University of Alberta

Functional characterization of a novel cell-wall annotated *PELPK1* gene in *Arabidopsis thaliana*

By

Abdur Rashid

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Plant Biology

Department of Biological Sciences

© Abdur Rashid
Spring 2011
Edmonton, Alberta

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only. Where the thesis is converted to, or otherwise made available in digital form, the University of Alberta will advise potential users of the thesis of these terms.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatsoever without the author's prior written permission.

Dedicated to my family:

My wife, Tahmina; my daughter, Lanita; and my son, Tahsin

Abstract

In silico analysis showed that *Arabidopsis thaliana* gene *AT5G09530* encodes a uniquely repetitive, proline-enriched protein that is conserved across species, and is likely secreted to the cell wall. Based on its most common amino acid repeat motif, I named the gene *PELPK1* and its putative paralog *PELPK2 (AT5G09520)*.

Reporter (GUS) expression showed that the *PELPK1* upstream genomic region is sufficient for expression in the aleurone layer during seed germination, and is induced throughout the plant by biotic factors (especially *Pseudomonas syringae* infection), defense chemicals (MeJa, salicylic acid), and mechanical wounding, consistent with the presence of conserved regulatory elements. Sub-cellular localization of a translational fusion of PELPK1 with GFP showed that the protein was secreted into seed-coat aleurone cells and to the cell walls of other tissues. Based on these results, it was concluded that the PELPK1 is a cell wall-associated protein and is most actively transcribed during radicle penetration of the seed coat and during pathogen and wounding responses. A proteomic survey of aleurone proteins failed to identify PELPK1, although several proteins not previously associated with this tissue were identified.

Mutational analysis demonstrated that RNAi silencing significantly down-regulated the transcript abundance of *PELPK1*. Phenotypic analysis showed that RNAi plants exhibited significantly slower germination and root growth when the medium was supplemented with sucrose (100mM). Conversely, constitutive overexpression (OX) of *PELPK1* enhanced seed germination and root elongation as compared to wild-type (WT). Analysis of soil-grown plants showed slower emergence and slower vegetative growth for RNAi lines, while OX plants exhibited faster emergence and enhanced vegetative growth and flowering as compared to WT. However, *PELPK1* RNAi and OX lines did not differ from WT in response to treatment with pathogens. These results show that the abundance of PELPK1 is positively correlated with plant growth rate under

some conditions. *PELPK1* may influence growth through CW modification or other independent pathways.

Acknowledgements

I am indebted to my supervisor, Dr. Michael Deyholos for excellent supervision, instruction, technical and financial supports, extremely effective comments on different chapters and very fast and excellent editing of my thesis. I would like to thank to my committee members, Dr. Nat Kav and Dr. George Owttrim for their guidance throughout the course of my graduate research. I would like to thank Dr. Nat Kav for allowing me to use his laboratory facilities and to Shiv Ganesh and Dr. M.H. Rahman for helping in proteomic work. My very special thank goes to Dr. Ajay Badhan of our lab for encouragement, emotional support, and for helping in immunolabeling and proteomic experiments. Dr. A. Hameed Khadhair, my former colleague and friend deserves appreciations for some microtome work. My cordial thanks to Dr. Hafiz Ahmed and Dr. Sheau-Fang Hwang of CDC North, and Dr. Khadhair for helping in pathogen identification in some of my plants. I would like to thank Mr. Rakesh Bhatnagar and Mrs Arlene Oatway for advanced microscopy support, and Dr. Anthony Cornish and Mr. Troy Locke for LCMS/MS and qRT-PCR supports. Thanks a lot to Tan Bao for some statistical help. Finally and most importantly, I would like to thank to all past and present members of Dr. Deyholos lab (Mary De Pauw, Disa Brownfield, Yuanqing Jiang, Mohsen Mohammadi, Melissa Roach, Manjeet Kumari, Neil Hobson, Ryan Mckenzie, Gord McNickle, Shanjida Khan, Leonardo Galindo, David Pinzon, and Dr. Joshua McDill) for their wonderful cooperation and support during my more than 6 years of stay in this lab.

The plant pathogens, *Pseudomonas syringae*, and *Pythium irregulare* were obtained from Dr. Stephen Strelkov.

Table of Contents

	Page
Chapter 1: Review of Literature	1
1. Plant cell wall	2
Cell wall compositions	2
Cell wall structural proteins (CWPs)	3
A. Hydroxyproline-Rich Glycoproteins (HRGPs)	4
General characteristics	4
Interaction with the cell wall	8
Developmental regulation and inducibility	12
B. Proline-Rich proteins (PRPs)	20
General characteristics	20
Interaction with the cell wall	21
Developmental regulation and inducibility	22
C. Glycine-Rich proteins (GRPs)	23
General characteristics	23
Interaction with the cell wall	24
Developmental regulation and inducibility	25
D. Arabinogalactan proteins (AGPs)	27
General characteristics	27
Interaction with the cell wall	27
Developmental regulation and inducibility	28
2. Application of reverse genetics methodologies in the functional characterization of novel *Arabidopsis* genes	29
3. Prospects of developing disease/insect resistant crops through genetic manipulation of CWPs	35

4. Conclusions and objectives 36

5. References 38

Chapter 2: *In silico* analysis of *PELPK1* (*AT5G09530*) gene 50

1. Introduction 51

2. Analysis of *PELPK1* bioinformatics databases 51

Computational analysis of primary amino acid sequences 51

Predictions of PELPK1 structure 56

Predictions of glycosylation 59

Identification of sequences related to *PELPK1* 61

Is *PELPK1* an HRGP, extensin, or periaxin? 68

Analysis of expression profiles of *PELPK1* 73

3. Conclusions 76

4. References 78

Chapter 3: Expression analysis of the *PELPK1* 81

1. Introduction 82

2. Materials and Methods 85

Extraction of Genomic DNA 85

Plasmid construction 85

Promoter::GUS fusion construct 85

CDS::GFP translational fusion construct 86

Agrobacterium transformation 87

Arabidopsis transformation 88

Selection of transgenic lines 88

Histochemical GUS staining 89

Stress treatments 90

Abiotic stress treatments 90

Biotic stress treatments 90

Hormonal treatments 91

Mechanical wounding 91

Microtomy 91

Microscopy 92

` GFP localization 92

Confocal microscopy 92

3. Results 94

Normal *PELPK1* promoter activity 94

PELPK1 promoter inducibility 99

Abiotic factors 99

Biotic factors 100

Defense hormones 100

Mechanical wounding 101

Computational analysis of the promoter sequence 101

Sub-cellular protein localization 108

4. Discussion 112

GUS reporter expression 112

GFP localization 114

5. Conclusions 116

6. References 117

Chapter 4: Mutational analysis of the *PELPK1* 122

1. Introduction 123

2. Materials and Methods 125

Plant growth conditions 125

Conditions in the MS-agar medium 125

Conditions in the soil 125

Abiotic stress treatments 125

Responses to hormones 126

Response to pathogen 126

Phenotypic and growth analyses 126

Analysis in the MS-agar medium 126

Analysis in the soil 127

Extraction of RNA 127

Synthesis of cDNA 127

Quantitative Real Time PCR (qRT-PCR) 128

Identification of homozygous T-DNA insertion lines 131

Plasmid construction 135

Construction of RNAi plasmids 135

Construction of over-expression plasmids 139

Arabidopsis transformation and selection 140

3. Results 144

Analysis of sequence-indexed insertion mutant lines 144

Analysis of RNA interference mutant lines 146

Analysis of over-expression mutant lines 154

4. Discussion 159

Seed germination and root elongation 160

Shoot growth and flowering 162

5. Conclusions 164

6. References 165

Chapter 5: Proteomic analysis of *Arabidopsis* seed-coat/whole seed 168

1. Introduction 169

2. Materials and Methods 172

Separation of seed coats 172

Extraction of total proteins 172

1-dimensional gel electrophoresis 173

2-dimensional gel electrophoresis 174

Liquid chromatography/Tandem mass spectrometry 175

3. Results and Discussion 177

Analysis of seed coat proteomic profile 177

Analysis of whole seed proteomic profile 185

4. Conclusions 187

5. References 188

Chapter 6: General discussion 190

1. General discussion 191

2. Future research 203

Site-directed mutagenesis of *PELPK1* 203

Heterologous expression of *PELPK1* 207

3. References 209

List of Tables

	Page
Table 1-1: Tissue localization of cell wall structural proteins (CWPs)	14
Table 1-2: Conditions regulating the expression of CWPs	15
Table 1-3: Distribution of T-DNA insertions in genes and intergenic regions	30
Table 2-1: AA profiles of the PELPK1 and other genes	54
Table 2-2: Comparative bioinformatics PELPK1 and others	55
Table 2-3: A summary of microarray-based expression profile of *PELPK1*	75
Table 3-1: Putative cis-regulatory elements of *PELPK1* promoter	107
Table 4-1: Primers used for transcript analysis by qRT-PCR	130
Table 4-2: Primers used for amplification of *PELPK1* and *PELPK2*	143
Table 4-3: Primers used for amplification of genes from cloned vectors	143
Table 5-1: Identification of seed coat protein spots from 2D gel by LC-MS/MS	179
Table 5-2: Identification of proteins from whole seed proteome	182

List of Figures

	Page
Figure 1-1: A scale model showing a portion of a primary cell wall	5
Figure 1-2: A model showing cross-linking of CW components	5
Figure 1-3: Primary structure of an extensin (AT1G76930)	7
Figure 1-4: Glycosylation of an extensin and an AGP	7
Figure 1-5: A model showing cross-linking of extensins with the cell wall	10
Figure 1-6: T-DNA insertion mutagenesis in *Arabidopsis* chromosomes	32
Figure 1-7: Mechanism of RNA interference	33
Figure 2-1: Amino acid sequence of PELPK1 (AT5G09530)	53
Figure 2 -2: PELPK1 sequence logo	53
Figure 2-3. Protein structure of PELPK1	57
Figure 2-4: Predicted glycosylation of PELPK1	60
Figure 2-5: A pair-wise alignment of the PELPK1 with PELPK2	62
Figure 2-6: A bootstrap consensus tree	64
Figure 2-7: Sequences from higher plants with similar motifs	65
Figure 2-8: Local protein disorder among all PRPs	70
Figure 2-9: Local protein disorder among extensins	71
Figure 2-10: Llocal protein disorder among PELPK1-like proteins	72
Figure 2-11: A growth-stage specific expression profile of PELPK1	76
Figure 3-1: *PELPK1*-promoter::GUS-reporter fusion construct	95
Figure 3-2: Normal *PELPK1*-promoter::GUS expression pattern	96
Figure 3-3: Normal promoter::GUS expression in seed coat aleurone layer	97
Figure 3-4: Aleurone cells intermingled with GUS crystals and granules	98
Figure 3-5: Disease infection symptoms in promoter:: GUS plants	103
Figure 3-6: GUS expression induced by pathogen infection	104

Figure 3-7: GUS expression induced by defense chemicals 105

Figure 3-8: GUS expression induced by mechanical wounding 106

Figure 3-9: Schematic diagram of CDS::GFP fusion 109

Figure 3-10: GFP emission in the aleurone layer 110

Figure 3-11: GFP emission in the cell walls 111

Figure 4-1: Agarose gel electrophoresis of RT-PCR products 129

Figure 4-2: Diagram of the *PELPK1* with T-DNA insertion sites 133

Figure 4-3: Homozygous T-DNA insertion by PCR 134

Figure 4-4: Gateway recombination cloning of RNAi fragment 138

Figure 4-5: Schematic diagram of an OX construct 141

Figure 4-6: Confirmation of RNAi and OX Insert constructs 142

Figure 4-7: Seedling morphology of a SAIL line and WT 145

Figure 4-8: qRT-PCR in T-DNA insertion lines 146

Figure 4-9: qRT-PCR in RNAi and OX-lines 150

Figure 4-10: Percentage of seed germination 152

Figure 4-11: Response of RNAi roots to sucrose 153

Figure 4-12: Quantitative root elongation assay of mutant and WT plants 156

Figure 4-13: Phenotype of mutant and WT plants in the soil 157

Figure 4-14: Height of inflorescence shoots of mutant and WT plants 158

Figure 4-15: Percentage of flowering of mutant and WT plants 158

Figure 5-1: Microtome sectioning through mature WT seed 171

Figure 5-2: 2D gel electrophoresis of the seed coat proteins 178

List of Abbreviations

AA	Amino acid
ABA	abscisic acid
ABRC	arabidopsis biological resource center
ABRE	ABA responsive element
ACN	acetonitrile
AGI	Arabidopsis genome initiative
AGPs	arabinogalactan proteins
ANOVA	analysis of variance
BLAST	Basic Local Alignment Search Tool
CAE	cis-acting element
CaMV35S	Cauliflower Mosaic Virus
CBS	colloidal blue staining
cDNA	complementary DNA
CDS	coding sequence
CHAPS	{3-(3-Cholamidopropyl)-dimethylammonio}-1-propanesulfonate
CRE	cis-regulatory element
CWPs	cell wall structural proteins
DAS	days after sowing
dC_t	delta threshold cycle
2D gel	2 dimensional gel electrophoresis
dsRNA	double-stranded RNA
DTT	dithiothreitol
ER	endoplasmic reticulum
GFP	Green Fluorescent Protein
GO	Gene Ontology
GOI	gene of interest
GP	Gateway primer
GRPs	glycine rich proteins
GUS	ß-glucuronidase
HPLC	high performance liquid chromatography
HPT	hygromycin phosphotransferase
HRGPs	hydroxyproline-rich glycoproteins
hyg	hygromycin
Hyp	hydroxyproline
IDPs	Intrinsically Disordered Proteins
IEF	isoelectric focusing
IPG	immobilized pH gradient
kan	Kanamycin
Kan^R	Kanamycin resistance
KO	knockout
LB	Left Border
LC-MS/MS	liquid chromatography-mass spectrometry
LP	left primer
LPS	lipopolysaccharides
MALDI-TOF-MS	matrix-assisted laser desorption Ionization with tandem time of flight- mass spectrometry
MeJa	methyl jasmonate
MS media	Murashige & Skoog media

MudPIT	multidimensional reverse-phase chromatography with on-line tandem mass spectrometry
MUT	mutant
MWCO	molecular weight cut-off
NCBI	National Cetner for Biotechnology Information
NPT	Neomycin phosphotransferase
OX	over expression
PAGE	polyacrylamide gel electrophoresis
PGA	polygalacturonic acid
PGN	Plant Genome Network
PBS	phohosphate buffered saline
PVSs	protein storage vacuoles
Q-dot	quantum dot
qRT-PCR	quantitative Real Time PCR
RCB	randomized complete block
RG	reference gene
RISC	RNA induced silencing complex
RP	right primer
RT-PCR	reverse-transcription polymerase chain reaction
SA	salicylic acid
SAIL	Syngenta Arabidopsis insertion library
SDS	Sodium dodecyl sulfate
siRNAs	small interfering RNA
SP	signal peptide
T_1 $T_2$$T_3$	transgenic progeny generation
TAIR	The Arabidopsis Information Resource
TBP	tyributyl phosphate
T-DNA	transferred-DNA
TEM	transmission electron microscopy
TIBA	triisobutylaluminum
TM	ransmembrane
UTR	untranslated region
WT	wild type

CHAPTER 1

REVIEW OF LITERATURE

CHAPTER 1

REVIEW OF LITERATURE

1. PLANT CELL WALL

The cell wall is a dynamic structure that can vary in its flexibility. The flexibility of cell walls is demonstrated for example during wilting, when stems begin to droop. The apparent rigidity of plant tissues is actually due to turgor pressure of the cells rather than rigid cell walls (Roberts, 1989, 1990; Howland, 2000). Cell walls contribute to turgor management, and provide cells with structural as well as mechanical supports (Brett and Warldron, 1996; Chivasa et al., 2002; Burgert, 2006; Roberts 2007). Being located on the outside of the plasma membrane, they also provide protection against abiotic and biotic stresses (Schindler, 1995; Hayashi 2006; Chen and Kim 2009). Thus, many aspects of plant function depend on the composition of cell walls.

Cell wall composition

The major components of cell wall are polysaccharides, phenolics, pectins and proteins, although composition varies between species, cell types, developmental stages and environmental conditions. A generalized cell wall model is depicted in Figure 1-1. Polysaccharides, phenolics, and pectins are the subjects of several comprehensive reviews (e.g. McNeil et al., 1984; Somerville, et al., 2004; Fry, 2004; Sandhua et al., 2009). Both enzymes and non-catalytic proteins are found in the cell wall (reviewed in Jamet et al., 2006, 2008). Examples of cell wall localized enzymes include glycosylhydrolases, proteases, and ascorbic acid oxidases (Bauer et al., 2006). The objective of this chapter was to review cell wall (CW) associated non-enzymatic structural proteins (CWPs) and their roles in plant growth and defense responses with a

view to functionally characterize a novel CW associated gene, *PELPK1* in *Arabidopsis thaliana*.

As I have elaborated in the following sections, CWPs are essential constituents of the plant cell wall. They provide structural support to plants during development and act in many other processes including defense against abiotic and biotic stresses, wound healing, signal transduction, and may interact with plasma membrane proteins (Jamet et al., 2008; Bradley et al., 1992; Trezzini et al., 1993; Showalter, 1993; Boudart et al. 1995; Ahn et al., 1996; Cassab, 1998; Jose-Estanyol and Puigdomenech, 2000; Chivasa et al., 2002; Ellis et al., 2002; Hall and Cannon, 2002; Merkouropoulos and Shirsat, 2003; Vogel, et al., 2004; Shirsat and Guo 2006; Basavaraju et al., 2009; Narvaez-Vasquez et al., 2005; Pearce et al., 2001; Ryan and Pearce, 2004; Matsubayashi, 2003).

CWPs and their roles in plants

Most CWPs have been suggested to be ubiquitous and relatively abundant in land plants and green algae (Cassab 1998). They may contain an N-terminal signal peptide, repetitive sequence motifs, and may be rich in one or more amino acid residues, variously glycosylated, and involved in growth and defense-related functions in plants (Showalter 1993; Cassab 1998; Kieliszewski and Shpak, 2001; Jamet et al., 2006; Albenne et al., 2009). Among the putative CWPs, two of the largest classes of proteins are hydroxyproline-rich glycoproteins (HRGPs) and glycine rich proteins (GRPs). Most authors divide the HRGP superfamily into at least three groups: proline-rich proteins (PRPs), arabinogalactan proteins (AGPs), and extensins including chimeric proteins containing extensin-like domains (Showalter 1993; Cassab 1998; Jose-Estanyol and Puigdomenech, 2000; Ringli et al., 2001; Ryser et al., 2004; Mousavi and Hotta, 2005; Xu et al., 2008). Extensins and PRPs have been suggested to form rods and GRPs to

form beta-pleated sheets. Some of these CWPs appear to cross-link with pectic substances. They may also serve as a scaffold for construction of other wall components. The cell wall structural model presented in Figure 1-2 shows cross-linking of cell wall components with some of the CWPs, particularly with extensins (Carpita, and Gibeaut, 1993; Showalter 1993; Cassab 1998; Jose-Estanyol and Puigdomenech, 2000; Ferris et al., 2001; Cannon 2008).

Below is a review of major CWPs encompassing their general characteristics, interaction with the cell wall components, expression during normal development of plants, and responses to abiotic and biotic stresses, and mechanical injuries in plants. I will first discuss extensins, followed by PRPs, GRPs, and AGPs.

A. Extensins

General characteristics

Among the HRGPs, extensins have been studied most extensively because of their involvement in cell wall extension (Lamport 1965, 1969; Wilson and Fry 1986; Cooper et al., 1987; Showalter 1993; Cassab 1998; Xu et al., 2008). They are present in a wide variety of plants including algae and have been proposed to be the major protein component of the cell walls of higher plants (Lamport 1965; Catt et al., 1979, Cassab 1998; García-HernándezI, and LópezII, 2005). Extensins (and other HRGPs) play a crucial role in early stages of cell wall assembly, and their precursors are seen early in cell wall formation (Ye and Varner, 1991). A large increase in the abundance of these proteins in the cell wall is also observed upon cessation of plant growth (Cassab 1998). Based on the above observations, it was suggested that incorporation of extensins into the cell wall matrix causes the cells to undergo cessation of growth and fix final shape (Carpita and Gibeaut, 1993).

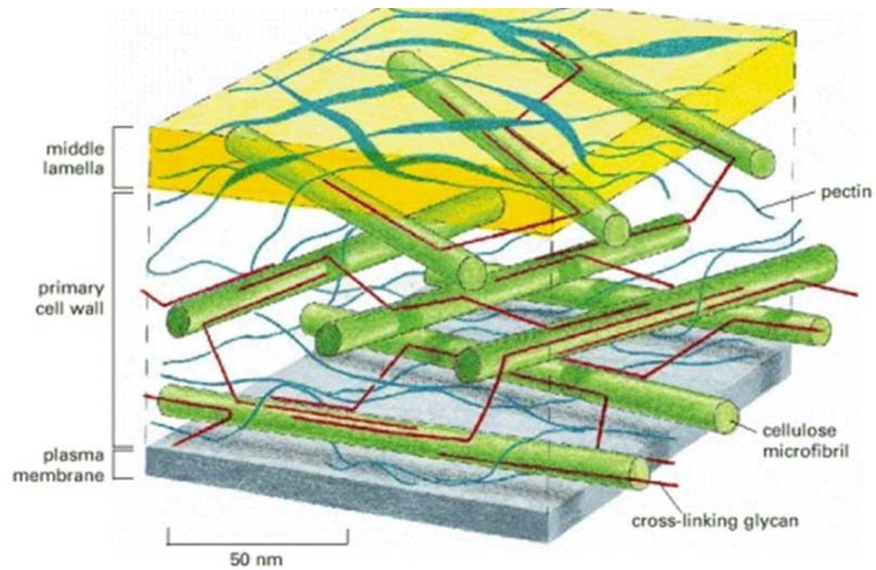


Figure 1-1: A scale model of a portion of a primary cell wall showing pectins and cross-linking glycans (from Roberts, 1994)

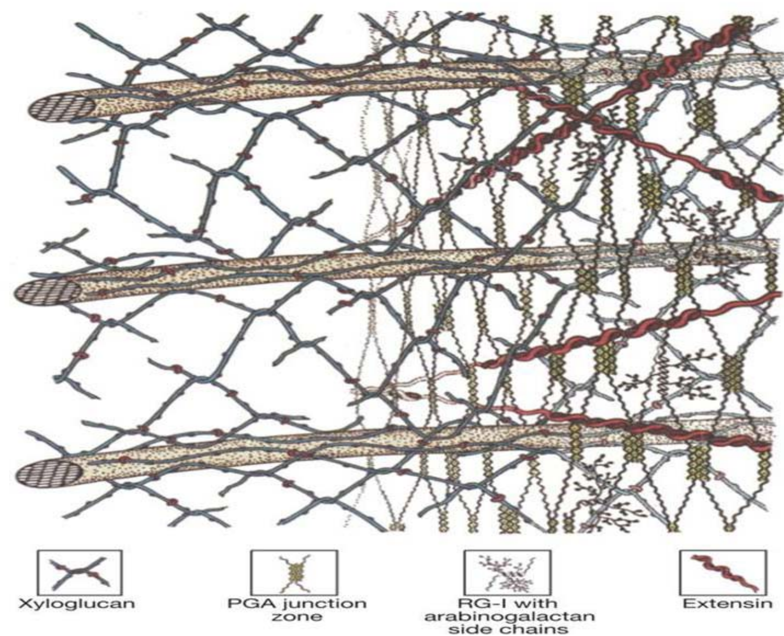


Figure 1-2: A model showing cross-linking of cell wall components including extensins (from Carpita and Gibeaut, 1993)

While it is not known whether extensins are present in all plant cell types, they appear to be more abundant in sclerenchyma cell walls (Cassab and Varner 1987). It has been suggested that formation of an extensin-cellulose framework in sclerenchyma as well as in the vascular cambium cell walls increases the tensile strength of these tissues (Cassab 1998). Since the sclerenchyma cells act as the skeletal elements of the plant and enable the plant body to withstand various strains including stretching, bending, compression, and tension (Esau 1965), the abundance of extensins in sclerenchyma cell walls appears to be consistent with a structural role for these proteins in plant cell walls (Cassab and Varner 1988).

Chemical characterization

Chemically, extensins are characterized by a N-terminal signal peptide, which is responsible for transporting extensins to the cell wall, a transmembrane domain that is completely over-lapped with the signal peptide, a repetitive region rich in Pro/Hyp residues with a main repeating pentapeptide motif of Ser-(Hyp)4, and an extended polyproline II helix (Figure 1-3) (Cassab and Varner, 1988; Showalter and Varner, 1989, Cassab 1998).

Most Pro residues in extensins are hydroxylated by prolyl hydroxylase to give Hyp and then O-glycosylated with one to four arabinosyl residues. Similarly, Ser residues in extensins are often O-glycosylated with a galactose unit (Figure 1-4), Jonson et al., (2003). Extensins have been reported to contain ~35% proteins and ~65% carbohydrate (~97% arabinose, and ~3% galactose; Lamport et al., (1973)). Most of the above features of extensins have been described in dicots, monocots, and also in one gymnosperm (Fong et al., 1992).

In addition to the Ser-(Hyp)4 pentapeptide motif, extensins may also contain other repetitive sequences, which can vary from plant to plant, or from monocots to dicots (Showalter, 1993; Kieliszewski and Lamport, 1994). In the monocot maize, for

*MGAPMASFLVLAFSLAFVSQTT*ANYFYS**SPPPP**V
KHY**SPPP**VYK**SPPPP**VKHY**SPPP**VYK**SPPP**VHYS
PPVYK**SPPPP**VKYY**SPPP**VYK**SPPPP**VYK**SPPPP**
VKHY**SPPP**VYK**SPPPP**VKHY**SPPP**VYK**SPPPP**VK
HY**SPP**VYK**SPPPP**VKHY**SPPP**SYTTLHHHRFTTH
LLQSYTTLHHHRFTTHLLQLYTTPHHHPRSTTNTN
LLLLRTPLLTSTPTLPLQITTSSSLLDYLIVSNGVDD
DTQENKESGDIIKDTRSLTGELKRIHMSKEKVSHF
NVFSLFINTLL

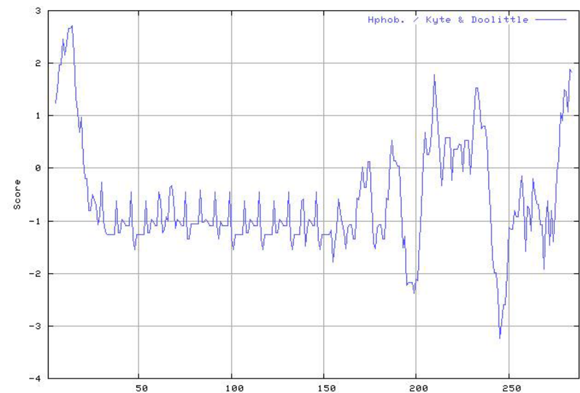
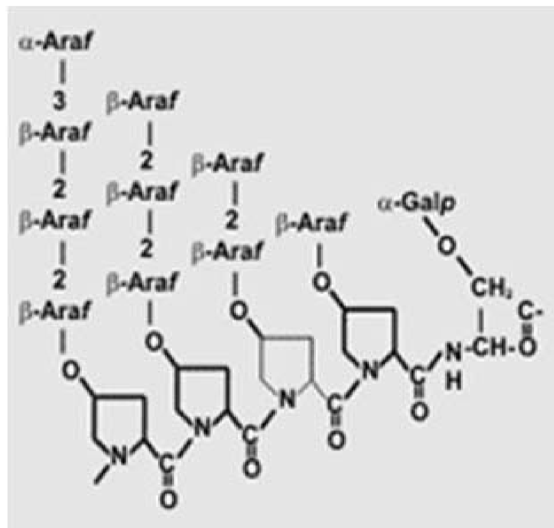


Figure 1-3: The primary structure (left) of an extensin (AT1G76930) showing a N-terminal signal sequence (italicized), a transmembrane region, (italicized and underlined), and protein motifs (bold and underlined); and a hydropathy plot (right)



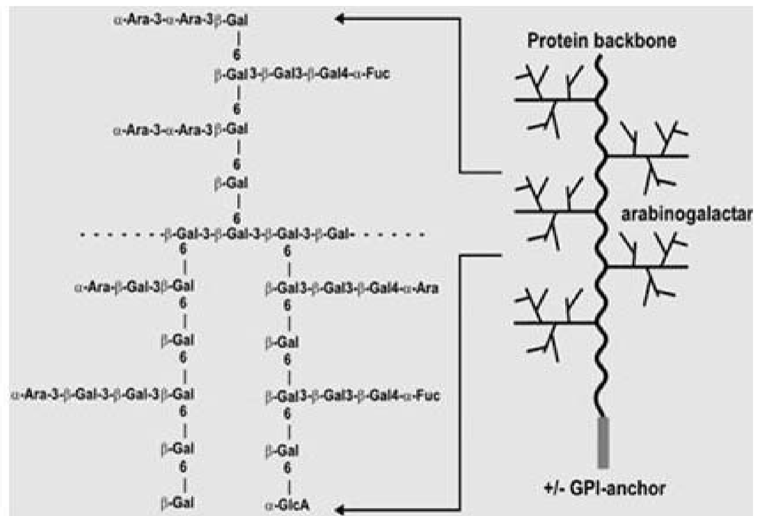


Figure 1-4. Glycosylation of an extensin (left) and an AGP (right). Adapted from Johnson et al., (2003)

example, extensin-like proteins with slightly different motifs such as Ser-Hyp-Lys-Pro-Hyp have been described (Kieliszewski et al., 1990). In addition to the abundance of Pro/Hyp residues, extensins may also be moderately rich in some other amino acid residues such as Ser, Val, Tyr, Lys and His (Showalter, 1993; Cassab 1998; Johnson et al., 2003).

