BIOCHEMICAL CHARACTERIZATION OF PLANT SMALL CTD PHOSPHATASES AND APPLICATION OF CTD PHOSPHATASE MUTANT IN HYPERACCUMULATION OF FLAVONOIDS IN *ARABIDOPSIS*

A Dissertation

by

YUE FENG

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2010

Major Subject: Molecular and Environmental Plant Sciences

Biochemical Characterization of Plant Small CTD Phosphatases and Application of CTD Phosphatase Mutant in Hyperaccumulation of Flavonoids in *Arabidopsis*

BIOCHEMICAL CHARACTERIZATION OF PLANT SMALL CTD PHOSPHATASES AND APPLICATION OF CTD PHOSPHATASE MUTANT IN HYPERACCUMULATION OF FLAVONOIDS IN ARABIDOPSIS

A Dissertation

by

YUE FENG

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Hisashi Koiwa
Committee Members,	Kendal D. Hirschi
	Scott A. Finlayson
	Mikhailo V. Kolomiets
Chair of Intercollegiate Faculty,	Jean Gould

August 2010

Major Subject: Molecular and Environmental Plant Sciences

ABSTRACT

Biochemical Characterization of Plant Small CTD Phosphatases and Application of CTD Phosphatase Mutant in Hyperaccumulation of Flavonoids in *Arabidopsis*.

(August 2010)

Yue Feng, B.S., China Agricultural University, China

Chair of Advisory Committee: Dr. Hisashi Koiwa

In addition to AtCPL1-4, the genome of *Arabidopsis thaliana* encodes a large number of putative acid phosphatases. The predicted *Arabidopsis* SCP1-like small phosphatases (SSP) are highly homologous to the catalytic domain of eukaryotic RNA polymerase II carboxyl terminal domain (pol II CTD) phosphatases. Among the family members, SSP4, SSP4b and SSP5 form a unique group characterized by long N-terminal extensions. These three SSPs showed similar and ubiquitous gene expression. SSP4 and SSP4b were localized exclusively in the nuclei, while SSP5 accumulated both in the nucleus and cytoplasm. In vitro observation revealed that SSP4 and SSP4b dephosphorylated the pol II CTD-PO_4 at both Ser2 and Ser5 in the conserved heptad repeats; however, SSP5 dephosphorylated only Ser5 of CTD-PO_4. These results indicate that *Arabidopsis SSP* family encodes active CTD phosphatases similarly to animal SCP1 family proteins and plant CPLs family proteins, but with distinct substrate specificities.

ssp mutants did not exhibit phenotypic abnormalities under normal growth conditions. However, *ssp5* single mutants and *ssp4 ssp4b ssp5* triple mutants showed enhanced

sensitivity to ABA and glucose during seed germination. Yet, increased ABA-inducible gene expressions were not distinguishable in triple mutants compared to wild type plants upon ABA treatment. Unlike the *ssp* mutations, the *cpl1* mutation strongly induced *RD29A* expression in response to cold, ABA and NaCl treatments. Thus, the *cpl1* mutant is an ideal genetic background for an inducible gene expression system, in which the detrimental effect to host plants caused by a conventional constitutive expression could be avoided.

Production of flavonoid such as anthocyanins in *Arabidopsis* is relatively easy to monitor and is regulated by transcription factors such as PAP1. *PAP1* activates the expression of multiple enzymes in the anthocyanin biosynthesis pathway; however, high level of flavonoid production could cause vegetative growth retardation. To optimize flavonoid accumulation, a three-component system was designed consisting of a cold inducible *RD29A-PAP1* expression cassette, a feedforward effector *RD29A-CBF3*, and a mutation in host repressor *CPL1*. Transgenic *cpl1* plants containing both homozygous *PAP1* and *CBF3* transgenes produced 30-fold higher level of total anthocyanins than control plants upon cold treatment. LC/MS/MS analysis showed the flavonoid profile in cold-induced transgenic plants resembled that of previously reported *pap1-D* plants but were enriched for kaempferol derivatives. Furthermore, PAP1 and environmental signals synergistically regulate flavonoid pathway to produce a flavonoid blend that has not been produced by PAP1 overexpression or cold treatment alone. These results delineate the usability of the three-component inducible system in plant metabolic engineering.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Hisashi Koiwa, for the great opportunity of my Ph.D. study in the exciting field of plant molecular biology. I am extremely thankful for his generous support and his patient guidance throughout the course of this research.

Special thanks go to my committee members, Dr. Kendal Hirschi, Dr. Scott Finlayson, and Dr. Mikhailo Kolomiets for their encouragement and advice. I also want to express my thanks to all my friends and colleagues in Koiwa lab, in Department of Horticultural Sciences, and Vegetable and Fruit Improvement Center. I also would like to extend my gratitude to NSF and USDA which support the researches described in this dissertation.

Finally, I would like to thank my parents for their endless support and all my friends for their encouragement throughout my doctoral program.

NOMENCLATURE

SSP	SCP1-like small phosphatase
pol II	RNA polymerase II
CTD	carboxyl terminal domain
ABA	abscisic acid
TFIIH	transcription factor IIH
CDK	cyclin-dependent kinase
AMP	Asp-based metal-dependent phosphatase
CPL	CTD-phosphatase-like
BRCT	Breast Cancer 1 C-terminal
ABI	ABA-insensitive
ABF	ABA-responsive element-binding factor
GFP	green fluorescent protein
*p*NPP	*para*-nitrophenyl phosphate
*p*NP	*para*-nitrophenol
RT-qPCR	quantitative reverse-transcription PCR
GST	glutathione S-transferase
HRP	horseradish peroxidase
BMP	bone morphogenetic protein
PAP	production of anthocyanin pigment
CBF	C-repeat binding factor

RD29A	responsive to dessication 29A
PAL	phenylalanine ammonia lyase
CHS	chalcone synthase
CHI	chalcone isomerase
F3’H	flavonoid 3’-hydroxylase
DFR	dihydroflavonol reductase
ANS	leucoanthocyanidin dioxygenase
FLS	flavonol synthase
UGT73B2	UDP-glucosyltransferase 73B2
UGT78D2	UDP-glucosyltransferase 78D2
HPLC	High performance liquid chromatography
HPLC-ESI-MS	High performance electrospray ionization mass spectrometry

TABLE OF CONTENTS

Page

ABSTRACT iii

ACKNOWLEDGEMENTS v

NOMENCLATURE vi

TABLE OF CONTENTS viii

LIST OF FIGURES x

LIST OF TABLES xii

CHAPTER

I INTRODUCTION 1

Background 1
Objective of This Work 12

II *ARABIDOPSIS* SCP1-LIKE SMALL PHOSPHATASES DIFFERENTIALLY DEPHOSPHORYLATE RNA POLYMERASE II C-TERMINAL DOMAIN 14

Introduction 14
Materials and Methods 16
Results 24
Discussion 50

III APPLICATION OF CTD PHOSPHATASE MUTANT IN HYPER-ACCUMULATION OF FLAVONOIDS IN TRANSGENIC *ARABIDOPSIS* 54

Introduction 54
Materials and Methods 55
Results 61
Discussion 77

IV CONCLUSIONS 80

Page

REFERENCES 82

VITA 102

LIST OF FIGURES

		Page
Figure 1.1	CTD phosphorylation status and different stages of the transcription cycle	2
Figure 2.1	Alignment of most conserved region of seventeen SSPs	25
Figure 2.2	Phylogenetic relationships among seventeen SSPs	26
Figure 2.3	Alignment of SSP4, SSP4b, SSP5 with SCP1 and FCP1.	27
Figure 2.4	Subcellular localizations of SSP4, SSP4b and SSP5	29
Figure 2.5	Expression profiles of *SSP4*, *SSP4b* and *SSP5* in different tissues	30
Figure 2.6	Phosphatase activities of recombinant SSPs	33
Figure 2.7	Phosphatase activities of SSP4, SSP4b and SSP5 in various conditions	34
Figure 2.8	Kinetic parameters determination for SSP4 and SSP5	36
Figure 2.9	Phosphatase activities of truncated SSPs	38
Figure 2.10	Kinetic parameters determination for SSP4b	40
Figure 2.11	CTD phosphatase activities of SSPs with synthetic phosphopeptide substrates and their specificities	42
Figure 2.12	Activities of SSPs with synthetic non-CTD phosphopeptide substrates	43
Figure 2.13	CTD phosphatase activities of SSPs with GST-CTD-PO_4 substrate in immunoblotting analyses and their substrate specificities	44
Figure 2.14	Dephosphorylation of GST-CTD_O by excessive amount of SSP5	45
Figure 2.15	Identification and characterization of *ssp* T-DNA insertion mutants	47

Page

Figure 2.16 Enhanced ABA sensitivity of *ssp5* single and *ssp4 ssp4b ssp5* triple mutant seed germination 48

Figure 3.1 Constructs for double transformation. 62

Figure 3.2 Relative gene expression in PAP1 and CBF3 transgenic plants 63

Figure 3.3 Design of the three component inducible system 65

Figure 3.4 Purple pigments on non-treated plants 67

Figure 3.5 Pigmentation hyperaccumulation upon cold treatment 68

Figure 3.6 Summary of the flavonoid biosynthesis pathway 70

Figure 3.7 Relative gene expression in PC_{cpl1} #21 and $B3_{cpl1}$ #12 lines under different length of cold treatment determined by qRT-PCR 71

Figure 3.8 Plant flavonoid profiling by HPLC/PDA chromatograms 75

LIST OF TABLES

		Page
Table 2.1	Sequences of primers in the project described in Chapter II.	16
Table 2.2	*SSP* gene expressions with or without ABA treatment	31
Table 2.3	Kinetic parameters of SSP	37
Table 2.4	Gene expressions in wild type and *ssp* triple mutant with or without ABA treatment.....	49
Table 3.1	Sequences of primers in the project described in Chapter III	56
Table 3.2	Anthocyanin content measured as cyanidin-3-glucoside equivalents	64
Table 3.3	Compound identification by LC-MS.....	74

CHAPTER I

INTRODUCTION

BACKGROUND

CTD and gene transcription regulation

C-terminal domain (CTD) of RNA polymerase II (Pol II) largest subunit Rpb1 plays an essential role in gene transcriptional regulation and mRNA processing. CTD has a very unique structure as it contains conserved tandem repeats of heptapeptide $(Y^1S^2P^3T^4S^5P^6S^7)_n$. The serines at position 2, 5 and 7 of CTD undergo reversible phosphorylation and dephosphorylation processes during the transcription cycles (Figure 1.1). Together with the repeat number *n* of the highly conserved heptapeptide (*n* varies from 15 to 52) (Corden, 1990; Chao and Young, 1991), CTD of different phosphorylation stages and site specificity can have up to 4^n different structures.

The complexity of CTD coordinates with dynamic regulation of gene transcriptions by a large variety of kinases and phosphatases. Phosphorylation and dephosphorylation are major post-translational modifications of protein structure and are central mechanisms to regulate cellular machineries. The attachment and removal of a phosphate group from a protein are associated with activation and inactivation of its biological function, and thus provide a molecular switch to promote or inhibit biological processes (Figure 1.1). For example, unphosphorylated Pol II CTD assembles with general transcription factors (GTF) into pre-initiation complexes (Palancade and Bensaude, 2003). Phosphorylation

---

This dissertation follows the style of The Plant Cell.

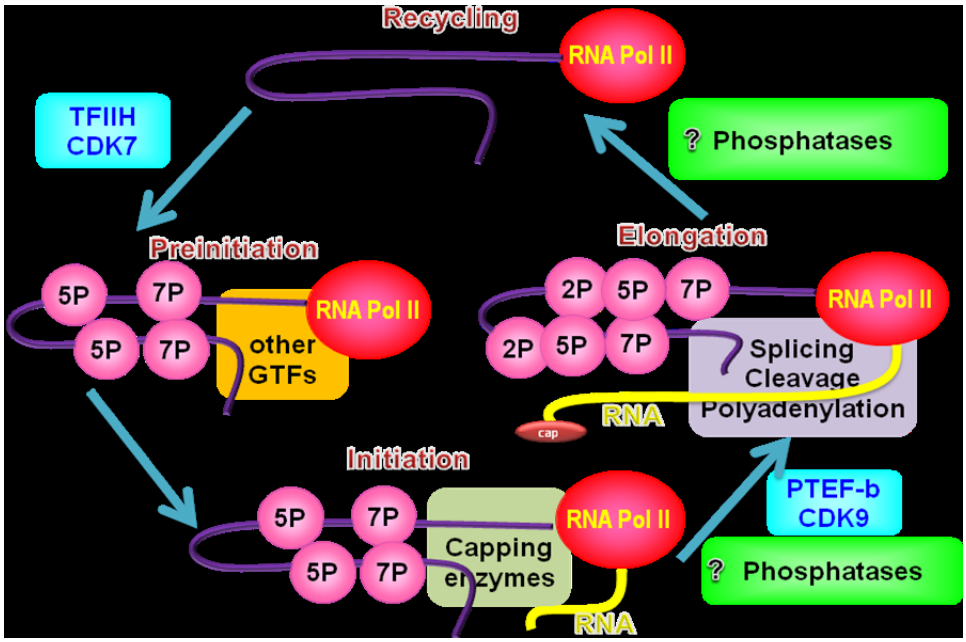

Figure 1.1. CTD phosphorylation status and different stages of the transcription cycle.

Only after Ser5 and Ser7 are phosphorylated by TFIIH kinases, RNA Pol II can enter initiation stage, after which mRNA capping enzyme is recruited and activated. The CTD Ser2 is phosphorylated by CDK9 upon entry into the elongation phase. During the transition from the early elongation to the productive elongation, and at the transcription termination, various CTD phosphatases dephosphorylate CTD and help advancing transcription cycles.

of Ser5 and Ser7 by TFIIH subunits CDK7/kin28 (metazoan/yeast) occurs before gene transcription begins in the initiation complex (Schroeder et al., 2000), whereas Ser2 is phosphorylated by CDK9/CTDK-1 (metazoan/yeast) upon entry into elongation (Lee and Greenleaf, 1997). While the CTD kinases promote gene transcription through serine phosphorylation, CTD phosphatases function reversely to remove phosphate from serine to maintain phosphorylation equilibrium as well as prepare Pol II to be recycled for next transcription cycle (Lin et al., 2002).

CTD phosphatases

Several classes of CTD phosphatases have been reported to be aspartic acid based metal-dependent phosphatases (AMPs) (Thaller et al., 1998). The signature motif of the AMPs is ψψψDXDX(T/V)ψψ (ψ represents a hydrophobic residue) with the first aspartic acid as a phosphoryl acceptor and the second stabilizing the leaving phosphate group (Kobor et al., 1999; Wang et al., 2002). The ψψψDXDX(T/V)ψψ motif is present in a number of acid phosphatases including vegetative storage proteins and several metabolic enzymes. FCP1 (TFIIF-interacting CTD-phosphatase) is the most characterized CTD phosphatase in recent decades. FCP1 has an essential breast cancer 1 C-terminus domain (BRCT) and a phosphatase catalytic domain (Hausmann and Shuman, 2002). FCP1 is highly conserved among eukaryotes and dephosphorylates Ser-PO_4 in the heptad repeat sequence of Pol II in vivo in yeast and metazoan and phosphopeptides in vitro (Suh et al., 2005) (Archambault et al., 1997; Archambault et al., 1998; Kobor et al., 1999). The yeast Fcp1 is essential for the survival, whereas a single nucleotide

substitution in human *FCP1* causes with congenital cataracts facial dysmorphism neuropathy (CCFDN) syndrome (Varon et al., 2003). In CCFDN patient cells, a single C to T substitution adds a new splicing donor site and causes in an insertion of 95 nucleotide sequences. This insertion results in premature termination and produce non-functional transcript. The wild type transcript in cells collected from CCFDN patients reaches only 15-35% of the level in control cells (Varon et al., 2003).

In addition to prototypical FCP1-family proteins, a number of small acid phosphatases homologous to FCP1 have been identified without any additional domains in various organisms. SCP1 family proteins containing only a FCP1-like catalytic domain have been discovered. In human, SCP1-SCP3 were first identified as CTD phosphatases that prefer Ser5-PO_4 of the pol II CTD heptad repeats as substrates (Yeo et al., 2003b). in vivo, SCP1 functions in silencing neuronal gene expression in non-neuronal cells presumably by interacting with repressor element 1 (RE1)-silencing transcription factor/neuron-restrictive silencer factor (REST/NRSF) complex, and phosphatase inactive forms of SCP1 promotes P19 stem cells in neuronal differentiation (Yeo et al., 2005; Visvanathan et al., 2007; Yeo and Lin, 2007). Another SCP homolog, SCP2 appears to function in similar pathway, and overexpression of SCP2/Os4 inhibits BMP signaling in xenopus embryo (Zohn and Brivanlou, 2001). SCP3 is identical to tumor suppressor HYA22A and overexpression of SCP3 decreases the level of phosphorylated RB (retinoblastoma) protein. In yeasts, membrane-bound acid phosphatases Psr1/2 and Nem1 regulate salt tolerance and nuclear morphology, respectively, and the latter has been shown to dephosphorylate phosphatidic acid phosphatase Smp2p. All SCP1, 2 and

3 but not FCP1 enhance transforming growth factor-β (TGF-β) signaling by dephosphorylating linker and N terminus region of Smad2/3 transcription factors (Sapkota et al., 2006; Wrighton et al., 2006). SCP1, 2 and 3 also mediate Smad1 linker as well as C-terminal region dephosphorylation in bone morphogenetic protein (BMP) signaling pathway and suppress BMP transcriptional responses (Knockaert et al., 2006; Sapkota et al., 2006). These studies suggested that SCP family proteins are associated with various biological processes in animals and regulate diverse cellular processes including cell recognition, proliferation, differentiation and apoptosis during embryogenesis as well as in mature tissues; however, specific roles of each SCP isoform have not been fully established, even in the animal systems.

Interestingly, FCP1 and SCP1 family proteins have distinct substrate preferences, despite of their high homology between the two families. FCP1 family proteins can dephosphorylate both Ser2-PO_4 and Ser5-PO_4 of CTD, with 6-fold preference of Ser2-PO_4 over Ser5-PO_4 (Hausmann et al., 2004; Ghosh et al., 2008). SCP1 family proteins have preference towards Ser5-PO_4 as exemplified by human SCP1, which is 70 fold more active at Ser5-PO_4 than Ser2-PO_4 (Zhang et al., 2006). Structural bases of distinct substrate specificities for yeast Fcp1 and human SCP1 were proposed based on three-dimensional models obtained by X-ray crystallographic data (Hausmann et al., 2004; Zhang et al., 2006; Ghosh et al., 2008). In the model for SCP1-CTD complex, Pro3 of CTD heptad is located inside of a hydrophobic pocket in the SCP1 active site and present Ser5-PO_4 for dephosphorylation. In contrast, a helical insertion and a BRCT domain in Fcp1 guided CTD substrate in a distinct path to an orthogonal orientation

compared with the substrate orientation found in SCP1-CTD complex. Both of these differences make its hydrophobic pocket inaccessible for CTD, and promote CTD conformation for Ser2-PO_4 dephosphorylation (Zhang et al., 2010). It's intriguing that in animals and fungi, CTD phosphatases with distinct substrate specificities toward Ser2-PO_4 and Ser5-PO_4 play diverse roles during the transcription elongation.

Plant CTD phosphatase

In Arabidopsis, classical protein phosphatases are classified into three distinct families (Luan, 2003). The PPP (serine/threonine-specific phosphoprotein phosphatase) and the PPM (metal-dependent protein phosphatases) families consist of Ser/Thr phosphatases and the PTP family consists of tyrosine specific phosphatases. Previous biochemical and molecular studies identified diverse functions of these phosphatases, including environmental and hormonal signaling, and development (Bellec et al., 2002; Camilleri et al., 2002; Fordham-Skelton et al., 2002; Monroe-Augustus et al., 2003; Schweighofer et al., 2004; Quettier et al., 2006; Pedmale and Liscum, 2007).

The plant homologs of FCP1 which contain catalytic FCP homology domain (FCPH) and BRCT domain are CPL3 and CPL4. The CPL3-like family is the closest to FCP1 and contains a C-terminal BRCT domain (Koiwa et al., 2002; Bang et al., 2006). The CTD phosphatase-like 1 (CPL1) family contains double-stranded-RNA binding domains in the C-terminus and is unique to plants (Koiwa et al., 2002; Koiwa et al., 2004). CPL1 and CPL2 have the same DXDXT motif as FCP1 that act metal-dependently via an aspartyl-phosphoenzyme intermediate, but they lack BRCT domain. Both CPL1 and

CPL2 preferentially dephosphorylate Ser5 of the CTD arrays in vitro. Families of SCP1-like small phosphatase also exist in planta, and they contain only a phosphatase catalytic domain. They were termed SSP (*S*CP1-like *s*mall *p*hosphatase) and they are most homologous to animal SCP1 (Bang et al., 2006).

There haven't been reports on the role of single catalytic domain containing plant phosphatases; however, biological functions of CPL family proteins have been characterized. They appeared to have critical functions in plant growth, development, and environmental responses (Koiwa et al., 2004). CPL1 and CPL2 form an essential paralogous gene pair (Koiwa et al., 2004). CPL4, too, is essential for normal plant development (Bang et al., 2006). Interestingly, both *cpl1* and *cpl3* mutations resulted in osmotic stress/ABA hypersensitivity (Koiwa et al., 2002), a phenotype often associated with mutations in RNA metabolism pathways. *cpl2* single mutation conferred auxin-insensitivity (Ueda et al., 2008b), whereas RNAi of CPL4 led to plant growth defects (Bang et al., 2006).

ABA signaling pathway

ABA is a central mediator of plant osmotic stress tolerance (Himmelbach et al., 2003) as well as a signal intermediate for glucose sensing pathways (Arenas-Huertero et al., 2000). In Arabidopsis, ABA is perceived by a family of intracellular receptor proteins, which regulate type 2C phosphatases, such as ABI1, ABI2 and HAB1 (Ma et al., 2009; Park et al., 2009). These phosphatases in turn regulate the activity of SNF1-related protein kinases 2 (SnRK2) family proteins (Fujii et al., 2009; Vlad et al., 2009).

Transcription activation of ABA-responsive genes is mediated by bZIP type transcription factors ABFs, such as ABI5 (Choi et al., 1999; Uno et al., 2000; Narusaka et al., 2003), and an AP2-type transcription factor ABI4 (Finkelstein et al., 1998).

In addition to the central pathway, ABA responses are modulated by RNA metabolism proteins (Kuhn and Schroeder, 2003). Besides of CPL family proteins, dsRNA-binding protein HYL1 (Lu and Fedoroff, 2000), mRNA cap binding proteins CBP80/ABH1 (Hugouvieux et al., 2001b) and CBP20 (Papp et al., 2004), Sm-like protein SAD1 (Xiong et al., 2001), and RNA helicases (Gong et al., 2002) all contribute to regulate the expression of osmotic-stress/ABA responsive genes. However, the extent of overlap between RNA metabolism and ABA signaling components, as well as the mechanisms that integrate these pathways, are not fully understood.

Plant CTD phosphatase mutant

Among four AtCPL, CPL1 is the most studied gene. The growth of *cpl1-2* is similar to wild type under normal condition, but upon cold/ABA/NaCl treatment, *cpl1-2* expresses 10-20 times higher expression of RD29A promoter. Thus *cpl1-2*/*fry2-1* mutation is an enhancer of RD29A promoter expression (Koiwa et al., 2002; Xiong et al., 2002). This makes *cpl1-2* mutant as an attractive host for engineering inducible gene expression system (Koiwa et al., 2002; Xiong et al., 2002; Koiwa et al., 2004).

