could not be translated but degraded rapidly, could partially account for the phenomenon; in addition, some posttranslational modifications or regulations, such as removal of signal peptides, phosphorylation and glycosylation, which can affect the proteins' subcellular localization, stability or enzyme activities, could occur after the protein synthesis without any detectable change in transcript abundance. Considering this significant distinctness, only the study of proteins themselves provides information on their real amount and activity under certain given conditions. Therefore, proteomic studies, which are able to detect the protein amount and their modifications directly, are necessary to compensate for the microarray studies (Zivy *et al.*, 2000).

1.3.3 Proteomic studies in plants

Most plant tissues do not provide a ready source of proteins and several factors severely affect the plant proteomic studies. Firstly, the protein may be only a small part of the plant biomass, since plant cell wall and the vacuole make up the major part, with the cytosol representing only 1–2% of the total cell volume. Secondly, plant cells contain many interfering substances such as phenolic compounds, proteolytic and oxidative enzymes, terpenes, pigments, organic acids, inhibitory ions, and carbohydrates which may be responsible for irreproducible and inferior results such as proteolytic breakdown, streaking, and charge heterogeneity in the plant proteomic studies (Carpentier *et al.*, 2005).

Although the aforementioned problems still remain, plant proteomics have gain great progress recently. In order to characterize the plant proteomic patterns, several methods were applied to various plant samples to enhance the quality of the proteomic results. Combination of selected proteomic methods, such as 2-DE,

multidimensional protein identification technology (MudPIT) and those for quantitative proteomics including DIGE, isotope coded affinity tag (ICAT), Multiplexed isobaric tagging technology (iTRAQ), were employed to facilitate the proteomic studies in plants. To enhance the identification of low abundance proteins, proteins from distinct plant organelles were extracted for following proteomic studies. With these efforts, several proteome analysis of cells, calli, seeds, roots, stem, leaves, xylem/phloem sap, pollen or whole seedlings, in relation to different aspects of plant biology, from growth and development to stress responses, have been carried out (Jorrín *et al.*, 2007).

Plants are not mobilized to avoid the stress they encounter in their growth condition, their response to these stresses are fascinating to many scientists, especially in some important agriculture plants. Proteomics has been applied to study the response of *Arabidopsis*, rice, wheat, barley, pea, and many other plants to various abiotic stress, including cold, drought, light, salt, metal and other stresses (Ndimba *et al.*, 2005; Jiang *et al.*, 2007; Jorrín *et al.*, 2007). Proteins that can counter these stresses directly (such as antioxidant enzymes and chapronins) or indirectly (such as enzymes in osmolyte synthesis) were reported to be up-regulated in the plants upon the stress. Proteomic studies of different plant varieties were also carried out, revealing that pathogenesis-, stress-related proteins and antioxidant enzymes were highly expressed in stress tolerant/resistant genotypes, while the enzymes of the photosynthesis and energy metabolism decreased in the susceptible ones (Jorrín *et al.*, 2007). These studies implied that 2D electrophoresis might also function as the DNA fingerprinting techniques, such as random amplification of polymorphic DNA (RAPD) or amplified fragment length polymorphism (AFLP), to visualize the genetic markers which are associated with the variation of quantitative traits in plant and are of importance in plant breeding (Zivy *et al.*, 2000).

As for the salinity stress, several proteins have been identified to be upregulated in plants under this kind of stress via the proteomic studies. Osmoprotectants which can regulate the osmotic pressure of the cells have been identified in many proteomics of plants with salinity stress. The most well known osmoprotectants are proline, glycine

betaine and osmotin containing the signature motif of soybean Kunitz trypsin inhibitor. Ascorbate peroxidase (APX) and other peroxidases used to reduce hydrogen peroxide in the cells were also found to be upregulated under the salinity stress. Several signaling pathways were also induced by the salinity stress, including salt-overly sensitive (SOS) pathway proteins, abscisic acid (ABA) and jasmonic acid (JA) signaling pathway. Other enzymes associated with metabolism, such as V-ATPase, β-glucosidase, glutamate ammonia ligase, adenosine kinase (ADK) and proteins in sulphur metabolism, all showed some relationships with the plants' response to salinity stress as revealed by many proteomic studies (Qureshi *et al.*, 2007). Together with all the proteomic discoveries, the processes involved in detoxification, homeostasis maintenances and growth regulation are definitely of importance in the plants' responses to the salinity stress (Zhu, 2001).

1.4 Epigenetics

1.4.1 Introduction to epigenetics

Nowadays, epigenetic mechanisms are considered as important regulation strategies in all organisms. Epigenetics, which is termed as 'heritable changes in gene expression not attributable to nucleotide sequence variation' (Murrell *et al.*, 2005), usually includes two mechanisms: DNA methylation and histone modifications and histone variants.

1.4.1.1 DNA methylation

DNA methylaiton can be found on cytosine. The 5'- position of cytosine is methylated in a reaction catalyzed by DNA methyltransferases (DNMTs) with S-adenosyl-methionine as the methyl donor. Symmetric DNA methylation occuring at both strands of CG dinucleotides is observed in both animals and plants. However, DNA methylation at CNG (where N is any base) and nonsymmetric CHH trinucleotides (where H is any base except G) is unique to plants. These types of DNA methylation are mediated by the plant-specific chromomethylase CMT3, as well as by DRM1 and DRM2 (Schöb *et al.*, 2006). In animal somatic cells, DNA methylation patterns are copied by the maintenance DNA methyltransferase I

(DNMTI) positioned at the replication forks, with some cooperation of DNMT3a and DNMT3b (Miranda *et al.*, 2007). Thus DNA methylation, maintained through mitosis, is considered a stable epigenetic mark.

In *Arabidopsis*, DNA methylation can be found in a significant fraction (20–33%) of genes besides the transposon rich heterochromatic regions (Zhang *et al.*, 2006). They are dynamic, increasing throughout *Arabidopsis* development, from cotyledons to vegetative organs to reproductive organs (Ruiz-García *et al.*, 2005). Although it is considered that DNA methylaiton is stable previously, active demethylation has been observed in both plants and animals recently. DNA glycosylases of the DEMETER (DME) family are responsible for removing methylcytosines from the 5' and 3' end of genes in *Arabidopsis* (Penterman J, *et al.*, 2007). About 179 loci of the genome are demethylated by DME (Penterman J, *et al.*, 2007).

It is well known that DNA methylation regulate gene expression by silencing genes and repetitive elements. It is suggested that DNA methylation can directly impede the binding of transcriptional factors to their target sites and recruits methyl-binding proteins (MBPs) that specifically bind to methylated CpG sites (cytosine-phosphate-guanine), thus prohibiting transcription. They also play important roles in chromatin organization and genomic imprinting. They are usually associated with the formation of heterochromatin by affecting histone modifications and nucleosome occupancy (Miranda TB, *et al.*, 2007).

1.4.1.2 Histone modifications and histone variants

The fundamental structural unit of chromatin in eukaryotic cells is the nucleosome that consists of 146 base pairs (bp) of DNA wrapped around a histone octamer, each of which is formed by two copies of H2A, H2B, H3 and H4 (Marino-Ramirez, *et al.*, 2005). An additional histone, H1 links these nucleosomes together along the chromatin chain. In general, the N terminus of histone H3 and H4, and N and C terminus of H2A and H2B are prone to being covalently modified by many enzymes, such as HMT (histone methyltransferase) and HAT (histone acetyltransferase). These modifications include methylation, acetylation, phosphorylation, ubiquitination, glycosylation, ADP ribosylation, carbonylation, sumoylation and biotinylation. By

using Western blotting and mass spectrometry, increasing histone modification sites are discovered in mouse, yeast, *Drosophila*, *Tetrahymena* and *Arabidopsis* (Figure 1.2) (Allis, *et al.*, 2007; Fuchs, *et al.*, 2006; Johnson, *et al.*, 2004).

In addition to the canonical histones which package the chromatin and whose transcription are tightly coupled to DNA replication, there are some other histone genes which are constitutively expressed and encode non-canonical histone variants with some differences in primary amino acid sequence from their canonical paralogues. Recent studies have showed that non-canonical variants are more diverse in their functions than the canonical histones, ranging from DNA repair, meiotic recombination, chromosome segregation, transcription initiation and termination, sex chromosome condensation to sperm chromatin packaging (Talbert and Henikoff, 2010). In both *D. melanogaster* and human cells, H3.3 is deposited into transcribed genes, associating with transcriptional induction and elongation (Henikoff, 2008). Consistent with their different functions between canonical and non-canonical histones, the histone posttranslational modifications in them also exhibit some differences. In Drosophila, marks associated with transcriptional activity are enriched in H3.3 rather than H3.1 (McKittrick *et al.*, 2004).

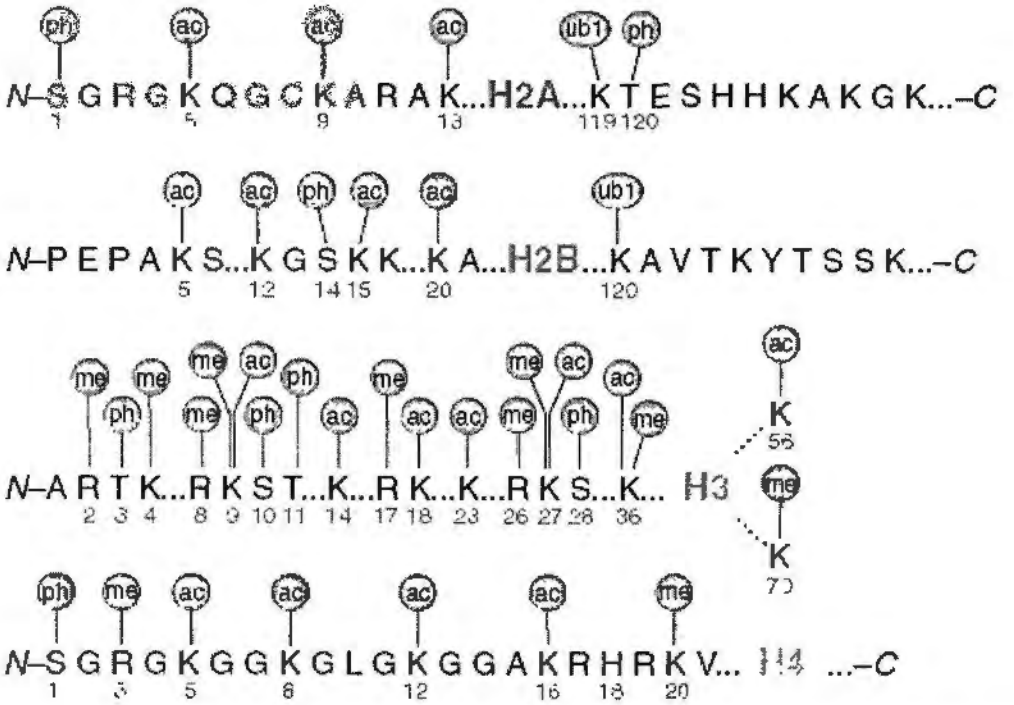


Figure 1.2: Posttranslational modifications of nucleosomal histones. These modifications include acetylation (ac), methylation (me), phosphorylation (ph) and ubiquitination (ub1). Globular domains of each core histone are represented as colored ovals (Bhaumik *et al.*, 2007).

1.4.1.3 Relationship between DNA methylation and histone modifications

Many reports have indicated that DNA methylation cooperated with histone modifications to participate in heterochromatin formation. DNA methyltransferases interact with the histone methyltransferases SUV39 and EZH2, which is responsible for H3K9 and H3K27 methylation respectively. Studies showed that SUV39 is important in genomic DNA methylation formation in both *Neurospora* and *Arabidopsis* and is required for Dnmt3b-dependent DNA methylation at pericentric repeats in mammal (Miranda TB, *et al.*, 2007).

Another protein that DNA methyltransferases interact with is the heterochromatin protein 1 (HP1), which specifically binds to methylated lysine 9 on histone H3 and is vital in the heterochromatin formation. It is likely that HP1 bind to H3K9 trimethylation and then recruit the DNA methyltransferases to these loci (Smallwood *et al.*, 2007). Taken together, these results suggest that trimethylation of H3K9 and

H3K27 are important markers of DNA methylation.

1.4.2 Histone modifications: functions, regulations and working mechanisms

1.4.2.1 The functions of histone modifications are diverse

Histone modifications play important roles in many fundamental biological processes by rearranging the structure and composition of chromatin. Inside the eukaryotic cells, such re-structuring events can help partition the genome into distinct domains such as euchromatin and heterochromatin and result in other processes, such as DNA transcription, DNA repair and DNA replication (Khorasanizadeh *et al.*, 2004; Komili *et al.*, 2008).

The "status" of chromatin might vary substantially between different chromosomal positions and different nuclear sites. Chromatin has been divided into two main classes based on structural and functional criteria: euchromatin and heterochromatin. Modifications on histones, like methylation and actylation, are fundamental for the formation of chromatin domains. They can be a landmark to indicate the environment of chromatin. For example, condensed chromatin is enriched in methylation of H3 lysine 9 and H3 lysine 27 (Nakayama *et al.*, 2001; Bártová *et al.*, 2008; Adcock *et al.*, 2006).

In addition, histone modifications have more roles in the cells. H3 Serine 10 phosphorylation is involved in the chromosome condensation, so they are also important in the cell cycle and cell mitosis (Li *et al.*, 2007; Berger, 2007). As the basic structure of chromatin, it is not surprising to find that histone modifications either involved in DNA repair, such as H3K79 methylation and H4K20 methylation (Zhou *et al.*, 2006; Huyen *et al.*, 2004; Sanders *et al.*, 2004). Moreover, several histone modifications have been correlated with gene transcription as the following discussions. While acetylation and H3K4 methylation are mainly associated with active genes; H3K9 and H3K27 methylation are leading to gene silencing (Li *et al.*, 2007; Berger, 2007).

1.4.2.2 Regulation of histone modifications: histone modifications are reversible

On one side, as aforementioned, histone modifications are functionally diverse, sometimes even leading to just opposite consequence, for example, transcription activation or inhibition. Interestingly, corresponding with their diverse functions, different histone modifications have distinct distribution patterns along the chromatin. For example, acetylated histones mainly located at the beginning of genes, while the distributions of histone methylation are very complicate. H3K4 trimethylation is localized to the 5' ends of ORFs. H3K36 trimethylation, however, is more broadly distributed throughout the ORF, peaking at the 3' end (Schübeler *et al.*, 2004). On the other side, histone modifications are dynamic, varying in different cell types and developmental stages (Shechter *et al.*, 2009; Nicklay *et al.*, 2009). Therefore, the histone modifications must be delicately modulated to make sure that they are working properly, just at the right time and at the right place. Previously, many enzymes responsible for modifying the histones have been identified, such as histone methyltransferase and histone acetyltransferase (Allis, 2007; Yang and Seto 2008). Nowadays, many studies have indicated that histone modifications, including methylation, acetylation, phosphorylation, ubiquitination, sumoylation and biotinylation, were also dynamic and reversible.

Numerous enzymes that can remove these modifications from histones have been isolated. It has long been known that histone acetylation is dynamic and generally removed by histone deacetylase (HDAC) (Allis, 2007; Yang and Seto 2008). Previously, it was anticipated that histone methylation was much more stable. However, this view was completely changed with the identification of the amine oxidase LSD1 as an H3K4me2/me1 specific demethylase (Shi *et al.*, 2004). Currently, more and more histone demethylase have been identified, including many Jumonji C (JmjC) domain containing proteins. They usually specifically remove certain kind of modifications. Similarly, other modifications can also be reversed by their correspondent enzymes (Agger *et al.*, 2008).

Recently, a more dramatic way through which multiple marks are simultaneously removed from the N tail of histone H3 is reported in mouse and yeast (Santos-Rosa *et al.*, 2009; Duncan *et al.*, 2008). Endopeptidase enzymes are responsible for this

mechanism. In mouse embryonic stem cells, cathepsin L cleaves after the first 21 amino acid residues of the N terminus of H3. And in yeast, the as-yet-unidentified enzyme is likely to be a serine protease which also cut H3 at the same site. Interestingly, the cleavage is regulated by covalent modifications present on the histone tail itself. By removing these modifications in the N terminus, H3-tail clipping can then regulate gene transcription and nucleosome displacement (Santos-Rosa *et al.*, 2009; Duncan *et al.*, 2008).

1.4.2.3 Working mechanisms of histone modifications: 'Histone code' hypothesis

In order to fulfill the multiple roles of histone modifications, many proteins that function as the effectors of histone modifications have been characterized. In 2001, the concept of "histone code" was raised to explain how the histone modifications work. It is speculated that those effectors of histone modifications will recognize some histone modifications and then execute their functions in many physiological processes (Jenuwein and Allis, 2001).

1.4.2.3.1 Histone modifications in nucleic structure: heterochromatin formation

Generally, the chromatin in the interphase can be classified into euchromatin and heterochromatin in the nucleic. Euchromatin contains almost all of the genes, both actively transcribed and quiescent. Heterochromatin is transcriptionally inert and is generally more condensed than euchromatin. Moreover, heterochromatin is able to silence euchromatic genes, which occurs when genes are brought into juxtaposition with heterochromatin by a chromosomal rearrangement or transposition (Bártová *et al.*, 2008; Adcock *et al.*, 2006; Talbert and Henikoff, 2006).

For a long time, the scientists have been fascinated to try to understand how the heterochromatin was formed. A well studied example is the position effect variegation (PEV), where gene silencing is observed in a subset of cells that normally express a given gene (Singh *et al.*, 2008). Phenomena of PEV have widely been discovered in many organisms, such as yeast, *Drosophila* and mammal. Several mechanisms for gene silencing associated with PEV have been raised, one of which is that the alterations in chromatin structure would render a gene inaccessible to

regulatory factors. This may be the results of the spreading of a compacted chromatin state from heterochromatin to adjacent genes. During this process, histone modifications play an important role (Ebert *et al.*, 2006; Wakimoto, 1998; Schotta *et al.*, 2002). A depletion in one of the structural building blocks of heterochromatin, heterochromatin protein 1 (HP1), leads to a suppression of PEV (Eissenberg *et al.*, 1990). Very interesting, it is necessary for HP1 to recognize methylated H3 Lysine 9 via its chromodomain to induce heterochromatin formation. More interestingly, subsequent studies have revealed that HP1 was able to recruit other proteins, which include histone methyltransferase (HMT), to methylate H3K9 in the next nucleosomes, thus spreading the heterochromatin structure to the adjacent regions (Johnson *et al.*, 2002). In addition, DNA methyltransferase (DNMT) responsible for DNA methylation was also recruited by HP1 to the heterochromatin region, which further facilitated the heterochromatin formation (Nakayama *et al.*, 2001). Therefore, histone modification is supposed to be one factor in regulating heterochromatin formation.

1.4.2.3.2 Histone modifications in modulating the gene transcription and DNA damage repair

Genes are transcribed from DNA that is in complex with proteins, as chromatin. Hence, chromatin structure imposes great obstacles on gene transcription. Generally, gene transcription only occurs when the chromatin structure is opened up, with loosening of the tight nucleosomal structure allowing RNA polymerase II and other basal transcription complexes to interact with DNA and initiate transcription. Many chromatin regulators can "open" the "closed" chromatin, and histone posttranslational modifications contribute greatly in this process (Berger, 2007; Li *et al.*, 2007). Many researches have shown that histone modifications serve as the binding sites for different effector proteins which will then mediate the biological functions of the histone modifications, for example, H3K4 methylation is able to recruite many "effector proteins" to the chromatin, which will be discussed in more detail later.

Another mechanism that histone modifications use to regulate gene transcription is to open the structure of chromatin directly. Histone acetylation neutralizes the positive charge of histone, reducing the strength of binding of histones-DNA, and finally opens the DNA-binding sites (Berger, 2007; Li *et al.*, 2007).

To repair DNA damage, H3K79 methylations can target p53-binding protein 1 (53BP1) to DNA double-strand breaks and activate the DNA damage checkpoint pathways (Huyen *et al.*, 2004). Another report found that chromatin assembly factor 1 (CAF-1) also interact with H3K79 to regulate telomeric silencing and DNA repair (Zhou *et al.*, 2006). In addition, histone H4 Lysine 20 methylation is evolutionarily conserved from yeast to mammals and is also very critical in DNA repair and genome integrity (Sanders *et al.*, 2004).

1.4.3 A paradigm of 'histone code': H3K4 methylation and PHD finger domain containing proteins

1.4.3.1 H3K4 methylation is a marker of transcriptionally active genes

The hypothesis of 'histone code' conceived that some proteins can interpret the code encoded by the histone modifications (Jenuwein and Allis, 2001). Subsequent studies with many chromatin proteins which were able to recognize the histone modifications contribute the brick to the hypothesis (Kim *et al.*, 2006; Taverna *et al.*, 2007). Among theses many histone modifications, H3K4 methylation is one of the most clearly annotated 'code'. H3K4 can be mono-, di- and tri-methylated in the ε-amine of the lysine by a series of enzymes such as MLL-family, ASH1, SET7/9, SMYD3, and Meisetz, containing the SET domain (Ruthenburg *et al.*, 2007). Genomic-scale analyses of H3K4 methylation revealed that H3K4 trimethylation is mainly located in the 5' regions of virtually all transcriptionally active genes and that this modification is strongly positively correlated with transcription rates, active polymerase II occupancy, and histone acetylation. In vertebrates, majority of H3K4 dimethylation colocalizes with H3K4 trimethylation, however, in *S. cerevisiae*, it appears to spread throughout genes, peaking toward the middle of the coding region. And monomethylation most abundant at 3' ends of genes (Schübeler *et al.*, 2004;

Bernstein *et al.*, 2005; Ruthenburg *et al.*, 2007). Despite of this difference in their distribution patterns, one conclusion that we can definitely get from these studies is that H3K4 methylation is associated with transcriptionally active genes.

As a mark of active genes, H3K4 methylation must be able to recruit its effector proteins and bring about its downstream biological events. So far, a large number of proteins that are able to interact with H3K4 methylation have been discovered (Table 1.1) (Ruthenburg *et al.*, 2007). Most of these proteins contain a specific domain through which the methylated H3K4 is recognized specifically. According to the domains the proteins contain, they can be classified into two distinct categories: the royal superfamily (containing the chromodomain or tudor domain) and the PHD-finger superfamily (containing the PHD finger domain). Although these domains belong to two different superfamilies of folds, following structure analysis showed that there are some commonalities between them. The first most striking commonality is that they contain an aromatic cage, which is composed by aromatic ring containing amino acids, mediating their interaction with H3K4 methylation. Another similarity is that H3R2 methylation, another histone modification very near H3K4 methylation in histone H3, can determine the interaction between H3K4 methylation and its effector proteins (Ruthenburg *et al.*, 2007).

Proteins	Recognized modification sites	Recognition domains	Associated complexes
CHD1	Di, tri	chromodomain	SAGA/SLIK, Paf1
BPTF	Tri	PHD finger domain	NURF
ING1	Di, tri	PHD finger domain	mSin3A HDAC; SWI/SNF
ING2	Tri	PHD finger domain	mSin3A HDAC; SWI/SNF
ING3	Di, tri	PHD finger domain	hNuA4/TIP60 HAT
ING4	Tri	PHD finger domain	HBO1 HAT
ING5	Tri	PHD finger domain	HBO1 HAT; MOZ/MORF HAT
JMJD2A	Tri	Double tudor	

WDR5	Di	WD40	SET1A, MLL1 and MLL2 H3K4 HMTs

Table 1.1: Identified proteins that can interact with histone H3K4 methylation.

The effector protein alone is able to recognize the histone modification, but it is far away from enough to regulate the gene transcription. In fact, many of the known H3K4 methylation readers reside within protein complexes associated with enzymatic activities operating on the chromatin template: Chromatin remodeling protein CHD1 (chromo-ATPase/helicase-DNA binding domain 1) is able to bind to the methylated H3K4 directly via its chromodomain (Pray-Grant, *et al.*, 2005). It is known that Chd1 is a component of the transcriptional co-activators SAGA (Spt-Ada-Gcn5 acetyltransferase) complex, a conserved multi-subunit histone acetyltrasferase (HAT) complex, which particularly induces acetylation in histone H3 and H2B (Daniel and Grant, 2007). Therefore, the Chd1 serves as a platform for assembly of SAGA to activate gene transcription. In addition, CHD1 is involved in ATP-dependent chromatin remodeling by recruiting the member of ISWI/SNF superfamily which was able to hydrolysis ATP to regulate chromatin structure (Neely and Workman, 2002). The WD40 repeats of WDR5 are also shown to directly associate with H3K4 methylation with the preference to H3K4 dimethylation (Wysocka *et al.*, 2005). Structural studies then exhibited that WDR5 is likely to present the methylated H3K4 side chain for further methylation by the MLL family HMTase complexes, of which WDR5 is a component (Dou *et al.*, 2006). With these proteins, H3K4 methylation may facilitate transcriptional activation by increasing the accessibility of the chromatin template to the transcriptional machinery (Ruthenburg *et al.*, 2007).

1.4.3.2 PHD finger domain containing proteins function as 'histone code' readers

As aforementioned, the plant homeodomain (PHD) finger domain is also widely used by many chromatin proteins to recognize H3K4 methylaitons. PHD finger domain was first discovered in the *Arabidopsis* HAT3.1 homeodomain protein over a decade

ago. It comprises about 60 amino acids and shows the characteristic structure of Cys4-His-Cys3 (C4HC3) with some other conserved amino acids, most notably a tryptophan or other aromatic amino acid (Figure 1.3) (Bienz, 2006; Lee *et al.*, 2009). It is a Zn^{2+} binding domain and structurally very similar to the Ring finger which can be found in many E3 ligase that mediate ubiquitination process. However, it lacks the E2 ligase-interacting surface that is characteristic of many RING domains. They are present in many nuclear proteins and predicted to be able to associate with chromatin previously. This domain is conserved throughout the eukaryotic proteomic, including a large number of chromatin regulatory factors such as recombination activating gene 2 (RAG2), the acetyltransferase proteins CBP/p300, the chromatin remodeling protein ACF, BPTF (bromodomain PHD finger transcription factor), and the putative tumor suppressors, ING (*in*hibitor of *g*rowth) family. So far, at least 14 PHD finger-containing proteins have been found in the budding yeast, 50 in the fruit fly, and up to several hundred in humans (Wysocka, *et al.*, 2006; Martin, *et al.*, 2006; Bienz, 2006; Baker *et al.*, 2008).

```
AIRE1      DECAVCRDGGELIC----CD--GCPR-AFHLACLSPLREIPSGT-------WRCSSCLQ
Mi-2       EFCRVCKDGGELLC----CD--ACPS-SYHLHCLNPPLPEIPNGE-------WLCPRCTC
NURF301    DHCRVCHRLGDLLC----CE--TCPA-VYHLECVDPPMNDVPTED-------WQCGLCRS
WSTF       ARCKVCRKKGEDDKLIL-CD--ECNK-AFHLFCLRPALYEVPDGE-------WQCPACQP
KAP-1      TICRVCQKPGDLVM----CN--QCEF-CFHLDCHLPALQDVPGEE-------WSCSLCHV
ATRX       EQCRWCAEGGNLIC----CD--FCHN-AFCKKCILNLLRKELSTIMDENNQWYCYICHP

ING2       TYC-LCNQVSYGEMIG--CDNEQCPIEWFHFSCVSLTYKPKGK---------WYCPKCRG

ACF1-1     SLCKVCRRGSDPEKMLL-CD--ECNA-GTHMFCLKPKLRSVPPGN-------WYCNDCVK
ACF1-2     KVCQKCFYDGGEIK----CV--QCRL-FFHLECVHLKRPPRTD---------FVCKTCKP

P300       HFCEKCFNEIQGES-32-CT--ECGR-KMHQICVLHHEIIWPAG--------FVCDGCLK

Consensus  C--C-----------C----C-----H--C------------------W-C--C
                     Loop 1                        Loop 2
```

Figure 1.3: Sequence alignment of ten different PHD fingers. The sequence signature of PHD finger domain is indicated as the consensus. All sequences are from human proteins, except for Pygopus, ACF1 and NURF301, which are from *Drosophila* (Bienz *et al.*, 2006).

Following studies confirmed that this domain bind to the chromatin through

recognizing histone. First, many of the PHD domain containing proteins can recognize the methylated lysine through the aromatic cage and these proteins include BPTF, ING superfamily member and RAG2 (Figure 1.4) (Ramón-Maiques *et al.*, 2007). An extensive network of hydrogen bonds and complementary surface interactions are responsible for the unique recognition of H3K4 methylation in the histone by the PHD finger (Li *et al.*, 2006; Mellor, 2006). More interestingly, they can also engage H3R2 methylation through another pocket and an invariant tryptophan will separate these two K4me and R2me binding pockets. In addition, the binding to H3R2 methylation in different proteins will regulate their affinity to H3K4 methylation, for example, recognition of H3R2 methylation in the ING2, BPTF will inhibit their binding to H3K4me, whereas this does not happen in RAG2 (Ramón-Maiques *et al.*, 2007; Baker *et al.*, 2008). Lately, the PHD finger domain is then found to be able to bind H3K4 without any modification (H3K4me0) in DNMT3L and BHC80, raising the possibility that histone without modifications can also serve as a code which need some other proteins to annotate them specifically. Structural studies reveal that these PHD fingers recognize the H3K4me0 through a way which is quite different from that they employ in the H3K4 methylation recognition. The establishment of PHD-H3K4me0 interaction is mainly through an electrostatic bridge between the unmodified epsilon amino group of H3K4me0 and an acidic residue in PHD finger (Asp90 in DNMT3L or Asp489 in BHC80), and methylation at H3K4 sterically excludes such interaction (Baker *et al.*, 2008). These structural studies also imply that the PHD fingers which lack these aforementioned characteristics may not be able to bind to H3K4. Indeed, emerging evidence shows that some of them associate with other methylated lysines, for example, some PHD fingers in yeast bind to H3K36me and PHD fingers in SMCX and ICBP90 to H3K9me. In addition, many PHD fingers may recognize modifications other than methyl-lysines or have unknown functions (Baker *et al.*, 2008).

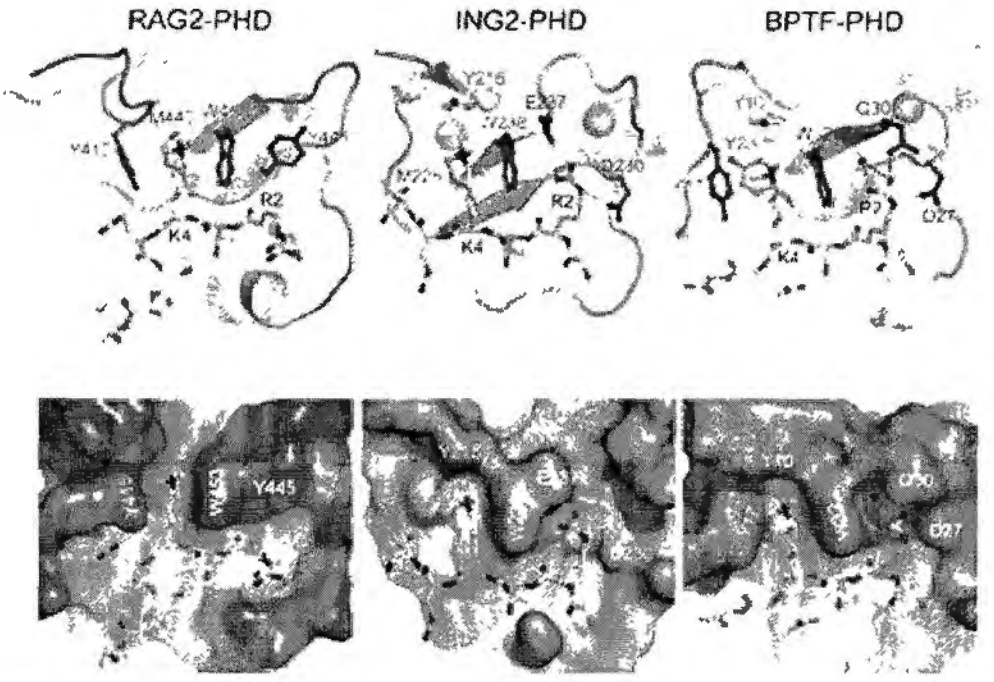


Figure 1.4: Ribbon diagrams (upper panel) and surface representation (lower panel) of PHD fingers of RAG2, ING2 (PDB entry 2G6Q), and BPTF (PDB entry 2F6J) bound to H3K4me3 peptide. RAG2-PHD has an open aromatic channel and does not possess a negatively charged residue equivalent to E237 in ING2 or D27 in BPTF to interact with R2 of the H3 peptide (Ramón-Maiques *et al.*, 2007).

In addition to recognize methylated lysine, PHD finger domains are thought to facilitate the interactions between the PHD containing proteins and other proteins. The ING protein family contains several nuclear proteins that share the highly conserved PHD finger domain at their carboxyl termini. Members of this family are widely found in multiprotein complexes that posttranslationally modify histones, such as histone acetylation and deacetylation (Table 1.1), suggesting that these proteins serve a general role in permitting various enzymatic activities to nucleosomes (Martin *et al.*, 2006). BPTF is important for the structural integrity of NURF (nucleosome remodeling factor), a protein complex using the energy of ATP hydrolysis to catalyze nucleosome sliding (Xiao *et al.*, 2001). Therefore, the NURF complex might be deposited to the beginning of active genes to regulate the chromatin structure and facilitate gene transcription by the binding of the BPTF to

the histone via its PHD domain recognizing H3K4 methylation and its bromodomain to lysines acetylation (Mellor, 2006; Wysocka *et al.*, 2006; Li *et al.*, 2006). In addition, some studies also suggested that PHD domain was able to bind phosphoinositide and served as a nuclear PtdInsP receptor. The PHD-phosphoinositide interaction then may directly regulate nuclear responses to DNA damage (Gozani *et al.*, 2003).

Since their abilities in recognizing histone code and involvement in many physiological processes, it is not surprising to find out that many PHD finger containing proteins were correlated with many diseases in mammals and human. In many immune diseases, mutations in the autoimmune regulator protein (AIRE) and the RAG2, which is the catalytic engine of V(D)J recombination and important for the B and T cell differentiation, have been reported. In addition, mutation in Inhibitor of Growth 1 (ING1) was also correlated with cancer development. More strikingly, many of these mutations were located in the PHD finger domain, which may subsequently result in the disruption of the structure of the PHD finger or interruption their interaction with H3K4 methylation. The functions of these proteins in the disease development further highlight the importance of these proteins and their interaction with H3K4 methylation (Baker *et al.*, 2008).

1.4.4 Studies of histone modifications in plants

Although histone modifications and their functions are well studied in yeast and mammals (Allis *et al.*, 2007; Fuchs *et al.*, 2006; Johnson *et al.*, 2004), such studies in plants are just at their infancy stage, focusing on the *Arabidopsis*. Recently, the variants of histone H2A, H2B, H3 and H4 and their modifications in *Arabidopsis* have been identified using mass spectrometry (Johnson *et al.*, 2004; Zhang *et al.*, 2007; Bergmuller *et al.*, 2007). These studies reveal some modification sites that are unique to plants (Zhang*et al.*, 2007). The genomic distribution patterns of several histone modifications (histone H3K4 di-methylation, H3K9 di-/tri-methylation, and H3K27 tri-methylation) in *A. thaliana* have been determined by microarray combined with chromatin immunoprecipitation (ChIP-chip), and those distribution

patterns are consistent with their functions (Zhang, 2008). However, the PTMs of histone in other plant species are still elusive, including several important crops, like soybean, rice and wheat.

Previous researches in plant reveal that histone modifications can regulate the plant's response to internal and external signals, such as cell differentiation, development, light, temperature, and abiotic and biotic stresses (Chen *et al.*, 2007). A well known process in plant controlled by epigenetics is vernalization, which requires the coordination between H3 acetylation and methylations of H3K4, H3K9 and H3K27. These modifications regulate the expression of *FRIGIDA (FRI), FLOWERING LOCUS C (FLC)* and other vernalization related gene to ensure flowering at proper time (Schmitz *et al.*, 2008; He *et al.*, 2003; Xu *et al.*, 2008; Pien *et al.*, 2008; He *et al.*, 2005). Histone H3 Serine 10 phosphorylation (H3S10p), H3 and H4 acetylation were up-regulated in response to high salinity and cold stress in tobacco and *Arabidopsis* cells (Sokol *et al.*, 2007). Phosphorylation at H3 threonine 11 (H3T11p) or threonine 3 (H3T3p) may serve as a signal for other proteins involved in chromosome condensation, while phosphorylation at H3 serine 10 (H3S10p) or serine 28 (H3S28p) is involved in sister chromatid cohesion (Houben *et al.*, 2007). In maize, histone acetylation and chromatin remodeling are important processes for their acclimation to UV-B (280-315nm) (Casati *et al.*, 2008). Histone acetylation can also affect cellular pattern in *Arabidopsis* root epidermis by regulating the expression of cellular patterning genes (Xu *et al.*, 2005).

Although many studies of PHD finger domain containing proteins have been done in animals, their roles in plant are still elusive. The functions of these PHD finger containing proteins in plants are diverse, ranging from regulation of male meiosis to specification of vasculature and primary root meristem, and embryogenesis and sister-chromatid cohesion, control of vernalization, disease resistance, apical meristem maintenance (Wei, *et al.* 2009). However, little is known about the detail mechanisms of PHD fingers in these important developmental or environmental regulators in plants (Lee *et al.*, 2009).

1.5 Objectives and significances of the present studies

Since soybean is an important economic crop whose growth and production could be affected by salinity stress, study of salt stress response in this crop plant may help to increase their tolerance to salinity stress and their yield on saline land. Furthermore, these findings may be applicable to crops that are more salt sensitive, such as carrot and oranges. Albeit of this importance, the information of the soybean genes which can confer salinity tolerance is still limited, suggesting the significances of investigating the soybean response to salinity stress.

Although it is known that plants will alter their protein expression patterns to adapt to the environmental stress they encounter, the detailed molecular mechanisms of how they sense the stress and then regulate their transcriptome are still not very clear. Over the past decades, it was suggested that epigenetics, including DNA methylation, histone variants and histone posttranslational modifications (PTMs), regulated many physiological processes. Recently, the "histone code" concept was raised which indicated that many protein "effectors" will function to decode the "histone code" and regulate gene transcription. This concept has been proven in animals; however, in plants it is still unknown. In our investigation, we are also trying to understand whether/how the histone modifications (in particular, methylation and methylation multiplicities) regulate soybean responds to salinity stress. Investigation of the histone PTMs under salinity stress can give us some hints to understand the epigenetic roles in gene transcription regulation and adaptation to the abiotic stresses in soybean and other plants. Our studies should broaden our understanding of the plant response to salinity stress from an epigenetic perspective.