Specialized approaches have been used to determine the structure of the extensin polypeptides. The first approach involved the removal of arabinose residues from extensins followed by trypsinization and sequencing (Lamport, 1980). Using this technique it was shown that removal of the most abundant neutral sugars, arabinose and galactose, from extensins resulted in the loss of the rod-like appearance of these glycoproteins, whereas addition of oligoarabinosides restored the extensin structures (van Holst et al., 1984; Stafstrom et al., 1986). Based on the above observations it was suggested that the arrangement of the arabinosides in the extensin molecules determines how these glycoproteins interact with other polymers of the cell wall (Stafstrom et al., 1986). The second approach involved inferring of the sequences of the proteins from cDNA nucleotide sequences (Showalter, 1993). This approach has limitations since post-translational modifications (i.e. hydroxylation of Pro to Hyp and then glycosylation) cannot be directly inferred from the nucleotide sequence. The third approach involved the isolation of extensin polypeptides before they are covalently incorporated into the wall, followed by deglycosylation, trypsinization, and sequencing (Smith et al., 1986). The second and third approaches have been suggested to be more effective in differentiating between species of extensin molecules in the cell wall, which cannot be done by the first approach (Qi et al., 1995),

Interaction of extensins with the cell wall

Although extensins have been proposed to be cell wall-associated structural proteins, the precise nature of their association with the cell wall components is unclear. To understand the nature of this interaction, researchers have used various methods to modify extensin content in the cell wall. For instance, it was shown that the formation of protein-bound Hyp in oat coleoptiles was inhibited when auxin was present in the medium; however when free Hyp was added to the medium, auxin-induced cell elongation was inhibited (Cleland, 1967). Another study showed that when etiolated pea seedlings were exposed to ethylene, the formation of protein-bound Hyp was increased (Ridge and Osborne, 1970). Based on the above observations, it was suggested that the inhibition of auxin-induced cell elongation due to the application of free Hyp, or the increase in the level of Hyp in the walls due to ethylene treatment might be related to the increase in the extensin content in the cell wall (Cleland 1967; Ridge and Osborne, 1970).

Formation of isodityrosine bridges

Based on several studies it has been proposed that the abundant Tyr residues in extensin polypeptides might be involved in creating isodityrosine cross-links with other extensins by covalent bonds (Fry 1986, Waffenschmidt et al., 1993; Held et al., 2004). The above suggestions appear to be consistent with the report indicating the isolation of a novel amino acid, di-isodityrosine, from hydrosylates of cell walls of tomato cell culture (Bradley et al., 1996). Several studies suggested that di-isodityrosine could also form interpolypeptide linkages between cell wall proteins (Bradley et al., 1996; Cassab 1998; Otte and Barz, 2000; Khashimova, 2003). Based on the above observations and suggestions, a model has been proposed showing that when evenly spaced Tyr residues of one extensin molecule bond with the Tyr residues of another extensin molecule, they can wrap around other cell wall constituents and knit the wall together (Figure 1-5).

Non-covalent interactions

Some studies suggested that because of high pI of extensins (pI = ~10 due to the high Lys content) they may interact non-covalently with the poly-anionic region of pectin in the cell wall (Cassab and Varner 1988; Cassab, 1998; MacDougall et al., 2001; Rose, 2003; Cannon et al., 2008). However, the following recent report does not appear to be consistent with the preceding hypothesis. Nuñez, et al., (2009) showed that when

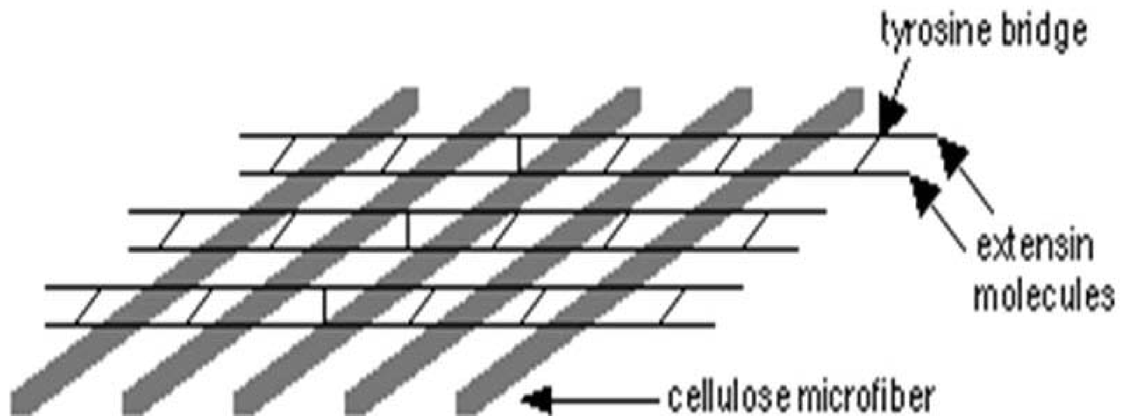


Figure 1-5: A model showing cross-linking of extensins in the cell wall using their Tyr residues. Tyr residues in extensins are more or less evenly spaced. When these residues on one extensin molecule bond with that on another extensin molecule, they can wrap around other cell wall components and knit the wall together (http://www.bio.indiana.edu/~hangarterlab/courses/b373/lecturenotes/cellwall/cellwall.html)

pectins extracted from sugar beet (*Beta vulgaris*) cell walls were subjected to MALDI-TOF-MS (matrix-assisted laser desorption ionization with tandem time of flight- mass spectrometry) following protease digestion, they produced peptide sequences that were highly consistent with the extensin protein of the sugar beet. However, when attempts were made to further disassociate proteins from pectin using 1M NaCl and a 100 kDa MWCO (molecular weight cutoff) membrane, no extensin peptides were detected. Based on the above observations, the authors suggested that pectin and extensin do not form a complex based on ionic interactions.

Covalent interactions

It has been shown that when extensins are incorporated into the cell wall matrix, they form covalent linkages with pectin. Although the exact nature of the linkage between the pectin and the extensin is unknown, the authors suggested that the linkage could be via either a 3,6-linked galactan (Keegstra et al., 1973) or a phenolic cross-link from a feruloylated sugar in the pectin to an amino acid in the extensin (Brownleader and Dey,1993; Qi et al., 1995). The latter suggestion was made based on the report suggesting that ferulic acid can be linked to galactose and arabinose moieties of plant cell-wall pectic polysaccharides (Bolwell, 1993). The cross-linking amino acid residues of extensins appear to be Ser and Hyp for galactose and arabinose, respectively (Qi et al., 1995). These observations appear to be consistent with the previous report suggesting that extensins can form covalent bonds between themselves as well as with other components of the cell wall (Cooper et al., 1983).

Cross-linking and insolubilization

Another line of research reported that in Boron-deficient root nodules, the insolubilization of extensins in the cell wall is significantly reduced (Bonilla et al., 1997). Since Boron has been shown to be linked to the rhamnogalacturonan II (RG II) fraction of pectin (Ishii and Matsunaga, 2001) while extensins were shown to be linked to the RG

I fraction (Qi el al., 1995), it appears that both fractions of pectin, RGII and RG I, are needed for the extensin molecules to become insolubilized in the cell wall. A number of studies have suggested that the insolubilization of extensins in the cell wall is mediated by the release of hydrogen peroxide from the cell wall followed by oxidative cross-linking mediated by extensin peroxidases (Cooper and Varner 1984; Bradley et al., 1992; Cassab 1998; Ribeiro et al. 2006; Bradley et al., 1992). On the other hand, it has been shown that complete deglycosylation of cell walls did not completely solubilize the extensins from the wall (Qi et al., 1995). It was therefore suggested that some form of protein-protein or protein-phenolic-protein cross-link may likely be present between extensin polypeptides in the cell wall (Qi et al., 1995). Another possible cross-link has been suggested to derive from the peroxidization of three readily peroxidizable amino acid residues, Val, Pro, and Lys, which are abundantly present in the motifs of Val-Tyr-Lys, and Val-Lys-Pro-Tyr-His-Pro of several extensins (Varner 1994).

From the above review, it appears that extensins are incorporated into the cell wall by forming some kind of cross-link with the other components of the cell walls. However, the mechanism by which this cross-link is formed is still unclear partly because in many cases extensins are recalcitrant to extraction from the cell wall once they are covalently cross-linked with the cell wall components (Watson and Sumner, 2006).

Developmental regulation and inducibility of extensins

The distribution pattern of extensins determined by analyzing the Hyp content in several organs and tissues of the soybean plants showed that the seed coat and root nodules contain the highest ratio of Hyp to dry weight compared to roots, leaves, stems, and flowers (Cassab et al., 1985; Cassab 1986). Cassab et al., (1985) showed that the seed coat contains close to 80% of the total Hyp in the seed at all stages of development, and extensins were primarily localized in the two external layers of the

seed coat. These authors further suggested that although the seed coat extensins were slightly different from that of other tissues with respect to amino acid composition and pattern of glycosylation, they play a similar structural role in seed coat like the extensins of other tissues.

To assign a possible function to a protein, it is useful to know in what cell type it is present, as well as its sub-cellular location. Therefore, considerable research has been carried out over the last two decades to determine the tissue localization of CWPs including extensins. The data presented in Table 1-1 show that the extensin gene expression and protein localization can vary between species and between cell and tissue types. Analysis of these data and the associated publications suggests that although extensins were commonly associated with phloem tissue and cambium cells, they can also be associated with other tissues as well (Showalter, 1993).
Expression of extensins is not only regulated by the developmental program of the plants but also by other factors including biotic and other environmental factors, Table 1-2 (Showalter 1993; Hong et al., 1994; Merkouropoulos et al.,1999; Merkouropoulos and Shirsat 2003).

Developmental regulation

Developmentally, extensins are expressed in many phases of plant growth including cell division, cell elongation, root initiation, stem thickening, pollen germination and fertilization, embryo development, fruit ripening, abscission formation, and senescence (Cleland and Karlsnes, 1967; Sadava and Chrispeels, 1973; Keller and Lamb, 1989; Ito et al., 1998; Wu et al., 2001, Hall and Cannon, 2002; Merkouropoulos and Shirsat, 2003; Robert and Shirsat, 2006, Zhang et al, 2008).

Table 1-1. Tissue localization of extensins, GRPs, and PRPs (Adapted from Showalter 1993)

Protein Class	Plant System	Tissue Localization	Reference
Extensins	Soybean stems and petioles	Cambium cells, in a few layers of cortex cells surrounding primary phloem, and in a some parenchyma cells around the primary xylem; abundant in hypocotyl apical regions	Ye and Varner (1991)
Extensins	Bean stems and petioles	Cambium cells in a few layers of cortex cells surrounding primary phloem	Ye and Varner (1991)
Extensins	Tomato stems and petioles	Outer and inner phloem	Ye et al. (1991)
Extensins	Petunia stems	Outer phloem	Ye et al. (1991)
Extensins	Tobacco stems	Outer phloem	Ye et al. (1991)
Extensins	Carrot stems	Phloem parenchyma walls	Stafstrom and Staehelin (1988)
Extensins	Soybean roots	Two to three layers of cortex cells around vascular bundles and in protoxylem; abundant in root apical regions	Ye and Varner (1991)
Extensins	Tomato root	Minor components of cortical and parenchyma cell walls	Benhamou et al. (1990)
Extensins	Tobacco with bean extension-GUS transgene	Subset of pericycle and endodermal cells involved with lateral root initiation	Keller and Lamb (1989)
Extensins	Rape with rape extension-GUS transgene	Phloem of rape roots	Shirsat et al. (1991)
Extensins	Soybean seeds	Sclerenchyma tissues (palisade epidermal and hourglass cells) of seed coats	Cassab and Varner (1987)
Extensins (i.e., THRGP)	Maize	Predominantly to sites of early vascular differentiation in embryos, leaves, and roots; these sites include xylem elements and surrounding sclerenchyma in leaves and metaxylem and protoxylem in roots	Stiefel et al. (1990)
GRPs	Soybean stems and petioles	Primary xylem, primary phloem, and in newly differentiated secondary xylem	Ye and Varner (1991)
GRPs	Bean stems	Protoxylem tracheary elements of the vascular system	Keller et al. (1989b)
GRPs	Bean stems	Protoxylem, primary xylem and phloem, and newly differentiated secondary xylem	Ye and Varner (1991)
GRPs	Bean stems	Protoxylem, cell corners of protoxylem and metaxylem elements, and phloem	Ryser and Keller (1992)
GRPs	Tomato stems	Xylem vessel elements and fibers; some in outer and inner phloem fibers	Ye et al. (1991)
GRPs	Petunia stems	Vascular tissue (phloem or cambium) and to a layer of cells at the epidermis	Condit et al. (1990)
GRPs	Petunia stems	Xylem vessel elements and fibers; some in outer phloem fibers	Ye et al. (1991)
GRPs	Tobacco stems	Xylem vessel elements and fibers	Ye et al. (1991)
GRPs	Tobacco with bean GRP 1,8-GUS transgene	Protoxylem tracheary elements of the vascular system	Keller et al. (1989a); Keller and Baumgartner (1991)
GRPs	Soybean roots	Primary xylem	Ye and Varner (1991)
GRPs	Bean seeds	Tracheary elements of the vascular tissue of seed coats	Keller et al. (1989b)
GRPs	Maize embryo	Scutellar epidermal cells surrounding embryo axis; epidermal cells of leaves	Gomez et al. (1988)
PRPs	Soybean stems	Xylem vessel elements of young stems	Ye et al. (1991)

		and in both phloem fibers and xylem vessel elements and fibers of older stems	
PRPs (SbPRP1)	Soybean stems	Phloem, xylem, and epidermis	Wyatt et al. (1992)
PRPs (SbPRP2)	Soybean stems	Epidermis, cortical cells, phloem, and pith parenchyma	Wyatt et al. (1992)
PRPs (SbPRP3)	Soybean stems	Endodermis	Wyatt et al. (1992)
PRPs	Tomato stems	Xylem vessel elements and fibers; some in outer and inner phloem fibers	Ye et al. (1991)
PRPs	Petunia stems	Xylem vessel elements and fibers; some in outer phloem fibers	Ye et al. (1991)
PRPs	Tobacco stems	Xylem vessel elements and fibers	Ye et al. (1991)
PRPs	Potato stems	Xylem vessel elements and fibers	Ye et al. (1991)
PRPs	Soybean roots	Corner walls of the cortex and in the protoxylem	Ye et al. (1991)
PRPs (SbPRP1)	Soybean seeds	Group of sclerid cells of the seed coat near the hilum	Wyatt et al. (1992)
PRPs (SbPRP2)	Soybean seeds	Primarily to the aleurone layer of the seed coat	Wyatt et al. (1992)
PRPs	Maize embryo	Scutellum and in nonvascular cells from the embryo axis	Jose-Estanyol et al. (1992)
PRPs (ENOD2)	Pea nodules	Nodule parenchyma (i.e., inner cortex)	van de Wiel et al. (1990)
PRPs (ENOD2)	Soybean nodules	Nodule parenchyma (i.e., inner cortex)	van de Wiel et al. (1990)

Table 1-2**.** Conditions regulating the expression of the five major classes of structural plant cell wall proteins (adapted from Showalter 1993)

Protein Class	Condition(s)
Extensins (dicot)	Wounding, fungal infection, viral infection, fungal elicitors, endogenous elicitors, ethylene, red light, heat shock, gravity, glutathione, cell culturing, development
“Extensins” (monocot)	Development, wounding
“Extensins” (Chlamydomonas)	Development
“Extensins” (Volvox)	Development
GRPs (dicot)	Development, viral infection, salicylic acid, abscisic acid, drought stress, wounding
GRPs (monocot)	Development, water stress, abscisic acid, mercuric chloride, wounding
PRPs	Wounding, endogenous elicitors, fungal elicitor, ethylene, cell culturing, light, red light, development
PRPs (nodulins)	Development
Solanaceous lectins	Wounding, viral infection
AGPs	Development, wounding

Inducibility by biotic factors – pathogenic induction

Induction of extensin expression has been shown to be mediated by diseases including fungal and bacterial infections and elicitations. These appear to be the most common factors that induce extensin expression in plants (Showalter 1993, Cassab 1998; Mazau and Esquerré-Tugayé, 1986; Merkouropoulos and Shirsat, 2003). Corbin

et al., 1987; Memelink et al., 1993; Jose-Estanyol and Puigdomènech, 1993; Tire´ et al. 1994; Hirsinger et al., 1997; Sommer-Knudsen et al., 1998; Merkouropoulos and Shirsat, 2003; Guo and Shirsat 2006; Basavaraju et al. 2009). It has been shown by immunochemical studies that extensins accumulate in cell walls close to sites where microbial growth was restricted by the plant (Esquerre-Tugaye et al. 1985). Another study reported RNA gel-blot and histochemical analyses of *Arabidopsis* transgenic lines expressing an extensin-GUS fusion construct that showed extensin induction by *Brassica* pathogen, *Xanthomonas campestris* pv. Campestris and that this induction was restricted to tissues close to the site of infection (Merkouropoulos and Shirsat, 2003). Wei and Shirsat (2006) reported that the over-expression of an extensin gene (*EXT1)* in *Arabidopsis* restricted the invasion of *Pseudomonas syringae* into the plant. They showed that the lesions on the transgenic plants due to the infection of the pathogen were five times smaller than those on the wild-type plants. It has recently been shown that in the resistant genotypes of sorghum plant, there was a significant correlation between induced accumulation of H_2O_2, and cell wall cross-linking, as evidenced by extensin accumulation, and cessation of pathogen growth (Basavaraju et al., 2009). Bradley (1982) reported that treatment of bean or soybean cells with fungal elicitor or glutathione caused a rapid insolubilization of two Hyp/Pro-rich structural proteins in the cell wall. Brisson et al., (1994) have shown that when cell walls were subjected to a short period (30 min) of elicitation, they became more resistant to enzyme digestion as indicated by the yield of protoplasts released. This effect was suggested to be the result of extensin cross-linking in the cell wall.

Elicitors and suppressors of extensin accumulation have been solubilized from plant cell walls by the pectinolytic enzyme, endopolygalacturonase (EPG), purified from bean pathogen, *Colletotrichum lindemuthianum* (Boudart et al., 1995). It has been shown that small galacturonides (elicitors), solubilized from the cell wall, can trigger

extensin gene expression and elicit a 40–70% Hyp increase in the cell wall. In contrast, pectic fragments of higher molecular masses had the ability to suppress Hyp deposition in the cell wall. The above observations suggest that elicitation and/ or suppression of Hyp are correlated with regulation of extensin gene expression in the cell wall (Boudart et al 1995). Mazau et al., (1987) suggested that extensin helps to produce an impenetrable physical barrier for the pathogens by binding to their surfaces. Mellon and Helgeson (1982) proposed that positively charged extensin molecules might also interact ionically with the negatively charged surfaces of plant pathogens causing agglutination of pathogens.

Inducibility by biotic factors – insect induction

Infestation of plants by insects can also induce extensin expression (Niebel et al., 1993; Lambert, 1995. van der Eycken et al., 1996; Williamson and Hussey, 1996; Cassab 1998; Gheysen, and Fenoll 2002). It has been reported that when sedentary endoparasitic root knot nematode, *Meloidogyne javanica* and cyst nematode *Globodera tabacum* ssp solanacearum infected tobacco roots, they induced the expression of an extensin gene (Niebel et al., 1993). van der Eycken et al. (1996) reported that in a compatible interaction, root-knot nematode (*Meloidogyne incognita*) induced the expression of two extensin genes at the feeding site of tomato.

Inducibility by wounding

Wounding is another cause of extensin expression in plants (Adams et al. 1992; Bown et al. 1993; Parmentier et al. 1995; Wycoff et al. 1995; Ahn et al.1996; Hirsinger et al. 1997, 1999; Merkouropoulos and Shirsat, 2003). It has been suggested that extensin expression as a result of insect attack might be related to wound formation caused by insect feeding. For example, it has been shown that the gene, 6PExt1.2 that encodes an extensin protein was induced by wounding in protoplasts and in leaf strips as well as due to infection of stem by *Agrobacterium tumefaciens* (Parmentier et al. 1995). Bown et al.,

(1993) reported that wounding of the potato tubers caused a significant increase in extensin-like mRNAs in the plants. In another study, it was reported that a chimeric extensin gene, SbHRGP3-GUS, constructed from a soybean extensin and expressed in tobacco plants required sucrose for its wound-inducible expression (Ahn et al. 1996). Merkouropoulos et al., (1999) reported that the *Arabidopsis* extensin gene, atExt1 was normally expressed in the root but remained silent in the leaf. However, when the plants were wounded, the above expression pattern was reversed.

From the above review, it can be concluded that increased synthesis of extensins occurs in response to biotic factors, and this is followed by their presumed deposition and oxidative cross-linking in the cell wall resulting in the formation of physical barriers against invading pathogens and insects (Showalter 1993).

<u>Inducibility by hormones and defense chemicals</u>

Extensin expression has also been reported to be strongly regulated by chemicals such as ethylene, abscisic acids (ABA), methyl jasmonate (MeJa), salicylic acid (SA). These are all hormones associated with various defense responses in plants (Ecker and Davis 1987; Tagu et al. 1992; Showalter et al. 1992; Josè and Puigdomènech, 1993; Memelink et al. 1993; Shirsat et al. 1996; Ahn et al. 1996; Hirsinger et al. 1999; Merkouropoulos et al. 1999; Merkop. and Shirsat 2003; Nik and Shirsat, 2005). A study showed that both wounding and ethylene treatments induced the accumulation of extensin mRNA in the carrot root; but the mRNA induced by ethylene treatment was different from that of wounding. Based on their results, the authors suggested that the two signals, ethylene and wounding, were distinctly different (Ecker and Davis 1987). Another study showed that the treatments of *Brassica napus* with ABA, MeJa, and SA induced the accumulation of two extensin mRNA transcripts. The expression was, however, detected 12-hr earlier for MeJa and SA than ABA. Based on the above observations, the authors suggested that these extensin glycoproteins were

associated with mechanisms related to plant defense (Nik and Shirsat 2005). Merkouropoulos and Shirsat (2003) showed that the *Arabidopsis* extensin gene, atExt1 was induced not only by wounding and pathogen infection but also by exogenously applied hormones such as SA, MeJa, auxins, and brassinosteroids. These observations suggest that the expression of extensins in plants is mediated in part by hormones in the defense signaling pathway, and as described above, by pectic fragments in the cell wall (Cassab 1998). Sucrose may also influence signal transduction during extensins induction, as sucrose has been shown to be required for wound-inducible expression of a soyben extensin gene (Ahn et al., 1996).

Although the structure and expression patterns of extensins give some clues to their possible roles within the plant, direct functional evidence of extensin function is still very limited. Research to date has established that extensins are cell wall associated structural glycoproteins that play a significant role in development, plant defense, and wound healing (Cassab 1998). To obtain direct functional evidence of extensins, anti-sense gene technology was applied (Memelink et al., 1993). Results demonstrated that tobacco transgenic plants over-expressing anti-sense extensin gene constructs exhibited significantly lower concentration of Hyp compared to transgenic plants over-expressing sense extensin construct. However, this reduction of Hyp (and presumably HRGP) concentration did not have significant effect on the phenotype of the plant or on the structure of cell wall (Memelink et al., 1993). Based on the these results it can be suggested that since only one cell wall structural protein (extensin) was diminished by anti-sense in this study, other cell wall proteins, such as other extensins, PRPs or GRPs may have compensated for the loss of extensin (Cassab 1998). Based on the above findings, it was suggested that further research was needed with suitable systems such as RNA interference (RNAi) for selective repression of CWPs to determine the precise function(s) of extensins in the plants. In subsequent research, Keskiaho et al., (2007)

reported that suppression of prolyl hydroxylases that catalyze the formation of Hyp by RNAi leads to a defective cell wall structure in the green algae *Chlamydomonas*. This observation suggests that simultaneous silencing of extensins and related genes by RNAi can more likely to provide mutant phenotype of cell wall extensins.

B. Proline-Rich Proteins (PRPs)

General characteristics

PRPs, in general, have been shown to contain a cell wall-directed signal peptide similar to extensins, followed by Pro-Pro repeats which occur within a variety of other larger repeat units. The most frequent repetitive element is Pro-Pro-Val-Tyr-Lys (Showalter, 1993; Jose-Estanyol and Puigdomenech 2000). Although PRPs do not contain a repeating pentapeptide motif similar to that of extensins, they contain repeating pentapeptide sequence of Pro-Pro-X-Y-Lys (where X and Y can be Va, Tyr, His, and Glu), Showalter, 1993. The main difference between PRPs and extensins is that in extensins, most of the Pro residues are believed to be post-translationally hydroxylated and the proteins are believed to be abundantly glycosylated. In contrast, in PRPs, only about half of the Pro residues are apparently hydroxylated, and the proteins are either not glycosylated or are only lightly glycosylated (Carpita and Gibeaut, 1993; Keller, 1993; Showalter, 1993; Cassab 1998; Jose-Estanyol and Puigdomenech 2000).

Amino acid sequence analysis of soybean cell wall PRPs showed that Hyp occurs only in the second position of a PRP pentapeptide repeat, Pro-Hyp-X-Y-Lys (Lindstrom and Vodkin, 1991). However, in gymnosperm PRPs, it occurs in the second and third positions as well as only in the third position of two hexapeptide repeats, Pro-Hyp-Hyp-Val-Tyr-Lys and Pro-Pro-Hyp-Val-Val-Lys, respectively (Kieliszewski et al., 1992). The maize (monocot) PRPs have been shown to contain an N-terminal Pro-rich domain with numerous Pro-Pro-Tyr-Val and Pro-Pro-Thr-Pro-Arg-Pro-Ser repeats.

Similar domain sequence has also been reported in two dicot PRPs, bean and tomato (Sheng et al., 1991; Salts et al., 1991). Thus, the PRPs that have been isolated and characterized until now are only lightly or not glycosylated, and contain approximately equimolar quantities of Pro and Hyp residues (Averyhart-Fullard et al., 1988; Datta et al., 1989; Kleis-San Francisco and Tierney, 1990).

Interaction of PRPs with the cell wall

As stated earlier, PRPs represent one of the four major families of CWPs that have been identified in higher plants (Carpita and Gibeaut, 1993; Showalter, 1993; Cassab, 1998). A review of literature suggests that the interaction of PRPs with the cell wall is comparable to that of extensins. For example, PRPs have been proposed to be insolubilized in the cell wall, form isodityrosine cross-links with wall components including PRPs, GRPs, and extensins, and involved in the process of ionic interaction with pectins, similar to that of extensins (Datta et al., 1989; Kleis-San Francisco and Tierney, 1990, Bradley et al., 1992, Ye et al., 1991). However, results from indirect immunological studies appear to have not provided strong evidence of PRP cross-linking with the wall (Bonilla et al. 1997; Bradley et al. 1992; Marcus et al., 1991). Protein localization studies (Table 1-1) suggest that PRPs may function in determining cell-type-specific wall structure during plant development. Immunohistochemical analysis using antibodies raised against soybean PRP (SbPRP2) observed PRP accumulation in soybean protoxylem cells within the root, and xylem and phloem fibers within the stem. Based on the above observations, it was suggested that these proteins were critical for maintaining structural integrity of mature tissues (Ye et al., 1991). PRPs may also play a similar role during seed development, since seed coat integrity appears to be altered in soybean lines that failed to accumulate these proteins within their cell walls (Nicholas et al., 1993).

Developmental regulation and inducibility of PRPs

PRPs were first identified when they accumulated in the cell wall in response to physical damage (Chen and Varner, 1985; Tierney et al., 1988). They were subsequently shown to be temporally regulated during plant development. PRPs have been proposed to play important roles during normal development, nodule formation, and defense responses of plants (Showalter 1993, Cassab 1998; Peng et al. 2008), although their precise functions are not clearly known.

Several studies have indicated that PRPs are implicated in various aspects of plant development. For example, PRP gene expression is associated with the early stages of nodule formation in legume roots (Franssen et al., 1987; van de Wiel et al., 1990; Wilson et al., 1994), development of soybean and bean seedlings, leaf, stem, and seed coat (Hong et al., 1989; 1990; Kleis-San Francisco and Tierney, 1990; Lindstrom and Vodkin, 1991; Ye et al., 1991, Sheng et al., 1991), and with the early stages of tomato fruit development (Santino et al., 1997). Analyses of *in situ* hybridization and reporter gene expression studies suggest that the spatial pattern of PRP gene expression is tightly regulated. For example, the soybean SbPRP1 and SbPRP2 transcripts have been shown to be localized to sclereids, the inner integument of the seed coat, and the epidermal, cortical, and endodermoidal cells of young seedlings (Wyatt et al., 1992; Suzuki et al., 1993). Fowler et al., (1999) reported from a study of expression patterns of four Arabidopsis PRP genes (AtPRP1 to AtPRP4) that these proteins are involved in specifying cell-type-specific wall structures. Thus, the PRPs appear to display tissue- and cell-specific patterns of expression, and are commonly present in the xylem, fibers, sclereids, epidermis, aleurone, and nodule parenchyma (van deWiel et al. 1990, Ye et al., 1991; Wyatt et al., 1992).

Similar to extensins, the expression of PRP genes is induced by wounding, pathogen infection and elicitation, and defense hormones such as ethylene (Marcus et al., 1991; Sheng et al., 1991; Mergold-Villasenor et al., 1996; Peng et al., 2008). PRPs are rapidly synthesized and insolubilized within the cell wall matrix in response to physical damage, treatment with fungal elicitors, and pathogen infection (Kleis-San Francisco and Tierney, 1990; Bradley et al., 1992; Brisson et al., 1994). Based on the above findings, it was suggested that PRPs contribute to defense reactions against physical damage and pathogen infection.

In conclusion, it appears that PRPs share many things in common with extensins with respect to their gene expression patterns. However, like extensins, the direct functional evidence of PRP function is still limited. Thus, further research is required, for instance by applying RNAi methodology, to elucidate the precise role of PRPs in plant cell wall architecture.