Inducible gene expression systems have been used to generate and study transgenic plants with genes whose constitutive expression is detrimental to the host plants (Guo *et al.*, 2003; Zuo *et al.*, 2006) . The most popular expression systems are promoters

activated by synthetic transcription factors co-expressed in the transgenic plants (Guo *et al.*, 2003; Zuo *et al.*, 2006) and native plant promoters activated by various environmental stimuli (Kasuga *et al.*, 1999) . Typically, the former systems can strongly induce transgenes but require application of chemical inducers, whereas the expression levels achieved by the latter are weaker. Enhancement of induction of plant promoter has been reported by Kasuga et al (Kasuga *et al.*, 1999), where dehydration/cold/salt-inducible *RD29A* promoter was used to drive *DREB1a/CBF3* transcription factor cDNA, which can activate, in addition to downstream stress tolerance determinants, *RD29A-DREB1a/CBF3* expression itself. A single-component, self-activation loop of *RD29A-CBF3* was sufficient to induce expressions of *CBF3* and cold-tolerance determinants specifically under low temperature.

Engineering of flavonoid production

Transgenic engineering of crop plants for enhancement of anthocyanin and other flavonoids is one of the current foci of plant biotechnology to produce health-promoting functional foods. Flavonoid are a big family of natural compounds that are produced in all tissues of higher plants, including leaves, stems, roots, flowers and fruits. They function in plant defense system including physical barriers and biochemical signaling (Dixon and Paiva, 1995; Shadle et al., 2003). Because their structures have multiple unsaturated double bond and hydroxyl group, they function as natural anti-oxidants in cells and thus are referred to as healthy phytochemicals providing capacity to scavenge free-radicals and preventing chronic degenerative diseases, like cancer, aging and

inflammations (Korkina, 2007; Iriti and Faoro, 2009). Phenylpropanoids may also function in visual capacity, brain cognitive function, obesity prevention, ulcer protection, and lowering cardiovascular risk (Galvano *et al.*, 2004). Anthocyanins are flavonoid pigments whose production is regulated by both developmental and environmental signals. Different level of anthocyanins and other flavonoids are produced under high light (Steyn *et al.*, 2009; Shi and Xie, 2010), salt stress (Piao *et al.*, 2001; Wahid and Ghazanfar, 2006; Keutgen and Pawelzik, 2007), nutrient starvation (Sanchez-Calderon *et al.*, 2006; Lea et al., 2007; Lillo et al., 2008; Lunde et al., 2008; Hsieh et al., 2009; Shi and Xie, 2010), and cold stress (Solecka et al., 1999; Lo Piero et al., 2005; Pan et al., 2008; Usadel et al., 2008; Toda et al., 2010).

Many gene families have been reported to regulate phenylpropanoid accumulation, including phenylpropanoid biosynthesis pathway rate limiting genes (Ferrer *et al.*, 2008), and transcription factors (Vom Endt *et al.*, 2002). Production of anthocyanin pigment 1-Dominant (pap1-D)' was found to be able to enhance the accumulation of purple pigment (Borevitz et al., 2000). Further study indicated that *Arabidopsis* PAP1 is a regulator that regulates the expression of multiple enzymes in phenylpropanoid biosynthetic pathway, especially in the branch of anthocyanin production (Tohge *et al.*, 2005). Overexpression of PAP1 in plants results in specific hyperaccumulation of certain cyaniding and quercetin derivatives but may cause retardation of vegetative growth (Tohge *et al.*, 2005).

Constitutive overproduction of anthocyanins was achieved in several studies using PAP1 and other related transcription factors (Borevitz et al., 2000; Verpoorte and

Memelink, 2002). On the other hand, anthocyanin and flavonoid accumulation varied substantially according to the growth condition, and cross-talk between PAP1-activation and endogenous regulation mechanism of anthocyanin production has not been fully understood. Furthermore, high level of anthocyanin/flavonoids could be inhibitory to plant growth (Geekiyanage *et al.*, 2007), likely due to the interference of auxin transport by flavonoids (Brown et al., 2001; Peer et al., 2001; Peer et al., 2004). In order to achieve high-level of anthocyanin production without causing growth defects, it is desirable to employ inducible production of phytochemicals that separates growth phase and production phase, the latter could be initiated by chemical or environmental stimuli, like cold treatment.

Low temperature treatment has been reported to increase phenylpropanoid accumulation in plant tissues, suggesting the role of phenylpropanoid in cold acclimation. Cold treatment (4˚C) increased major anthocyanin contents in *Arabidopsis* leaves of the ecotypes with better tolerant towards cold stress (Marczak *et al.*, 2008). Cold temperature increased yield of soluble phenylpropanoid in winter oilseed rape leaves, especially when the tissues were subjected to a period of cold acclimation prior to treatment (Solecka *et al.*, 1999). Low temperature induced anthocyanin accumulation were also reported in vegetable and fruits. Red color appeared quickly after low temperature in apple and pears peels and induced anthocyanins served a photoprotective function (Steyn *et al.*, 2009). Seventy-five days of low temperature (4˚C) storage induced 8 times higher anthocyanins in red orange juice vesicles, as well as 40 times more expression of phenylpropanoid biosynthesis pathway genes compared with normal

temperature (25˚C) storage (Lo Piero *et al.*, 2005). Transcriptomic and metabolomic studies have shown both phenylpropanoid biosynthesis and metabolism have been induced by six hours of cold treatment (Usadel *et al.*, 2008).

Majority of the plant cold stress response including the phenylpropanoid accumulation are orchestrated by CBF family transcription factors (Kaplan *et al.*, 2004; Hannah *et al.*, 2006). CBF1-3, together with upstream ICE1 transcription factor, globally regulate plant protection against cold stress, and are critical for plant cold acclimation (Zhu *et al.*, 2007). Ectopic expression of CBF genes in plants in normal temperature results in similar metabolic response as cold treated plants (Cook *et al.*, 2004). RD29A is a target gene of CBF3, and expression of CBF3 driven by inducible RD29A promoter has been shown to enhance stress tolerance without compromising growth whereas constitutive expression of CBF3 can cause growth retardation as well as conferring stress tolerance. (Koiwa et al., 2002; Xiong et al., 2002).

OBJECTIVE OF THIS WORK

There have been no reports on the role of single catalytic domain containing plant phosphatases. Understanding the biochemical functions of the small CTD phosphatase family will present valuable information to elucidate RNA pol II transcription regulation. CTD substrate specificities for Ser2-PO_4 and/or Ser5-PO_4 will provide insights into phosphatases' interaction with pol II CTD in each particular stage of the transcription cycle. Moreover, with the help of loss of function mutants, SSP biological function could be evaluated. Furthermore, utilization of the already characterized CTD phosphatase mutant

cpl1-2 and its hypersensitive phenotype to environmental stimuli opens a new way to construct inducible systems that respond to those stimuli and help plant engineering to promote certain aspect of plant growth and development.

CHAPTER II

ARABIDOPSIS SCP1-LIKE SMALL PHOSPHATASES DIFFERENTIALLY DEPHOSPHORYLATE RNA POLYMERASE II C-TERMINAL DOMAIN*

INTRODUCTION

The carboxyl terminal domain (CTD) of the largest subunit of eukaryotic RNA polymerase II (pol II) contains conserved sequences of tandem heptad repeats $(Y^1S^2P^3T^4S^5P^6S^7)_n$. The number of repetitions correlates with evolutional complexity and ranges from 15 in a microsporidian parasite, *Encephalitozoon Cuniculi*, 26 in budding yeast, 34 in *Arabidopsis* to 52 in human (Hausmann et al., 2004; Phatnani and Greenleaf, 2006). The CTD undergoes reversible phosphorylation and dephosphorylation at its serine residues during the transcription cycles and recruits various RNA processing factors depending on its distinct phosphorylation status at Ser2 and Ser5. Indeed, the phosphorylation status of pol II CTD is a hallmark for different transcription stages of pol II complexes (Phatnani and Greenleaf, 2006).

Dephosphorylation of CTD is catalyzed by CTD phosphatases, many of which belong to the family of Asp-based metal-dependent phosphatases (AMPs) with ψψψDXDX(T/V)ψψ (ψ represents a hydrophobic residue) as their signature

---

* Part of this chapter is reprinted with permission from "*Arabidopsis* SCP1-like small phosphatases differentially dephosphorylate RNA polymerase II C-terminal domain" by **Feng, Y., Kang J.S., Kim S., Yun D.J., Lee S.Y., Bahk J.D., and Koiwa H.** (2010). Biochem. Biophys. Res. Commun. **397,** 355-60 © 2010 with kind permission from Elsevier B.V.

motif (Thaller et al., 1998). The prototypical CTD phosphatase FCP1, containing a phosphatase catalytic domain and a BRCT (*Br*east *C*ancer 1 *C-t*erminal) domain, is highly conserved among eukaryotes (Archambault et al., 1997; Archambault et al., 1998; Kobor et al., 1999). A large number of FCP1-like phosphatase genes exist in the genome of *Arabidopsis thaliana* (Koiwa, 2006). These phosphatases are classified based on domain structures, namely, CPL1-like, CPL3-like, and SSPs (Bang et al., 2006). SSP (*S*CP1-like *s*mall *p*hosphatase) family proteins consist solely of a phosphatase catalytic domain and thus resemble animal SCP1 (Bang et al., 2006). The understanding on specific roles for each plant SSP is still in its infancy (Ji et al., 2010) and to date, no CTD substrate specificity of any plant SSP family phosphatases were reported.

In this chapter, in vitro and in vivo characterizations of highly homologous SSP isoforms SSP4, SSP4b, and SSP5 are described. SSP isoforms were ubiquitously expressed, but exhibited different subcellular localizations. Biochemical analyses of recombinant SSP proteins showed their highly specific CTD phosphatase activities with distinct Ser-PO_4 position preferences. SSP4 and SSP4b harbored both Ser2-PO_4 and Ser5-PO_4 phosphatase activity, while SSP5 was only capable of Ser5-PO_4 dephosphorylation. Analyses of single and combined *ssp* mutant plants indicated that SSPs redundantly function in ABA and glucose signaling during seed germination and implied the importance of CTD Ser5 in the response of seeds to ABA.

MATERIALS AND METHODS

Preparation of SSP cDNA clones and GFP expression constructs

Table 2.1. Sequences of primers used in the project described in Chapter II.

Name	Sequences (5'—>3')
Y854	AAAAAGCAGGCTCCATGCCATATATAGAAATGAAGAGC
Y855	AGAAAGCTGGGTTTTCCTTGTTACCAAATCTCC
Y856	AAAAAGCAGGCTCCATGGAAGATGATGACTCTCC
Y857	AGAAAGCTGGGTTCCTTTCGAAAGGATCTCCCG
Y858	GTTGTTGTATTTCTAGAGACT
Y859	CTTAGACAAAACCTTAAGAGTC
Y860	AAAAAGCAGGCTCCATGCCAATGCCATTTTTGAA
Y861	AGAAAGCTGGGTTATCAAAAGAACTTCTTGTTG
619	ACCCGGGGCATGCCATATATAGAAATGAAGAGCA
548	CCGCTCGAGTTATTCCTTGTTACCAAATCTCCT
Y904	GACCTCCCGGGAAATGCCATTTTTGAAAATGA
Y905	GCGATGAGCTCTTACGCCTAATCAAAAGAACTTC
623	ACCCGGGGCATGGAAGATGATGACT
550	CCGCTCGAGTTACCTTTCGAAAGGATCTCC
Y927	GATGGTTCTCAATCTGTCGAAA
Y928	TGTAAATACTCGACCAGTTCCAT
Y929	CAAGGACTCGTTAAGAGAGAAA
Y930	CCTAGTAGAAGAATCCAAACCTATA
Y923	GATTGCTTGCTCTAGATACAAGTA
Y924	GAAGTAATAAGGGAATCTTGACCA
S29	AGCATCTATGCCTCACAACTT
S28	ATTTATTTGCAATCTGTACAC
Y951	CAACTGATGATGGGTTGATTACTA
Y952	CAACTGATGATGGTCTGATTACAT
Y953	CATCAGATAAAGAACTCCTTCACT
Y899	TGATAGGCCGAACATCATCA
Y900	TGACAGGACGAACATCGTTG
Y897	GATGTGGTATTCAATTATTTCC
Y903	TATTGGCCTCGGCAAACCCGAAGCTGTCCTCCTA
Y940	AGCTGGGTTCTCCCTTAGATTGAATTTCTTGGC
M916	TGTGGACAGTTGTCAACTAT

Table 2.1. Continued.

Y937	CGGACAGTTGTCAACAATGATAACACGAGATAAATC
K64	ATATTGACCATCATACTCATTGC
188	TGGTTCACGTAGTGGGCCATCG
Y901	AAAAAGCAGGCTCCGATGAGAGTGGAAACAACGA
Y865	ACGTTGTAAAACGACGGCCAG
Y914	CCAGGAAACAGCTATGACCATG
Y935	ATCAAAGTCTTCGGAGCCTGCTTTTTTGTACAAAGT
Y934	AGCAGGCTCCGAAGACTTTGATCCTCAGCT
Y938	TCTTAGGGTATCGGAGCCTGCTTTTTTGTACAAAGT
Y939	AAAGCAGGCTCCGATACCCTAAGAAAGAAGTCTGTG
Y949	TTCGGAGAGTCGGAGCCTGCTTTTTTGTACAAAGT
Y948	AAAGCAGGCTCCGACTCTCCGAAAAGGAAGGCGGTA
Y947	GCTTCGGGTTTGGGAGCCTGCTTTTTTGTACAAAGT
Y946	AAAGCAGGCTCCCAAACCCGAAGCTGTCCTCCTATT
Y980	CGGCTGTTTCAGACTATCTT
Y981	ACTCCTCCAGTTTCTTCTTTG
Y984	AAAGAATCAGGAGACGGAGA
Y985	AACACCTAAGTGGGATGTCA
Y986	TCACCTGATGAAAGCAACTG
Y987	GGCGATCAAAGTGTCTGTAG
Y988	TGGGTTCTTCACAGGGATAT
Y989	CCATGTAGCAAATACCTGATGA
Y992	GGAAGAGGAAGCAACAGTATTT
Y993	TCCGACTTCTTCTCCATCTT
Y994	GTGCTGTCCTTATCATTGACT
Y995	GTAGTGGCATCCATCTTGTTA
E155	GTTCTCACTGGTATGGCTTC
E156	GATGTTGCCGTCACCTTT
E161	TTCCCTCAGCACATTCCA
E162	CCCATTCATAAAACCCCAGC

Primer sequences used in this research are listed in Table 2.1. *SSP4* cDNA clone was obtained from the Arabidopsis Biological Resource Center (U19564) (Yamada et al., 2003) and prepared by PCR with primer pairs [619, 548]. A cDNA fragment for *SSP5* was obtained by RT-PCR with [623, 550]. Both cDNA fragments were cloned into pBCSK vector (Agilent Technologies, TX) to produce pBCSSP4 and pBCSSP5. cDNA fragment for *SSP4b* was prepared by nested RT-PCR with [Y858, Y859] for the first round PCR and with [Y860, Y861] for the second, and cloned into pDONR201 using BP clonase (Life Technologies, CA) to produce pDonSSP4b.

In order to prepare GFP-fusion constructs, pBCSSP4/SSP5 were first digested with SmaI/XhoI and ligated with pET44a fragments digested with SmaI/XhoI. SSP fragments were then excised from pET44a-SSP4/SSP5 by digestion with SmaI/NotI, and ligated with a pEnEOTG fragment digested with SmaI/NotI to produce pENSSP4 and pENSSP5. For SSP4b, cDNA fragment was amplified by PCR using [Y904, Y905], digested by SmaI/SacI and ligated with a pEnEOTG fragment digested with SmaI/SacI to produce pENSSP4b.

Protoplast transient expression of GFP-SSP

Preparation of *Arabidopsis* protoplasts and transformation by polyethylene glycol were performed as described previously (Koiwa et al., 2004). Ten microgram pENSSP4, pENSSP4b or pENSSP5 plasmids were transformed into 2×10^5 protoplasts. The protoplasts were then incubated for 16 hr at 23°C in dark, and examined using a BX51 epi-fluorescence microscope (Olympus, CA).

Gene expression analyses by RT-qPCR

For *SSP* expression in different tissues, total RNA was isolated from flowers, rosette leaves, cauline leaves, stems, seeding shoots and seeding roots of 9-day-old *Arabidopsis* using Trizol reagent (Life Technologies). For ABA responses, total RNA was obtained from 9-day-old seedlings 3 hr after the spray application of 100 μM ABA. RNA samples resuspended in 50 μl were treated with 7.5 unit of DNase I (Qiagen, MD) for 60 min at 37°C, and re-purified with the RNeasy plant mini kit (Qiagen). Quantitative reverse-transcription PCR (RT-qPCR) was performed as described previously (Ueda et al., 2008a). Absence of genomic DNA contamination was confirmed using minus-reverse-transcriptase controls. The data were processed as described previously (Chi et al., 2009). Specific primer pairs for RT-qPCR analyses were: [Y980, Y981] for RD29A, [E155, E156] for COR15A, [Y984, Y985] for ABF2, [Y992, Y993] for ABI5, [Y988, Y989] for CAK4, [Y986, Y987] for CDKC2, [Y994, Y995] for EF1α, and [E161, E162] for ACT2.

Preparation of recombinant SSP proteins

To produce NusA-SSP fusion proteins, cDNA fragments encoding full-length SSP4 and SSP5 coding sequences were amplified with primer pairs [Y854, Y855] and [Y856, Y857], respectively. Resulting PCR products were cloned into pDONR201 using BP clonase to produce pDonSSP4/SSP5. A protein expression vector pET44a (EMD chemicals, MD) was modified by inserting a Gateway (GW) cassette and a strep-tagII sequence between EcoRI-SalI sites. pDonSSP clones were recombined with

pET44aGW-strep using LR clonase (Life Technologies) to produce pETSSP clones. For expression of truncated SSP except $SSP4^{172-453}$, overlapping PCR were performed with pDonSSP templates and primer pairs [Y865, Y938] and [Y939, Y914] for $SSP4^{274-453}$, [Y865, Y949] and [Y948, Y914] for $SSP4b^{262-446}$, and [Y865, Y947] and [Y946, Y914] for $SSP5^{105-305}$ to generate attL1-(truncated SSP)-attL2 fragments, which were then recombined with pET44aGW-strep using LR clonase. $pETSSP4^{172-453}$ was prepared by the same way as full length SSP4, namely, amplified with primer pairs [Y901, Y855], recombined into pDONR201 and then into pET44aGW-strep.

Purification of SSP proteins

pETSSP4, 4b and 5 were transformed into *Escherichia coli* BL21SR (Koiwa et al., 2004). Recombinant SSP proteins were induced by 0.4 mM isopropyl β-d-thiogalactoside for 16 hr at 18°C in 200 ml LB media containing 70 μg/μl ticarcillin and 20 μg/μl chloramphenicol. Cells were harvested by centrifugation at 4000 x g for 20 min at 4°C, and the cell pellets were resuspended in 8 ml buffer A (20 mM Tris·HCl pH7.9, 500 mM NaCl). Cells were lysed by sonication on ice, and the lysate was cleared by centrifugation at 10000 x g for 20 min at 4°C. The supernatant fractions containing soluble recombinant SSPs were applied to Ni^{2+}-NTA column (IMAC Sepharose 6 fast flow, GE Healthcare) (bead volume 1 ml) equilibrated with buffer A. The column was washed with 10-column volumes of buffer A and then 10-column volumes of washing buffer (buffer A + 40 mM imidazol). The bound recombinant proteins were eluted with 100 mM imidazol in buffer A. The eluates were combined and loaded onto StrepTactin-

sepharose column (bead volume 1 ml) (GE Healthcare) equilibrated with buffer B (20 mM Tris·HCl pH7.9, 300 mM NaCl). After washing with 10-column volumes of buffer B, the recombinant proteins were eluted with 2.5 mM desthiobiotin (Life Technologies) in buffer B. Glycerol (final concentration 10 %) was added to the protein eluates, which were then kept frozen at -80°C until use. Purity of recombinant SSP4, SSP4b and SSP5 was assessed by 7.5% SDS-PAGE.

Phosphatase assays

Phosphatase assays of SSPs with *para*-nitrophenol phosphate (*p*NPP) (Sigma, MO) were performed under the conditions described previously (Hausmann et al., 2005), and 1 unit (U) of activity was defined as producing 1 nmol *p*-nitrophenol/min in the standard conditions (100 µl reaction mixture containing 100 mM Tris·acetate pH5.5, 10 mM $MgCl_2$, and 10 mM *p*NPP, incubated for 5 min at 37 °C). To determine the optimal pH and metal dependency, 2.32 U SSP4, 1.12 U SSP4b or 4.90 U SSP5 were used in standard conditions except with different pH or metal ion concentrations as indicated. To determine kinetic parameters, 2.90 U SSP4, 1.74 U SSP4b or 5.39 U SSP5 were assayed in standard conditions with varying *p*NPP concentrations. At the end of incubation, reactions were stopped by adding 0.9 ml of 1 M Na_2CO_3. Released *p*-nitrophenol (*p*NP) was measured as ABS_{410} and the amount was calculated according to a *p*NP standard curve. Data at each concentration were collected in triplicate and fit to the Michaelis-Menten equation ($v_0 = V_{max}[S]/K_m + [S]$) using the nonlinear least-squares-fitting.

For assays with phosphopeptide substrates, CTD phosphopeptides were synthesized by EZ Biolab (IN), and standard phosphopeptides pS [RRApSVA] and pY [RRLIEDAEpYAARG] were obtained from Upstate (NY). Reaction mixtures (25 µl) containing 50 mM Tris·acetate pH 7.0, 10 mM $MgCl_2$, 25 µM phosphopeptide, and SSP were incubated for 30 min at 37 °C. The reactions were quenched by adding 0.3 ml malachite green reagent (BIOMOL Research Laboratories, PA). Release of phosphate was determined by measuring ABS_{650} and the values were interpolated to a phosphate standard curve.

Preparation of phosphorylated GST-CTD and dephosphorylation assays by SSP were performed as described previously (Koiwa et al., 2004). Phosphorylation status was determined by immunoblotting analysis with H5 (anti-CTD-Ser2-PO_4) or H14 (anti-CTD-Ser5-PO_4) monoclonal antibodies (Covance, NJ). Total GST-CTD was detected by immunoblotting using anti-GST-HRP conjugate (GE Healthcare, NJ).

Identification and molecular genetic analyses of SSP mutants

T-DNA insertion lines *ssp4-1* (SALK_087127) and *ssp4b-1* (SALK_151923) were provided by the *Arabidopsis* Biological Resource Center (Alonso et al., 2003). *ssp4-2* (GABI_224B07) and *ssp5-1* (GABI_405H06) were obtained from Genomanalyse im biologischen System Pflanze via European *Arabidopsis* Stock Centre (NASC) (Scholl et al., 2000; Rosso et al., 2003). The homozygous mutants were identified by PCR-based diagnoses and the triple mutant *ssp4-1 ssp4b-1 ssp5-1* was prepared by genetic crosses. PCR was used to determine T-DNA insertion alleles with primer pairs as following:

[S29, S28] for wild type *SSP4*, [S29, 188] for *ssp4-1* and [Y916, K64] for *ssp4-2*; [Y897, M916] for wild type *SSP4b* and [Y897, 188] for *ssp4b-1*; [Y903, Y937] for wild type *SSP5* and [Y903, K64] for *ssp5-1*.