In particular, our objectives are:

1. to investigate the soybean response to salinity stress through proteomic studies
2. to identify the histone posttranslational modifications in soybean
3. to elucidate the roles of histone modifications in soybean salinity stress response

Chapter 2 Proteomic studies in soybean

2.1 Introduction

Many proteomic studies have been performed in plants, but most of them were carried out in the *Arabidopsis* and rice. The proteomic studies in soybean are just at their infancy stage. The first proteomic study of soybean appeared as early as 2005, where the protein profiles of soybean seed filling was investigated (Hajduch *et al.*, 2005). Later on, proteomic studies of different soybean organelles and tissues were carried out, including seed, leaf, hypocotyls, root, nodule, xylem sap, apoplast and peroxisomes (Natarajan *et al.*, 2006; Djordjevic *et al.*, 2007; Arai *et al.*, 2008; Oehrle *et al.*, 2008; Brechenmacher *et al.*, 2008; Afroz *et al.*, 2009). Ahsan *et al.* compared the soybean leaves and flowers at different developmental stages, revealing several organ specific functional differentiation proteins (Ahsan *et al.*, 2009). In addition, the soybean responses to some abotic stresses were also studied by proteomics recently. Cadmium induced the expression of several proteins in soybean cells, such as superoxide dismutase, histone H2B, chalcone synthase and glutathione transferase (Sobkowiak *et al.*, 2006). Danchenko *et al.* revealed the changes of protein profiles of soybean seed from the Chernobyl area under the radioactive environment, most of which were involved in protein destination and storage followed by disease and defense, suggesting that the soybean have adapted to the radioactive regions (Danchenko *et al.*, 2009). Proteome analysis of root and early stage soybean seedling and plasma membrane proteins under the water stress have been performed (Komatsu *et al.*, 2009; Hashiguchi *et al.*, 2009; Yamaguchi *et al.*, 2009). Investigation the impact of solar ultraviolet-B (UV-B) radiation on the soybean leaf proteome showed that higher levels of flavonoids would lead to a reduction in UV-B sensitivity at the proteomic level (Xu *et al.*, 2008).

In order to increase the soybean tolerance to salinity stress, it is imperative to identify genes whose products confer improved salt tolerance. However, limited information is available about salt-response genes in soybean at present, and the study of protein expression in response to salinity may therefore help identify the genes responsible

for salt adaptation mechanisms in this important crop. Recently, proteome analysis was performed in the hypocotyls and root of soybean to identify the salt stress related proteins of soybean. Instead of using MALDI-TOF/TOF to identify the differentially expressed proteins, these proteins were identified by Edman sequencing. Their results indicated that LEA, β-conglycinin, elicitor peptide and basic/helix -loop-helix protein were up-regulated and protease inhibitor, lectin and stem 31-kDa glycoprotein were down-regulated in the salinity stress (Aghaei *et al.*, 2009). In order to further understand the salinity stress response in other tissues such as leaves, I compared the soybean leaf protein profiles under salinity stress.

Moreover, treated the plant with salt stress intensively in laboratory may only activate the plant acute response to the stress and some novel processes or mechanisms unique to naturally stress-tolerant plants could be difficult to be revealed by this way (Riccardi *et al.*, 1998; Zhu, 2001). Therefore, several groups have compared the proteomics of different varieties of barley and *Populus xeuramericana*, including varieties that were sensitive and tolerant to the abiotic stresses (Bonhomme *et al.*, 2009; Witzel *et al.*, 2009). In order to understand the adaptation and tolerance mechanisms to salinity in the naturally salinity stress tolerant soybeans, I compared the proteomic profiles of two soybean varieties, wild soybean and cultivated soybean (Union) by two-dimensional gel electrophoresis and indentified the differentially expressed proteins with MALDI-TOF/TOF. Although the cultivated soybean was sensitive to the salinity stress, the wild type soybean was originally grown well in saline land and tolerant to the stress. I also tried to use proteomic tools to understand the soybean development, by identifying some development related proteins. All of these results provided some potential breeding targets for improvement of salt tolerance in the cultivated soybean and some proteins for regulating root development.

2.2 Material and methods

2.2.1 Plant materials and stress treatment

Cultivated soybean (*Glycine max,* Union) and wild type soybean were germinated

and grown in Metromix-200 soil (Hummert International Horticultural Supplier, Early City, MO, USA) in the green house of the Department of Biology at the Chinese University of Hong Kong. The temperature was kept between 25-30 °C, 60-90% humidity, natural light, 500-2000 $\mu EM^{-2}h^{-2}$. During germination stage and seedling stage, soybeans were collected and divided into two portions: roots and above ground tissues, including leaves and shoots. In order to investigate the leaves response to salinity stress, cultivated soybean (*Glycine max,* Union) were first germinated in Metromix-200 soil and then transferred to 1× Hoagland's solution to grow into 3-4 trifoliates under the same condition as mentioned before. The plantlets were then treated with 0.9% sodium chloride for 48 hours before the leaves were collected. Samples without salt treatment were collected as control. All the samples were stored at -80℃ until used.

2.2.2 Protein extraction and two dimensional gel electrophoresis

For protein extraction, the collected soybean tissues were first ground into fine powder in liquid nitrogen and total proteins were extracted with TCA/acetone extraction method. After overnight precipitation in 10% TCA/acetone at -20℃, proteins were dissolved in rehydration buffer, which contained 6 M urea, 2 M thiourea and 4% 3-[(3-cholamidopropyl) dimethylammonio]-1- propanesulfonate (CHAPS). Finally, 0.2% (w/v) dithiothreitol (DTT) and 0.5% IPG buffer were added into the samples before IEF. The protein concentration was determined by 2-D Quant kit (GE healthcare) and total 250 μl samples containing about 200 μg proteins were applied to the dry IPG strips (13cm, pH 3-10 nonlinear, GE healthcare). The program of IEF was as followed: rehydration at 20℃ for 7 hrs, 30 V for 7 hrs, 150 V for 2 hrs, 500 V for 0.5 hrs, 1000 V for 0.5 hrs, 4500 V for 3000 vhs, 8000 V for 66000 vhs. Focused strips were first equilibrated by incubating in equilibration buffer (6 M urea, 30% v/v glycerol, 2% w/v SDS, 50 mM Tris-HCl, pH 8.8, and 1% w/v DTT) for 15min, followed by incubation in the same equilibration buffer containing 4% w/v iodoacetamide (IAA) instead of DTT for another 15 min. For the second dimension, the focused IPG strips were laid horizontally on the 12% acrylamide SDS-PAGE

separating gels, and sealed with 1% agarose, and electrophoresis were run at 85 V.

2.2.3 Gel staining and image analysis

After SDS-PAGE, the PAGE gels were stained with silver and the image were captured by magic scanner. The gel images were then analyzed using ImageMaster™ 2D Platinum 5.0 software (GE Healthcare). The protein expression profiles were compared and the protein spot volumes were normalized automatically against the total spot volume of the gel using the software. The relative volume (%Vol) for each protein spot was obtained and analyzed. At least two independent experiments of each sample were performed to ensure technical reproducibility and the differentially expressed protein spots which were reproducibly detected to be changed over 1.5 fold in these repeated gels or whose $p \leq 0.05$ in the two-sample Student's *t* test analysis were considered to be significantly different and selected out for further analysis.

2.2.4 Tryptic in-gel digestion

Protein spots of interest were manually excised out from the 2-D gels. They were destained with an equal volume mixture of 30 mM potassium ferricyanide and 100 mM sodium thiosulfate until the color disappeared. After washing in Milli-Q water twice, the gel plugs were equilibrated in 50 mM ammonium bicarbonate for 10 min twice. The destained gels were finally soaked in 100% acetonitrile to be dehydrated. Thereafter, vacuum-dried gel plugs were rehydrated with 10 mg/ml of trypsin in 25 mM ammonium bicarbonate (pH 8.0) for 30 min on ice. Proteins were proteolyzed for 16-18 h at 30°C and digested peptides were extracted from the gels with 80% acetonitrile/2.5% trifluoroacetic acid.

2.2.5 Protein identification by tandem mass spectrometry

Mass spectrometric analysis was carried out using a MALDI-TOF/TOF mass spectrometer 4700 Proteomics Analyzer (Applied Biosystems, USA). For acquisition of mass spectra, 0.5μl of overnight digested peptides were spotted onto the

MALDI-plate, followed by adding 0.5 μl matrix solution (5 g/L α-cyano-4-hydroxycinnamic acid in 50% acetonitrile/0.1% trifluoroacetic acid) and allowed to dry in air. Mass data acquisitions were piloted by 4000 Series Explorer™ software using batched-processing and automatic switching between MS and MS/MS modes. All MS survey scan were acquired over the mass range *m/z* 700-3500 Da in the reflectron positive-ion mode and accumulated from 1000 laser shots with acceleration of 20 kV. The MS spectra were internally calibrated using porcine trypsin autolysis products (*m/z* 805.417, *m/z* 906.505, *m/z* 1153.574, *m/z* 2163.057 and *m/z* 2273.160) resulted in mass errors of less than 20 ppm. Peptide precursor ions corresponding to contaminants were excluded and the filtered precursor ions with a user-defined threshold (S/N ratio ≥50) could be selected for a MS/MS scan. Fragmentation of precursor ions were performed using argon as the collision gas and MS/MS spectra were accumulated from 3000 laser shots with a default calibration.

2.2.6 Database analysis

MS and MS/MS spectra from the 4700 Proteomics Analyzer were loaded into GPS Explorer™ software and searched against the soybean database which was downloaded from http://www.phytozome.net/soybean using the MASCOT search engine for combined MS and MS/MS analysis. The following search parameters were used: 700-3500 Da; monoisotopic peptide mass (MH^+); pI 0-14; taxonomy, *Glycine max.*; enzyme, trypsin; precursor-ion mass tolerance, 50ppm; MS/MS fragment-ion mass tolerance, 0.1 Da; variable modifications, oxidation for methionine and carboxyamidomethylation of cysteine

2.3 Results

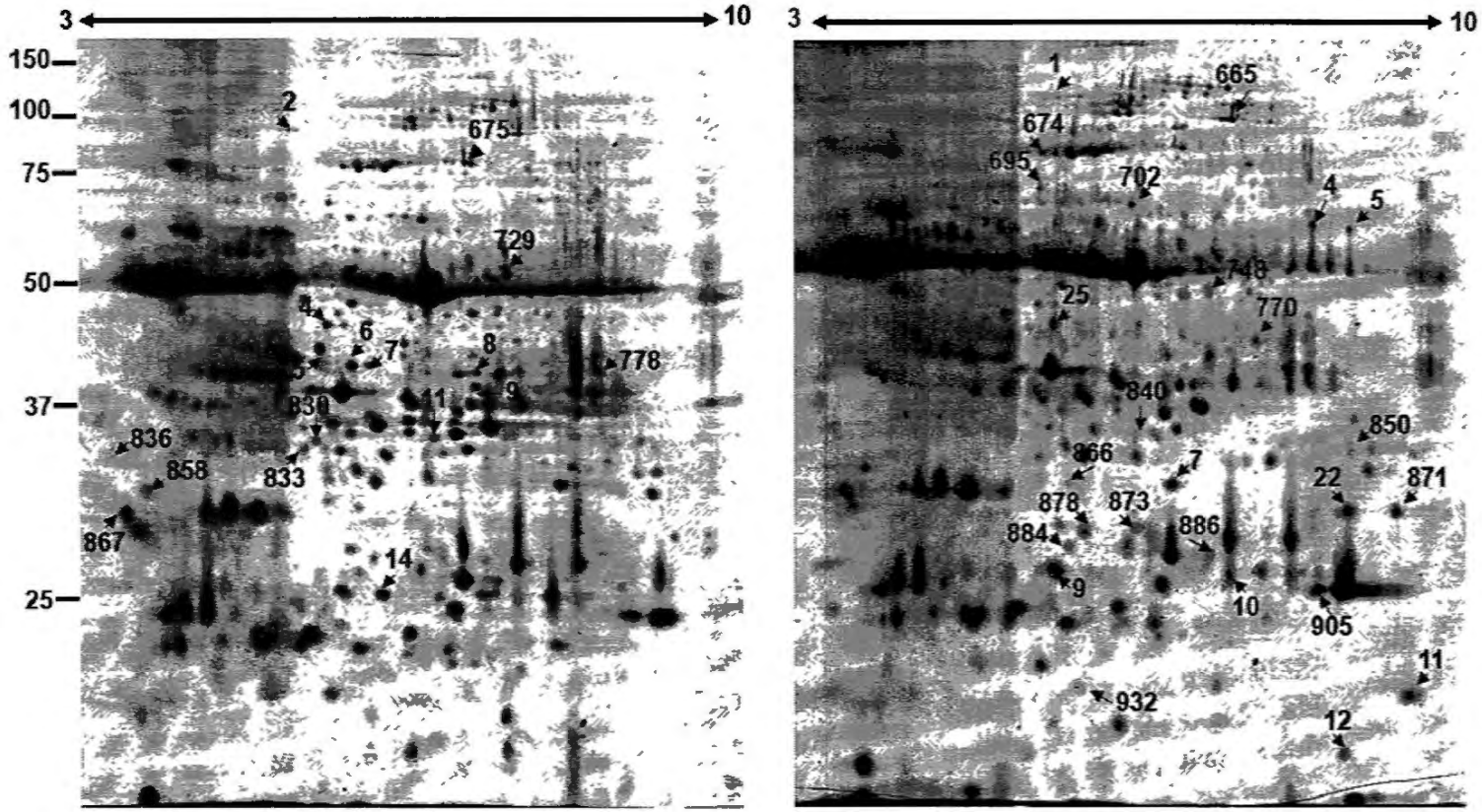
2.3.1 Comparative proteome study of soybean in different variants and developmental stages

In order to identify the differentially expressed proteins which may associate with the salinity tolerance ability of the wild type soybean, proteomic profiles of the wild type and cultivated soybean were compared. These soybean samples were collected at

germination stages and seedling stages and total proteins were extracted from the roots and above ground tissues at each stage with TCA/acetone method, separated in 2-D gels and stained with silver. Gels were analyzed and total over 300 spots were reproducibly detected in each gel (Figure 2.1-2.4). My results showed that total about 117 spots were significantly changed between the wild type and cultivated soybean. While 72 spots were up-regualted or *de novo* synthesis in the wild type soybeans, the rest 45 spots were down-regulated (Table 2.1).

Comparative analysis the protein profiles of the roots of the cultivated soybean at germination stage and seedling stage were conducted, in an effort to identify some proteins and physiological processes that may be specifically function in the root at certain developmental stage (Figure 2.5). These comparisons displayed that 53 spots were changed over 1.5 fold at different developmental time. Of these spots, 31 spots were up-regulated or *de novo* synthesis in the germination stage while 22 spots were down-regulated (Table 2.2).

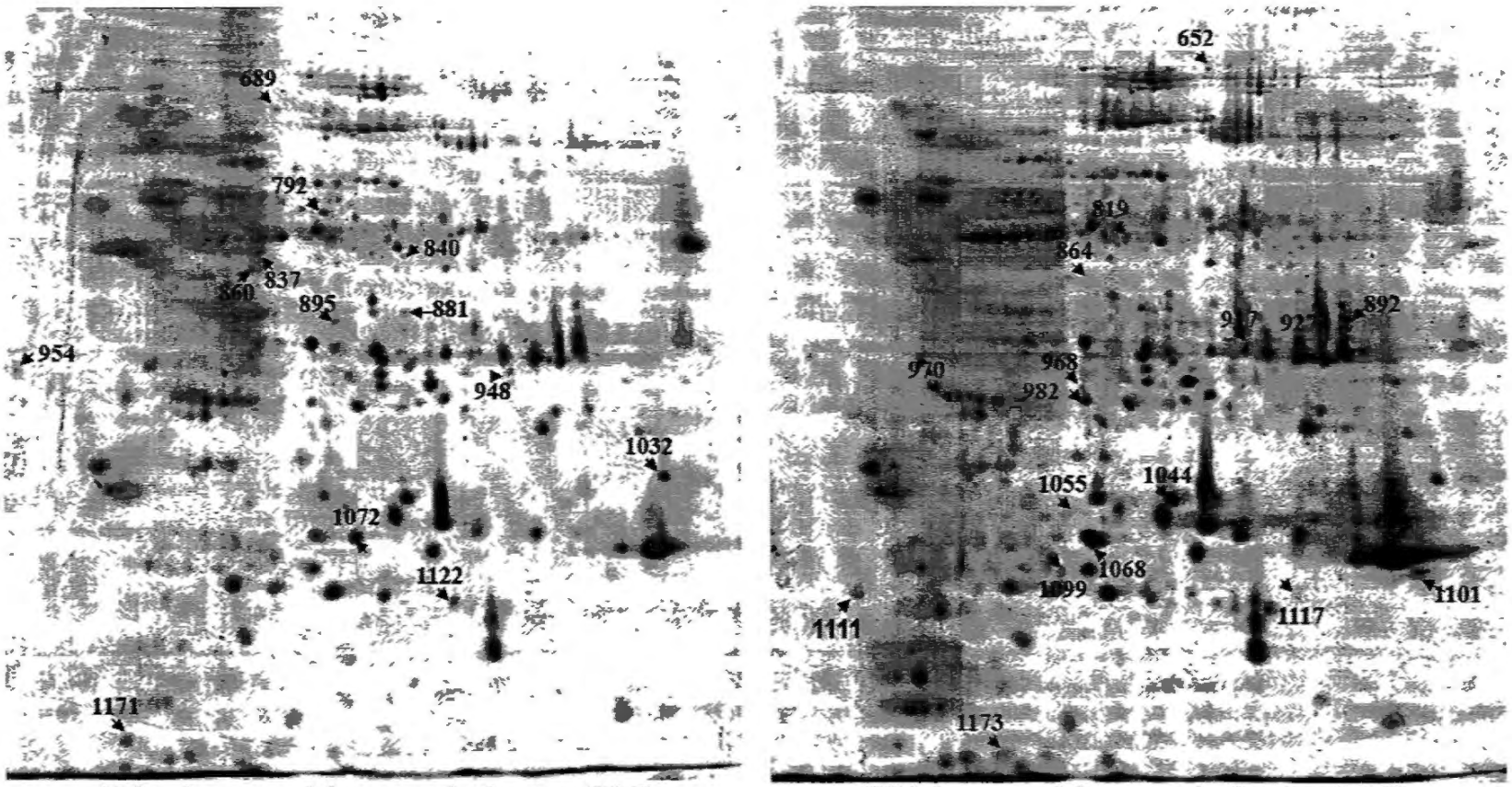
Similar experiment was carried out to identify the proteins associated with soybean response to salinity stress in the leaves (Figure 2.6). Proteins from leaf samples with and without salinity treatment were extracted and the results show that 18 spots were greatly down-regulated under the salinity stress while 4 spots were up-regulated (Table 2.3).



Union above ground tissues seedling stage (ULS)

Wild above ground tissues germination stage (MLS)

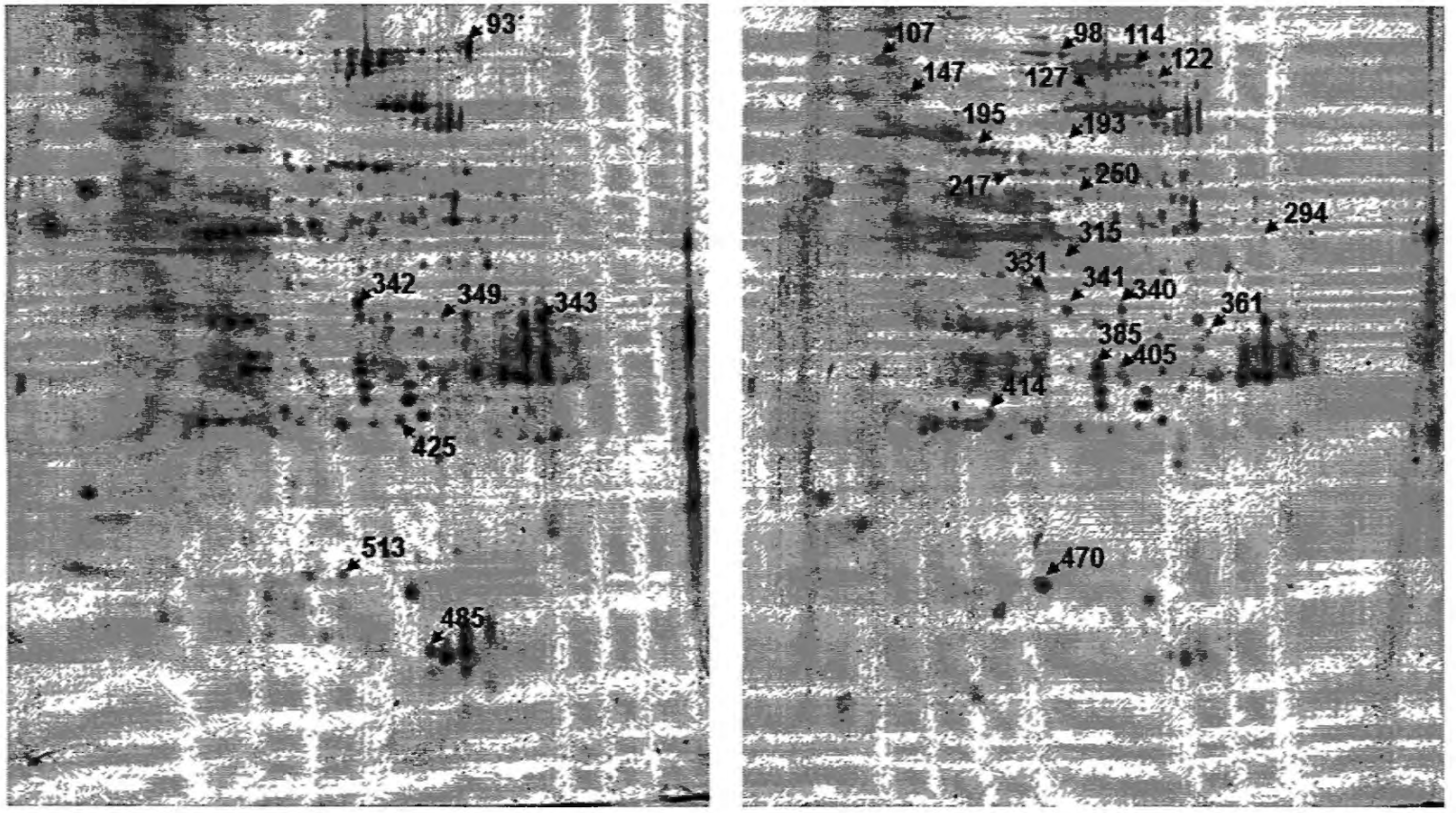
Figure 2.1: 2-DE profiles of proteins extracted from the above ground tissues of union and wild type soybean at the seedling stage. Up-regulated proteins in the union soybean were indicated by arrows in the ULS gel, while up-regulated proteins in the wild type soybean were indicated in the MLS gel.



Union above ground tissues germination stage (ULG)

Wild above ground tissues germination stage (MLG)

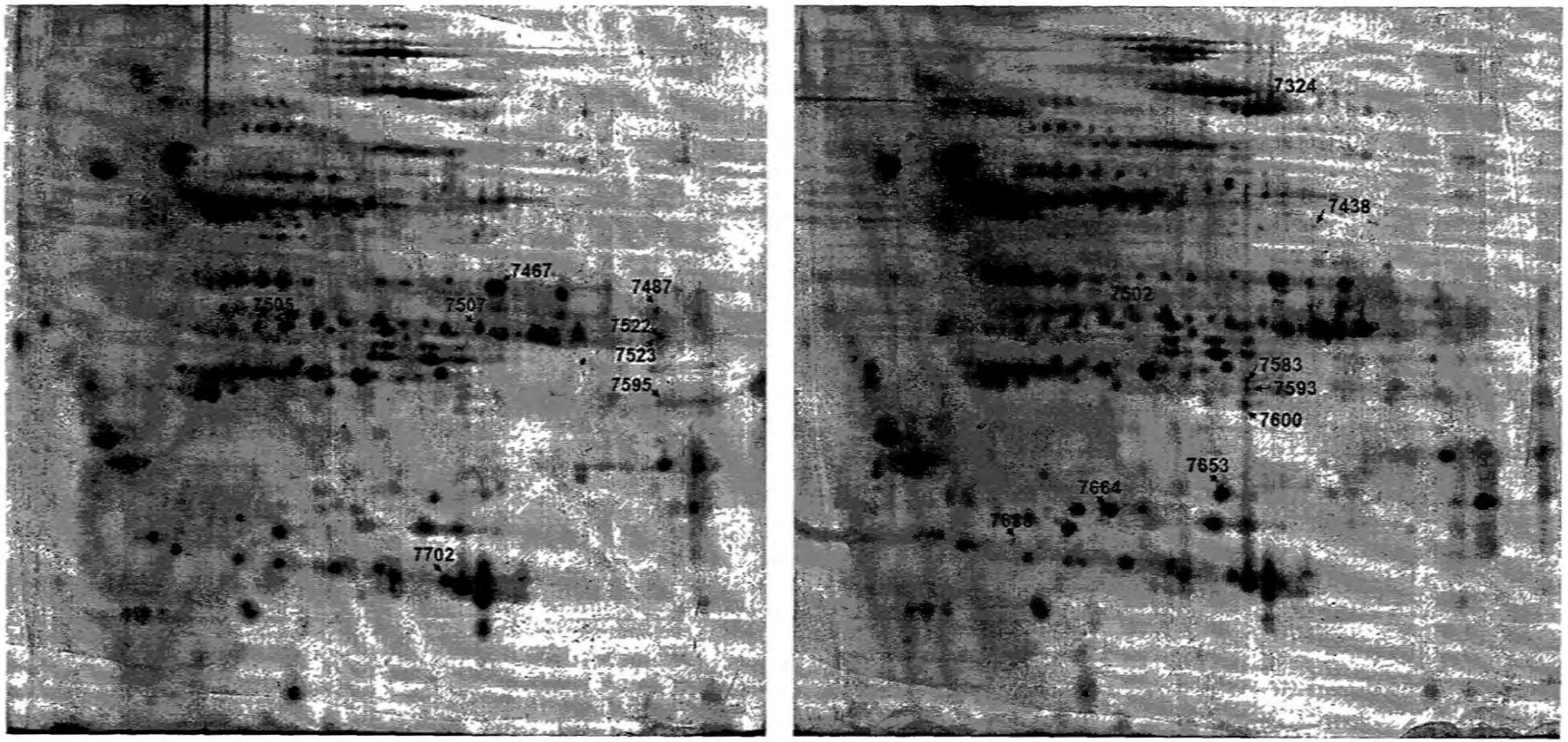
Figure 2.2: 2-DE profiles of proteins extracted from the above ground tissues of union and wild type soybean at the germination stage. Up-regulated proteins in the union soybean were indicated by arrows in the ULG gel, while up-regulated proteins in the wild type soybean were indicated in the MLG gel.



Union root seedling stage (URS)

Wild root seedling stage (MRS)

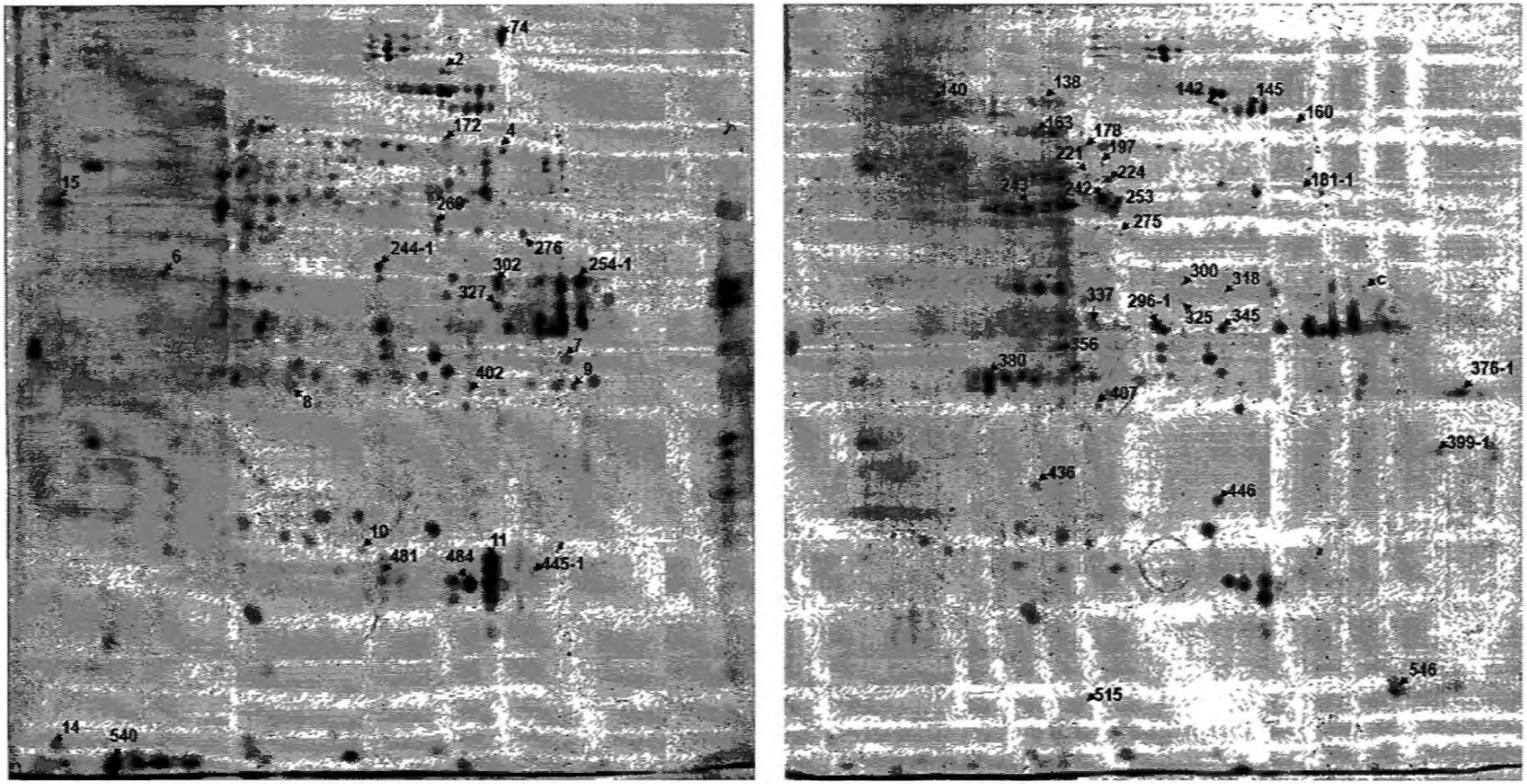
Figure 2.3: 2-DE profiles of proteins extracted from the roots of union and wild type soybean at the seedling stage. Up-regulated proteins in the union soybean were indicated by arrows in the URS gel, while up-regulated proteins in the wild type soybean were indicated in the MRS gel.



Union root germination stage (URG)

Wild root germination stage (MRG)

Figure 2.4: 2-DE profiles of proteins extracted from the roots of union and wild type soybean at the germination stage. Up-regulated proteins in the union soybean were indicated by arrows in the URG gel, while up-regulated proteins in the wild type soybean were indicated in the MRG gel.



Seedling stage

Germination stage

Figure 2.5: 2-DE profiles of proteins extracted from the roots of union at the germination stage and seedling stage. Up-regulated proteins in the seedling stage were indicated by arrows in the seedling stage gel, while up-regulated proteins in the wild type soybean were indicated in the germination stage gel.

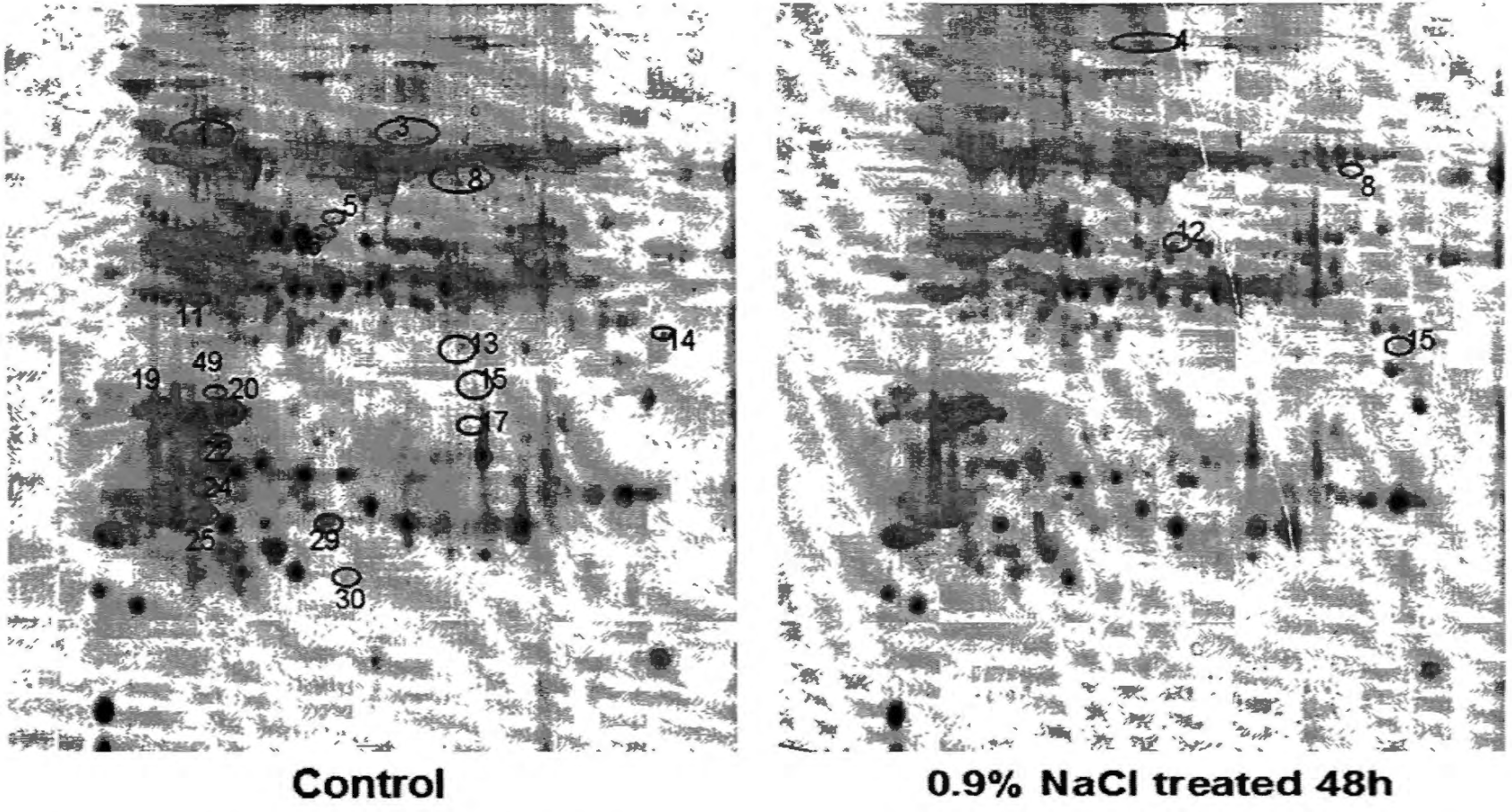


Figure 2.6: Protein profiles of soybean leaves under control and salt treated condition. Control: total proteins from soybean leaves without salt treatment. 0.9% NaCl treated 48h: total proteins from soybean leaves treated with 0.9% NaCl for 48 hours. Proteins were down-regulated after salt treatment were circled in the control gel image, while up-regulated proteins were circled in the treated gel image.

2.3.2 Identification of differentially expressed proteins in different varieties and developmental stages

Differentially expressed spots were excised from the gels and applied for trypsin digestion. The digested peptides were extracted from the excised gels and analyzed by MALDI-TOF/TOF. The peptide mass lists were searched against the data base which was downloaded from soybean genome data base (http://www.phytozome.net/soybean). Since most of the proteins were still not experimentally annotated in soybean, their functions were predicted according to their homologs in other species using blast to search against the NCBI data base. Following this way, I confidently identified 112 spots out of the 117 changed spots in the wild type soybean and cultivated soybean (Table 2.1). While at different developmental stages, of the 53 spots which were changed in the roots of the cultivated soybean as they grew, 51 spots were confidently identified by MALDI-TOF/TOF (Table 2.2). Of the 22 spots that were changed under the salinity stress, 18 spots were also confidently identified (Table 2.3).

Interestingly, comparison of the proteomics of the wild type and cultivated soybean showed that several differentially expression spots were finally turned out to be the same protein, such as aminomethyltransferase, cytosolic phosphoglycerate kinase, and isocitrate dehydrogenase (NADP). Meanwhile, several spots which all represented for stem 28 kDa glycoprotein or stem 31 kDa glycoprotein were detected in the proteomic profiles of the above ground tissues of the wild type soybean at the seedling stage. This phenomenon may be result from the difference in the posttranslational modifications of these proteins, since their pI and molecular weight varied as displayed in the gels (Figure 2.1-2.4). Although whether the modifications of these proteins have any effects on their activities need further studies, this result supported that post-transcriptional regulation played important roles in the plants' response to stress, and highlighted the necessary to combine transcriptomic and proteomic analyses (Jiang *et al.*, 2007). Similar results can also be obtained by comparing the proteomic profiles of soybean roots at different developmental stages, like peroxidase precursor.

The expression of different isoforms of a gene may be dynamic according to their growth condition (Espartero *et al.*, 1994). I observed that the isoforms of chalcone isomerase, glyceraldehyde-3-phosphate dehydrogenase, isocitrate dehydrogenase, methionine synthase, peroxisomal voltage-dependent anion-selective channel protein, S-adenosylmethionine synthetase and transketolase were differentially expressed in the two soybean genotypes, indicating that these two soybean varieties have distinct regulation mechanisms to control these isoforms' expression. However, it is still unknown whether there are any differences in the activities of these isoforms.

The differentially expressed proteins between wild type and cultivated soybean could be classified into 7 categories according to their probable biological functions. About 23% of these proteins contributed directly to the stress/defense pathway, which is one of the major parts of the differentially expressed proteins. Moreover, most of the proteins belonging to this category were found to be up-regulated in the wild type soybean except several peroxidases. Other functional categories of energy metabolism, protein and amino acid metabolism, secondary metabolism and protein storage were represented at 23%, 20%, 5% and 5%, respectively (Figure 2.7A). The above observation might suggest that the wild type soybeans were actively engaged in the plant defense responses which usually be observed when the plants were treated with salinity stress.