C. Glycine-Rich proteins (GRPs)

General characteristics

The GRP family of proteins has been classified as a third group of cell wall associated structural proteins, like extensins and PRPs. GRPs, in general, are characterized by their high content of Gly residues (Ringli et al. 2001; Mousavi and Hotta, 2005). Although some RNA-binding proteins are also rich in Gly residues, this review will focus only on cell wall associated structural GRPs (Ryser and Keller 1992, Ryser et al. 1997; 2004; Sturm 1992; Parsons and Mattoo, 1994; Ringli et al. 2001). These are characterized by their repetitive structure containing 60-70% Gly residues arranged in Gly-X repeat units (where X is often Gly, but can also be Ala, Ser, Val, His, Phe, Tyr and Glu or other amino acids). GRPs usually contain an N-terminal signal peptide suggesting that these proteins may be transported to the cell wall. For example,

analysis of cell wall protein fraction of pumpkin seed coat showed 47% Gly content suggesting that GRPs are cell wall proteins (Varner and Cassab, 1986). However, some of the dicot GRPs apparently do not contain signal peptides. Other reports showed that neither the rice GRP nor the maize GRP includes a signal peptide. By contrast, another rice GRP includes a signal peptide and the mature protein contains 67% Gly residues (Lei and Wu, 1991). Based on the above observations, it was suggested that a subset of GRPs might have been localized in the cytoplasm (Showlter 1993). The GRPs appear to be only slightly glycosylated, and predicted to form a β-pleated sheet, although this has not been confirmed (Ringli et al, 2001).

Interaction of GRPs with cell walls

Structural GRPs have been suggested to be insolubilized in the cell wall (Varner, 1994) similar to extensins and PRPs. This suggestion was made based on the finding that the bean GRP 1.8 is insolubilized in the cell wall (Keller et al., 1989). However, whether this is true for all cell wall GRPs, is not known. It has been suggested that some GRPs containing Tyr residues might become linked to the aromatic residues of lignin. This hypothesis is based on the observation that when Tyr residues were added to synthetic proteins (poly-Lys/Tyr), the cross-linking of these proteins into lignin-like dehydrogenation products was enhanced (McDougall et al., 1996). It was also suggested that the GRPs that do not contain Tyr residue might utilize His, Val, Glu, and Gln residues to develop linkages with lignin similar to those of extensins. For example, transglutamylation has been suggested to be a possible source of cross-linking for both GRPs and nodulin PRPs, since both proteins contain high levels of Glu. However, the extraction of GRP from soybean aleurone layers with hot water suggested that GRP may be associated with cell wall polysaccharides by nonionic bonds (Matsui et al., 1995).

Developmental regulation and inducibility of GRPs

Expression of GRPs is regulated both developmentally and by environmental stimuli (Showalter, 1993; Cassab, 1998). Developmental regulation of GRPs has been implicated in a number of growth and development processes including vascular development, nodule formation, and flower development (Condit et al., 1990; Ryser et al., 1992; Oliveira et al. 1993; Kuster et al. 1995; Ryser et al., 1997; Ringli et al., 2001; Kevei et al., 2002; Tao et al., 2006; Chen et al. 2007; Fusaro and Sachetto-Martins, 2007). Although GRPs have been assumed to be associated with lignified cells, ultrastructural studies revealed that GRP proteins are localized mainly in the primary walls of non-lignified protoxylem cells (O'Brian,1981; Rouser et al., 1997). It has been suggested that GRPs appear to mediate a repair process in the development of protoxylem cells. For example, it was shown that after the death of living protoxylem and xylem parenchyma cells, where GRPs are mainly synthesized, additional GRPs are exported from neighboring xylem parenchyma cells to the protoxylem wall (Rouser et al., 1997; Ringli et al., 2001). Based on the above observations, the authors proposed that GRPs are part of a repair system of the plant during the stretching phase of protoxylem.

Tao et al., (2006) studied the regulation of the expression of a maize silk-specific GRP gene, zmgrp5. They reported that the GRP was secreted into the extracellular matrix and was localized in the cell wall fraction mainly through interactions mediated by covalent disulphide bridges. Chen et al., (2007) analyzed the expression pattern of an *Arabidopsis* GRP gene, *AtGRP9*. Analysis of GUS or GFP expression under the control of the *AtGRP9* promoter showed that AtGRP9 was expressed in the vascular tissue of the root. Sub-cellular localization analysis further demonstrated that *AtGRP9* protein was localized in the cell wall and in the cytoplasm. Kevel (2002) studied four genes encoding

small GRP proteins and reported that expression of these genes was undetectable in flower, leaf, stem, and hypocotyl cells, whereas expression was highly induced during root nodule development, suggesting that GRP genes act as nodulins. Fusaro and Sachetto-Martins (2007) reported that the expression of genes encoding GRPs was developmentally regulated and involved in several independent physiological processes. They further reported from expression analysis that the *AtGRP2* gene was active in meristematic tissues and modulated during flower development. Down-regulation of *AtGRP2* gene using gene silencing techniques resulted in early flowering, altered stamen number and affected seed development. These results suggest that RNAi is relatively more suitable technique for functional characterization of cell wall associated genes including genes encoding GRPs.

GRPs appear to be induced by a number of factors including infection by pathogens, SA, ABA, wounding and drought stress, similar to extensins and PRPs (Table1-2). For example, Molina et al. (1997) reported that the transcript level of two barley GRP genes was increased by cold treatment and due to infection by fungal pathogens, *Erysiphe graminis* and *Rhynchosporium secalis,* in both compatible and in incompatible interactions. Kevel et al., (2002) demonstrated that bacterial infection induced the expression of the nodule-specific GRP genes. Park et al. (2008) demonstrated that certain *Arabidopsis* GRP genes such as *AtGRP5* and *AtGRP23* were induced by 16-hydroxypalmitic acid (HPA), a major component of cutin. According to them, these GRPs might play important roles against pathogen invasion mediated by cutin monomers in the cell wall. They showed that ABA and SA treatments enhanced the transcript levels of these GRPs, and HPA treatment effectively elicited the accumulation of H_2O_2 in rosette leaves of Arabidopsis. Several other studies demonstrated that GRPs may also play important roles in wound healing and freezing tolerance (Castonguay et

al., 1993; Condit et al., 1990; Oliveira et al., 1993; Kuster et al., 1995; Ryser et al., 1992; 1997).

The above review suggests that although GRPs are structurally less related than extensins and PRPs, their regulatory and functional properties are comparable to these two other classes of CWPs. Thus, further research is also required as suggested for extensins and PRPs to reveal the precise role of GRPs in plant cell wall.

D. Arabinoglactan proteins (AGPs)

General characteristics of AGPs

AGPs have been described as a class of Hyp-rich glycoproteins which may contain Ala-Hyp repeats. In addition, their protein moiety can also be rich in Ser, Ala, Thr, Gly and Lys residues (Gleeson et al., 1985; 1989, Yang and Showalter 2007). Unlike extensins, PRPs and GRPs, AGPs are usually very soluble proteins. They are widely distributed in plants, and are heavily glycosylated. They may contain <10% protein, and >90% carbohydrate (Nothnage, 1997; Showalter and Varner, 1989; Showalter, 1993; Showalter 2001). As their name implies, AGPs contain O-linked galactose and arabinose as the major carbohydrate constituents. The carbohydrate side chains, which may contain more than 50 residues, are primarily linked by O-glycosylation to the OH group of Ser and Hyp (Figure 1-5). Their molecular weights have been reported to be very variable because of their different extents of glycosylation (Keegstra et al., 1973, Showalter 1993). Although their solubility properties greatly facilitated their extraction, their extensive glycosylation have made them resistant to proteolytic cleavages. AGPs have isoelectric points in the range of pH 2 to 5 (Showalter 1993; Cassab 1998).

Interaction of AGPs with cell wall

Although conclusive cellular localization of many AGPs has not been possible because of their extreme solubility, they are often found as constituents of the cell wall as well as localized in plasma membrane, periplasm, and extracellular secretions (Showalter 2001; Lamport et al. 2006). It has been suggested that most AGPs possess highly labile glycosylphosphatidylinositol (GPI) lipid anchors, which transiently attach AGPs to the plasma membrane before they are released into the cell wall following the hydrolysis of GPI anchor (Gilson et al., 2001). In the cell wall, AGPs do not appear to be covalently linked to other wall components and therefore they may not play a structural function. In the cell walls, AGPs may be involved in directing planes of growth and development and participating in cell shape (Cassab 1998). Several studies suggested that AGP may mediate signal transduction at the cell wall-plasma membrane interface (Kjellbom, 1997; Gao and Showalter, 1999a; Showalter, 2001).

Developmental regulation and inducibility of AGPs

AGPs are expressed in various organs of plants including seedlings, leaves, stems, roots, flowers and seeds. Studies have suggested that AGPs are involved in a number of plant growth and developmental processes (Sardar et al., 2006). Examples include female gametogenesis (Acosta-Garcia and Vielle-Calzada, 2004), cell proliferation (Serpe and Nothnagel, 1994; Langan and Nothnagel, 1997), vascular differentiation (Schindler et al., 1995), somatic embryogenesis (Thompson and Knox, 1998; van Hengel et al., 2001), cell expansion (Ding and Zhu, 1997), pollen germination and growth (Cheung et al., 1995), root regeneration and seed germination (van Hengel and Roberts, 2003), hormone responses (Park et al., 2003) and programmed cell death (Chaves et al., 2002)

AGPs also accumulate in response to wounding (Fincher et al., 1983). Narayanasamy, (2006) reported that upon wounding, AGPs were secreted in large

amounts at the wound site, suggesting that they may act as a physical barrier for pathogens by producing a gel plug. The above observations suggest that AGPs may be involved in defense responses in plants similar to that of other CWPs. However, another study showed that two orthologus AGPs, NaAGP4 and LeAGP-1 respectively from tobacco and tomato, were rapidly suppressed by tissue wounding and by pathogen infection (Gilson et al., 2001).

From the above review it appears that although AGPs are less likely to provide structural support to plant cells because of their lack of insolubilization and cross-linking in the cell wall, they appear to play defense responses like other CWPs. Further research is necessary to precisely understand the regulatory functional properties of AGPs.

2. APPLICATION OF REVERSE GENETICS IN THE FUNCTIONAL CHARACTERIZATION OF *Arabidopsis* NOVEL GENES

Since the completion of the *Arabidopsis thaliana* genome sequence and the identification of approximately 26,000 predicted genes (Arabidopsis Genome Initiative, 2000), functional characterization of these genes has become a challenge to the research community. Reverse genetics methodologies such as sequence-indexed insertional mutagenesis, gene silencing by RNA interference (RNAi), and gene over expression are being routinely used for the gene characterization. This review briefly discusses the advantages and limitations of these techniques.

Although numerous sequence-indexed knock out mutant lines of many *Arabidopsis* genes have been developed with insertion sites located at different positions of the genes, Table 1-3 (Alonso et al., 2003; Sessions et al., 2002), the majority of these loss-of- function mutants do not show detectable phenotypes (TAIR), Figure 1-5. This might also be due in part to the high degree of gene redundancy within extensive

regions of duplication (AGI 2000) within the *Arabidopsis* genome. About 17% of Arabidopsis genes have been reported to belong to tandemly repeated families (Tantikanjana et al., 2004) that do not segregate because of linkage. This presents a serious obstacle to some types mutational analyses (Jander and Barth 2007).

Table 1-3. Distribution of T-DNA insertions in genes and intergenic regions of *Arabidopsis* chromosomes (adapted from Alonso et al., (2003).

	Chr. 1	Chr. 2	Chr. 3	Chr. 4	Chr. 5	Total
Promoter	5,488	3,376	4,452	3,076	4,900	21,292
5'UTR	1,243	737	951	680	1,099	4,710
Coding exon	5,089	2,960	3,988	2,871	4,440	19,348
Intron	2,663	1,507	1,840	1,681	2,284	9,975
3'UTR	1,621	914	1,263	966	1,535	6,299
Intergenic regions	6,861	4,323	5,180	3,813	6,321	26,498

Furthermore, many insertion mutations can be lethal causing the death of mutants at gametophytic or embryonic stages, and some may give rise to weak phenotypes particularly if they are in the promoters or 3'-UTR or close to the 3'-end of the genes (Krysan et al., 1999; Helliwell et al., 2002). Therefore, in recent years, post-transcriptional gene silencing technology, RNAi, has been used to simultaneously silence functionally redundant genes (Vance and Vaucheret 2001; Wesley et al., 2001; Jones 2002; Helliwell et al., 2002; Diallo and Schepers 2003; Kinder and Martienssen 2003; Helliwell and Waterhouse 2003; Wielopolska et al., 2005; Fire et al., 1998).

RNAi is also considered to be a normal defense mechanism inherent in both plants and animals to protect their genomes against certain viruses or transposable elements that form double-stranded RNAs (dsRNAs) inside the host cell nuclei (Vance and Vaucheret 2001; Stram and Kuzntzova, 2006; Blevins et al., 2006). Briefly, the

mechanism of RNAi is invoked when a dsRNA is formed inside a host cell nucleus. After being transported to the cytoplasm, it is cleaved by RNase III-like enzyme (RNase III helicase) called DICER into 20-25 nucleotide long small interfering RNAs (siRNAs) with 3' dinucleotide overhangs. These siRNAs are then assembled into endoribonuclease-containing complexes known as RNA-induced silencing complexes or RISCs. The siRNA strands are then unwound by the RISCs. One of the siRNA strands, also referred to as the guide strand, subsequently guides the RISC to a complementary mRNA molecule, where it cleaves and destroys the cognate mRNA causing the silencing of the gene. The siRNA is thought to provide target specificity to RISC through base pairing of the guide strand with the target mRNA. Cleavage of cognate mRNA takes place near the middle of the region bound by the siRNA strand (Diallo and Schepers 2003).

Although RNAi is an effective method of studying the phenotype of a gene without disrupting its integrity, it may not totally eliminate the expression of the gene. Therefore, RNAi mutagenesis is sometimes referred to as gene knockdown (Voorhoeve and Agami, 2003) as compared to gene knockout in the case of T-DNA insertion mutagenesis. Furthermore, RNAi methodology has other reported limitations including instability of its phenotypes, variable levels of residual gene activity and the inability to simultaneously silence several unrelated genes (Hannon 2002; Bargmann 2001; Wesley et al., 2001; Kamath et al., 2001).

Another reverse genetic approach that has been used since plant transformation protocols have become available is the over-expression of genes using strong constitutive promoters such as CaMV35S (Benfey, Chua 1990; Benfey et al., 1990; Bert et al., 1999; Jackson et al. 2002). Gene over-expression is particularly important in situations where there are functionally redundant genes or when a knockout mutation is deleterious. In case of functional redundancy, this approach can be used to specifically up-regulate a specific gene with minimal interference with

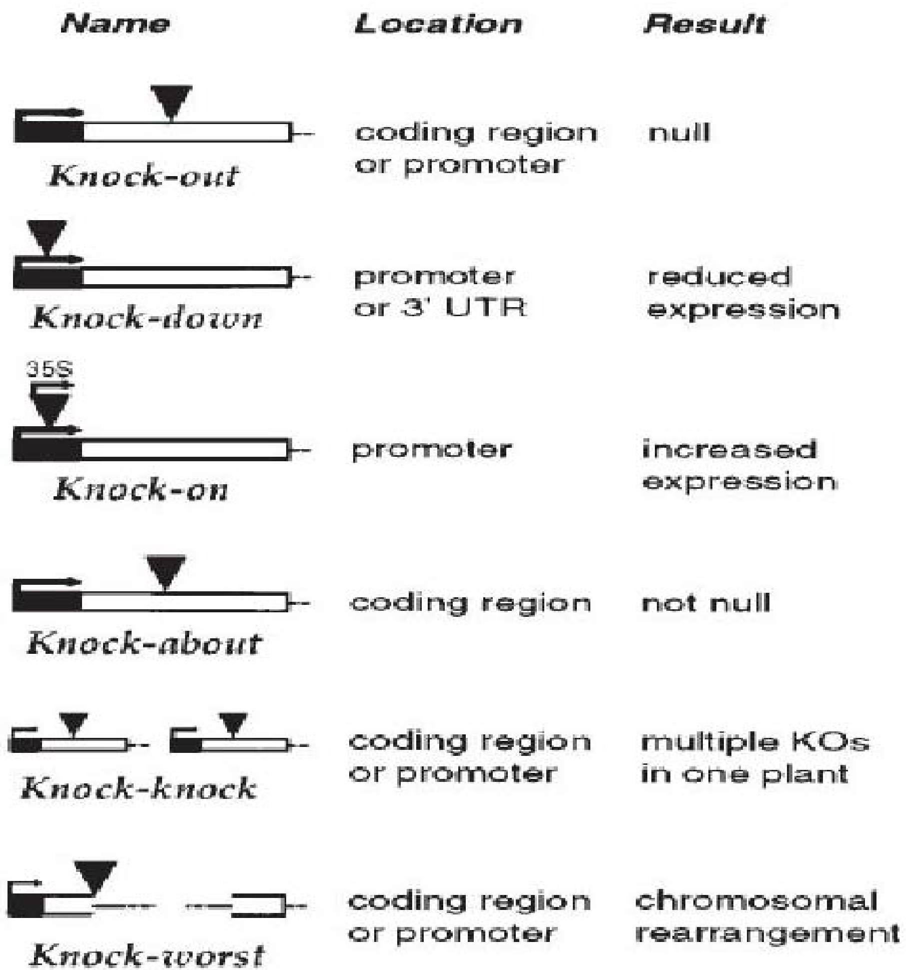


Figure 1-6 Different outcomes of T-DNA insertion mutagenesis in *Arabidopsis* chromosomes. White box represents the coding region of the gene, black region with the arrow indicates the promoter, and the black triangle represents T-DNA element. KOs, knockouts; UTR, un-translated region (from Krysan et al., 1999)

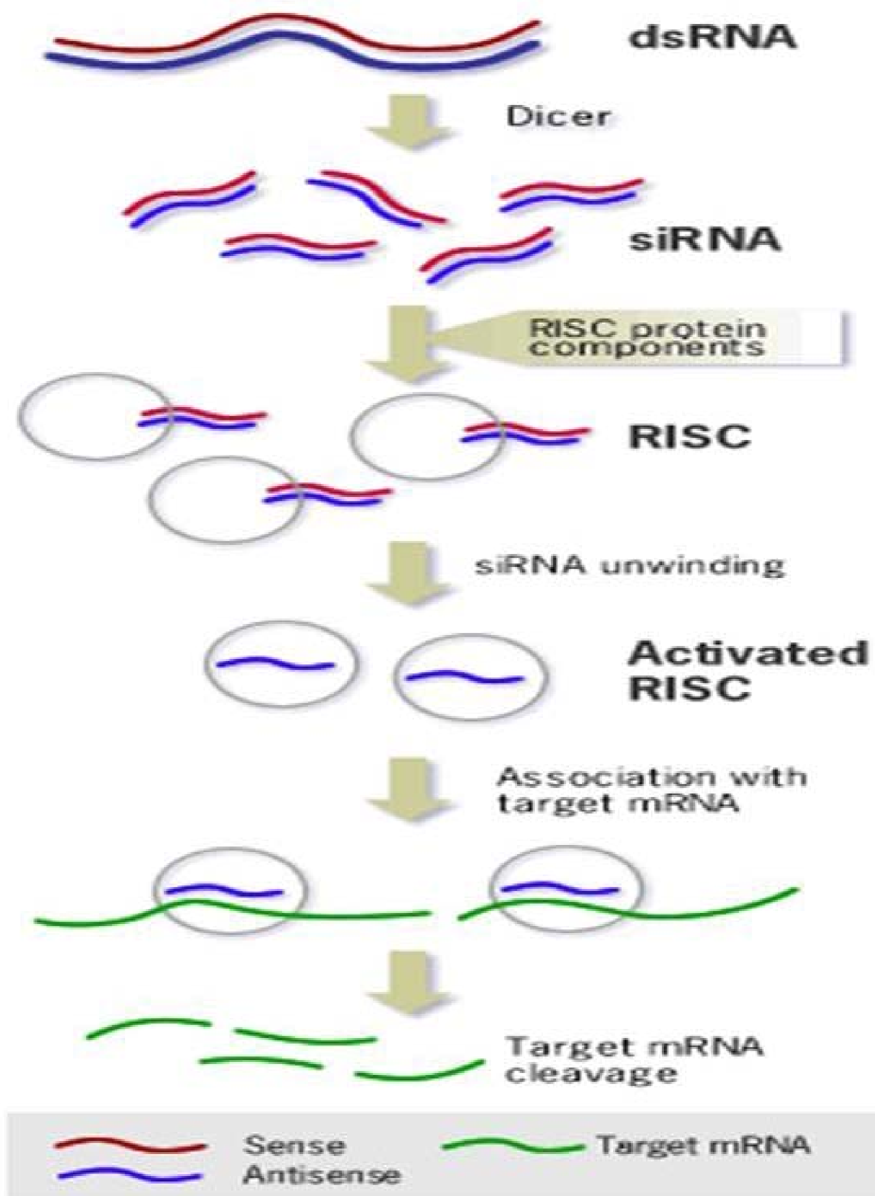


Figure 1-7: Diagrammatic representation of the mechanism of RNA interference (from Diallo and Schepers 2003). dsRNA = double-stranded RNA, siRNA = small interfering RNA, RISC = RNA-induced silencing complex

related genes (Martienssen and Irish 1999; Vision et al. 2000). It is also useful in the case of polyploidy, where knockout mutation is not very effective (Llyod 2003).

Over-expression has been used extensively in research and biotechnology. For example, relevant to CWPs, a study reported that over-expression of a cell wall associated native extensin gene, EXT1, in *Arabidopsis* plants by constitutive CaMV35S promoter significantly reduced the invasion capability of the bacterial pathogen, *P. syringe* to the host plant (Wei and Shirsat 2006).

It is clear that some functional information can be obtained for a particular gene simply by over-expressing it. However, there are some limitations to this strategy. For example, expression of a gene outside of its natural context may cause phenotypic effects that are not relevant to its normal function. Furthermore, the over-expression transgene-construct itself can induce phenotypic changes due to insertion in another native gene causing knockout of the inserted native gene. Therefore, it is necessary to produce multiple independent transgenic lines with unlinked insertion positions to circumvent this problem. If most or all of the transgenic over-expression lines give a similar phenotype, then it can be assumed that it is due to the over-expression of the gene of interest. Producing multiple transgenic lines is also desirable to obtain a set of lines with varying levels of transgene expression.

In conclusion, despite some limitations discussed above, reverse genetics methodologies will continue to play important roles in functional genomics particularly in the characterization of known sequences.

3. PROSPECTS OF DEVELOPING DISEASE AND INSECT RESISTANT CROPS THROUGH GENETIC MANIPULATION OF CWPS

Plants, in their natural growing environment, are frequently subjected to various types of stress including disease and insect infestations. This causes detrimental effects on plant growth and productivity. For instance, it has been estimated that one fifth or more of the world's food grains are damaged each year due to infestation of insects and diseases in many parts of the world (Bergvinson and García-Lara, 2004). One way to increase the quantity and quality of food grains is to reduce the damage caused by these organisms. One of the strategies to accomplish the above objective is to develop disease and insect resistant/tolerant plants/crops.

There are two main strategies in developing disease or insect resistant/tolerant plants/crops: (i) conventional breeding technique using hybridization, back-crossing, and selection methods, and (ii) applying transgenic technologies either through introduction of novel genes from other sources or through manipulation of the expression level of the endogenous/native genes.

The conventional breeding strategy is time-consuming and can be compounded by undesirable traits because of lack of segregation of the traits causing linkage drag. Of the two transgenic strategies, transferring of disease/insect resistance/tolerance genes from other sources to target plants/crops may have some limitations. Following are two examples: (a) if the resistance/tolerance character is polygenic i.e. controlled by more than one gene, transferring of one gene by genetic engineering may not provide expected resistance/tolerance to the target plant/crop, and (b) the introduced foreign gene may not be properly expressed in the new host.

On the other hand, manipulation of endogenous defense-related genes that are induced in target plants in response to pathogen or insect attack can more likely increase the tolerance levels of the plants to the incident pests (Huckelhoven, 2007). For

example, it has been shown that over-expression of a tomato native gene, Pto, that encodes a Ser/Thr kinase and confers resistance against bacterial strain of *Pseudomonas syringae* pv tomato under the control of CaMV35S promoter, activated the defense responses in tomato even in the absence of pathogen inoculation (Tang et al., 1999). Thus, transgenic approaches in the manipulation of cell wall associated structural genes (Table 1-1) can confer increased tolerance against pathogens and insects.

It is known that cell walls provide basal defense to plants against invading pathogens and insects (Huckelhoven, 2007). It has also been established that among the components of the cell wall, CWPs play a direct role in this defense mechanism (Showalter 1993; Cassab 1998). Thus, the future prospect of developing disease/insect resistant/tolerant plants/crops through genetic manipulation of endogenous cell wall structural genes is promising.

4. CONCLUSIONS AND OBJECTIVES

CWPs belong to four major classes of multi-gene family glycoproteins including HRGPs (extensins), PRPs, GRPs, and AGPs. As reviewed above, each of these families of protein has been reported to possess specific characteristics, and contribute to cell wall architecture, and plant defense against pathogens, insects, and mechanical injuries (Showalter 1993, Cassab 1998; Jose-Estanyol and Puig 2000; Jamet et al., 2006, Hayashi 2006). For example, extensins have been reported to contribute to the construction of cell wall by forming cross links among themselves and with the other cell wall components, play important roles during development and wound healing, and act as a barrier against invading pathogens and insects (Jamet et al., 2008). Because of my interest in cell wall associated structural proteins, the current project was undertaken with an objective to characterize functionally an uncharacterized novel gene,

AT5G09530, of *Arabidopsis thaliana*. This gene is referred to as my gene of interest (GOI) and designated as *PELPK1* in Chapter 2 based on thorough bioinformatics analysis. *PELPK1* has been annotated by others as a HRGP- family protein containing a Pro-rich extensin domain (TAIR) and as a cell wall-associated glycoprotein. The extensin domain, as reviewed above has been shown to contain a repetitive region rich in Pro/Hyp with a main repeating pentapeptide motif of Ser-(Hyp)4 residues (Showalter 1993; Cassab 1998).

For functional characterization of *PELPK1*, which has been the primarily objective of this project, the following investigations were carried out: (i) bioinformatics analysis, (ii) expression analysis (promoter fusion and translational fusion), (iii) mutational analysis (sequence-indexed T-DNA insertion, over-expression, RNA interference), and (iv) proteomic analysis. I addressed the following questions: (i) Does *PELPK1* encode a HRGP? (ii) Does it contain an extensin domain? (iii) What is the biochemical function of the *PELPK1*? (iv) Where is the PELPK1-protein localized in the cell? (v) When and where does PELPK1-protein act during normal development of *Arabidopsis* plants? (vi) Does the PELPK1-protein contribute to cell wall architecture? (vii) Does the PELPK1-protein contribute to stress responses?

5. REFERENCES

Acosta-Garcia G, Vielle-Calzada JP (2004) A classical arabinogalactan protein is essential for the initiation of female gametogenesis in Arabidopsis. Plant Cell 16: 2614-2628

Adams CA, Nelson WS, Nunberg AN, Thomas TL (1992) A wound-inducible member of the hydroxyproline-rich glycoprotein gene family in sunflower. Plant Physiol 99:775–776

Ahn JH, Choi Y, Kwon YM, Kim S-G, Choi YD, Lee JS (1996) A novel extensin gene encoding a hydroxyproline-rich glycoprotein requires sucrose for its wound-inducible expression intransgenic plants. Plant Cell 8:1477–1490

Albenne C, Canut H, Boudart G, Zhang Y, San Clemente H, Pont-Lezica R, Jamet E. (2009). Plant cell wall proteomics: mass spectrometry data, a trove for research on protein structure/function relationships. Mol Plant. 2009 Sep;2(5):977-89

Alberts, B., D. Bray, J. Lewis, M. Raff, K. Roberts, and J. D. Watson. 1989. Molecular Biology of the Cell, 2nd edition. New York: Garland Publishing.

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., Ecker, J.R. (2003) "Genome-wide insertional mutagenesis of Arabidopsis thaliana" Science 301: 653-657

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

Averyhart-Fullard, V., Datta, K., and Marcus, A. (1988). A hydroxyroline-rich protein in the soybean cell wall. Proc. Natl. Acad. Sci. USA, 85: 1082-1085.

Bargmann, CI, (2001). Genome Biol. 2, Reviews-1005

Basavaraju P, Shetty NP, Shetty HS, de Neergaard E, Jørgensen HJ.(2009). Infection biology and defence responses in sorghum against Colletotrichum sublineolum. J Appl Microbiol. 2009 Aug;107(2):404-15.

Bauer, S, Vasu, P, Persson, S, Mort, AJ and Somerville CR (2006). Development and application of a suite of polysaccharide-degrading enzymes for analyzing plant cell walls. PNAS 103: 11417–11422

Benfey, PN, Chua, N-H (1990). The Cauliflower Mosaic Virus 35S Promoter: Combinatorial Regulation of Transcription in Plants. Science 250: 959-966

Benfey, PN, Ren, L., and Chua, N. H. (1990) Combinatorial and synergistic properties of CaMV 35S enhancer sub-domains. EMBO J. 9, 1685–1696.

Bergvinson, D., and García-Lara, S. (2004). Genetic approaches to reducing losses of stored grain to insects and diseases Current Opinion in Plant Biology 7: 480-485

Bert J. van der Zaal, Leon W. Neuteboom, Johan E. Pinas, Agnes N. Chardonnens, Henk Schat, Jos A.C. Verkleij, and Paul J.J. Hooykaas (1999). Over-expression of a Novel Arabidopsis Gene Related to Putative Zinc-Transporter Genes from Animals Can Lead to Enhanced Zinc Resistance and Accumulation. Plant Physiol. 119: 1047–1056

Blevins T, Rajeswaran R, Shivaprasad P, Beknazariants D, Si-Ammour A, Park H, Vazquez F, Robertson D, Meins F, Hohn T, Pooggin M (2006). Four plant Dicers mediate viral small RNA biogenesis and DNA virus induced silencing. Nucleic Acids Res 34: 6233–4246.