For analyses of *SSP* transcripts in T-DNA insertion lines, RT-PCR was performed. Primer pairs used in the analyses were as follows. For SSP4, [Y927, Y928] were used for 5' region, [S29, 188] for *ssp4-1* chimeric transcript (cDNA-T-DNA), [S29, K64] for *ssp4-2* chimeric transcript, and [Y951, Y899] for 3' region. For *SSP4b*, [Y929, Y930] were used for 5' region, [Y897, 188] for *ssp4b-1* chimeric transcript, and [Y952, Y900] for 3' region. For *SSP5*, [Y923, Y924] were used for 5' region, [Y903, K64] for *ssp5-1* chimeric transcript, and [Y953, Y940] for 3' region. Mock cDNA templates prepared without reverse transcriptase were used as negative controls.

Stress response assays

Germination responses to exogenous chemicals were measured by placing imbibed seeds on 1/4 × MS media (pH 5.7) containing NaCl (0–140 mM), mannitol (0–300 mM) or ABA (0–1.5 μM). After stratification for 3 days at 4°C, seeds were transferred to room temperature (22°C) and the successful germination was recorded after two weeks as the rate of seedlings with fully expanded cotyledons.

RESULTS

Bioinformatic identification of SSP4, SSP5 and SSP4b

The *Arabidopsis* genome has 17 single-domain acid phosphatases homologous to SCP1 family proteins. These proteins were named *S*CP1-like *s*mall *p*hosphatase 1-17 (SSP1-17) because they contain the same single ψψψDXDX(T/V)ψψ motif (ψ represents hydrophobic residue) (Figure 2.1), which was previously shown as a Mg^{2+}-binding catalytic motif of CTD phosphatases (Hausmann and Shuman, 2002). All SSP shared high homology among themselves in the family in the predicted phosphatase catalytic domains (Figure 2.2). Additional search in *Arabidopsis* sequence databases identified a full-length cDNA (GenBank accession #AK228630) that encodes a protein containing LVLDLDETLV at amino acid position 272-281. Because of the high degree of identity with SSP4 (see below), this protein was termed SSP4b (Figure 2.3). SSP4b is encoded by the At4g18140 locus; however, TAIR genome annotation (ver. 8) of At4g18140 lacks the C-terminal catalytic domain because of a frame-shift at position 1325, caused by a deletion of single thymine, upstream of the catalytic motif. The sequences of SSP4b cDNA and genomic DNA from Col-0 wild type were experimentally determined and they were consistent with the AK228630 entry. Therefore, At4g18140 encodes a SSP family phosphatase, SSP4b.

Among eighteen SSP members, SSP4 (At5g46410), SSP4b (At4g18140) and SSP5 (At5g11860) are unique as they contain long N-terminal extensions of 280, 267 and 110 residues. They are also closest in terms of evolutionary connections and relationships (Figure 2.2). Amino acid sequences of SSP4, SSP4b and SSP5 were aligned with the

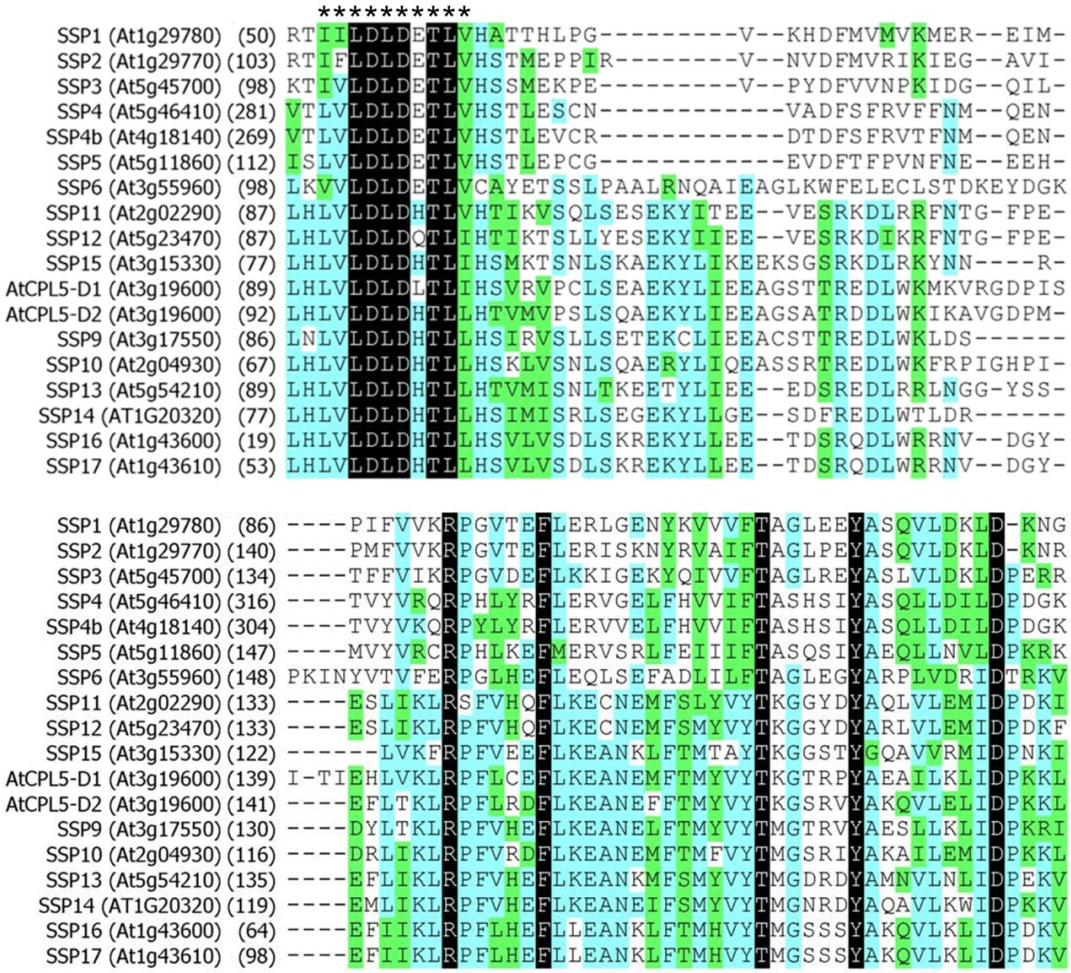

Figure 2.1. Alignment of most conserved region of seventeen SSPs.

Shaded boxes in black represent the identical residues among all SSPs. Shaded boxes in blue represent the identical residues among individual SSPs. Shaded boxes in green represent the conserved residues. Asterisks indicate the conserved acid phosphatase motif ψψψDXDX(T/V)ψψ.

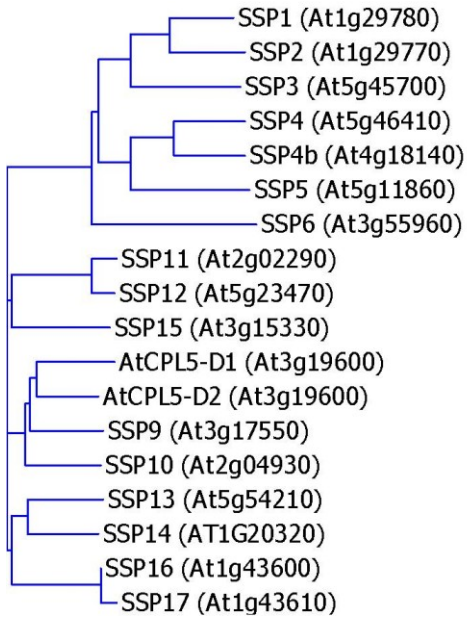

Figure 2.2. Phylogenetic relationships among seventeen SSPs.

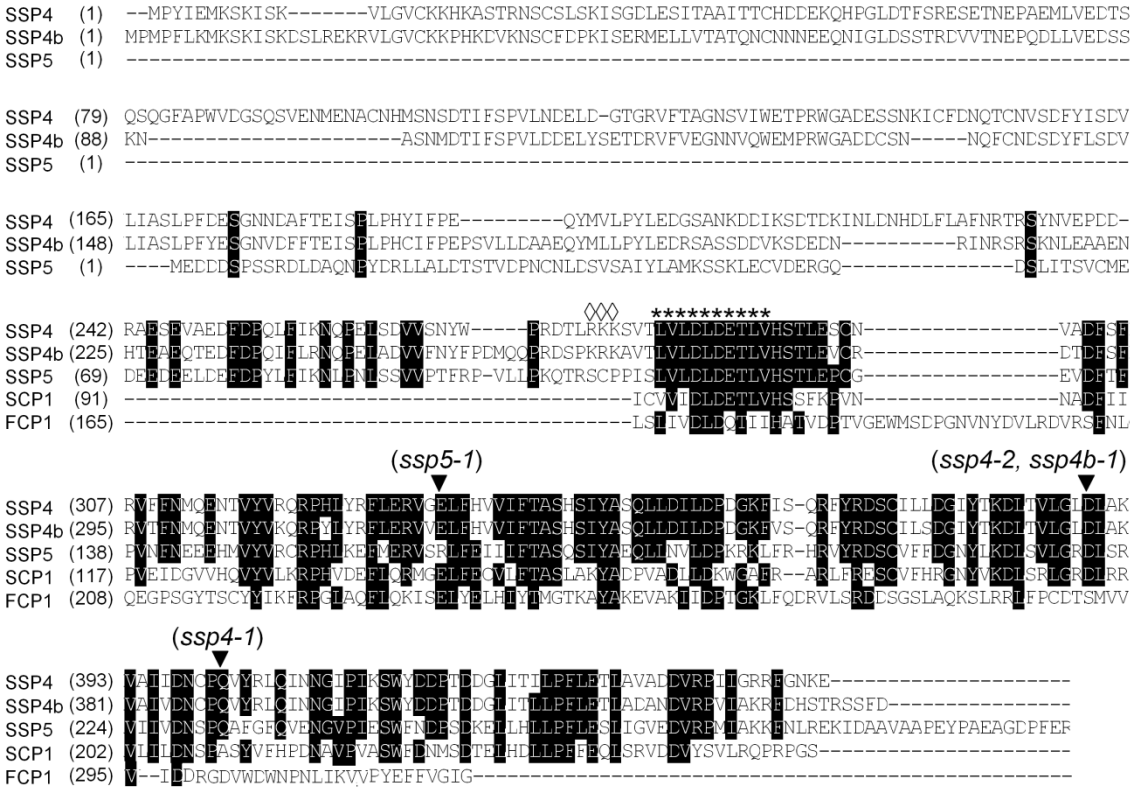

Figure 2.3. Alignment of SSP4, SSP4b, SSP5 with SCP1 and FCP1.

Full length SSP4, SSP4b and SSP5 were aligned with the conversed catalytic domains of human SCP1 and yeast spFcp1. Shaded boxes represent the conserved residues. Diamonds indicate the consecutive basic amino acids in SSP4 and SSP4b. Asterisks indicate the conserved acid phosphatase motif ψψψDXDX(T/V)ψψ. Filled triangles indicate the T-DNA insertion sites for *ssp* mutant alleles. Note *ssp4-2 and ssp4b-1* insertions truncate the proteins at the same position.

catalytic domain of human SCP1 and *Schizosaccharomyces pombe* Fcp1 (Figure 2.3). Homologies among SSPs in their catalytic domains are 91.3% (SSP4/SSP4b), 71.9% (SSP4/SSP5), and 71.9% (SSP4b/SSP5). Homologies between each SSPs (SSP4, SSP4b, and SSP5) and human SCP1 catalytic domain are 60.6%, 58.1%, and 67.5 %, and those between SSPs and spFCP1 catalytic domain are 36.5%, 33.1%, and 37.6%, respectively. Because of the high homologies between SSPs and the known CTD phosphatases (Hausmann et al., 2004), various analyses were performed to determine the biochemical functions of SSP.

Localization and expression of SSP isoforms

The sequence comparison among three SSP isoforms revealed a consecutive basic amino acid stretch only in SSP4 and SSP4b (at amino acid 277-279, and 265-267, respectively), but not in SSP5 (Figure 2.4). Although this is different from prototypical bipartite nuclear localization signals, a basic amino acid stretch is often associated with nuclear localization. In order to determine whether SSP4 and SSP4b show distinct subcellular localization from SSP5, GFP-SSP fusion proteins in *Arabidopsis* protoplasts were constructed, transiently expressed, and their localizations were examined. GFP fluorescence was observed in nuclei of the cells expressing GFP-SSP4 and GFP-SSP4b, but in both cytoplasm and nuclei of the cells expressing GFP-SSP5 (Figure 2.4). The patterns were consistent throughout the time course after transformation and among the cells with different levels of GFP-fluorescence intensity. This result indicates that SSP4 and SSP4b have a nuclear localization signal that is absent in SSP5. The distinct

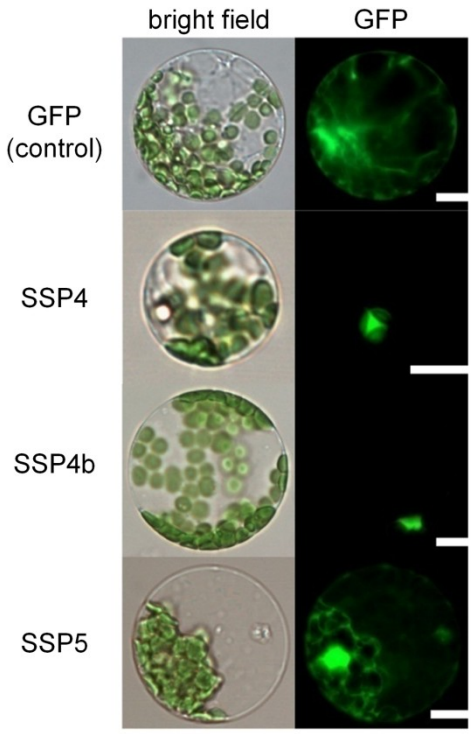

Figure 2.4. Subcellular localizations of SSP4, SSP4b and SSP5.

Plasmids encoding GFP vector control, GFP-SSP4, GFP-SSP4b and GFP-SSP5 were introduced into *Arabidopsis* protoplasts via PEG-mediated transformation. Images were taken 16 hours after transformation. Scale bars indicate 10 μm.

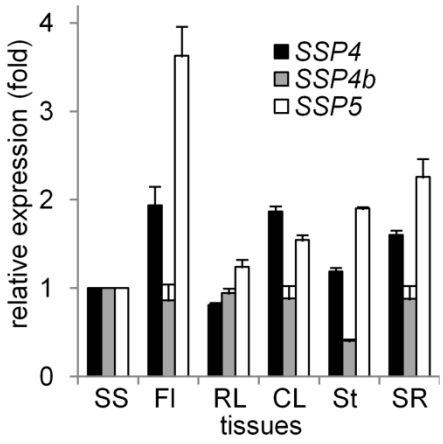

Figure 2.5. Expression profiles of *SSP4*, *SSP4b* and *SSP5* in different tissues.

Relative transcript levels in flowers (Fl), rosette leaves (RL), cauline leaves (CL), stems (St) in mature *Arabidopsis* plants, and 9-day-old seedling shoots (SS) and seedling roots (SR) were determined using quantitative reverse-transcription PCR (RT-qPCR). Values were expressed relative to seedling shoots. Bars indicate standard errors.

Table 2.2. *SSP* gene expressions with or without ABA treatment.

Gene	Col-0	Col-0 (ABA)	
SSP4	1	1.964	(0.270)
SSP4b	1	1.913	(0.268)
SSP5	1	0.945	(0.130)

The numbers show relative gene expressions (fold) compared with wild type plants without ABA treatment using RT-qPCR, with standard errors in parentheses. For ABA treatment, 100μM ABA was sprayed and total RNA was extracted after 3 hr. Experiments were repeated twice with similar results.

nuclear-cytoplasmic localizations among SSP isoforms are similar to those observed in different CPL isoforms (Koiwa et al., 2004).

Expression profiles of *SSP* isoforms were analyzed using Genevestigator (https://www.genevestigator.com/gv/index.jsp). *SSP* expressions were detected in all tissues, including root, stem, leaf, apex, floral organs, flower, seed and the whole plant. Various treatments including hormone, light, abiotic stress and pathogen infection did not significantly change gene expression levels of *SSP*s (data not shown). To confirm the microarray data, tissue specific expressions of *SSP*s by RT-qPCR were analyzed (Figure 2.5). Consistent with the database, three *SSP* genes were expressed in all tissues examined, with slightly higher *SSP4* and *SSP5* expressions in flowers and seedling roots (Figure 2.5). This may indicate that *SSP* expressions are associated with active growth and development in plants. Furthermore, *SSP* transcripts were shown to be weakly induced by ABA (Table 2.2), suggesting a possible role of SSPs in plant ABA response.

SSP4, SSP4b and SSP5 are CTD phosphatases with distinct activities

In order to establish catalytic activities of SSPs, recombinant SSPs as NusA-fusion proteins were expressed with an N-terminus $(His)_6$-tag and a C-terminal StrepTagII sequences (Figure 2.6A), and purified by consecutive Ni^{2+}-NTA and StrepTactin column chromatographies (Figure 2.6B). The protein amounts were determined using Bio-Rad protein assay reagent (Bio-Rad, CA) and confirmed with SDS-PAGE. Phosphatase activity of purified SSPs was evaluated with *para*-nitrophenyl phosphate (*p*NPP). Surprisingly, despite the high degree of similarities among SSP sequences, the activities

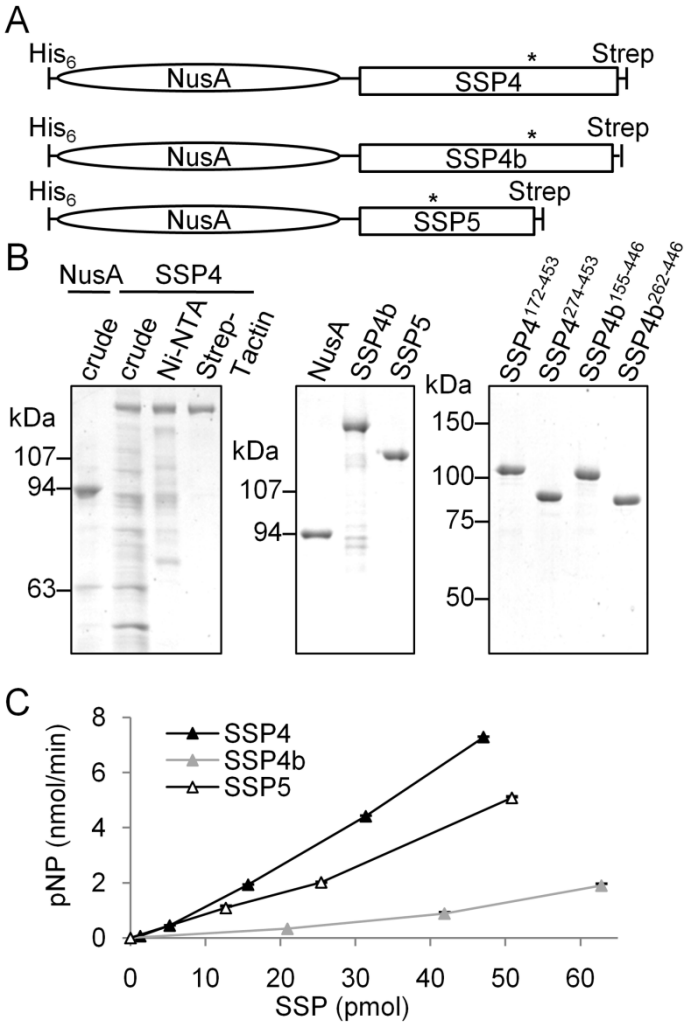

Figure 2.6. Phosphatase activities of recombinant SSPs.

(A) Structure of NusA-SSP4, -SSP4b and -SSP5 fusion proteins with an N-terminal $(His)_6$ tag and a C-terminal StrepTagII sequences. Asterisks indicate positions of the catalytic motif.
(B) (Left) Two-step purification of SSP4. Bacterial crude extracts containing control (NusA) or NusA-SSP4, and NusA-SSP4 fractions after each step were analyzed by SDS-PAGE. (Middle) Purified NusA, SSP4b and SSP5 fusion proteins. (Right) Purified N-terminally truncated SSP4 and SSP4b fusion proteins.
(C) Phosphatase activity of purified recombinant proteins. SSPs were incubated in reaction mixtures containing 100 mM Tris·acetate pH 5.5, 10 mM $MgCl_2$, and 10 mM *p*NPP for 5 min at 37˚C. The bars represent standard errors.

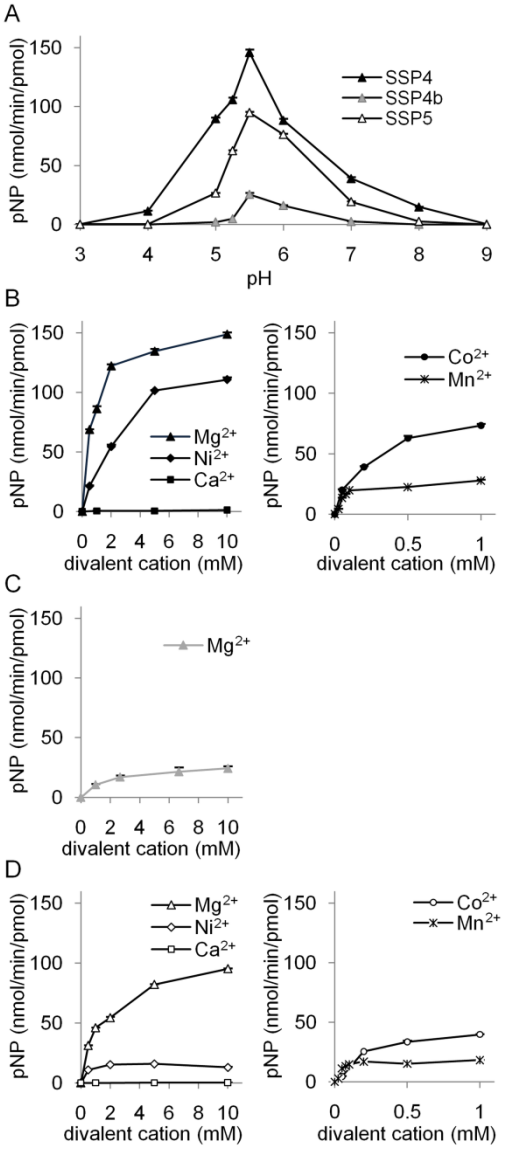

Figure 2.7. Phosphatase activities of SSP4, SSP4b and SSP5 in various conditions.

(A) pH dependence of SSP phosphatase activity. Reaction mixtures containing 100 mM Tris·acetate pH 3.0-9.0, 10 mM $MgCl_2$, 10 mM *p*NPP and SSP were incubated for 5 min at 37˚C. The amounts of proteins used were 12.2 pmol SSP4, 34.9 pmol SSP4b or 55.9 pmol SSP5.
(B to D) SSP metal dependence. Reaction mixtures containing 100 mM Tris·acetate pH 5.5, 10 mM *p*NPP, and indicated concentrations of $MgCl_2$, $NiCl_2$, $CaCl_2$, $MnCl_2$ or $CoCl_2$ were incubated for 5 min at 37˚C. The amounts of proteins used were 11.3 pmol SSP4 **(B)**, 43.6 pmol SSP4b **(C)** or 54.9 pmol SSP5 **(D)**. Bars indicate standard errors.

of SSP isoforms varied substantially (Figure 2.6C). Specific activities of full-length SSP4, SSP4b and SSP5 were 148, 26.7, and 95.4 unit/nmol, respectively.