The developmental dependent changed proteins could also be sorted into 6 functional categories (Figure 2.7B). Surprisingly, the proteins involved in the stress/defense pathway also took the dominant part, about 34% of the differentially expressed proteins. The expressions of other proteins involved in energy metabolism, protein and amino acid metabolism, and secondary metabolism were also found to be changed at different developmental stages, which took 27%, 17% and 4% of these changed proteins respectively. Storage proteins and adenosine kinase involved in the purine metabolism were up-regulated in the germination stage.

When the soybeans were treated with salinity stress, several differentially expression spots with different pI and molecular weight were also turned out to be the same proteins, such as ribulose 1,5-bisphosphate carboxylase/oxygenase and oxygen

evolving enhancer protein 1 (OEE1). Those salt inducible proteins can be separated into 5 groups (Figure 2.7C), with the proteins involved in photosynthesis taking the dominant part (about 61%). The rest of the proteins were involved in other processes such as protein and amino acid synthesis, stress and defense and primary metabolism.

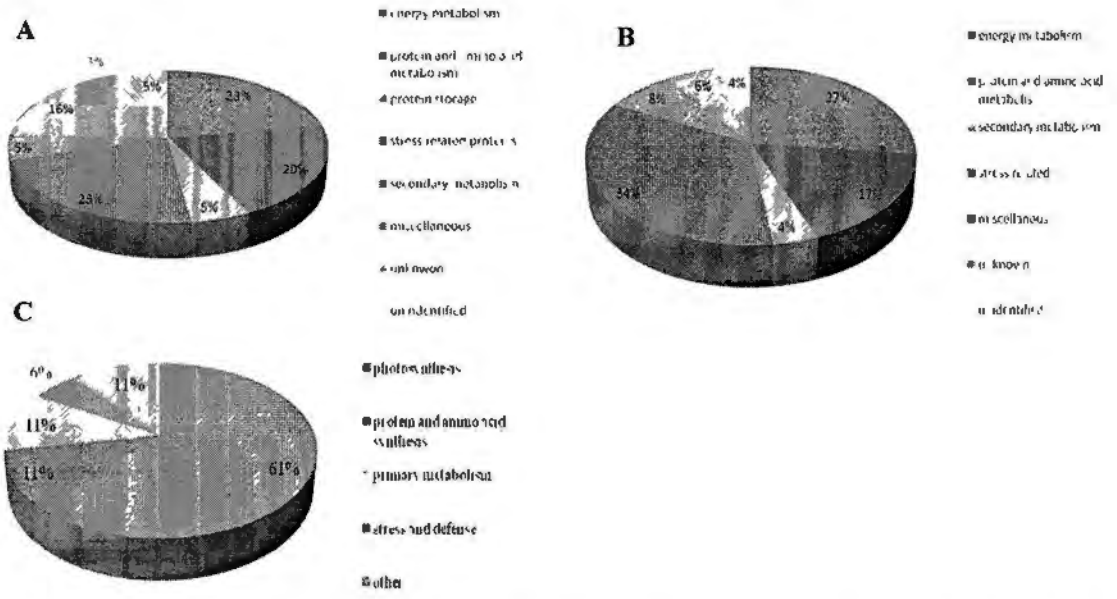


Figure 2.7: Functional classification of differentially expressed proteins based on known as well as putative functions. A: in the wild type and union soybean, total 117 proteins, representing 94 nonredundant protein species, were categorized into 7 functional classes. B: in different developmental stages, total 53 proteins, representing 52 nonredundant protein species, were categorized into 6 classes. C: identified differential expression proteins in the salt treatment were classified into 5 catalogs.

2.4 Discussions

2.4.1 Comparative proteomic studies of cultivated and wild type soybean

Upon comparative proteomic studies, several distinguish physiological processes between wild type soybean and cultivated soybean were identified: energy metabolism, detoxification and defense, and protein metabolism. Although in my studies, these proteins were identified to be changed in the soybean varieties without

salinity stress, many of them were confidently to be involved in salinity stress as showed in many other studies (Zhu, 2001; Shulaev *et al.*, 2008). Therefore, I believed that this set of proteins could contribute to the differences in the abilities of the two soybean varieties to tolerant the salinity stress. By regulating these processes, the wild type soybean may yield a physiological advantage under the salinity stress.

2.4.1.1 Energy metabolism

The energy metabolism was actively regulated in the wild type soybean. Glycolysis and tricarboxylic acid cycle (TCA cycle) were the most important pathways that produce ATP to support most of the physiological processes. My proteomic studies showed that many proteins involved in glycolysis and TCA cycles, such as phosphoglycerate mutase, aconitate hydratase and malic enzyme, were up-regulated in wild type soybean. Pyruvate metabolism was essential in the TCA cycles, in which the pyruvate dehydrogenase complex catalyzed the overall conversion of pyruvate to acetyl-CoA to drive the TCA cycles. Two enzymes of the pyruvate dehydrogenase complex, dihydrolipoyllysine residue acetyltransferase and pyruvate decarboxylase isozyme 1, were unpregulated in the wild type soybean in my studies. Ferredoxin--NADP reductase, which played a key role in regulating the cyclic and non-cyclic electron flow to meet the demands of the plant for ATP and reducing power, was found to be highly expressed in the wild type soybean. Other pathways including the oxidative pentose phosphate pathway (PPP) and ketogenesis were also involved in energy production. The PPP pathway was central in carbohydrate metabolism in plants and provides both reducing power and precursors for biosynthesis of other cell components (Liska *et al.*, 2004). 6-phosphogluconate dehydrogenase, a key enzyme of the PPP pathway catalyzing the conversion from 6-phospho-D-gluconate to D-ribulose 5-phosphate, was significantly increased in the wild type soybeans. Previous studies in wheat, rice and Dunaliella also showed that the expression of 6-phosphogluconate dehydrogenase was induced by salinity stress (Liska *et al.*, 2004; Huang *et al.*, 2003). Acetoacetyl-CoA thiolase involved in the ketogenesis was also up-regulated in the wild type soybean. Acetoacetyl-CoA

thiolase catalyzed the first step of ketogenesis through which the acetyl-CoA was converted to acetoacetate, which sometimes served as important metabolic fuels.

Some previous studies have already indicated that proteins involved in glycolysis, TCA cycle, and/or the oxidative pentose phosphate pathway were induced by prolong and progressive stress treatment (water stress and salinity stress) in cells of rice, maize and Dunaliella (Riccardi *et al.*, 1998; Liska *et al.*, 2004; Huang *et al.*, 2003). Under those abiotic stresses, in addition to keep their normal development and growth, the plants also needed large amount of energy to eliminate the adverse effect of such stress conditions. For example, they required heat shock proteins to keep the normal structure of the protein under the stress condition which usually needed large amount of energy to function (Tapley *et al.*, 2009). They also need lots of the intermediates of these metabolism pathways to synthesis other osmolytes like glycerol (Liska *et al.*, 2004). Therefore, it was likely that the plants had to activate these pathways to fulfill these requirements because some other physiological processes such as photosynthesis might be inhibited owing to stomatal closure and limited carbon dioxide uptake under the water stress (Hashiguchi *et al.*, 2009; Bailey-Serres *et al.*, 2008; Zhu, 2001). So it was not surprising to find out the simultaneously induction of these pathways to generate energy and maintain homeostasis for survival in the wild type soybean.

2.4.1.2 Detoxification and defense pathway

Stress including salinity stress induced the accumulation of reactive oxygen species (ROS) such as superoxide, hydrogen peroxide, and hydroxyl radicals in many plants (Hasegawa *et al.*, 2000). A proper amount of ROS was acquired as substrates and signals not only in normal cell metabolism, growth, and differentiation, but also in inducing ROS scavengers and other protective mechanisms against the stress. Therefore, copper amino oxidase which contributed to the synthesis of ROS was up-regulated in wild type soybean. Alternatively, quinone oxidoreductase and benzoquinone reductase in the wild type soybean could also produce hydrogen peroxide and superoxide anion by oxidizing quinones (Vianello *et al.*, 1991).

Strangely, several peroxidases were found to be up-regulated in the root of cultivated soybean in my studies. Peroxidase were not only important in removing certain types of ROS, but also responsible for producing some kinds of them and inducing root growth as the followed discussion, suggesting that they might induce root growth in cultivated soybeans, which was evidenced by the observation of the faster growth of the root in the cultivated soybeans than the wild type soybean.

Despite of their positive roles in plant growth, the homeostasis of ROS must be kept in the cells under the stress condition, since they could either induce stress injury when their concentration was elevated by the abiotic stress. It was well documented that plants would activate the detoxification and defense pathway to eliminate the over-produced ROS under stress (Zhu, 2001; Shulaev *et al.*, 2008). One of the important mechanisms was to upregulate ascorbate peroxidase (APX) which could reduce hydrogen peroxide to H_2O with the concomitant generation of monodehydroascorbate dependent on ascorbate. Previous studies in barley, wheat and *Arabidopsis* have confirmed that APX was a key component in scavenging ROS under various stresses (Witzel *et al.*, 2009; Shigeoka *et al.*, 2002; Jiang *et al.*, 2007; Caruso *et al.*, 2008; Wang *et al.*, 2008). The ROS scavenging system would collapse in the *Arabidopsis* absent of the cytosolic APX (Davletova *et al.*, 2005). Transgenic plants overexpressing ROS scavengers or mutants with higher ROS scavenging ability showed increased tolerance to environmental stresses (Xiong *et al.*, 2002). Hence, it was not surprising to found that the APX was up-regulated to counter its active ROS-producing system in the wild type soybean. However, I found that other spots (ULG1072, ULS14, URS513) which also represent for the APX but with a bigger pI were greatly induced in the cultivated soybean. Previous reports showed that some soybean varieties may have two isoforms of APXs and their catalytic activities were different (Caldwell *et al.*, 1998). Moreover, those soybean varieties with only one APX were more tolerable to chilling treatment (Funatsuki *et al.*, 2003), suggesting that the lacking of one APX in the wild type soybean might partially correlate with its salinity tolerance. However, the underlying mechanisms needed further studies. As another product of the reaction catalyzed by APX, monodehydroascorbate can be

reversed to ascorbate by the enzyme monodehydroascorbate reductase, which was also increased in the wild type soybean, probably to maintain the concentration of the reduced agents.

I also identified GST as well as actoylglutathione lyase (glyoxalase I) were up-regulated in the wild type soybean. GST was another protein well-known for its ability to conjugate reduced glutathione with various compounds and involvement in oxidative stress responses (Droog, 1997). Glutathione was also necessary in the lactoylglutathione lyase mediated detoxification process of the methylglyoxal (MG), the accumulation of which was indicative of abiotic stress conditions, such as salinity, drought, and cold. Recent studies showed that overexpression of GST or glyoxalase I in tobacco both could enhance their tolerance to salinity during germination (Singla-Pareek *et al.*, 2003; Witzel *et al.*, 2009).

As a well known stress marker gene, alcohol dehydrogenase 1 which catalyzed the oxidation of alcohols to aldehydes or ketones was up-regulated in wild type soybean. Large number of previous work have showed that its expression was stress-related and it was likely that they were required to remove the alcohols, which were commonly produced through lots of metabolic pathways with deleterious effects on many cellular components such as nucleic acids, proteins and carbohydrates (Sengupta *et al.*, 2009). I also found that the enzyme aldehyde dehydrogenase and an aldo-keto reductase which involved in oxidative stress tolerance by detoxifying reactive aldehydes derived from lipid peroxidation during abiotic stress were induced in wild type soybean. Previously, Oberschall *et al.* also indicated that an aldose/aldehyde reductase was able to provide protection to transgenic plants against lipid peroxidation under chemical and drought stresses (Oberschall *et al.*, 2000).

Wild type soybean accumulated much more lipoxygenase than the cultivated soybean. Lipoxygenase initiated the synthesis of a series of oxylipins, which were acyclic or cyclic compounds with diverse functions, particularly in responses to herbivory and pathogen invasion (Feussner *et al.*, 2002; Wang *et al.*, 2009). In addition, they also associated with many other stress stimuli, including cold and radiation stress (Danchenko *et al.*, 2009).

Abiotic stress including salinity stress would cause proteins to denature. Similar to other plants response to the salinity stress, heat shock proteins 90 (Hsp90) and cyclophilin were found to be up-regulated in wild type soybean. Heat shock proteins were among the most well-known stress-related proteins in plants which could be induced by several types of stresses such as heat, osmotic and salt stress. These proteins could act as chaperones to accelerate protein folding and protect them from denaturing under stress conditions (Aghaei *et al.*, 2008). In addition, they could also decrease the intracellular level of reactive oxygen species (ROS). Cyclophilin, another ubiquitously distributed protein which also catalyzed the folding of proteins, could be induced by abiotic stresses, such as drought stress, too (Hajheidari *et al.*, 2005).

2.4.1.3 Protein metabolism pathway

Compared with cultivated soybeans, the protein and amino acid metabolism might be less active in the wild type soybeans. Enzymes such as cysteine synthase and glutamine synthase were down-regulated in wild type soybean. It was known that these enzymes were important in amino acid synthesis and protein synthesis (Aghaei *et al.*, 2008; Teixeira *et al.*, 2006). This result indicated that the wild type soybean might be inclined to decrease the level of free amino acid pool. Consistent with this, the protein synthesis was suppressed in the wild type soybean, since that several proteins involved in this process, such as ribosomal protein L12-A, translational elongation factor 1 subunit B, and nascent polypeptide associated complex alpha subunit, were down-regulated in them. This observation was also very similar to the plants' responses to salt stress, where the transcription of ribosome was severely depressed in *Arabidopsis* roots (Jiang and Deyholos, 2006).

Another interesting result was that more storage proteins, including stem 28 kD glycoprotein and 31 kD glycoprotein, were accumulated in the wild type soybean during their germination and seedling stage. These glycoproteins were among the major proteins in leaves and most other vegetative tissues of soybean. Studies have showed that they may be rapidly degraded according to the need for nutrients by

other tissues of soybean plants (Ahsan *et al.*, 2009; Staswick, 1989). The more accumulation of these storage proteins in the wild type soybean might be caused by that the system of protein degradation was less efficient in them than in cultivated soybean, since I have observed that proteins which played roles in this process, such as leucine aminopeptidase, Clp protease ATP-binding subunit clpC, and prolyl oligopeptidase family protein, were down-regulated in wild type soybean.

Different isoforms of serine hydroxymethyltransferase and methionine synthase variedly expressed in wild type soybean and cultivated soybean, which might be account for their multiple functions. On one side, they were involved in the amino acid synthesis; on the other side, they were also revealed to be associated with plant stress response. For example, serine hydroxymethyltransferase could catalyze the interconversion of serine and glycine and several recent studies have indicated that it was important in photorespiration (Xu *et al.*, 2010; Roth *et al.*, 2006; Moreno *et al.*, 2005). Another study found that the mutant plants without serine hydroxymethyltransferase accumulated more H_2O_2 than wild-type plants when subjected to salt stress, indicating that serine hydroxymethyltransferase was part of the dissipater mechanisms of plants to minimize production of ROS at the chloroplast and to minimize oxidative damage provoked by abiotic stresses (Moreno *et al.*, 2005). Similarly, methionine synthase, alongside with S-adenosylmethionine synthetase were also involved in the lignifications of the cell wall for turgor maintenance to avoid water loss from the cells under stresses. During this process, methylation of the lignin monomer was necessary, which was carried out by *S*-adenosylmethionine synthetase transferring the methyl group from *S*-adenosylmethionine, a common methyl group donor. To fulfill the demand for more methyl groups for lignin methylation under stresses, more methionine synthase was necessary to produce more methionine for *S*-adenosylmethionine generation (Bhushan *et al.*, 2007).

2.4.1.4 Other proteins

In addition to these three main types of proteins, some other proteins that have been

reported to be involved in the plant responses to salinity stress were also up-regulated in the wild type soybean. Under the salinity stress, plants activated several ion transporters to maintain the homeostasis of ion. In the wild type soybean, the V type proton ATPase was up-regulated, in agreement with that V-ATPase underwent moderate changes in expression of its subunits and modulations of its structure to regulate its activity to adapt to the environmental stress (Ratajczak, 2002). The V-ATPase created an electrochemical proton gradient which was the driving force for a variety of transport events of ions and metabolites by pumping protons into the vacuolar lumen with ATP hydrolysis.

In plant, salt stress increased sucrose and other osmolytes to reduce water loss. UDP-glucose pyrophosphorylase, a key enzyme in carbohydrates metabolism, was up-regulated in the wild type soybean. UDP-glucose pyrophosphorylase was responsible for the synthesis and pyrophosphorolysis of UDP-glucose, the key precursor of sucrose and cell wall components (*e.g.* cellulose, β-glucans). In *Arabidopsis*, the UDP-glucose pyrophosphorylase was induced by light and cold stress (Yan *et al.*, 2005). In addition, my results showed that sucrose synthase was also greatly increased in the wild type soybean. Sucrose synthase catalyzed the reversible conversion of sucrose and was also known for its role in energy metabolism especially under energy limiting conditions during stresses (Fernandes *et al.* 2004; Sengupta *et al.*, 2009). In addition, a portion of sucrose synthase existed as a membrane-bound form in association with the cellulose/callose synthase complex, contributing to cell wall biosynthesis (Sengupta *et al.*, 2009). Thus the highly expressed UDP-sugar pyrophosphorylase and sucrose synthase together with methionine synthase, and S-adenosylmethionine synthetase in wild type soybean might function coordinately in strengthening their cell wall, maintaining the cell turgor to avoid water loss and regulating their growth when they encounter the salinity stress in their natural growth condition.

Trypsin inhibitor was also reported to be induced by the stress condition in many plants. Although their roles in the salt stress response remain unclear, they constitute an important defense line to fight against pathogens probably by neutralizing the

proteases produced by the pathogens (Aghaei *et al.*, 2009). Lectin was also induced by wounding and viral infection in plants (Showalter, 1993). Moreover, they could be regulated by drought stress and might be involved in stabilizing the cytoskeleton structure under stress conditions (Bhushan *et al.*, 2007). Previous study in some cultivar soybean which could not appropriately respond to salinity showed that some protease inhibitors and lectin were down regulated after salinity stress (Aghaei *et al.*, 2009), and now my result indicated that wild type soybean would increase their expressions, implying that they were likely to help the plant to adapt to the salinity stress appropriately.

2.4.2 Comparative proteomic studies between different developmental stages in soybean

In order to identify any proteins that were important in certain developmental stage, I compared the proteomic patterns of the soybean root in germination stage and seedling stage. Surprisingly, the most dramatically changed proteins between these two developmental stages were the proteins involved in the stress and defense pathway.

Proteins, such as peroxidases, laccase 110a as well as several quinone oxidoreductases, which accumulated abundantly in the root at the seedling stage, all led to the production of the ROS. Aside from their 'notorious' roles in inducing the stress damage in the cells under the stress condition, recently, ROS were realized to have a 'benign' facet with essential roles in cell growth and cell shape fomation.

Peroxidases, owing to their great catalytic versatility to eliminate H_2O_2 or inversely to produce ROS, played a prominent role in apoplastic ROS metabolism. It was present in almost all plant tissues, but particularly abundant in roots. Plants contained numerous peroxidase genes and it was predicted that these various genes had different functions and distribution patterns along the roots (Tognolli *et al.*, 2002; Passardi *et al.*, 2006). In the elongation and meristematic zone where $O_2^{\bullet -}$ accumulated, peroxidases worked with $O_2^{\bullet -}$ to produce OH• necessary for cell wall loosening and induce root growth, while in the differentiation zone where H_2O_2

predominated, peroxidases together with H_2O_2 to induce root hair formation and inhibited cell elongation because of their abilities to oxidize the growth promoting hormone auxin and to stiffen cell walls by crosslinking between phenolic groups (tyrosine, phenylalanine, ferulic acid) in wall proteins, pectins, and other wall polymers, which further led to the formation of lignin (Dunand *et al.*, 2007; Ros Barceló, 1997; Foreman *et al.*, 2003). In the soybean genome, there were at least 351 loci encoding the putative peroxidase and my results showed that at least 6 proxidases were up-regulated in the root at seedling stage while another one was up-regulated at the germination stage. I predicted that these peroxidases had distinct distribution patterns in the soybean root at different developmental stages and function in various physiological processes to properly regulate root growth. Similarly, other proteins involved in the production of ROS, including laccase 110a as well as several quinone oxidoreductases, might also contribute to the root growth through the ROS as the peroxidase, however, further studies were required to test this hypothesis.

Consistent with the increasing of ROS production in the seedling stage, it was not surprising to find that many proteins involved in the detoxification pathway were also up-regulated simultaneously to keep the homeostasis of ROS during root growth. These proteins included GSH-dependent dehydroascorbate reductase 1 and monodehydroascorbate reductase. Therefore, the plants have evolved perfect mechanisms to fine tune the effects of ROS during their growth.

2.4.3 Comparative proteomic studies of soybean with salinity stress

Using the soybean leaves to study the soybean responses to salinity stress revealed that the expressions of the proteins involved in the photosynthesis were significantly suppressed under the stress. Ribulose 1,5-bisphosphate carboxylase/oxygenase (rubisco), rubisco activase, chlorophyll a-b binding protein 3, HCF136 (High chlorophyll fluorescence 136), oxygen evolving enhancer protein 1 (OEE1) and ATP synthase CF1 alpha subunit were the important components of the photosynthesis machine of the plants, functioning in light absorption, carbon dioxide fixation and so

on. The 50S ribosomal protein L1, which was located in chloroplasts and contributed to chloroplast proteins synthesis, and glutamate-1-semialdehyde 2, 1-aminomutase involved in porphyrin and chlorophyll biosynthesis were both down-regulated after salt treatment. Obviously, these proteins were of importance in the chloroplast formation. This results indicated that salt stress not only inhibit photosynthesis process, but also affect the biosynthesis of the chloroplast. Together with previous results that chlorophyll and total carotenoid contents of leaves decreased in general under salt stress (Parida and Das, 2005), my results therefore indicated that chloroplast was one of the major organelles that can be severely affected by the salt stress. Meanwhile, lipoxygenase and serine hydroxymethyltransferase involved in the stress and defense response, as mentioned before, were up-regulated in the soybean leaves by the salt treatment.

In order to further understand the influence of the salt stress on the chloroplasts, I have tried to isolate the chloroplasts and performed the organelle proteomic studies. Although some proteins involved in the photosynthesis were again identified to be affected by the salt stress (such as rubisco and rubisco activase), the protocols still needed further optimization.

Table 2.1: Differentially expressed proteins between wild type and cultivated soybean identified by MALDI-TOF/TOF. 'up' means the protein was upregulated in the wild type soybean; 'down' means the protein was downregulated in the wild type soybean.

Spots No.	Accession No.	Protein names	theoretial MW/pI	match peptide	protein score	CI%	changed
primary metabolism							
MLS702	Glyma01g01180.1	NADP dependent malic enzyme	64985.8/ 5.83	18	253	100	up
MLS1	Glyma01g36750.1	aconitate hydratase, cytoplasmic	98459.8/ 5.56	20	184	100	up
MRS340	Glyma02g40820.1	isocitrate dehydrogenase (NADP) (EC 1.1.1.42)	46050.5/ 5.87	14	123	100	up
URS342	Glyma02g40820.1	isocitrate dehydrogenase (NADP) (EC 1.1.1.42)	46050.5/ 5.87	18	156	100	down
ULS8	Glyma04g01750.1	glyceraldehyde-3-phosphate dehydrogenase B subunit	48143.8/ 7.1	18	314	100	down
MLG892	Glyma04g36870.1	glyceraldehyde-3-phosphate dehydrogenase	36818.1/7.1	12	251	100	up
ULS9	Glyma06g01850.1	glyceraldehyde-3-phosphate dehydrogenase B subunit	48401.9/ 6.76	20	444	100	down
MLG927	Glyma18g01330.1	glyceraldehyde 3-phosphate dehydrogenase	36973.2/6.97	13	279	100	up
MRS217	Glyma09g40690.1	2,3-bisphosphoglycerate-independent phosphoglycerate mutase	60831/ 5.51	12	106	100	up
MRS98	Glyma11g08550.2	aconitate hydratase, cytoplasmic	98383.7/ 5.75	16	145	100	up
MLG652	Glyma12g32000.1	aconitase	107173.8/8.12	22	104	100	up
MRS250	Glyma13g30490.1	Pyruvate decarboxylase isozyme 1	63371.2/ 5.73	9	83	100	up
MRS361	Glyma14g00760.2	acetoacetyl-CoA thiolase	38938.1/ 5.91	6	70	99.2	up
MRS341	Glyma14g39160.3	isocitrate dehydrogenase (NADP) (EC 1.1.1.42)	43833.4 6.13	12	118	100	up

ULS11	Glyma15g11490.2	ATP synthase gamma chain 2, chloroplast, putative	41303.9/ 6.69	12	259	100	down
MLS25	Glyma15g41540.1	phosphoglycerate kinase, putative	50035.9/ 7.79	14	84	100	up
MLG917	Glyma15g41550.1	cytosolic phosphoglycerate kinase	42408.6/5.96	11	228	100	up
MRS385	Glyma15g41550.1	cytosolic phosphoglycerate kinase	42408.6/ 5.96	10	103	100	up
URG7507	Glyma15g41550.1	cytosolic phosphoglycerate kinase	42408.6/ 5.96	9	102	100	down
MLS840	Glyma16g34570.1	Ferredoxin--NADP reductase, leaf isozyme	36066.6/ 6.34	11	129	100	up
MRS315	Glyma18g51260.1	6-phosphogluconate dehydrogenase	53617.6/ 5.68	13	130	100	up
MLS748	Glyma20g24830.2	LTA2; dihydrolipoyllysine-residue acetyltransferase	48936.5/ 8.78	8	72	99.5	up
MLS674	Glyma20g38720.1	transketolase	80028.5/ 6.03	16	159	100	up
URS349	Glyma03g03200.1	transketolase, putative	78857.2/ 6.4	7	68	98.8	down
protein and amino acid synthesis							
ULS836	Glyma06g14250.1	nascent polypeptide associated complex alpha subunit	24247.8/ 4.27	7	273	100	down
ULS830	Glyma03g40490.1	cysteine synthase	34413.2/ 5.53	13	272	100	down
ULS833	Glyma02g15640.1	OAS-TL4 cysteine synthase	41621.6/ 6.91	11	181	100	down
ULS867	Glyma04g37120.1	translational elongation factor 1 subunit Bbeta	24972.7/ 4.42	9	149	100	down
ULS2	Glyma06g17360.1	ATP-dependent Clp protease ATP-binding subunit clpC homolog, chloroplastic	102433.8/ 6.09	32	444	100	down
ULG860	Glyma07g01720.1	mitochondrial processing peptidase alpha subunit	54420.9/ 5.89	13	122	100	down
ULS778	Glyma07g39190.1	Aminomethyltransferase, mitochondrial	44308.9/ 8.68	17	297	100	down
MLS770	Glyma07g39190.1	Aminomethyltransferase, mitochondrial	44308.9/ 8.68	17	327	100	up
ULS729	Glyma17g04210.1	dihydrolipoamide reductase	52854.5/ 6.9	17	214	100	down

MRS294	Glyma08g11490.2	serine hydroxymethyltransferase 1/2	51733.2/ 7.59	15	146	100	up
MLS4	Glyma08g37270.1	Serine hydroxymethyltransferase, mitochondrial	57319.6/ 8.6	15	127	100	up
ULG792	Glyma10g40750.1	d-3-phosphoglycerate dehydrogenase	62573.4/ 6.32	7	62	95.4	down
ULG1171	Glyma11g10550.1	RPL12-A (RIBOSOMAL PROTEIN L12-A)	19762.7/ 5.81	3	82	100	down
URG7505	Glyma11g33560.1	cytosolic glutamine synthase	38966.5/ 5.48	12	128	100	down
MRS331	Glyma1337s00200.1	methionine adenosyltransferase	43027.7/ 5.65	9	98	100	up
ULS6	Glyma15g10890.2	glutamine synthetase precursor	47662 6.73	18	469	100	down
MRS127	Glyma16g04240.3	methionine synthase	84229.6/ 5.93	9	68	98.7	up
MLS5	Glyma18g27710.1	Serine hydroxymethyltransferase, mitochondrial	61900.8/ 8.49	15	107	100	up
ULG837	Glyma19g22730.1	leucine aminopeptidase, putative	60796.5/ 6.38	9	65	97.7	down
ULS675	Glyma19g29180.2	methionine synthase	84218.7/ 5.97	25	563	100	down
ULG689	Glyma20g23350.1	prolyl oligopeptidase family protein	86444.8/ 5.8	17	186	100	down
protein storage and target							
MLS878	Glyma01g34300.1	narbonin domain containing protein	31239.2/ 6	11	144	100	up
MLS850	Glyma07g01730.1	Stem 28 kDa glycoprotein	29046.8/ 8.75	14	273	100	up
MLS905	Glyma07g01730.1	Stem 28 kDa glycoprotein	29046.8/ 8.75	18	205	100	up
MLS22	Glyma07g01730.1	Stem 28 kDa glycoprotein	29046.8/ 8.75	16	195	100	up
MLS10	Glyma07g01730.1	Stem 28 kDa glycoprotein	29046.8/ 8.75	8	57	83.8	up
MLG970	Glyma08g13440.2	nutrient reservoir, putative	33426.7/5.16	3	93	100	up
MLG1055	Glyma08g21410.1	Stem 31 kDa glycoprotein	29261.8/6.72	13	344	100	up
MLS886	Glyma08g21410.1	Stem 31 kDa glycoprotein	29261.8/6.72	14	235	100	up

MLS7	Glyma08g21410.1	Stem 31 kDa glycoprotein	29261.8/ 6.72	13	516	100	up
MRG7653	Glyma08g21410.1	Stem 31 kDa glycoprotein	29261.8/ 6.72	8	102	100	up
MLS932	Glyma08g21410.2	Stem 31 kDa glycoprotein	25580.9/ 6.66	12	245	100	up
stress and defense							
URS93	Glyma03g15800.4	laccase 110a	63370/ 6.51	5	160	100	down
MRG7583	Glyma01g24950.2	aldo-keto reductase	34853.1/ 6.19	16	271	100	up
MRS405	Glyma06g12780.3	alcohol dehydrogenase-1F	36891.4/ 5.77	9	96	100	up
MLS665	Glyma07g00890.1	lipoxygenase	96275.7/ 6.22	33	620	100	up
MRS114	Glyma07g00900.1	Lipoxygenase	96757/ 5.78	18	137	100	up
MRG7438	Glyma08g02100.2	monodehydroascorbate reductase	52130.1/ 8.36	18	504	100	up
MLG819	Glyma08g04380.1	aldehyde dehydrogenase,	54802.2/5.71	14	147	100	up
MLG1173	Glyma08g14550.1	dehydration stress-induced protein	20758.1/6.04	4	130	100	up
MRS147	Glyma08g44590.1	heat shock protein 90	80133.8/4.98	10	97	100	up
MRS195	Glyma07g30290.1	hsp70	72424.5/ 5.82	12	100	100	up
MLS866	Glyma09g00660.1	glyoxylase I	32389.6/ 5.76	10	145	100	up
URG7522	Glyma09g02680.1	peroxidase	37990.5/ 6.94	6	114	100	down
MLS12	Glyma12g02790.1	putative cyclophilin	18114.9/ 7.68	9	305	100	up
MLG1068	Glyma12g07780.2	cytosolic ascorbate peroxidase 2	27108.8/5.65	14	316	100	up
MLS9	Glyma12g07780.2	cytosolic ascorbate peroxidase 2	27108.8 5.65	14	502	100	up
MRS470	Glyma12g07780.2	cytosolic ascorbate peroxidase 2	27108.8/ 5.65	9	97	100	up
ULG1072	Glyma12g07780.2	cytosolic ascorbate peroxidase 2	27108.8/ 5.65	14	398	100	down

ULS14	Glyma12g07780.2	cytosolic ascorbate peroxidase 2	27108.8/ 5.65	19	810	100	down
URS513	Glyma12g07780.2	cytosolic ascorbate peroxidase 2	27108.8/ 5.65	12	159	100	down
MRG7664	Glyma12g07780.2	cytosolic ascorbate peroxidase 2	27108.8/5.65	14	316	100	up
URG7467	Glyma12g32160.1	peroxidase precursor	35644.1/ 7.12	12	122	100	down
MLG1117	Glyma13g36000.1	benzoquinone reductase	21725/6.43	6	206	100	up
URG7487	Glyma15g13500.1	peroxidase precursor	38547.7/6.65	8	97	100	down
URS343	Glyma09g02600.1	peroxidase	38497.5/ 5.66	5	142	100	down
MRG7324	Glyma17g02260.1	copper amino oxidase; diamine oxidase	75776/ 6.21	15	142	100	up
MRS107	Glyma17g37820.1	heat shock protein 90	93262.3/ 4.88	14	121	100	up
MLG1099	Glyma18g41410.1	glutathione S-transferase GST 8	25679.6/5.58	9	169	100	up
MRS414	Glyma19g01120.1	quinone oxidoreductase	35056.4/ 5.53	9	118	100	up
URG7523	Glyma20g30910.1	peroxidase	38890.8/ 6.88	7	121	100	down
secondary metabolism							
MLG982	Glyma04g01380.2	isoflavone reductase homolog 2	27862.8/6.92	8	62	95.1	up
ULS5	Glyma16g10880.3	1-deoxy-D-xylulose 5-phosphate reductoisomerase	50558.4/ 5.89	19	397	100	down
MRG7600	Glyma18g52250.1	chalcone reductase	35114.4/ 6.13	14	211	100	up
URG7702	Glyma20g38560.1	chalcone isomerase	23250.2/ 6.23	5	223	100	down
URS485	Glyma20g38560.1	chalcone isomerase	23250.2/ 6.23	7	78	100	down
ULG1122	Glyma20g38560.1	Chalcone--flavonone isomerase 1A	23250.2/ 6.23	5	223	100	down
MRG7688	Glyma20g38570.1	Chalcone--flavonone isomerase 1B-1	25011.8/ 5.26	6	156	100	up
other proteins							

URG7595	Glyma03g28850.1	Glucan endo-1,3-beta-glucosidase	38088.3/ 8.72	10	116	100	down
MRS193	Glyma04g42370.1	UDP-glucose pyrophosphorylase	66085.7/ 5.62	9	79	99.9	up
MRS122	Glyma13g17420.2	sucrose synthese	92185.7/ 6.04	26	229	100	up
ULS4	Glyma15g21890.2	S-adenosylmethionine synthetase	43025.8/ 5.5	24	443	100	down
MLG864	Glyma1337s00200	S-adenosylmethionine synthetase	43027.7/5.65	7	142	100	up
MLG1044	Glyma02g18090.1	lectin	30070.4/ 5.93	10	206	100	up
MLS873	Glyma02g18090.1	lectin	30070.4 5.93	9	135	100	up
URS425	Glyma03g23890.1	unknown protein	37896.5/ 5.94	8	101	100	down
ULG948	Glyma03g40860.4	putative alto/keto reductase	26293.6/ 5.74	8	251	100	down
ULG895	Glyma04g00420.1	Glutamate-1-semialdehyde 2,1-aminomutase	50010.7/6.05	12	236	100	down
ULS7	Glyma04g00420.1	Glutamate-1-semialdehyde 2,1-aminomutase	50010.7 6.05	13	199	100	down
MLG968	Glyma06g46680.1	hsr203J homolog	37709.9/5.53	6	111	100	up
MLS884	Glyma07g15960.1	unknown protein	35648.7/ 8.91	15	219	100	up
MLS695	Glyma08g23990.1	V-type proton ATPase catalytic subunit A	68734.9/ 5.48	29	433	100	up
MLG1101	Glyma08g45610.1	trypsin inhibitor 26 kDa isoform	25930.5/8.57	8	228	100	up
ULG1032	Glyma09g37570.1	peroxisomal voltage-dependent anion-selective channel protein	29737.6/ 8.57	8	90	99.9	down
MLS11	Glyma10g39460.1	photosystem I subunit PsaD	22989/ 9.63	11	370	100	up
ULS858	Glyma15g00560.1	Unknown protein	31273.4/ 5.13	9	233	100	down
MRG7593	Glyma17g36230.1	26S proteasome non-ATPase regulatory subunit	34754.6/ 6.22	10	206	100	up
MLS871	Glyma19g02380.1	peroxisomal voltage-dependent anion-selective channel protein	29686.5/ 9.06	6	78	99.9	up

MRG7502		unidentified					up
MLG1111		unidentified					up
ULG840		unidentified					down
ULG954		unidentified					down
ULG881		unidentified					down

Table 2.2: Differentially expressed proteins from roots of cultivated soybean at germination stage and seedling stage identified by MALDI-TOF/TOF. 'up' means the protein was upregulated in the seedling stage; 'down' means the protein was downregulated in the seedling stage.