Bolwell GP (1993) Dynamic aspects of plant extracellular matrix. Int Rev Cytol 146: 261-324
Bonilla I, Mergold-Villase˜nor C, Campos M, S´anchez N, P´erez H, et al. 1997. The aberrant cell walls of boron deficient bean root nodules have no covalently bound hydroxyproline-proline-rich proteins. Plant Physiol. 115:1329–40
Boudart G, Dechamp-Guillaume G, Lafitte C, Ricart G, Barthe J-P, et al.,1995. Elicitors and suppressors of hydroxyproline-rich glycoprotein accumulation are solubilized from plant cell walls by endopolygalacturonase. Eur. J. Biochem. 232:449–57
Bown DP, Bolwell GP, Gatehouse JA (1993) Characterisation of potato (Solanum tuberosum L.) extensins: a novel extensin-like cDNA from dormant tubers. Gene 134:229–233
Bradley DJ, Kjellbom P, Lamb CJ (1992) Elictor- and wound- induced oxidative cross-linking of a proline-rich plant cell wall protein: a novel, rapid defense response. Cell 70: 21-30
Bradley DJ, Sadler IH, Fry SC. 1996. Diisodityrosine a novel tetrameric derivative of tyrosine in plant cell wall proteins: a new potential cross-link. Biochem. J.315:323–27
Brett, C. Waldron, K. (1996). Physiology and biochemistry of plant cell walls. Second Edition, Published by Chapman and Hall, London, UK
Brisson LF, Tenhaken R, Lamb CJ. 1994. Function of oxidative cross-linking of cell wall structural protein in plant disease resistance. Plant Cell 6:1703–12
Brownleader MD, Dey PM (1993). Purification of extensin from cell walls of tomato hybrid of Lycopersicon esculentum and L. peruviunum) cells in suspension culture. Planta 191: 457469
Burgert, I. (2006), Exploring the micromechanical design of plant cell walls. American Journal of Botany. 2006; 93:1391-1401
Cannon, M.C., Terneus, K, Hall, Q., Tan, L., Wang, Y. Wegenhart, B.L. Chen, L. Lamport, T.A., Chen, Y. and Kieliszewski, M.J. (2008). Self-assembly of the plant cell wall requires an extensin scaffold. PANS, 105, 2226–2231
Carpita NC, Gibeaut DM. Structural models of primary cell walls in flowering plants: Consistency of molecular structure with the physical properties of the walls during growth. Plant J. 1993; 3: 1–30. (Extensin-pectin model)
Cassab, G. I . 1986. Arabinogalactan proteins during the development of soybean root nodules. Planta 168 :441-46
Cassab, GI (1998). Plant cell wall proteins. Annual review of plant physiology and plant molecular biology. 49: 281-309
Cassab, GI, Nieto-Sotelo, J, Cooper, JB, Van Holst, G, Varner, JE. (1985). A developmentally regulated hydroxyproline-rich glycoprotein from the cell walls of soybean seed coats. Plant Physiol. 77:532-35
Cassab GI, Varner JE. 1987. Immunolocalization of extensin in developing soybean seed coats by immunogold-silver staining and tissue printing on nitrocellulose paper. J. Cell Biol. 105: 2581–88
Cassab GI, Varner JE. 1988. Cell wall proteins. Annu. Rev. Plant Physiol. Plant Mol. Biol. 39:321–53
Castonguay Y, Nadeau P, Laberge S. 1993. Freezing tolerance and alteration of translatable mRNAs in alfalfa (Medicago sativa L.) hardened at subzero temperatures. Plant Cell Physiol. 34:31–38
Catt, JW, Hills, GJ, Roberts, K. (1976). A structural glycoprotein, containing hydroxyproline, isolated from the cell wall of Chl amydomon as reinh ardtii. Pl ant a 131:165-71

Chen AP, Zhong NQ, Qu ZL, Wang F, Liu N, Xia GX.(2007). Root and vascular tissue-specific expression of glycine-rich protein AtGRP9 and its interaction with AtCAD5, a cinnamyl alcohol dehydrogenase, in Arabidopsis thaliana Plant Res. 120 :337-43.

Chen J, and Varner J.E. (1985a). Isolation and characterization of cDNA clones for carrot extensin and a proline-rich 33-kDa protein. Proc. Natl. Acad. Sci. USA 82, 4399-4403

Chen XY, Kim JY (2009). Callose synthesis in higher plants. Plant Signal Behav. 2009 Jun;4(6):489-92.

Cheung AY, Wang H, Wu HM (1995) A floral transmitting tissue-specific glycoprotein attracts pollen tubes and stimulates their growth. Cell 82: 383-393

Chivasa, S. Ndimba, B. K., Simon, W.J., Robertson, R. Yu X-L. , Knox, J.P., Bolwell, P., Slabas, A.R. (2002). Proteomic analysis of the Arabidopsis thaliana cell wall. Electrophoresis 23, 1754–1765

Clarke AE, Anderson RL, Stone BA. (1979). Form and function of arabinogalactans and arabinogalactan-proteins. Phytochemistry 18:521–40

Cleland R. (1967). Inhibition of cell elongation in Avena coleoptile by hydroxyproline. Plant Physiol. 42:271–274.

Cleland R. (1967). Inhibition of formation of protein-bound Hydroxyproline by free yydroxyproline in Avena coleoptiles. Plant Physiol. 42:1165–1170.

Cleland, R. (1968). Hydroxyproline formation and its relation to auxin-induced cell elongation in the Avena coleoptile. Plant Physiol. 43: 1625–1630.

Cleland R, and Karlsnes, A.M. (1967). A possible role of hydroxyproline-containing proteins in the cessation of cell elongation. Plant Physiol. 42:669–671.

Cooper, JB, Chen, J. A . , Van Holst, G. J, Varner, J. E. 1987. Hydroxyproline-rich glycoproteins of plant cell walls. Trends Biochem. Sci. 12:24-27

Cooper JB, Varner JE. 1983. Insolubilization of hydroxyproline-rich glycoprotein in aerated carrot root slices. Biochem. Biophys.Res. Commun. 112:161–67

Cooper, JB and Varner, J.E.(1984). Cross-Linking of Soluble Extensin in Isolated Cell Walls. Plant Physiology 76:414-417

Condit CM, McLean M, Meagher RB. 1990. Characterization of the expression of the petunia glycine-rich protein-1 gene product. Plant Physiol. 93:596–602

Corbin DR, Sauer N, Lamb CJ (1987) Differential regulation of a hydroxyproline-rich glycoprotein gene family in wounded and infected plants. Mol Cell Biol 7:4337–4344

Datta, K., Schmidt, A., and Marcus, A. (1989). Characterization of two soybean repetitive proline-rich proteins and a cognate cDNA from germinated axes. Plant Cell 1, 945-952.

Ding L, Zhu JK (1997). A role of arabinogalactan-proteins in root epidermal cell expansion. Planta 203: 289-294

Ecker JR, Davis RW (1987) Plant defense genes are regulated by ethylene. Proc Natl Acad Sci USA 84:5202–5206

Ellis, C. et al. (2002) The arabidopsis mutant cev1 links cell wall signaling to jasmonate and ethylene responses. Plant Cell 14, 1557-1566

Esau , K . 1965. Plant Anatomy. New York: Wiley. 767 pp

Esquerre-Tugaye M-T, Mazau D, Pelissier B, Roby D, Rumeau D, Toppan A. 1985. Induction by elicitors and ethylene of proteins associated to the defense of plants. In Cellular Biology of Plant Stress, ed. JL Key, T Kosuge, pp. 459–73. New York: Liss

Ferris PJ, Woessner JP, Waffenschmidt S, Kilz S, Drees J, Goodenough UW (2001). Glycosylated polyproline II rods with kinks as a structural motif in plant hydroxyproline-rich glycoproteins. Biochemistry. 40:2978-87
Fincher GB, Stone BA, Clarke AE (1983) Arabinogalactan-proteins: structure, biosynthesis, and function. Ann Rev Plant Physiol 34: 47-70
Fong, C., Kieliszewski, M.J., de Zacks, R., Leykam, J.F., and Lamport, D.T.A. (1992). A gymnosperm extensin contains the serinetetrahydroxyproline motif. Plant Physiol. 99, 548-552
Fire A, Xu S, Montgomery M, Kostas S, Driver S, Mello C (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature 391: 806–811
Fowler, T.J., Bernhardt, C., and Tierney, M.L. (1999). Characterization and expression of four Proline-Rich Cell Wall Protein Genes in Arabidopsis Encoding Two Distinct Subsets of Multiple Domain Proteins. Plant Physiol, December 1999, Vol. 121, pp. 1081-1091
Franssen HJ, Nap J-P, Gloudemans T, Stiekema W, van Dam H, et al. 1987. Characterization of cDNA for nodulin-75 of soybean: a gene product involved in early stages of root nodule development. Proc. Natl. Acad. Sci. USA 84:4495–99
Fry SC (1986). Cross-linking of matrix polymers in the growing cell walls of angiosperms. Annu. Rev. Plant Physiol. 37:165–86
Fry, SC (2004) Primary cell wall metabolism: tracking the careers of wall polymers in living plant cells. New Phytologist 161, 641-675
Fusaro AF, Sachetto-Martins G. (2007). Blooming time for plant glycine-rich proteins. 5, 386-387
Gao M, Kieliszewski MJ, Lamport DTA, Showalter AM (1999) Isolation, characterization and immunolocalization of a novel, modular tomato arabinogalactanprotein corresponding to the LeAGP-1 gene. Plant J 18: 43-55
García-HernándezI, E.L. and López, G.I.C. (2005) Structural cell wall proteins from five pollen species and their relationship with boron. Braz. J. Plant Physiol., 17:375-381, 2005
Gheysen, G., and C. Fenoll (2002). Gene expression in nematode feeding sites. Annu. Rev. Phytopathol. 2002. 40:191–219
Gilson P, Gaspar YM, Oxley D, Youl JJ, Bacic A. (2001). NaAGP4 is an arabinogalactan protein whose expression is suppressed by wounding and fungal infection in Nicotiana alata. Protoplasma. 215:128-39
Gleeson, PA, McNamara, M., Wettenhall, E.H., Stone, B.A., and Fincher, G.B. (1989). Characterization of the hydroxyproline-rich protein core of an arabinogalactan-protein secreted from suspension cultured Lolium multiflorum (Italian ryegrass) endosperm cells. Biochem. J. 264, 857-862.
Hall, Q. and Cannon, M.C. (2002) The cell wall hydroxyproline-rich glycoprotein RSH is essential for normal embryo development in Arabidopsis. Plant Cell, 14, 1161–1172.
Hannon, GJ (2002). RNA interference. Nature 418: 244 -251
Hayashi, T. (2006). The Science and lore of the plant cell wall: biosynthesis, structure and function (refer to different chapters)
Held, MA, Tan, L., Kamyab, A., Hare, M. Shpak, E., and Kieliszewski, M.J. (2004). Di-isodityrosine Is the Intermolecular Cross-link of Isodityrosine-rich Extensin Analogs Cross-linked in Vitro. J. BIOL. CHEM. 279: 55474–55482
Helliwell, C. and Waterhouse, P. (2003) Constructs and methods for high-throughput gene silencing in plants. Methods, 30, 289–295

Helliwell, CA, Wesley, S.V., Wielopolska, A.J. and Waterhouse, P.M. (2002) High-throughput vectors for efficient gene silencing in plants. Funct. Plant Biol. 29, 1217–1255
Hirsinger C, Parmentier Y, Durr A, Fleck J, Jamet E (1997) Characterization of a tobacco extensin gene and regulation of its gene family in healthy plants and under various stress conditions. Plant Mol Biol 33:279–289
Hirsinger C, Salval, Marbach J, Durr A, Fleck J, Jamet E (1999) The tobacco extensin gene Ext 1.4 is expressed in cells submitted to mechanical constraints and in cells proliferating under hormone control. J Exp Bot 50:343–355
Hong JC, Cheon YW, Nagao RT, Bahk JD, Cho MJ, Key JL. 1994. Isolation and characterization of three soybean extension cDNAs. Plant Physiol. 104:793–96
Hong JC, Nagao RT, Key JL. 1987. Characterization and sequence analysis of a developmentally regulated putative cell wall protein gene isolated from soybean. J. Biol. Chem. 262:8367–76
Hong, JC, Nago, R. T. and KEY, J. L. (1989). Developmental regulated expression of soybean proline-rich cell wall protein genes. The Plant Cell 1, 937-943.
Hong, JC., R.T. Nagao and J.L. Key (1990). Characterization of a proline-rich cell wall protein gene family of soybean. A comparative analysis. J. Biol. Chem. 265 (1990), pp. 2470–2475.
Howland, J.L. (2000). The surprising Archaea: Discovering another domain of Life. Oxford: Oxford Univ. Press. pp. 69–71.
Huckelhoven, R. (2007) Cell Wall–Associated Mechanisms of Disease Resistance and SusceptibilityPhysiology and Biochemistry Annu. Rev. Phytopathol. 2007. 45:101–27
Ishii, T., and Matsunagab, T. (2001). Pectic polysaccharide rhamnogalacturonan II is covalently linked to homogalacturonan. Phytochemistry 57: 969-974
Ito M, Kodama H, Komamine A, Watanabe A. (1998). Expression of extensin genes is dependent on the stage of the cell cycle and cell proliferation in suspension-cultured Catharanthus roseus cells. Plant Molecular Biology 36, 343–351.
Jackson RG, Kowalczyk M, Li Y, Higgins G, Ross J, Sandberg G, Bowles DJ (2002). Over-expression of an Arabidopsis gene encoding a glucosyltransferase of indole-3-acetic acid: phenotypic characterisation of transgenic lines. Plant J. 2002 Nov;32(4):573-83
Jamet, E., Canut, H, Boudart, G., and Pont-Lezica, R. F. (2006). Cell wall proteins: a new insight through proteomics. Trends in Plant Science 11: 33-39
Jamet, E., Albenne, C., Boudart, C., Irshad, M., Canut, H., and Pont-Lezica, R. (2008) Recent advances in plant cell wall proteomics. Proteomics 8, 893–908
Jander, G and Barth, C (2007). Tandem gene arrays: a challenge for functional genomics. Trends in Plant Science, 12: 203-210
Johnson et al (2003) Non-enzymic cell wall (glyco)proteins. In The Plant Cell Wall (Rose JKC ed) Blackwell, pp. 111-154.
Jones, L. (2002). Revealing micro-RNAs in plants. Trends in Plant Science. 7:473-475
Jose-Estanyol, M. and P. Puigdomènech (1993). Structure and expression of genes coding for structural proteins of the plant cell wall. New Phytol. 125:259–282.
Jose-Estanyol, M. and P. Puigdomènech (2000). Plant cell wall glycoproteins and their genes. Plant Physiol. Biochem. 38:97-108.
Jung, U-J, Choi, C-S, Park, J-H, Kang, H-W, Choi, J-E, Lee, S-Y, and Kang, KK. (2008). Over-expression of the pineapple fruit bromelain gene (BAA) in transgenic Chinese cabbage (Brassica rapa) results in enhanced resistance to bacterial soft rot. Electron. J. Biotechnol. 11 n.1

Kamath, RS, Martinez-Campos, M, Zipperlen, P, Fraser, AG, and Ahringer, J. (2001). Genome Biol. 2. Research - 0002

Keegstra K, Talmadge KW, Bauer WD, Albersheim P. 1973. The structure of plant Cell walls. III.Amodel of the walls of suspension cultured sycamore cells based on the interconnections of the macromolecular components. Plant Physiol. 51:188–96

Keller, B. (1993) Structural cell wall proteins. Plant Physiol. (1993) 101: 1127-1 130

Keller B, Lamb CJ (1989) Specific expression of a novel cell wall hydroxyproline-rich glycoprotein gene in lateral root initiation. Genes Devel 3:1639–1646

Keller, B., Templeton, M.D., and Lamb, C.J. (1989) Specific localization of a plant cell wall glycine-rich protein in protoxylem cells of the vascular system. Proc. Natl. Acad. Sci. USA 86, 1529-1533

Kevei Z, Vinardell JM, Kiss GB, Kondorosi A, Kondorosi E. (2002). Glycine-rich proteins encoded by a nodule-specific gene family are implicated in different stages of symbiotic nodule development in Medicago spp. Mol Plant Microbe Interact. 15:922-31.

Kidner, CA and Martienssen, RA (2003). Macro effects of MicroRNAs in plants. TRENDS in Genetics. 19: 13-16

Khashimova, Z.S. (2003). Hydroxproline containing proteins. Chem. Natural Compounds, 39, 229- 236

Kieliszewski MJ, Lamport DTA. 1994. Extensin: repetitive motifs functional sites post-translational codes and phylogeny. Plant J. 5:157–72

Kieliszewski, MJ, and Shpak, E. (2001). Synthetic genes for the elucidation of glycosylation codes for arabinogalactan-proteins and other hydroxyproline-rich glycoproteins. Cell. Mol. Life Sci. 58 1386–1398.

Kjellbom P, Snogerup L, Stohr C, Reuzeau C, McCabe PF, Pennell RI (1997) Oxidative cross-linking of plasma membrane arabinogalactan proteins. Plant J 12: 1189-1196

Kleis-San Francisco, SM, and Tierney, M.L. (1990). Isolation and characterization of a proline-rich cell wall protein from soybean seedling. Plant Physiol. 94, 1897-1902

Krysan, P.J., Jeffery C. Young, and Michael R. Sussman (1999). T-DNA as an Insertional Mutagen in Arabidopsis. The Plant Cell, 11: 2283–2290

Kuster H, Schr¨oeder G, Fr¨uhling M, Pich U, Rieping M, et al. 1995. The nodulespecific VfENOD-GRP3 gene encoding a glycine-rich early nodulin is located on chromosome I of Vicia faba L. and is predominantly expressed in the interzone II-III of root nodules. Plant Mol. Biol. 28:405–21

Lambert, KN (1995). Isolation of Genes Induced Early in the Resistance Response to *Meloidogyne javanica* in Lvcopersicon esculentum. PhD Dissertation (Davis: University of California-Davis).

Lamport, DTA. (1965). The protein component of primary cell walls. Adv. Bot. Res. 2: 15 -18

Lamport, DTA. (1969). The isolation and partial characterization of hydroxyproline-rich glycopeptides obtained by enzymatic degradation of primary cell walls. Biochemistry 8: 1155-63

Lamport, DTA. (1980). Structure and function of plant glycoproteins. In The Biochemistry of Plants, ed. P. K. Stumpf, E. E . Conn, 3:501-36. New York: Academic. 639 pp

Lamport, DTA, Katona, L, Roering, S. (1973). Galactosylserine in extensin. Biochem. J. 133: 125-31

Lamport DTA, Kieliszewski MJ, Showalter AM (2006). Salt stress upregulates periplasmic arabinogalactan proteins: using salt stress to analyse AGP function. New Phytol 169:479–492

Langan KJ, Nothnagel EA (1997) Cell surface arabinogalactan-proteins and theirrelation to cell proliferation and viability. Protoplasma 196: 87-98

Lei M. and Wu R. (1991) A novel glycine-rich cell wall protein gene in rice. Plant Mol. Biol. 16: 187–198

Lindstrom JT, Vodkin LO. 1991. A soybean cell wall protein is affected by seed color genotype. Plant Cell 3:561–71

Lloyd, A. (2003) Vector Construction for gene Over-expression as a tool to elucidate gene function. Methods in Molecular Biology, vol. 236: Plant Functional Genomics: Methods and Protocols, Edited by: E. Grotewold © Humana Press, Inc., Totowa, NJ

MacDougall, AJ, Brett, G.M.,. Morris, VJ, Rigby, N.M., Ridout, M.J., Ring, S.G. (2001). The effect of peptide–pectin interactions on the gelation behaviour of a plant cell wall pectin. Carbohydrate Research 335:115–126

Marcus A, Greenberg J, Averyhart-Fullard V. 1991. Repetitive proline-rich proteins in the extracellular matrix of the plant cell. Physiol. Plant. 81:273–79

Martienssen, R. and Irish, V. (1999) Copying out our ABCs, the role of gene redundancy in interpreting genetic hierarchies. Trends Genet. 15, 435–437.

Matsubayashi Y (2003). Ligand-receptor pairs in plant peptide signaling. Cell Sci.116: 3863-70.

Matsubayashi, Y.and Sakagami, Y (2006). Peptide Hormones in Plants. Annual Review of Plant Biol. 57:649-674

Matsui M, Toyosawa I, Fukuda M. 1995.Purification and characterization of a glycine-rich protein from the aleurone layer of soybean seeds. Biosci. Biotechnol. Biochem. 59:2231–34

Mazau D, Rumeau D, Esquerr´e-Tugaye, M-T. 1987. Molecular approaches to understanding cell surface interactions between plants and fungal pathogens. Plant Physiol. Biochem. 25:337–43

McDougall GJ, Stewart D, Morrison IM. 1996. Tyrosine residues enhance cross-linking of synthetic proteins into lignin-ike dehydrogenation products. Phytochemistry 41:43–47

McNeil, M., Darvill, A.G, Fry, S.C and Albersheim, P. (1984). Structure and function of the primary cell walls of plants. Annu. Rev. Biochem. 53: 625–663.

Mellon JE, Helgeson JP. 1982. Interaction of a hydroxyproline-rich glycoprotein from tobacco callus with potential pathogens. Plant Physiol. 70:401–5

Memelink J, Swords KMM, de Kam RJ, Schilperoort RA, Hoge JHC, Staehelin LA (1993) Structure and regulation of tobacco extensin. Plant J 4:1011–1022

Mergold-Villasenor, CA, Castrej´on L, Cassab GI. 1996. The role of turgor pressure in the cross-linking of plant cell wall proteins. Plant Physiol. 111:670 (Abstr

Merkouropoulos, G., Barnett, D.C. and Shirsat, A.H. (1999) The Arabidopsis extensin gene is developmentally regulated, is induced by wounding, methyl jasmonate, abscisic and salicylic acid, and codes for a protein with unusual motifs. Planta, 208, 212–219.

Merkouropoulos, G. and Shirsat, A.H. (2003) The unusual Arabidopsis extensin gene EXT1 is expressed throughout plant development and is induced by a variety of biotic and abiotic stresses. Planta, 217, 356–366.

Molina, A., Mena, M., Carbonero, P., and Garcıa-Olmedo, F. (1997). Differential expression of pathogen-responsive genes encoding two types of glycine-rich proteins in barley. Plant Molecular Biology 33: 803–810

Mousavi A, Hotta Y, 2005. Glycine-rich proteins: a class of novel proteins. Appl Biochem Biotechnol. 2005 Mar;120(3):169-74.

Narayanasamy, P. (2006). Immunology in plant health and its impact on food safety. Products Press,. An Imprint of the Haworth Press, New York – London; springerlink.com
Narváez-Vásquez J, Pearce G, abd Ryan CA (2005). The plant cell wall matrix harbors a precursor of defense signaling peptides. Proc Natl Acad Sci. 102:12974-12977
Nicholes, CD, Lindstrom, J.T. and Vodkin, L.O. (1993) Variation of proline rich cell wall proteins in soybean lines with anthocyanin mutations. Plant Mol. Biol. 21,145-156.
Niebel A, Engler JA, Tire´ C, Engler, G, Van Montagu, M, Gheysen, G (1993) Induction patterns of an extensin gene in tobacco upon nematode infection. Plant Cell 5:1697–1710
Nik Marzuki, S. and Shirsat, A.H. 2005. A Brassica napus extensin gene is induced by abscisic acid, methyl jasmonate and salicylic acid. Malays. J. Biochem. Mol. Biol. 11: 1-5.
Nothnagel EA (1997) Proteoglycans and related components in plant cells. Int Rev Cytol174: 195-291
Nuñez, A., Fishman, M.L., Fortis, L.L., Cooke, PH, Hotchkiss, AT, Jr.(2009). Identification of extensin protein associated with sugar beet pectin. J Agric Food Chem. 57:10951-10958
O'Brian TP. 1981. The primary xylem. In Xylem Cell Development, ed. JR Barnett, pp. 14–46. Turnbridge Wells, Kent, UK: Castle House Publ.
Oliveira DD, de Franco L, Simoens C, Seurinck J, Coppitiers J, et al. (1993). Inflorescence-specific genes from Arabidopsis thaliana encoding glycine-rich proteins. Plant J. 3:495–507
Otte, O., and Barz, W.(2000) Characterization and oxidative in vitro cross-linking of an extensin-like protein and a proline-rich protein purified from chickpea cell walls. Phytochem. 53, 1-5
Park JH, Suh MC, Kim TH, Kim MC, Cho SH (2008). Expression of glycine-rich protein genes, AtGRP5 and AtGRP23, induced by the cutin monomer 16-hydroxypalmitic acid in Arabidopsis thaliana. Plant Physiol Biochem 46, 015-1018
Park MH, Suzuki Y, Chono M, Knox JP, Yamaguchi I (2003) CsAGP1, a gibberellin responsive gene from cucumber hypocotyls, encodes a classical arabinogalactan protein and is involved in stem elongation. Plant Physiol 131: 1450-1459
Parmentier Y, Durr A, Marbach J, Hirsinger C, Criqui M-C, Fleck J, Jamet E (1995) A wound-inducible extensin gene is expressed early in newly isolated protoplasts of Nicotiana sylvestris. Plant Mol Biol 29:279–292
Parsons BL and Mattoo AK. (1994) A wound-repressible glycine-rich protein transcript is enriched in vascular bundles of tomato fruit and stem. Plant Cell Physiol. 35: 27–35
Pearce, G, Moura, D., Stratmann, J. and Ryan, C A. (2001) Production of multiple plant hormones from a single polyprotein precursor. Nature 411, 817-820.
Peng J, Lai L, Wang X (2008). A cell wall proline-rich protein containning GASA domain in Gerbera hybrida. Sci China C Life Sci. 51:520-5. PRGL
Qi XY, Beherens BX, West PR, Mort AJ. (1995). Solubilization and partial characterization of extensin fragments from cell walls of cotton suspension cultures. Plant Physiol. 108:1691–701
Reymond P, Farmer EE (1998). Jasmonate and salicylate as global signals for defense gene expression. Curr Opin Plant Biol. 1998 Oct;1(5):404-11.
Ribeiro, JM., Silva Pereira, C., Soares, N.C., Vieira, AM., Feijó, J.A. and Jackson, P.A.(2006). The contribution of extensin network formation to rapid, hydrogen

peroxide-mediated increases in grapevine callus wall resistance to fungal lytic enzymes. Journal of Experimental Botany, 57, 2025-2035
Ridge I, Osborne DJ. 1970. Hydroxyproline and peroxidases in cell walls of Pisum sativum: regulation by ethylene. Journal of Experimental Botany 21, 843–856.
Ringli C, Keller B, Ryser U (2001). Glycine-rich proteins as structural components of plant cell walls Cell Mol Life Sci. 58:1430-41.
Roberts, K. (1994). The plant extracellular matrix: in a new expansive mood. Current Opinions Cell Biology 6: 688–694
Roberts, K. (1989) The plant extracellular matrix. Curr. Opin. Cell Biol. 1, 1020-1027
Roberts, K. (1990) Structures at the plant cell surface. Curr. Opin. Cell Biol. 2, 920-928
Robert, K.(2007) Hand book of plant Science Vol 2; John Wiley and Sons Ltd, England)
Roberts, K and Shirsat, A.H. (2006) Increased extensin levels in Arabidopsis affect inflorescence stem thickening and height J. of Expt. Botany 57:537-545
Rosegrant MW, Paisner MS, Meijer S, Witcover J (2001). Projections of global food supply and demand and child malnutrition. In Global Food Projections to 2020: Emerging Trends and Alternative Futures. Washington, DC: IFPRI; 58-81.
Rose, J.K.C. (2003). The plant cell wall. Annual Plant Review, 8:122
Ryan, C A. and Pearce, G. (2004). in Plant Hormones: Biosynthesis, Signal Transduction, Action, ed. Davies, P. J. (Springer, Secaucus, NJ) pp. 654-670.
Ryser, U, and Keller, B. (1992). Ultrastructural localization of a bean glycine-rich protein in unlignified primary walls of protoxylem cells. Plant Cell 4, 773-783.
Ryser U., Schorderet M., Zhao G.-F., Studer D., Ruel K., Hauf G. et al. (1997) Structural cell-wall proteins in protoxylem development: evidence for a repair process mediated by a glycinerich protein. Plant J. 12: 97–111
Ryser U, Schorderet M, Guyot R, Keller B. (2004). A new structural element containing glycine-rich proteins and rhamnogalacturonan I in the protoxylem of seed plants. J Cell Sci. 117:1179-90.
Sadava, D. and M.J. Chrispeels (1973) Hydroxyproline-rich cell wall proteins (extensin): Role in the cessation of elongation in excised pea epicotyls. Devel. Biol. 30: 49-55.
Sadava, D, F. Walker and M.J. Chrispeels (1973) Hydroxyproline-rich cell wall proteins (extensin): Biosynthesis and accumulation in growing pea epicotyls. Devel. Biol. 30: 42-48.
Salts, Y, Wachs, R, Gruissem, W, and Barg, R (1991). Sequence coding for a novel proline-rich protein preferentially expressed in young tomato fruit. Plant. MOI. Biol. 17, 149-150.
Sandhua, APS., Randhawab, G.S. and Dhuggaa, K.S. (2009) Plant Cell Wall Matrix Polysaccharide Biosynthesis. Molecular Plant 2:840-850
Santino CG, Stanford GL, Connor TW (1997) Developmental and transgenic analysis of two tomato fruit enhanced genes. Plant Mol Biol 33: 405-416
Sardar, HS, Yang, J., Showalter, A.M. (2006). Molecular Interactions of Arabinogalactan-proteins (AGPs) with Cortical Microtubules and F-actin in Bright Yellow-2 (BY-2) Tobacco Cultured Cells. Plant Physiology Preview. 10.1104/pp.106.088716
Schindler T, Bergfeld R, Schopfer P (1995) Arabinogalactan proteins in maize coleoptiles: developmental relationship to cell death during xylem differentiation but not to extension growth. Plant J 7: 25-36
Schindler TM. 1998. The new view of the primary cell wall. Zeitschrift für Pflanzenernährung und Bodenkunde 161, 499–508.
Serpe MD, Nothnagel EA (1994) Effects of Yariv phenylglycosides on Rosa cellsuspensions: evidence for the involvement of arabinogalactan-proteins in cell proliferation. Planta 193: 542-550