When the activity was tested in buffers with a pH range of 3-9, all SSPs exhibited highest activity at pH5.5 (Figure 2.7A). Interestingly, SSP4 was active in a wide range of pH whereas SSP4b and SSP5 didn't have detectable activities at pH4 and pH8. All SSP phosphatase activities were dependent on divalent cations, with the highest at 10 mM Mg^{2+}. Weaker activities were detected with SSP4 and SSP5 in the presence of 10 mM Ni^{2+}, 1 mM Mn^{2+} and 1 mM Co^{2+} (Figure 2.7, B to D), and no activities were found with 1-10 mM Ca^{2+}, Zn^{2+} and Cu^{2+}. SSP4b, however, did not show any measurable activity with these cations except Mg^{2+} (data now shown).

In the optimal conditions (pH5.5, 10 mM Mg^{2+}), K_m and k_{cat} toward *p*NPP were 1.6 mM and 5.09 s^{-1} for SSP4, 15.5 mM and 0.83 s^{-1} for SSP5 (Figure 2.8). k_{cat}/K_m values were 3.24×10^3 $M^{-1} \cdot s^{-1}$ for SSP4 and 53.54 $M^{-1} \cdot s^{-1}$ for SSP5, suggesting SSP4 has 60 times higher catalytic efficiency than SSP5 (Table 2.3). It was not possible to determine kinetic parameters of full-length SSP4b due to its weak activity. These results establish that SSP4, SSP4b and SSP5 are metal dependent acid phosphatases and that SSP4 is more tolerable towards unfavorable conditions than SSP4b and SSP5.

Extended N-terminus is not required for phosphatase activity

The long N-terminal extension is a unique feature of SSP4, SSP4b and SSP5. The length and sequence of the N-terminal region vary among SSP4, SS4b and SSP5, which

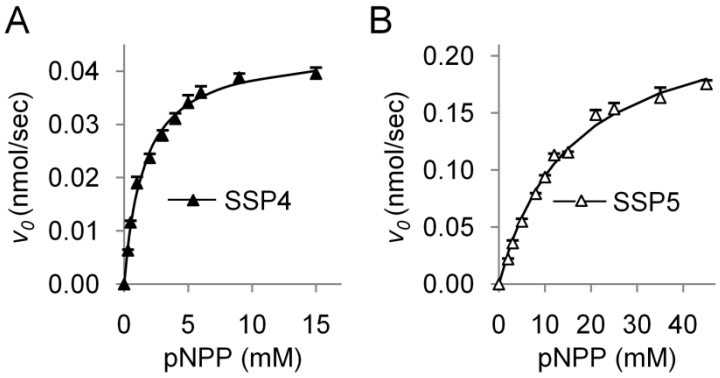

Figure 2.8. Kinetic parameters determination for SSP4 **(A)** and SSP5 **(B)**.

The reactions were carried under the condition of 100 mM Tris·acetate pH 5.5, 10 mM $MgCl_2$, and various *p*NPP concentrations at 37˚C. Protein amounts used were 15.7 pmol SSP4, and 59.0 pmol SSP5. The bars represent standard errors.

Table 2.3. Kinetic parameters of SSP.

SSP	K_m (mM)	V_{max} (nmol·sec^{-1})	k_{cat} (s^{-1})	k_{cat}/K_m ($M^{-1}\cdot s^{-1}$)
SSP4	1.6	0.044	2.82	1763
SSP4b$^{262-446}$	9.1	0.074	2.99	328
SSP5	15.5	0.242	4.10	265

The values represent three measurements of SSP activities.

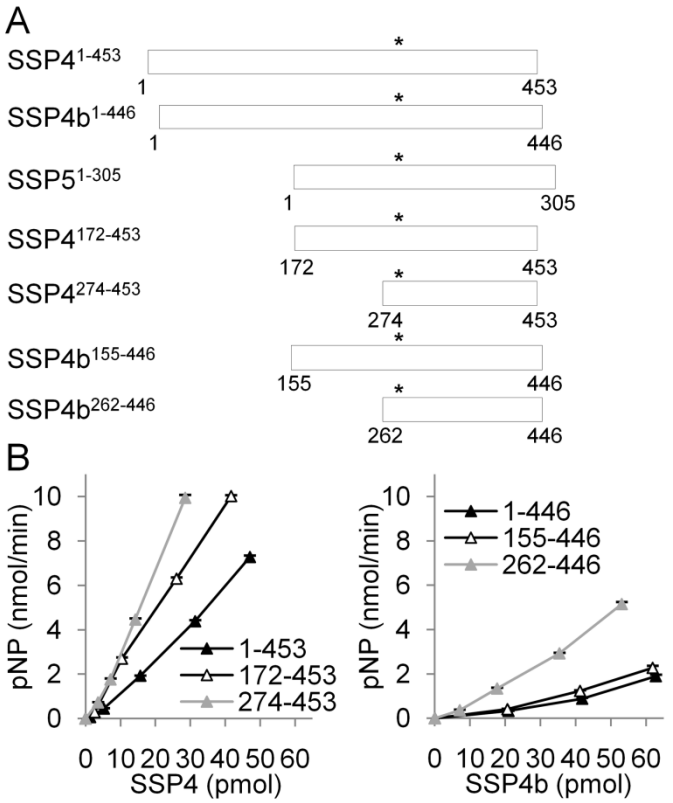

Figure 2.9. Phosphatase activities of truncated SSPs.

(A) Schematic illustration of SSP truncations. Asterisks indicate the positions of catalytic motif.
(B) Phosphatase activity of purified SSP4 (left) and SSP4b (right) deletion mutants. The reaction condition was same as Fig. 3C. The bars represent standard errors.

implied that this region contributes to differential activities among SSP isoforms (Figure 2.3). To test the role of N-terminal extensions, truncated SSPs were prepared by PCR-based mutagenesis (Figure 2.9A). For SSP4 and SSP4b, truncation of the N-terminal region increased induction of the soluble fraction as well as specific activity of the recombinant proteins. $SSP4^{274-453}$and $SSP4b^{262-446}$, both of which have only nine amino acids before the catalytic motif, showed more than two-fold increases of phosphatase activity compared to the respective full-length proteins (Figure 2.9B).

The K_m, k_{cat} and k_{cat}/K_m values for $SSP4b^{262-446}$ were 9.1 mM, and 0.36 s^{-1}, and 39.33 $M^{-1}\cdot s^{-1}$, which were comparable to the values of full-length SSP4 and SSP5 (Figure 2.10). Since the activity of full-length SSP4b was too weak, the $SSP4b^{262-446}$ variant was used for further characterizations. Given both improved activity, truncated SSP4 still harbored higher activities than truncated SSP4b. An equivalent N-terminal truncated $SSP5^{105-305}$ improved the yield of the soluble recombinant protein, but decreased its activity to less than 3% of full length SSP5 (data not shown). Similarly, removal of a short C-terminal extension in SSP5 improved the yield but abolished the activity of $SSP5^{1-284}$ (data not shown). Therefore, N-terminal extension of SSP might not be necessary for the phosphatase activity; neither does it contribute to the phosphatase activity differences among SSPs.

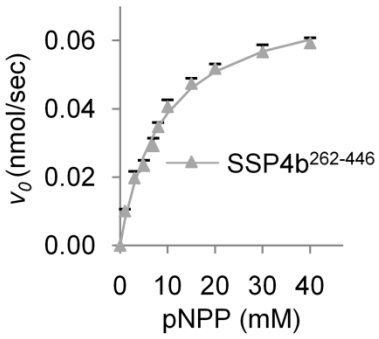

Figure 2.10. Kinetic parameters determination for SSP4b.

The reactions were carried under the condition of 100 mM Tris·acetate pH 5.5, 10 mM $MgCl_2$, and various *p*NPP concentrations at 37˚C. 24.8 pmol SSP4b$^{262-446}$ was used in the reaction. The bars represent standard errors.

SSP4 and SSP4b dephosphorylate both Ser2 and Ser5 of CTD-PO_4, while SSP5 is specific to Ser5

Phosphatase specificities of SSPs were determined using various substrate systems. Phosphorylated peptide substrates include phosphoserine substrates CTD-Ser2-PO_4 (S2) $[YpSPTSPS]_3$, CTD Ser5-PO_4 (S5) $[YSPTpSPS]_3$, pS [RRApSVA: from type L pyruvate kinase (Rayapureddi et al., 2005)], and a phosphotyrosine substrate pY [RRLIEDAEpYAARG: from Tyrosine 419 phosphorylation site in Src (Rayapureddi et al., 2005)]. In order to test the phosphatase activities in a condition similar to the native environment, the assays were performed at pH 7.0. SSPs exhibited distinct preferences toward S2 and S5 peptide substrates (Figure 2.11). SSP4 and SSP4b could dephosphorylate both S2 and S5 substrates, and exhibited 8 and 16 times higher activity towards S5 than towards S2, respectively. However, SSP5 dephosphorylated only S5, but not S2 (Figure 2.11). All SSPs tested here were virtually inactive toward pS and pY (Figure 2.12). These results indicate that SSP4, $SSP4b^{262-446}$ and SSP5 are pol II CTD phosphatases with distinct position preferences.

The abilities of SSPs to dephosphorylate the heptad repeat substrate in a native CTD context were tested using recombinant pol II CTD with 34 repeats fused to glutathione S-transferase (GST-CTD). As shown in Figure 2.13, GST-CTD-PO_4 prepared using activated AtMPK3 contained both Ser2 and Ser5 phosphorylations, which were visualized by monoclonal antibodies H5 and H14, respectively. In addition, phosphorylation of GST-CTD decreased its mobility in SDS-PAGE gels. The hyperphosphorylated low-mobility form was termed as CTD_O and unphosphorylated

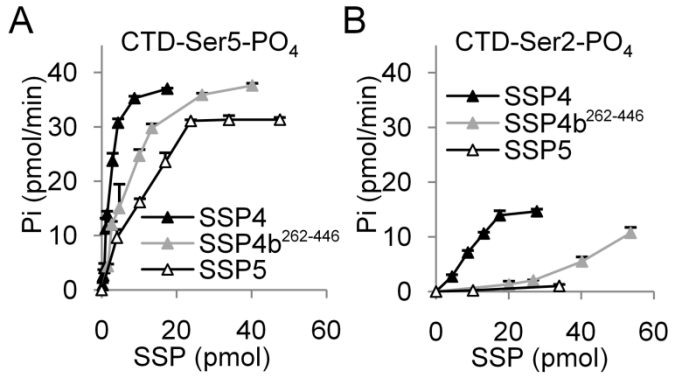

Figure 2.11. CTD phosphatase activities of SSPs with synthetic phosphopeptide substrates and their specificities.

(A) Activities of SSPs with CTD-Ser5-PO_4 substrates. SSPs were incubated in reaction mixtures containing 50 mM Tris·acetate pH 7.0, 10 mM $MgCl_2$, and 25 μM synthetic phosphopeptide substrates for 30 min at 37°C. The bars represent standard errors.
(B) Activities of SSPs with CTD-Ser2-PO_4 substrates. Conditions were same as **(A)**.

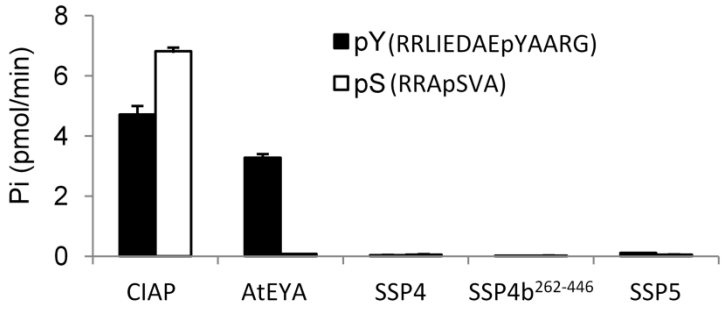

Figure 2.12. Activities of SSPs with synthetic non-CTD phosphopeptide substrates.

Reaction mixtures containing 50 mM Tris·acetate pH 7.0, 10 mM $MgCl_2$, and 25 μM synthetic pS or pY substrates were incubated for 30 min at 37°C. The amounts of proteins used were 21pmol of SSP4, 60pmol of $SSP4b^{262-446}$, 51pmol of SSP5, 21pmol AtEYA or 21pmol of CIAP. Bars indicate standard errors.

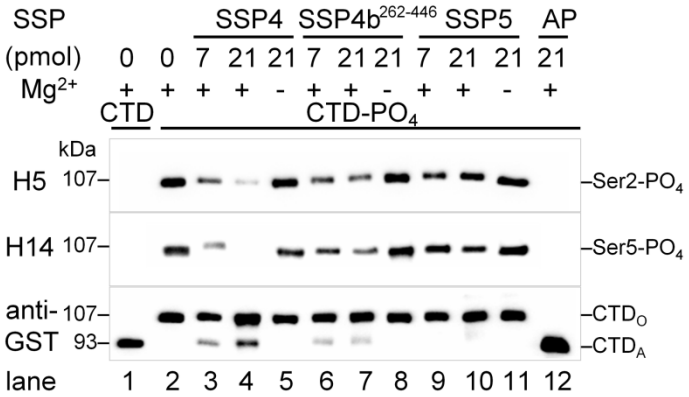

Figure 2.13. CTD phosphatase activities of SSPs with GST-CTD-PO_4 substrate in immunoblotting analyses and their substrate specificities.

GST-CTD was phosphorylated by activated *Arabidopsis* MPK3. One-hundred nanogram of GST-CTD_O was incubated in reaction mixtures containing 50 mM Tris·acetate pH 7.0, with or without 10 mM $MgCl_2$, and indicated amount of SSP or alkaline phosphatase (AP). Aliquots of the mixtures were analyzed by immunoblotting. Ser2-PO_4 and Ser5-PO_4 of GST-CTD were detected by H5 and H14 antibody, respectively, and the total GST-CTD substrate was visualized by anti-GST antibody. Lane 1, unphosphorylated GST-CTD (GST-CTD_A). Lane 2-11, phosphorylated GST-CTD (GST-CTD_O) incubated with 0-21 pmol of SSP4, SSP4b or SSP5, with or without Mg^{2+}. Lane 12, GST-CTD_O incubated with 21 pmol alkaline phosphatase (positive control).

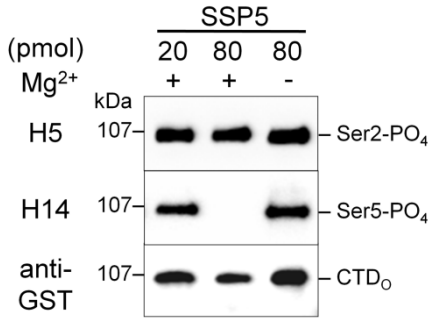

Figure 2.14. Dephosphorylation of GST-CTD_O by excessive amount of SSP5.

The CTD phosphatase assay was performed as described in Figure 2.13 with indicated amount of SSP5 protein.

high-mobility form was termed CTD_A. Upon incubation of CTD_O with SSP4 in the presence of Mg^{2+}, CTD_O lost reactivity with both H5 and H14, indicating SSP4 dephosphorylated both Ser2-PO_4 and Ser5-PO_4 of CTD_O. The loss of immunoreactivity was concomitant with an appearance of the CTD_A form. Similar result was obtained with $SSP4b^{262-446}$, albeit its activity was less robust (Figure 2.13). In contrast, SSP5 was able to dephosphorylate only Ser5-PO_4 but not Ser2-PO_4 of GST-CTD_O (Figure 2.13 and 2.14). These results are consistent with the data obtained with the peptide substrates and indicate that these SSPs are functional pol II CTD phosphatases.

Identification and phenotypic characterization of *ssp* T-DNA insertion mutants

In order to test *in vivo* function of SSP isoforms, T-DNA insertion mutants were identified in T-DNA Express (http://signal.salk.edu) and homozygous mutant lines were established (Figure 2.15, A to B). The insertion positions were confirmed by sequencing the PCR products for T-DNA-genome junctions. RT-PCR analyses revealed that *ssp* T-DNA insertion lines produces *SSP-T-DNA* chimeric transcripts, which encode partial SSP peptides fused to short translation products of T-DNA left border sequence, at the level comparable to corresponding wild type *SSP* transcripts (Figure 2.15, C to E). Double and triple *ssp* mutant combinations were prepared by genetic crosses. *ssp* mutants did not show defects under normal growth condition, except for a slight delay in flowering with *ssp4* and *ssp4 ssp4b ssp5* triple mutant (data not shown). However, *ssp5* single mutant and *ssp4 ssp4b ssp5* triple mutant exhibited enhanced ABA sensitivity during germination (Figure 2.16). *ssp4 ssp4b* double mutant did not show the same

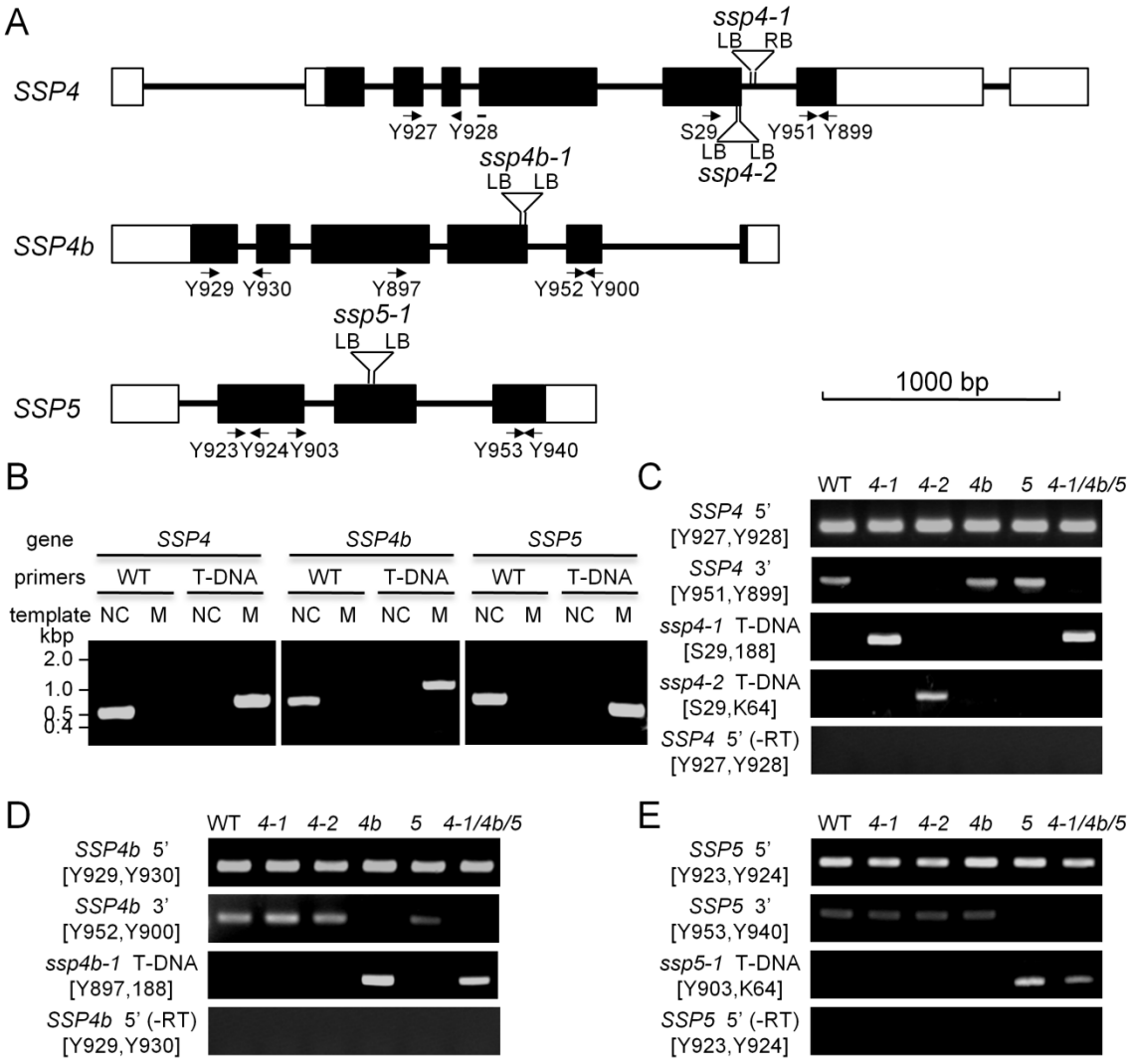

Figure 2.15. Identification and characterization of *ssp* T-DNA insertion mutants.

(A) T-DNA insertion positions in *ssp* mutant alleles. White boxes represent UTRs and black boxes indicate coding regions. The primers used to identify T-DNA insertion positions and in RT-PCR analyses were shown as arrows.
(B) PCR confirmation of T-DNA insertions in *ssp4 ssp4b ssp5* triple mutant line. WT and T-DNA primers amplify corresponding SSP sequences without or with the T-DNA insertions, respectively. NC template: wild type DNA; M template: *ssp4 ssp4b ssp5* triple mutant.
(C to E) RT-PCR analyses of *SSP4*, *SSP4b*, *SSP5* transcripts in *ssp* single and triple mutants using primers marked in **(A)**.

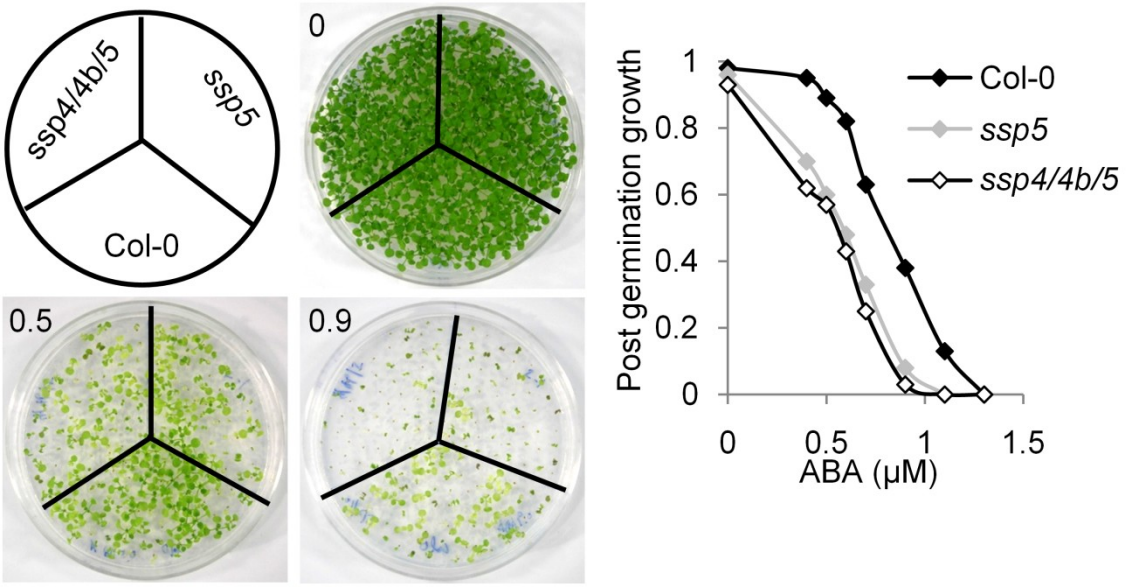

Figure 2.16. Enhanced ABA sensitivity of *ssp5* single and *ssp4 ssp4b ssp5* triple mutant seed germination.