Spots No.	Accession No.	Protein names	theoretial MW/pI	match peptide	protein score	CI%	changed
energy metabolism							
G178	Glyma09g40690.1	2,3-bisphosphoglycerate-independent phosphoglycerate mutase	60831/ 5.51	7	182	100	down
G224	Glyma10g40750.1	d-3-phosphoglycerate dehydrogenase	62573.4/ 6.32	10	169	100	down
G275	Glyma18g51260.1	6-phosphogluconate dehydrogenase	53617.6/ 5.68	13	130	100	down
G181-1	Glyma15g11490.2	ATP synthase subunit alpha, mitochondrial	41303.9/ 6.69	12	259	100	down
G380	Glyma13g41960.1	fructokinase 2	35375.4/ 5.29	10	93	100	down
G243	Glyma10g41330.2	mitochondrial ATPase beta subunit	58664.8/ 8.83	15	414	100	down
G318	Glyma14g00760.2	acetyl-CoA C-acetyltransferase	38938.1/ 5 91	8	103	100	down
G345	Glyma08g17600.1	phosphoglycerate kinase	42365.6/ 6.28	16	416	100	down
G296-1	Glyma15g41550.1	cytosolic phosphoglycerate kinase	42408.6/ 5.96	10	103	100	down
G138	Glyma05g34790.1	cytosolic phosphoglucomutase	63474.1/ 5.33	19	303	100	down
S244-1	Glyma02g40820.1	isocitrate dehydrogenase (NADP) (EC 1.1.1.42)	46050.5/ 5.87	18	156	100	up
S172	Glyma02g40550.1	oxalyl-CoA decarboxylase	60638.4/ 6	12	133	100	up
S9	Glyma15g13680.1	ferredoxin NADP oxidoreductase	42164.3/ 8 52	18	311	100	up
S6	Glyma18g53700.1	transaldolase	48032.9/ 6.14	9	132	100	up
protein and amino acid metabolism							
G242	Glyma05g28480.2	adenosylhomocysteinase	53156.1/ 5.6	12	196	100	down

Gc	Glyma06g08670.1	aspartate transaminase	49687.8/ 8.67	18	310	100	down
G300	Glyma09g01270.1	Fumarylacetoacetase	45769.5/ 5.84	7	87	100	down
G407	Glyma11g00810.2	cysteine synthase	34248.2/ 5.69	5	74	99.7	down
G142	Glyma16g04240.3	methionine synthase	84229.6/ 5.93	9	68	98.7	down
G253	Glyma18g04940.1	glutamate decarboxylase	57212.2/ 5.59	12	356	100	down
S2	Glyma19g29180.2	methionine synthase	84218.7/ 5.97	23	384	100	up
G221	Glyma20g26530.1	d-3-phosphoglycerate dehydrogenase	62779.5/ 6.17	8	96	100	down
G325	Glyma19g33570.1	elongation factor tu	49078.6/ 6.4	13	262	100	down
secondary metabolism							
G337	Glyma06g14200.1	caffeoyl-CoA-O-methyltransferase	39869.3/ 5.57	5	82	99.9	down
G436	Glyma07g33780.1	caffeoyl-CoA-O-methyltransferase	28053.4/ 5.46	12	220	100	down
stress and defense related							
S74	Glyma03g15800.4	laccase 110a	63370/ 6.51	5	160	100	up
G376-1	Glyma05g22180.1	peroxidase2	35475/ 9.03	12	282	100	down
G163	Glyma07g30290.1	hsp70	72424.5/ 5.82	12	100	100	down
S540	Glyma07g37240.3	Stress-induced protein SAM22	15524.9/ 4.74	8	228	100	up
S276	Glyma08g02100.2	monodehydroascorbate reductase	52130.1/ 8.36	18	504	100	up
S254-1	Glyma09g02600.1	peroxidase precursor	38497.5/ 5.66	5	142	100	up
S15	Glyma09g02650.1	peroxidase1A	37236.2/ 4.14	3	111	100	up
S481	Glyma10g43730.1	GSH-dependent dehydroascorbate reductase 1	23434.4/ 5.98	8	102	100	up
G546	Glyma12g02790.1	cyclophilin	18114.9/ 7.68	9	296	100	down

S302	Glyma12g32160.1	peroxidase precursor	35644.1/ 7.12	13	226	100	up
S327	Glyma12g32160.1	peroxidase precursor	35644.1/ 7.12	8	68	98.9	up
S445-1	Glyma12g34500.1	benzoquinone reductase	21754/ 6.43	7	191	100	up
S484	Glyma15g07040.1	quinone oxidoreductase/Minor allergen Alt a	21685/ 6.43	9	372	100	up
S269	Glyma15g13560.1	peroxidase	39066.7/ 7.57	5	147	100	up
G140	Glyma16g00410.1	Chaperone DnaK/chloroplast heat shock protein 70-2	73709.4 5.2	14	171	100	down
S7	Glyma16g27880.1	peroxidase	37750.3/ 6.99	11	324	100	up
G145	Glyma17g02260.1	copper amino oxidase; diamine oxidase	75776/ 6.21	14	147	100	down
G197	Glyma18g53590.1	chaperonin containing t-complex protein 1, zeta subunit	58984.7/ 5.68	6	68	98.9	down
S8	Glyma19g01150.1	quinone oxidoreductase	34308.2/ 5.54	4	76	99.8	up
miscellanous							
G356	Glyma10g06580.1	adenosine kinase	37538.9/ 5.32	4	135	100	down
G446	Glyma08g21410.1	Vegetative storage protein A	29261.8/ 6.72	8	84	100	down
G399-1	Glyma09g37570.1	peroxisomal voltage-dependent anion-selective channel	29737.6/ 8.57	9	254	100	down
S402	Glyma17g36230.1	26S proteasome non-ATPase regulatory subunit	34754.6/ 6.22	10	206	100	up
S14	Glyma05g30780.1	unknown	12313/ 4.29	2	130	100	up
S10	Glyma17g15690.1	unknown	27650.4/ 6.3	8	266	100	up
G515	Glyma11g21020.1	unknown	21216.9/ 5.9	8	381	100	down
S11		unidentified					up
G160		unidentified					down

Table 2 3 Differentially expressed proteins under the salinity stress identified by MALDI-TOF/TOF 'down' indicated that the protein was downregulated under the salinity stress, while 'up' indicated that the protein was upregulated under the salinity stress

Spots No	Accession No	Protein names	Species	theoretial MW/pI	matched peptides	protein score	CI%	changed
Photosynthesis								
C3	gi|20257348	ribulose 1,5-bisphosphate carboxylase/oxygenase	*Madhuca microphylla*	50890 6/6 14	21	154	100	down
C8	gi|6691141	ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit	*Olinia ventosa*	50710 6/6 6	10	87	99 9	down
C15	gi|9909636	ribulose 1,5-bisphosphate carboxylase/oxygenase	*Elatine hydropiper*	26562 4/7 01	8	84	99 9	down
C17	gi|2980643	ribulose 1,5-bisphosphate carboxylase/oxygenase	*Phylica pubescens*	32221 2/6 7	8	93	100	down
C30	gi|3115154	ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit	*Limonium mouretii*	46606 4/5 86	8	86	100	down
C6	gi|10720249	rubisco activase precursor	*Vigna radiata*	47871 3/7 57	5	101	100	down
C24	gi|115800	chlorophyll a-b binding protein 3, chloroplast precursor	*Glycine max*	27844 1/5 46	6	158	100	down
C11	gi|15237225	HCF136 (High chlorophyll fluorescence 136)	*Arabidopsis thaliana*	44076 4/6 79	8	155	100	down
C19	gi|131384	oxygen evolving enhancer protein 1, chloroplast precursor	*Pisum sativum*	34871 8/6 25	9	208	100	down
C20	gi|131384	oxygen evolving enhancer protein 1, chloroplast precursor	*Pisum sativum*	34871 8/6 25	11	327	100	down
C49	gi|131384	oxygen evolving enhancer protein 1, chloroplast precursor	*Pisum sativum*	34871 8/6 25	6	66	97 0	down
Primary metabolism								
T12	gi|435103	glyceraldehyde-phosphate dehydrogenase	*Pinus sylvestris*	44205 8/7 55	3	85	99 9	up
C1	gi|13518443	ATP synthase CF1 alpha subunit	*Lotus japonicus*	55746 3/5 22	18	180	100	down
Stress and defense								
T4	gi|118138511	lipoxygenase-B	*Glycine max*	96597 5/5 63	25	393	100	up

Protein and amino acid synthesis									
	C14	gi|1350625	50s ribosomal protein L1, chloroplast precursor (CL1)	*Pisum sativum*	23481.1/10.23	3	114	100	down
	T8	gi|11762130	serine hydroxymethyltransferase	*Arabidopsis thaliana*	51764.3/7.12	6	60	89.4	up
Others									
	T15	gi|1706546	glucan endo-beta-1,3-glucosidase precursor	*Glycine max*	38088.3/8.72	14	510	100	up
	C5	gi|1170031	glutamate-1-semialdehyde 2,1-aminomutase	*Glycine max*	49614.3/5.67	10	79	99.8	down
	C22		unidentified						down
	C25		unidentified						down
	C29		unidentified						down
	C13		unidentified						down

Chapter 3 Identification of histone modifications and histone variants in soybean

3.1 Introduction

My proteomic studies in soybean and many other studies in plants have indicated that plants regulate their gene transcriptions to adapt to the abiotic stress. However, how can these plants perceived these stress signals and then regulate their transcriptions are still not very clear. As aforementioned, histone modifications play important roles in many physiological processes, including regulating gene transcription. Therefore, I wonder whether these histone modifications also involved in the plants response to abiotic stress by regulating their gene transcription.

Epigenetic studies of chromatin in model organisms have provided insights into the modifications of histones, ranging from the identification of several enzymes and related effectors associated with histone modifications to their biological functions in cell (Kouzarides *et al.*, 2007; Simon *et al.*, 2007). Although numerous works have been done, information of the plant histone modifications was still limited. Moreover, most of them were mainly focusing on *Arabidopsis*, rice and maize. Previously, no study about histone modifications in soybean, an important economic crop with a diploidized tetraploid genome (~950Mb) which is much larger than that of *A. thaliana* (125Mb) (*Arabidopsis* Genome Initiative, 2000), has been reported. In order to facilitate my following studies about understanding the relationships between histone modifications and soybean salinity stress response, I first tried to reveal the histone modification patterns and histone variants of soybean using matrix-assisted laser desorption/ionization-time-of-flight mass spectrometry (MALDI-TOF MS), in combination with nano-liquid chromatography (nano-LC). I also hypothesized that histone modifications and histone variants might change when the soybeans were

treated with salt, leading me to investigate their dynamic changes under the salinity stress.

3.2 Material and methods

3.2.1 Plant materials

Soybean (*Glycine max* [L.] Merr. Cultivar Union) was germinated in soil under greenhouse conditions. One week later, the plants were transferred and cultured in 1× Hoagland nutrient solution. At the growth stage with 3-4 leaves, the leaves were harvested, frozen immediately in liquid nitrogen and stored at -80 ˚C.

3.2.2 Nuclei extraction and histone isolation

Soybean tissues were ground into powder in liquid nitrogen, and suspended in nuclei isolation buffer (NIB) containing 20 mM Tris-HCl (pH 7.5), 10 mM KCl, 10 mM $MgCl_2$, 6% sucrose, 0.6% Triton X-100, 0.05% β-mercaptoethanol, 1 mM phenylmethylsulfonyl fluoride (PMSF), as described (with some modifications) previously (Calikowski *et al.*, 2006). After being homogenized on ice bath, the tissue was filtered using filter paper (pore size 30 μm). The resulting nuclei fraction was harvested by centrifugation at 4000 *g* for 10 min, and then washed twice with NIB. Core histones were extracted in 0.4 M hydrochloric acid (HCl). The extract was then centrifuged at 12000 *g* for 10 min. The core histones in the supernatant were precipitated by TCA. The pellet of core histone mixture was collected and stored at -20 ˚ C until use.

In order to separate the individual histone molecules, 100 μg of purified core histones were dissolved in water and separated by reversed–phase high performance liquid chromatography (RP-HPLC) (Agilent 1100 series) using C4 column (4.6×250mm; 5 μm) at a flow rate of 1 ml/min (Shechter *et al.*, 2007). The running program was:

buffer A 10 min; 35% to 65% buffer B in 100 min, then 65% to 100% buffer B in 10 min. Buffer A was 0.1% TFA in water; Buffer B was 0.05% TFA in ACN. The eluted fractions were lyophilized and stored at -80° C until use.

3.2.3 Histone protein in-gel digestion and nano-liquid chromatography

The purified histone powder was re-dissolved using 1 × SDS-PAGE sample loading buffer and subjected to 15% SDS-PAGE for analysis. Corresponding histone bands were excised and cut into small pieces. The gel was de-stained twice using the destaining buffer (50% methanol, 50mM Na_2CO_3 in water), dehydrated using ACN and then dried by SpeedVac for 5 min. The de-stained gel slices were immersed in 10-15 μl endoproteinase (15 ng/μl trypsin (Promega) or 5 ng/μl Lys-c (Roche)) and after overnight digestion at 30 °C, the gel was sonicated (135W, 42 KHz) for 10 min to extract the digested peptides. After centrifugation, 0.8 μl aliquots of the supernatants were spotted onto the MALDI sample plate and dried in air, followed by adding 0.5 μl of the matrix solution containing α-yano-4-hydroxycinnamic acid in 50% ACN / 0.1% TFA for MS analysis.

The purified histone powder was also subjected to in solution digestion. The powder was suspended in 20 μl endoproteinase and digested at 30 °C overnight. The solution was then adjusted to 40 μl with buffer A (2 % ACN, 0.05 % TFA in water) and separated by Nano-LC which was automatically performed using the C18 microcolumn (PrepMap100 3μm, 15cm×75μm, LC Packings, Dionex) on the nano-LC Packings UltiMate™ systems (UltiMate System SwitchosII, Advanced Microcolumn Switching Unit, FAMOSII™ Microautosampler, Probot™ MicroFraction Collector). The elution of peptides was accomplished adopting a linear gradient from 30 % mobile phase buffer A to 90 % buffer B (80% ACN, 0.05 % TFA in water) in 90 rnin at a flow rate of 0.3μl/min. Each fraction was

autocollected on the MALDI-TOF sample plate.

3.2.4 Mass spectrometry

Mass spectrometric analysis was carried out using a MALDI-TOF/TOF tandem mass spectrometer ABI 4700 proteomics analyzer (Applied Biosystems, USA). Mass data acquisitions were piloted by 4000 Series Explorer™ Software v3.0. Linear mode MS were operated over the mass range 5k-25k *m/z* for full protein detection. Reflector mode MS survey scan were acquired over the mass range 600-3500 *m/z* in the positive-ion mode and accumulated from 2000 laser shots with acceleration of 20 kV. The MS spectra were internally calibrated using porcine trypsin autolytic products (*m/z* 842.509, *m/z* 1045.564, *m/z* 1940.935 and *m/z* 2211.104) resulted in mass errors of less than 30 ppm. The MS peaks (MH^+) were detected on minimum S/N ratio≥20 and cluster area S/N threshold ≥25 without smoothing and raw spectrum filtering. Peptide precursor ions corresponding to contaminants including keratins and the trypsin autolytic products were excluded. The filtered precursor ions with a user-defined threshold (S/N ratio ≥50) were selected for the MS/MS scan. Fragmentation of precursor ions was performed using MS-MS 1kV positive mode with CID on and argon/air as the collision gas. MS/MS spectra were accumulated from 3000 laser shots using default calibration with Glu-Fibrinopeptide B from 4700 Calibration Mixture (Applied Biosystems, USA). The MS/MS peaks were detected on minimum S/N ratio ≥3 and cluster area S/N threshold ≥15 with smoothing.

3.2.5 Database search

The MS and MS/MS data were loaded into the GPS Explorer™ software v3.5 (Applied Biosystems, Foster City, USA) and searched against NCBI database by Mascot search engine v1.9.05 (Matrix science, London, UK) using combined MS

(peptide-mass-fingerprint approach) with MS/MS (*de novo* sequencing approach) analysis for protein identification. The following search parameters were used: monoisotopic peptide mass (MH^+); 700-3500 Dalton; one missed cleavage per peptide; enzyme, trypsin/Lys-C; taxonomy, all taxonomy and green plants; p*I*, 0-14; precursor ion mass tolerance, 50ppm; MS/MS fragment-ion mass tolerance, 0.1 Da; variable modifications, carbamidomethylation of cysteine, oxidation of methionine, acetylation of Lysine and arginine, mono-, di- and tri-methylation of Lysine were allowed. Known contaminant ions corresponding to trypsin and keratins were excluded from the peak lists before database searching. Top ten hits for each protein search were reported. For PTMs confirmation by MS/MS analysis, *De novo* Explorer™ software (Applied Biosystems, Foster City, USA) was used to deduce the amino acid sequence of the selected peptide.

3.2.6 Western blotting

Ten μg core histone mixtures were separated in SDS-PAGE gel, and transferred to a polyvinylidene difluoride (PVDF) membrane. The membranes were first blocked in 5% not-fat milk in TBS, and probed with specific primary antibody (1:1000). After three washes with TBST, the membranes were incubated with alkaline phosphatase-conjugated secondary antibody (goat-Anti-rabbit IgG-AP, Santa Cruz Biotechnology) at 1:2000 dilution in TBS. The signal is developed by the NBT/BCIP (Roche). Specific antibodies used in the experiments included: H3K18 acetylation (Upstate, 07-354), H3K23 acetylation (Upstate, 07-355), H3K4 trimethylation (MILLIPORE, 04-745), H3K27 trimethylation (LPBio, AR-0171), H3K36 trimethylation (Upstate, 05-801), H3K79 monomethylation (LPBio, AR-0172) and H3K79 dimethylation (LPBio, AR-0177), H4K8 acetylation (Upstate, 07-328), H4K12 acetylation (Upstate, 04-119).

3.3 Results

3.3.1 Isolation and identification of core histones of soybean

Using reversed phase high-performance liquid chromatography (RP-HPLC), core histones of soybean were separated and eluted in the order of H2B, H4, H2A and H3 between 38-55% of buffer B, and collected according to the UV signal (210 nm) (Figure 3.1). MALDI-TOF MS (linear mode) was employed to monitor the isolated histones in the collected fractions and the calculated mass of histone H4, H3, H2A and H2B were approximately 11.3, 15.2, 15.3 and 16.1 kDa, respectively. According to the results of the RP-HPLC analysis (Figure 3.1), several variants of H2B and H2A were detected. Triton-acetic acid-urea (TAU) gel indicated that at least 5 variants of histone H2B and 4 variants of histone H2A were present in soybean (Figure 3.13). By extending the slope of gradient of buffer B from 35% to 65% ACN in 100 min, two variants of histone H3, H3.1 and H3.2, were also separated (Figure 3.1).

Individual histone protein was also separated via SDS-PAGE. Protein bands containing the corresponding core histones were excised and followed by endoproteinase in-gel digestion. Each histone protein band was divided into two portions and subjected to trypsin or Lys-C digestion respectively before MS analysis. MS analysis covered most of the amino acid sequence of histone H3, which consists of 135 amino acid residues. Most of the 102 amino acid residues in soybean histone H4 were also identified using MS analysis.

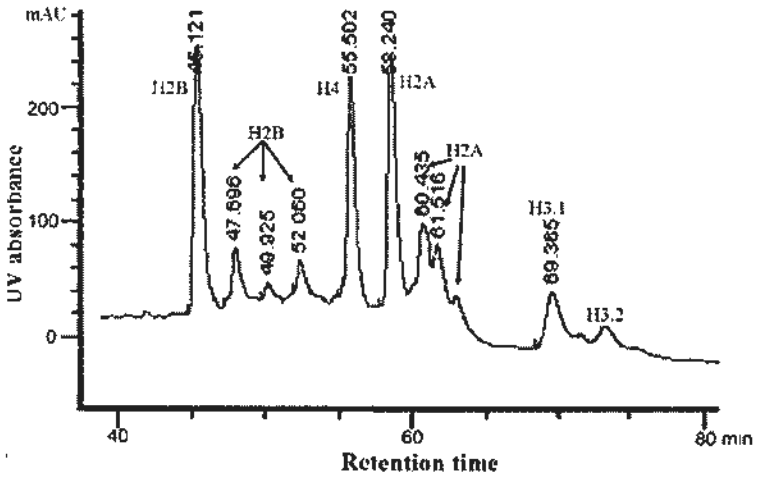


Figure 3.1: Spectrum of histone isolation with RP-HPLC. The core histones were extracted in acid and separated by RP-HPLC. They were eluted in the sequence of histone H2B, H4, H2A and H3. Several variants of histone H2B, H2A and H3 were separated and their retention times were labeled on the top of their corresponding peaks.

3.3.2 Histone modifications of soybean histone H3 and its variants

Two variants of histone H3 were determined in soybean. Although the amino acid sequences of the two variants of *D. melanogaster* histone H3 were very similar and with only four amino acid differences, they could be separated by extending the slope of gradient of buffer B during RP-HPLC separation (McKittrick *et al.*, 2004). Similar methods were adopted to isolate soybean histone H3 variants (Figure 3.1). Two consecutive peaks were eluted between 46.2% - 47.2% of buffer B. These two peaks were collected, digested by trypsin and analyzed by nano-LC/MS-MS separately. In the mass spectrum of the first peak, the histone peptide with the mass of 929.53 containing ^{27}KSAPA^{31}TGGVK^{36} was detected (Figure 3.2). In the mass spectrum of the second peak, another histone peptide with the mass of 959.58, corresponding to ^{27}KSAPT^{31}TGGVK^{36} was identified (Figure 3.3). These two histone peptides were

different in the amino acid residue 31, so the first and second peaks were designated histone H3.1 and H3.2, respectively. I further analyzed the variants of histone H3 using the information from soybean genome database (http://www.phytozome.net/soybean). Data from soybean genome showed that these two histone H3 variants in soybean differed in four amino acids at the position of amino acid 31, 41, 87 and 90. They were $A^{31}F^{41}S^{87}S^{90}$ and $T^{31}Y^{41}H^{87}L^{90}$ in histone H3.1 and H3.2, respectively. Three more peptides from my MS analysis further confirmed this conclusion: peptide precursor ion at *m/z* 3396.60 containing ^{84}FQSS^{87}AVS^{90}ALQEAAEAYLV^{115} and peptide precursor ion at *m/z* 1016.57 containing ^{41}FRPGTVALR^{49} in the mass spectrum of histone H3.1, peptide precursor ion at *m/z* 1032.60 corresponding to ^{41}YRPGTVALR^{49} in the mass spectrum of histone H3.2 (Figure 3.4). In the soybean genome, I also found another histone H3 variant, centromere specific histone H3, which differed greatly in amino acid sequence from the other two variants (Figure 3.5).

Next, the modifications of histone H3 were investigated. Modifications of histone H3 were complicated due to its high abundance of both Lysine and arginine in its primary amino acid sequence (Table 3.1). From the MS analysis, mono-, di- and tri-methylation of Lysine 27 were detected in both histone H3 variants; with mono-methylation as the predominant modification (Figure 3.2 and 3.3). In the trypsin digestion, peptide precursor ions with the mass of *m/z* 959.58, 973.59 and 987.61 represented the mono-, di-, and tri-methylated peptides ^{27}KSAPTTGGVK^{36} of histone variant H3.2 respectively (Figure 3.3). Although such peptide contained two potential methylation sites (Lysine 27 and Lysine 36), *de novo* sequencing clearly indicated that methylation were mainly located at Lysine 27 (Figure 3.3). Methylated Lysine 36 was determined by other peptides whose mass were *m/z* 1349.81, 1363.83 and 1377.84 containing ^{28}SAPTTGGVKKPHR^{40} of histone variant

H3.2. *De novo* sequencing showed that it could also be mono-, di- and tri-methylated (Figure 3.6). More interestingly, most of histone H3 Lysine 36 methylation did not appear in those peptides which contained histone H3 Lysine 27 methylation, since only two very small peaks whose mass were *m/z* 1001.59 and 1015.61 were detected in the MS spectrum (Figure 3.3A), which may be corresponding to the peptides containing methylation at both Lysine 27 and Lysine 36. In addition, no peptide that contained both tri-methylated Lysine 27 and Lysine 36 was identified because of the absence of peptide precursor ion at *m/z* 1029 in Figure 3.3A. Similar results were also obtained in histone variant H3.1 (Figure 3.2). Other PTMs were also observed in the peptides of histone H3. Peptide ^{3}TKQTAR^{8} containing mono-, di- and tri-methylated histone H3 Lysine 4, of which mass were *m/z* 718.43, 732.44 and 746.46 respectively, were detected (Figure 3.7A). Of these three modifications, histone H3 Lysine 4 mono-methylation was the dominant one, and this result was similar to that in *A. thaliana* (Zhang *et al.*, 2007). Lysine acetylation in soybean histone H3 was also identified. Peptides ^{10}STGGK^{14Ac}APR^{17} at the *m/z* 815.40 and ^{18}K^{Ac}QLATK^{23} at the *m/z* 730.42 containing acetylated Lysine 14 and Lysine 18 respectively were shown in Figure 3.7B and 3.7C. Another peptide at the *m/z* 1028.57 containing acetylated Lysine 23 was also detected, which was ^{19}QLATK^{23Ac}AARK^{27} (Figure 3.7D). Since the mass shift of acetylation and tri-methylation were very similar (~ 42Da), Western blotting with specific antibodies to these acetylation and tri-methylation sites was performed and further confirmed my MS results (Figure 3.8A).

Methylation of histone H3 Lysine 79 was observed in my studies. Such methylation was frequently found in mammals (Barski *et al.*, 2007). Compared with the mass of the peptide at *m/z* 1335.66, the mass of the peptides at *m/z* 1349.68 and 1363.69 shifted about 14Da and 28Da (Figure 3.9A). This indicated that these peptides might

be methylated. Fragmentation of the methylated peptide at *m/z* 1349.68 resulted in a MS-MS spectrum containing both complete b-ion series and y-ion series. According to this spectrum (Figure 3.9B), the amino acid sequence of ^{73}EIAQDFK^{79Mono}TDLR^{83} could be assigned to this peptide, which revealed that there was mono-methylation at Lysine 79 in soybean histone H3. Western blotting was performed to confirm this result (Figure 3.8A). Consequently, the peptide with the mass 1363.69 should contain di-methylated histone H3 Lysine 79. Due to their low abundance, *de novo* sequence was not successful; however, Western blotting supported this prediction (Figure 3.8A).

The differences of the modification patterns found in these histone H3 variants were obvious. Although most of their acetylation patterns were similar, their methylation patterns exhibited several differences. Almost all of Lysine 27 in histone variant H3.2 were methylated, whereas some histone variant H3.1 were not methylated at Lysine 27. A peptide precursor ion at *m/z* 915.49 which contained the unmethylated Lysine 27 was detected in histone H3.1 (Figure 3.2A) while the peptide containing unmethylated Lysine 27 of histone H3.2 (with a theoretical mass about 945) were absent in Figure 3.3A. While Lysine 36 methylation can be easily detected in histone H3.2 (Figure 3.6), such methylation was not detected in histone H3.1. Another difference between these two variants was that mono-, di- and tri- methylated Lysine 4 were also only present in histone H3.2 (Figure 3.7A). Although the modifications of the soybean centromere specific histone H3 were not identified in this study, the amino acid residues at all the acetylated sites and two methylated sites (Lysine 27 and Lysine 79) of histone H3.1 and H3.2 were different in the centromere specific histone H3 (Figure 3.5), indicating that the centromere specific histone H3 might have distinct histone modification patterns from that of H3.1 and H3.2.

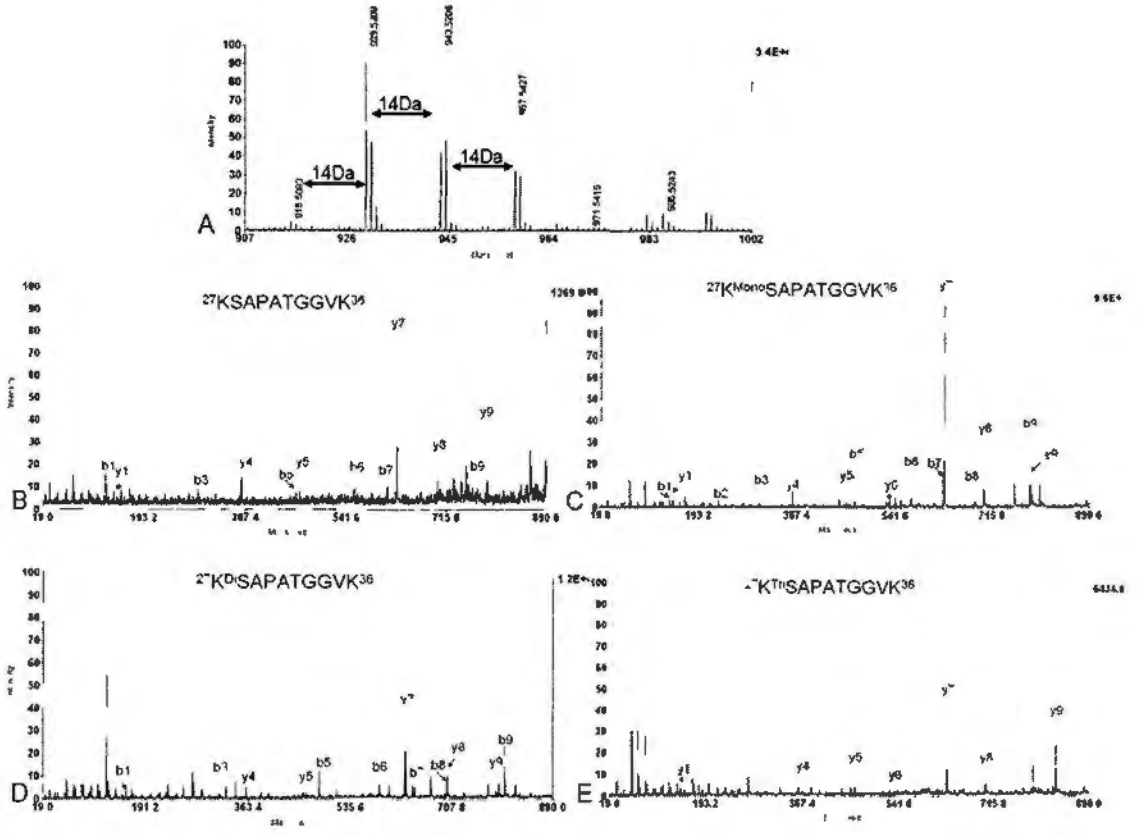


Figure 3.2: Determination of histone variant H3.1 and identification of methylation at Lysine 27 of histone variant H3.1. A. MALDI-TOF mass spectrum showing non- (*m/z* 915.52), mono- (*m/z* 929.53), di- (*m/z* 943.53) and tri- (*m/z* 957.55) methylation at Lysine 27 in the peptide ^{27}KSAPATGGVK^{36} of histone H3.1. B, C, D and E. MS/MS spectrum of the peptide precursor ion at *m/z* 915.52, 929.53, 943.53 and 957.55 determining non-, mono-, di- and tri-methylation at Lysine 27 in the peptide of ^{27}KSAPATGGVK^{36} of histone H3.1, respectively. These results clearly showed that the amino acid sequence of this peptide was KSAPATGGVK and only Lysine 27 was methylated, but not Lysine 36.

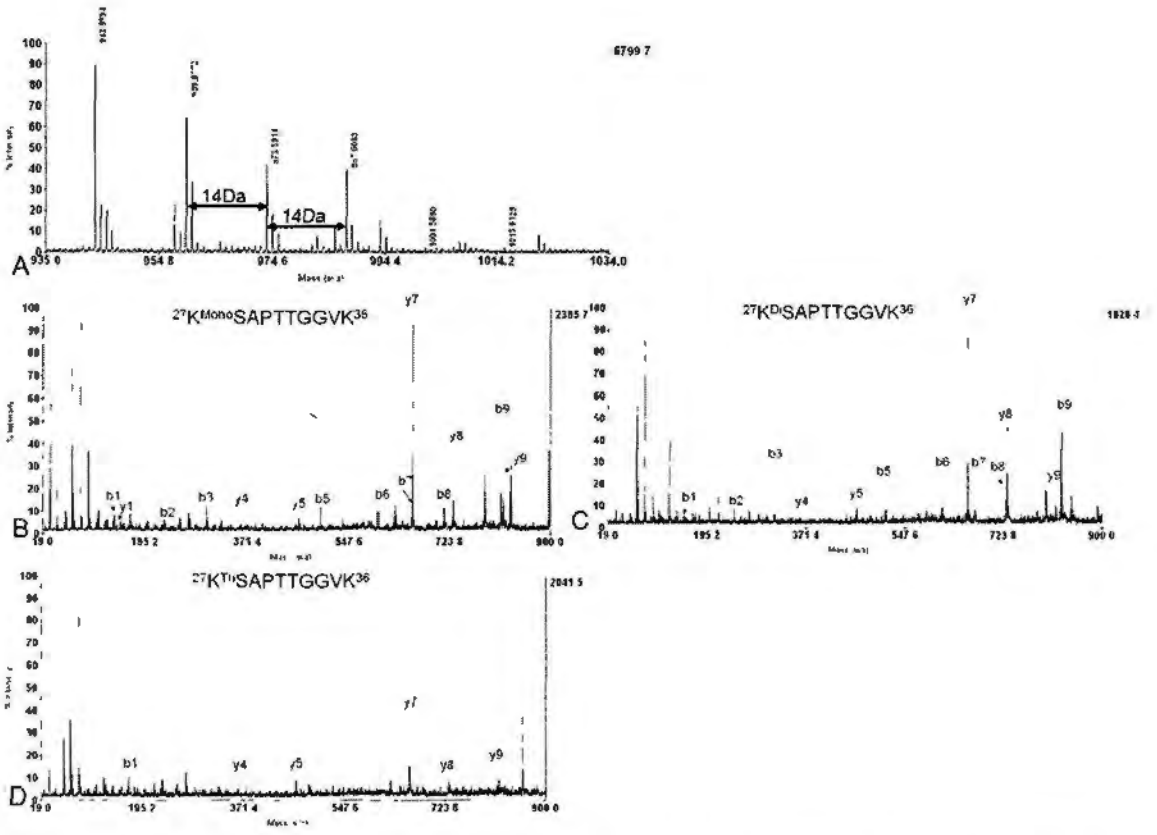


Figure 3.3: Determination of histone variant H3.2 and identification of methylation at Lysine 27 of histone variant H3.2. A. MALDI-TOF mass spectrum showing mono- (*m/z* 959.58), di- (*m/z* 973.59) and tri- (*m/z* 987.61) methylation at Lysine 27 in the peptide ^{27}KSAPTTGGVK^{36} of histone H3.2, but without non-methylation (about *m/z* 945) at this site. B, C and D. MS/MS spectrum of the peptide precursor ion at *m/z* 959.58, 973.59 and 987.61 respectively determining mono-, di- and tri-methylation at Lysine 27 in the peptide of ^{27}KSAPTTGGVK^{36} of histone H3.2. B, C, and D indicated that the amino acid sequence of this peptide was KSAPTTGGVK and only Lysine 27 was methylated, but not Lysine 36.

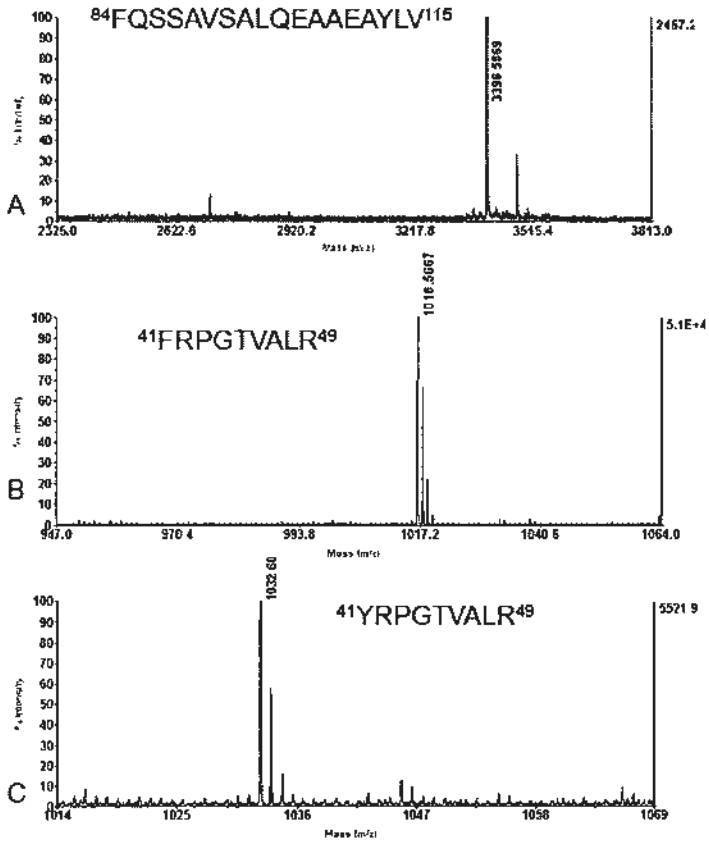


Figure 3.4: Confirmation of two variants of histone H3 of soybean. A and B. MALDI-TOF mass spectrum showing the peptide precursor ion at *m/z* 3396.60 and 1016.57 corresponding to the peptide ^{84}FQSSAVSALQEAAEAYLV^{115} and ^{41}FRPGTVALR^{49} of histone variant H3.1 respectively. C. MALDI-TOF mass spectrum showing the peptide precursor ion at *m/z* 1032.60 corresponding to the peptide ^{41}YRPGTVALR^{49} of histone variant H3.2.

Figure 3.5: Protein sequence alignment of the three variants of histone H3 in soybean. The cetromere specific histone H3 (centro. H3) was very different from the other two histone variants H3.1 and H3.2, while H3.1 and H3.2 were different from each other in only 4 amino acids, $A^{31}F^{41}S^{87}S^{90}$ in H3.1 and $T^{31}Y^{41}H^{87}L^{90}$ in H3.2, which were indicated by red triangle in the figure. Sequences were downloaded from soybean genome database (http://www.phytozome.net/soybean). Accession numbers were as follows: Glyma11g37960.1 for histone H3.2; Glyma11g1 3940.1 for histone H3.1; Glyma07g06310.1 for centro.H3.

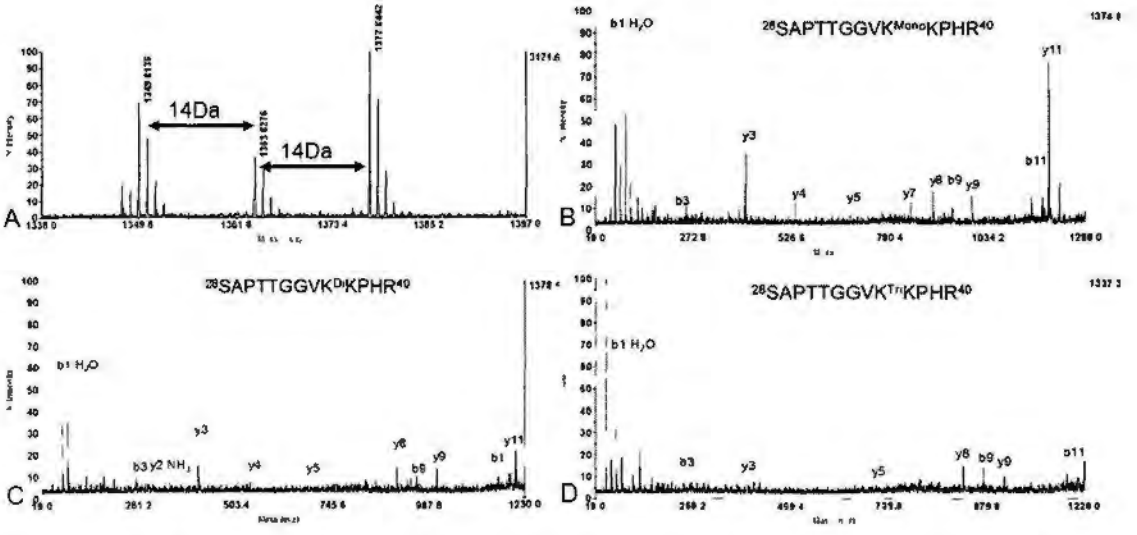


Figure 3.6: Identification of methylation of Lysine 36 of histone variant H3.2. A.

MALDI-TOF mass spectrum showing mono- (*m/z* 1349.81), di- (*m/z* 1363.83) and tri- (*m/z* 1377.84) methylation at Lysine 36 of histone H3.2. B, C and D. MS/MS spectrum of the peptide precursor ion at *m/z* 1349.81, 1363.83 and 1377.84 which determined mono-, di- and tri-methylation at Lysine 36 of histone H3.2, respectively.