Sessions, A, Burkea, E, Prestinga, G, Auxb, G, McElverb, J, Pattonb, D, Dietrichb, B, Hoa, P, Bacwadena, J, Koa, C, Clarkea, CD, Cottona, D, Bullisa, D, Snella, J, Miguela, T, Hutchisona, D, Kimmerly, B, Mitzelc, T, Katagiria, F, Glazebrooka, J, Lawb, M and Goffa, SA. (2002). A High-Throughput Arabidopsis Reverse Genetics System. The Plant Cell, 14: 2985-2994.
Sheng J, D'Ovidio R, Mehdy MC (1991) Negative and positive regulation of a novel proline-rich protein mRNA by funga1 elicitor and wounding. Plant J 1: 345-354
Shirsat AH, Wieczorek D, Kozbial P (1996b) A gene for Brassica napus extensin is differentially expressed on wounding. Plant Mol Biol 30:1291–1300
Showalter AM. 1993. Structure and function of plant cell wall proteins. Plant Cell 5:9–23
Showalter AM (2001) Arabinogalactan proteins: structure, expression and function. Cell Mol Life Sci 58: 1399-1417
Showalter, A.M., Butt, A.D. and Kim, S. (1992). Molecular details of tomato extensin and glycine-rich protein gene expression Plant Molecular Biology 19: 205-215
Showalter, A.M., and Varner, J.E. (1989). Plant hydroxyproline-rich glycoproteins. In The Biochemistry of Plants, Vol. 15, P.K. Stumpfand E.E. Conn, eds (New York: Academic Press), pp. 485-520.
Smith JJ, Muldoon EP, Willard JJ, Lamport DTA (1986). Tomato extensin precursors P1 and P2 are highly periodic structures. Phytochemistry 25: 1021-1030
Somerville, C. et al. (2004) Toward a systems approach to understanding plant cell walls.Science 306, 2206-2211
Sommer-Knudsen, J., A. Bacic and A.E. Clarke, Hydroxyproline-rich plant glycoproteins. Phytochemistry 47 (1998), pp. 483–497.
Stafstrom JP, Staehlin LA. 1986. Crosslinking patterns in salt-extractable extension from carrot cell walls. Plant Physiol. 81:234–41
Stram Y, Kuzntzova L (2006). Inhibition of viruses by RNA interference. Virus Genes 32 (3): 299–306.
Sturm A. (1992) A wound-inducible glycine-rich protein from Daucus carota with homology to single-stranded nucleic acidbinding proteins. Plant Physiol. 99: 1689–1692
Suzuki H, Fowler TJ, Tierney ML (1993) Deletion analysis and localization of SbPRP1, a soybean cell wall protein gene, in roots of transgenic tobacco and cowpea. Plant Mol Biol 21: 109-119
Tagu D, Walker N, Ruiz-Avila L, Burgess S, Martı´nez-Izquierdo JA, Leguay JJ, Netter P, Puigdome`nech P (1992) Regulation of the maize HRGP gene expression by ethylene and wounding. mRNA accumulation and qualitative expression analysis of the promoter by microprojectile bombardment. Plant Mol Biol 20:529–538
Tao TY, Ouellet T, Dadej K, Miller SS, Johnson DA, Singh J. (2006). Characterization of a novel glycine-rich protein from the cell wall of maize silk tissues. Plant Cell Rep. 25:848-58
Tang X, Xie M, Kim YJ, Zhou J, Klessig DF, Martin GB (1999) Overexpression of Pto activates defense responses and confers broad resistance. Plant Cell 11: 15-29
Tantikanjana, T, Mikkelsen, MD, Hussain, M, Halkier, BA, and Sundaresan, V (2004). Functional Analysis of the Tandem-Duplicated P450 Genes SPS/BUS/CYP79F1 and CYP79F2 in Glucosinolate Biosynthesis and Plant Development by Ds Transposition-Generated Double Mutants. Plant Physiology, 135: 840–848
Thompson HJM, Knox JP (1998) Stage-specific responses of embryogenesis carrot cell suspension cultures to arabinogalactan protein-binding β-glucosyl Yariv reagent. Planta 205: 32-38

Tierney, M.L., Wiechert, J., and Pluymers, D. (1988). Analysis of the expression of extensin and p33-related cell wall proteins in carrot and soybean. MOI. Gen. Genet. 211, 393-399

Tire´ C, De Rycke R, De Loose M, Inze´ D, Van Montagu M, Engler G (1994) Extensin gene expression is induced by mechanical stimuli leading to local cell wall strengthening in Nicotiana plumbaginifolia. Planta 195:175–181

Treuini GF, Horrichs A, Somisich IE (l993): Isolation of putative defense related genes from Arabidopsis thaliana and expression in fungal elicitor-treated cells. Plant MO/ Biol, 21:385-389.

Trezzini, G.F., Horrichs, A. and Somssich, I.E. (1993) Isolation of putative defense-related genes from Arabidopsis thaliana and expression in fungal elicitor-treated cells Plant Molecular Biology 21: 385-389

Vance, V and Vaucheret, H. (2001). RNA silencing in plants – defense and counter-defense. Science 292: 2277-2280

van der Eycken, W, de Almeida Engler, J, Inzé, D, van Montagu, M. and Gheysen, G. (1996). A molecular study of root-knot nematode-induced feeding sites. Plant J. 9: 45–54.

van deWiel C, Scheres B, Franssen H, van Lierop M-J, van Lammeren A, et al. (1990). The early nodulin transcript ENOD2 is located in the nodule parenchyma (inner cortex) of pea and soybean root nodules. EMBO J. 9:1–7

van Hangel AJ, Tadesse Z, Immerzeel P, Schols H, Van Kammen A, de Vries SC (2001) N-acetlyglucosamine and glucosamine-containing arabinogalactan proteins control somatic embryogenesis. Plant Physiol 125: 1880-1890

van Hengel AJ, Roberts K (2003) AtAGP30, an arabinogalactan-protein in the cell walls of the primary root, plays a role in root regeneration and seed germination. Plant J 36: 256-270

van Holst G-J, Varner JE. 1984. Reinforces polyproline II conformation in a hydroxyproline-rich cell wall glycoprotein from carrot root. Plant Physiol.74:247–51

Varner JE. 1994. Cell wall structural proteins - can we deduce function from sequence? Biochem. Soc. Symp. 60:1–4

Varner JE, and Lin LS (1989). Plant cell wall architecture. Cell. 56:231–239

Vision, T. J., Brown, D. G., and Tanksley, S. D. (2000) The origins of genomic duplications in Arabidopsis. Science 15, 2114–2117

Vogel, J.P. et al. (2004) Mutations in PMR5 result in powdery mildew resistance and altered cell wall composition. Plant J. 40, 968-978

Waffenschmidt, S, Woessner, JP, Beer, K, and Goodenough UW. 1993. Isodityrosine cross-linking mediates insolubilization of cell walls in Chlamydomonas. Plant Cell. 5: 809–820.

Watson, B.S. and Sumner, L.W. (2006). Isolation of Cell Wall Proteins from Medicago sativa Stems. In Plant Proteomics: Methods and Protocols. Series.Methods in Molecular Biology Vol.: 355 pages 79-92; The Samuel Roberts Noble Foundation, Ardmore, Oklahoma

Wei, G. and Shirsat, A.H., (2006). Extensin over-expression in Arabidopsis limits pathogen invasiveness. Molecular Plant Pathology. 6: 579–592

Wesley, S.V., Helliwell, C.A., Smith, N.A., Wang, M.-B., Rouse, D.T., Liu, Q., Gooding, P.S., Singh, S.P., Abbott, D., Stoutjesdijk, P.A., Robinson, S.P., Gleave, A.P., Green, A.G. and Waterhouse, P.M. (2001) Construct design for efficient, effective and high-throughput gene silencing in plants. Plant J. 27, 581–590.

Wesley, SV, Christopher A, Helliwell C, Smith NA. (2001). Construct design for efficient, effective and high-throughput gene silencing in plants. Plant J. 27: 581-590

Wielopolska, A., Townley, H., Moore, I., Waterhouse, P and Helliwell, (2005). A high-throughput inducible RNAi vector for plants. Plant Biotechnology Journal. 3: 583–590

Williamson VM, Hussey RS. 1996. Nematode pathogenesis and resistance in plants. Plant Cell 8:1735–45

Wilson, F. Long, E.M. Maruoka and J.B. Cooper, A (1994) New proline-rich early nodulin from Medicago truncatula is highly expressed in nodule meristematic cells. Plant Cell 6:1265–1275.

Wilson, LG, Fry C (1986). Extensin-a major cell wall glycoprotein. Plant Cell Environ. 9:239-60

Wu H, de Graaf B, Mariani C, Cheung AY. (2001). Hydroxyproline-rich glycoproteins in plant reproductive tissues: structure, functions and regulation. Cell Mol Life Sci. 58:1418-29.

Wyatt RE, Nagao RT, Key JL (1992) Patterns of soybean proline-rich protein gene expression. Plant Cell 4: 99-110

Wycoff KL, Powell PA, Gonzales RA, Corbin DR, Lamb C, Dixon RA (1995) Stress activation of a bean hydroxyprolinerich glycoprotein promoter is superimposed on a pattern of tissue-specific developmental expression. Plant Physiol 109:41–52

Xu J, Tan L, Lamport D.T, Showalter A.M, Kieliszewski M.J 2008. The O-Hyp glycosylation code in tobacco and Arabidopsis and a proposed role of Hyp-glycans in secretion. Phytochemistry. 2008 May;69:1631-40.

Yang J. and A. M. Showalter (2007) Expression and localization of AtAGP18, a lysine-rich arabinogalactan-protein in Arabidopsis. Planta, 226, 169-179

Ye Z-H, Song Y-R, Marcus A, Varner JE. 1991. Comparative localization of three classes of cell wall proteins. Plant J. 1:175–83

Ye Z-H, Varner JE. 1991. Tissue-specific expression of cell wall proteins in developing soybean tissues. Plant Cell 3:23–37

Zhang, X., Ren, Y., and Zhao,J. (2008). Roles of extensins in cotyledon primordium formation and shoot apical meristem activity in Nicotiana tabacum. J. Expt. Bot. 59:4045-4058.

CHAPTER 2

IN SILICO ANALYSIS OF *PELPK1* (*AT5G09530)*

CHAPTER 2

IN SILICO ANALYSIS OF *PELPK1* (AT5G09530)

1. INTRODUCTION

When the *Arabidopsis* genome sequence was first published, the majority of its predicted genes (~70%) were assigned to categories based on homology to genes of known function, while the remaining 30% of genes lacked significant sequence similarity and have remained largely unclassified (AGI 2000; Gutiérrez et al. 2004). Based on protein alignments, *Arabidopsis* gene *At5g09530* has been variously annotated as either an extensin-like protein (TAIR database prior to 2010), HRGP (hydroxyproline-rich glycoprotein) family protein (TAIR current annotation), an HRGP family member containing Pro-rich extensin domains (NCBI REFSEQ: NP_196515), or a periaxin-like protein (NCBI accession: AAK96839). Furthermore, a recently developed algorithm for the classification of cell wall proteins (Showalter et al., 2010) has designated *At5g09530* as *PRP10*, a member of the PRP (proline-rich protein) family. An objective of this chapter is to evaluate these classifications and determine the most accurate annotation for *At5g09530* based on evidence from bioinformatics analyses. *At5g09530* is here named *PELPK1*, according to its unique sequence motif as described below.

2. ANALYSIS OF *PELPK1* (*At5g09530*) BIOINFORMATICS DATABASES

Computational analysis of PELPK1 primary amino acid sequence

The *At5g09530* coding region consists of a single exon that encodes a predicted 370aa protein with pI=5.9 and molecular weight of 41.6kDa (Figure 2-1). This gene model is supported by full-length cDNA clones (e.g. NM_120990, AAK96839) and many ESTs (indexed at www.arabidopsis.org).

The amino acid composition of At5g09530 is particularly rich in five amino acid residues: Pro (29%); charged amino acids Lys (16%) and Glu (16%); and branched-chain amino acids Leu (12%), Ile (6%), and Val (5%), (Table 2-1). Visual inspection of At5g09530 showed that these amino acids are frequently arranged in a repeated motif: Pro-Glu-(Leu/Ile/Val)-Pro-Lys (Figure 2-1). This motif is here named PELPK, using the single-letter codes for each of the amino acids, with the branched-chain amino acids (Leu/Ile/Val) represented as L in the motif name. An unbiased scan using SeqWord confirms that P-E-L-P-K (and its permutation K-P-E-L-P) is the most abundant 5-mer word in Atg509530 (data not shown; Ganesan et al., 2008). A sequence logo created from At5g09530 shows the frequency distribution of amino acids within the PELPK motif (Figure 2-2). The PELPK motif is also found as a component of all of the major repeats identified by automated repeat-detection programs: P-E-I-Q-K-P-E-L-P-K, P-E-I-P-K-E-L, P-E-I-P-K, P-E-L-P-K (http://www.ebi.ac.uk/Tools/Radar/index.html, Table 2-2).

The PELPK motif, as defined above, occurs 36 times in Atg509530 (Figure 2-1). Allowing for up to one mismatch at any position of the PELPK motif, there are 52 occurrences of the motif within At5g09530. Together the 52 PELPK motifs and their close variants encompass essentially the entire coding region of At5g09530 outside of the signal peptide (280/341 residues, 82%, Figure 2-1). Because of the abundance of the novel PELPK motif within *At5g09530*, we have named this gene *PELPK1*.

Hydropathy plots (e.g. SOSUI, http://bp.nuap.nagoya-u.ac.jp/sosui/sosui_submit.html) predict a short, highly hydrophobic region between residues 9 and 29, near the N-terminus of the PELPK1 protein (Figure 2-1). This is a potential transmembrane domain. The hydrophobicity index (GRAVY score) of the first 30 amino acids from the N-terminus of the PELPK1 protein is 1.60. In contrast, the remainder of the protein is on average highly hydrophilic, with a GRAVY score of -0.86. The GRAVY score of the entire PELPK1 protein is -0.66 (Table 2-2). The overall

>AT5G09530.1
MALMKKSLSAALLSSPLLIICLIALLADPFSVGARRLLEDPKPEIPKLPELPKFEVPKLPEFPKPELPKLPE
FPKPELPKIPEIPKPELPKVPEIPKPEETKLPDIPKLELPKFPEIPKPELPKMPEIPKPELPKVPEIQKPEL
PKMPEIPKPELPKFPEIPKPDLPKFPENSKPEVPKLMETEKPEAPKVPEIPKPELPKLPEVPKLEAPKVPEI
QKPELPKMPELPKMPEIQKPELPKLPEVPKLEAPKVPEIQKPELPKMPELPKMPEIQKPELPKMPEIQKPEL
PKVPEVPKPELPTVPEVPKSEAPKFPEIPKPELPKIPEVPKPELPKVPEITKPAVPEIPKPELPTMPQLPKL
PEFPKVPGTP*

Figure 2-1: Amino acid sequence of *PELPK1* (AT5G09530). Conserved PELPK motifs (i.e. P-E-(L/I/V)-P-K) are shown in blue, and variants of the PELPK motif are shown in magenta. The putative signal peptide is italicized in bold and the predicted transmembrane region is underlined.

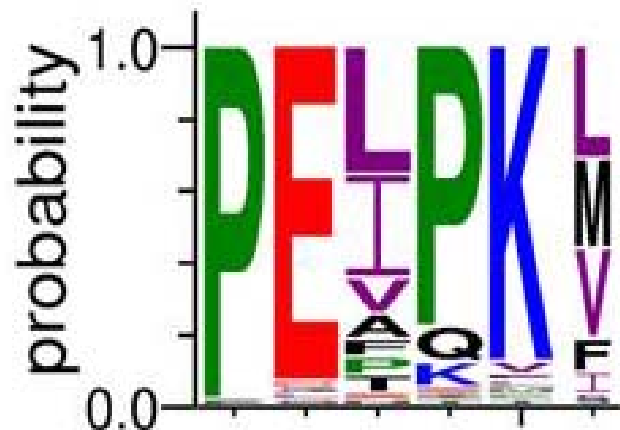


Figure 2-2. A PELPK1 sequence logo. The logo was created by dividing the 341 aa, repeat-rich region of the At5g09530 amino acid sequence into successive fragments of 5 aa or 6 aa (a gap was introduced as necessary at the end of 5 aa fragments to maintain the alignment of the conserved domain).

Table 2-1: Amino acid composition of the PELPK1 and other proteins and groups for comparison. Values shown are (%) composition for individual proteins (PELPK1, periaxin), or median values for all *Arabidopsis* extensins and Arabidopsis PRPs (family membership as defined by Showalter et al., 2010) and for proteins (PELPK-like) from various plant species that have been assigned by Phytozome to the same protein family (22878593) as PELPK1

	PELPK1	PELPK-like	extensins	PRPs	periaxin
Ala (A)	3	3.4	3.9	3.4	3.1
Cys (C)	0.3	0.4	1.0	2.2	0.8
Asp (D)	1.1	0.7	2.6	2.0	0.8
Glu (E)	15.7	12.6	2.3	2.1	12.3
Phe (F)	2.4	2.5	3.3	2.6	7.7
Gly (G)	0.5	0.7	3.9	3.3	1.5
His (H)	0	2.3	1.6	2.0	0
Ile (I)	6.2	3.4	3.2	5	4.6
Lys (K)	15.9	8.05	4.3	9.2	12.3
Leu (L)	12.2	14.2	6.3	7.4	14.6
Met (M)	3	2.3	1.1	0.7	3.8
Asn (N)	0.3	0.5	3.7	2.2	0
Pro (P)	28.6	27.5	24.2	25.4	23.8
Gln (Q)	1.9	1.0	2.2	1.9	0.8
Arg (R)	0.5	1.1	2.7	0.9	1.5
Ser (S)	1.9	4.5	12.7	5.4	5.4
Thr (T)	1.6	3.5	4.1	7.3	3.8
Val (V)	4.9	5.6	6.4	8.1	3.1
Trp (W)	0	0	0.3	0	0
Tyr (Y)	0	0.2	3.8	2.6	0

Table 2-2: Comparison of sequence characteristics of PELPK1 and other proteins and protein families. Values shown are for individual proteins (PELPK1, periaxin), or median and standard deviation for all *Arabidopsis* extensins and *Arabidopsis* PRPs (family membership as defined by Showalter et al., 2010, excluding *PELPK1* and *PELPK2*) and for proteins (PELPK-like) from various plant species that have been assigned by Phytozome to the same protein family

	PELPK1	PELPK-like	extensins	PRPs	periaxin
Proteins (n)	1	30	62	16	1
Length (aa)	370	195±74	474±253	317±141	865
MW (kDa)	41.6	20.9±8.3	52.6±28	34.2±15	-
Isoelectric point (pI)	5.9	5.7±1.6	8.4±2.0	9.9±2.0	-
Hydropathy (gravy score)	-0.66	-0.43±0.2	-0.54±0.3	-0.29±0.4	-0.02
PE(L|I|V)PK	34	8.6	0	0	0
(PP(V|Y|H|E)(V|Y|H|E)K)	0	0	0.13	2.0	2
PP	0	2.9	24.7	27.5	0
SPPPP	0	0.03	16.3	0.13	0

hydrophilic nature of the PELPK1 protein is consistent with the enrichment of the PELPK domain in charged amino acids and Pro, despite the presence also of highly hydrophobic, branched chain amino acids.

Protein localization algorithms predict with high confidence that PELPK1 contains an N-terminal signal peptide (Figure 2-1). This putative signal peptide overlaps the hydrophobic, N-terminal domain described above. SignalP 3.0 predicts secretory pathway localization for PELPK1 (Emanuelsson et al., 2007; http://www.cbs.dtu.dk/services/SignalP/; score: 0.98). WoLF-POSORT (Horton et al., 2007; http://wolfpsort.org/) reports extracellular localization (score 5.0) as most likely, with localization to vacuole (4.0) or ER (2.0) also predicted, but with slightly lower confidence. Conversely, PSORT (http://psort.ims.u-tokyo.ac.jp/form.html) predicts vacuolar (0.90) localization, with slightly higher confidence than the extracellular

compartment (0.82) or ER (0.10). Although the vacuole and extracellular compartment are physically distinct, both are supplied in part by the secretory pathway, and it is possible for a single species of protein to be targeted to either compartment. Thus, the various protein localization predictions reported above are generally consistent and indicate that PELPK1 is most likely localized to either the cell wall or vacuole, or both.

Predictions of PELPK1 structure

The secondary structure of PELPK1 is predicted with very high confidence (9/9 psipred confidence score) to consist entirely of a random coil, with the exception of the N-terminal hydrophobic region, which likely forms a helix (not presented) (psipred, version 2.6, http://bioinf.cs.ucl.ac.uk/psipred; phyre, version 2.0 by Kelley and Sternberg, 2009; http://www.sbg.bio.ic.ac.uk/phyre). Other software programs likewise predict a lack of regular ordered structure within PELPK1 and classify it as an intrinsically disordered protein (IDP) (e.g. DisEMBL, Figure 2-3).

IDPs are active proteins that do not form regular 3-dimensional (3D) globular structures (Wright and Dyson, 1999 and Dunker et al., 2002). These differ from globular proteins in that they are rich in disorder-promoting amino acid residues, such as Glu, Pro, Lys, Ser, Gly, and Gln, and depleted in order-promoting amino acid residues, such as Trp, Tyr, Phe, Cys, Ile, Leu, and Val (Tompa 2002; Dunker et al., 2001). As reported in Table 2-1, PELPK1 is rich in disorder-promoting residues such as Pro, (29%), Lys (16%), and Glu (16%), and almost completely deficient in order-promoting residues such as Trp (0.0%), Tyr (0.0%), and Cys (0.3%).

Although there is no universally-accepted definition of protein disorder, according to thermodynamics, disorder in a polypeptide chain likely includes random coiling.

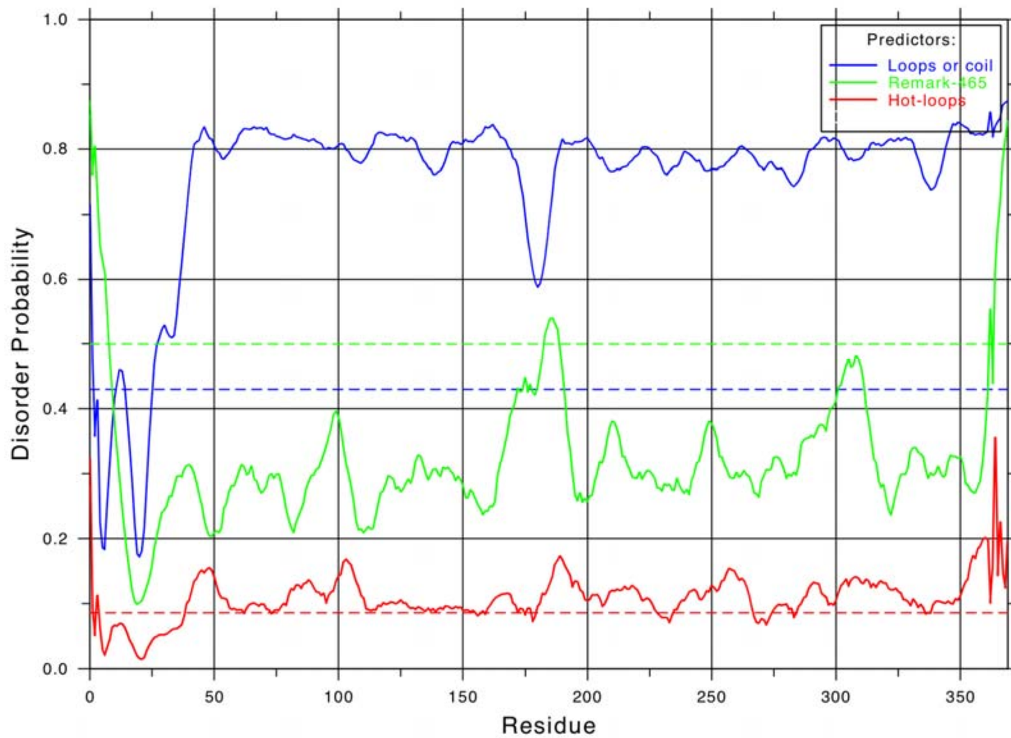


Figure 2-3. DisEMBL analysis of PELPK1 amino acid sequence. Three measures of disorder-related probabilities (trained on different predictors) are plotted here. Thresholds: Loops/Coils = 0.516, Remark 465 = 0.6, Hot loops = 0.1204. Loops/Coils refer to the probability of the structure forming a random coil. The Loops/Coils predictor is most generally relevant the probability of forming a random coil.

However, this theoretical state has not been observed even under extremely denaturing conditions. Therefore, it has been suggested that proteins in solution always tend to keep some form of residual structure (Shortle and Ackerman, 2001, Ackerman and Shortle, 2002, and Klein-Seetharaman et al., 2002). Once, it was thought that proteins with low complexity regions were structurally disordered. The strongest evidence came from the observation that low-complexity regions were only rarely detected in proteins with 3D structures (Saqi and Sternberg, 1994). However, a strong correlation between the two has not been established as regions of low sequence complexity were not found to be always disordered (Dunker et al., 2002). Although a number of experimental methodologies including NMR-, Raman-, and CD-spectroscopy and hydrodynamic measurements have been applied to indirectly determine protein disorder (Smyth et al., 2001 and Dunker et al., 2001), analysis of hydrophobicity appears to provide a better indication of IDPs. As the residues of IDPs are more exposed to an aqueous environment they are less hydrophobic than the residues in globular proteins, which are less exposed to aqueous solution (Dysen and Wright, 2005). The highly hydrophilic character PELPK1 described above is therefore consistent with this characteristic of IDPs.

Studies have demonstrated that protein misfolding can cause aggregation and thereby pathogenecity in humans (Schweers et al., 1994, Kaplan et al., 2003, and Bates, 2003). Some IDPs have been reported to be associated with human diseases (Uversky et al., 2009). In plants, however, IDPs have been reported to act as a chaperone against abiotic stresses (Kovacs et al., 2008). They are also reported to be involved in the regulation of signal transduction and gene expression (Tompa 2002; Dyson and Wright 2005).

Predictions of glycosylation

The PELPK1 protein sequence was analyzed by NetOGlyc 3.1 (http://www.cbs.dtu.dk/services/NetOGlyc/), and NetNGlyc 1.0 (http://www.cbs.dtu.dk/services/NetNGlyc/) to predict sites for O-linked, and N-linked glycosylations, respectively (Figure 2-4). The total number of predicted O-linked (Thr + Ser), and N-linked (Asn-Xaa-Ser/Thr) glycosylations above the threshold level (G-score >0.5), were 7, and 0 respectively (Table 2-2). These preliminary analyses therefore indicate that PELPK1 is unlikely to be N-glycosylated, and may have limited O-glycosylation.

PELPK2. *At5g09520/PELPK2* has also been labeled as *PRP*9 (Showalter et al., 2010). Compared to *PELPK1*, *PELPK2* is much smaller (130 aa, predicted 14.7kDa) and more basic (predicted pI = 8.5). No full length cDNAs have been described for *PELPK2*, and the limited number of published ESTs (e.g. AV548955) only partially overlap this coding region. *PELPK2* shares many of the same sequence motifs as *PELPK1* (Figure 2-5). Because *PELPK1* and *PELPK2* are highly similar and occupy adjacent loci, they likely evolved through tandem duplication of a common ancestor and can be considered paralogs. The differences in length of *PELPK1* and *PELPK2* can be explained by either internal duplications or deletions of the PELPK motif, subsequent to gene duplication.

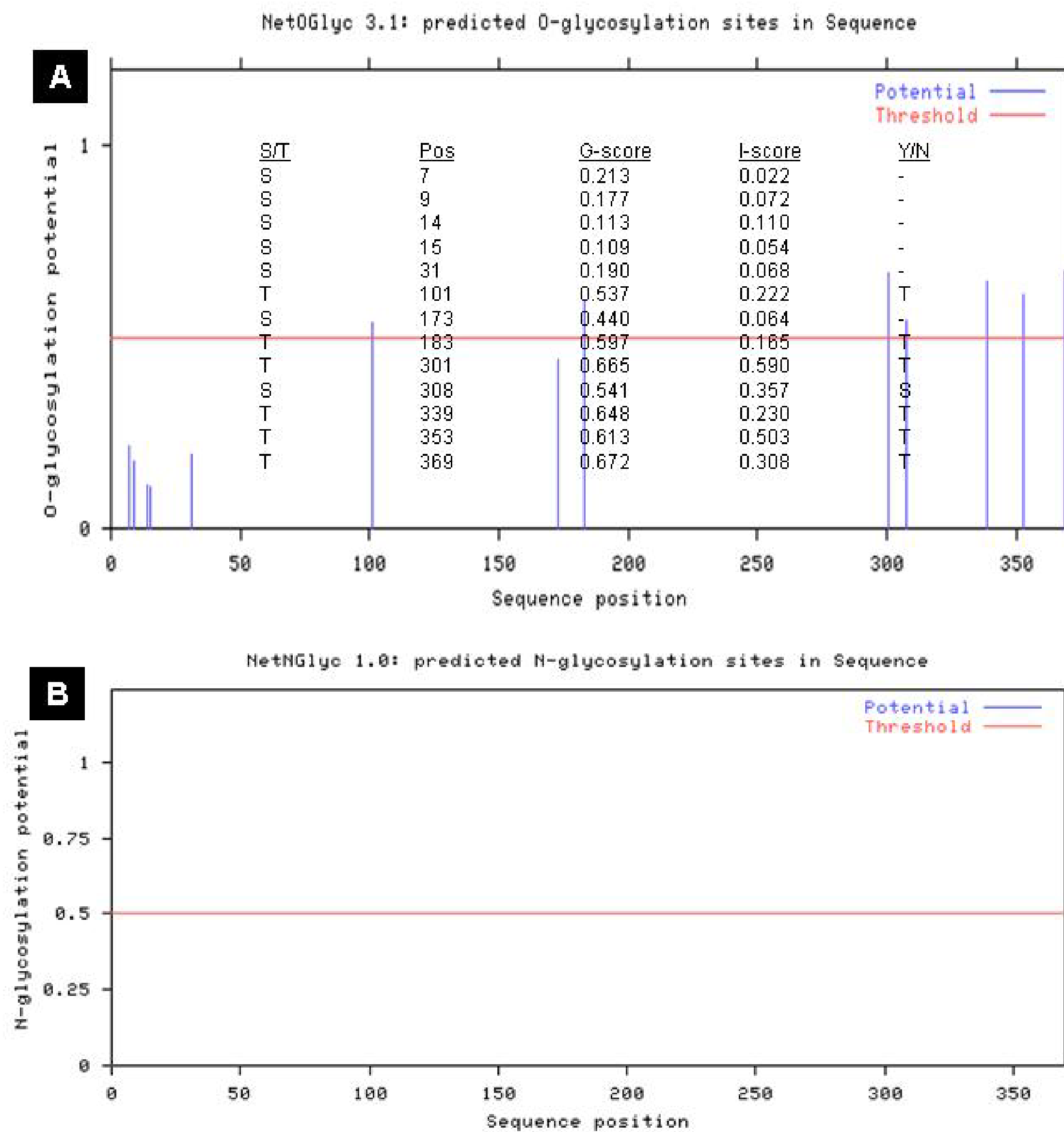


Figure 2-4. Predicted glycosylation sites in PELPK1 protein sequence. A: O-linked glycosylation sites; B: N-linked glycosylation sites. The horizontal line within the plot indicates threshold level. The table inserted in “A” shows output scores.