Sixty seeds of wild type (Col-0), single mutant(*ssp5)* and the triple mutant (*4/4b/5*) were sown on 1/4 x MS salt, 0.8 % agar plates containing various concentration of ABA. After stratification, plates were incubated in 25°C under 16 hr/8 hr light/dark cycle for two weeks before documentation and scoring.

Table 2.4. Gene expressions in wild type and *ssp* triple mutant with or without ABA treatment.

Gene	Col-0	ssp4/4b/5		Col-0 (ABA)		ssp4/4b/5 (ABA)	
SSP4	1	0.043	(0.006)	1.964	(0.270)	0.084	(0.012)
SSP4b	1	0.009	(0.000)	1.913	(0.268)	0.080	(0.015)
SSP5	1	0.190	(0.004)	0.945	(0.130)	0.073	(0.006)
RD29A	1	0.776	(0.035)	108.031	(18.886)	92.345	(5.792)
COR15A	1	0.571	(0.015)	143.570	(19.737)	100.924	(6.440)
ABF2	1	1.278	(0.055)	8.175	(1.147)	8.470	(0.531)
ABI5	1	1.237	(0.143)	42.257	(5.793)	38.391	(2.502)
CAK4	1	0.940	(0.018)	0.758	(0.105)	0.679	(0.087)
CDKC2	1	0.874	(0.017)	0.617	(0.085)	0.479	(0.031)
EF1a	1	1.112	(0.021)	0.566	(0.080)	0.580	(0.036)
ACT2	1	1.047	(0.030)	0.519	(0.072)	0.490	(0.031)

The numbers show relative gene expressions (fold) compared with wild type plants without ABA treatment using RT-qPCR, with standard errors in parentheses. For ABA treatment, 100μM ABA was sprayed and total RNA was extracted after 3 hr. Experiments were repeated twice with similar results.

enhancement (data not shown), and addition of *ssp4 ssp4b* double mutant to *ssp5* single mutant background didn't change seed germination significantly (Figure 2.16). These data suggest that *SSP5* suppress ABA responses during seed germination, and that *SSP4 and SSP4b* may redundantly regulate ABA response with *SSP5*. Unlike prototypical ABA-hypersensitive mutants (Hugouvieux et al., 2001a), root growth responses of *ssp4 ssp4b ssp5* triple mutant seedlings to ABA, NaCl, and mannitol were not distinguishable from those of wild type. RT-qPCR analyses indicated that expressions of several constitutive and ABA-responsive genes were not altered in the triple mutant (Table 2.4). Further analyses are necessary to elucidate the mechanism of SSP on ABA response.

DISCUSSION

Recent studies have established the central ABA signal transduction pathway in which perception of ABA molecule leads to transcriptional activation of ABA-responsive genes and stomatal movement (Fujii et al., 2009; Ma et al., 2009; Park et al., 2009; Vlad et al., 2009). Still, the role of RNA metabolism in ABA signaling remains unclear. This is partially due to the fact that only a limited number of vast general RNA metabolism proteins have been shown to associate with ABA signaling. In this chapter a group of *Arabidopsis* small CTD-phosphatase-like proteins, SSP were analyzed. Consistent with the predicted transcriptional functions, SSPs were mainly localized to nuclei, however, a proportion of SSP5 also detected in the cytoplasm. Different subcellular localization profiles between SSP5 and SSP4/SSP4b are consistent with the fact that SSP4 and SSP4b are more homologous to each other than to SSP5. Similarly, these SSP isoforms

showed different dephosphorylation specificities toward CTD substrates. SSP5 exclusively dephosphorylated Ser5-PO_4 of CTD, which was similar to the cases of *Arabidopsis* CPL1 and CPL2. In contrast, SSP4 and SSP4b could dephosphorylate both Ser2-PO_4 and Ser5-PO_4, and convert hyperphosphorylated CTD_O to hypophosphorylated CTD_A. This conversion had never been achieved in previously reported plant CTD phosphatase-like proteins. Interestingly, while both the N- and C- terminal regions were important for SSP5 activity in vitro, the long N-terminal extensions of SSP4 and SSP4b were not necessary for, but rather inhibitory to, their phosphatase activity. Since our attempt to inhibit truncated $SSP4^{274-453}$ by N-terminal peptide *in trans* resulted in very weak inhibition of the phosphatase activity (data not shown), it is not clear whether the N-terminal region functions as a regulatory sequence.

To date, no substantial developmental defects in *Arabidopsis* plants have been observed with single and combined *ssp* mutations, and no other *Arabidopsis ssp* mutants have been identified in forward genetic screening. This may indicate a higher order of genetic redundancy in developmental functions of SSPs. It's also possible that SSPs function in a very specific stage of plant development, which requires careful inspection of particular tissues. A recent study in rice of a *cpl1* mutant implied the latter (Ji et al.). OsCPL1 is most homologous to *Arabidopsis* SSP5, and the *cpl1* mutation in rice promoted differentiation of the abscission layer during panicle development in rice. Notably, T-DNA insertions in *Arabidopsis* SSPs described in this study locate upstream of recessive loss of function mutations in OsCPL1. Also, a short C-terminal truncation of SSP5 abolished its in vitro activity. These support our conclusion that the *ssp* mutant

lines produce dysfunctional transcripts. The repressive role of OsCPL1 is similar to those of human SCPs, *Arabidopsis* CPL family proteins, and SSPs, which repress ABA responses in wild type plants.

How do CTD phosphatases specifically affect phenotypic outputs? The recent studies on animal SCPs have indicated more than one mechanism for CTD phosphatases to regulate cellular signaling. SCP1-SCP3 are similar to SSP4 and SSP4b because they could dephosphorylate both Ser2-PO_4 and Ser5-PO_4 but prefer Ser5-PO_4 of CTD as substrates (Yeo et al., 2003a). In vivo, SCPs were shown to function in silencing neuronal gene expression likely through CTD dephosphorylation (Yeo et al., 2005; Visvanathan et al., 2007; Yeo and Lin, 2007). Also, SCP2 has been shown to regulate gene expression at the level of promoter clearance of pol II (Thompson et al., 2006). This is consistent with the current plant model that CTD phosphatases collectively dephosphorylate pol II CTD and regulate transcription elongation. In our study, Ser5 rather than Ser2 phosphatase activity is likely the cause of observed phenotype because pyramiding all three SSPs was necessary to detect the enhanced ABA sensitivity, and only Ser5 phosphatase activity was common in the three SSPs. Still, the phenotype of *ssp* triple mutant is relatively mild, and its ABA-induced gene expression was not distinguishable from wild type. Indeed, additional proteins, such as CPL1 and CPL2, other SSP isoforms, and an SSU72 homolog likely have the Ser5 phosphatase activity (Koiwa et al., 2004).

In addition, human SCPs could also regulate development by dephosphorylating non-CTD substrates. SCPs could exert regulatory dephosphorylation of the linker region of

Smad transcription factors, which mediated BMP (Bone Morphogenetic Protein) signaling pathway (Knockaert et al., 2006; Sapkota et al., 2006; Wrighton et al., 2006). Furthermore, SCP3 is identical to tumor suppressor HYA22A and has been implied to regulate the level of phosphorylated retinoblastoma protein (Gangopadhyay et al., 2008). By analogy, SSP family proteins may function not only via regulating CTD dephosphorylation but also affecting other cellular components. Indeed, the central ABA signaling pathway has multiple phosphorylation /dephosphorylation components. For example, SSPs may dephosphorylate the ABF family transcription factors and attenuate their functions in transcriptional activation of target genes. Further genetic and biochemical studies are necessary to decipher the mode of SSP functions in ABA signaling.

CHAPTER III

APPLICATION OF CTD PHOSPHATASE MUTANT IN HYPERACCUMULATION OF FLAVONOIDS IN TRANSGENIC *ARABIDOPSIS THALIANA*

INTRODUCTION

Flavonoids are a family of compounds that are produced in both vascular and non-vascular plants. The functions of flavonoids include forming physical barriers, biochemical and visual signals to symbiotic partners and pollinators, protection from UV damage, and regulation of auxin transport during development (Dixon and Paiva, 1995; Shadle et al., 2003). For animal consumption, flavonoids are known for health-promoting effects, featuring their antioxidant activity and prevention of chronic degenerative diseases and obesity (Korkina, 2007; Iriti and Faoro, 2009). Anthocyanin, kaempferol and Quercetin are several major types of compounds that have been most extensively studied in *Arabidopsis* and other species. Several transcription factors including myb-type transcription factors PAP1 and PAP2 (*Production of Anthocyanin Pigment 1/2*) (Kranz et al., 1998; Borevitz et al., 2000; Stracke et al., 2001; Gao et al., 2008), homeobox gene ANL2 (*Anthocyaninless2*) (Kubo *et al.*, 1999) have been identified in anthocyanin branch of flavonoid pathways.

Ectopic overexpression of *PAP1* and other MYB transcription factors has been successfully enhanced biosynthesis of anthocyanins in various plant species (Vom Endt *et al.*, 2002; Xie et al., 2006; Zhou et al., 2008; Peel *et al.*, 2009; Li et al., 2010; Velten et al., 2010) . Transcriptomic analysis of activation tagging mutant of *PAP1* in

Arabidopsis (*pap1-D*) revealed that *PAP1* strongly upregulate anthocyanin biosynthesis pathway genes specifically, without effective interference with the expression of early phenylpropanoid pathway and flavonoid pathway (Tohge *et al.*, 2005). While constitutively overexpression of *PAP1* may cause vegetative growth retardation, an inducible system could be applied to achieve *PAP1* upregulation at a proper stage of plant development.

In this chapter, a three-component gene expression system is illustrated as well as its application to cold-inducible anthocyanin production. A gene of interest (*PAP1*) was cloned downstream of cold-inducible *RD29A* promoter, and *Arabidopsis* plants were co-transformed with *RD29A-PAP1* and a feedforward effecter gene of the cold signal (*RD29A-CBF3*). A mutation in host repressor *CPL1* (Koiwa et al., 2002; Xiong et al., 2002) seemed to be an essential third component for the success of this expression system. Cold induction activated expressions of *PAP1* and anthocyanin biosynthetic genes, which was accompanied with overproduction of anthocyanins. Resulting phytochemical profile of transgenic plants showed synergism of native and PAP1-induced anthocyanin production. The results establish that a three-component system using a native plant promoter is sufficient to drive high expression of transgene upon induction.

MATERIALS AND METHODS

Construction of expression cassettes

Primer sequences used in this research are listed in Table 3.1. A cDNA fragment encoding *Arabidopsis PAP1*and *CBF3* were amplified using primer pairs [680, 681] and

Table 3.1. Sequences of primers in the project described in Chapter III.

Name	Sequences (5'—>3')
680	AAGGTCGACCATGGAGGGTTCGTCCA
681	TTTCCCGGGCTAATCAAATTTCACAG
678	AGGGTCGACCATGAACTCATTTTCTG
679	TTTCCCGGGTTAATAACTCCATAACG
Y888	CGGAGCATGAGTCATTGAACACTCCG
Y889	ATTCATCATCACCACTTGGAGCA
Y882	CGGTTATCTGAGTAAGAAACATGAACCG
Y883	TTTAGTGCCGGTGTTGTAGGAAT
Y884	CGTGTATGGTGGAGGCTATTTACACG
Y885	CCAACAAACTCGGCATCTCAAA
Y976	TCTAGGGAATAGGTGGTCTTTA
Y977	CACGGTTCATGTTTCTTACTC
Y978	GAAATGTTTGGCTCCGATTAC
Y979	TCACGAAACTTCTTACGACC
Y980	CGGCTGTTTCAGACTATCTT
Y981	ACTCCTCCAGTTTCTTCTTTG
Y958	GTATGAACAAAGGCACTGATAG
Y959	CGGCGTTAAGGAATCTAATAAG
Y960	ACCTCAAGGAGAAGTTCAAG
Y961	TTTCCTTGAGGAATTCCTCC
Y968	ACTTACCATTTCTCAGCCAA
Y969	TTCTCCACCTTCAGTTTCTC
Y962	AAGATCCTGAGAACGAAGTG
Y963	CTTTGCCTTAACACATGCTT
Y970	CTTACCTTCAGGCGGTTATC
Y971	GTTCGTCAATAGAGTCGATCC
Y1006	CTTCTTTCATCTTGCGTATCC
Y1007	CTCGTTGCTTCTATGTAATCAC
Y996	CACTCTTCTTGTTCCTAACGA
Y997	CCCATTACTCAACCTCAGAATC
Y1000	CCTCCAGAATACATTGAAGTGA
Y1001	ATGCCTTAAACCTAGTCCTTC
Y1002	CAAATTCAGTCATTTATGTTTCCTTT
Y1003	CATTCTTCTCTATCATCTTTGGTTT
Y1004	GAGATAAATGCGTCGTGGATT
Y1005	GCTCACCTACTTCTTTGACAC

[678,679], respectively. The entry plasmid pEnRD29A-LUC was prepared by inserting an *RD29A-LUC* expression cassette (Ishitani *et al.*, 1997) into pEntr2B (Life Technologies, CA). pEnRD29A-PAP1 and pEnRD29A-CBF3 were prepared by replacing luciferase ORF (*LUC*) with *PAP1* and *CBF3* coding sequences, respectively. A plasmid vector pFAJ3163 containing *BAR* gene was provided by Dr. Cammue, and pFAJGW was prepared by replacing *35S*::*GUS* cassette with a gateway cassette. A plasmid vector pMDC99 (Curtis and Grossniklaus, 2003) containing *HPT* gene was provided by Arabidopsis Biological Resource Center. In order to prepare plant transformation binary plasmids, pEnRD29A-PAP1 and pEnRD29A-CBF3 were recombined using LR clonase (Invitrogen, CA) with pFAJGW and pMDC99, respectively.

Plant growth conditions

For in vitro culture, surface-sterilized seeds were sown on media containing 1/4 × MS salts, 0.5% sucrose and 0.8% agar. After stratification for 2-4 days plates were incubated at 23°C for 7 days under the 16 hr light/8 hr dark cycle. For cold treatment, plates then were moved to 4°C and incubated for additional 4 days.

For growth and induction of anthocyanin accumulation on soil, seeds were sown directly on Metromix 366 potting media. After 2-4 days stratification, plants were grown at 23°C for 3 weeks under the 16 hour light/8 hour dark cycle. For cold treatment, plants were then moved to 4°C and grown for specified periods under the 16 hour light/8 hour dark cycle.

Arabidopsis double transformation

Binary plasmids pMDC-CBF3 and pFAJ-PAP1 were transformed into *Agrobacterium tumefaciens* GV3101 and ABI, respectively. Empty vector controls (pBIB-HYG, pFAJ3163) were also transformed into Agrobacteria. In order to transform Arabidopsis, bacterial suspensions were prepared in solution containing 5% sucrose and 0.03% Silwet L-77. Mixtures (1:1) of suspensions were prepared in following combinations: [pMDC-CBF3 and pFAJ-PAP1 (PC)], [pFAJ-PAP1 and pBIB (PB)], [pMDC-CBF3 and pFAJ3163 (C3)], or [pBIB and pFAJ3163 (B3)] and applied to flower buds of *Arabidopsis* wild type and *cpl1-2* mutants. Resulting T_1 seeds of eight genotypes were harvested separately.

For selection of double transformants, first, hygromycin-resistant transformants were selected on media containing 1/4 x MS salts, 30 µg/ml hygromycin B, 100 µg/ml cefotaxim and 0.8% agar. Sixty lines of each selected genotype of T_1 plants were then transplanted to the soil and sprayed with 30 µg/ml phosphinothricine to identify PPT^R transformants. Thirty T_1 double transformants of each combination were harvested and subjected to Hyg^R and PPT^R selection again to obtain single copy T_2 transformants. T_3 plants were tested again to identify transformants homozygous for both Hyg^R and PPT^R. T_4 plants that contained single copy T-DNA for both transgenes as homozygous state were used for further analysis.

Gene expression analyses by RT-qPCR

Total RNA was isolated using Trizol reagent (Life Technologies, CA). RNA samples resuspended in 50 µl were treated with 7.5 unit of DNase I (Qiagen, MD) for 60 min at 37°C, and re-purified with the RNeasy plant mini kit (Qiagen). Quantitative reverse-transcription PCR (RT-qPCR) was performed as described previously. The absence of genomic DNA contamination was confirmed using minus-reverse-transcriptase controls. The data were processed as described previously. For screening of transgenic plants, specific primer pairs for RT-qPCR analyses were: [Y888, Y889] for RD29A, [Y882, Y883] for PAP1 and [Y884, Y885] for CBF3. For time-course analyses, [Y980, Y981] for RD29A, [Y976, Y977] for PAP1, [Y978, Y979] for CBF3, [Y958, Y959] for PAL1, [Y960, Y961] for CHS, [Y968, Y969] for CHI, [Y962, Y963] for DFR, [Y970, Y971] for F3'H, [Y1006, Y1007] for ANS, [Y996, Y997] for FLS1, [Y1000, Y1001] for FLS3, [Y1002, Y1003] for UGT73B2, and [Y1004, Y1005] for UGT78D2.

Phytochemical identification and quantification by LC-MS

For spectrophotometric quantification of total anthocyanin content, one gram of leaf samples were processed as described (Fuleki and Francis, 1968). The anthocyanin contents were calculated as cyanidin-3-glucoside equivalent.

For LC-MS identification and quantification of separated anthocyanin peaks, one gram of leaf samples were ground in liquid nitrogen, and extracted with five grams of methanol: water: acetic acid (9:10:1) at 4 ˚C for 24 h in dark on shaker at 120 rpm. Extracts were centrifuged at 10,000 g at 4 ˚C for 20 min. Supernatant was filtered with a

0.45 µm nylon filter (Fisherbrand, Houston, TX) and 20 µl were injected into the LC-MS. Individual compounds were identified based on retention time, UV spectra and their mass per charge ratio using LC-MS as described previously (Tohge *et al.*, 2005). Chromatographic separations were performed in reverse phase with a 150 x 2.00 mm Synergi 4µ Hydro RP 80A column (Phenomenex, Torrance, CA) and a guard column of the same chemistry. The mobile phase consisted of 1% formic acid in water (solvent A) and HPLC-grade acetonitrile (solvent B) at a flow rate of 200 µl/min. Solvent gradient started with 1% B, increasing linearly to 25% B at 20 min, reaching 55% B at 30 min and 90% B at 31 min. Column was equilibrated for 10 minutes in between samples.

The LC-MS system consisted of a triple quadruple ion trap LCQ Deca XP Max (Thermo Finnigan, San Jose, CA), a Surveyor P2000 quaternary pump, an autosampler, and a UV 2000 PDA detector (Thermo Finnigan, San Jose, CA). Anthocyanins and flavonoids were detected at 520 and 320 nm, respectively. Sample was delivered to the MS by electrospray ionization (ESI). Mass spectrometry analysis were done in the positive ion mode, with a mass range was 80-1500 *m/z*. The MS conditions were as follows: 50 arb units sheath gas, 0 arb units auxiliary gas, temperature 275 ºC, spray voltage 4 kV, capillary voltage -21 V, and tube lens offset -60 V. Nitrogen was used as the sheath gas and helium as the collision gas. MS^2 and MS^3 analysis were used during the identification. Collision energy of 50% was used for the MS^n analysis.

RESULTS

Designing osmotic stress inducible transcription factor cassettes

Toward the goal of on-demand phenylpropanoid production by exogenous stimuli, cold-regulated *PAP1* overexpression cassette was prepared by placing *PAP1* cDNA downstream of *RD29A* promoter and TMV *Omega* sequence (Figure 3.1). In order to enhance the efficiency of induction, an *RD29A-CBF3* effecter was prepared. Since CBF3 binds to and promotes expression of *RD29A* promoter, *RD29A-CBF3* functions as a cold-induced self-amplicon, which will feedforward the expression of *RD29A* promoter. Furthermore, the effecter has a protective function during the cold treatment.

Co-expression of *RD29A-PAP1* and *RD29A-CBF3* was not sufficient to induce anthocyanin accumulation

PAP1 and/or *CBF3* expression cassettes as well as vector control constructs were introduced into wild type *Arabidopsis thaliana* plants. These lines were designated as B3 (vector control), PB (*RD29A-PAP1* only), PC (*RD29A-PAP1* and *RD29A-CBF3*). Plants containing single copy of each expression cassette were selected based on hygromycin (for pMDC-CBF3) and Liberty resistance (for pFAJ-PAP1), and homozygous T3 lines were identified. The homozygous lines were screened for the transgene expression and anthocyanin contents before and after activation of *RD29A* promoter by cold stress (Table 3.2, Figure 3.2). Compared to the unstressed vector control lines, cold-treated PB transformants (*RD29A-PAP1* only) expressed 7-10 fold higher level of *PAP1*. Pyramiding *RD29A-CBF3* on top of *RD29A-PAP1* (PC lines)

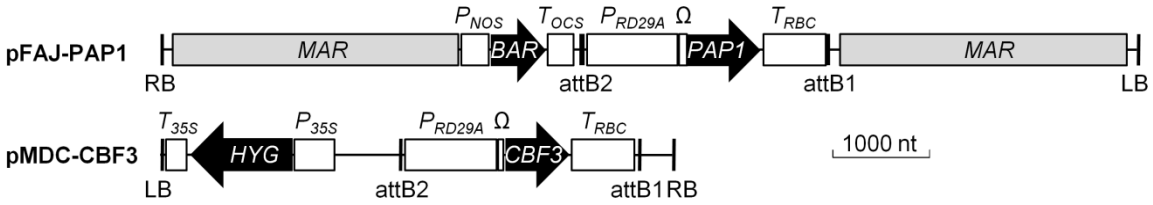

Figure 3.1. Constructs for double transformation.

RD29A-PAP1 construct was recombined with pFAJ vector carrying *BAR* resistant marker gene and RD29A-CBF3 with pMDC vector harboring *HPT* gene. Double transformation was performed using two constructs as well as different combination of control vectors.