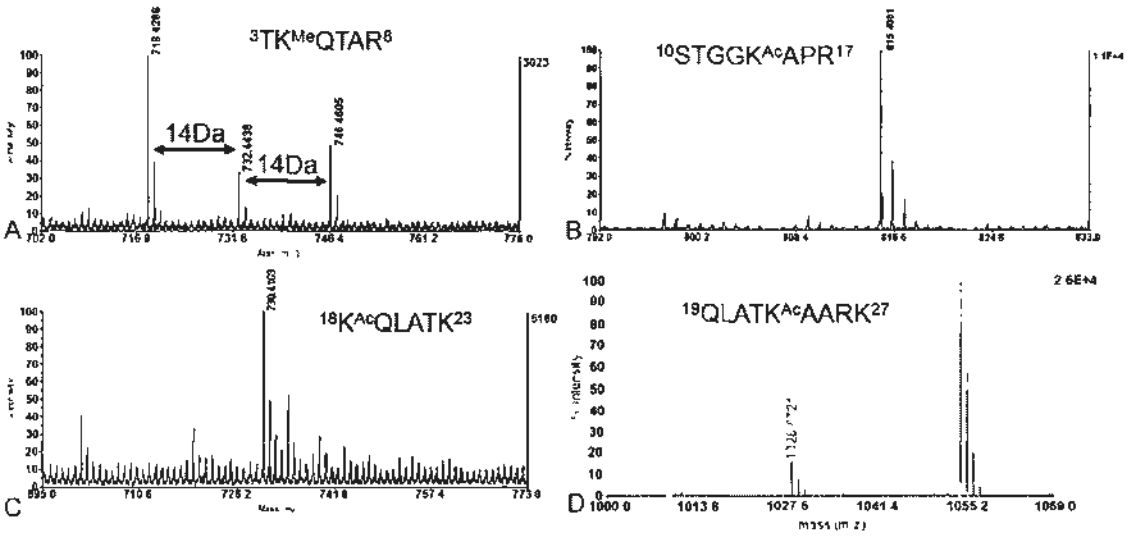


Figure 3.7: Identification of modification sites of histone H3. A. MALDI-TOF mass spectrum showing mono- (*m/z* 718.43), di- (*m/z* 732.44), tri- (*m/z* 746.46) methylation at Lysine 4 of histone H3. B. MALDI-TOF mass spectrum showing acetylation (*m/z* 815.40) at Lysine 14 of histone H3. C. MALDI-TOF mass spectrum showing acetylation (*m/z* 730.42) at Lysine 18 of histone H3. D. MALDI-TOF mass spectrum showing acetylation (*m/z* 1028.57) at Lysine 23 of histone H3. Me: methylation; Ac: acetylation.

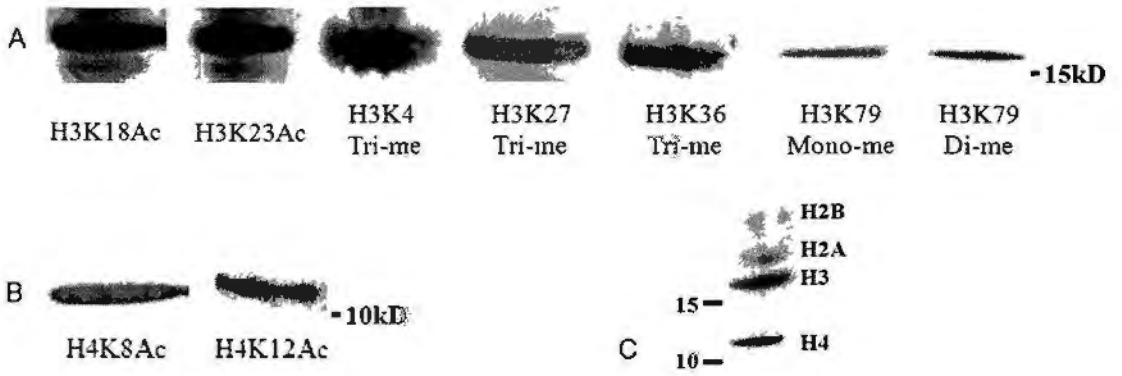


Figure 3.8: Identification of histone modifications in histone H3 and H4 by Western Blotting. Ten μg soybean core histone mixtures were separated in 15% SDS-PAGE gel, and transferred to a PVDF membrane (one μg samples were used when antibodies that recognized H3K18Ac and H3K23Ac were used). A. Western blotting showed the presence of H3K18Ac, H3K23Ac, H3K4Tri-me, H3K27Tri-me, H3K36Tri-me, H3K79Mono-me and H3K79Di-me in histone H3. B. Western blotting showed the presence of H4K8Ac and H4K12Ac in histone H4. C. Coomassie stained SDS-PAGE gel showed the soybean core histone H2A, H2B, H3 and H4. Specific antibodies used were marked under their corresponding figure. Ac: acetylation; Me: methylation.

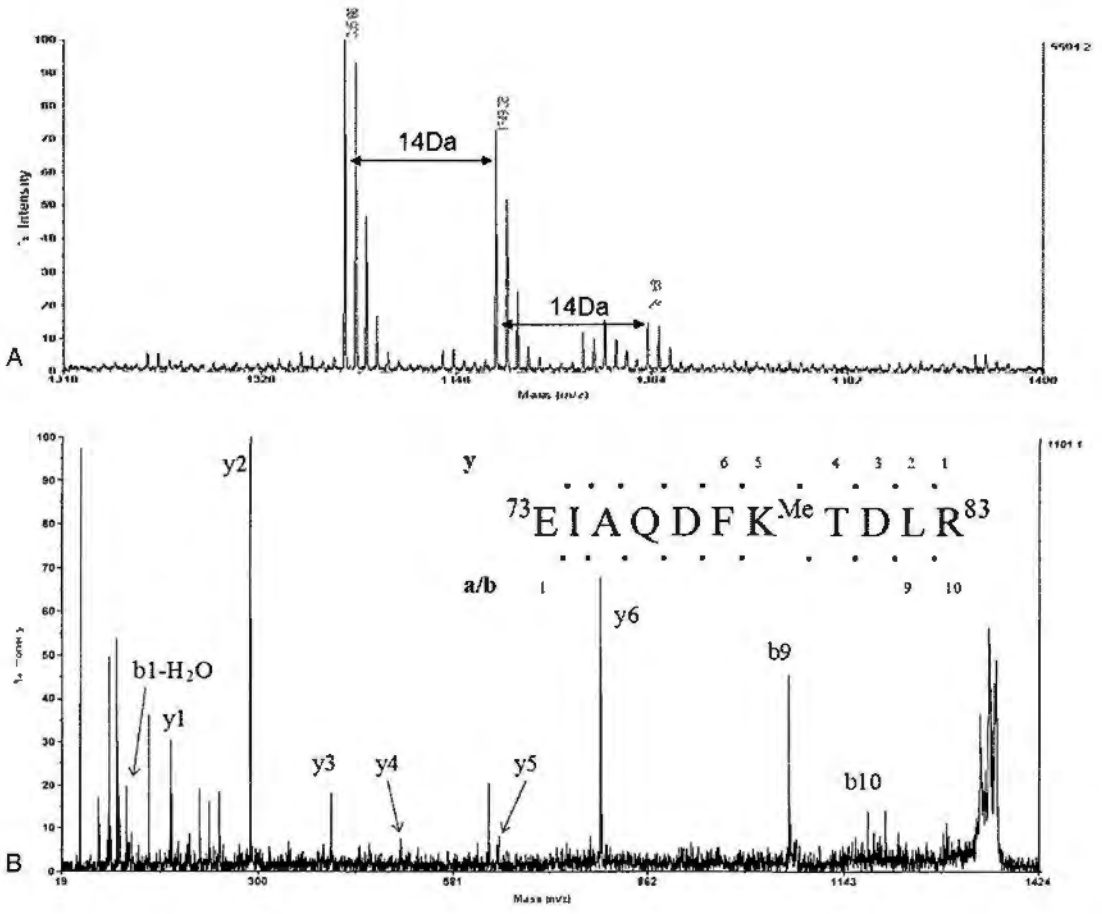


Figure 3.9: Identification of methylation at Lysine 79 of histone H3. A. MALDI-TOF mass spectrum showing non- (*m/z* 1335.66), mono- (*m/z* 1349.68), and di- (*m/z* 1363.69) methylation at Lysine 79 of histone H3. B. MS/MS spectrum of the peptide precursor ion with the mass 1349.68, demonstrating mono-methylation at Lysine 79 in the peptide ^{73}EIAQDFK^{79}TDLR^{83}. However, my data did not indicate that whether histone H3 Lysine 79 methylation was located in certain histone H3 variant.

Table 3.1: Comparison of PTMs of histone H3 in *Glycine max*, *A. thaliana* and mammals

		Modification Sites			Functions
		Mammals	*A. thaliana*	*G. max*	
Acetylation	K9	+	+	nd	Transcriptional activation
	K14	+	+	+	Transcriptional activation
	K18	+	+	+	Transcriptional activation
	K23	+	+	+	Transcriptional activation
	K56	nd	+	nd	
Methylation	K4	+	+	+	Transcriptional activation
	K9	+	+	nd	Transcriptional repression
	K27	+	+	+	Transcriptional repression
	K36	+	+	+	Transcriptional activation
	K64	+	nd	nd	
	K79	+	nd	+	Telomere silencing
	K122	+	nd	nd	

nd, not detected; +, modification present.

3.3.3 Histone modifications of soybean histone H4 and its variants

Purified histone H4 was digested separately with either trypsin or Lys-C and the corresponding digested fractions were separated and analyzed by nano-LC combined with MS/MS. Most of the potential PTM sites were examined and compared to other species. Acetylation of histone H4 was observed. As shown in Table 3.2, Lysine 8 of histone H4 was acetylated in the peptide ^{6}GGK^{8Ac}GLGK^{12} with the mass of 658.37 (Figure 3.10A). Lysine 12 was acetylated in the histone H4 peptide ^{9}GLGK^{12Ac}GGAK^{16} with mass at *m/z* 729.42 (Figure 3.10B). None of the two unacetylated or di-acetylated peptide precursor ions was detected. I also detected a peptide precursor ion with mass at *m/z* 1456.92, which corresponded to the peptide ^{1}SGRGKGGKGLGK^{12Ac}GGAK^{16} (Figure 3.10C), and further proved that Lysine 12 could be acetylated. Similarly, these acetylation sites were further verified by Western blotting with specific antibodies to histone H4 Lysine 8 acetylation and Lysine 12 acetylation (Figure 3.8B). However, acetylation of Lysine 5 and 16 were not detected. My data thus indicated that Lysine 8 and 12 were the main acetylation sites in the N terminus of soybean histone H4 and their acetylation might not happen simultaneously; a result that is differed from those found in histone H4 of *A. thaliana* and mammals (Zhang *et al.*, 2007). In my MS analysis, I cannot detect histone H4 Lysine 20 modification, whereas the Western blotting results showed that histone H4 Lysine 20 methylation did present in soybean.

Two variants of histone H4 were identified (designated as H4.1 and H4.2), which varied at the amino acid residue I^{60} of histone H4.1 and V^{60} of histone H4.2 (Figure 3.11). The trypsin digested peptides of histone H4 were directly applied to MALDI-TOF/TOF analysis and after peptide mass fingerprinting search, the peptide precursor ion at *m/z* 1003.65 was readily detected. Further *de novo* sequencing showed that it contained the amino acid sequence of ^{60}IFLENVIR^{67}. Meanwhile, in the nano-LC fractionated histone H4 peptides, another peptide with the amino acid

sequence of ^{60}VFLENVIR^{67} with the mass of 989.55 was detected. Although only one peak representing histone H4 was observed in the RP-HPLC spectrum (Figure 3.1B), it may be due to the high similarity in the hydrophobicity of the two variants so that they can not be separated using such method.

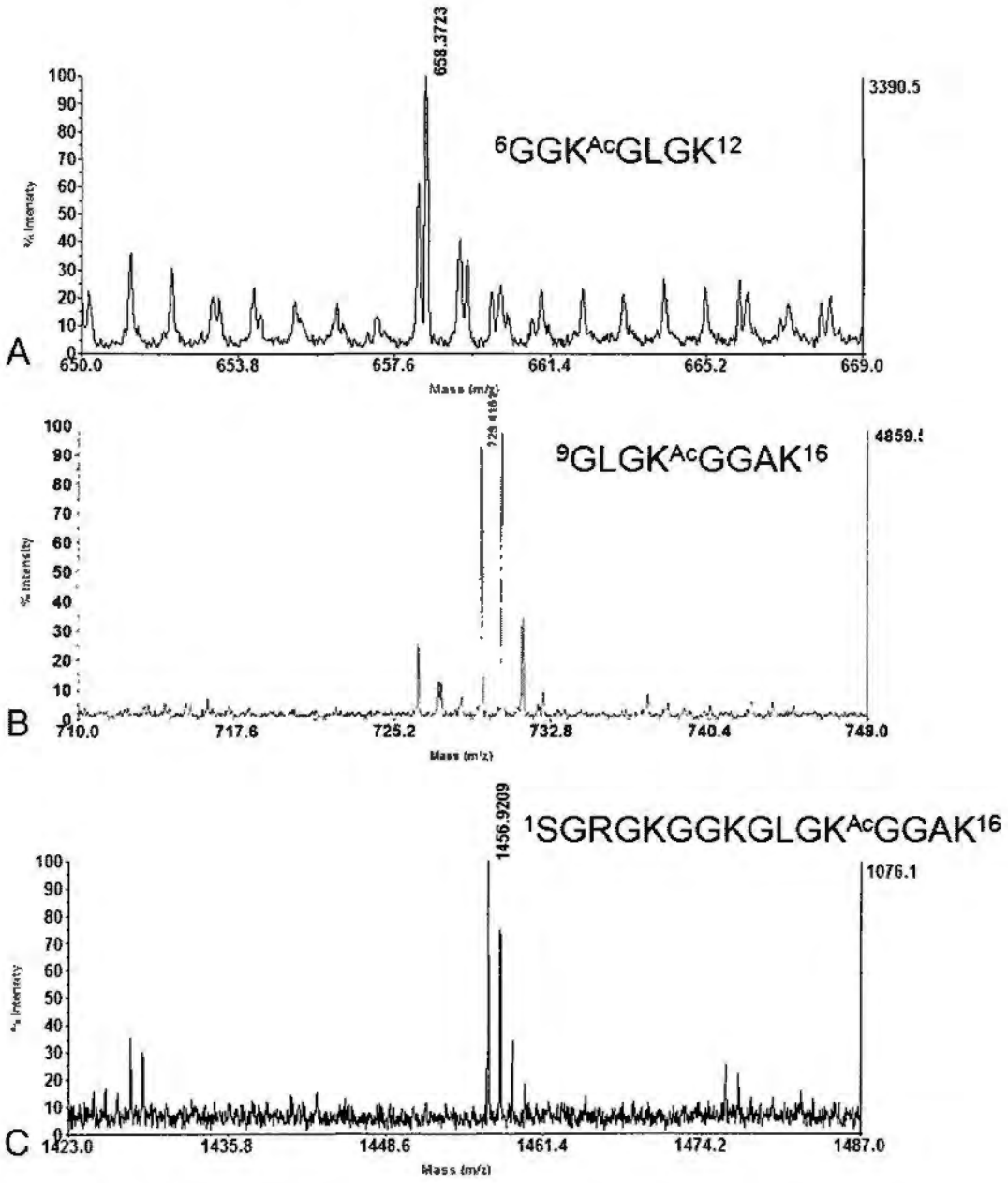


Figure 3.10: Identification of acetylation sites in histone H4. A. MALDI-TOF mass spectrum showing the acetylation (*m/z* 658.37) at Lysine 8 of histone H4. B and C. MALDI-TOF mass spectrum showing the acetylation (*m/z* 729.42 and 1456.92 respectively) at Lysine 12 of histone H4. Ac: acetylation.

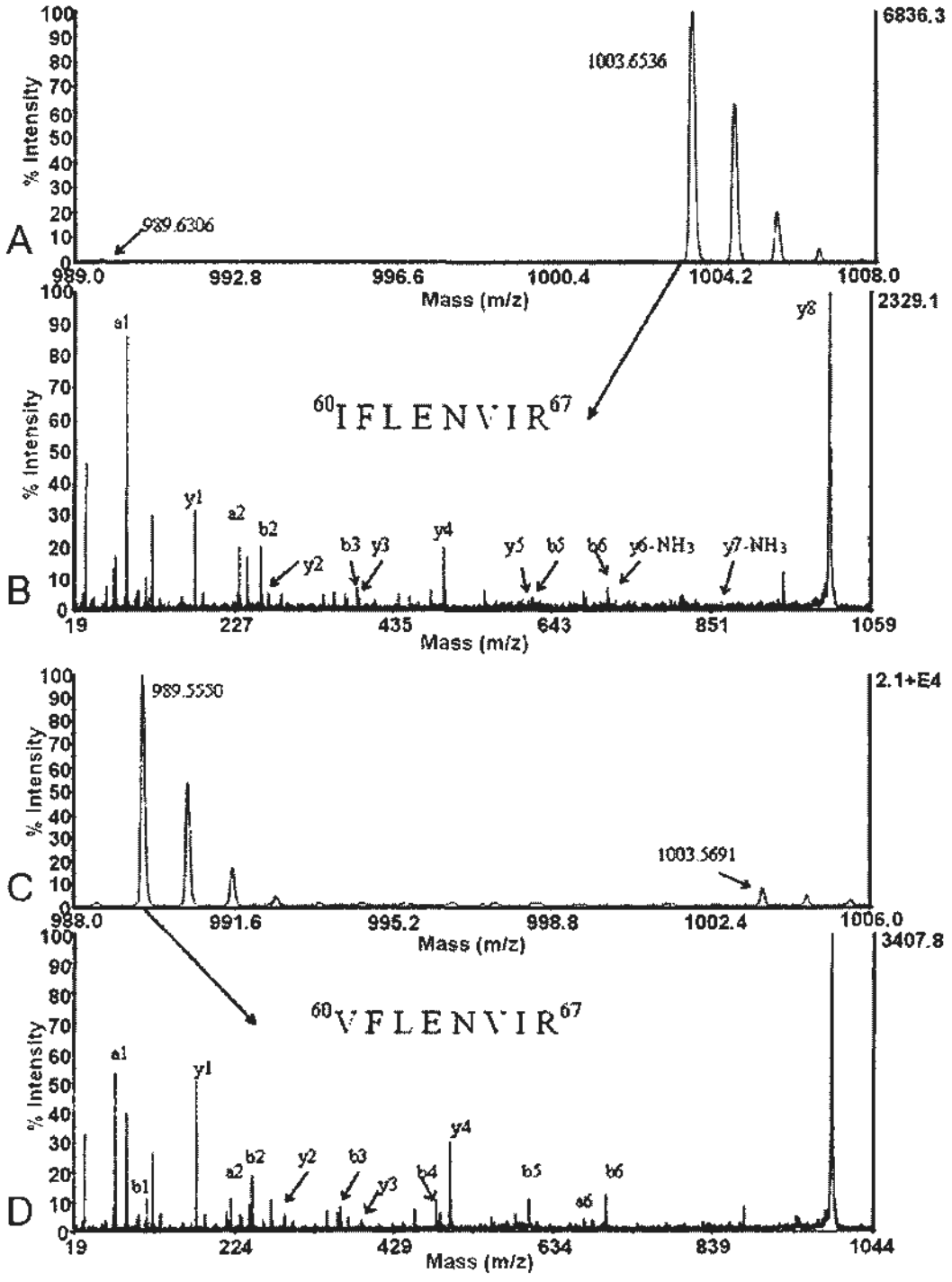


Figure 3.11: Identification of the two variants of histone H4. A. MALDI-TOF mass spectrum showing that the amount of the peptide with calculated mass of *m/z* 1003.6 from histone H4.1 was much more than that of the peptide (*m/z* 989.6) from histone H4.2 in the peptide mass fingerprinting of trypsin digested histone H4. B. MS/MS spectrum showing peptide (*m/z* 1003.6) corresponding to ^{60}IFLENVIR^{67} of histone

variant H4.1. C. MALDI-TOF mass spectrum showing the peptide (*m/z* 989.6) from histone H4.2 after nano-LC separation. D. MS/MS spectrum showing the peptide (*m/z* 989.6) corresponding to ^{60}VFLENVIR^{67} of histone variant H4.2.

Table 3.2: Comparison of PTMs of histone H4 in *Glycine max*, *A. thaliana* and mammals

		Modification Sites			Functions
		Mammals	*A. thaliana*	*G. max*	
Acetylation	K5	+	+	nd	Transcriptional repression
	K8	+	+	+	Transcriptional activation
	K12	+	+	+	Transcriptional activation
	K16	+	+	nd	Transcriptional activation
	K20	nd	+	nd	
Methylation	K20	+	nd	+	Heterochromatin silencing

nd, not detected; +, modification present .

3.3.4 Dynamic changes of histone modifications and histone variants under the salinity stress

With these preliminary results, I then tried to correlate these soybeans' histone modifications and variants with salinity stress by monitoring their dynamic changes under the salinity stress.

Western blotting results showed that salt treatment did not induce obvious changes of

most of the detected histone modifications including histone H3K4 trimethylation, H3K27 trimethylation and H3K36 trimethylation (Figure 3.12). However, compared with the control histone, H3K79 mono-methylaiton and di-methylation increased under the salinity stress. With TAU-SDS PAGE gels, histone variants were separated and their amount was compared. Unfortunately, I did not observe any dramatic changes in these variants when the soybeans were treated with salt (Figure 3.13). Similar results can also be obtained from the RP-HPLC spectrum.

I then tried to work out the roles of the histone H3K79 methylation in the soybean response to salinity stress. According to the "histone code", I struggled to identify whether there were any proteins that would be recruited by this kind of modifications. I synthesized the peptides which contained the modified or unmodified lysine 79 and performed the peptide pull down assays. Unlikely, I could not detect any proteins which exhibited the potential to interact with histone H3K79 methylation specifically (Figure 3.14).

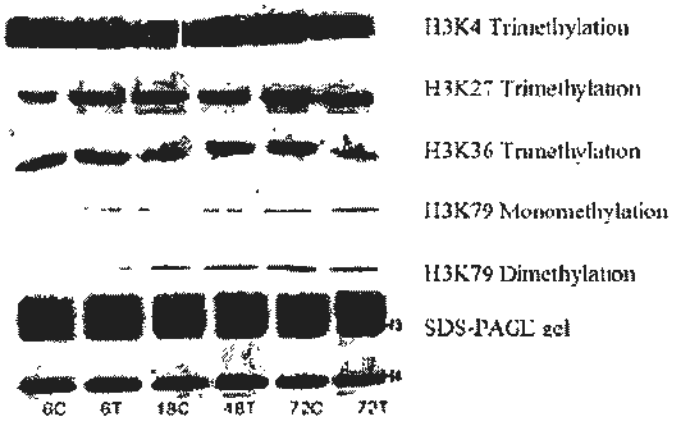


Figure 3.12: Determine the dynamic change of histone modifications under the salinity stress by western blotting. No obvious change of histone H3K4 trimethylation, H3K27 trimethylation and H3K36 trimethylation was observed, while H3K79 methylation (mono- and di-) were increased under the salinity stress. Coomassie blue staining gel was showed as loading control. 6C, 48C and 72C:

control for 6 hours, 48 hours and 72 hours respectively; 6T, 48T and 72T: samples were treated with 0.9% NaCl for 6 hours, 48 hours and 72 hours respectively.

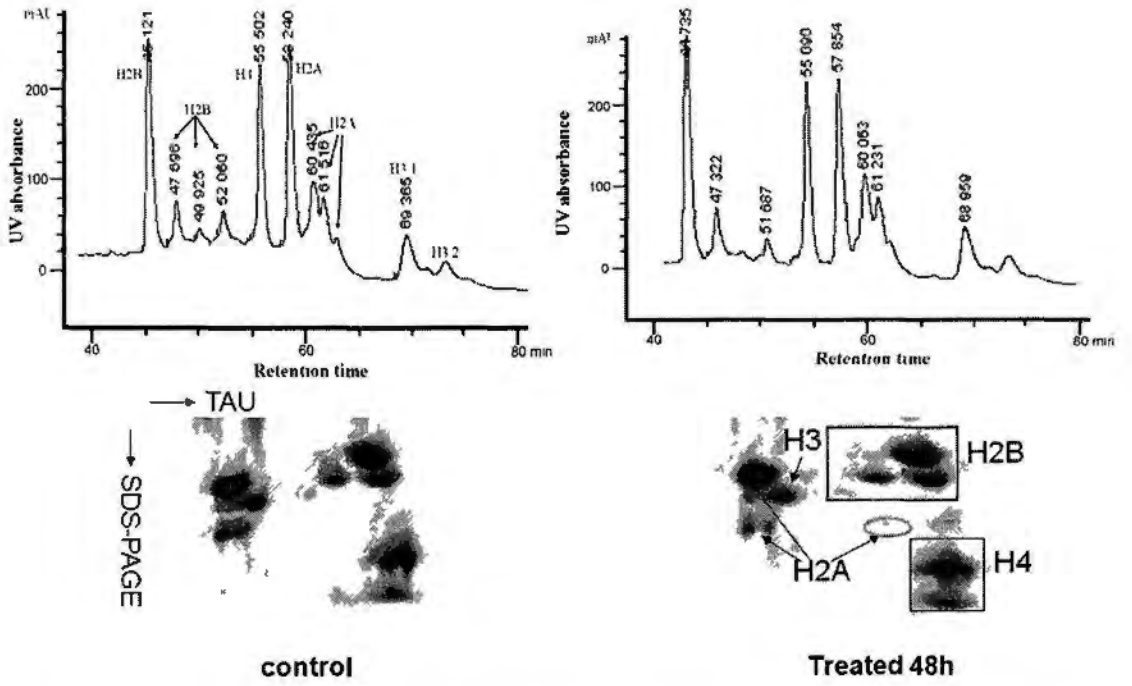


Figure 3.13: Histone variants did not change dramatically under the salinity stress. Upper panel: Spectrum of histone isolation with RP-HPLC. Control: histone was extracted from untreated soybean leaves. Treated 48h: histone was extracted from soybean leaves treated with 0.9% NaCl for 48 hours. Lower panel: TAU-SDS PAGE gel images of the control and treated histone. Equal amount of control and treated histone were applied to the TAU-SDS PAGE gels and histone variants were separated. Different histone variants did not change greatly after salt treatment in consistent with the RP-HPLC results.

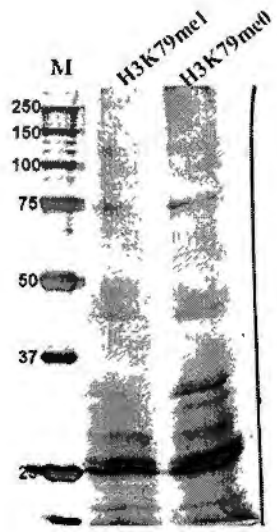


Figure 3.14: Peptide pull down assay with peptide containing H3K79 methylation. Peptide containing H3K79 monomethylation (H3K79me1) or un-modified H3K79 (H3K79me0) were immobilized onto the beads and the nucleic proteins pulled down by these peptides were separated in SDS-PAGE gel and stained by silver. No specific band was present in the H3K79me1 peptide pulled down samples. M: marker.

3.4 Discussions

In general, the amino acid sequences of histones in eukaryote are highly conserved and the posttranslational modification (PTM) patterns on specific amino acid residues are also quite similar. Characterization of histone modifications of histones H3 and H4 in soybean showed similarities to that of *A. thaliana* and other organisms. High density acetylations in the N-terminal tails of histone H3 and H4 were detected in both soybean and other organisms (Allis *et al.*, 2007; Fuchs *et al.*, 2006; Johnson *et al.*, 2004, Berger, 2007). It is suggested that these acetylations play important roles in the transcriptional regulation of many physiological processes in plants, including cold tolerance, floral development and light responsiveness (Earley *et al.*, 2007; Zhu *et al.*, 2008).

However, histone modification patterns in different eukaryotes may also have some

distinct properties. For example, previous studies indicated that histone H4 Lysine 20 modifications were quite distinct between animal and plant. Histone H4 Lysine 20 methylation is evolutionarily conserved from yeast to mammals and is very critical in DNA repair and genome integrity (Sanders *et al.*, 2004). However, histone H4 Lysine 20 was acetylated in *A. thaliana* (Zhang *et al.*, 2007). My results also showed some differences that exist between soybean and the model dicot *A. thaliana*: mono- and di- methylation of Lysine 79 were detected in soybean but such PTMs were not found in *A. thaliana* (Zhang *et al.*, 2007). Western blotting results also showed that methylated histone H3 Lysine 79 might not be widely distributed throughout the whole soybean genome, since when equal amount of histone was applied, the signals of histone H3 Lysine 79 methylation were much weaker than that of other modifications of histone H3 (Figure 3.8A). Studies in yeast and mammals show that histone H3 Lysine 79 is hypermethylated at silenced loci and is important in DNA repair and genome stability (Allis *et al.*, 2007; Ng *et al.*, 2003). Whether this modification is also crucial in maintaining soybean genome integrity requires further investigations.

The patterns of histone H3 Lysine 27 and Lysine 36 methylation were also different between soybean and *A. thaliana*. Previous studies indicate that methylation of Lysine 27 and Lysine 36 carry independent functions: Histone H3 Lysine 27 methylation is mainly involved in gene silencing and heterochromatin formation while methylated histone H3 Lysine 36 is found to be associated with the phosphorylated CTD of Pol II, suggesting a role in gene expression and elongation (Li *et al.*, 2003). In *A. thaliana*, the MADS-box transcription repressor FLOWERING LOCUS C (FLC) is a crucial regulator in controlling flowering time. Histone H3 Lysine 27 methylation usually represses FLC expression while histone H3 Lysine 36 methylation has an opposite effect, suggesting that the modifications at

these two sites must be carefully regulated in order to flower properly (Bastow *et al.*, 2004; He *et al.*, 2005; Xu *et al.*, 2008). In *A. thaliana*, it was reported about 15% of the peptides from histone variant H3.2 were modified with both histone H3 Lysine 27 di-methylation and Lysine 36 mono-methylation (Johnson *et al.*, 2004). So it seems that methylated Lysine 27 and Lysine 36 can coexist on the same histone H3 N-terminus in *A. thaliana*. However, my present MS data revealed that most of the methylated Lysine 27 and methylated Lysine 36 were unlikely to coexist on the same histone H3 molecule in soybean. Therefore, I speculate that soybean and *Arabidopsis* may regulate the occurrence of histone H3 Lysine 27 and Lysine 36 methylation by different ways, although so far little about the relationship between histone H3 Lysine 27 and Lysine 36 has been known.

Analysis of the public database of soybean genome revealed that at least 14 variants of H2A and 12 variants of H2B were present in soybean. It may be due to the gene duplications and reshuffling events happened during soybean diploidized tetraploid genome formation, which occurred at about 8-10 million years ago and 40-50 million years ago respectively (http://soybeangenome.siu.edu). However, I have not identified any PTMs of soybean histone H2B and H2A in my studies so far.

Genomic analysis also found 3 variants of histone H3 in soybean: H3.1, H3.2 and centromere specific histone H3, but I could not isolate the centromere specific histone H3. It may account for the absence of centromere specific histone H3 in soybean leaves which do not undergo active cell division, since other studies indicate that the expression of this variant peaks in late S/G2 phase and it is mainly deposited at functional centromeres (Jansen *et al.*, 2007; Bernstein *et al.*, 2006). The modification patterns of the other two histone H3 variants in soybean were different from those in *A. thaliana*. Only tri-methylation at histone H3 Lysine 36 was found in histone H3.1 of *A. thaliana* (Johnson *et al.*, 2004) while methylated histone H3

Lysine 36 including tri-methylation was absent in soybean histone H3.1 and mono-, di- and tri-methylation of histone H3 Lysine 36 were found in soybean histone H3.2. Besides, histone H3 Lysine 4 methylation was only detected in histone variant H3.2. Histone H3 Lysine 4 methylation is suggested to be associated with euchromatin region and viewed as a marker of transcriptionally active genes (Schübeler *et al.*, 2004; Li *et al.*, 2007). In addition, methylated Lysine 36 is also associated with gene transcription (Li *et al.*, 2003). Previous studies suggested that different variants of histone H3 might carry different functions (Ahmad *et al.*, 2002; Hake *et al.*, 2006). In *D. melanogaster* and *A. thaliana*, the replication-independent histone H3 variants which are usually associated with actively transcribing regions are rich in active modifications, including histone H3 Lysine 4 methylation and acetylations ((Johnson *et al.*, 2004; McKittrick *et al.*, 2004). The presence of modifications (methylation at Lysine 4 and Lysine 36 and acetylation) in soybean histone H3.2 suggested that the soybean histone H3.2 might also be related to actively transcribing genes.

Two soybean histone H4 variants were identified in my study, although histone H4 was the most conserved core histone, and no variant of histone H4 was found previously (Marino-Ramirez *et al.*, 2005). The significance of these two novel histone H4 variants of soybean awaits further investigations.

Western blotting indicated that histone H3K4 trimethylation, H3K27 trimethylation and H3K36 trimethylation did not increase dramatically under the salinity stress. However, these modifications still may be dynamic, since they may increase in some regions of the genome while decrease in other regions, which will not to be reflected in the western blotting where the total histone are used. Therefore, it is necessary to determine whether there are any changes in their distribution patterns and location sites in the genome after the salt treatment. Indeed, some recent reports showed that some histone modifications did increase in some genome regions, such as histone

acetylation and H3S10 phosphorylation (Sokol *et al.*, 2007; Kim *et al.*, 2008).

Interestingly, H3K79 methylation was increased under the salinity stress. Lysine 79 methylation (H3K79me) on histone H3 is a highly conserved modification that exists in many organisms, from fungi to mammal. Dot1 (disruptor of telomeric silencing-1) is responsible for this modification. In yeast, H3K79 methylation is widely distributed across genome but depleted from telomeic loci and other sites where heterochromatin form (Steger *et al.*, 2008). Currently, it is known that the formation of heterochromatin is because of H3K79 methylation refining SIR proteins at heterochromatic sites where repressive complexes assemble (Im *et al.*, 2003). In *Drosophila* and mammals, H3K79 methylation is also located at euchromatin, which indicates H3K79 methylation may associate with gene transcription (Steger *et al.*, 2008). Chromatin assembly factor 1 (CAF-1) interacts with H3K79 to regulate telomeric silencing and DNA repair (Zhou *et al.*, 2006). Another report has found that H3K79 methylation targets p53-binding protein 1 (53BP1) to DNA double-strand breaks and activates the DNA damage checkpoint pathways (Huyen *et al.*, 2004). Therefore, it is predicted that H3K79 methylation may has multiple biological functions, ranging from gene transcription, heterchromatin formation to DNA double-strand breaks (DSBs) repair.

I have made efforts to isolate some effector proteins that may specifically recognize this modification via peptide pull down assay, but disappointingly, I could not get any candidates. Actually, many research groups have failed in trying to find these kinds of proteins in animals. Some researchers supposed that 53BP1 might function as the histone code reader of this modification (Huyen *et al.*, 2004), but recently it turned out to that 53BP1 would recognize H4K20 methylation rather than H3K79 methylation (Botuyan *et al.*, 2006). With so many efforts in vain, some scientists proposed that there were no protein that might specifically recognize this

modification. H3K79 methylation was a quite special histone modification. Rather than located at the N termini of histone H3 where most of the histone modifications located, it was at the core domain of histone, indicating that its surrounding amino acids and structures might be also important for its functions and being recognized by other proteins *in vivo*. However, the protein structure could not be mimicked *in vitro* by peptides which were used in the pull down assay. This finding also indicated that in order to find out the true effector proteins of the H3K79 methylation, the nucleosomes containing only H3K79 methylation should be used to perform the *in vitro* pull down assay.

Chapter 4 GmPHD1 and histone H3K4 methylation

4.1 Introduction

Previous studies demonstrated that histone modifications, such as H3 and H4 acetylation, H3S10 phosphorylation were increased and involved in the plant salinity stress (Sokol, *et al.*, 2007). ChIP studies indicated that levels of H3K4me3, H3K9ac, H3K14ac, H3K23ac and H3K27ac were altered on the coding regions of drought stress-responsive genes, including RD29A, RD29B, RD20, when they were activated under drought stress conditions (Kim *et al.*, 2008). However their detail molecular mechanisms in these processes remain elusive.

H3K4 methylation is widely considered as a marker of active transcriptional genes, since it can recruit many proteins to active gene transcription. Previous reports indicated that PHD containing proteins can read the histone code by recognizing H3K4 methylation via its PHD finger domain. Several proteins containing such domain have been reported in plants; VIN3, VIL1 and VRN5 are involved in vernalization-mediated epigenetic silencing and regulate the flowering time of *Arabidopsis* (Sung and Amasino, 2004; Sung *et al.*, 2006; Greb *et al.*, 2007; Lee *et al.*, 2009). Another PHD containing protein ORC1, the large subunit of the origin recognition complex, is involved in defining origins of DNA replication, and could bind to H3K4me3 with its PHD domain and regulate transcription in *Arabidopsis* (De *et al.*, 2009). Interestingly, some other studies suggested that the expression of alfalfa Aflin1 and *Arabidopsis* Alfin1-like (AL) gene, which all contained a PHD finger in their C terminus, were increased under salinity stress (Seki *et al.*, 2002; Winicov, 1993). A recent study indicated that its homolog in soybean was located in the nuclei and its level was up-regulated under salinity stress (Wei *et al.* 2009). Recently, the PHD fingers of the Alfin-like proteins from *Arabidopsis* have been

found to bind to histone posttranslational modifications H3K4me3/2 (Lee *et al.*, 2009). In this study, I find that GmPHD1 proteins may function as the 'code reader' of H3K4 methylation to regulate the gene expression and further the salinity stress response in the soybean.