Identification of sequences related to PELPK1

The gene adjacent to *PELPK1* on *Arabidopsis* chromosome 5, *At5g09520*, has very high similarity to *PELPK1* (Figure 2-5; BLASTP e-value: $3.4e^{-29}$), and is here named as PELPK2. We scanned the *Arabidopsis* genome (TAIR9 release) for other proteins that contain the PELPK motif (i.e. P-E-L/I/V-P-K) using a custom Perl script. As described above, the PELPK motif occurs 34 times in the PELPK1 protein, and 4 times in PELPK2. The motif is not found to be repeated within any of the other 33,410 Arabidopsis protein sequences surveyed, although a single occurrence of the motif occurs in each of 35 different Arabidopsis proteins. Reducing the stringency of the motif search, to allow any amino acid in the second position of the motif, and any basic amino acid in the last position (i.e. P-(E/D)-(L/I/V)-P-(K/R/H)) had almost no impact on the frequency of detection of the motif in the Arabidopsis genome. Furthermore, when stringency was further reduced to allow up to one mismatch in any position of the motif, 52 and 12 occurrences were found in PELPK1, and PELPK2 respectively (see Figures 2-1 and 2-5), but only one other protein (At5g09480, putative HRGP) had 6 or more occurrences under these criteria. The PELPK motif therefore appears to be specific to PELPK1 and PELPK2.

We also searched the *Arabidopsis* proteome for proteins similar to PELPK1 by using BLASTP. A neighbor joining tree was constructed from the 20 Arabidopsis genes with the highest protein similarity to PELPK1 (BLASTP e-values $<1e^{-4}$) using MEGA 4 (http://www.megasoftware.net/mega.html), (Figure 2-6). The tree is also labeled with current annotations for each of these proteins. According to this inferred phylogeny, *PELPK1* and *PELPK2* are most closely related to other genes annotated as HRGP and PRP family members, including *At5g09480*, which we identified

```
At5g09530_protein      MALMKKSLSAALLSSPLLIICLIALLADPFSVGARRLLEDPKPEIPKLPELPKFEVPKLP
At5g09520_protein      --MTLKKSFSASLLSPFLIICLIALLSVPVSVGARRLLEEPKPEIPTFPELPKPEMPKLP
                         :  *.  :* * **:*********: *.*********:******.:***** *:****

At5g09530_protein      EFPKPELPKLPEFPKPELPKIPEIPKPELPKVPEIPKPEETKLPDIPKLELPKFPEIPKP
At5g09520_protein      EFPKLELPKLPEIPKPEMPKLPEIQKPELPTFPELP----------------KMPEFPKF
                       **** *******:****:**:*** *****..**:*                *:**:**

At5g09530_protein      ELPKMPEIPKPELPKVPEIQKPELPKMPEIPKPELPKFPEIPKPDLPKFPENSKPEVPKL
At5g09520_protein      DFPKLPELPKPEETKVPAFTMPKFPGSP--------------------------------
                       ::**:**:**** .*** :  *::*  *

At5g09530_protein      METEKPEAPKVPEIPKPELPKLPEVPKLEAPKVPEIQKPELPKMPELPKMPEIQKPELPK
At5g09520_protein      ------------------------------------------------------------

At5g09530_protein      LPEVPKLEAPKVPEIQKPELPKMPELPKMPEIQKPELPKMPEIQKPELPKVPEVPKPELP
At5g09520_protein      ------------------------------------------------------------

At5g09530_protein      TVPEVPKSEAPKFPEIPKPELPKIPEVPKPELPKVPEITKPAVPEIPKPELPTMPQLPKL
At5g09520_protein      ------------------------------------------------------------

At5g09530_protein      PEFPKVPGTP
At5g09520_protein      ----------
```

Figure 2-5: A pair-wise alignment of the PELPK1 (AT5G09530) with PELPK2 (AT5G09520). PELPK domains in AT5G09520 are highlighted in blue if they are fully conserved, and in magenta if they contain one mismatch. (*) indicates positions which have a single, fully conserved residue; (:) indicates that one of the following 'strong' groups is fully conserved: STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY, FYW; (.) indicates that one of the following 'weaker' groups is fully conserved: CSA, ATV, SAG, STNK, STPA, SGND, SNDEQK, NDEQHK, NEQHRK, FVLIM, HFY

above, based on the presence of 6 PELPK-like motifs. Although inferring relationships of highly repetitive proteins is complicated, and not well-suited to standard tools of phylogenetics, these results do show that PELPK1 and PELPK2 do appear to form a distinct group within the Arabidopsis genome.

Beyond the *Arabidopsis* genome, there is evidence for conservation of the PELPK motif and PELPK1 protein. A gene cluster that includes *PELPK1* and *PELPK2* has been defined by the Phytozome database (gene family 22878593, www.phytozome.org, Figure 2-7). The proteins in this list that are most similar to PELPK1 are from its close relative *Arabidopsis lyrata* and, surprisingly from the much more distantly related *Glycine max*. Indeed, the PELPK motif is found repeated in many species, including both monocots and dicots. Although the evolutionary history of this repeat-rich protein is still unclear, the conservation of PELPK1-like sequences over large evolutionary distances suggests that there is a specific function associated with these repeated motifs.

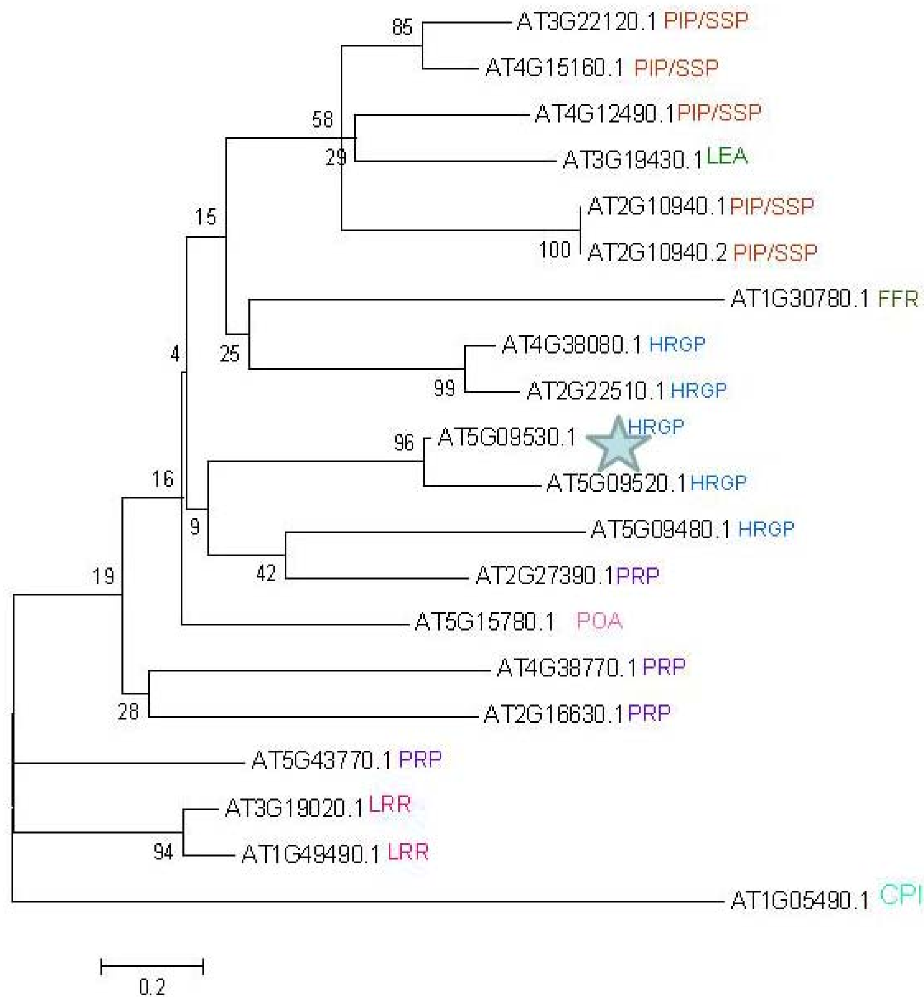


Figure 2-6: A bootstrap neighbor joining tree constructed using MEGA4 is showing the relationship between *PELPK1* and twenty other closely related *Arabidopsis* genes. The tree was drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method (Tamura et al., 2007) and are in the units of the number of amino acid substitutions per site. All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option). Bootstrap values (in %) on the branches are calculated as the number of times that a particular grouping of sequences appears during the bootstrap analysis. A 96% bootstrap value for the grouping of PELPK1 (AT5G09530) with PELPK2 (AT5G09520) indicates that in the 1000 bootstrap replicates selected, that grouping was found 960 times. HRGP, Hydroxyproline rich glycoproteins; PRP, Proline rich proteins; PIP, Protease inhibitor proteins; SSP, Seed storage proteins; LEA, Late embryogenesis abundant; LRR, Leucine rich repeats; CPI, C-protein immunoglobulin; FFR, F-box family protein; POA, pollen Ole e 1 allergen.

>Ath:AT5G09530.1|AT5G09530.1|AT5G09530|hydroxyproline-rich
MALMKKSLSAALLSSPLLIICLIALLADPFSVGARRLLEDPKPEIPKLPELPKFEVPKLPEFPKPELPKLPE
FPKPELPKIPEIPKPELPKVPEIPKPEETKLPDIPKLELPKFPEIPKPELPKMPEIPKPELPKVPEIQKPEL
PKMPEIPKPELPKFPEIPKPDLPKFPENSKPEVPKLMETEKPEAPKVPEIPKPELPKLPEVPKLEAPKVPEI
QKPELPKMPELPKMPEIQKPELPKLPEVPKLEAPKVPEIQKPELPKMPELPKMPEIQKPELPKMPEIQKPEL
PKVPEVPKPELPTVPEVPKSEAPKFPEIPKPELPKIPEVPKPELPKVPEITKPAVPEIPKPELPTMPQLPKL
PEFPKVPGTP*
>Ath:AT5G09520.1|AT5G09520.1|AT5G09520|hydroxyprolineproline-rich
MTLKKSFSASLLSPFLIICLIALLSVPVSVGARRLLEEPKPEIPTFPELPKPEMPKLPEFPKLELPKLPEIP
KPEMPKLPEIQKPELPTFPELPKMPEFPKFDFPKLPELPKPEETKVPAFTMPKFPGSP*
>Aly:487770|487770|synonym:fgenesh2_kg.6__904__AT5G09530.1
MALMKQSLSAALLSSPLLIICLIALLADPVSVGARRLLEDPKPEIPKLPELPKFEVPKLPEFPKPELPKLPE
FPKPELPKIPEIPKPELPKVPEIPKPEEAKLPEIPKPELPKFPEIPKPELPKIPEIPKPELPKVPEIQKPEL
PKVPEIPKPELPKFPEIPKPELPKFPENSKPEVPKLMETEKPEAPKVPEIPKPELPKMPEVPKLEAPKLPDI
PKPELPKIPEPKVPEIQKPELPKMPEIQKPELPKMPEIQKPELPKVPEVPKPELPTVPEVPKPEAPKLPEIP
KPELPKVPEIQKPELPKIPEIPKPAVPEIPKPELPKMPELPKLPEFP*
>Aly:487769|487769|synonym:fgenesh2_kg.6__903__AT5G09520.1
MALKKILSASLLSPFLIICIIALLSVPVSVGARRLLEDPKPEIPTLPELPKLEMPKLPEFPKPELPKIPEIP
KPELPKMPEIPKPDLPTFPEMPKMPEFPKFDAPKLPELPKPEETKVPAFTMPKFPGSP*
>Gma:Glyma14g10200.1|Glyma14g10200.1|Glyma14g10200
MATYSWSTEIFSFLVLTLLSTSSHTMVAGARNLLESTLSKPEVPTLPKPHELPPLPEIPELPKFESPKIPAL
PKPELPKVPELSKPDMSKVPELPKVPERLKVPEISKILELSKPELPKGPELLKPELPSVPNIPKVPELPKPE
LPEVPKLPKPELPKVPELPKPELPKVPEIPELPNLELPKVTQLPKSKLPKVPEIPKVPEFPKPELPKVPELS
KPELPKAPEIPKVPEFPKPELPKVPELSKPELPKAPEIPKVPEFPKPELPKVPELPKPELPKIPEIPKVPEF
PKPELPKVPEVPKPELSKVPKLPKSELPKVPEIPELPKSELPKIPEIPKVPEFPKVPKAFSTTNP*
>Gma:Glyma0041s00240.1|Glyma0041s00240.1|Glyma0041s00240
MATCGLSTVIFSFVVLMLLSASSHTMVAGARNLLESTLPKPEVPQLPKPELPPLPKVPELPKAELPKIPTFP
KPELPKVPELPKLEKPKVPELPKVSEIPKVPKLPKAPEMFKVSELPKLELPKVAEISKILELSKPELPKIPK
LPKIHELSKVPELSNPELSKVSELPKVPELPKVPELSKPELPKVPELPKVRELPKPEFSKVPELPKVHELPK
PEIPKILEFPKVPQFPKAFSTSNP*
>Lus_g7973translation
MASLVGCVPLLLMLAFTLVPGNGASRRLLDVELPKVELPPLPKLDFPQLPKPELPKPNIPEMPKPETPQLPK
PEIPQVPKPEVPQLPKPDIPQVPKPELPQVPKHEIPQLPKPEVPELPKPEIPQLPKPEVPQLPKPDLPVPPH
NAKPTELPSPSKKFEVPKLPKPDIPSLPKPELPDFPKVEIPQFPKPEFPELPRPEMPHIPELPKPEMPHIPE
LPKPEMPHIPELPKLPFPTLPKETPLPSVPHPTTTAP*
>Lus_g21250translation
MASLIGCIPLFLMLAFSLVPGNGASRRLLDVELPKPQLPSLVLPKVERPPLPKLDFPQLPKPELPKPNISEI
PKPQVPQLPKPEIPQVPKPEFPQVPEPEIPQLPKPEIPQVPKPEVPQLAKPEIPQLPKSEVPQLGKPEIPQL
PKPEVLELPQPELPVPPQNVKPTKLPTPSAKFEVPELPKPDIPSLPKPELPEFPKVEIPQFPKPEFPELPKP
EMPQIPEFPKPELPSFPKPEMPHIPELPKLPFPTLPKETPLPYVPHPTTTSP*
>Rco:29838.m001685|29838.m001685|29838.t000055|Gamma-gliadin
MACFCFRSSFFLLLLIALSIHVDARQLLETKLPEVPKPEFPKPELPKPELPTLPKTELPELPKPEVPKLPEL
QVPELPKPEVPKVPELPKLEIPKVPELPKFPELPKLEVPKVLKMPKPEMPKVPELPKPELPPLPHFPELPKP
TLPITPTVPKDIKPPQSTTTP*
>Rco:29838.m001687|29838.m001687|29838.t000057|conserved
MAYHRFLLFMLPLVWISMAVLNSQTILVNARQLLEITFPEIPELPKPELPSFPKVELPPLPEIPTLPKLKFP
DLPKPELPDLPEPEVPELPSIPHFPDLSKPTLPAIPSIPKDINPFHAATTP*
>Ptr:POPTR_0007s03360.1|POPTR_0007s03360.1|POPTR_0007s03360
MANHIFPLFISPLIVIMSMSLIISQTILVEARQLLEVTLPELPKPEFPELPKPELPKLPEFPIPELPKFEIP
KLPELPKPEFPELPKPEFPKLPEFPKPELPKFEIPKLPELPPFPHFPDLTKPTLPTIPSHSTT
>Ptr:POPTR_0007s03340.1|POPTR_0007s03340.1|POPTR_0007s03340
MANHIFPLFISPLIVIMSMSLIISQTILVEARQLLEVTLPELPKPEFPELPKPELPKLPEFPIPELPKFEIP
KLPELPPFPHFTDLTKPTLPTIPKDINPSHSTTSP*
>Ptr:POPTR_0017s07240.1|POPTR_0017s07240.1|POPTR_0017s07240
MNSQTILVEARQLLEAPLPELPKPELPKPELPELPKPEFPELPPKPELPKFEVPQLPELPTFPHLPELPKPT
LPTIPKDINPSHSTAISGGEAARHLLQFPPLPSVPNLPKPTLPSMPTLPQLTLPTSPKSQLSLPKPTLPPLP
SLPTMPSLPKVALPPLPTMPSIPFLSPPPGN*
>Ptr:POPTR_0007s03370.1|POPTR_0007s03370.1|POPTR_0007s03370|synonym:euge
ne3.01640065MASLRFLTFLSPLLLITLSMVDNTCRTEARRILETTLPMVPELPKPELPEMPPLPKVELP
TIPKPELPELPKPEVPKMPELPSFPHFPELPKTTLPTIPALPKDIKPPQSTTSP*
>Vvi:GSVIVT00031600001|GSVIVT00031600001|GSVIVT00031600001
MAHHHEPSILLPLLLITLSLMSGKEVLASRHLLETTLSTVPELHKVELPPLPTLPTLPKFELPPLPKVEIPS
LPHVPTLQEPQLPTLPKPELPTVPHVPALEKPELPELPSLPHLPDLPKLTLPTIPTLPKDIPFPSLSPPHST
TSP

>Cpa:evm.model.supercontig_42.15|evm.model.supercontig_42.15|evm.TU.supe
rcontig_42.15
MAFMTTNTFLLVGARRLLETKEPELPKLPELPHLTKIPTELPKPELPKKPIELPKVPELPKVPELPKLGIPT
VPVELPKPEMPKIPELPKPELPKPEVPKMPELPKPEVSKMTTELPKPELPKIPELPEMPELPKPETPKTPTE
LPKPELPKMPELPKLEMPKIPGELPKPELPELPELPKPELPKMPELPKPEIPAELPKPELPKTSSSP*
>Mgu:mgf020799m|mgf020799m|mgf020799m
MALIDIRSSVFLLLLITLSSITGNTNQAEAARHLLEIPKPEIPAFPKPELPTLPKPELPTLPKPELPAFPKL
ELPAFPKPELPAFPKPELPAFPKPELPAFPTPQVTLEKAELPSLPKPEIPTFPKPELPAFPKPELPTLPKPE
VPKLPKLPALPKPELPAFPKPELPSLPKPELPAFPKPEIPTLPKPELPAIPKPELPAFPKPEVPKLPELPKF
PELPKPTLPEGPKHP*
>Csa:Cucsa.218330.1|Cucsa.218330.1|Cucsa.218330
MATLRFSTLVFPLILGVFMLMNFDISIMTVDARYLLETPLPEELPKPEIPSLPNLEFPPLPKVELPPLPHLP
PLEKPKLPELPKVPNLDHSKDSSNVPKLSEVSELPKPEFPKFPELPNFPKPEVPKLPELPNVPKPELPKLPE
FPPLPKPAFASLPKDIPYPSFNSPHTTTINP*
>Sbi:Sb02g038440.1|Sb02g038440.1|Sb02g038440|similar
MASSSRSTVSSFLLVALLLCCSRMSSAARLLEEAPPKEEYPHPAVPELPKPELPPHPTDVVPPEVPHPVPEQ
PKPELPPHPAVPELPKPEVPHTVPEQPKAELPPHPAVPELPKPEVPHPVPELPTPELLPPHPAVPELPKPEV
PPHPVAPELPKPELPPHPTVPELPHPDVPEVPNHELPPLPKAELPPKPEGHYPEPEAKP*
>Sbi:Sb02g038450.1|Sb02g038450.1|Sb02g038450|similar
MASKSTMHSLVLLMGLLLLSCSGTSSAARLLEEVAAPKEEYPHPEVPELPKPELPPHPTVPEFPKPELPPHP
VLPELPKPEVPVHPAVPEKLPKPELPKPEVPEHPAAVPELPKPELPPHPTPTVPELPKPEVPEHPAAVPELP
KPELPPHPTVPELPKPEVPEHPAAVPELPKPEMPEHPTVPELPPLAEPELPVPPKAESHYPEPETKP*
>Sbi:Sb02g038460.1|Sb02g038460.1|Sb02g038460|similar
MAAKSTMIPSLVLLIALLLSCSGISGAARLLEEAEYPHPATPAELPKPELPPHPTVPELPKPEVPAHPATPE
LPKPEIPEHLPELPKPELPPPHPAAVPELPKPEVVPEHPAVPELPKPELPPHPAVPELPKPEVPAAHPTVPE
LPKPELPPHPTVPELPKPEVPKHELPPKPESHYPVPETKP*
>Sbi:Sb02g038470.1|Sb02g038470.1|Sb02g038470|similar
MPLFSNHHDKNLQLSSDTQIEIAAAQKHKRFDMVSKRTMSSSLAFLMALLLSCSTMSSAARYLEETKPAEEY
PPHPTVPEIPKPELPPHPTVPELPKPELPPHPTVPELPKAELPPHPAVPEHPKPEPPPHPTVPEHHPTVPEL
PKPELPHPTVPEVPKPELPHPAVPELPHSAVPEVPKPELPHPAVPEVPKPELPPHPAAPEVPKPEMPHPAVP
EVQKPELPHPEVPKPELPPHPTVPEVPEVPKHELPPKPESHYPVPEAKP*
>Sbi:Sb02g038480.1|Sb02g038480.1|Sb02g038480|similar
MSSLALVLALLLSSGAIGSAARQLQELPPLPKPEIPPRPDLPPLPKPEEQPPLPKPELPVPPQPLPKPELPV
PPQPLPVPPQPLPKPELPVPPEPLPKPELPVPPQPLPKPEVPVTPEPLPKPELPVPPQPLPKPELPVPPEPL
PKPELPVPPEPLPKPELPVPPQPLPKPELPPPVLAGELPPKP*
>Bdi:Bradi1g22440.1|Bradi1g22440.1|Bradi1g22440
MGCKIVFVMALLLLSCSSMSSAARHLEEIVPKENQHPSYPTVPKPEVPAHPAVPELPKPEMLHPVVPEAPKE
HQVPHSVVPEVPKEHEVPHPVVPELPKPELPPHLTVPELPKAELPHPASLPEVPHPVVPEVPKEHEVPHLAV
PELPRPEMPRPAMPEVPKEPHVSHPVVPELPKPEEPKHELPPFPKAELPPKPEFHFPEPEAKP*
>Bdi:Bradi1g22450.1|Bradi1g22450.1|Bradi1g22450
MDAKSRVSSVFFLVVLLLSCSCMTRAARYLEEKVPKEEVPKMPELPHPVVPEVPKKPEESHPVVPELPKPEL
PHPVMPEVPKMPEVPHVSIPKVPKSEIPHLVVPEVPKAPEVPHPTVPEVPKMPEVPHLSTPEAPKVSEVPHP
TMSGVPKMPELSHPVMPEVPKVSEVPHPAVPEVPKMPEVPHPAIPEMPKASELPHPEVPEAPKIREVPHPAV
PELPKMPEMPHLTMPEVPKLPEVPHVSIPEIPHPAIPEVPKHELPPVPKVEVPPKPEGIPQYPKPEAKP*
>Bdi:Bradi1g22470.1|Bradi1g22470.1|Bradi1g22470
MPEMPNPVVPEVPKMPEELHPVVPELPKPELPHPGTSEVPKMPEVPRASMPEVPNLEIPHLIVSKVPKTPEV
THPIVSEVPKMPELPRVTAPEATKMPKAPHPIVPEATKMPEVPHPIVPEVPKMPEVPHPTVLEVPKMPEVSH
PTVPEVPKIPELAHPAAAEVPKMPEVPHPTMPELPKVPEVSHLTIPKAPEESHPVVPEAPKTRLPHPVVPEV
PKMPEMPHNTMPNVPKMIEVPRPTMPHVPKPKIPHPTMPEVPKHELPRVPKVELPPKPEDTLHYSEPEAKP*
>Osa:LOC_Os07g40830.1|LOC_Os07g40830.1|LOC_Os07g40830|synonym:13107.m042
48|app1,MSSSSSSSSALLLMAALLLSCGAMGSTARHLEEKAPHFPAVPELPPHPELPELPKPELPPPLPE
LPRPVVPELPPHPAVPELPPLPKPELPPHPVVPEMLPHPVVPELPHYPAVPGFPKHGLPPKPELPPLPTAEL
PPEHEVHDPEPETKQP*
>Osa:LOC_Os07g40870.1|LOC_Os07g40870.1|LOC_Os07g40870|synonym:13107.m042
53|igAMASSHTKPSILLLAAALLLLSCSSIGGAARYLEEAAPAAAAAEEEEHPAHPAVPEIPKPELPELPK
VPELPHPVVPELPKPELPKIPEVPHLAVPELPKPEVPEIPKAELPPLPKFELPPKPEFHFPEPEAKP*
>Osa:LOC_Os07g40890.1|LOC_Os07g40890.1|LOC_Os07g40890|synonym:13107.m042
55|igAMASKNSMSSSLLFLMALLLSWSSISSAARYLEEEAAPKEEYPELPKPELPPHLAVPELPKPELPHG
AAVPEFPKVPELPHPEVPELPKPELPEHPAVPELPKPELPSLPKVELPPLPKPEFHFPEPEAKP*
>Zma:GRMZM2G019373_P01|GRMZM2G019373_T01|GRMZM2G019373
MASSSRSTVFSLLLLVALLLSCSGMSSAARLLEEAPPKEEHPHPAVPELPEPELPPHPTDVVPPELPKPELP
PHPAVVPELPKPEVPHPVPEQPKPELTPHPAAVPELPKPEVPHPTAVPELPKPEVPHPVPELPKPEVPPHPT
AVPELPKPEVPYPVPELPKPELTPHPAAVPELPKSEVPHPVPELPKPEVPPQPTAVPELPKPEVPHPVPELP

KPELPPHPAV**PELPKPEVPH**QVA**PELPKPELPP**H**PTVPK**LPH**PEVPE**VPNHEL**PPLPK**AELPPKPEGHYPEP
EAKP*

Figure 2-7 Sequences from higher plants with repeated patterns of similar motifs to PELPK1 (At5g09530). The PELPK motif (where L is Leu or any branched chain amino acid, E is any acidic residue, and K is any basic residue) is shown in blue, and motifs with one variation from the PELPK consensus are shown in magenta. Ath= *Arabidopsis thaliana*; Aly= *Arabis lyrata*; Gma= *Glycine max*; Lus= *Linum usitatissimum*; Rco= *Ricinus communis*; Ptr= *Populus trichocarpa*; Vvi = *Vitis vinifera*; Cpa= *Carica papya*; Csa= *Cucumis sativus*; Sbi= *Sorghum bicolor*, Bdi = *Brachypodium distachyon*; Osa= *Oryza sativa*; Zma = *Zea mays*.

Is PELPK1 an HRGP, extensin, or periaxin?

As described in the introduction to this chapter, *PELPK1/At5g09530* has been assigned several different annotations in public databases, including descriptions referring to HRGPs, extensins, PRPs, and periaxins. The HRGP-family proteins, of which extensins are a member, are characterized by a repetitive region rich in Pro/Hyp, and a main repeating pentapeptide motif consisting of Ser-Hyp4, and are reported to be

associated with the cell wall (Showalter 1993; Cassab 1998). Periaxin, on the other hand, has been reported to function in maintaining the myelin, a protective substance that covers nerves and promotes efficient transmission of nerve impulses in animals including humans (Gillespie et al., 1994). A pair-wise alignment of the deduced protein sequence of *PELPK1* with that of an extensin, PRP, and a human periaxin showed only limited and scattered sequence similarity (not shown). The BLAST alignment algorithm reports these alignments as significant because of all three proteins contain proline-rich repeats, however, as described below, the composition of these repeats is very different in each protein. Therefore, standard interpretations of BLAST results used to generate automated annotations may overstate the similarity of extensins, periaxins, and proline-rich proteins to *PELPK1*.

PELPK1 is defined by the following predicted features: a highly repetitive sequence containing a unique motif, all of which forms a highly hydrophilic, intrinsically disordered protein. We have already shown that the repeated PELPK motif is unique to PELPK1 and PELPK2, and is therefore not present in extensins or PRPs. To test for similarities in higher-level structures, we used the VSL2 algorithm of DisProt to predict the degree of order or disorder in each of the *Arabidopsis* extensins and PRPs (as defined by Showalter et al., 2010) and compared these to PELPK1-like proteins identified in various organisms (Figure 2-8 gene family 22878593, www.phytozome.org). Figures 2-8 and 2-9 show that most PRPs and extensins are predicted to contain well-ordered domains with regular structures at both the N-terminus and elsewhere in the protein. The ordered domains are especially prevalent among the extensins, which is consistent with what is known about the regular, rod-like structures characteristic of most (but not all) proteins called extensins. This can be contrasted with the distribution of disordered domains among the PELPK1-like proteins (Figure 2-10). Although among

even extensins and PRPs, there are some proteins that are predicted to be largely disordered, disorder appears to be highly prevalent in all PELPK1-like proteins (Figure 2-10), suggesting that this is a characteristic that distinguishes PELPK1 from other types of HRGPs.