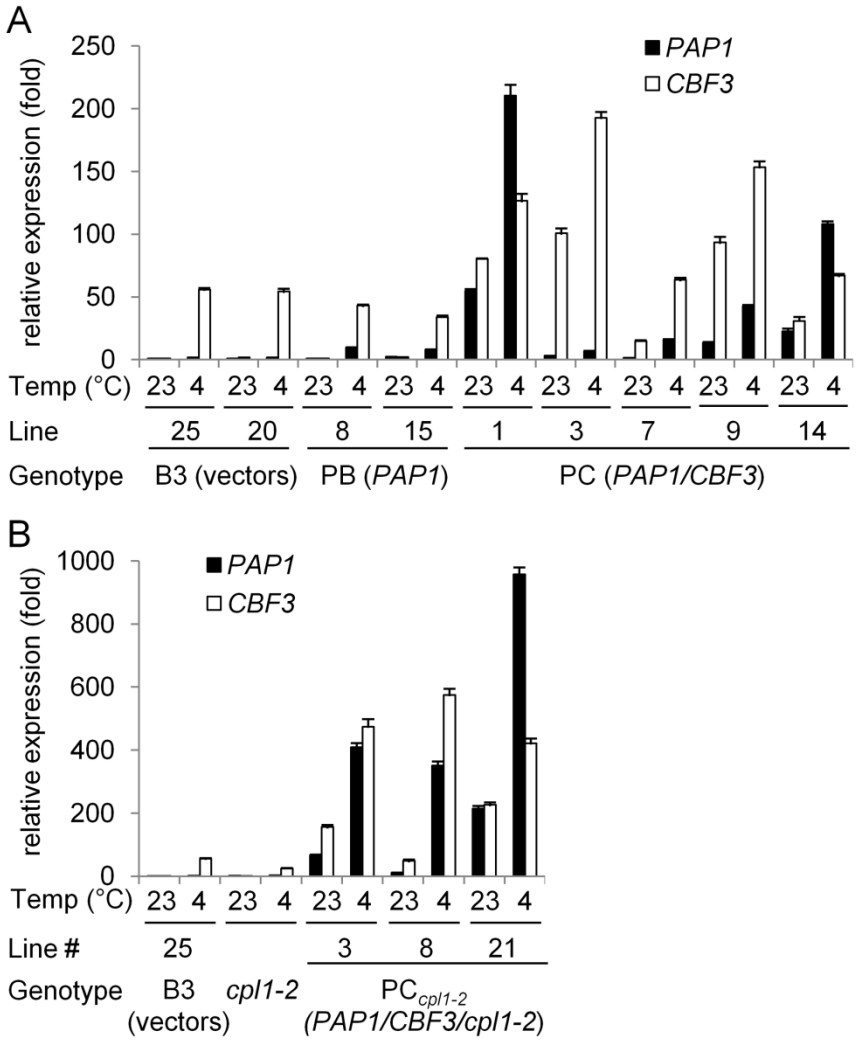

Figure 3.2. Relative gene expression in PAP1 and CBF3 transgenic plants.

PAP1 and *CBF3* expression levels were assayed using double homozygous transgenic T2 lines. cDNA was prepared with total RNA from 7-day-old seedling at room temperature (23˚C) and from seedlings treated with low temperature (4˚C) for 4 extra days. Expression folds of each gene are shown relative to the levels of double vector control line B3#25 grown at room temperature (23˚C). Bars indicate standard errors.

Table 3.2. Anthocyanin content measured as cyanidin-3-glucoside equivalents.

Plant extract was obtained from plants at 23˚C for 3 weeks (middle column) and at 4˚C for extra 3 weeks (right column). Absorbance of clear extract was measured at both 535 and 700 nm, and the value was converted into cyanidin-3-glucoside amount. Standard errors were shown in parentheses.

Line	Anthocyanin (±std) (μg cyanidin-3-glucoside/g fresh tissue)	
	23°C	4°C
B3#11	65 (7)	15 (10)
B3#18	79 (26)	37 (0)
B3#20	31 (10)	44 (2)
B3#25	21 (6)	51 (26)
B3#29	60 (24)	58 (2)
PB#3	56 (19)	18 (1)
PB#4	81 (13)	23 (0)
PB#8	149 (4)	31 (7)
PB#13	85 (12)	23 (10)
PB#15	71 (8)	39 (2)
PB#17	85 (9)	22 (20)
PC#1	34 (3)	41 (19)
PC#2	58 (1)	33 (12)
PC#3	13 (1)	70 (5)
PC#7	57 (22)	54 (26)
PC#9	95 (1)	37 (21)
PC#10	70 (1)	Not detected
PC#14	56 (14)	51 (11)
PC#26	40 (17)	61 (8)
PC$_{cpl1}$#3	330 (3)	468 (13)
PC $_{cpl1}$#8	455(0)	1371 (36)
PC $_{cpl1}$#21	523 (0.7)	929 (22)

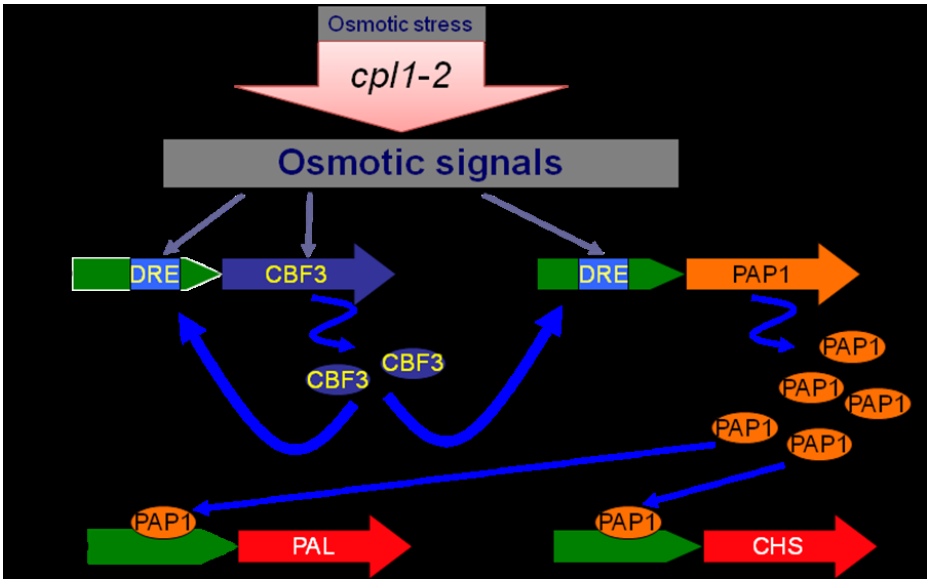

Figure 3.3. Design of the three component inducible system.

PAP1 transcription factor is placed downstream of a cold inducible RD29A promoter. In order to have better elevation of PAP1 expression, a cold responsive transcription factor CBF3 was also placed downstream of cold inducible RD29A promoter, to create a self amplification loop. Upon cold stress, this feed forward cassette will exert its affect on the same RD29A upstream of PAP1 as well. After the plants are grown big and healthy, cold treatment can then be applied leading to the overexpression of *PAP1* and hyperaccumulation of flavonoids.

enhanced the *PAP1* expression level up to 200-fold over unstressed vector control plants, indicating the *RD29A-CBF3* effecter indeed feedforwarded the *RD29A* promoter activity (Figure 3.2A). Surprisingly, B3, PB and PC lines showed similar level of transcripts encoding upstream phenylpropanoid pathway enzymes, such as *PAL* and *CHS* (data not shown). Total anthocyanin contents (as cyanidine 3-glucoside equivalent) of PB/PC lines were not substantially higher than vector control lines even after cold treatment for 3 weeks (Table 3.2).

Three-component system with *cpl2-1* background induced anthocyanin production under low temperature

It's conceivable that the lack of anthocyanin accumulation in PB and PC lines was due to insufficient level of *PAP1* expression even after cold-induction. To increase the efficiency of cold induction, the third component, *cpl1-2* mutation were incorporated. *Arabidopsis* host plants with *cpl1-2* mutation, which could induce *RD29A* promoter up to 10 fold higher than wild type, was used as a recipient of the *RD29A-PAP1* and *RD29A-CBF3* transgenes (Figure 3.3). These lines and vector control lines were designated as PC_{cpl1} and $B3_{cpl1}$, respectively, and homozygous plants were selected. During the selection, a high frequency of PC_{cpl1} lines with spotty pigmentations on the leaf surface was observed. Microscopic observations showed that these spots were trichomes accumulating anthocyanin pigments (Figure 3.4). Partial coloration was observed in leaf veins as well. Some individuals show high level of pigments in entire plant bodies and grew very slowly, which were not included in further analyses. RT-qPCR analyses indicated that 2 days cold treatment induced *PAP1* up to 950-fold in

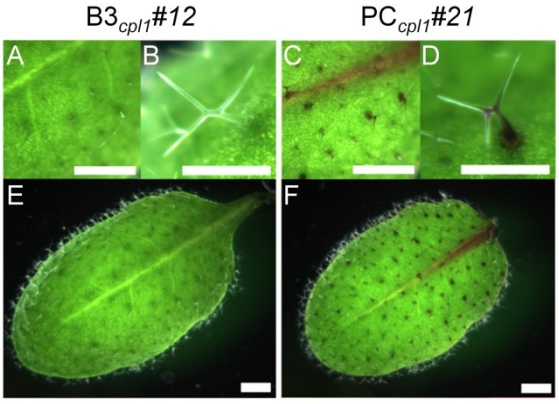

Figure 3.4. Purple pigments on non-treated plants.

Purple pigments were only observed in PC*cpl1-21* plants under normal growth condition (23˚C) for three weeks, but not in double homozygous transgenic B3*cpl1-12* (control) and. Scale bars are 1 mm in A, C, E and F, and 50 µm in B and D.

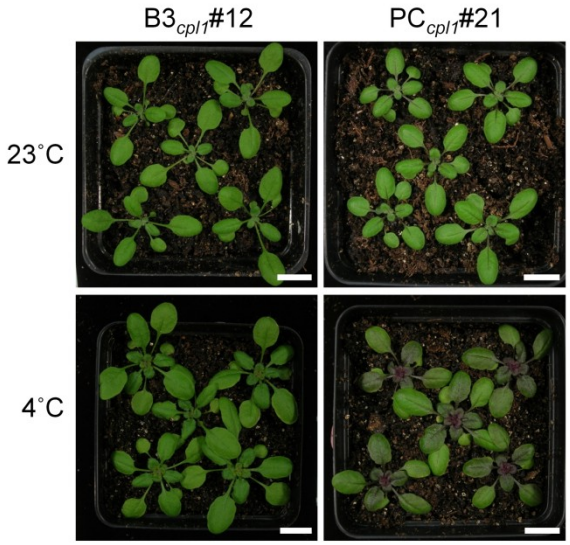

Figure 3.5. Pigmentation hyperaccumulation upon cold treatment.

Purple pigments in double homozygous transgenic B3$_{cpl1}$*#12* (control) and in PC$_{cpl1}$*#21* plants under normal growth condition (23˚C) for three weeks as well as cold treatment (4˚C) for an extra of three weeks. Scale bars are 1 cm.

PC_{cpl1} line over the vector control lines (Figure 3.2). Total anthocyanin analyses showed that PC_{cpl1} plants accumulated up to 30-fold more anthocyanin than vector control plants did (Table 3.2). The levels of anthocyanin produced in PC_{cpl1} lines were comparable to the level produced in constitutive overexpression of *PAP1* by CaMV 35S promoter (Tohge *et al.*, 2005). These results indicate that the three-component system is necessary to induce anthocyanin biosynthetic pathway above the threshold level. Since PC_{cpl1} line 21 consistently induced *PAP1* and anthocyanin to high level, this line was used for further analysis. For negative control purposes, double vector control plants (B3) in *cpl1-2* background were selected.

Gene expression profile of three-component transgenic plants during cold activation

In order to understand the efficiency of the cold-inducible three-component system, a time course of gene expression was determined during a long-term cold induction. Three-week old PC_{cpl1} and $B3_{cpl1}$ plants were exposed to 4°C for up to additional 3 weeks. Cold treatment greatly enhanced the purple pigmentation accumulation in PC_{cpl1} line, especially in newly developed rosette leaves near the center meristem (Figure 3.5). Cold treatments longer than 3 weeks induced senescence of plants and therefore were not included in the analysis. To determine the effect of *PAP1* overexpression on anthocyanin biosynthesis pathway genes, expression of the transgenes were analyzed, as well as some rate limiting genes in overall flavonoid biosynthesis pathway where flavonol synthases branch out both kaempferol and quercetin production from

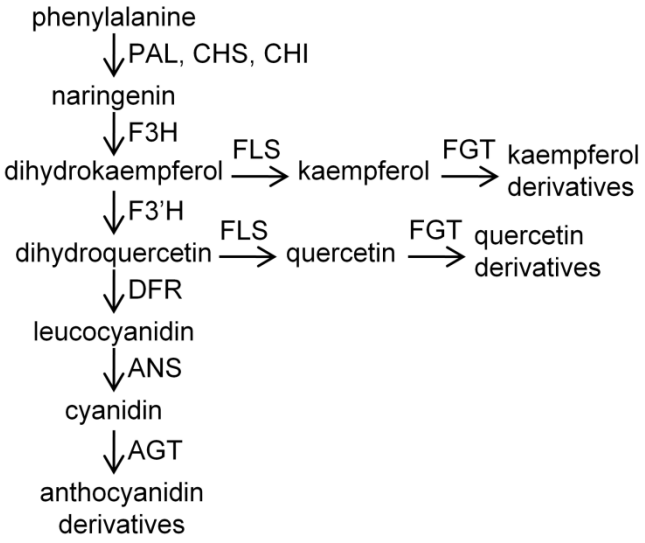

Figure 3.6. Summary of the flavonoid biosynthesis pathway.

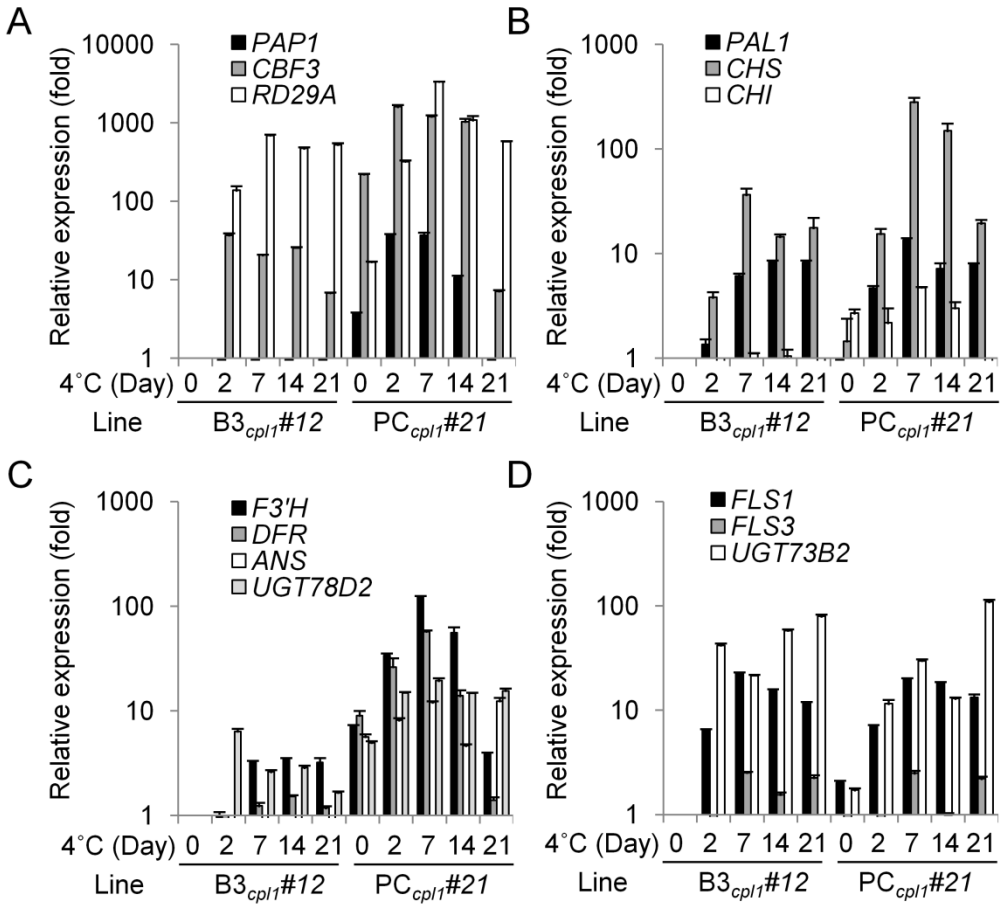

Figure 3.7. Relative gene expression in PC_{cpl1} #21 and $B3_{cpl1}$ #12 lines under different length of cold treatment determined by qRT-PCR.

cDNA was prepared with total RNA from shoots of 3 week old plant leaves grown at room temperature (23˚C) and from plants treated with low temperature (4˚C) for addition 2 days, 1 week, 2 weeks and 3 weeks. Expression folds of each gene were shown relative to the levels of double vector control line B3*cpl1*-12 grown at room temperature (23˚C). Bars indicate standard errors. PAL1: phenylalanine ammonia lyase, CHS: chalcone synthase, CHI: chalcone isomerase, F3'H: flavonoid 3'-hydroxylase, DFR: dihydroflavonol reductase, ANS: leucoanthocyanidin dioxygenase, FLS1: flavonol synthase 1, FLS3: flavonol synthase 3, UGT73B2: UDP-glucosyltransferase 73B2 (flavonol 3-O-glucosyltransferase activity), UGT78D2*:* UDP-glucosyltransferase 78D2 (anthocyanidin 3-O-glucosyltransferase).

anthocyanin pathway and can affect the downstream anthocyanin accumulation accordingly (Figure 3.6). In PC_{cpl1}, expression of *CBF3* and *PAP1* reached their highest levels (1602 fold and 37 fold, respectively) after 2 days and slowly declined after 1 week (Figure 3.7A). After 3 weeks of cold treatment, the *PAP1* level was similar to that of vector control plants. Expression of *PAL* was induced both in cold-treated PC_{cpl1} and $B3_{cpl1}$ plants, albeit PC_{cpl1} plants showed slightly faster response and higher expression level (Figure 3.7B). Genes that lead to anthocyanin biosynthesis, such as *CHS, CHI, F3'H, DFR*, and *ANS* were all expressed higher in cold-treated PC_{cpl1} plants (Figure 3.7C). In contrast, cold treatment induced expression of synthase (FLS1) and flavonol glucosyltransferases (UGT73B2) both in PC_{cpl1} and $B3_{cpl1}$ plants to the similar levels (Figure 3.7D). These results indicated that the three-component system effectively activated anthocyanin biosynthesis pathway, whereas cold treatment itself induced flavonol biosynthesis pathway genes independent of the three-component system. In addition, gene expression data indicated that induction of anthocyanin biosynthesis pathway persisted until plants start to senesce after three weeks of cold treatment.

Identification of hyperaccumulated flavonoid compounds by LC-MS

In PC_{cpl1} plants, cold induction of phenylpropanoid pathway genes were accompanied with accumulation of anthocyanin pigments throughout the aerial part of plant bodies, indicating that the cold-induction system indeed increased biosynthetic capacity of flavonoids in transgenic plants. In order to determine whether anthocyanins phytochemicals produced via the three-component system are similar to those produced

by constitutive overexpression of PAP1, profiles of anthocyanins and other flavonoids produced in transgenic plants were analyzed. Phytochemicals were extracted from 3 weeks old PC_{cpl1} and $B3_{cpl1}$ plants grown at room temperature (23˚C) and plants with three extra weeks of cold treatment (4˚C). Putative flavonoid compounds were identified by their molecular mass, their UV-visible absorption spectra in LC and their fragmentation patterns by MS-MS, which were compared to the reported profiles. The order of the compound eluates was also taken into consideration of their identification.

A total of five anthocyanin (cyanidin derivatives) (A5, A8-A11) have been identified in PC_{cpl1} plants after three weeks of cold treatment (Table 3.3). The nomenclatures for the numbers after letter A were consistent with previous reports on *PAP1* overexpression plants (Tohge et al., 2005; Rowan et al., 2009; Li et al., 2010; Shi and Xie, 2010). Amount of each compounds were determined as cyanidin-3-glucoside equivalent. Among the five anthocyanin compounds detected at 520nm in cold induced PC_{cpl1} plants (Table 3.3), A11 was the most abundant, accounting for 62-63% of the total anthocyanins under cold treatment. A11, which contains 3 acyl moieties and 4 glycosides, is also the most highly modified among the identified anthocyanins.

Several kaempferol and quercetin derivatives were also identified in this study in cold treated PC_{cpl1} #21 (Figure 3.8). At 21.91 min, F1 was eluted. F1 had a molecular mass of 579 (positive mode) and gave two daughter ion with an m/z of 287 and 433. Compared with the published HPLC/PDA chromatograms of PAP1 overexpression Arabidopsis, F1 could be Kaempferol 3-*O*-rhamnoside 7-*O*-rhamnoside. At 19.21min, a

Table 3.3. Compound identification by LC-MS.

Plant extract was obtained from PC_{cpl1}#21 lines grown at 23˚C for 3 weeks and at 4˚C for extra 3 weeks. Filtered extract was analyzed via LC-MS. MS2 and MS3 fragmentation were performed to identify the daughter ions.

	Compound	λ_{max}	Rt (min)	$[M+H]^+$	Fragmention $[M+H]^+$		
A8	Cyanidin 3-*O*-[2"-*O*-(xylosyl) 6"-*O*-(*p*-*O*-(glucosyl) *p*-coumaroyl) glucoside] 5-*O*-[6'"-*O*-(malonyl) glucoside]	282, 523	17.55	1137	889	535	287
A10	Cyanidin 3-*O*-[2"-*O*-(2'"-*O*-(sinapoyl) xylosyl) 6"-*O*-(*p*-*O*-(glucosyl) *p*-coumaroyl) glucoside] 5-*O*-glucoside	280, 532	18.27	1257	1095	449	287
A11	Cyanidin 3-*O*-[2"-*O*-(6'"-*O*-(sinapoyl) xylosyl) 6"-*O*-(*p*-*O*-(glucosyl)-*p*-coumaroyl) glucoside] 5-*O*-(6""-*O*-malonyl) glucoside	284, 535	18.91	1343	1095	535	287
F1	Kaempferol 3-*O*-rhamnoside 7-*O*-rhamnoside	264, 336	20.91	579	433	287	
A5	Cyanidin 3-*O*-[2"-*O*-(xylosyl)-6"-*O*-(*p*-coumaroyl) glucoside] 5-*O*-malonylglucoside	303, 525	21.24	975	727	535	287
A9	Cyanidin 3-*O*-[2"-*O*-(2'"-*O*-(sinapoyl) xylosyl) 6"-*O*-(*p*-*O*-coumaroyl) glucoside] 5-*O*-[6""-*O*-(malonyl) glucoside]	298, 532	22.58	1181	993	535	287

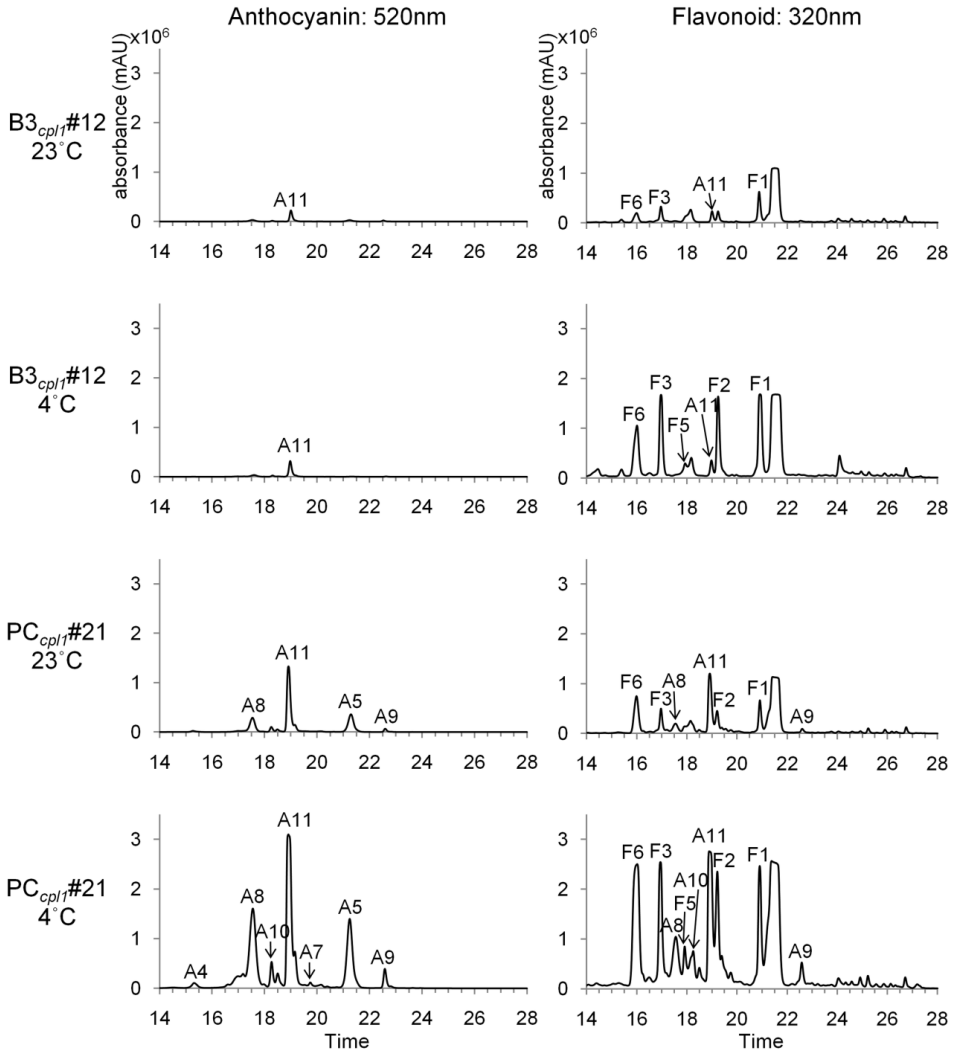

Figure 3.8. Plant flavonoid profiling by HPLC/PDA chromatograms.