4.2 Materials and methods

4.2.1 Gene cloning and plasmid construction

Synthesised soybean cDNA was gifts from Miss Fuk Ling Wong from Professor Hon-ming Lam's lab. The PCR reaction was set as follow: 2 μl template was mixed with 0.5 μl dNTP (10 mM), 0.5 μl forward primer (10 μM) and 0.5 μl reverse primer (10 μM), 2.5 μl 10 × PCR buffer, 0.3 μl high fidelity platinum Taq polymerase (invitrogen) and 1.5 μl $MgCl_2$ (25 mM). The final volume was adjusted to 25 μl by distilled milliQ water. The following primers were used in cloning the gene *GmPHD1, GmISWI1, GmISWI2, GmGNAT, GmElongin* and constructing the expression vectors:

Genes		Primers
GmPHD1	Forward	5' AGTGGATCCGAAGGAGTACCGCACCCAA 3'
	Reverse	5' TCAGTCGACCTCAAACTCTAACCCTCTTGT 3'
GmISWI1	Forward	5' AGTGGATCCCAGAAAATGAAGAAACAGAAG 3'
	Reverse	5' TCAGTCGACCTCATAAATAATCTTCGAGTATGTC 3'
GmISWI2	Forward	5' AGTGGATCCATGTTTCGTGGATATCAAT 3'
	Reverse	5' TCAGTCGACCTTATTTTCTTCTCTTCCCC 3'
GmGNAT	Forward	5' AGTGGATCCATGGCTGCAGCATCATCAA 3'
	Reverse	5' TCAGTCGACCTCACATAGTCTTTTGCTCAT 3'
GmElongin	Forward	5' AGTGGATCCATGATGAGAAGAGATCAA 3'

	Reverse	5' TCAGTCGACCCTAAAATACCCTCTTTGT 3'
GmPHD1-C	Forward	5' AGTGGATCCACATGTGGTGCTTGCGGTG 3'
	Reverse	5' TCAGTCGACCTCAAACTCTAACCCTCTTGT 3'
GmPHD1-N	Forward	5' AGTGGATCCGAAGGAGTACCGCACCCAA 3'
	Reverse	5' TCAGTCGACCTCATGCACCCTGTTCATCATCT 3'

In order to subclone these genes into the expression vector, specific restriction enzyme recognize sequences (*BamHI*: GGATCC, *SalI*: GTCGAC) were introduced into the primers. The PCR was performed in the following program:

Number of cycles	Length of time	Temperature
1 cycle	5 minutes	94℃
5 cycles	30 seconds	94℃
	30 seconds	45℃
	1 minute	72℃
25 cycles	30 seconds	94℃
	30 seconds	50℃
	1 minute	72℃
1 cycle	10 minutes	72℃

In order to construct the plasmid for recombinant expression, the PCR products were separated in 1% agarose gel and the DNA fragments were purified from the gel followed with overnight digestion with specific restriction enzymes (*BamHI* and *SalI*, New England Biolabs). Expression vector plasmid (GST: pGEX-4T-1, GE healthcare; MBP: pMAL-C2, New England Biolabs) was prepared from *E. coli* and subsequently digested with the same restriction enzymes. The isolated candidate DNA fragments

were ligated with the digested expression vector in a molar ratio of 10:1 (insert:vector) by adding 1×ligase buffer (Promega) and 3U of T4 DNA ligase (Promega). *GmPHD1* was inserted into the GST expression vectors, while other cloned genes, *GmISWI1*, *GmISWI2*, *GmGNAT*, and *GmElongin*, were ligated into the MBP expression vector.

4.2.2 Transformation and positive clone screening by PCR

After overnight ligation at 16℃, the ligated PCR products were transformed into *E. coli* (bacteria strain: DH5α). The $CaCl_2$ treated competent cells were thawed on ice and gently mixed by tapping the tube. Ten μl of the above ligation mix was added to 100μl of competent cells and the mixture was placed on ice for 20 minutes, then subjected to a heat pulse at 42℃ for 2 minutes and cooling on ice for another 2 minutes, followed by addition of 1 ml LB broth to rescue the cells. The transformed cells were incubated at 37℃ with shaking for 1 hour and then spread on LB agar plates supplemented with 100 μg/ml of ampicillin. The plates were incubated at 37℃ overnight.

The colonies from the above plates were picked by a toothpick, inoculated into a LB agar plate supplemented with 100 μg/ml of ampicillin, and the remaining cells were washed out into a PCR reaction mix with 2.5 μl of 10×PCR reaction mix, 0.5 μl of 10 mM dNTPs mix, 0.5 μl of 10 μM specific forward primer and 0.5 μl of 10 μM specific reverse primer and 0.5 μl Taq DNA polymerase (Promega). The final volume was made up to 25μl by double distilled water. It was subjected to the following PCR profile. PCR products were tested in the 1% agarose gel.

Number of cycles	Length of time	Temperature
1 cycle	5 minutes	94℃

30 cycles	30 seconds	94℃
	30 seconds	52℃
	1 minute	72℃
1 cycle	10 minutes	72℃

4.2.3 Plasmid extraction and sequencing

One positive colony was inoculated into 5 ml of LB broth supplemented with 100 μg/ml of ampicillin and grown at 37℃ overnight with shaking. The recombinant plasmid was extracted by using Wizard™ Plus Minipreps DNA purification systems (Promega). The cells were pelleted and resuspended in 200μl of Cell Resuspension Solution. Two hundreds μl of Cell Lysis Solution was added and the solution was mixed by gently inverting the tubes for several times till the suspension turned clear. Then, 200 μl of Neutralization Solution was added. The solution was inverted gently four times and centrifuged at 10,000 g for 5 minutes. The supernatant was transferred into a barrel of the Minicolumn/Syringe assembly and centrifuged for 2 minutes. One ml of Column Wash Solution was applied to wash the column twice. The Minicolumn was then transferred to a new microcentrifuge tube and 80 μl of double distilled water was added to elute the bound plasmids by centrifuging at 10,000 g for 1 minute. DNA concentration and quality were determined by spectrophotometric measurements at optical density 260nm and 280nm.

The ABI PRISM™ dRhodamine Terminator Cycle Sequencing Ready Reaction kit (Perkin Elmer 402078) was used to make single-stranded DNA by PCR for sequencing. Two μl of 3 M sodium acetate, pH5.2 and 50 μl 95% ethanol were added to the PCR products. The mixture was kept on ice overnight and centrifuged at 14,000g for 30 minutes then. The DNA pellet was washed in 70% ethanol. The washed and air-dried pellet was then resuspended in 15 μl Template Suppression

Reagent (Perkin Elmer), denatured at 95℃ for 2 minutes and placed on ice immediately. The sample was then applied to the Genetic Analyzer ABI prism 310 to resolve the cycle sequencing product. Raw data of the sequencing reaction were collected by ABI PRISM 310 Genetic Analyzer Data Collection software and analyzed by ABI PRISM 310 Genetic Analyzer Sequencing Analysis software.

4.2.4 Recombinant protein expression in *E. coli*

Plasmids with the correct sequence were transformed into the expression host bacterial cells (bacteria strain: DE3). The transformed bacteria were then inoculated into LB broth supplemented with 100 μg/ml of ampicillin and grown at 37℃ for 2.5-3 hours when the OD reached about 0.6-0.8. IPTG was added to get a final concentration of 1 mmol/l to induce the expression of the recombinant proteins at 25℃. After overnight expression, bacteria were collected, suspended in PBS and lysised with 1 mg/ml lysosome by incubating the cells on ice for at least 1 hour. Supernatant were collected after centrifugation at 4℃ for 15 minutes at maximum speed and stored at -80℃ until use.

4.2.5 Soybean nucleic proteins extraction

Soybean nuclei were isolated from leaves as mentioned in Chapter 3. Isolated nuclei were swelled in low salt buffer (20 mM Tris-HCl, pH 7.6, 10 mM KCl, 2.5 mM $MgCl_2$, 2 mM DTT and 0.5mM PMSF), and total nuclear proteins were then extracted by high salt extraction buffer (500 mM NaCl, 25% glycerol in low salt buffer) (Wysocka, 2006). The concentration of the NaCl in the extracted proteins was diluted to 250 mM before use.

4.2.6 Isolation and identification of the interaction proteins of GmPHD1

Recombinant GST fusion protein GST-PHD1 was first bound to the GST column (GE health) by incubating the protein with GST agarose beads at room temperature for 30 minutes. Following that, the extracted nucleic proteins were applied to the beads and incubated at 4℃ overnight. The beads were then washed 10 times with wash buffer (25 mM Tris-HCl (pH 8.0), 10% glycerol, 1 mM EDTA, 200 mM NaCl, 1 mM PMSF, 1 mM DTT, 0.1% Triton X-100) and subsequently boiled with SDS-PAGE gel loading buffer at 99℃ for 10 minutes before applying for SDS-PAGE gel separation. SDS-PAGE gel was stained with silver and the different protein bands were excised and destained as described in Chapter 3. The proteins were identified by MALDI-TOF/TOF as aforementioned.

4.2.7 *In vitro* GST pull down assay

Recombinant GST-PHD1 protein was incubated with MBP-ISWI, MBP-ISWI2, MBP-GNAT, MBP-elongin in the GST column at 4℃ overnight, respectively. Then the individual column was washed extensively with buffer containing 25 mM Tris buffer (pH 8.0), 10% glycerol, 1mM EDTA, 500 mM NaCl, 1 mM PMSF, 1 mM DTT, 1% Triton X-100 twice, followed with another 6 washes with buffer containing 25mM Tris buffer (pH 8.0), 10% glycerol, 1 mM EDTA, 150 mM NaCl, 1 mM PMSF, 1 mM DTT, 0.1% Triton X-100. Finally, the beads were recovered from each of the column and they were boiled separately with SDS page gel loading buffer at 99℃ for 10 minutes and the western blotting was performed as aforementioned with anti-MBP antibody.

4.2.8 *In vitro* acetyltransferase activity assay

MBP-GNAT protein was mixed with 125 μM Acetyl-Coenzyme A (GE health), 60 μg histone extracted from soybean or other tested proteins, 1.5 mM DTT, 10%

glycerol, 0.15 mM EDTA, 15 mM sodium butyl, 15 mM nicotiamide, 1 mM PMSF, 1 mM protease inhibitor and then incubated at 30℃ overnight. The reaction was then concentrated and protein acetylation was tested by western blotting with the anti-acetyl-K antibody (Millipore, 05-515).

4.2.9 Chromatin immunoprecipitation

ChIP was performed using the chromatin immunoprecipitation (ChIP) assay kit (Millipore). Briefly, the soybean leaves were first fixed in 1% formaldehyde for 15 minutes. The fixation was then terminated by adding glycine to a final concentration of 125 mM. Nuclei were then extracted from the fixed leaves and resuspend in SDS lysis buffer and incubated for 10 minutes on ice. The lysate were sonicated to shear the genome DNA to lengths between 200-1000 bp. After sonication, the samples were centrifuged for 10 minutes at maximum speed at 4℃. The supernatant was then diluted 10 folds with ChIP dilution buffer and 1% of the sample were aliquot as input sample. The diluted samples were subsequently pre-cleared with protein A agarose/salmon sperm DNA (50% slurry) for 1 hour at 4℃ with agitation. Immunoprecipitating antibody was then added into the pre-cleared sample and incubated overnight at 4℃ with rotation. Then protein A agarose/salmon sperm DNA (50% slurry) was used to precipitate the antibody/protein/DNA complex, following washed with low salt wash buffer, high salt wash buffer, LiCl wash buffer and TE buffer sequentially. Bound protein/DNA complex was then eluted from the beads with freshly prepared elution buffer (1% SDS, 0.1 M $NaHCO_3$). To reverse the protein-DNA crosslinks, 5 M NaCl was applied to the elution sample to get a final concentration of 200 mM and heating at 65℃ for over 4 hours. The DNA was finally recovered from the sample by phenol/chloroform extraction and ethanol precipitation. The primers for PCR were as follow: *HML806*-P1:

5'CGTTTAACTGTTTAAGGAA3' and 5'AAATGGGTAGGAGACGAT3'; *HML806*-P2: 5'GGAAAAGAAGAAAGCCACACTCTGA3' and 5' AATTGGACATTGATCGATTGATGA 3'; *HML806*-EX: 5' AGGCCAGGTGCTGCATAATCT 3' and 5' ATGGATTGCCACCAGTGCAA 3'; *HML806*-L: 5' AGTCCTTCTTTCCTGTTCT 3' and 5' TACTATCTACAACGATTTAC 3';

HML1107-P: 5'TTAGGGCGGTGTTACTC3' and 5'CGTATTATCGCTCTTCTT3'.

ChIP-PCR reaction was set up as followed: 4 μl template was mixed with 0.4 μl dNTP (10 mM), 0.4 μl forward primer (10 μM) and 0.4 μl reverse primer (10 μM), 2 μl 10×PCR buffer, 0.25 μl Taq polymerase (Promega) and 1 μl $MgCl_2$ (25mM). The final volume was adjusted to 20 μl by distilled milliQ water. PCR was carried out using the following parameter:

Number of cycles	Length of time	Temperature
1 cycle	5 minutes	94℃
38 cycles	30 seconds	94℃
	30 seconds	42℃
	1 minute	72℃
1 cycle	10 minutes	72℃

The PCR product was then tested in 2% agarose gel.

4.2.10 Peptide synthesis and antibody production

Peptides (GKNERKRLFQMINDLPT and TPAKAEHIKQYK from GmPHD1, GEEATAELDAKMKKFTEDAIK from GmISWI) were synthesized using the standard procedures of F-moc solid-phase peptide synthesis protocol on the Applied Biosystems 433A solid-phase peptide synthesizer. The synthesized peptides were dissolved in milli-Q water, and then purified by standard reversed-phase HPLC. The

homogeneity of the purified peptides was determined by MALDI-TOF mass spectrometry (ABI, 4700 proteomics analyzer).

The purified peptides were conjugated to KLH (Keyhole Limpet Hemocyanin, Sigma, H8283). About 1 mg purified peptides were mixed with 1 ml KLH solution and the mixture was sonicated for 30 minutes. NHS (5 mg/ml, Pierce) and EDAC (10 mg/ml, Sigma) were added into the mixture which was stirred for another 30 minutes. Meanwhile, G-25 Sephadex desalting column (PD-10, GE healthcare) was equilibrated with 3 volumes of PBS. Then peptide-KLH solution was loaded onto the desalting column and eluted with PBS. Cloudy elution was collected and kept at -20℃ until used.

Equal amount of complete Freunds adjuvant (Sigma) was mixed with purified peptide-KLH solution (contain about 100 μg peptide). The mixture was then emulsified manually until the emulsion becomes very thick and did not disperse when a drop of it was placed on the surface of a saline solution. The 6-8 weeks old rabbits were immunized with these emulsions subcutaneously. After the priming immunization, rabbits were boost with 100μg antigen emulsified in incomplete Freunds adjuvant (Sigma) (1:1) for 3 times at 2 weeks intervals. Finally, the serum was collected and tested with western blotting. Control serum was collected before the priming immunization. All the rabbits were raised in the animal centre of The Chinese University of Hong Kong according to the animal ethics.

4.2.11 Peptide pull down assay

Biotin-conjugated peptides containing H3K4 mono-, di- and tri-methylation were purchased [Millipore, catalog number: 12-563 (mono-), 12-460(di-), and 12-564(tri-)]. Biotin conjugated peptides containing H3K9 trimethylation (Millipore, 12-568) were used as control. They were immobilized onto the avidin agarose beads

(Pierce, 20219). Recombinant GST-GmPHD1 was then incubated with these beads at 4℃ overnight. The beads were then washed with buffer containing 25 mM Tris buffer (pH 8.0), 10% glycerol, 1 mM EDTA, 500 mM NaCl, 1 mM PMSF, 1 mM DTT, 1% Triton X-100 twice, followed with another 6 washes with buffer containing 25 mM Tris buffer (pH 8.0), 10% glycerol, 1 mM EDTA, 150 mM NaCl, 1 mM PMSF, 1 mM DTT, 0.1% Triton X-100. Finally, the beads were boiled with SDS-PAGE gel loading buffer at 99℃ for 10 minutes and western blotting was performed as aformentioned with anti-GST antibody (Sigma, G7781).

4.3 Results

4.3.1 *GmPHD1* is a PHD finger domain containing protein

The soybean *GmPHD1* was cloned from the soybean leaves cDNA, the full length of which is 756bp, encoding a protein containing 251 amino acid (Figure 4.1A). SMART analysis (http://smart.embl-heidelberg.de/) showed that it contained a PHD finger domain in its C termini which had the typical character of the PHD finger domain, C4HC3 (Figure 4.1B)

More interestingly, sequence alignment analysis showed that the PHD finger domain of the GmPHD1 also contained the aromatic amino acids which have been demonstrated to be conserved and important for the PHD finger to recognize histone H3K4 methylation by forming a pocket (Figure 4.1C). Another pocket which was composed by the negative charge containing amino acid, such as aspartic acid and glutamic acid, and of importance in holding the H3R2 methylation was also found in the GmPHD1 (Figure 4.1C). Therefore, it was likely that the GmPHD1 might also be able to recognize the H3K4 methylation and H3R2 methylation as most of the PHD finger containing proteins.

A

```
ATGGAAGGAGTACCGCACCCAATACCCAGAACTGTCGAAGAGGTTTTCACCGATTTTAAG
 M  E  G  V  P  H  P  I  P  R  T  V  E  E  V  F  T  D  F  K
GGCAGACGCGCTGGTTTGATTAAGGCCCTCACTACTGACGTTGAAAAGTTTTACCAGCAG
 G  R  R  A  G  L  I  K  A  L  T  T  D  V  E  K  F  Y  Q  Q
TGCGATCCCGAGAAGGAGAATTTGTGTCTATATGGGTTTCCAAATGAAACATGGGAAGTG
 C  D  P  E  K  E  N  L  C  L  Y  G  F  P  N  E  T  W  E  V
AATTTGCCTGTTGAGGAAGTGCCTCCTGAACTTCCTGAGCCAGCATTAGGTATAAACTTT
 N  L  P  V  E  E  V  P  P  E  L  P  E  P  A  L  G  I  N  F
GCCAGGGACGGCATGCAAGAGAAGGACTGGCTATCACTGGTTGCAGTTCACAGTGACTCA
 A  R  D  G  M  Q  E  K  D  W  L  S  L  V  A  V  H  S  D  S
TGGCTGCTTGCTGTTGCTTTCTATTTTGGTGCCCGCTTTGGATTTGGTAGGAATGAAAGG
 W  L  L  A  V  A  F  Y  F  G  A  R  F  G  F  G  R  N  E  R
AAAAGGCTTTTTCAGATGATAAATGATCTGCCGACAATCTTCGAACTTGTGACAGGAAGT
 K  R  L  F  Q  M  I  N  D  L  P  T  I  F  E  L  V  T  G  S
GCTAAGCAATTAAAGGATCAACCAGCTGCTCACAACAATGGTAGCAAATGCAAATCAAGT
 A  K  Q  L  K  D  Q  P  A  A  H  N  N  G  S  K  C  K  S  S
GGAAAGTCCCATCAGTCTGAGTCTCAGGCCAAGGGGATGAAGATGTCTGCACCACCCAAA
 G  K  S  H  Q  S  E  S  Q  A  K  G  M  K  M  S  A  P  P  K
GAAGAGGATGAGAGTGGAGAAGAAGAAGAAGATGATGAACAGGGTGCAACATGTGGTGCT
 E  E  D  E  S  G  E  E  E  E  D  D  E  Q  G  A  T  C  G  A
TGCGGTGATAATTATGGCACTGATGAATTCTGGATCTGTTGTGATATGTGTGAAAGATGG
 C  G  D  N  Y  G  T  D  E  F  W  I  C  C  D  M  C  E  R  W
TTCCATGGTAAATGTGTTAAAATTACTCCTGCTAAGGCTGAGCACATCAAGCAATACAAG
 F  H  G  K  C  V  K  I  T  P  A  K  A  E  H  I  K  Q  Y  K
TGCCCTAGCTGCAGTAACAAGAGGGTTAGAGTTTGA
 C  P  S  C  S  N  K  R  V  R  V  *
```

B

C

```
atING1      YC-ICNQVSYGE---MVACDNNACKIEWFHFGCVGLKEQPKG---KWYCPECA
msAlfin1    TCGACGDN-YGTDEFWICCD--MC--EKWFHGKCVKITPAKAEHIKQYKCPGCS
atAL6       VCGACGDN-YGTDEFWICCD--AC--EKWFHGKCVKITPAKAEHIKHYKCPTCS
hsBPTF      YC-ICKTP-YDESKFYIGCD--RC--QNWYHGRCVGILQSEAELIDEYVCPQCQ
hsING2      YC-LCNQVSYGE---MIGCDNDECPIEWFHFSCVGLNHKPKG---KWYCPKCR
GmPHD1      TCGACGDN-YGTDEFWICCD--MC--EKWFHGKCVKITPAKAEHIKQYKCPSCS
```

Figure 4.1: Soybean *GmPHD1* was a PHD finger domain containing protein. A: Nucleotide and amino acid sequences of soybean *GmPHD1*. The amino acids in the

red rectangle were the PHD finger domain. B: The GmPHD1 contained a PHD finger domain in its C termini, which has the typical C4HC3 structure, as highlighted in pink and blue rectangles. C: Alignment of the PHD finger domain of *At*ING1, *Ms*Alfin 1, *At*AL6, *Hs*BPTF, *Hs*ING2, *Gm*PHD1. The red rectangle indicated the conserved aromatic amino acids which composed the pocket recognizing H3K4 methylation. The blue rectangle indicated the conserved amino acids which composed the pocket recognizing H3R2 methylation. *At*ING1:at3g24010; *Ms*Alfin 1:AAA20093.2; *At*AL6: at2g02470; *Hs*BPTF: NP_872579.2; *Hs*ING2: NP_001555.1. *At: Arabidopsis thaliana*; *Ms: Medicago sativa; Hs: Homo sapiens; Gm: Glycine max.*

4.3.2 The expression of GmPHD1 in soybean

The synthetic peptides were used to immunize the rabbits and the antibody that could specifically recognize the GmPHD1 protein was raised. Compared with the pre-immune sera, the anti-PHD1 antibody could recognize a protein whose molecular weight was about 35 kD. Meanwhile, the recombinant proteins could also be recognized by the anti-PHD1 antibody (Figure 4.2A and 4.2B). Theoretically, the GmPHD1 was about 28 kD, which was smaller than the observed molecular weight. Considering that many proteins have been posttranslationally modified, such as glycosylation, this difference was acceptable.

With this antibody, I tried to find out the distribution patterns of the GmPHD1 in soybean with western blotting. Total proteins were extracted from leaves and roots and the results indicated that it was ubiquitously expressed in both leaves and roots (Figure 4.2C). In addition, when the soybeans were treated with salt stress, the expression of GmPHD1 would increase both in leaves and roots, as expected (Figure 4.2D) (Seki, *et al.*, 2002; Winicov, 1993).

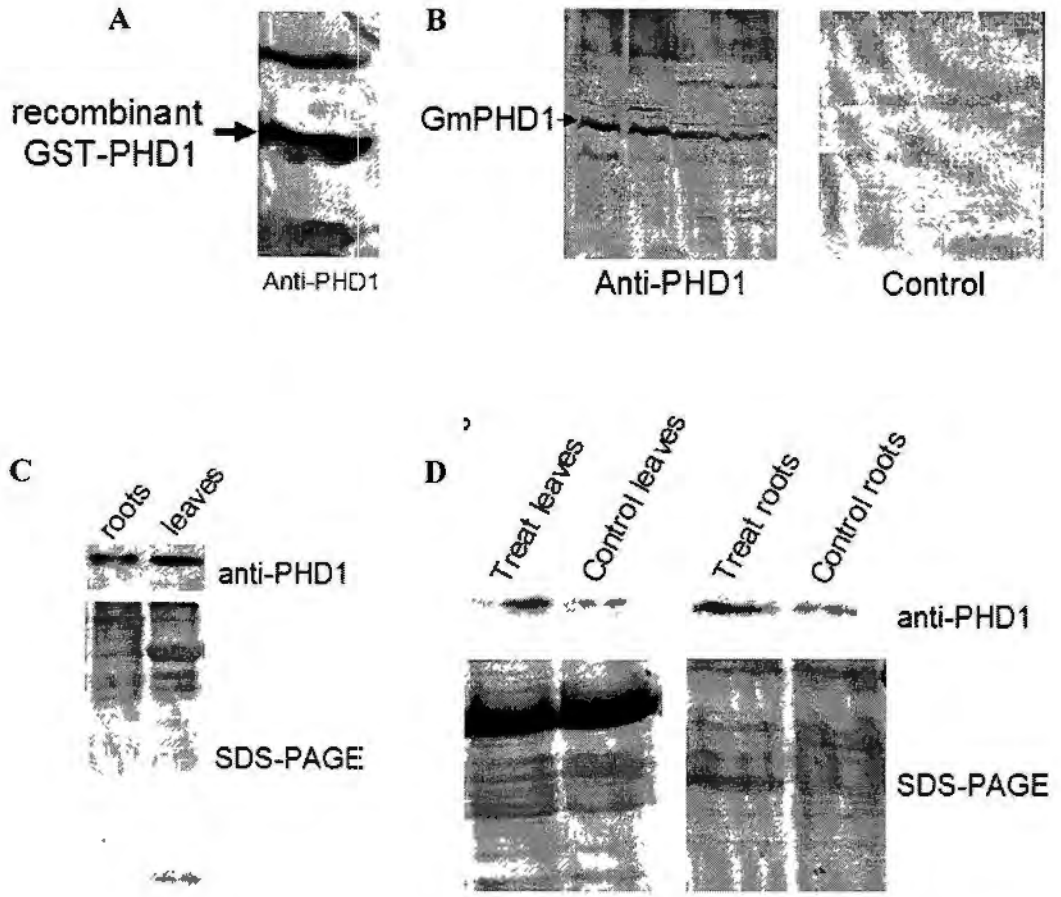
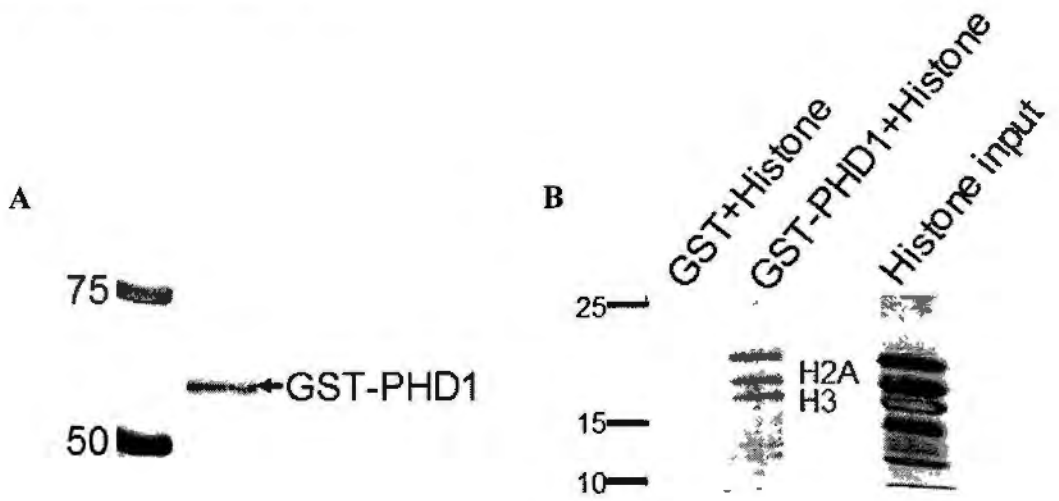


Figure 4.2: GmPHD1 was ubiquitously expressed and its expression was up-regulated by salinity stress in soybean. A: The specificity of anti-PHD1 antibody was tested with recombinant GST-PHD1 by western blotting. B: The specificity of anti-PHD1 antibody was tested with soybean total proteins by western blotting. Anti-PHD1: anti-PHD1 antibody, Control: preimmune antiserum. C: GmPHD1 was expressed in roots and leaves. Upper panel: western blotting results. Lower panel: SDS-PAGE gel image of the total proteins from soybean roots and leaves. D: GmPHD1 was upregulated by salinity stress in soybean. Upper panel: western blotting results. Lower panel: coomassie blue staining SDS-PAGE gels of soybean total proteins were showed as loading control.

4.3.3 GmPHD1 recognized histone H3K4 methylation

Sequence alignment analysis suggested that GmPHD1 might interact with histone H3K4 methylation (Figure 4.1C). For validation, I first expressed the GmPHD1 fused with GST in *E. coli* (Figure 4.3A). The fusion proteins were then incubated with histone extracted from soybean leaves to test their ability to interact with soybean histone. The results clearly demonstrated that histone H3 and H2A can be pulled down by the GST-PHD1 (Figure 4.3B) and histone H3K4 methylation was present in these pulled down histone H3 (Figure 4.3C).

Since H3K4 can be mono-, di-, and tri- methylated, I then wondered which modification was mainly recognized by GmPHD1. Similar to other PHD finger containing protein, GmPHD1 have the preference for H3K4 dimethylation in the peptide pull down assay (Figure 4.3D). However, GmPHD1 would recognize H3K4 trimethylation with very low affinity, even lower than that to H3K4 monomethylation (Figure 4.3D), which is quite different from other PHD finger containing protein such as ING protein and BPTF.



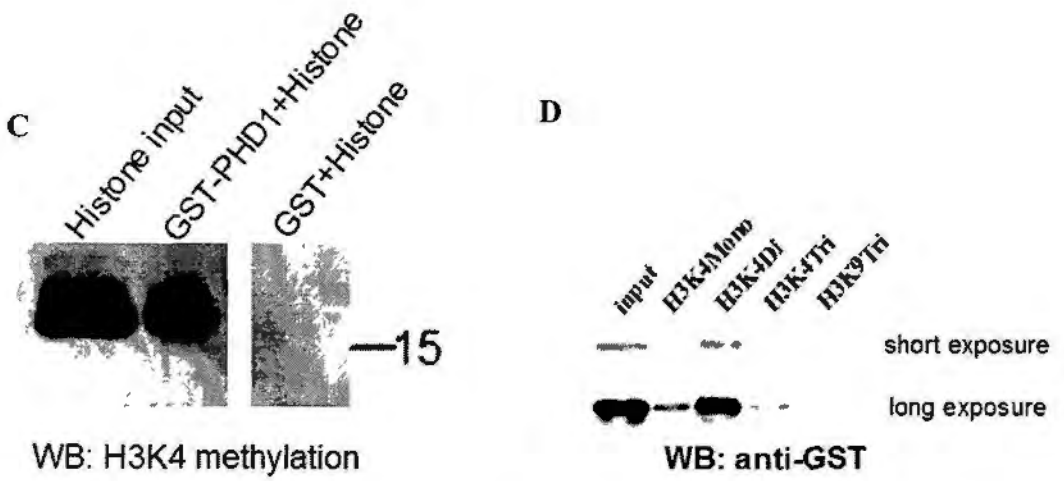


Figure 4.3: GmPHD1 interacted with histone H3 and recognized H3K4 methylation. A: GST-PHD1 fusion proteins were expressed in *E. coli* and purified. B: SDS-PAGE gel showed that histone H3 was pulled down by the GmPHD1 in the GST pull down assay. C: Western blotting showed that H3K4 methylation was present in the histone H3 pulled down by GmPHD1. D: peptide pull down assay indicated that GmPHD1 recognized histone H3K4 methylation with the preference to H3K4 dimethylation.

4.3.4 GmPHD1 located in the promoter region and the body of some salt stress inducible genes

Since GmPHD1 can interact with H3K4 methylation, which was widely distributed along transcriptionally active genes, I then curious about the distribution patterns of GmPHD1 in the soybean genome.

ChIP was then performed with GmPHD1 antibody and the location of GmPHD1 along the *HML806,* a soybean salt stress inducible gene, was then selected for further analysis. The PCR product of *HML806*-P1 located in the far promoter region (from -886bp to -685bp upstream of the start codon); while the PCR product of *HML806*-P2 was in the near promoter region (-188bp upstream of the start codon and contained partial of the first exon of the *HML806*, including the start codon). The

PCR product using primer *HML806*-EX contained partial of intron 1 and exon 2 and the set of primer *HML806*-L was in the 3' UTR of the genes, +3582 downstream of the start codon. ChIP-PCR clearly showed that GmPHD1 was mainly located in the near promoter region (primer *HML806*-P2) and the body of the genes (primer *HML806*-EX), while in the far promoter region (primer *HML806*-P1) and the 3'UTR (primer *HML806*-L), the GmPHD1 was hard to be detected (Figure 4.4A). In another salt stress inducible gene, *HML1107,* the GmPHD1 was also detected to be located in the near promoter region (Figure 4.4B). Meanwhile, the signal in the *Actin* was very weak. These results indicated that the GmPHD1 was not uniformly distributed in the genome; instead it might localize in certain kinds of genes and regulate their transcription, such as some salt stress inducible genes.

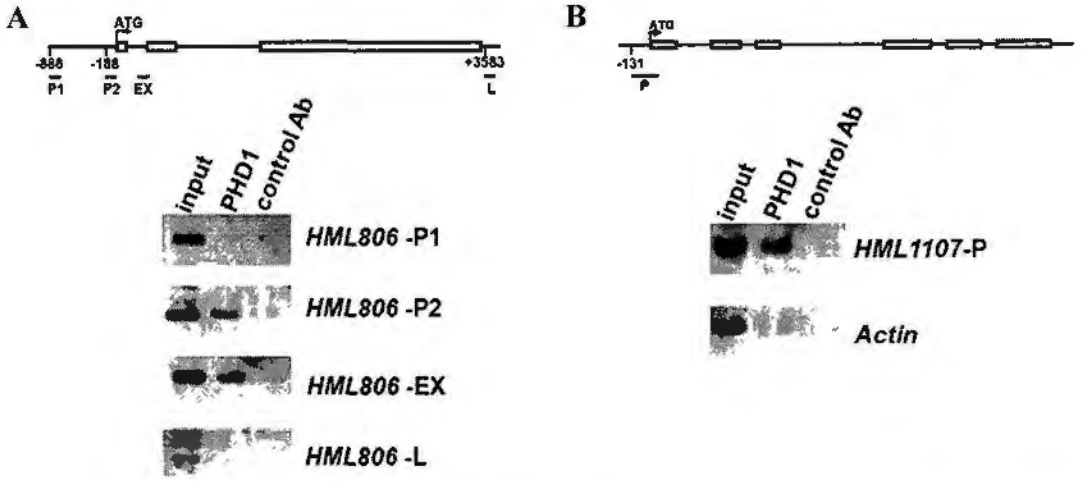


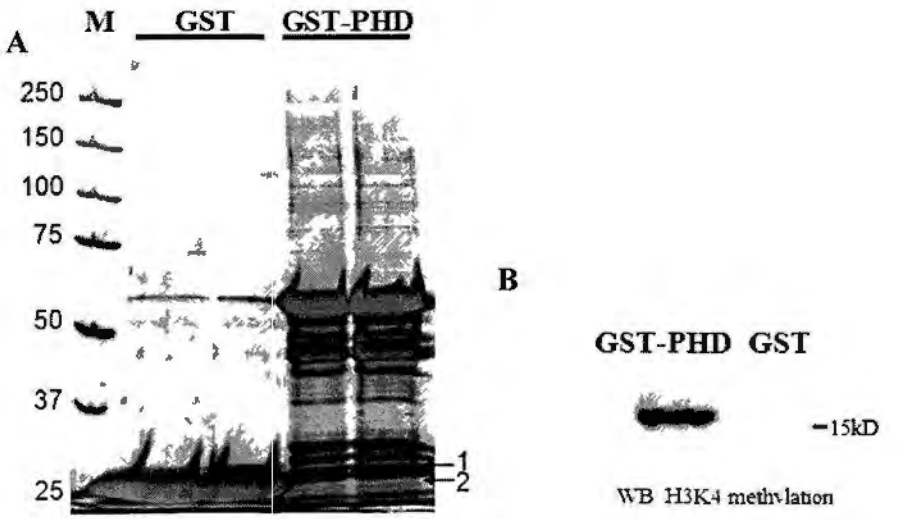
Figure 4.4: GmPHD1 located in the promoter and body of some salt stress inducible genes. A: ChIP results showed that GmPHD1 was mainly located in the near promoter region (*HML806*-P2) and the body (*HML806*-EX) of *HML806* while low abundance of GmPHD1 was located in the far promoter (*HML806*-P1) and 3'UTR (*HML806*-L) of *HML806*. B: GmPHD1 also located in the near promoter of another salt stress inducible gene, *HML1107* (*HML1107*-P). A very small amount of

GmPHD1 located in the body of *actin*. Control Ab: preimmune antiserum.

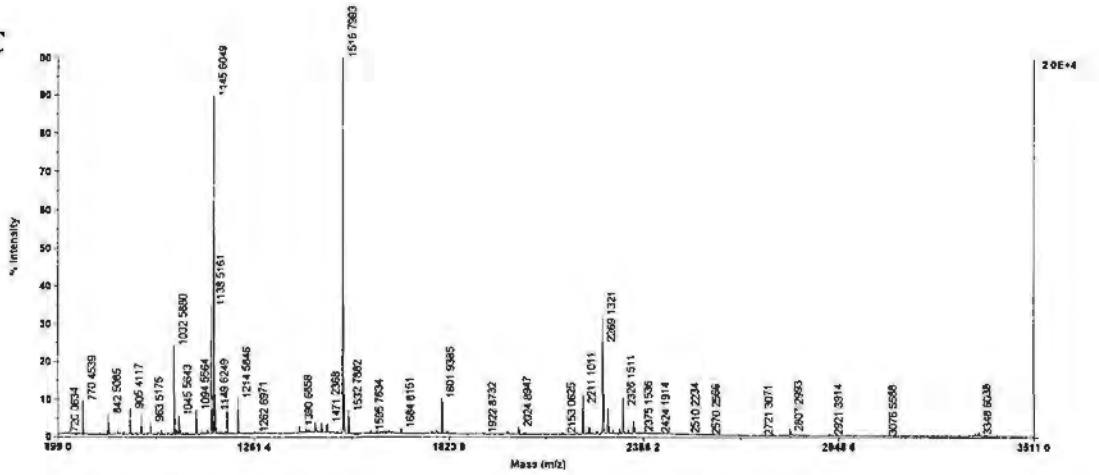
4.3.5 Identification of proteins that interacted with GmPHD1 in soybean

The following question intrigued me most was that whether and how the GmPHD1 can regulate the gene transcription. Obviously, there was no known domain or structure in the GmPHD1 implying that they could regulate gene transcription by themselves. Therefore, it was very likely that they would work together with other proteins.

I incubated the recombinant GST-GmPHD1 proteins with extracted soybean nucleic proteins and found that several proteins could be pulled down in this way (Figure 4.5A). Western blotting with anti-H3K4 methylation showed that histone H3 was also pulled down by the recombinant GmPHD1 (Figure 4.5B), validating the conclusion that GmPHD1 could recognize histone H3K4 methylation and suggesting that some interaction patterns were indeed pulled down in this experiment. Therefore, I tried to identify the pulled down proteins by mass spectrometry, with two of them were successfully determined, which were elongin A and GNAT (GCN5-related N-acetyltransferase (GNAT) family protein) (Figure 4.5C and 4.5D).