Glycosylation is also a defining feature of some cell wall proteins. Predicted glycosylation sites for PELPK1 are described above (Figure 2-4). A comparative analysis indicated (i) that extensins contained the highest O-linked glycosylation sites (2+16=18), followed by the PRPs (7 + 4 = 11), PELPK1 (6+1=7), PELPK2 (5+1=6) and periaxin (2+0), and (ii) that none of the above proteins contained N-linked glycosylation sites, except periaxin (which contained one) (Table 2-2). In addition, the Pro residues in the repetitive motifs could also be the target for O-linked glycosylation provided that they undergo prior hydroxylation, for which a Ser-Pro or Ala-Pro dimer is thought to be needed (Shpak et al., 2001). Visual analysis of the primary structure of PELPK1 (and its paralog, PELPK2) did not show any Pro residues in either of these two arrangements (not presented), suggesting a lack of Hyp-based O-linked glycosylation within the PELPK1 protein. In contrast, the extensin contained >15 Ser-Pro dimers in the repeat motifs (not shown), suggestive of extensive Hyp-based O-linked glycosylation

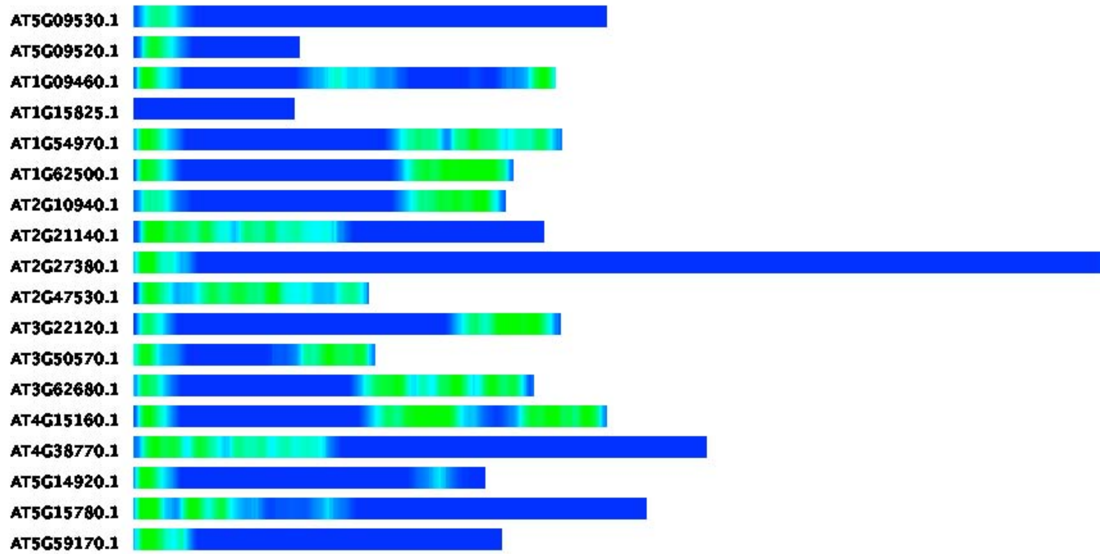


Figure 2-8. Heat map of probability of local protein disorder among all PRPs. The membership of the PRP family is shown as defined by Showalter et al., 2010). Protein disorder was predicted by the VSL2 algorithm of DisProt. Blue color indicates high probability of disorder, and green color indicates low probability of disorder. PELPK1 (At5g09520) and PELPK2 (At5g09520) appear at the top of the figure.

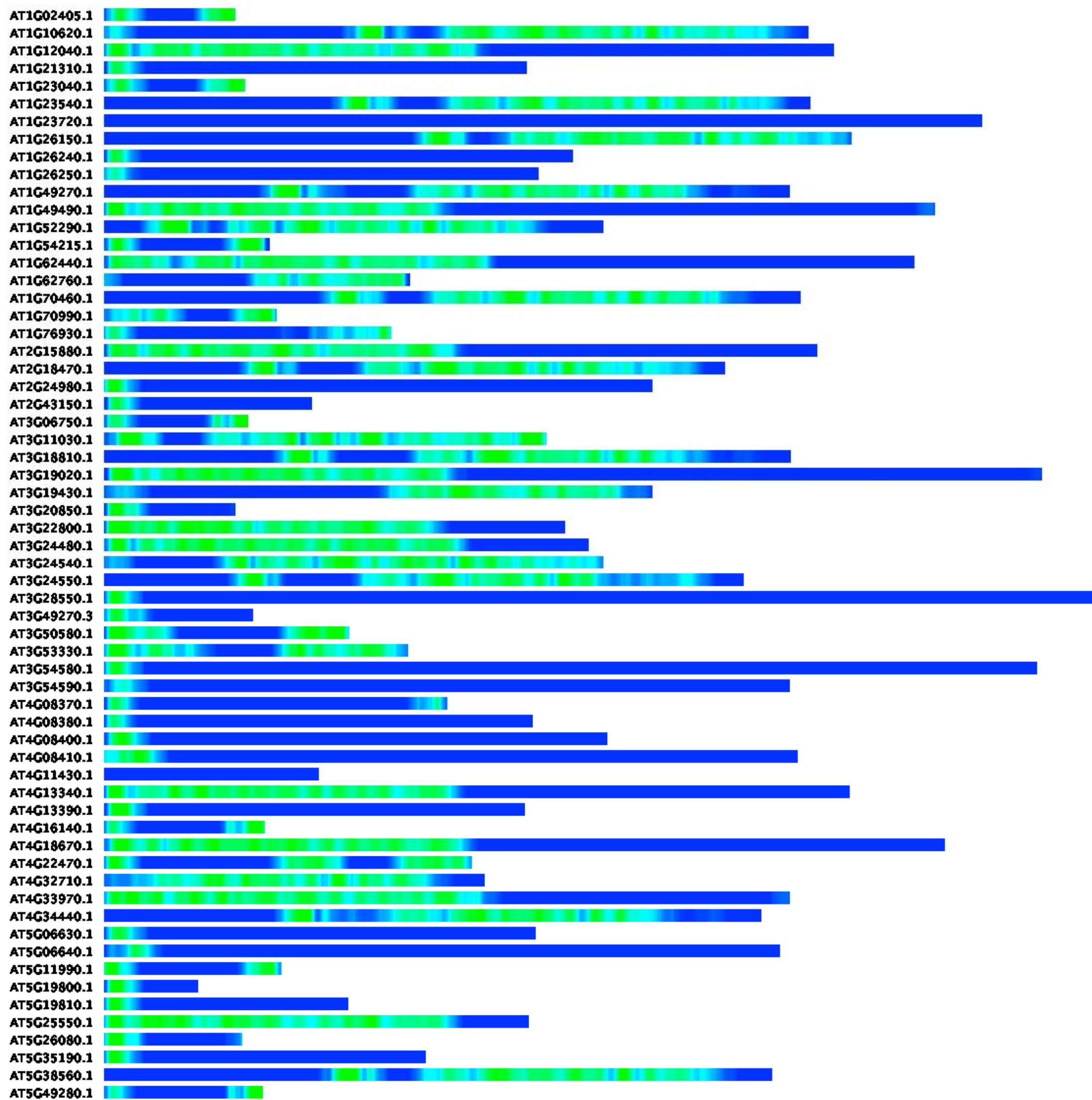


Figure 2-9. Heat map of probability of local protein disorder among all extensins. The membership of the extensin family is shown as defined by Showalter et al., (2010). Protein disorder was predicted by the VSL2 algorithm of DisProt. Blue color indicates high probability of disorder, and green color indicates low probability of disorder.

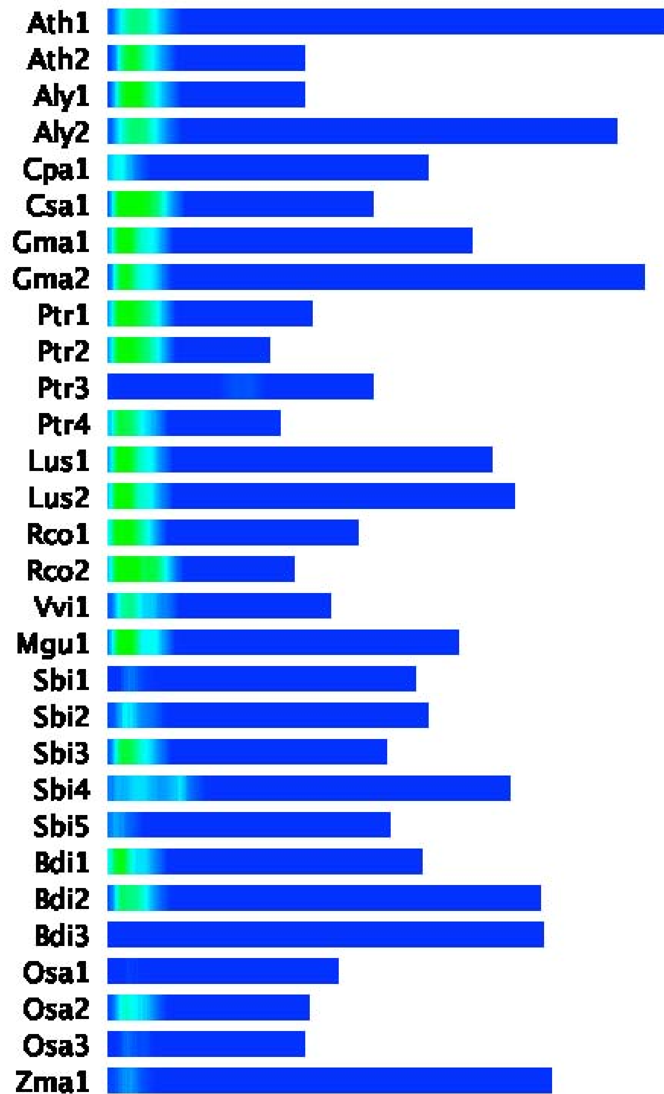


Figure 2-10. Heat map of probability of local protein disorder among PELPK1-like proteins from various species. Species are abbreviated as shown in Figure 2-7. Protein disorder was predicted by the VSL2 algorithm of DisProt. Blue color indicates high probability of disorder, and green color indicates low probability of disorder. PELPK1 (At5g09520) and PELPK2 (At5g09520) appear at the top of the figure as Ath1 and Ath2.

Analysis of expression profiles of PELPK1

A survey of microarray expression profiles from the Genevestigator V3 database (ATH1: 22k full genome Affymetrix GeneChip; https://www.genevestigator.com/gv/user/serveApplet.jsp; Zimmermann et al., 2004) showed that *PELPK1* (probe set: 250500_at) transcripts expression varies depending on the tissue examined and various external factors (e.g. exogenous hormones, environmental stress). These observations are summarized in the following two subsections:

Tissues - this expression profile of PELPK1 showed the following ranking of transcript abundance (arbitrary units in parenthesis): radicle (22.0)> root (15.0)> hypocotyl (6.5)> seed (5.5)> seedling (4.0)> silique (3.0)> inflorescence /leaf/ root-hair (~1.5)> others (Table 2-3). A growth-stage specific transcript expression profile however showed highest expression during flowering and silique stages followed by seedlings or germinated seeds and rosette (Figure 2-11). Microarray expression profiles of *PELPK1* from a seed-specific database (http://seedgenenetwork.net/) further showed that *PELPK1* transcripts were maximally expressed in the seed coat of maturation green stage seeds (not presented).

External factors – the *PELPK1* gene is induced by a variety of factors including abiotic, biotic, environmental, and hormonal factors. Based on the impact of these factors on PELPK1 expression, they can be roughly summarized as follows: biotic factors >>>> elicitors >>> nutritional factors >> abiotic factors ≥ hormones/defense chemicals (Table 2-3). These observations show that *PELPK1* is maximally induced by biotic factors. Among the biotic factors, *Pseudomonas syringae* induced the highest level of *PELPK1* expression (infected/uninfected signal ratio = 70). More specifically, the factors responsible for significant up-regulation of *PELPK1* in microarray analysis are as follows: *Pseudomonas syringae* (biotic) >>>>>>> CalCuV, cabbage leaf curl virus (biotic) >> CS-

137 (nutrient) ≥ elevated CO_2 (abiotic) > glucose (nutrient) ≥ nitrate starvation (nutrient) > syringolin (elicitor/chemical) ≥ *Blumeria graminis* (biotic) ≥ LPS (elicitor) > drought (abiotic) > cold (abiotic) ≥ salt (abiotic) > FLG-22 (eliciror) > iron deficiency (nutrient) > ABA (hormone) > BL/H_3BO_3 (hormone) > wounding (abiotic) > salicylic acid (defense hormone) ≥ LPS (elicitor) ≥ oxidative stress (abiotic) ≥ low nitrogen > others.

Table 2-3. A summary of microarray-based expression profile of the PELPK1 and PELPK2. DAS, days after sowing; >, indicates higher expression

Expression profile		PELPK1	PELPK2
Organ-specific expression		Root> Seed> Hypocotyl> Radicle> Others	Root> Petal> Hypocotyl> Radicle> Others
Abiotic stress:	treated/untreated ratio		-
	Low light	5.24	
	Elevated CO_2	11.27	
	Cold (4°C)	8.21	
	Drought	8.58	
	Salt	8.17	
	Wounding	5.58	
	Oxidative stress	4.65	
	Temperature (25°C)	2.47	
Biotic factors:	infected/uninfected ratio		-
	Blumeria graminis	10.32	
	CalCuV	23.47	
	P. syringae	69.91	
Elicitors:	treated/untreated ratio		-
	FLG-22	8.01	
	HrpZ	4.72	
	LPS	10.22	
	Syringolin	10.37	
Hormonal factors/defence chemicals: as above			-
	ABA	6.72	
	BL/$H3BO_3$	6.21	
	Ethylene	1.93	
	MeJa	2.09	
	SA	2.63	
	TIBA	3.77	
Nutritional factors: as above			-
	CS-137	11.36	
	Glucose	10.67	
	Iron deficiency	7.3	
	Low nitrogen	4.64	
	Mannitol (3%)	5.42	
	Nitrate starvation	10.51	
	Sucrose	3.82	

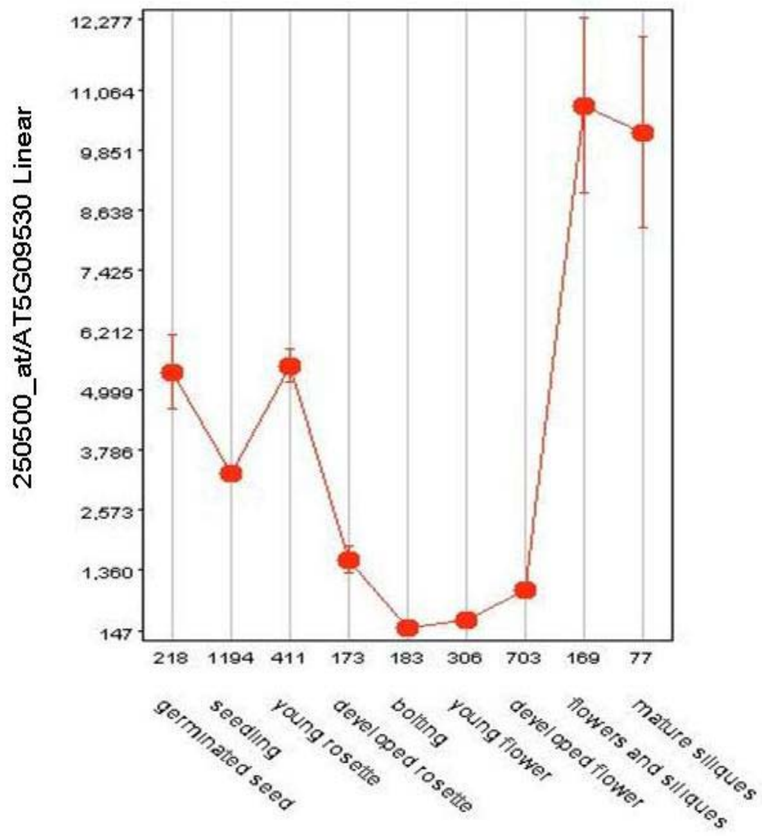


Figure 2-11: Microarray-based growth-stage specific expression profile of the PELPK1 (numbers on the y-axis indicate mean values)

3. CONCLUSIONS

In summary, the bioinformatics analyses presented above indicates that *PELPK1* and its presumptive paralog, *PELPK2* form a distinct subgroup of HRGP-family proteins. PELPK1, like extensins and PRPs, is highly repetitive and rich in Pro residues and contains a transmembrane domain and a predicted transit sequence consistent with targeting to the secretory pathway. However, my analyses do not support the description of PELPK1 as an extensin-like protein or periaxin-like protein, because motif analysis showed that repeated motifs in these proteins are not conserved (Table 2-2). Furthermore, PELPK1 differs from PRPs, extensins, and periaxins in the frequency of their predicted glycosylation sites, and in the extent of disordered domains within the predicted protein structures. I conclude, based on *in silico* analyses that, PELPK1 is most likely a secreted, intrinsically disorded protein that is not highly glycosylated, and is distinct from extensins and typical PRPs, and is expressed in a tissue-specific manner during development and in response to exogenous factors, including pathogen attack.

4. REFERENCES

Ackerman, M. S. and Shortle, D. (2002). Robustness of the long-range structure in denatured staphylococcal nuclease to changes in amino acid sequence. Biochemistry, 41:13791-13797.

AGI (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408: 796–815

Attwood,TK. (2000). The quest to deduce protein function from sequence: the role of pattern databases. Int. J. Biochem. Cell Biol., 32, 139–155.

Bates, G. (2003). Huntingtin aggregation and toxicity in Huntington's disease. Lancet, 361:1642-16424.

Cassab, G.I. (1998). Plant cell wall proteins. Annual review of plant physiology and plant molecular biology. 49: 281-309

Clare, A. and King,RD. (2003) Predicting gene function in it Saccharomyces cerevisiae. Bioinformatics, 19, ii42–ii49.

Dunker, A. K., Brown, CJ, Lawson, J D. Iakoucheva, LM, and Obradovic, Z. (2002). Intrinsic disorder and protein function. Biochemistry, 41:6573-6582.

Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., Oldfield, C. J., Campen, A. M., Ratliff, C. M., Hipps, K. W., Ausio, J., Nissen, M. S., Reeves, R., Kang, C., Kissinger, C. R., Bailey, R. W., Griswold, M. D., Chiu, W., Garner, E. C., and Obradovic, Z. (2001). Intrinsically disordered protein. J. Mol. Graph Model., pages 26-59.

Dyson, H.J. and Wright, P.E. (2005). Intrinsically unstructured proteins and their functions. Nature Reviews: Molecular Biology, 6:197-208

Emanuelsson, O, Brunak, S, von Heijne, G, Nielsen, H (2007). Locating proteins in the cell using TargetP, SignalP, and related tools. Nature Protocols 2, 953-971 (2007).

Ganesan H, Rakitianskaia AS, Davenport CF, Tümmler B, Reva ON (2008). The SeqWord Genome Browser: an online tool for the identification and visualization of atypical regions of bacterial genomes through oligonucleotide usage. BMC Bioinformatics 7: 333.

Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A.; Protein Identification and Analysis Tools on the ExPASy Server; (In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press (2005). pp. 571-607

Gillespie, C. S., Sherman, D. L., Blair, G. E. and Brophy, P. J. (1994). Periaxin, a novel protein of myelinating Schwann cells with a possible role in axonal ensheathment. Neuron 12, 497-508.

Gunasekaran, K., Tsai, C., Kumar, S., Zanuy, D., and Nussinov, R. (2003). Extended disordered proteins: targeting function with less scaffold. Trends Biochem. Sci., 28:81-85.

Gutie´rrez,R.A. et al. (2004) Phylogenetic profiling of the Arabidopsis thaliana proteome: what proteins distinguish plants from other organisms? Genome Biol., 5, R53.

Hulo N, Bairoch A, Bulliard V, Cerutti L, Cuche BA, de Castro E, Lachaize C, Langendijk-Genevaux PS, Sigrist CJ.The 20 years of PROSITE. Nucleic Acids Res. 2008; 36(Database issue):D245-9

Hvidsten,T.R. et al. (2001) Predicting gene function from gene expressions and ontologies. Pac. Symp. Biocomput., 299–310.

Josè M., and Puigdomènech P. (1993). Structure and expression of genes coding for structural proteins of the plant cell wall, New Phytol. 125, 259–282.

Kaplan, B., Ratner, V, and Haas, E. (2003). Alpha-synuclein: Its biological function and role in neurodegenerative diseases. J. Mol. Neurosci., 20:83-92.
Kelley LA and Sternberg MJE (2009). Protein structure prediction on the web: a case study using the Phyre server. Nature Protocols. 4: 363 - 371
King, R., Karwath, A, Clare, A. and Dehaspe, L (2000) Accurate prediction of protein functional class in the M. tuberculosis and E. coli genomes using data mining. Comp. Funct. Genomics,17, 283–293.
King, R., Karwath, A, Clare, A. and Dehaspe, L (2001) The utility of different representations of protein sequence for predicting functional class. Bioinformatics, 17, 445–454.
King,R., Karwath,A., Clare,A. and Dehaspe,L. (2000) Genome scale prediction of protein functional class from sequence using data mining. In Proceedings of theACM International Conference KDD 2000, ACM, Boston, USA.
Klein-Seetharaman, J., Oikawa, M., Grimshaw, S. B., Wirmer, J., Duchardt, E., Ueda, T., Imoto, T., Smith, L. J., Dobson, C. M., and Schwalbe, H. (2002). Long-range interactions within a nonnative protein. Science, 295:1719-1722.
Marcotte,E. et al. (1999) A combined algorithm for ?genome-wide prediction of protein function. Nature, 402, 83–86.
Plaxco, K. W. and Gross, M. (2001). Unfolded, yes, but random? never! Nat. Struct. Biol., 8:659-660.
Pavlidis,P., Weston,J., Cai,J. and Grundy,W. (2001) Gene functional classification from heterogenous data. In Proceedings of RECOMB 2001.
Radhakrishnan, I., Perez-Alvarado, G. C., Parker, D., Dyson, H. J., Montminy, M. R., and Wright, P. E. (1997). Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: a model for activator:coactivator interactions. Cell, 91:741-752.
Saqi, M. A. and Sternberg, M. J. (1994). Identification of sequence motifs from a set of proteins with related function. Protein Eng., 7:165-171.
Schlueter, S. D., Wilkerson, M.D., Huala, E., Rhee, S.Y., Brendel, V. (2005). Community-based gene structure annotation. Trends in Plant Science 10, 9-14
Schweers, O., Schonbrunn-Hanebeck, E., Marx, A., and Mandelkow, E. (1994). Structural studies of tau protein and Alzheimer paired helical filaments show no evidence for beta-structure. J. Biol. Chem., 269:24290-24297.
Shortle, D. and Ackerman, M. S. (2001). Persistence of native-like topology in a denatured protein in 8 M urea. Science, 293:487-489.
Showalter AM. 1993. Structure and function of plant cell wall proteins. Plant Cell 5:9–23
Showalter, AM, Keppler, B, Lichtenberg, J, Gu, D, and Welch, LR (2010). A bioinformatics approach to the identification, classification, and analysis of hydroxyproline-rich glycoproteins. Plant Physiology 153:485-513
Shpak E, Barbar E, Leykam JF, Kieliszewski MJ (2001) Contiguous hydroxyproline residues direct hydroxyproline arabinosylation in Nicotiana tabacum. J Biol Chem 276: 11272–11278
Syed,U. and Yona,G. (2003) Using a mixture of probabilistic decision trees for direct prediction of protein function. In Proceedings of RECOMB 2003, ACM, Berlin, Germany.
Smyth, E., Syme, C. D., Blanch, E. W., Hecht, L., Vasak, M., and Barron, L. D. (2001). Solution structure of native proteins with irregular folds from Raman optical activity. Biopolymers, 58:138-151.
Tamura K, Dudley J, Nei M & Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution 24:1596-1599

Tompa, P. (2002). Intrinsically unstructured proteins. Trends Biochem. Sci., 27:527-533.
Uversky, VN. (2002). Natively unfolded proteins: a point where biology waits for physics. Protein Sci., 11:739-756.
Uversky VN, Oldfield, CJ, Midic, U, Xie, H, Xue, B, Vucetic, S, Iakoucheva, LM, Obradovic, Z, and Dunker, AK (2009). Unfoldomics of human diseases: linking protein intrinsic disorder with diseases. BMC Genomics 10 (Suppl 1):S7
Verkhivker, G. M., Bouzida, D., Gehlhaar, D. K., Rejto, P. A., Freer, S. T., and Rose, P. W. (2003). Simulating disorder-order transitions in molecular recognition of unstructured proteins: where folding meets binding. Proc. Natl. Acad. Sci. U.S.A., 100:5148-5153.
Wortman, J.R., Haas, B.J., Hannick, L.I., Smith, R.K.,Jr., Maiti, M., Ronning, C.M., Chan, A.P., Yu, C., Ayele, M., Whitelaw, C.A., White, O.R., and Town, C.D. (2003). Annotation of the Arabidopsis Genome1. Plant Physiology, 132, 461-468
Wright, P. E. and Dyson, H. J. (1999). Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. J. Mol. Biol., 293:321-33
Zimmermann, P, Hirch-Hoffman, M, Henning, L, and Gruissem, W. (2004). Genevestigator, Arabidopsis microarray database and analysis toolbox. Plant Physiol., 136:2621-2632
Zimmermann, P, Hirch-Hoffman, M, Henning, L, and Gruissem, W. (2004). Genevestigator, Arabidopsis microarray database and analysis toolbox. Plant Physiol., 136:2621-2632

CHAPTER 3

EXPRESSION ANALYSIS OF *PELPK1*

CHAPTER 3

EXPRESSION ANALYSIS OF *PELPK1*

1. INTRODUCTION

Gene expression describes the spatial and temporal pattern of gene activity, as well as its inducibility by abiotic, biotic or chemical signals and other environmental factors. Knowledge of the expression pattern of a gene can contribute to understanding the gene's function. The term expression can be used to refer to either transcription, or translation, or to the actual functional activity of a gene of interest. Frequently, measurement of transcript abundance is used to study gene expression, however it must be noted that transcript expression patterns does not always equate with gene activity. For functional expression of a gene, it has to be transcribed, translated, post-translationally modified, and then the processed protein may also need to be transported to its site of action. Therefore, the expression pattern of a gene should be defined based on a number of complementary experiments including analysis of transcript or protein accumulation patterns, analysis of promoter activity, sub-cellular localization, and functional (e.g. enzymatic) assays of the protein.

In the present study, to help characterize the expression pattern of *PELPK1*, a promoter-reporter fusion analysis as well as sub-cellular localization of the PELPK1-encoded protein through translational reporter gene fusions, were carried out. Protein expression of the *PELPK1* is addressed indirectly in a subsequent chapter of seed coat proteomics (Chapter 5). It is not possible to assay the functional activity of the *PELPK1* protein in isolation, as this activity is still unknown.

While the upstream intergenic DNA often contains most of the regulatory elements responsible for driving the expression of the native gene, studies have shown that some regulatory elements are present within the transcribed region of the gene

(Sieburth and Meyerowitz, 1997; Tylor 1997), or even in downstream intergenic DNA. It was therefore suggested that results of other independent techniques must be combined with the promoter::reporter (β-glucoronidase, GUS) fusion technique to demonstrate accurate gene expression patterns (Tylor 1997*).* It was however noted that despite this limitation, promoter::reporter fusion analysis will continue to play a useful role in elucidating the mechanisms underlying the transcriptional regulation of plant genes as long as they are performed carefully and the data interpreted critically (Tylor 1997).

Furthermore, it has been suggested that reporter gene expression can be confounded by the interactions between the test promoter and those driving the expression of other genes within the same construct (Yoo et al., 2005). The pCAMBIA vector series website (www.cambia.org/daisy/cambia/585) has also acknowledged the problems with some of their promoter-testing vectors. The constitutive CaMV35S promoter, in its various forms, contains one or more enhancer elements that can bi-directionally drive transcription and/or supplement the ectopic expression of noncontiguous nearby genes (Xie et al., 2001; Yoo et al., 2005). It has further been demonstrated that transcriptional interference can occur between promoters within a construct, depending on their separation distance and relative orientation (Yoo et al., 2005). Based on the above information, it has been speculated that the enhancers, irrespective of their origin, may cause unintended misexpression of other transgenes included in a transformation vector or they may even modify the transcription of endogenous sequences nearby the T-DNA insertion site. These problems can be overcome by using a co-transformation strategy.

Annotations of the *PELPK1* sequence in public databases (e.g. TAIR, NCBI) describes it as a HRGP-family protein containing a Pro-rich extensin-like domain that is likely to be cell wall related. However, a detailed analysis of *PELPK1* bioinformatics databases (Chapter 2) showed that it also differs from extensins in a number of

characteristics including the structure of its repetitive motifs and frequency of predicted glycosylation sties. Bioinformatics analysis showed *PELPK1*-encoded protein to contain a N-terminal and secretory pathway directed signal peptide, a TM domain, highly repetitive sequence motifs, and was highly rich in Pro and moderately rich in Glu, Lys and Leu residues. Computational analysis further predicts that *PELPK1* encodes an intrinsically disordered protein (Chapter 2) which does not form regular 3D structure (Wright and Dyson, 1999 and Dunker et al., 2001, 2002, Dyson et al., 2005). Based on the above information, it was hypothesized that *PELPK1* encodes a CW protein with a function that has diverged from that of canoncial extensins. To determine sub-cellular localization of *PELPK1*-encoded protein, studies were carried out using a CDS (coding sequence)::GFP (Green Fluorescent Protein) translational fusion.

The overall objectives were (i) to analyze the activity of a putative *PELPK1* promoter fragment during development and in response to various stress factors, and (ii) to analyze the sub-cellular localization of the PELPK1.