PC_{cpl1}#21 and $B3_{cpl1}$#12 plants were grown under normal growth condition (23˚C) or additional 3 weeks of cold treatment (4˚C). A5, A8-A11 stands for cyanidin derivatives, as listed in Table 3.3. A4 and A7 were speculated as cyanidin 3-*O*-[2"-*O*-(2"'-*O*-(sinapoyl) xylosyl) glucoside] 5-*O*-glucoside, and cyanidin 3-*O*-[2"-*O*-(2"'-*O*-(sinapoyl) xylosyl) 6"-*O*-(*p*-coumaroyl) glucoside] 5-*O*-glucoside. F1, F2 and F3 were speculated as kaempferol 3-*O*-rhamnoside 7-*O*-rhamnoside, kaempferol 3-*O*-glucoside 7-*O*-rhamnoside, and kaempferol 3-*O*-[6"-*O*-(rhamnosyl) glucoside] 7-*O*-rhamnoside. F5 and F6 were speculated as quercetin 3-*O*-glucoside 7-*O*-rhamnoside, and quercetin 3-*O*-[6"-*O*-(rhamnosyl) glucoside] 7-*O*-rhamnoside. All these speculation about cyanidin derivatives, kaempferol glycosides and quercetin glycosides are according to their order of elution, spectrum, molecular mass, and partial fragmentation pattern information in accordance with references.

compound with an m/z of 595 was eluted. Because the fragmentation pattern of this ion couldn't be obtained, F2 could be a kaempferol derivative, kaempferol 3-*O*-glucoside 7-*O*-rhamnoside, or a quercetin derivative, quercetin 3-*O*-rhamnoside 7-*O*-rhamnoside. It's assigned as F2, due to the assumption that it's more likely being a kaempferol derivative, as other major speculated kaempferol derivatives are cold inducible in both vector control lines and PC_{cpl1} lines (Figure 3.8). F3, eluted at 16.95 between the peaking containing F6 and peak A8, had a molecular mass of 741, was hypothesized as kaempferol 3-*O*-[6"-*O*-(rhamnosyl) glucoside] 7-*O*-rhamnoside, in accordance to previous publication. F5, with an m/z of 611, could be quercetin 3-*O*-glucoside 7-*O*-rhamnoside, but with the limited amount of F5, clear fragmentation pattern couldn't be acquired. F6, on the other hand, represented a mixture of at least two compounds. One of the mixtures, with an m/z of 757, could be quercetin 3-*O*-[6"-*O*-(rhamnosyl) glucoside] 7-*O*-rhamnoside. However, better separation of the peaks is necessary to fully evaluate the peak composition and quantity of the quercetin derivatives.

Given the same injection amount of the extraction from 4 different samples, at 520 nm, only small amount of A11 were accumulated in uninduced $BC3_{cpl1}$ plants (Figure 3.8). At 320 nm, minimum amount of F1, F3 and F6 formed peaks in LC chromatograms. Cold treatment greatly induced hypothetical kaempferol derivatives (F1, F2, F3), but had little effect on anthocyanin accumulation. Cold treatment also moderately induced hypothetical quercetin F6, and F5 started to form a peak. PAP1 overexpression in PC_{cpl1} plants specifically induced anthocyanins (A8, A11, A5 and A9) and quercetin derivative F6 compared with $B3_{cpl1}$ plants. However, PC_{cpl1} plants had

similar hypothetical kaempferol amounts (F1, F2, and F3) as B3$_{cpl1}$ plants. This agreed with the expression profile of anthocyanin biosynthesis pathway genes (Figure 3.7). However, major anthocyanins A8, A11, and A5 were most accumulated in the three component inducible system after the cold stimulus. Synergistic effect of both PAP1 overexpression and cold treatment on quercetin (F5 and F6) was observed.

DISCUSSION

Inducible production of phytochemicals is a strategy commonly used in industrial culturing processes, however, the concept has not been widely adapted in genetic engineering of plant metabolism. In this study, cold induction was chosen to activate production of subset of flavonoids and regulate anthocyanin production. Compared to the anthocyanin profile obtained from previous constitutive overexpression (Tohge et al., 2005; Rowan et al., 2009), the amount of individual anthocyanins in cold treated PC$_{cpl1}$ was up to 5 fold higher (Figure 3.8 and 3.9). This was similar to the level obtained when *pap1-D* plants were exposed to light stress (Shi and Xie, 2010). These observations indicated that ectopic PAP1 expression did not fully activate the all rate limiting step of anthocyanin biosynthesis, and further activation of phenylpropanoid pathway require additional environmental signals. In this case, cold treatment induced kaempferol biosynthesis both in B3$_{cpl1}$ and PC$_{cpl1}$ plants, albeit PC$_{cpl1}$ plants produced less kaempferol and more anthocyanins and quercetins upon cold treatment. Apparently, cold treatment in PC$_{cpl1}$ induced a sufficient metabolic flow to dihydrokaempferol, for which F3'H successfully compete with FLS. This contrasts with the case of isoflavone

synthase overexpression, where isoflavone was overproduced only when competing pathway was turned off (Lozovaya *et al.*, 2007). It has been proposed that flavonoid pathway enzymes form supercomplex and channel metabolites (Sweetlove and Fernie, 2005). Perhaps, sufficient amount of native F3'H, which were induced by PAP1, can associate with the complex even in the presence of FLS. At the downstream, in contrast, FLS successfully competed with DFR and produced quercetin. The resulting phytochemical profile showed higher level of anthocyanins, kaempferols, and quercetins. This contrast with the case of *pap1-D*, in which anthocyanin accumulation was accompanied with substantial decrease of kaempferols and only small amount of quercetin was produced (Tohge *et al.*, 2005).

In this system, it was necessary to have all three components and cold induction to produce large amount of anthocyanins and flavonoids. Plants with only two components did not produce anthocyanins more than control plants. Indeed, although the PAP1 expression levels of some cold-induced PC plants and uninduced PC_{cpl1} plants were similar, only uninduced PC_{cpl1} plants showed elevated anthocyanin accumulations. It's conceivable that these differences were caused by different expression levels of *PAP1* in each cell. Perhaps, although cold-induced PC plants and uninduced PC_{cpl1} plants showed similar total *PAP1* mRNA levels, distribution of *PAP1* transcripts are different between these plants. *PAP1* expression likely is more restricted to specific tissues, such as trichomes, in uninduced PC_{cpl1} plants, whereas lower but even expression occurs in cold-induced PC plants. Such difference could render above-threshold level *PAP1* expression in some uninduced PC_{cpl1} tissues but not in cold-induced PC plants.

In conclusion, three-component system consisting of RD29a-PAP1, RD29a-CBF3, and the cpl1 mutation was shown to effectively enhance specific anthocyanin accumulation. Expression of *PAP1* using inducible three-component system can minimize severe vegetative growth inhibition caused by the constitutive expression of transgenes. Unlike several inducible systems, such as dexamethasone inducible system, the three component system described here do not require any constitutive expression of the system components, therefore, expected to be more resistant to gene silencing. Since *cpl1* mutation can enhance expression of other inducible promoters in addition to osmotic stress pathway genes (Koiwa et al., 2004), the three-component system with *cpl1* are applicable for other inducible promoter-transcription factor combinations.

CHAPTER IV

CONCLUSIONS

Plant CTD phosphatases have very important roles in RNA polymerase II transcription regulation. The first section of this project focused on understanding the biochemical and biological function of the small SCP1 family proteins and describing the phosphorylation status of the CTD heptapeptide arrays at different stages of transcription. Ubiquitously expressed and highly homologous SSP isoforms exhibited different subcellular localizations. Notably, recombinant SSP proteins showed highly specific CTD phosphatase activities with distinct Ser-PO_4 position preferences. SSP4 and SSP4b had both Ser2-PO_4 and Ser5-PO_4 phosphatase activity while SSP5 was only capable of Ser5-PO_4 dephosphorylation. Previous to these results, no Ser2-specific CTD phosphatase activities had been reported and complement the current knowledge in plant transcription elongation processes. This system will be essential to produce active recombinant SSPs for further dissection of substrate specificities in SSP family members and the mechanism contributing to those specificities.

Phenotypic abnormality in *ssp5* and *ssp4 ssp4b ssp5* mutants is relatively mild under stress, thus other phosphatase mutants with obvious phenotypes could provide deeper knowledge when used as a platform to study plant inducible gene expression systems. To study this, a single-component self-activation loop of *RD29A-CBF3* in plant CTD phosphatase mutant *cpl1-2* background was shown to sufficiently induce expression of *CBF3* and cold-tolerance phenotype specifically under low temperature. *RD29A-PAP1* was selected as another component that induces anthocyanin hyperaccumulation and that

can easily be monitored. The results show that a three-component system using a native plant promoter is capable of promoting high expression of transgenes upon induction without compromising normal plant growth and development, resulting in enhanced performance of engineered plants.

Compared with constitutive overexpression systems, the proposed three-component system is capable of producing a unique profile of phytochemicals, featuring higher anthocyanin accumulations. Improved separation of plant extracts and enhanced daughter ion signals from fragmentation would provide better understanding of the kaempferol and quercetin profile of PC_{cpl1} plants, especially upon cold treatment. It is highly possible that this system and its variants will be valuable tools that contribute to a broad range of activities, such as metabolic and physiological engineering, and heterologous protein expression strategies for plant environmental responses studies.

REFERENCES

Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., Shinn, P., Stevenson, D.K., Zimmerman, J., Barajas, P., Cheuk, R., Gadrinab, C., Heller, C., Jeske, A., Koesema, E., Meyers, C.C., Parker, H., Prednis, L., Ansari, Y., Choy, N., Deen, H., Geralt, M., Hazari, N., Hom, E., Karnes, M., Mulholland, C., Ndubaku, R., Schmidt, I., Guzman, P., Aguilar-Henonin, L., Schmid, M., Weigel, D., Carter, D.E., Marchand, T., Risseeuw, E., Brogden, D., Zeko, A., Crosby, W.L., Berry, C.C., and Ecker, J.R. (2003). Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. Science **301,** 653-657.

Archambault, J., Chambers, R., Kobor, M.S., Ho, Y., Bolotin, D., Andrews, B., Kane, C.M., and Greenblatt, J. (1997). An essential component of a C-terminal domain phosphatase that interacts with transcription factor IIF in *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA **94,** 14300-14305.

Archambault, J., Pan, G., Dahmus, G.K., Cartier, M., Marshall, N., Zhang, S., Dahmus, M.E., and Greenblatt, J. (1998). Fcp1, the Rap74-interacting subunit of a human protein phosphatase that dephosphorylates the carboxyl-terminal domain of RNA polymerase II. J. Biol. Chem. **273,** 27593-27601.

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

Bang, W., Kim, S., Ueda, A., Vikram, M., Yun, D., Bressan, R.A., Hasegawa, P.M., Bahk, J., and Koiwa, H. (2006). *Arabidopsis* carboxyl-terminal domain phosphatase-like isoforms share common catalytic and interaction domains but have distinct in planta functions. Plant Physiol. **142,** 586-594.

Bellec, Y., Harrar, Y., Butaeye, C., Darnet, S., Bellini, C., and Faure, J.D. (2002). *Pasticcino2* is a protein tyrosine phosphatase-like involved in cell proliferation and differentiation in *Arabidopsis*. Plant J. **32,** 713-722.

Borevitz, J.O., Xia, Y., Blount, J., Dixon, R.A., and Lamb, C. (2000). Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. Plant Cell **12,** 2383-2394.

Brown, D.E., Rashotte, A.M., Murphy, A.S., Normanly, J., Tague, B.W., Peer, W.A., Taiz, L., and Muday, G.K. (2001). Flavonoids act as negative regulators of auxin transport in vivo in *Arabidopsis*. Plant Physiol. **126,** 524-535.

Camilleri, C., Azimzadeh, J., Pastuglia, M., Bellini, C., Grandjean, O., and Bouchez, D. (2002). The *Arabidopsis TONNEAU2* gene encodes a putative novel protein phosphatase 2A regulatory subunit essential for the control of the cortical cytoskeleton. Plant Cell **14,** 833-845.

Chao, D.M., and Young, R.A. (1991). Tailored tails and transcription initiation: the carboxyl terminal domain of rna polymerase II. Gene Expr. **1,** 1-4.

Chi, Y.H., Salzman, R.A., Balfe, S., Ahn, J.E., Sun, W., Moon, J., Yun, D.J., Lee, S.Y., Higgins, T.J., Pittendrigh, B., Murdock, L.L., and Zhu-Salzman, K.

(2009). Cowpea bruchid midgut transcriptome response to a soybean cystatin--costs and benefits of counter-defence. Insect Mol. Biol. **18,** 97-110.

Choi, H., Hong, J., Ha, J., Kang, J., and Kim, S.Y. (1999). ABFs, a family of ABA-responsive element binding factors. J. Biol. Chem. **275,** 1723-1730.

Cook, D., Fowler, S., Fiehn, O., and Thomashow, M.F. (2004). A prominent role for the CBF cold response pathway in configuring the low-temperature metabolome of *Arabidopsis*. Proc. Natl. Acad. Sci. USA **101,** 15243-15248.

Corden, J.L. (1990). Tails of RNA Polymerase II. Trends Biochem. Sci. **15,** 383-387.

Curtis, M.D., and Grossniklaus, U. (2003). A gateway cloning vector set for high-throughput functional analysis of genes in planta. Plant Physiol. **133,** 462-469.

Dixon, R.A., and Paiva, N.L. (1995). Stress-induced phenylpropanoid metabolism. Plant Cell **7,** 1085-1097.

Ferrer, J.L., Austin, M.B., Stewart, C., Jr., and Noel, J.P. (2008). Structure and function of enzymes involved in the biosynthesis of phenylpropanoids. Plant Physiol. Biochem. **46,** 356-370.

Finkelstein, R.R., Wang, M.L., Lynch, T.J., Rao, S., and Goodman, H.M. (1998). The *Arabidopsis* abscisic acid response locus ABI4 encodes an APETALA 2 domain protein. Plant Cell **10,** 1043-1054.

Fordham-Skelton, A.P., Chilley, P., Lumbreras, V., Reignoux, S., Fenton, T.R., Dahm, C.C., Pages, M., and Gatehouse, J.A. (2002). A novel higher plant protein tyrosine phosphatase interacts with SNF1-related protein kinases via a KIS (kinase interaction sequence) domain. Plant J. **29,** 705-715.

Fujii, H., Chinnusamy, V., Rodrigues, A., Rubio, S., Antoni, R., Park, S.Y., Cutler, S.R., Sheen, J., Rodriguez, P.L., and Zhu, J.K. (2009). In vitro reconstitution of an abscisic acid signalling pathway. Nature **462,** 660-664.

Fuleki, T., and Francis, F.J. (1968). Quantitative methods for anthocyanins. J. Food. Sci. **33,** 78-83.

Galvano, F., La Fauci, L., Lazzarino, G., Fogliano, V., Ritieni, A., Ciappellano, S., Battistini, N.C., Tavazzi, B., and Galvano, G. (2004). Cyanidins: metabolism and biological properties. J. Nutr. Biochem. **15,** 2-11.

Gangopadhyay, S.S., Gallant, C., Sundberg, E.J., Lane, W.S., and Morgan, K.G. (2008). Regulation of $Ca^{2+/}$calmodulin kinase II by a small C-terminal domain phosphatase. Biochem. J. **412,** 507-516.

Gao, P., Xin, Z.Y., and Zheng, Z.L. (2008). The OSU1/QUA2/TSD2-encoded putative methyltransferase is a critical modulator of carbon and nitrogen nutrient balance response in *Arabidopsis*. PLoS One **3,** e1387

Geekiyanage, S., Takase, T., Ogura, Y., and Kiyosue, T. (2007). Anthocyanin production by over-expression of grape transcription factor gene *VlmybA2* in transgenic tobacco and *Arabidopsis*. Plant Biotechnol. Rep. **1,** 11-18.

Ghosh, A., Shuman, S., and Lima, C.D. (2008). The structure of Fcp1, an essential RNA polymerase II CTD phosphatase. Mol. Cell **32,** 478-490.

Gong, Z., Lee, H., Xiong, L., Jagendorf, A., Stevenson, B., and Zhu, J.K. (2002). RNA helicase-like protein as an early regulator of transcription factors for plant chilling and freezing tolerance. Proc. Natl. Acad. Sci. USA **99,** 11507-11512.

Guo, H.S., Fei, J.F., Xie, Q., and Chua, N.H. (2003). A chemical-regulated inducible RNAi system in plants. Plant J. **34,** 383-392.

Hannah, M.A., Wiese, D., Freund, S., Fiehn, O., Heyer, A.G., and Hincha, D.K. (2006). Natural genetic variation of freezing tolerance in *Arabidopsis*. Plant Physiol. **142,** 98-112.

Hausmann, S., and Shuman, S. (2002). Characterization of the CTD phosphatase Fcp1 from fission yeast. J. Biol. Chem. **277,** 21213-21220.

Hausmann, S., Schwer, B., and Shuman, S. (2004). An *Encephalitozoon cuniculi* ortholog of the CTD serine phosphatase Fcp1. Biochemistry **43,** 7111-7120.

Hausmann, S., Erdjument-Bromage, H., and Shuman, S. (2004). *Schizosaccharomyces pombe* carboxyl-terminal domain (CTD) phosphatase Fcp1: distributive mechanism, minimal CTD substrate, and active site mapping. J. Biol. Chem. **279,** 10892-10900.

Hausmann, S., Koiwa, H., Krishnamurthy, S., Hampsey, M., and Shuman, S. (2005). Different strategies for carboxyl-terminal domain (CTD) recognition by serine 5-specific CTD phosphatases. J. Biol. Chem. **280,** 37681-37688.

Himmelbach, A., Yang, Y., and Grill, E. (2003). Relay and control of abscisic acid signaling. Curr. Opin. Plant biol. **6,** 470-479.

Hsieh, L.C., Lin, S.I., Shih, A.C.C., Chen, J.W., Lin, W.Y., Tseng, C.Y., Li, W.H., and Chiou, T.J. (2009). Uncovering small RNA-mediated responses to phosphate deficiency in *Arabidopsis* by deep sequencing. Plant Physiol. **151,** 2120-2132.

Hugouvieux, V., Kwak, J.M., and Schroeder, J.I. (2001). An mRNA cap binding protein, ABH1, modulates early abscisic acid signal transduction in *Arabidopsis*. Cell **106,** 477-487.

Iriti, M., and Faoro, F. (2009). Bioactivity of grape chemicals for human health. Nat. Prod. Commun. **4,** 611-634.

Ishitani, M., Xiong, L., Stevenson, B., and Zhu, J.-K. (1997). Genetic analysis of osmotic and cold stress signal transduction in *Arabidopsis*: interactions and convergence of abscisic acid-dependent and abscisic acid-independent pathways. Plant Cell **9,** 1935-1949.

Ji, H., Kim, S.R., Kim, Y.H., Kim, H., Eun, M.Y., Jin, I.D., Cha, Y.S., Yun, D.W., Ahn, B.O., Lee, M.C., Lee, G.S., Yoon, U.H., Lee, J.S., Lee, Y.H., Suh, S.C., Jiang, W., Yang, J.I., Jin, P., McCouch, S.R., An, G., and Koh, H.J. (2010). Inactivation of the CTD phosphatase-like gene *OsCPL1* enhances the development of the abscission layer and seed shattering in rice. Plant J. **61,** 96-106.

Kaplan, F., Kopka, J., Haskell, D.W., Zhao, W., Schiller, K.C., Gatzke, N., Sung, D.Y., and Guy, C.L. (2004). Exploring the temperature-stress metabolome of *Arabidopsis*. Plant Physiol. **136,** 4159-4168.

Kasuga, M., Liu, Q., Miura, S., Yamaguchi-Shinozaki, K., and Shinozaki, K. (1999). Improving plant drought, salt, and freezing tolerance by gene transfer of a single stress-inducible transcription factor. Nat. Biotech. **17,** 287-291.

Keutgen, A.J., and Pawelzik, E. (2007). Modifications of strawberry fruit antioxidant pools and fruit quality under NaCl stress. J. Agr. Food Chem. **55,** 4066-4072.

Knockaert, M., Sapkota, G., Alarcon, C., Massague, J., and Brivanlou, A.H. (2006). Unique players in the BMP pathway: small C-terminal domain phosphatases dephosphorylate Smad1 to attenuate BMP signaling. Proc. Natl. Acad. Sci. USA **103,** 11940-11945.

Kobor, M.S., Archambault, J., Lester, W., Holstege, F.C.P., Gileadi, O., Jansma, D.B., Jennings, E.G., Kouyoumdjian, F., Davidson, A.R., Young, R.A., and Greenblatt, J. (1999). An unusual eukaryotic protein phosphatase required for transcription by RNA Polymerase II and CTD dephosphorylation in *S. cerevisiae*. Mol. Cell **4,** 55-62.

Koiwa, H. (2006). Phosphorylation of RNA polymerase II C-terminal domain and plant osmotic-stress responses. In abiotic stress tolerance in plants: Toward the improvement of global environment and food, T.T. Ashwani K. Rai, ed (Dordrecht, The Netherlands: Springer), pp. 47-57.

Koiwa, H., Barb, A.W., Xiong, L., Li, F., McCully, M.G., Lee, B.-h., Sokolchik, I., Zhu, J., Gong, Z., Reddy, M., Sharkhuu, A., Manabe, Y., Yokoi, S., Zhu, J.-K., Bressan, R.A., and Hasegawa, P.M. (2002). C-terminal domain phosphatase-like family members (AtCPLs) differentially regulate *Arabidopsis thaliana* abiotic stress signaling, growth, and development. Proc. Natl. Acad. Sci. USA **99,** 10893-10898.