C



>GNAT

MAAASSTLFSFSFPFDPQIQTLSFPLSSAKPRLSPKPFSTFKLTSTHSS

SSSSLTVSDDPFNTGRFLSNDELRRLRLLETFLYRCDLPSGGSLSVRVM

RPHETDPTVLLLAASFAESMLLPQPYVKFLAFLVKQYLLDRRSLMPHTA

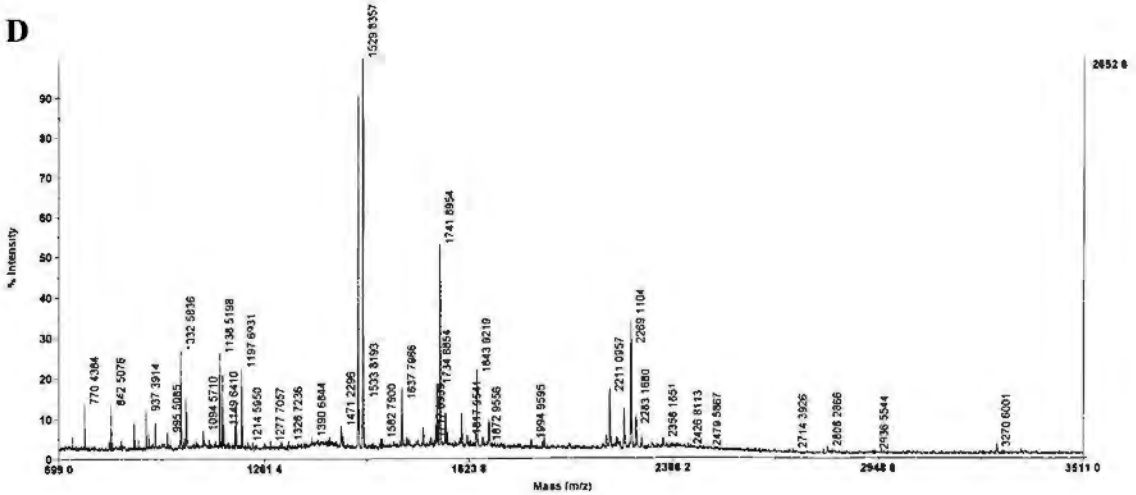
TLVAFYTQTAAAAADQDQEQEQEARLAGTVELSFDIRGANATVPSPTPP

RDKPYICNMAVRKSLRRRGIGWHLLRASEELISQMSSAREVYLHCRIID

EAPFNMYTKADYKIVKTDSILVLLTLQRRKHLMCKKLPLLSTPPETDLS

VSDEQKTM

D



>Elongin A

MMRRDQIRAREMRTPSLVDLCVQKVIDNVRYLGNVGSLDQHLLEQILPH

CTADQLMHVEKSTKGRNLSPVTDKLWKKFYERQFGTNNTNEVIKRMKEK

GVNFRWMQLYEAKGKERAQAENEALDRIRQLYKKEDARKQSRQVRTCTK

VPPSSKRRFWGDNGPGYNVSNVKSNIMKKAKIEFLKSHEVKNLAAMKNK

SIQRNPSSSSFKKTGSISGIGSTSKDPKPTKRVF

Figure 4.5: Identification of GmPHD1 interaction proteins. A: silver staining gel of GST-PHD1 pulled down proteins. Proteins only present in the GST-PHD1 pulled down samples were picked out for mass spectrometry analysis. Proteins with confident identifications were indicated in the gel (band 1 and 2). B: histone H3 were pulled down by GST-PHD1 in this experiment as determined by western blotting. C: protein 1 was identified as GNAT by mass spectrometry. Upper panel: the mass spectrum of protein 1. Middle panel: the amino sequence of the GmGNAT. The peptide highlighted in red was the detected peptides by mass spectrometry. Lower panel: the structure of the GmGNAT. D: protein 2 was identified as Elongin A by mass spectrometry. Upper panel: the mass spectrum of protein 2. Middle panel: the amino sequence of the GmElongin A. The peptide highlighted in red was the detected peptides by mass spectrometry. Lower panel: the structure of the GmElongin A.

Next, I tried to validate the interaction of GmPHD1 with GmGNAT and GmElongin A. Both genes (GmGNAT and GmElongin A) were firstly cloned from soybean. Luckily, two isoforms of GmGNAT were cloned in this process, which were named GmGNAT1 and GmGNAT2, separately (GmGNAT1 is the originally identified protein in the mass spectrometry). These two isoforms displayed 89% identities in their nucleotide sequences and 87% identities in their amino acid sequences (Figure

4.6). An acetyltransferase domain was present in these proteins (Figure 4.5C and 4.6). The Gmelongin A was a subunit of RNA polymerase II transcription factor SIII (Elongin) with a characteristic structure in its N termini (Figure 4.5D).

These two genes were subsequently cloned into the MBP vector for recombinant expression (Figure 4.7A) and these expressed MBP fusion proteins were then incubated with GST-PHD1 for GST pull down assay. These assays definitely supported our previous mass spectrometry results, as MBP-GNAT1 and MBP-elongin A can be pulled down by GST-PHD1 (Figure 4.7B).

The GmPHD1 was further investigated to find out which part of it was responsible for its interaction with GmGNAT1 and GmElongin A. The N termini without the PHD finger domain and the C termini which contained only the PHD finger was fused with GST and expressed (Figure 4.7C). Similarly, GST pull down assays were performed and it showed that the N termini of GmPHD1 were enough for its interaction with GmGNAT1 although the PHD finger had some weak interactions with GmGNAT1 (Figure 4.7D). However, the truncated GmPHD1 would severely impair its interaction with GmElongin A (Figure 4.7D), indicating the importance of the full length of GmPHD1 in its interaction with GmElongin A.

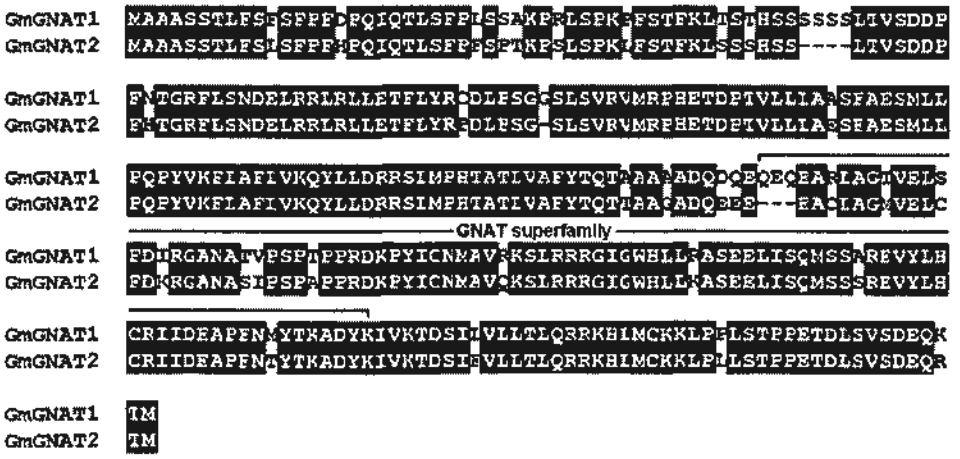


Figure 4.6: Alignment of the two GmGNATs of soybean. These two isoforms of

GmGNAT displayed 89% identities in their nucleotide sequences and 87% identities in their amino acid sequences. The GCN5-related N-acetyltransferase domain (GNAT superfamily) was indicated in this figure.

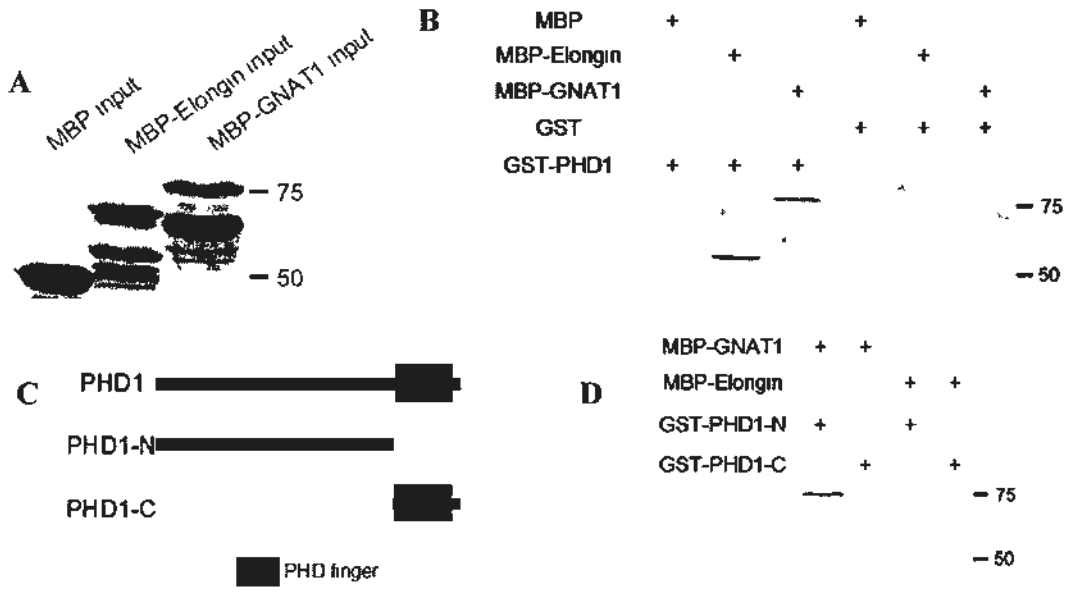


Figure 4.7: Validation of the interaction between GmPHD1 and GmGNAT1, GmElongin A by GST pull down assay. A: Inputs of the GST pull down assay. B: Western blotting results of the GST pull down assay. C: The diagram of the construction of the truncated GmPHD1. The N termini of the GmPHD1 without its PHD finger domain and the C termini of the GmPHD1 with only the PHD finger domain were inserted into the GST expression vector. D: GST pull down assay with the truncated GmPHD1.

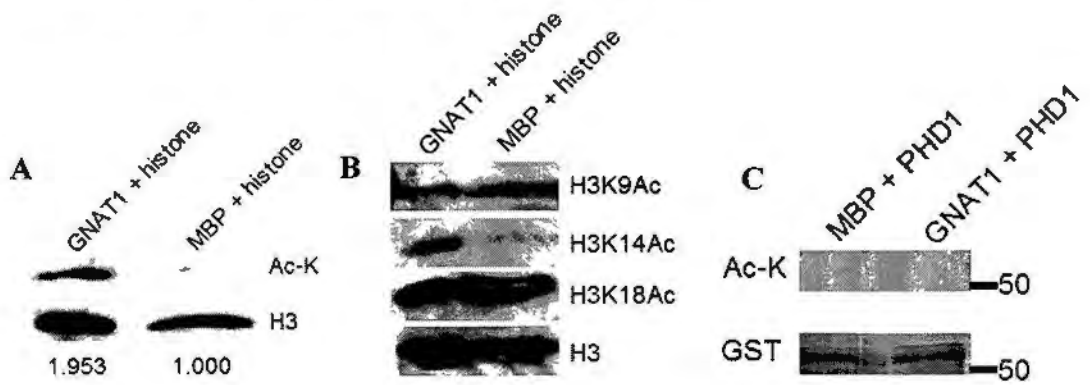
4.3.6 GmGNAT1 is an acetyltransferase

The presence of the acetyltransferase domain in the GmGNAT1 suggested that it might transfer the acetyl group to its substrates from acetyl-CoA. However, its substrates remained elusive.

As the GmGNAT1 could interact with GmPHD1 and might be recruited to the

histone H3 by GmPHD1, I then tested whether GmGNAT1 could acetylate GmPHD1 and histone H3. With the antibody which could specifically recognize acetylated lysine, I found that acetylation in the extracted soybean histone H3 increased after the *in vitro* acetyltransferase assays with GmGNAT1 (Figure 4.8A). Since several lysines in the histone H3 could be acetylated, including lysine 9, 14 and 18, I then used the antibody which specifically recognized these acetylated sites to detect if the GmGNAT1 could specifically modify certain sites in histone H3. While the acetylation at histone H3 lysine 9 and lysine 18 did not change much, the H3K14 acetylation increased significantly after the acetyltransferase assay (Figure 4.8B). However, no obvious acetylation signal were observed in the GmPHD1 with similar assays (Figure 4.8C), stating that GmPHD1 might not be the substrates of GmGNAT1 although they could interact with each other directly.

More interestingly, I also found that GmGNAT1 could be self acetylated, since the acetylation signal can be detected only when the GmGNAT1 was present (Figure 4.8D). This result thus intrigued me to study the roles of GmGNAT1 self-acetylation. Although so far I have not identified which amino acid of the GmGNAT1 was acetylated by itself, the acetyl group in the GmGNAT1 would impair the interaction between GmGNAT1 and GmPHD1 (Figure 4.8E).



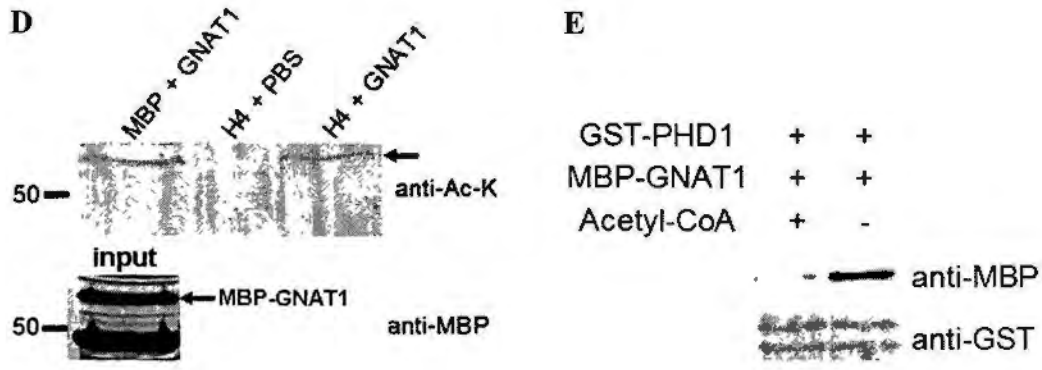


Figure 4.8: GmGNAT1 acetylated histone H3 and itself. A: *In vitro* acetyltransferase assay indicated that GmGNAT1 acetylated histone H3. B: GmGNAT1 acetylated histone H3 mainly at histone H3K14. C: GmGNAT1 could not acetylate GmPHD1. D: GmGNAT1 was self-acetylated. E: GmGNAT1 self-acetylation inhibited its interaction with GmPHD1.

4.3.7 GmPHD1 also interacted with GmISWI

ISWI (imitation switch) is a much conserved protein from yeast to mammal (He *et al.*, 2008). It contains several domains such as DEXDx, HELICs and SANT domain and functions in remodelling the chromatin structure by hydrolysis ATP. Previous reports showed that some PHD finger containing proteins, such as ING1 and ING2, could interact with these proteins to facilitate gene transcription (Ruthenburg *et al.*, 2007). Therefore, I conceived that GmPHD1 could also recruit GmISWI to remodel chromatin structure. I then cloned the GmISWI and expressed them in *E. coli* as the MBP fusion proteins. Since the full length of GmISWI contained 972 amino acids and the full protein was not suitable for expression in *E. coli*, it was divided into two portions, GmISWI1 which contained the first DEXDx domain, and GmISWI2 which contained the rest of GmISWI. Similarly, I found that MBP-ISWI1 could be pulled down by the GST-GmPHD1 in the GST pull down assay, while GmISWI2 was not able to be pulled down (Figure 4.9B). Therefore, GmISWI might interact with

GmPHD1 through its N termini.

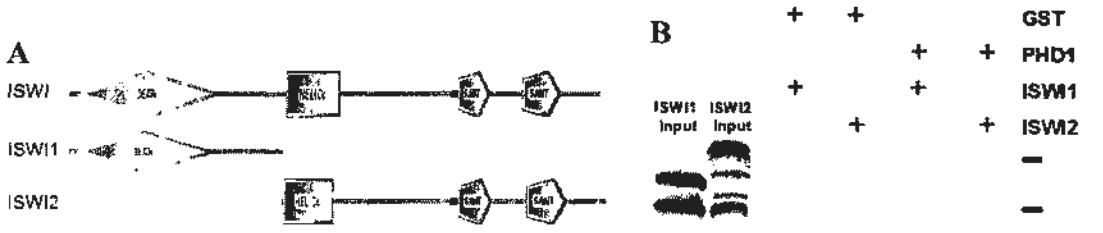


Figure 4.9: GmPHD1 could interact with GmISWI. A: The structure of GmISWI and the two constructed vectors which were expressed in *E. coli*. B: GST pull down assay indicated that GmISWI interacted with GmPHD1 through its N termini.

4.4 Discussions

I confirmed that GmPHD1 was up-regulated under the salinity stress with western blotting, as expected. Wei *et al.* also showed that its expression was induced to a higher intensity in the drought- and salt-tolerant soybean variety than that in the sensitive ones under the stresses, further correlating this protein with stress response and tolerance in soybean (Wei *et al.*, 2009). Actually, there are six homologs of PHD1 in soybean and their response to abiotic stresses, such as salinity, cold and drought stress, were different, suggesting that although these six GmPHDs were highly conserved, their activities might not be the same (Wei *et al.*, 2009). Whether other GmPHDs had the similar working mechanisms or would work as the ING family members by recruiting different protein complexes is still unknown and more efforts are required.

GmPHD1 could interact with H3K4 methylation and show some similarity to Alfin1 in alfalfa and Alfin1-like protein in *Arabidopsis* in their amino acid sequences. Alfin1 was characterized as a transcriptional factor that could bind to the promoter of salt-inducible MsPR2 and enhance its expression at the transcriptional level in alfalfa

roots (Bastola *et al.*, 1998). Recently, soybean PHD type transcription factor was also reported to be able to bind to the *cis*-element "GTGGAG" directly (Wei *et al.*, 2009). On this point, the interaction between GmPHD1 and histone H3K4 methylation might not play the essential roles in initiating the recruitment of GmPHD1 to its target DNA region. Therefore, the major roles of their interaction might stabilize or enhance the interaction between DNA/chromatin and GmPHD1.

Many studies showed that the PHD finger domain managed to distinguish the state of lysine methylation, for example, BPTF, ING superfamily members and RAG2 mainly recognized histone H3K4 di- and tri-methylation, while DNMT3L and BHC80 bound to H3K4me0 (Baker *et al.*, 2008). However, none of the paper has stated that the PHD finger proteins were able to distinguish H3K4 dimethylation and trimethylation. In our studies, I found that GmPHD1 has the affinity to histone H3K4 methylation in the order of dimethylation > monomethylation > trimethylation. How is able for GmPHD1 to distinguish the minute difference between dimethylation and other methylation states remains an interesting question and need further structure analysis. Interestingly, some other proteins without the PHD finger domain showed the preference to H3K4 dimethylation, rather than trimethylation. The WDR5, which contained WD40 repeats, would associate with H3K4 dimethylation through the WD40 repeats (Wysocka *et al.*, 2005). Some histone demethylases also conferred the similar selectivity for distinct methylation states. Histone demethylase JMJD2A only functioned as trimethylation-specific demethylases, but not be able to remove the methyl group from di- and mono-methylation (Whetstine *et al.*, 2006). Structural analysis then revealed that its specificity may be dependent on the CH···O hydrogen bonding between the lysyl ζ-methyl groups and the oxygens in the active site in its methylammonium-binding pocket. Amino acid substitution which would disrupt this

bond would increase the JMJD2A's preference to dimethylation (Couture *et al.*, 2007).

My present studies have identified different proteins that can interact with GmPHD1, including GmGNAT, GmElonging A and GmISWI.

GmGNAT belongs to the GNAT family, which catalyzes the transfer of an acetyl group from acetyl coenzyme A to a primary amine. The most well studied histone acetyltransferase in the GNAT family is the GCN5. Yeast GCN5 (yGCN5) regioselectively transferred acetyl group to K14 of histone H3 and to K8 and 16 of histone H4 (Vetting *et al.*, 2005). In addition, they interacted with enhancer-binding factors and linked the upstream activating sequence (UAS) and the basal transcription machinery in yeast to facilitate gene transcripiton. My results showed that GmGNAT1 was an acetyltransferase with the ability to acetylate histone H3, mainly at the H3 lysine 14. Many studies have suggested that histone acetylation is an integral part of transcriptional regulatory systems. They can neutralize the positive charge of the histone and attenuate the DNA-histone contacts, eventually loose the structure of chromatin and induce gene transcription; at the same time, they can affect the interaction of the amino-terminal tails with non-histone chromatin proteins, such as those bormodomain containg proteins (Yang, 2004; Ruthenburg *et al.*, 2007; Berger, 2002; Brownell and Allis, 1996; Benhamed *et al.*, 2006). Therefore, the GmGNAT1 may modulate the chromatin structure and induce gene transcription via histone acetylation.

In addition to the N-acetyltransferase domain which serves as the catalytic domain, GCN5 contains a bromodomain, which is able to tether them to specific chromosomal sites directly by binding to the acetylated lysine as other transcription

factors that contain separable activation and sequence-specific DNA-binding domains (Brownell and Allis, 1996). As for the GmGNAT1, since its lack of the bromodomain, it may not be able to bind to the chromatin directly just like GCN5. Therefore, its interaction with GmPHD1 which can directly bind to the histone and/or DNA would be very crucial for its location to the appropriate sites in the chromatin. Actually, GCN5 also exists as a component of large complexes, for example, in yeast, Gcn5p existed in a heteromeric complex with at least two additional partners, ADA2 and ADA3. Gcn5p were then recruited to a specific promoter or chromatin domain by these trans-acting factors (Brownell and Allis, 1996).

Many reports demonstrate that there are some crosstalks between different histone modifications, suggesting that one histone modification promotes or inhibits the generation of another one (Suganuma and Workman, 2008). Phosphorylation of serine 10 on histone H3 by the Snf1 kinase promoted the acetylation of H3 lysine 14 by the Gcn5 acetyltransferase. Ubiquitination of H2B lysine 123 was a prerequisite for H3 lysine 4 methylation (Sun and Allis, 2002). H3K4 methylation, especially dimethylation and trimethylation, have long been viewed as a marker of actively transcription genes and correlated with histone acetylation (Schübeler *et al.*, 2004; Li *et al.*, 2007; Ruthenburg *et al.*, 2007). Actually, several Yng1 (another PHD finger domain containing protein) containing protein complexes can acetylate histone lysine, including H3K14, and positively regulate gene transcription at different loci (Rando, 2007; Taverna *et al.*, 2006). My results showed that protein GmPHD1 bind to the histone H3K4 methylation, and then recruit GmGNAT1 to acetylate H3K14. Therefore, it represented for another kind of histone modification crosstalk between H3K4 methylation and H3K14 acetylaiton to cooperate in regulating gene transcription.

GmElongin A is a member of the RNA polymerase II transcription factor SIII (Elongin) subunit A. In mammals, the Elongin complex activates elongation by suppressing transient pausing of the RNA polymerase II during transcription. Elongin is a heterotrimer composed of A, B, and C subunits. Subunit A has been shown to function as the transcriptionally active component of Elongin (Aso *et al.*, 1995).

ISWI uses the energy of ATP hydrolysis to alter nucleosome position and/or structure. In human, the homologs of ISWI are SNF2h (sucrose nonfermenting 2 homologue) and SNF2L (sucrose nonfermenting 2-like), both of which can be found in several small protein complexes, such as hACF (human ATP utilizing chromatin assembly and remodeling factor, consisting of hACF1 and SNF2h), RSF (remodeling and spacing factor, consisting of Rsf1 and SNF2h), CHRAC (chromatin accessibility complex, consisting of hACF1, SNF2h, p15, and p17), WICH (Williams syndrome transcription factor-imitation switch, consisting of WSTF and SNF2h), NoRC (nucleolar remodeling complex, consisting of human TIP5 and SNF2h) and WCRF (Williams syndrome transcription factor-related chromatin remodeling factor, consisting of WCRF180 and SNF2h). Most of these complexes play important roles in the establishment and maintenance of heterochromatin and, by extension, gene repression (Barak *et al.*, 2003). More interestingly, SNF2L can also associate with BPTF (bromodomain PHD finger transcription factor) and RbAP46/48 (retinoblastoma-associated protein 48 and 46) to form the NURF (nucleosome remodeling factor) complex, which can regulate chromatin structure and induce gene transcription. Human NURF (hNURF) is enriched in brain and interacts with the promoter of some developmentally important genes to activate their expression with its nucleosome-stimulated ATPase activity in remodeling the chromatin template

(Barak *et al.*, 2003).

ChIP results suggested that GmPHD1 may be mainly located in the near promoter and the body of some genes. Some studies have showed that H3K4 methylation also mainly located in a similar region (Schübeler *et al.*, 2004; Bernstein *et al.*, 2005). This is in agreement with that GmPHD1 prefer to recognize H3K4 dimethylation. In addition to the salt inducible genes (*HML806 and HML1107*), the GmPHD1 also located in other genes, such as *actin*. However, the signal was much weaker than that in the *HML806 and HML1107*; indicating that may be less amount of the GmPHD1 was located in the *actin* body. Therefore, it is likely that GmPHD1 may be widely distributed in some salt stress inducible genes and regulate their expression. Transgenic *Arabidopsis* with GmPHD increased the expression of several proteins involved in ROS scavenging, such as monodehydroascorbate reductase, peroxidase and some other proteins in the stress signal transduction like the ABI5, an ABA-responsive basic leucine zipper transcription factor (Wei *et al.*, 2009), supporting my conception that GmPHD1 may play important roles in regulate the salt stress responsive gene expression.

At this point, it seems to be controversial when there are reports indicating that GmPHD can repress several gene transcriptions, such as some negative regulators of stress tolerance in plants (Wei *et al.*, 2009). Although previously the NURF was found to mainly involve in the transcription activation, now it is known that the Drosophila NURF complex could be recruited by transcriptional repressor Ken protein to the STAT-binding sequence overlapped by the Ken-binding sites in hemocytes, then repressed STAT responders and regulated the expression of genes involved in Drosophila innate immunity (Kwon *et al.*, 2008). Therefore, the NURF

complex could active and repress the gene transcription, depending on where the nucleosome is positioned by the complex (Kwon *et al.*, 2008). It is likely that the GmPHD-GmISWI complex will work in a similar way to both active and repress gene transcription.

In addition to histone, other non-histone proteins can also be acetylated. So far, acetylation in the non-histone proteins is reported to be able to regulate the proteins' activity in several ways. First, they can regulate the proteins localization and stability, as the observation in p53 where lysine acetylation maintains it in the nucleic and prevents it from exposure to proteosome in the cytosol. Second, lysine acetylation also induces the protein-protein interaction, for example, the acetyl group supply docking sites in p53 for recruitment of transcriptional co-activators, such as TAF1, a TFIID subunit. Third, lysine acetylation has also been found in many enzymes, including acetyl-CoA synthase, nitric oxide synthase and other metabolic enzymes and these modifications usually modulate their activities. More interestingly, several acetyltransferase themselves are auto-acetylated, such as p300 and CBP (Yang and Seto, 2008). In my studies, I also presented that GmGNAT1 themselves were auto-acetylated and this acetylation would obviously prohibit its interaction with GmPHD1. As a histone acetyltransferase which may play roles in regulating gene transcription, the auto-acetylation of GmGNAT1 might function as a brake in the gene transcription and therefore represented for a negative regulation mechanism to prevent over-response to the stresses, for example keeping a homeostasis of ROS in plant.

In conclusion, GmPHD1 might directly bind to some DNA sequence with its N termini (Wei *et al.*, 2009), and I proposed that GmPHD1 recognized the histone

H3K4 methylation, especially dimethylation, to stabilize or enhance its interaction with DNA at the same time. Subsequently, GmPHD1 was able to recruit several interaction patterns, such as GmElongin A, GmISWI, and GmGNAT1 and they might facilitate the chromatin remodeling and further recruit gene transcription machinery to induce gene transcription. Under the salinity stress, the GmPHD1 was up-regulated. Transgenic *Arabidopsis* with GmPHD1 showed greater stress tolerance than the wild type plants; likely through upregulating the ROS scavenge system (Wei *et al.*, 2009) which played important roles in the plant's response to salinity stress. Chromatin immunoprecipitation results indicated that GmPHD1 could locate in the promoter and body of several salt stress inducible genes in soybean. Self acetylation of GmGNAT1 might play some self-regulation mechanisms to minimize the overreaction of the salinity stress response by prohibiting its interaction with GmPHD1. Therefore, GmPHD1 played a rather important role in regulating the transcription of stress related genes under salinity stress and salt stress response in soybean. In this way, through the GmPHD1, I managed to correlate epigenetic (histone H3K4 methylation) to soybean salinity stress (Figure 4.10).

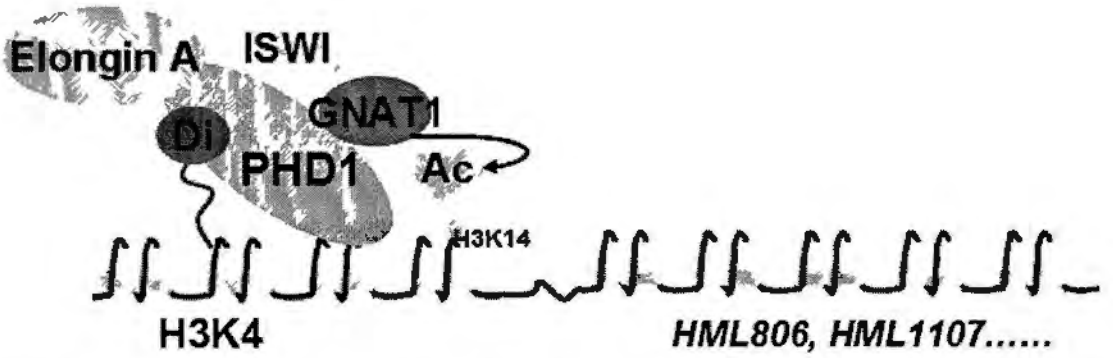


Figure 4.10: A proposed model for GmPHD1 in regulating gene expression. GmPHD1 bound to the promoter region and body of some salt stress inducible genes and recruited GmGNAT, GmElongin A and GmISWI to regulate gene transcription in

soybean.

Chapter 5 Conclusions and perspectives

Salinity stress is a kind of abiotic stress which severely affects plant growth and crop production. My present studies aimed to understand the physiological processes involved in the salinity stress and the regulation mechanisms in the salinity stress response in soybean. In addition, I believed my data may also be applicable for enhancing soybean tolerance to such abiotic stress.

Previous investigations have confirmed that plants could regulate their metabolome and transcriptome to adapt to the salinity stress. My proteomic studies in the soybean leaves showed that plant photosynthesis was severely impaired; and the chloroplasts were one of the major organelles to be damaged by the salinity stress. Some studies also proposed that salt stress could inhibit photosynthesis by reducing water potential (Parida and Das, 2005). So one efficient way to increase the soybean tolerance to salinity stress is to maintain their water potential, increase their water consumption efficiency and induce their photosynthesis under salinity.

Moreover, my comparative proteomic studies on the salt sensitive cultivated (Union) and salt tolerant wild type soybean revealed other biological processes/mechanisms that may altered in the affected plant upon salinity stress. In the present study, wild type soybean was found to be capable of regulating the expression of several salt stress response proteins to control the energy homeostasis, detoxification processes and growth even when they did not encounter the salinity stress, which was quite similar to the plants' response to salinity stress when they were treated with such stress (Urano *et al.*, 2010). This phenomenon could represent a pre-existing tolerance mechanism in the wild type soybean. Considering the long history of soil salinity even more ancient than humans and agriculture (Zhu, 2001), it is likely that the wild

type soybean achieves this ability during evolution because of the natural selection pressure resulted from the surrounding salty environment where they grow. On the other hand, some important stress tolerant genes may be lost upon the long term man-made selection during artificial breeding, which finally resulted in the genotype and phenotype differences between wild type and cultivated soybean.

Soybean is originated from China, where it has been domesticated for more than 5000 years. A large collection of soybean germplasms including wild type and cultivar are available in China, with observable differences in the salt tolerance capabilities in different varieties (Shao, *et al.*, 1993). In addition, previous genetic analysis by the cross between a salt tolerance variety (Wenfeng7) and a salt sensitive variety (Union) suggests that the salt tolerance ability in soybean is inheritable (Shao *et al.*, 1994). All these basic researches have implied that it is possible for me to obtain new varieties with high productivity and salt tolerance property by artificial crossing. My proteomic studies therefore could provide a platform and generate protein markers for high-output screening in soybean breeding.

Interestingly, proteomic comparisons between different developmental stages also reveal that proteins involved in the stress response are also regulated in the soybean root as they grow. These proteins may play important roles in plant root development, as the peroxidases which produce the ROS to regulate root growth. In addition, the plant growth may result in a protein redox pressure in the plant cell and thus need the detoxification system to maintain the homeostasis of protein redox. However, clear understanding these proteins in the plant growth needs further studies.

I present the first report of histone H3 and H4 variants and their PTMs in the legume plant soybean using nano-LC combined with mass spectrometry, mainly focusing on

the acetylation and methylation of histone H3 and H4 and their variants (Wu *et al*, 2009). Significant differences are found in histone modifications between soybean and *A. thaliana*, especially in histone H3K79 methylation and the modification patterns of the two histone H3 variants, suggesting that although the amino acid sequences of histones are conserved in evolution, their modification patterns can be quite different. The modifications of the variants of soybean histone H3 are also different, further proving that histone variants have distinct biological functions which are consistent with their specific modification patterns. In addition, the dynamic changes of these histone modifications and histone variants upon the salinity stress are also investigated. These results present comprehensive information for my following studies on understanding the biological functions of histone modifications in regulating the DNA transcription in soybean.

Although several histone modifications have been correlated with plants' response to abiotic stresses, the roles they played in this process are still not clear. In my studies, I propose a model through which the H3K4 methylation is able to regulate the salinity stress response in soybean. I have identified a soybean PHD finger domain containing protein, GmPHD1, which is up-regulated under the salinity stress and widely expressed in many organs of soybean and it can recognize histone H3K4 methylation, especially H3K4 dimethylaiton. After binding to the chromatin, GmPHD1 then functions as a histone code reader to decode the signal under H3K4 methylation by recruiting several other proteins, such as GmElongin A, GmISWI, and GmGNAT1, to remodel the chromatin structure and facilitate gene transcription. With chromatin immunoprecipitation assay, the GmPHD1 are found to be located in the promoter and body of several salt stress inducible genes in soybean, such as *HML806, HML1107*. Moreover, in the transgenic plants, the GmPHD1 increases the

plants' tolerance to salinity stress by regulating the transcription of genes in ROS scavenge system (Wei *et al*, 2009), which have been confirmed to be important in the soybean response to salinity stress in my proteomic studies. Therefore, GmPHD1 may work as a transcription co-activator to regulate the expression of such salt stress responsive genes under the salinity stress. At the same time, the activity of the GmPHD1 protein complex may be auto-regulated by self-acetylation of the GmGNAT1. Together with previous studies in soybean histone modifications and variants, I believe that my present investigations shed some light to better understanding the mechanism and functional significance of epigenetics in plants and the GmPHD1 may be a target with which the plants' tolerance to salinity stress can be modulated.

In order to validate the roles of GmPHD1 in regulating the expression of salt stress related genes, more regions where the GmPHD1 located in the genome are needed to be identified in the following studies. My proteomic studies have applied several potential candidate regions which can be confirmed by chromatin immunoprecipitation. Of course, ChIP-sequencing or ChIP–chip analysis using the tiling array will enable the genome-wide identification of the GmPHD1 location and supply us more information (Kim *et al*, 2010).

Referring to 'Epigenetics', which is the 'heritable changes in gene expression not attributable to nucleotide sequence variation' (Murrell *et al*., 2005); an important but unclear question can be raised: As a mechanism of epigenetics, can the histone modifications induced by the salinity stress be inherited by the next generation in plants? If such stress memory does exist, then we may increase the salt tolerance in the offspring by monitoring the epigenetic changes in their parents via simple salt treatment. However, at present, it is still a task of impossible, since so far we have

not understood how these histone modifications pass from the parents to their offsprings. In addition, we are still not sure whether epigenetic changes induced by abiotic stress might have an adaptive advantage for stress tolerance. Moreover, we have no idea what kind of histone modifications are likely to work in this way (Kim *et al*, 2010). Altogether, more efforts are needed in this interesting field in future.

References

Adcock IM, Ford P, Barnes PJ, Ito K. (2006). Epigenetics and airways disease. *Respir Res.* 7, 21-40.

Afroz A, Hashiguchi A, Khan R, Komatsu S. (2010). Analyses of the proteomes of the leaf, hypocotyl, and root of young soybean seedlings. *Protein & Peptide Letters.* 17, 319-331.

Agger K, Christensen J, Cloos PA, Helin K. (2008). The emerging functions of histone demethylases. *Curr Opin Genet Dev.* 18, 159-68.

Aghaei K, Ehsanpour AA, Komatsu S. (2008). Proteome analysis of potato under salt stress. *Journal of Proteome Research.* 7, 4858-4868.

Aghaei K, Ehsanpour AA, Shah AH, Komatsu S. (2009). Proteome analysis of soybean hypocotyl and root under salt stress. *Amino Acids.* 36, 91-98.

Ahmad K, Henikoff S. (2002). Histone H3 variants specify modes of chromatin assembly. *Proc. Natl. Acad. Sci. USA.* 99(Suppl 4), 16477-16484.

Ahsan N, Komatsu S. (2009). Comparative analyses of the proteomes of leaves and flowers at various stages of development reveal organ-specific functional differentiation of proteins in soybean. *Proteomics.* 9, 4889-4907.

Allis CD, Jenuwein T, Reinberg D. (2007). *Epigenetics.* New York: Cold Spring Harbor Laboratory Press.

Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature.* 408, 796-815.

Arai Y, Hayashi M, Nishimura M. (2008). Proteomic analysis of highly purified peroxisomes from etiolated soybean cotyledons. *Plant and Cell Physiology.* 49, 526-39.

Ashraf M. (1994) Breeding for salinity tolerance in plants. *Crit. Rev. Plant Sci.* 13, 17-42.

Aso T, Lane WS, Conaway JW, Conaway RC. (1995). Elongin (SIII): a multisubunit regulator of elongation by RNA polymerase II. *Science.* 269, 1439-1443.

Baginsky S. (2009). Plant proteomics: concepts, applications, and novel strategies for data interpretation. *Mass Spectrom Rev.* 28, 93-120.

Bailey-Serres J, Voesenek LA. (2008). Flooding stress: acclimations and genetic diversity. *Annual Review of Plant Biology.* 59, 313-339.

Baker LA, Allis CD, Wang GG. (2008). PHD fingers in human diseases: disorders arising from misinterpreting epigenetic marks. *Mutat Res.* 647, 3-12.

Barak O, Lazzaro MA, Lane WS, Speicher DW, Picketts DJ, Shiekhattar R. (2003). Isolation of human NURF: a regulator of Engrailed gene expression. *EMBO J.* 22, 6089-6100.

Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell.* 129, 823-837.

Bártová E, Krejcí J, Harnicarová A, Galiová G, Kozubek S. (2008). Histone modifications and nuclear architecture: a review. *J Histochem Cytochem.* 56, 711-721.

Bastola DR, Pethe VV, Winicov I. (1998). Alfin1, a novel zinc-finger protein in alfalfa roots that binds to promoter elements in the salt-inducible MsPRP2 gene. *Plant Mol Biol.* 38, 1123-1135.

Bastow R, Mylne JS, Lister C, Lippman Z, Martienssen RA, Dean C. (2004). Vernalization requires epigenetic silencing of FLC by histone methylation. *Nature.* 427,164-167.

Benhamed M, Bertrand C, Servet C, Zhou DX. (2006). Arabidopsis GCN5, HD1, and TAF1/HAF2 interact to regulate histone acetylation required for light-responsive gene expression. *Plant Cell.* 18, 2893-2903.

Berger SL. (2007). The complex language of chromatin regulation during transcription. *Nature.* 447, 407-412.

Berger SL, Kouzarides T, Shiekhattar R, Shilatifard A. (2009). An operational definition of epigenetics. *Genes Dev.* 23, 781-783.

Berger SL. (2002). Histone modifications in transcriptional regulation. *Curr. Opin. Genet. Dev.* 12, 142–148.

Bergmuller E, Gehrig PM, Gruissem W. (2007). Characterization of post-translational modifications of histone H2B-variants isolated from Arabidopsis thaliana. *J Proteome Res.* 6, 3655-3668.

Bernstein BE, Kamal M, Lindblad-Toh K, Bekiranov S, Bailey DK, Huebert DJ, McMahon S, Karlsson EK, Kulbokas EJ 3rd, Gingeras TR, Schreiber SL, Lander ES. (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell.* 120, 169-181.

Bernstein E, Hake SB. (2006). The nucleosome: a little variation goes a long way. *Biochem Cell Biol.* 84, 505-517.

Bhaumik SR, Smith E, Shilatifard A. (2007). Covalent modifications of histones during development and disease pathogenesis. *Nat Struct Mol Biol.* 14, 1008-1016.

Bhushan D, Pandey A, Choudhary MK, Datta A, Chakraborty S, Chakraborty N. (2007). Comparative proteomics analysis of differentially expressed proteins in chickpea extracellular matrix during dehydration stress. *Molecular Cellular Proteomics.* 6, 1868-1884.

Bienz M. (2006). The PHD finger, a nuclear protein-interaction domain. *Trends Biochem Sci.* 31, 35-40.

Blumwald E, Aharon GS, and Apse MP. (2000). Sodium transport in plant cells. *Biochim. Biophys. Acta.* 1465, 140-151.

Bonhomme L, Monclus R, Vincent D, Carpin S, Lomenech AM, Plomion C, Brignolas F, Morabito D. (2009). Leaf proteome analysis of eight Populus xeuramericana genotypes: genetic variation in drought response and in water-use efficiency involves photosynthesis-related proteins. *Proteomics.* 9, 4121-4142.

Botuyan MV, Lee J, Ward IM, Kim JE, Thompson JR, Chen J, Mer G. (2006). Structural basis for the methylation state-specific recognition of histone H4-K20 by 53BP1 and Crb2 in DNA repair. *Cell.* 127, 1361-1373.

Brechenmacher L, Lee J, Sachdev S, Song Z, Nguyen TH, Joshi T, Oehrle N, Libault M, Mooney B, Xu D, Cooper B, Stacey G. (2009). Establishment of a protein reference map for soybean root hair cells. *Plant Physiology.* 149, 670-682.

Brownell JE, Allis CD. (1996). Special HATs for special occasions: linking histone acetylation to chromatin assembly and gene activation. *Curr Opin Genet Dev.* 6, 176-184.

Caldwell CR, Turano FJ, McMahon MB. (1998). Identification of two cytosolic ascorbate peroxidase cDNAs from soybean leaves and characterization of their products by functional expression in *E. coli*. *Planta.* 204, 120-126.

Calikowski TT, Meier I. (2006). Isolation of nuclear proteins. *Methods Mol Biol.* 323, 393-402.

Carpentier SC, Witters E, Laukens K, Deckers P, Swennen R, Panis B. (2005). Preparation of protein extracts from recalcitrant plant tissues: an evaluation of different methods for two-dimensional gel electrophoresis analysis. *Proteomics.* 5, 2497-2507.

Caruso G, Cavaliere C, Guarino C, Gubbiotti R, Foglia P, Lagana A. (2008). Identification of changes in *Triticum durum* L. leaf proteome in response to salt stress by two-dimensional electrophoresis and MALDI-TOF mass spectrometry. *Analytical and Bioanalytical Chemistry.* 391, 381-390.

Casati P, Campi M, Chu F, Suzuki N, Maltby D, Guan S, Burlingame AL, Walbot V. (2008). Histone acetylation and chromatin remodeling are required for UV-B-dependent transcriptional activation of regulated genes in maize. *Plant Cell.* 20, 827-842.

Chen ZJ, Tian L. (2007). Roles of dynamic and reversible histone acetylation in plant development and polyploidy. *Biochim Biophys Acta.* 1769, 295-307.

Couture JF, Collazo E, Ortiz-Tello PA, Brunzelle JS, Trievel RC. (2007). Specificity

and mechanism of JMJD2A, a trimethyllysine-specific histone demethylase. *Nat Struct Mol Biol.* 14, 689-695.

Danchenko M, Skultety L, Rashydov NM, Berezhna VV, Mátel L, Salaj T, Pret'ová A, Hajduch M. (2009). Proteomic analysis of mature soybean seeds from the Chernobyl area suggests plant adaptation to the contaminated environment. *Journal of Proteome Research.* 8, 2915-2922.

Daniel JA, Grant PA. (2007). Multi-tasking on chromatin with the SAGA coactivator complexes. *Mutat Res.* 618, 135-148.

Dat J, Vandenabeele S, Vranova E, Van Montagu M, Inze D and Van Breusegem F. (2000). Dual action of the active oxygen species during plant stress responses. *Cell. Mol. Life Sci.* 57, 779-795.

Davletova S, Rizhsky L, Liang H, Shengqiang Z, Oliver DJ, Coutu J, Shulaev V, Schlauch K, Mittler R. (2005). Cytosolic ascorbate peroxidase 1 is a central component of the reactive oxygen gene network of Arabidopsis. *Plant Cell.* 17, 268-281.

De la Paz Sanchez M, Gutierrez C. (2009). Arabidopsis ORC1 is a PHD containing H3K4me3 effector that regulates transcription. *Proc. Natl. Acad. Sci. USA.* 106, 2065–2070.

Djordjevic MA, Oakes M, Li DX, Hwang CH, Hocart CH, Gresshoff PM. (2007). The glycine max xylem sap and apoplast proteome. *Journal of Proteome Research.* 6, 3771-3779.

Dou Y, Milne TA, Ruthenburg AJ, Lee S, Lee JW, Verdine GL, Allis CD and Roeder RG. (2006). Regulation of MLL1 H3K4 methyltransferase activity by its core components. *Nat. Struct. Mol. Biol.* 13, 713–719.

Droog F. (1997). Plant glutathione S-transferases, a tale of theta and tau. *Journal of Plant Growth Regulation.* 16, 95-107.

Dunand C, Crevecoeur M, Penel C. (2007). Distribution of superoxide and hydrogen peroxide in Arabidopsis root and their influence on root development: possible interaction with peroxidases. *New Phytologist.* 174, 332-341.

Duncan EM, Muratore-Schroeder TL, Cook RG, Garcia BA, Shabanowitz J, Hunt DF, Allis CD. (2008). Cathepsin L proteolytically processes histone H3 during mouse embryonic stem cell differentiation. *Cell.* 135, 284-294.

Earley KW, Shook MS, Brower-Toland B, Hicks L, Pikaard CS. (2007). In vitro specificities of Arabidopsis co-activator histone acetyltransferases: implications for histone hyperacetylation in gene activation. *Plant J.* 52, 615-626.

Ebert A, Lein S, Schotta G, Reuter G. (2006). Histone modification and the control of heterochromatic gene silencing in Drosophila. *Chromosome Res.* 14, 377–392.

Eissenberg JC, James TC, Foster-Hartnett DM, Hartnett T, Ngan V, Elgin SC. (1990). Mutation in a heterochromatin-specific chromosomal protein is associated with suppression of position-effect variegation in Drosophila melanogaster. *Proc. Natl. Acad. Sci. USA.* 87:9923-9927.

Espartero J, Pintor-Toro JA, Pardo JM. (1994). Differential accumulation of S-adenosylmethionine synthetase transcripts in response to salt stress. *Plant Molecular Biology.* 25, 217-227.

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

Fernandes FM, Arrabaça MC, Carvalho LMM. (2004). Sucrose metabolism in *Lupinus albus* L. under salt stress. *Biologia Plantarum.* 48, 317-319.

Feussner I, Wasternack C. (2002). The lipoxygenase pathway. *Annual Review of Plant Biology.* 53, 275-297.

Foreman J, Demidchik V, Bothwell JHF, Mylona P, Miedema H, Torres MA, Linstead P, Costa S, Brownlee C, Jones JDG, Davies JM, Dolan L. (2003). Reactive oxygen species produced by NADPH oxidase regulate plant cell growth. *Nature.* 422, 442–446.

Fuchs J, Demidov D, Houben A, Schubert I. (2006). Chromosomal histone modification patterns--from conservation to diversity. *Trends Plant Sci.* 11, 199-208.

Funatsuki H, Kurosaki H, Murakami T, Matsuba S, Kawaguchi K, Yumoto S, Sato Y. (2003). Deficiency of a cytosolic ascorbate peroxidase associated with chilling tolerance in soybean. *TAG Theoretical and Applied Genetics.* 106, 494-502.

Gozani O, Karuman P, Jones DR, Ivanov D, Cha J, Lugovskoy AA, Baird CL, Zhu H, Field SJ, Lessnick SL, Villasenor J, Mehrotra B, Chen J, Rao VR, Brugge JS, Ferguson CG, Payrastre B, Myszka DG, Cantley LC, Wagner G, Divecha N, Prestwich GD, Yuan J. (2003). The PHD finger of the chromatin-associated protein ING2 functions as a nuclear phosphoinositide receptor. *Cell.* 114, 99-111.

Hajduch M, Ganapathy A, Stein JW, Thelen JJ. (2005). A systematic proteomic study of seed filling in soybean. Establishment of high-resolution two-dimensional reference maps, expression profiles, and an interactive proteome database. *Plant Physiology.* 137, 1397-1419.

Hajheidari M, Abdollahian-Noghabi M, Askari H, Heidari M, Sadeghian SY, Ober ES, Salekdeh GH. (2005). Proteome analysis of sugar beet leaves under drought stress. *Proteomics.* 5, 950-960.

Hake SB, Garcia BA, Duncan EM, Kauer M, Dellaire G, Shabanowitz J, Bazett-Jones DP, Allis CD, Hunt DF. (2006). Expression patterns and post-translational modifications associated with mammalian histone H3 variants. *J Biol Chem.* 281, 559-568.

Hamada A, Shono M, Xia T, Ohta M, Hayashi Y, Tanaka A, and Hayakawa T. (2001). Isolation and characterization of Na^+/ H^+ antiporter gene from the halophyte *Atriplex gmelini*. *Plant Mol. Biol.* 46, 35-42.

Hasegawa PM, Bressan RA, Zhu JK, and Bohnert HJ. (2000). Plant cellular and molecular responses to high salinity. *Annual Review of Plant Physiology and Plant Molecular Biology.* 51, 463-499.

Hashiguchi A, Sakata K, Komatsu S. (2009). Proteome analysis of early-stage soybean seedlings under flooding stress. *Journal of Proteome Research* 8, 2058-2069.

He X, Fan HY, Garlick JD, Kingston RE. (2008). Diverse regulation of SNF2h chromatin remodeling by noncatalytic subunits. *Biochemistry.* 47, 7025-7033.

He Y, Amasino RM. (2005). Role of chromatin modification in flowering-time control. *Trends Plant Sci.* 10, 30-35.

He Y, Michaels SD, Amasino RM. (2003). Regulation of flowering time by histone acetylation in Arabidopsis. *Science.* 302, 1751-1754.

Henikoff S. (2008). Nucleosome destabilization in the epigenetic regulation of gene expression. *Nature Rev. Genet.* 9, 15–26.

Houben A, Demidov D, Caperta AD, Karimi R, Agueci F, Vlasenko L. (2007). Phosphorylation of histone H3 in plants--a dynamic affair. *Biochim Biophys Acta.* 1769, 308-315.

Huang J, Zhang H, Wang J, Yang J. (2003). Molecular cloning and characterization of rice 6-phosphogluconate dehydrogenase gene that is up-regulated by salt stress. *Molecular Biology Reports.* 30, 223-227.

Huyen Y, Zgheib O, Ditullio RA Jr, Gorgoulis VG, Zacharatos P, Petty TJ, Sheston EA, Mellert HS, Stavridi ES, Halazonetis TD. (2004). Methylated lysine 79 of histone H3 targets 53BP1 to DNA double-strand breaks. *Nature.* 432, 406-411.

Im H, Park C, Feng Q, Johnson KD, Kiekhaefer CM, Choi K, Zhang Y, Bresnick EH. (2003). Dynamic regulation of histone H3 methylated at lysine 79 within a tissue-specific chromatin domain. *J Biol Chem.* 278, 18346-18352.

Jain RL and Selvaraj G. (1997). Molecular genetic improvement of salt tolerance in plants. *Biotechnology Annual Review.* 3, 245-267.

Jansen LE, Black BE, Foltz DR, Cleveland DW. (2007). Propagation of centromeric chromatin requires exit from mitosis. *J Cell Biol.* 176, 795-805.

Jenuwein T, Allis CD. (2001). Translating the histone code. *Science.* 293, 1074-1080.

Jiang Y, Yang B, Harris NS, Deyholos MK. (2007). Comparative proteomic analysis of NaCl stress-responsive proteins in Arabidopsis roots. *Journal of Experimental Botany.* 58, 3591-3607.

Jiang YQ, Deyholos MK. (2006). Comprehensive transcriptional profiling of NaCl-stressed Arabidopsis roots reveals novel classes of responsive genes. *BMC Plant Biology.* 6, 25.

Johnson L, Cao X, Jacobsen S. (2002). Interplay between two epigenetic marks. DNA methylation and histone H3 lysine 9 methylation. *Curr Biol.* 12, 1360-1367.

Johnson L, Mollah S, Garcia BA, Muratore TL, Shabanowitz J, Hunt DF, Jacobsen SE. (2004). Mass spectrometry analysis of Arabidopsis histone H3 reveals distinct combinations of post-translational modifications. *Nucleic Acids Res.* 32, 6511-6518.

Jorrín JV, Maldonado AM, Castillejo MA. (2007). Plant proteome analysis: a 2006 update. *Proteomics.* 7, 2947-2962.

Kavallaris M, Marshall GM. (2005). Proteomics and disease: opportunities and challenges. *Med J Aust.* 182, 575-579.

Khorasanizadeh S. (2004). The nucleosome: from genomic organization to genomic regulation. *Cell.* 116, 259-272.

Kim J, Daniel J, Espejo A, Lake A, Krishna M, Xia L, Zhang Y, Bedford MT. (2006). Tudor, MBT and chromo domains gauge the degree of lysine methylation. *EMBO Rep.* 7, 397-403.

Kim JM, To TK, Ishida J, Morosawa T, Kawashima M, Matsui A, Toyoda T, Kimura H, Shinozaki K, Seki M. (2008). Alterations of lysine modifications on the histone H3 N-tail under drought stress conditions in Arabidopsis thaliana. *Plant Cell Physiol.* 49, 1580-1588.

Kim JM, To TK, Nishioka T, Seki M. (2010). Chromatin regulation functions in plant abiotic stress responses. *Plant Cell Environ.* 33, 604-611.

Komatsu S, Wada T, Abaléa Y, Nouri MZ, Nanjo Y, Nakayama N, Shimamura S, Yamamoto R, Nakamura T, Furukawa K. (2009). Analysis of plasma membrane proteome in soybean and application to flooding stress response. *Journal of Proteome Research.* 8, 4487-4499.

Komili S, Silver PA. (2008). Coupling and coordination in gene expression processes:

a systems biology view. *Nat Rev Genet.* 9, 38-48.

Kouzarides T. (2007). Chromatin modifications and their function. *Cell.* 128, 693-705.

Kozlowski TT. (2000). Responses of woody plants to human-induced environmental stresses: Issues, problems, and strategies for alleviating stress. *Crit. Rev. Plant Sci.* 19, 91-170.

Kwon SY, Xiao H, Glover BP, Tjian R, Wu C, Badenhorst P. (2008). The nucleosome remodeling factor (NURF) regulates genes involved in Drosophila innate immunity. *Dev Biol.* 316, 538-547.

Lane CS. (2005). Mass spectrometry-based proteomics in the life sciences. *Cell Mol Life Sci.* 62, 848-869.

Lee JD, Shannon JG, Vuong TD, Nguyen HT. (2009). Inheritance of salt tolerance in wild soybean (Glycine soja Sieb. and Zucc.) accession PI483463. *J Hered.* 100, 798-801.

Lee WY, Lee D, Chung WI, Kwon CS. (2009). Arabidopsis ING and Alfin1-like protein families localize to the nucleus and bind to H3K4me3/2 via plant homeodomain fingers. *Plant J.* 58, 511-524.

Li B, Carey M, Workman JL. (2007). The role of chromatin during transcription. *Cell.* 128, 707-719.

Li B, Howe L, Anderson S, Yates JR 3rd, Workman JL. (2003). The Set2 histone methyltransferase functions through the phosphorylated carboxyl-terminal domain of RNA polymerase II. *J Biol Chem.* 278, 8897-8903.

Li H, Ilin S, Wang W, Duncan EM, Wysocka J, Allis CD, Patel DJ. (2006). Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature.* 442, 91-95.

Li B, Carey M, Workman JL. (2007). The role of chromatin during transcription. *Cell.* 128, 707-719.

Liska AJ, Shevchenko A, Pick U, Katz A. (2004). Enhanced photosynthesis and redox energy production contribute to salinity tolerance in Dunaliella as revealed by homology-based proteomics. *Plant Physiology.* 136, 2806-2817.

Maas EV and Hoffman GJ. (1977). Crop salt tolerance-current assessment. *J. Irrig. Drainage Div. Am. Soc. Civ. Eng.* 103, 115.

Marino-Ramirez L, Kann MG, Shoemaker BA, Landsman D. (2005). Histone structure and nucleosome stability. *Expert Rev Proteomics.* 2, 719-729.

Martin DG, Baetz K, Shi X, Walter KL, MacDonald VE, Wlodarski MJ, Gozani O, Hieter P, Howe L. (2006). The Yng1p plant homeodomain finger is a methyl-histone binding module that recognizes lysine 4-methylated histone H3. *Mol Cell Biol.* 26, 7871-7879.

McKittrick E, Gafken PR, Ahmad K, Henikoff S. (2004). Histone H3.3 is enriched in covalent modifications associated with active chromatin. *Proc. Natl. Acad. Sci. USA.* 101, 1525-1530.

Mellor J. (2006). It takes a PHD to read the histone code. *Cell.* 126, 22-24.

Miranda TB, Jones PA. (2007). DNA methylation: the nuts and bolts of repression. *J Cell Physiol.* 213, 384-390.

Moreno JI, Martín R, Castresana C. (2005). Arabidopsis SHMT1, a serine hydroxymethyltransferase that functions in the photorespiratory pathway influences resistance to biotic and abiotic stress. *Plant Journal.* 41, 451-463.

Murrell A, Rakyan VK, Beck S. (2005). From genome to epigenome. *Hum Mol Genet.* 14, R3-R10.

Nakayama J, Rice JC, Strahl BD, Allis CD, Grewal SI. (2001). Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science.* 292, 110-113.

Natarajan SS, Xu C, Bae H, Caperna TJ, Garrett WM. (2006). Characterization of storage proteins in wild (*Glycine soja*) and cultivated (*Glycine max*) soybean seeds using proteomic analysis. *Journal of Agricultural and Food Chemistry.* 54,

3114-3120.

Ndimba BK, Chivasa S, Simon WJ, Slabas AR. (2005). Identification of Arabidopsis salt and osmotic stress responsive proteins using two-dimensional difference gel electrophoresis and mass spectrometry. *Proteomics.* 5, 4185–4196.

Neely KE, Workman JL. (2002). The complexity of chromatin remodeling and its links to cancer. *Biochim Biophys Acta.* 1603, 19-29.

Ng HH, Ciccone DN, Morshead KB, Oettinger MA, Struhl K. (2003). Lysine-79 of histone H3 is hypomethylated at silenced loci in yeast and mammalian cells: a potential mechanism for position-effect variegation. *Proc. Natl. Acad. Sci. USA.* 100, 1820-1825.

Nicklay JJ, Shechter D, Chitta RK, Garcia BA, Shabanowitz J, Allis CD, Hunt DF. (2009). Analysis of histones in Xenopus laevis. II. mass spectrometry reveals an index of cell type-specific modifications on H3 and H4. *J Biol Chem.* 284, 1075-1085.

Oberschall A, Deak M, Torok K, Sass L, Vaas I, Kovacs I, Fehér A, Dudits D, Horváth GV. (2000). A novel aldolase/aldehyde reductase protects transgenic plants against lipid peroxidation under chemical and drought stresses. *Plant Journal.* 24, 437–446.

Oehrle NW, Sarma AD, Waters JK, Emerich DW. (2008). Proteomic analysis of soybean nodule cytosol. *Phytochemistry.* 69, 2426-2438.

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

Owens S. (2001). Salt of the earth: genetic engineering may help to reclaim agricultural land lost due to salinisation. *EMBO Rep.* 2, 877-879.

Parida AK, Das AB. (2005). Salt tolerance and salinity effects on plants: a review. *Ecotoxicol Environ Saf.* 60, 324-349.

Passardi F, Tognolli M, De Meyer M, Penel C, Dunand C. (2006). Two cell wall associated peroxidases from Arabidopsis influence root elongation. *Planta.* 223,

965–974.

Penterman J, Zilberman D, Huh JH, Ballinger T, Henikoff S, Fischer RL. (2007). DNA demethylation in the Arabidopsis genome. *Proc. Natl. Acad. Sci. USA.* 104, 6752-6757.

Pien S, Fleury D, Mylne JS, Crevillen P, Inzé D, Avramova Z, Dean C, Grossniklaus U. (2008). ARABIDOPSIS TRITHORAX1 Dynamically Regulates FLOWERING LOCUS C Activation via Histone 3 Lysine 4 Trimethylation. *Plant Cell.* 20, 580-588.

Pray-Grant MG, Daniel JA, Schieltz D, Yates JR, Grant PA. (2005). Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation. *Nature.* 433:434-438.

Qureshi MI, Qadir S, Zolla L. (2007). Proteomics-based dissection of stress-responsive pathways in plants. *J Plant Physiol.* 164, 1239-1260.

Ramón-Maiques S, Kuo AJ, Carney D, Matthews AG, Oettinger MA, Gozani O, Yang W. (2007). The plant homeodomain finger of RAG2 recognizes histone H3 methylated at both lysine-4 and arginine-2. *Proc. Natl. Acad. Sci. USA.* 104, 18993-18998.

Rando OJ. (2007). Global patterns of histone modifications. *Curr. Opin. Genet. Dev.* 17, 94–99.

Ratajczak R. (2000). Structure, function and regulation of the plant vacuolar H^+-translocating ATPase. *Biochimica et Biophysica Acta.* 1465, 17-36.

Riccardi F, Gazeau P, de Vienne D, Zivy M. (1998). Protein changes in response to progressive water deficit in maize. Quantitative variation and polypeptide identification. *Plant Physiology.* 117, 1253-1263.

Ros Barceló A. (1997). Lignification in plant cell walls. *International Review of Cytology.* 176, 87–132.

Roth U, von Roepenack-Lahaye E, Clemens S. (2006). Proteome changes in Arabidopsis thaliana roots upon exposure to Cd^{2+}. *Journal of Experimental Botany.* 57, 4003-4013.

Ruiz-García L, Cervera MT, Martínez-Zapater JM. (2005). DNA methylation increases throughout Arabidopsis development. *Planta.* 222, 301-306.

Ruthenburg AJ, Allis CD, Wysocka J. (2007). Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. *Mol Cell.* 25, 15-30.

Sanders SL, Portoso M, Mata J, Bahler J, Allshire RC, Kouzarides T. (2004). Methylation of histone H4 Lysine 20 controls recruitment of Crb2 to sites of DNA damage. *Cell.* 119, 603-614.

Santos-Rosa H, Kirmizis A, Nelson C, Bartke T, Saksouk N, Cote J, Kouzarides T. (2009). Histone H3 tail clipping regulates gene expression. *Nat Struct Mol Biol.* 16, 17-22.

Schmitz RJ, Sung S, Amasino RM. (2008). Histone arginine methylation is required for vernalization-induced epigenetic silencing of FLC in winter-annual Arabidopsis thaliana. *Proc. Natl. Acad. Sci. USA.* 105, 411-416.

Schöb H, Grossniklaus U. (2006). The first high-resolution DNA "methylome". *Cell.* 22, 1025-1028.

Schotta G, Ebert A, Krauss V, Fischer A, Hoffmann J, Rea S, Jenuwein T, Dorn R, Reuter G. (2002). Central role of Drosophila SU(VAR)3-9 in histone H3-K9 methylation and heterochromatic gene silencing. *EMBO J.* 21, 1121-1131.

Schübeler D, MacAlpine DM, Scalzo D, Wirbelauer C, Kooperberg C, van Leeuwen F, Gottschling DE, O'Neill LP, Turner BM, Delrow J, Bell SP, Groudine M. (2004). The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. *Genes Dev.* 18, 1263-1271.

Seki M, Narusaka M, Ishida J, Nanjo T, Fujita M, Oono Y, Kamiya A, Nakajima M, Enju A, Sakurai T, Satou M, Akiyama K, Taji T, Yamaguchi-Shinozaki K, Carninci P, Kawai J, Hayashizaki Y, Shinozaki K. (2002). Monitoring the expression profiles of 7000 Arabidopsis genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. *Plant J.* 31, 279-292.

Sengupta S, Majumder AL. (2009). Insight into the salt tolerance factors of a wild

halophytic rice, *Porteresia coarctata*: a physiological and proteomic approach. *Planta.* 229, 911-929.

Shao GH, Chang RZ and Chen YW. (1993). Recent development of study on salt tolerance in soybean. *Soybean Sci.* 12, 244-248.

Shao GH, Chan RZ, Chen YW, Yan SR. (1994). Study on inheritance of salt tolerance in soybean. *Acta Agron. Sin.* 20, 721–726.

Shechter D, Dormann HL, Allis CD, Hake SB. (2007). Extraction, purification and analysis of histones. *Nat Protoc.* 2, 1445-1457.

Shechter D, Nicklay JJ, Chitta RK, Shabanowitz J, Hunt DF, Allis CD. (2009). Analysis of histones in Xenopus laevis. I. A distinct index of enriched variants and modifications exists in each cell type and is remodeled during developmental transitions. *J Biol Chem.* 284, 1064-1074.

Shi Y, Lan F, Matson C, Mulligan P, Whetstine JR, Cole PA and Casero RA. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell.* 119, 941–53.

Shi HZ, Ishitani M, Kim C and Zhu JK. (2000). The *Arabidopsis thaliana* salt tolerance gene *SOS1* encodes a putative Na^+/H^+ antiporter. *Proc. Natl. Acad. Sci. USA.* 97, 6896-6901.

Shigeoka S, Ishikawa T, Tamoi M, Miyagawa Y, Takeda T, Yabuta Y, Yoshimura K. (2002). Regulation and function of ascorbate peroxidase isoenzymes. *Journal of Experimental Botany.* 53, 1305-1319.

Showalter AM. (1993). Structure and function of plant cell wall proteins. *Plant Cell.* 5, 9-23.

Shulaev V, Cortes D, Miller G, Mittler R. (2008). Metabolomics for plant stress response. *Physiologia Plantarum.* 132, 199-208.

Simon MD, Chu F, Racki LR, de la Cruz CC, Burlingame AL, Panning B, Narlikar GJ, Shokat KM. (2007). The site-specific installation of methyl-Lysine analogs into recombinant histones. *Cell.* 128, 1003-1012.

Singh J, Freeling M, Lisch D. (2008). A position effect on the heritability of epigenetic silencing. *PLoS Genet.* 4, e1000216.

Singla-Pareek SL, Reddy MK, Sopory SK. (2003). Genetic engineering of the glyoxalase pathway in tobacco leads to enhanced salinity tolerance. *Proc. Natl. Acad. Sci. USA.* 100, 14672-14677.

Smallwood A, Esteve PO, Pradhan S, Carey M. (2007). Functional cooperation between HP1 and DNMT1 mediates gene silencing. *Genes Dev.* 21, 1169-1178.

Sobkowiak R, Deckert J. (2006). Proteins induced by cadmium in soybean cells. *Journal of Plant Physiology.* 163, 1203-1206.

Sokol A, Kwiatkowska A, Jerzmanowski A, Prymakowska-Bosak M. (2007). Up-regulation of stress-inducible genes in tobacco and Arabidopsis cells in response to abiotic stresses and ABA treatment correlates with dynamic changes in histone H3 and H4 modifications. *Planta.* 227, 245-254.

Staswick PE. (1989). Developmental regulation and the influence of plant sinks on vegetative storage protein gene expression in soybean leaves. *Plant Physiology.* 89, 309-315.

Steger DJ, Lefterova MI, Ying L, Stonestrom AJ, Schupp M, Zhuo D, Vakoc AL, Kim JE, Chen J, Lazar MA, Blobel GA, Vakoc CR. (2008). DOT1L/KMT4 recruitment and H3K79 methylation are ubiquitously coupled with gene transcription in mammalian cells. *Mol Cell Biol.* 28, 2825-2839.

Suganuma T, Workman JL. (2008). Crosstalk among histone modifications. *Cell.* 135, 604-607.

Sun ZW, Allis CD. (2002). Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast. *Nature.* 418, 104-108.

Sun H. (1987). Good crops from salty soils. *Feeding a Billion. Frontiers of Chinese Agriculture.* East Lansing, Michigan State University Press: 83-92.

Talbert PB, Henikoff S. (2010). Histone variants--ancient wrap artists of the

epigenome. *Nat Rev Mol Cell Biol.* 11, 264-275.

Talbert PB, Henikoff S. (2006). Spreading of silent chromatin: inaction at a distance. *Nat Rev Genet.* 7, 793–803.

Tanaka K, Waki H, Ido Y, Akita S, Yoshida Y and Yoshida T. (1988). Protein and polymer analyses up to m/z 100,000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* 2, 151–153.

Tapley TL, Korner JL, Barge MT, Hupfeld J, Schauerte JA, Gafni A, Jakob U, Bardwell JC. (2009). Structural plasticity of an acid-activated chaperone allows promiscuous substrate binding. *Proc. Natl. Acad. Sci. USA.* 106, 5557-5562.

Taverna SD, Ilin S, Rogers RS, Tanny JC, Lavender H, Li H, Baker L, Boyle J, Blair LP, Chait BT, Patel DJ, Aitchison JD, Tackett AJ, Allis CD. (2006). Yng1 PHD finger binding to H3 trimethylated at K4 promotes NuA3 HAT activity at K14 of H3 and transcription at a subset of targeted ORFs. *Mol Cell.* 24, 785-796.

Taverna SD, Li H, Ruthenburg AJ, Allis CD, Patel DJ. (2007). How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat Struct Mol Biol.* 14, 1025-1040.

Teixeira J, Pereira S, Queiros F, Fidalgo F. (2006). Specific roles of potato glutamine synthetase isoenzymes in callus tissue grown under salinity: molecular and biochemical responces. *Plant Cell, Tissue and Organ Culture.* 87, 1-7.

Tognolli M, Penel C, Greppin H, Simon P. (2002). Analysis and expression of the class III peroxidase large gene family in *Arabidopsis thaliana*. *Gene.* 288, 129–138.

Tuteja N. (2007). Mechanisms of high salinity tolerance in plants. *Methods in Enzymology.* 428, 419-438.

Urano K, Kurihara Y, Seki M, Shinozaki K. (2010). 'Omics' analyses of regulatory networks in plant abiotic stress responses. *Current Opinion in Plant Biol.* 13, 132-138.

Vetting MW, S de Carvalho LP, Yu M, Hegde SS, Magnet S, Roderick SL, Blanchard JS. (2005). Structure and functions of the GNAT superfamily of acetyltransferases.

Arch Biochem Biophys. 433, 212-226.

Vianello A, and Macri F. (1991). Generation of superoxide anion and hydrogen peroxide at surface of plant cells. *Journal of Bioenergetics and Biomembranes.* 23, 409–423.

Wakimoto BT. (1998). Beyond the nucleosome: epigenetic aspects of position-effect variegation in Drosophila. *Cell.* 93, 321-324.

Wang MC, Peng ZY, Li CL, Li F, Liu C, Xia GM. (2008). Proteomic analysis on a high salt tolerance introgression strain of Triticum aestivum/Thinopyrum ponticum. *Proteomics.* 8, 1470-1489.

Wang W, Vinocur B, Altman A. (2003). Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. Planta. 218, 1-14.

Wang X, Yang P, Zhang X, Xu Y, Kuang T, Shen S, He Y. (2009). Proteomic analysis of the cold stress response in the moss, *Physcomitrella patens*. *Proteomics.* 9, 4529-4538.

Wei W, Huang J, Hao YJ, Zou HF, Wang HW, Zhao JY, Liu XY, Zhang WK, Ma B, Zhang JS, Chen SY. (2009). Soybean GmPHD-type transcription regulators improve stress tolerance in transgenic Arabidopsis plants. *PLoS One.* 4, e7209.

Whetstine JR, Nottke A, Lan F, Huarte M, Smolikov S, Chen Z, Spooner E, Li E, Zhang G, Colaiacovo M, Shi Y. (2006). Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases. *Cell.* 125, 467-481.

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

Winicov I. (1993). cDNA encoding putative zinc finger motifs from salt-tolerant alfalfa (Medicago sativa L.) cells. *Plant Physiol.* 102, 681–682.

Witzel K, Weidner A, Surabhi GK, Borner A, Mock HP. (2009). Salt stress-induced

alterations in the root proteome of barley genotypes with contrasting response towards salinity. *Journal of Experimental Botany.* 60, 3545-3557.

Wu T, Yuan T, Tsai SN, Wang C, Sun SM, Lam HM and Ngai SM. (2009). Mass spectrometry analysis of the variants of histone H3 and H4 of soybean and their post-translational modifications. *BMC Plant Biol.* 9:98.

Wysocka J, Swigut T, Milne TA, Dou Y, Zhang X, Burlingame AL, Roeder RG, Brivanlou AH, Allis CD. (2005). WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell.* 121, 859–872.

Wysocka J, Swigut T, Xiao H, Milne TA, Kwon SY, Landry J, Kauer M, Tackett AJ, Chait BT, Badenhorst P, Wu C, Allis CD. (2006). A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature.* 442, 86-90.

Xiong L, Schumaker KS, Zhu JK. (2002). Cell signaling during cold, drought, and salt stress. *Plant Cell.* 14 Suppl, S165-183.

Xiong L and Zhu JK. (2002). Molecular and genetic aspects of plant responses to osmotic stress. *Plant Cell Environ.* 25, 131-139.

Xu C, Huang B. (2010). Differential proteomic response to heat stress in thermal *Agrostis scabra* and heat-sensitive *Agrostis stolonifera*. *Physiologia Plantarum.* (Epub ahead of print).

Xu C, Sullivan JH, Garrett WM, Caperna TJ, Natarajan S. (2008). Impact of solar ultraviolet-B on the proteome in soybean lines differing in flavonoid contents. *Phytochemistry.* 69, 38-48.

Xu CR, Liu C, Wang YL, Li LC, Chen WQ, Xu ZH, Bai SN. (2005). Histone acetylation affects expression of cellular patterning genes in the Arabidopsis root epidermis. *Proc. Natl. Acad. Sci. USA.* 102, 14469-14474.

Xu L, Zhao Z, Dong A, Soubigou-Taconnat L, Renou JP, Steinmetz A, Shen WH. (2008). Di- and tri- but not mono-methylation on histone H3 Lysine 36 marks active transcription of genes involved in flowering time regulation and other processes in Arabidopsis thaliana. *Mol Cell Biol.* 28, 1348-1360.

Yamaguchi M, Valliyodan B, Zhang J, Lenoble ME, Yu O, Rogers EE, Nguyen HT, Sharp RE. (2010). Regulation of growth response to water stress in the soybean primary root. I. Proteomic analysis reveals region-specific regulation of phenylpropanoid metabolism and control of free iron in the elongation zone. *Plant Cell Environ.* 33, 223-243.

Yan S, Tang Z, Su W, Sun W. (2005). Proteomic analysis of salt stress-responsive proteins in rice root. *Proteomics.* 5, 235-244.

Yang XJ, Seto E. (2008). Lysine acetylation: codified crosstalk with other posttranslational modifications. *Mol Cell.* 31, 449-461.

Yang XJ. (2004). Lysine acetylation and the bromodomain: a new partnership for signaling. *Bioessays.* 26, 1076–1087.

Yeo A. (1998). Molecular biology of salt tolerance in the context of whole-plant physiology. *J. Exp. Bot.* 49, 915-929.

Zhang K, Sridhar VV, Zhu J, Kapoor A, Zhu JK. (2007). Distinctive Core Histone Post-Translational Modification Patterns in Arabidopsis thaliana. *PLoS One.* 2:e1210.

Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW, Chen H, Henderson IR, Shinn P, Pellegrini M, Jacobsen SE, Ecker JR. (2006). Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. *Cell.* 126, 1189-1201.

Zhang X. (2008). The epigenetic landscape of plants. *Science.* 320, 489-492.

Zhang HX, Hodson JN, Williams JP, and Blumwald E. (2001). Engineering salt-tolerant Brassica plants:characterization of yield and seed oil quality in transgenic plants with increased vacuolar sodium accumulation. *Proc. Natl. Acad. Sci. USA.* 98, 12832-12836.

Zhou H, Madden BJ, Muddiman DC, Zhang Z. (2006). Chromatin assembly factor 1 interacts with histone H3 methylated at lysine 79 in the processes of epigenetic silencing and DNA repair. *Biochemistry.* 45, 2852-2861.

Zhu J, Jeong JC, Zhu Y, Sokolchik I, Miyazaki S, Zhu JK, Hasegawa PM, Bohnert HJ, Shi H, Yun DJ, Bressan RA. (2008). Involvement of Arabidopsis HOS15 in histone deacetylation and cold tolerance. *Proc. Natl. Acad. Sci. USA*, 105, 4945-4950.

Zhu JK. (2001). Plant salt tolerance. *Trends Plant Science*. 6, 66-71.

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