Koiwa, H., Hausmann, S., Bang, W.Y., Ueda, A., Kondo, N., Hiraguri, A., Fukuhara, T., Bahk, J.D., Yun, D.J., Bressan, R.A., Hasegawa, P.M., and Shuman, S. (2004). *Arabidopsis* C-terminal domain phosphatase-like 1 and 2 are essential Ser-5-specific C-terminal domain phosphatases. Proc. Natl. Acad. Sci. USA **101,** 14539-14544.

Korkina, L.G. (2007). Phenylpropanoids as naturally occurring antioxidants: from plant defense to human health. Cell Mol. Biol. **53,** 15-25.

Korn, M., Peterek, S., Mock, H.P., Heyer, A.G., and Hincha, D.K. (2008). Heterosis in the freezing tolerance, and sugar and flavonoid contents of crosses between *Arabidopsis thaliana* accessions of widely varying freezing tolerance. Plant Cell Envir. **31,** 813-827.

Kranz, H.D., Denekamp, M., Greco, R., Jin, H., Leyva, A., Meissner, R.C., Petroni, K., Urzainqui, A., Bevan, M., Martin, C., Smeekens, S., Tonelli, C., Paz-Ares, J., and Weisshaar, B. (1998). Towards functional characterisation of the members of the R2R3-MYB gene family from *Arabidopsis thaliana*. Plant J. **16,** 263-276.

Kubo, H., Peeters, A.J.M., Aarts, M.G.M., Pereira, A., and Koornneef, M. (1999). *ANTHOCYANINLESS2*, a homeobox gene affecting anthocyanin distribution and root development in *Arabidopsis*. Plant Cell **11,** 1217-1226.

Kuhn, J.M., and Schroeder, J.I. (2003). Impacts of altered RNA metabolism on abscisic acid signaling. Curr. Opin. Plant Biol. **6,** 463-469.

Lea, U.S., Slimestad, R., Smedvig, P., and Lillo, C. (2007). Nitrogen deficiency enhances expression of specific MYB and bHLH transcription factors and accumulation of end products in the flavonoid pathway. Planta **225,** 1245-1253.

Lee, J.M., and Greenleaf, A.L. (1997). Modulation of RNA polymerase II elongation efficiency by C-terminal heptapeptide repeat domain kinase I. J. Biol. Chem. **272,** 10990-10993.

Li, X., Gao, M.J., Pan, H.Y., Cui, D.J., and Gruber, M.Y. (2010). Purple canola: *Arabidopsis PAP1* increases antioxidants and phenolics in *Brassica napus* leaves. J. Agric. Food Chem. **58,** 1639-1645.

Lillo, C., Lea, U.S., and Ruoff, P. (2008). Nutrient depletion as a key factor for manipulating gene expression and product formation in different branches of the flavonoid pathway. Plant Cell Environ. **31,** 587-601.

Lin, P.S., Marshall, N.F., and Dahmus, M.E. (2002). CTD phosphatase: role in RNA polymerase II cycling and the regulation of transcript elongation. Prog. Nucleic Acid Res. Mol. Biol. **72,** 333-365.

Lo Piero, A.R., Puglisi, I., Rapisarda, P., and Petrone, G. (2005). Anthocyanins accumulation and related gene expression in red orange fruit induced by low temperature storage. J. Agric. Food. Chem. **53,** 9083-9088.

Lozovaya, V.V., Lygin, A.V., Zernova, O.V., Ulanov, A.V., Li, S.X., Hartman, G.L., and Widholm, J.M. (2007). Modification of phenolic metabolism in soybean hairy roots through down regulation of chalcone synthase or isoflavone synthase. Planta **225,** 665-679.

Lu, C., and Fedoroff, N. (2000). A mutation in the *Arabidopsis HYL1* gene encoding a dsRNA binding protein affects responses to abscisic acid, auxin, and cytokinin. Plant Cell **12,** 2351-2366.

Luan, S. (2003). Protein phosphatases in plants. Annu. Rev. Plant Biol. **54,** 63-92.

Lunde, C., Zygadlo, A., Simonsen, H.T., Nielsen, P.L., Blennow, A., and Haldrup, A. (2008). Sulfur starvation in rice: the effect on photosynthesis, carbohydrate metabolism, and oxidative stress protective pathways. Plant Physiol. **134,** 508-521.

Ma, Y., Szostkiewicz, I., Korte, A., Moes, D., Yang, Y., Christmann, A., and Grill, E. (2009). Regulators of PP2C phosphatase activity function as abscisic acid sensors. Science **324,** 1064-1068.

Marczak, L., Kachlicki, P., Kozniewski, P., Skirycz, A., Krajewski, P., and Stobiecki, M. (2008). Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry monitoring of anthocyanins in extracts from *Arabidopsis thaliana* leaves. Rapid. Commun. Mass. Sp. **22,** 3949-3956.

Monroe-Augustus, M., Zolman, B.K., and Bartel, B. (2003). IBR5, a dual-specificity phosphatase-like protein modulating auxin and abscisic acid responsiveness in *Arabidopsis*. Plant Cell **15,** 2979-2991.

Narusaka, Y., Nakashima, K., Shinwari, Z.K., Sakuma, Y., Furihata, T., Abe, H., Narusaka, M., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2003). Interaction between two *cis*-acting elements, ABRE and DRE, in ABA-

dependent expression of *Arabidopsis RD29A* gene in response to dehydration and high-salinity stresses. Plant J. **34,** 137-148.

Palancade, B., and Bensaude, O. (2003). Investigating RNA polymerase ii carboxyl-terminal domain (ctd) phosphorylation. Eur. J. Biochem. **270,** 3859-3870.

Pan, H., Wang, Y., Zhang, Y., Zhou, T., Fang, C., Nan, P., Wang, X., Li, X., Wei, Y., and Chen, J. (2008). Phenylalanine ammonia lyase functions as a switch directly controlling the accumulation of calycosin and calycosin-7-*O*-β-D-glucoside in *Astragalus membranaceus* var. *mongholicus* plants. J. Exp. Bot. **59,** 3027-3037.

Papp, I., Mur, L.A., Dalmadi, A., Dulai, S., and Koncz, C. (2004). A mutation in the cap binding protein 20 gene confers drought tolerance to *Arabidopsis*. Plant Mol. Biol. **55,** 679-686.

Park, S.Y., Fung, P., Nishimura, N., Jensen, D.R., Fujii, H., Zhao, Y., Lumba, S., Santiago, J., Rodrigues, A., Chow, T.F., Alfred, S.E., Bonetta, D., Finkelstein, R., Provart, N.J., Desveaux, D., Rodriguez, P.L., McCourt, P., Zhu, J.K., Schroeder, J.I., Volkman, B.F., and Cutler, S.R. (2009). Abscisic acid inhibits type 2C protein phosphatases via the PYR/PYL family of START proteins. Science **324,** 1068-1071.

Pedmale, U.V., and Liscum, E. (2007). Regulation of phototropic signaling in *Arabidopsis* via phosphorylation state changes in the phototropin 1-interacting protein NPH3. J. Biol. Chem. **282,** 19992-20001

Peel, G.J., Pang, Y.Z., Modolo, L.V., and Dixon, R.A. (2009). The LAP1 MYB transcription factor orchestrates anthocyanidin biosynthesis and glycosylation in *Medicago*. Plant J. **59,** 136-149.

Peer, W.A., Brown, D.E., Tague, B.W., Muday, G.K., Taiz, L., and Murphy, A.S. (2001). Flavonoid accumulation patterns of transparent testa mutants of *Arabidopsis*. Plant Physiol. **126,** 536-548.

Peer, W.A., Bandyopadhyay, A., Blakeslee, J.J., Makam, S.I., Chen, R.J., Masson, P.H., and Murphy, A.S. (2004). Variation in expression and protein localization of the PIN family of auxin efflux facilitator proteins in flavonoid mutants with altered auxin transport in *Arabidopsis thaliana*. Plant Cell **16,** 1898-1911.

Phatnani, H.P., and Greenleaf, A.L. (2006). Phosphorylation and functions of the RNA polymerase II CTD. Genes Dev. **20,** 2922-2936.

Piao, H.L., Lim, J.H., Kim, S.J., Cheong, G.W., and Hwang, I. (2001). Constitutive over-expression of *AtGSK1* induces NaCl stress responses in the absence of NaCl stress and results in enhanced NaCl tolerance in *Arabidopsis*. Plant J. **27,** 305-314.

Quettier, A.L., Bertrand, C., Habricot, Y., Miginiac, E., Agnes, C., Jeannette, E., and Maldiney, R. (2006). The *phs1-3* mutation in a putative dual-specificity protein tyrosine phosphatase gene provokes hypersensitive responses to abscisic acid in *Arabidopsis thaliana*. Plant J. **47,** 711-719.

Rayapureddi, J.P., Kattamuri, C., Chan, F.H., and Hegde, R.S. (2005). Characterization of a plant, tyrosine-specific phosphatase of the aspartyl class. Biochemistry **44,** 751-758.

Rosso, M.G., Li, Y., Strizhov, N., Reiss, B., Dekker, K., and Weisshaar, B. (2003). An *Arabidopsis thaliana* T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics. Plant Mol. Biol. **53,** 247-259.

Rowan, D.D., Cao, M., Lin-Wang, K., Cooney, J.M., Jensen, D.J., Austin, P.T., Hunt, M.B., Norling, C., Hellens, R.P., Schaffer, R.J., and Allan, A.C. (2009). Environmental regulation of leaf colour in red 35S:PAP1 *Arabidopsis thaliana*. New Phytol. **182,** 102-115.

Sanchez-Calderon, L., Lopez-Bucio, J., Chacon-Lopez, A., Gutierrez-Ortega, A., Hernandez-Abreu, E., and Herrera-Estrella, L. (2006). Characterization of low phosphorus insensitive mutants reveals a crosstalk between low phosphorus-induced determinate root development and the activation of genes involved in the adaptation of *Arabidopsis* to phosphorus deficiency. Plant Physiol. **140,** 879-889.

Sapkota, G., Knockaert, M., Alarcon, C., Montalvo, E., Brivanlou, A.H., and Massague, J. (2006). Dephosphorylation of the linker regions of Smad1 and Smad2/3 by small C-terminal domain phosphatases has distinct outcomes for bone morphogenetic protein and transforming growth factor-β pathways. J. Biol. Chem. **281,** 40412-40419.

Scholl, R.L., May, S.T., and Ware, D.H. (2000). Seed and molecular resources for *Arabidopsis*. Plant Physiol. **124,** 1477-1480.

Schroeder, S.C., Schwer, B., Shuman, S., and Bentley, D. (2000). Dynamic association of capping enzymes with transcribing RNA Polymerase II. Genes Dev. **14,** 2435-2440.

Schweighofer, A., Hirt, H., and Meskiene, I. (2004). Plant PP2C phosphatases: emerging functions in stress signaling. Trends Plant Sci. **9,** 236-243.

Shadle, G.L., Wesley, S.V., Korth, K.L., Chen, F., Lamb, C., and Dixon, R.A. (2003). Phenylpropanoid compounds and disease resistance in transgenic tobacco with altered expression of L-phenylalanine ammonia-lyase. Phytochemistry **64,** 153-161.

Shi, M.Z., and Xie, D.Y. (2010). Features of anthocyanin biosynthesis in *pap1-D* and wild-type *Arabidopsis thaliana* plants grown in different light intensity and culture media conditions. Planta **231,** 1385-1400.

Solecka, D., Boudet, A.M., and Kacperska, A. (1999). Phenylpropanoid and anthocyanin changes in low-temperature treated winter oilseed rape leaves. Plant Physiol. Biochem. **37,** 491-496.

Steyn, W.J., Wand, S.J., Jacobs, G., Rosecrance, R.C., and Roberts, S.C. (2009). Evidence for a photoprotective function of low-temperature-induced anthocyanin accumulation in apple and pear peel. Physiol. Plant **136,** 461-472.

Stracke, R., Werber, M., and Weisshaar, B. (2001). The R2R3-MYB gene family in *Arabidopsis thaliana*. Curr. Opin. Plant Biol. **4,** 447-456.

Suh, M.H., Ye, P., Zhang, M., Hausmann, S., Shuman, S., Gnatt, A.L., and Fu, J. (2005). Fcp1 directly recognizes the C-terminal domain (CTD) and interacts with

a site on RNA polymerase II distinct from the CTD. Proc. Natl. Acad. Sci. USA **102,** 17314-17319.

Sweetlove, L.J., and Fernie, A.R. (2005). Regulation of metabolic networks: understanding metabolic complexity in the systems biology era. New Phytol. **168,** 9-23.

Thaller, M.C., Schippa, S., and Rossolini, G.M. (1998). Conserved sequence motifs among bacterial, eukaryotic, and archaeal phosphatases that define a new phosphohydrolase superfamily. Protein Sci. **7,** 1651-1656.

Thompson, J., Lepikhova, T., Teixido-Travesa, N., Whitehead, M.A., Palvimo, J.J., and Janne, O.A. (2006). Small carboxyl-terminal domain phosphatase 2 attenuates androgen-dependent transcription. EMBO J. **25,** 2757-2767.

Toda, K., Takahashi, R., Iwashina, T., and Hajika, M. (2010). Difference in chilling-induced flavonoid profiles, antioxidant activity and chilling tolerance between soybean near-isogenic lines for the pubescence color gene. J. Plant Res. (In press)

Tohge, T., Nishiyama, Y., Hirai, M.Y., Yano, M., Nakajima, J., Awazuhara, M., Inoue, E., Takahashi, H., Goodenowe, D.B., Kitayama, M., Noji, M., Yamazaki, M., and Saito, K. (2005). Functional genomics by integrated analysis of metabolome and transcriptome of *Arabidopsis* plants over-expressing an MYB transcription factor. Plant J. **42,** 218-235.

Ueda, A., Li, P., Feng, Y., Vikram, M., Kim, S., Kang, C.H., Kang, J.S., Bahk, J.D., Lee, S.Y., Fukuhara, T., Staswick, P.E., Pepper, A.E., and Koiwa, H. (2008).

The *Arabidopsis thaliana* carboxyl-terminal domain phosphatase-like 2 regulates plant growth, stress and auxin responses. Plant Mol. Biol. **67,** 683-697.

Uno, Y., Furihara, T., Abe, H., Yoshida, R., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2000). *Arabidopsis* basic leucine zipper transcription factors involved in an abscisic acid-dependent signal transduction pathway under drought and high-salinity conditions. Proc. Natl. Acad. Sci. USA **97,** 11632-11637.

Usadel, B., Blasing, O.E., Gibon, Y., Poree, F., Hohne, M., Gunter, M., Trethewey, R., Kamlage, B., Poorter, H., and Stitt, M. (2008). Multilevel genomic analysis of the response of transcripts, enzyme activities and metabolites in *Arabidopsis* rosettes to a progressive decrease of temperature in the non-freezing range. Plant Cell Environ. **31,** 518-547.

Varon, R., Gooding, R., Steglich, C., Marns, L., Tang, H., Angelicheva, D., Yong, K.K., Ambrugger, P., Reinhold, A., Morar, B., Baas, F., Kwa, M., Tournev, I., Guerguelcheva, V., Kremensky, I., Lochmuller, H., Mullner-Eidenbock, A., Merlini, L., Neumann, L., Burger, J., Walter, M., Swoboda, K., Thomas, P.K., von Moers, A., Risch, N., and Kalaydjieva, L. (2003). Partial deficiency of the C-terminal-domain phosphatase of RNA polymerase II is associated with congenital cataracts facial dysmorphism neuropathy syndrome. Nat. Genet. **35,** 185-189.

Velten, J., Cakir, C., and Cazzonelli, C.I. (2010). A spontaneous dominant-negative mutation within a 35S::*AtMYB90* transgene inhibits flower pigment production in tobacco. PLoS One **5,** e9917

Verpoorte, R., and Memelink, J. (2002). Engineering secondary metabolite production in plants. Curr. Opin. Biotech. **13,** 181-187.

Visvanathan, J., Lee, S., Lee, B., Lee, J.W., and Lee, S.K. (2007). The microRNA miR-124 antagonizes the anti-neural REST/SCP1 pathway during embryonic CNS development. Genes Dev. **21,** 744-749.

Vlad, F., Rubio, S., Rodrigues, A., Sirichandra, C., Belin, C., Robert, N., Leung, J., Rodriguez, P.L., Lauriere, C., and Merlot, S. (2009). Protein phosphatases 2C regulate the activation of the Snf1-related kinase OST1 by abscisic acid in *Arabidopsis*. Plant Cell **21,** 3170-3184.

Vom Endt, D., Kijne, J.W., and Memelink, J. (2002). Transcription factors controlling plant secondary metabolism: what regulates the regulators? Phytochemistry **61,** 107-114.

Wahid, A., and Ghazanfar, A. (2006). Possible involvement of some secondary metabolites in salt tolerance of sugarcane. J. Plant Physiol. **163,** 723-730.

Wang, W., Cho, H.S., Kim, R., Jancarik, J., Yokota, H., Nguyen, H.H., Grigoriev, I.V., Wemmer, D.E., and Kim, S.H. (2002). Structural characterization of the reaction pathway in phosphoserine phosphatase: crystallographic "snapshots" of intermediate states. J. Mol. Biol. **319,** 421-431.

Wrighton, K.H., Willis, D., Long, J., Liu, F., Lin, X., and Feng, X.H. (2006). Small C-terminal domain phosphatases dephosphorylate the regulatory linker regions of Smad2 and Smad3 to enhance transforming growth factor-β signaling. J. Biol. Chem. **281,** 38365-38375.

Xie, D.Y., Sharma, S.B., Wright, E., Wang, Z.Y., and Dixon, R.A. (2006). Metabolic engineering of proanthocyanidins through co-expression of anthocyanidin reductase and the PAP1 MYB transcription factor. Plant J. **45,** 895-907.

Xiong, L., Gong, Z., Rock, C.D., Subramanian, S., Guo, Y., Xu, W., Galbraith, D., and Zhu, J.-K. (2001). Modulation of abscisic acid signal transduction and biosynthesis by an Sm-like protein in *Arabidopsis*. Dev. Cell **1,** 771-781.

Xiong, L.M., Lee, H., Ishitani, M., Tanaka, Y., Stevenson, B., Koiwa, H., Bressan, R.A., Hasegawa, P.M., and Zhu, J.K. (2002). Repression of stress-responsive genes by FIERY2, a novel transcriptional regulator in *Arabidopsis*. Proc. Natl. Acad. Sci. USA **99,** 10899-10904.

Yamada, K., Lim, J., Dale, J.M., Chen, H., Shinn, P., Palm, C.J., Southwick, A.M., Wu, H.C., Kim, C., Nguyen, M., Pham, P., Cheuk, R., Karlin-Newmann, G., Liu, S.X., Lam, B., Sakano, H., Wu, T., Yu, G., Miranda, M., Quach, H.L., Tripp, M., Chang, C.H., Lee, J.M., Toriumi, M., Chan, M.M., Tang, C.C., Onodera, C.S., Deng, J.M., Akiyama, K., Ansari, Y., Arakawa, T., Banh, J., Banno, F., Bowser, L., Brooks, S., Carninci, P., Chao, Q., Choy, N., Enju, A., Goldsmith, A.D., Gurjal, M., Hansen, N.F., Hayashizaki, Y., Johnson-Hopson, C., Hsuan, V.W., Iida, K., Karnes, M., Khan, S., Koesema, E.,

Ishida, J., Jiang, P.X., Jones, T., Kawai, J., Kamiya, A., Meyers, C., Nakajima, M., Narusaka, M., Seki, M., Sakurai, T., Satou, M., Tamse, R., Vaysberg, M., Wallender, E.K., Wong, C., Yamamura, Y., Yuan, S., Shinozaki, K., Davis, R.W., Theologis, A., and Ecker, J.R. (2003). Empirical analysis of transcriptional activity in the *Arabidopsis* genome. Science **302,** 842-846.

Yeo, M., and Lin, P.S. (2007). Functional characterization of small CTD phosphatases. Methods Mol. Biol. **365,** 335-346.

Yeo, M., Lin, P.S., Dahmus, M.E., and Gill, G.N. (2003). A novel RNA polymerase II C-terminal domain phosphatase that preferentially dephosphorylates serine 5. J. Biol. Chem. **278,** 26078-26085.

Yeo, M., Lee, S.K., Lee, B., Ruiz, E.C., Pfaff, S.L., and Gill, G.N. (2005). Small CTD phosphatases function in silencing neuronal gene expression. Science **307,** 596-600.

Zhang, M., Liu, J., Kim, Y., Dixon, J.E., Pfaff, S.L., Gill, G.N., Noel, J.P., and Zhang, Y. (2010). Structural and functional analysis of the phosphoryl transfer reaction mediated by the human small C-terminal domain phosphatase, Scp1. Protein Sci. **19,** 974-86

Zhang, Y., Kim, Y., Genoud, N., Gao, J., Kelly, J.W., Pfaff, S.L., Gill, G.N., Dixon, J.E., and Noel, J.P. (2006). Determinants for dephosphorylation of the RNA polymerase II C-terminal domain by Scp1. Mol. Cell **24,** 759-770.

Zhou, L.L., Zeng, H.N., Shi, M.Z., and Xie, D.Y. (2008). Development of tobacco callus cultures overexpressing *Arabidopsis* PAP1/MYB75 transcription factor and characterization of anthocyanin biosynthesis. Planta **229,** 37-51.

Zhu, J.H., Dong, C.H., and Zhu, J.K. (2007). Interplay between cold-responsive gene regulation, metabolism and RNA processing during plant cold acclimation. Curr. Opin. Plant Biol. **10,** 290-295.

Zohn, I.E., and Brivanlou, A.H. (2001). Expression cloning of Xenopus Os4, an evolutionarily conserved gene, which induces mesoderm and dorsal axis. Dev. Biol. **239,** 118-131.

Zuo, J., Hare, P.D., and Chua, N.H. (2006). Applications of chemical-inducible expression systems in functional genomics and biotechnology. Methods Mol. Biol. **323,** 329-342.

VITA

Name: Yue Feng

Address: MS2133, Dept of Hort. Sci., Texas A&M University
College Station, TX, 77843-2133, USA

Email Address: acaugirl@hotmail.com

Education: B.S., Biological Sciences, China Agricultural University, 2005
Ph.D., Molecular and Environmental Plant Sciences , Texas A&M University, 2010

Publications:

Feng Y., Kang J.S., Kim S., Yun D.J., Lee S.Y., Bahk J.D., Koiwa H. (2010). *Arabidopsis* SCP1-like small phosphatases differentially dephosphorylate RNA polymerase II C-terminal domain. Biochem. Biophys. Res. Commun. **397,** 355-60

Kang, C.H., Feng, Y., Vikram, M., Jeong, I.S., Lee, J.R., Bahk, J.D., Yun, D.J., Lee, S.Y., and Koiwa, H. (2009). *Arabidopsis thaliana* PRP40s are RNA polymerase II C-terminal domain-associating proteins. Arch. Biochem. Biophys. **484,** 30-38

Ueda, A., Li, P., Feng, Y., Vikram, M., Kim, S., Kang, C.H., Kang, J.S., Bahk, J.D., Lee, S.Y., Fukuhara, T., Staswick, P.E., Pepper, A.E., and Koiwa, H. (2008). The *Arabidopsis thaliana* carboxyl-terminal domain phosphatase-like 2 regulates plant growth, stress and auxin responses. Plant Mol. Biol. **67,** 683-97.