2. MATERIALS AND METHODS

Extraction of genomic DNA

Genomic DNA was extracted from WT *Arabidopsis* plants (Col) grown under green house conditions (day/night temperatures of 22°/19°C and a light/dark cycle of 16h/8h with approximate photosynthetic photon flux density of 160 μmol m^{-2} s^{-1}) following Doyle and Doyle (1987) with modifications. 4-6 leaves of rosette plants were ground into powder using liquid nitrogen, mixed with 7mL of extraction buffer containing 1M Tris-HCl (pH 8.0), 0.25M NaCl, 5% (w/v) CTAB, 0.2% β-mercaptoethanol, and then incubated at 65°C for 30-60 min. After adding an equal volume of chloroform: isoamyl alcohol (24:1), the mixture was centrifuged at 10,000x g for 20 min. The upper aqueous phase was carefully removed and the DNA was precipitated by adding ½ x volume of cold isopropanol. The solution was centrifuged at 10,000x g for 15 min and the pellet was washed by adding wash buffer containing 76% ethanol, and 10 mM NH_4OAc. The mixture was centrifuged again as above, and the pellet containing DNA was air dried, and suspended in 0.5mL of TE buffer (pH 8.0). The DNA preparation was then treated with RNase A (10 μg/mL), incubated at 37°C for 30 min, precipitated again as above, air dried, and finally resuspended in 0.2 mL TE buffer. DNA concentration was determined by a NanoDrop spectrophotometer (NanoDrop ND 1000 V3 7.1).

Plasmid construction

Promoter::GUS fusion construct

For the construction of a *PELPK1*-promoter::GUS fusion construct, the binary vector, pCAMBIA-1391Z (www.cambia.org, AF234312) containing a promoter-less GUS gene interrupted by a catalase intron, and a double CaMV35S-driven hygromycin B phosphotransferase (hpt) gene in the T-DNA region was used. The double CaMV35S promoter was deleted to eliminate the influence of this promoter, which was positioned in

a reverse orientation, on the nearby *PELPK1* promoter to be investigated. This was accomplished by excision of the *Bst*XI - *Xho*I fragment of 1881 bp containing the CaMV35S- hpt-CaMV35S Terminator fragment from pCAMBIA1391Z. The resulting vector was blunted by T4 DNA polymerase (Fermentas) and self-ligated to obtain the modified vector, p0381Z.

The putative promoter fragment consisting of 1.085 kb upstream intergenic sequence from the translational start site of the *PELPK1* (AT5G09530: 79 bp 5'UTR + 944 bp upstream intergenic sequence + 62 bp 3'UTR sequence of the immediate upstream gene, AT5G09540) was amplified by PCR using WT genomic DNA as the template. The PCR cycling program comprised the following: 94°C for 2 min, followed by 35 cycles involving 94°C for 30 sec, 55°C for 20 sec, 72°C for 40 sec, and then final extension at 72°C for 10 min. The primer pair, designed by incorporating *Hin*dIII and *Nco*I restriction sites to the forward (Fwd) and reverse (Rvs) primers, respectively, were as follows: 5'-CCC-AAGCTT-GGAAACTGACCTAATT-3' (Fwd), and 5'-CATG-CCATGG-GTTTGAGCTTGCTTGA-3' (Rvs). Following gel purifications (Promega) of the resulting PCR product as well as the mini-prep of the newly constructed binary vector, p0381Z, the promoter fragment and the binary vector were digested with *Hin*dIII and *Nco*I enzymes. The ligation reaction for the insertion of the promoter fragment into the above vector was performed using T4 DNA ligase (New England BioLabs).

CDS::GFP translational fusion construct

To study sub-cellular localization of the PELPK1 protein, a CDS::GFP translational fusion construct under the control of a 2x CaMV35S promoter was developed as follows. A binary vector, pCsGFPBT (Acc # DQ370426) was first digested with *Nco*I, then dephosphorylated with TSAP (thermo-sensitive alkaline phosphatase) to prevent religation using Promega protocol II (sequential digestion and

dephosphorylation), and then purified with a PCR purification kit (Qiagen). To amplify the coding region of *PELPK1*, primers were designed by incorporating *Nco*I restriction sites into both forward and reverse primers as follows:
5'-CATGCCATGGCACTAATGAAGAAGAGTCTCTCTGC-3' (Fwd); 5'-CATGCCATGGCAGCTCCACCTCCACCTCCAGGAGTTCCGGGAACTTTTGGGAATTCCG-3' (Rvs). The amplified CDS fragments from WT genomic DNA were purified using a Qiagen PCR purification kit, restricted by *Nco*I, purified again, and then cloned into *Nco*I digested pCsGFPBT vector using T4 DNA ligase and the Promega protocol I.

***Agrobacterium* transformation**

Agrobacterium tumefaciens (GV3101) competent cells were prepared by inoculating LB medium (200 mL) with *Agrobacterium* cells and incubating overnight at 37°C with constant shaking at 200 rpm. The cells were harvested by centrifugation at 5000x g for 10 min at room temperature. The pellet was washed once with sterile TE buffer (pH 8.0), resuspended in the same buffer, frozen in liquid nitrogen, and then stored at -80°C in 0.2 mL aliquots.

Transformation of competent cells with plasmid constructs was carried out using a freeze-thaw method (Weigel and Glazebrook, 2002). Cells were thawed on ice, incubated with the plasmid construct (10 µL) for 5 min on ice, transferred to liquid nitrogen for 5 min, incubated in a 37°C water bath for 5 min, added 1 mL of LB medium, and then put on a shaker at room temperature for 4 hrs. The cells were spread on LB plates containing 50µg/mL kanamycin, 50µg/mL gentamycin, and 15µg/mL rifampicin and incubated at 28°C for two days with constant shaking at 200 rpm. Positive clones were identified by colony PCR using either promoter-specific primers for the cells transformed with the promoter::GUS fusion construct, and CDS-specific primers for the cells transformed with CDS::GFP translational fusion construct. The PCR cycling

conditions were the same as mentioned above except that the initial denaturation time was 10 min instead of 2 min.

Arabidopsis transformation

Arabidopsis thaliana plants were transformed following a modified floral dip method (Martinez-Trujllo et al., 2004). WT *Arabidopsis* (Col) seeds were sown in pots containing wet potting soil (Sunshine mix, Sungro), incubated in the dark at 4°C for two days, and then transferred to the greenhouse containing the conditions as described above. After 10 days, the pots were thinned to keep five plants in each pot.

Because the plant selectable marker (*hpt*) was excised from the binary vector used for developing promoter::GUS fusion construct, a co-transformation protocol was used. *Agrobacterium* cells, harboring either the promoter::GUS fusion construct or a pCAMBIA1300 vector containing only CaMV35S-hpt-Ter within T-DNA region, were cultured separately to an OD_{600} of ~0.6. The two cultures were combined in a ratio of 2:1 (v/v), centrifuged, and the cells were resuspended in infiltration medium (½x MS, 5% sucrose (pH 5.75 by KOH), and 0.5µL Silwett L-77 added just before use) for plant transformation. The CDS::GFP translational fusion construct was transformed using a similar method, except that a single plasmid, rather than co-transformation, was used.

Selection of transgenic lines

Bulked T_1 seeds, collected from T_0 plants, were surface-sterilized by 50% (v/v) commercial bleach (Javex, Clorox), cold incubated at 4°C for two days in the dark, and then sown in petri-dishes (100 x15 mm) containing ½ x MS medium supplemented with 50 µg/mL hygromycin B (Sigma-Aldrich). After 8 -10 days, putative T_1 seedlings were transplanted to the pots containing potting medium as mentioned above. T_2 seeds, collected separately from each T_1 plant were sown in petri-dishes as above. They were

transplanted to the pots, and the T_3 seeds from each plant were collected separately and used for further experimentation.

For the confirmation of transgene constructs, genomic DNA isolated separately from the leaves of T_3 rosette plants harboring either the promoter::GUS fusion construct or the CDS::GFP translational fusion construct, was used as a template in PCR with either a promoter-specific forward (described above), and a GUS gene-specific reverse primer (5'-TGCCCAACCTTTCGGTATAA-3') for promoter::GUS fusion lines, or a vector-specific forward (5'-CGAATCTCAAGCAATCAAGC-3') and a CDS-specific reverse primer (5'-CATGCCATGGCAGCTCCACCTCCACCTCCAGGAGTTCCGGGAACTTTTGGGAATTCCG-3') for the CDS::GFP translational fusion lines.

Histochemical GUS staining

GUS staining was conducted following the protocol described by Salinas and Sanchez-Serrano (2006). T_3 and WT plant tissues were either untreated or treated with various stress factors and were subjected to GUS staining at different stages of growth and development including dry seeds, imbibed seeds, imbibed seeds at different stages of germination, seedlings, rosette plants, stems, inflorescence, and developing, immature, and mature siliques. The tissues were separately incubated at 37°C in the staining solution containing 0.1M sodium phosphate (pH 7.2), 2 mM $K_4Fe(CN)_6$, 2 mM $K_3Fe(CN)_6$, 0.2% Triton X-100, 10 mM EDTA, 2 mM X-Gluc (Rose Scientific) for 4 to 24 h depending on the softness of the tissues. Following GUS staining, the tissues were transferred to 70% ethanol (v/v) and incubated for 2 to 4 h with gentle shaking at room temperature to remove chlorophyll. They were finally fixed in FAA (formalin 5%: acetic acid 5%: ethanol 50%) solution.

Stress treatments

Abiotic stress treatments

T_3 seeds of independent transgenic lines bearing the promoter::GUS fusion construct as well as WT seeds, were surface sterilized and cold-incubated as described above. They were then sown in petri-dishes containing ½ x MS medium supplemented with different concentrations of either sucrose (0, 50, 100, 150, 200 mM), or NaCl (0, 50, 100, 150 mM). Following germination, the seedlings of transgenic and WT plants were up-rooted from the phytablend plate and subjected to histochemical GUS staining at different time intervals.

Biotic stress treatments

Pathogen infection in MS medium**:** T_3 sterilized seeds sown in ½ X MS medium in the presence of 50 mM sucrose were transferred to room temperature following 2 days of cold incubation as mentioned above. After germination, the plants were cocultivated with *Pseudomonas syringae,* grown and suspended as described previously (Yucel et al., 1989), for 24-48 hrs at different stages of growth before subjected to GUS staining. Cocultivation was performed by applying the bacterial suspension (OD_{660} ~ 0.1) at the hypocotyl region of the seedlings on the surface of the MS media. T_3 and WT plants grown in similar plates but without pathogen cocultivation were used as control.

Pathogen infection in the soil: T_3 seeds were sown in sterile pots filled with moist double sterilized soil and transferred to growth chamber following two days of cold incubation as described earlier. The soil was then inoculated with the suspension of either *Pseudomonas syringae* (OD_{660} ~0.5) or *Pythium irregulare*, and maintained in moist conditions throughout the experimental duration. Similar pots without pathogen

inoculation were used as controls. Plants from pathogen inoculated and un-inoculated soils were subjected to GUS staining at different stages of growth.

Hormone treatments

The experimental procedure was similar to the one as described above for abiotic stress treatment. The ½ X MS medium containing 50 mM sucrose was supplemented by different concentrations of hormones as follows: IAA (0, 10, 20, 50 μM), GA_3 (0, 10, 30, 50, 100 μM), ABA (0, 0.1, 0.3, 0.5, 1, 10 μM), MeJa (0, 1, 5, 10, 20, 50 μM), salicylic acid (0, 0.1, 0.2, 0.5, 1 mM). The plants were subjected to GUS staining as above.

Mechanical wounding

Mechanical wounding was carried out using the following two methods. (i) T_3 stems and leaves of soil grown plants were excised with a sterilized scissor while the plant organs were attached with the main shoot (in plant wounding). After 5 days, the wounded tissues were cut off from the main shoot and subjected to GUS staining. (ii) As described in Nishiuchi et al., (1997) with modifications. Intact stems and leaves of soil grown plants were cut into 1cm long sections using a sterile blade, and immediately soaked into 50 mM sodium phosphate buffer (pH 7.0). They were then transferred to petri-dishes lined with four layers of filter paper (Whatman # 1) soaked with the above buffer, exposed to continuous light at room temperature, and subjected to GUS staining at different time intervals.

Microtomy

Plant samples were first fixed in a FAA (Formalin-Aceto-Alcohol) solution containing 5% (v/v) formalin, 5% (v/v) acetic acid and 50% (v/v) ethanol. They were then

dehydrated following the procedure as described by Disbrey and Rack (1970) with modifications as follows: transferred to 70% ethanol for (1+1) 2 hrs, 100% ethanol for 1hr, ethanol: toluene 1:1 for 1hr, tolune-1 for ½ hr, tolune-2 for ½ hr, paraffin-1 for 2 hrs, paraffin-2 for 3 hrs, and then embedded in paraffin blocks. The blocks were subjected to microtome (AO-820) sectioning to obtain slices of 8-12 µm thickness. The sections were put on slides, incubated overnight at 37°C to dry and attach to the slides, and then processed to stain with safranin as follows: incubated in toluene for 10 min, 100% ethanol for 10 min, 95% ethanol for 2 min, and in safranin O Stain for 30 min. The sections were then subjected to washing by 100% ethanol for 2 min, toluene (2 + 2) for 4 min and then mounted by adding DPX.

Microscopy

Tissue samples were viewed and photographed either using a Leica dissecting microscope (Leica Microsystems) equipped with a digital camera, or by a Olympus BX51 microscope (Olympus Optical Co.) equipped with a Photometrix CoolSnap fx digital camera (Roper Scientific) and a MicroColor liquid crystal tunable RGB filter (Cambridge Research & Instrumentation, Inc.).

GFP localization

Confocal microscopy

T_3 plants bearing CDS::GFP translational fusion construct were subjected to both epifluorescent as well as confocal-laser scanning microscopy at different stages of growth. For epifluorescent microscopy, the samples were observed using an Olympus BX51 microscope as described above. For confocal-laser scanning microscopy, a Leica DM IRBE microscope equipped with TCS-SP laser scanning module was used. GFP was visualized with an Argon laser at 25% power output (488 nm excitation line) and

signal was detected at an emission wavelength between 505 to 530 nm. Images were captured at a fixed resolution of 1024 x 1024 and later processed with Adobe Photoshop CS3. Hydrated samples of germinated seeds, young seedlings as well as various plant parts including, seed coats of germinated seeds, cross-sections of stem, petiole, flower, and longitudinal sections of siliques were examined to observe GFP emission. Plasmolysis of stem sections was performed using 0.8M sorbitol.

3. RESULTS

For promoter activity analysis of the *PELPK1* gene, an upstream 1.085kb genomic fragment from translational start site including a predicted 5'UTR (79 bp), upstream intergenic sequence (944 bp), and part (62 bp) of the predicted 3'UTR of the immediate upstream gene (AT5G09540) were ligated to the reporter gene, ß-glucuronidase, GUS (Figure 3-1) following Jefferson et al., (1987). To help describe the expression pattern of the *PELPK1*, at least 10 independent T_3 lines harboring the promoter::GUS fusion construct were examined. The transgenic plants were produced using a co-transformation strategy so that the constitutive CaMV35S promoter driving selectable marker expression would not interfere with the GUS reporter gene.

PELPK1 promoter activity during normal development of plants

It was observed that in plant tissues, in the absence of stress or hormone treatments, the GUS reporter gene was strongly expressed only in the aleurone layer of the seed coat of germinating seed (Figure 3-2). The expression was mostly localized in the micropylar region through which the radicle emerged from the germinating seed (Figure 3-2 B to E). No expression was detected in the testa of the germinated seed (Figure 3-2) or in the young seedling including cotyledons, hypocotyl, and root (Figure 3-2F).

Further analysis of the seed coat aleurone layer of germinated seed by microtome sectioning (Figure 3-3), and by tissue maceration (Figure 3-4) confirmed that the GUS reporter was specifically expressed in the cells of the aleurone layer. Blue, needle-shaped GUS precipitates were observed within the aleurone cells (Figure 3-3 B & C). In the macerated intact aleurone cells, spindle-shaped GUS crystals were observed intermingled with the presumed protein storage vacuoles, PSVs, starch

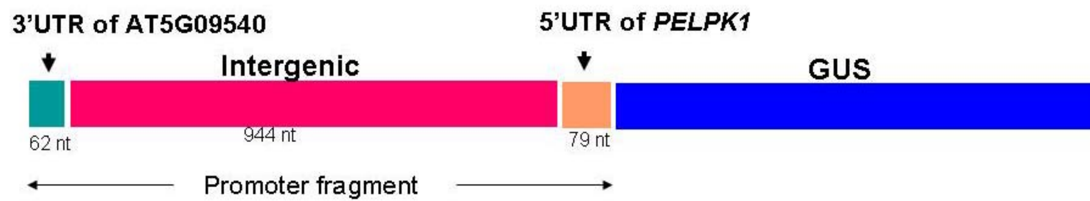


Figure 3-1. The *PELPK1*-promoter::GUS-reporter fusion construct showing upstream 1.085 kb fragment containing putative cis-regulatory elements. More than ten T3 trasgenic lines were developed harboring the above construct.

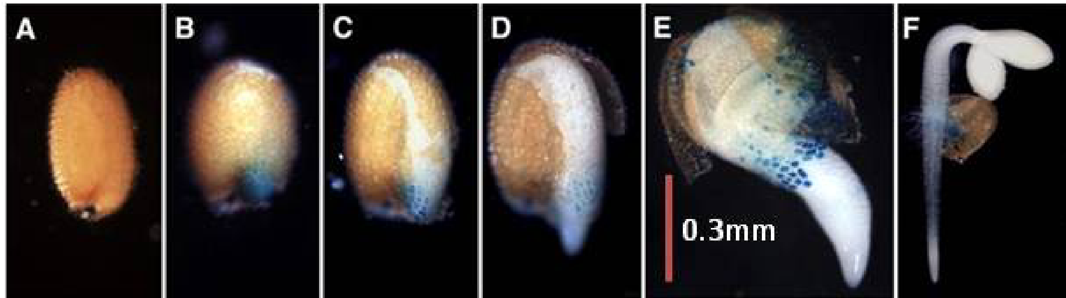


Figure 3-2: Normal GUS expression pattern in T3 seed harboring a promoter::GUS fusion construct. **A,** a seed just before imbibition showing no GUS expression; **B**, a germinating seed with protruding radicle showing GUS expression at the micropylar end; **C**, a germinating seed with fractured seed coat (testa) showing GUS expression in the inner cell layer at the micropylar end surrounding the embryo; **D**, a germinating seed with partially emerged radicle showing GUS expression in the aleurore layer of micropylar end; **E**, a germinating seed with completely emerged radicle showing GUS expression in the aleurone layer; **F**, a young seedling showing no GUS expression in the cotyledons, hypocotyl, and root. More than ten independent T3 lines were developed and tested for GUS expression and all of them showed similar pattern of GUS expression.

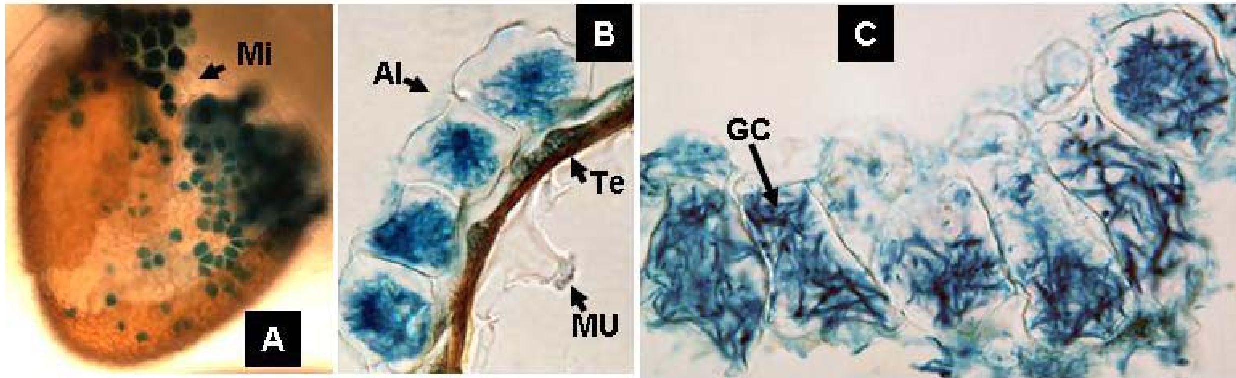


Figure 3-3: Normal GUS expression in the seed coat aleurone layer. **A**, a seed coat following germination showing intense GUS expression at the micropylar region (Mi); **B**, a microtome section through the seed coat showing an outer mucilage layer (Mu), an inner testa (Te), and a single-layered aleurone/endosperm (Al); **C**, a microtome section through an aleurone layer showing needle-shaped GUS crystals (GC) inside the cells. Other conditions are as mentioned in Figure 3-2 legend.

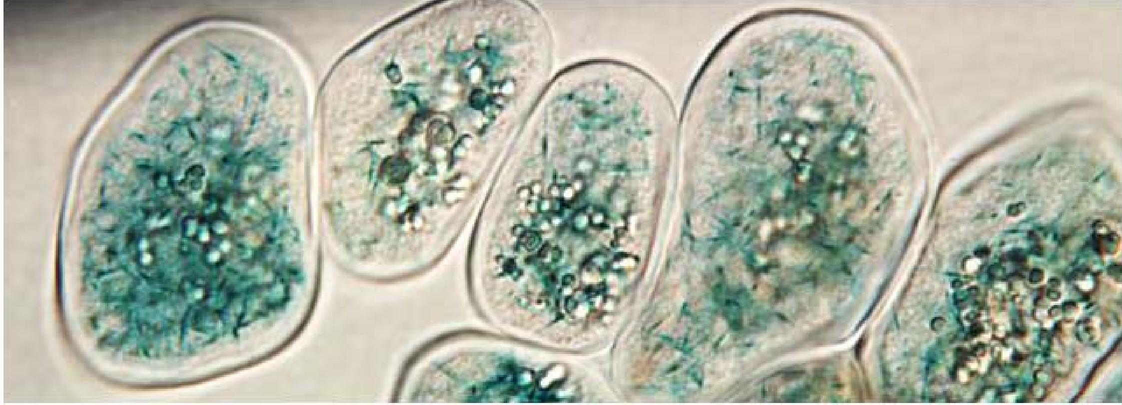


Figure 3-4: Aleurone/endosperm cells macerated out from the seed coat aleurone layer are showing GUS crystals inside the cells along with PSVs, starch granules, and lipid bodies. These entities are the characteristics of aleurone cells.

granules, and lipid bodies (Morrison el al. 1978; Muntz, 1998; Bethke et al., 2007), (Figure 3-4).

Under normal growing conditions, GUS reporter expression was not detectable in any other plant parts/tissues tested including seed coat of ungerminated seeds, developing embryos, cotyledons, hypocotyl, young seedlings, rosette plants, inflorescence-shoots, flowers, immature siliques and developing seeds, and mature siliques and mature seeds (data not shown).

PELPK1 promoter inducibility

Database analyses for microarray-based expression profiles of *PELPK1* showed that this gene is induced by a variety of factors including abiotic, biotic, elicitors, nutritional, hormonal, defense, and other chemical factors (Chapter 2). To determine the inducibility of the *PELPK1* promoter, experiments were performed by treating T3 plants with some of the common stress-inducing factors and hormones as described below.

Abiotic factors - because *PELPK1* has been reported to be induced by osmotic and salt stresses and ABA in microarray experiments (Seki et al., 2002), these factors were tested for GUS reporter expression by treating T3 plants in MS medium supplemented with sucrose (100-200 mM), NaCl (50-150 mM) and ABA (1 - 10 μM) for up to 3 weeks as described above. These plants did not show any GUS expression following GUS staining (data not shown). Similarly, plants grown under the above conditions but supplemented with GA_3 or IAA as described above did not show any evidence of GUS expression (data not shown).

Biotic factors – database analysis (Chapter 2) showed that the *PELPK1* was maximally induced by the bacterial pathogen, *Pseudomonas syringae.* In the present investigation,

disease-inducibility of the *PELPK1* promoter was tested by subjecting the T_3 plants to the bacterium *Pseudomonas syringae* and the oomycete, *Pythium irregulare,* as described above. Visual analysis of the disease symptoms observed in the soil-grown adult plants showed that *P. syringae* was more active in causing pathogenecity in *Arabidopsis* plants than *P. irregulare* (Figure 3-5).

Histochemical GUS staining of plants either cocultivated in MS-agar medium or grown in the soil inoculated with the *P. syringae* showed that in both growing conditions, this pathogen induced intense GUS expression. Since the expression pattern in both conditions was very similar, only the expression pattern in the soil-grown plants is presented (Figure 3-6). Figure 3-6B showed that *P. syringae*-induced GUS expression was observed both in the roots as well as in younger and older leaves. In the older leaves, the expression was prominent in the veins, lamina, and the petioles. Similar staining of plants grown in *P. irregulare*-inoculated soil (Figure 3-6C) showed that the expression was mostly localized in the main veins and was relatively less intense compared to plants grown in *P. syringae* inoculated soil (Figure 3-6B). Furthermore, in the MS-agar medium, *P.syringae* induced GUS expression was noticed in less than 12 hrs of cocultivation (not presented). In the soil conditions also, the plants grown in the *P.syrangae* inoculated soil exhibited GUS expression much earlier than that of the plants grown in *Pythium irregulare* inoculated soil.

Defense chemicals – experiments were conducted in MS-agar medium supplemented with different concentrations of either MeJa or SA as described above. Since both defense hormones similarly induced GUS expression in the leaves of T3 plants, only MeJa-induced GUS expression pattern is presented (Figure 3-7). The expression was predominantly observed in the leaf veins (Figure 3-7 B&C). No reporter expression was observed in the roots (Figure 3-7B) although the hormones were added to the growth

medium. Anatomical analysis of the microtome sections through the leaf veins showed that the GUS precipitates were deposited into the vascular xylem tissues (Figure 3-7C).

Mechanical wounding - T_3 plants grown in MS-agar medium in the presence of 50 mM sucrose as well as in double sterilized soil as mentioned above were used to determine the inducibility of the *PELPK1* promoter by mechanical wounding caused by stem and leaf sectioning as described above. Both types of injuries induced GUS expression in plants grown under both conditions. Only the results of soil grown plants are presented here (Figure 3-8). In both stem and leaf sections, GUS expression was observed in <12 hrs after sectioning. In the stem sections, the expression was noticed close to the cut sites in the form of rings (Figure 3-8A). In the leaf sections, it was induced at the cut sites and then appeared to spread throughout the vein networks (Figure 3-8B). In plant wounding by leaf sectioning as mentioned above also induced similar GUS expression (Figure 3-8C. GUS reporter was also induced by leaf-tip and leaf-margin burning perhaps due to nutritional deficiencies (Figure 3-8D).

Computational analysis of the promoter fragment

To identify conserved cis-regulatory elements present in the upstream genomic fragment used in the development of promoter::GUS fusion construct (Figure 3-1), promoter analysis was carried out using PlantCare (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/; Lescot et al., 2002). Both positive and negative cis-acting DNA elements may be needed for normal spatial and temporal regulation of gene expression (Butos et al., 1991). Putative cis-acting elements detected in the *PELPK1* upstream sequence are presented (Table 3-1). Analysis of these results showed that among other regulatory elements, the *PELPK1* upstream sequence contained: two endosperm-specific regulatory elements (TGTGTCA,

GTCAT), one of which (GTCAT) is repeated three times in the sequence; one defense and stress-responsive (GTTTTCTTAC), one salicylic acid-responsive (CCATCTTTTT), one ABA-responsive (TACGTG), one auxin- responsive (AACGAC), and one light responsive (AATCTAATCT) element (Lescot et al., 2002). There were no osmotic- and/or salt-stress specific regulatory elements present in this promoter fragment (Table 3-1)

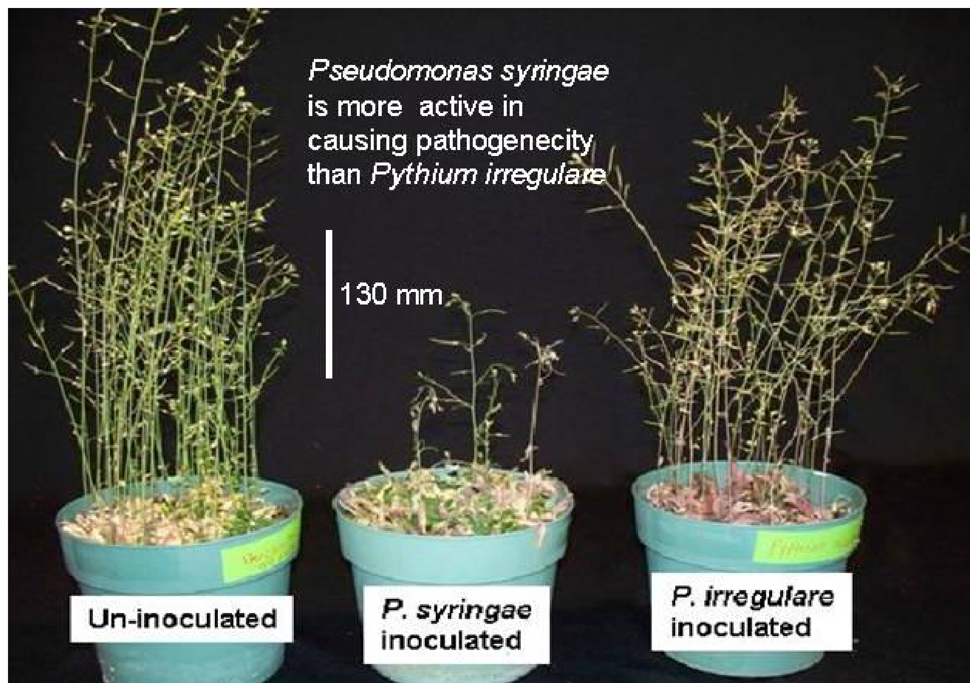


Figure 3-5. Disease symptoms in T3 plants harboring a *PELPK1*-ptomoter::GUS-reporter fusion construct induced by pathogens in inoculated soil. The pathogens were inoculated into the double sterilized soil. Plants grown in uninoculated soil were considered as controls. Three independent T3 lines were tested in the pathogen-inoculated soil, and all of them exhibited similar disease symptoms. The photograph was taken 30 days after emergence of the seedlings from the soil. The size of the pots = 180 x 130 mm.

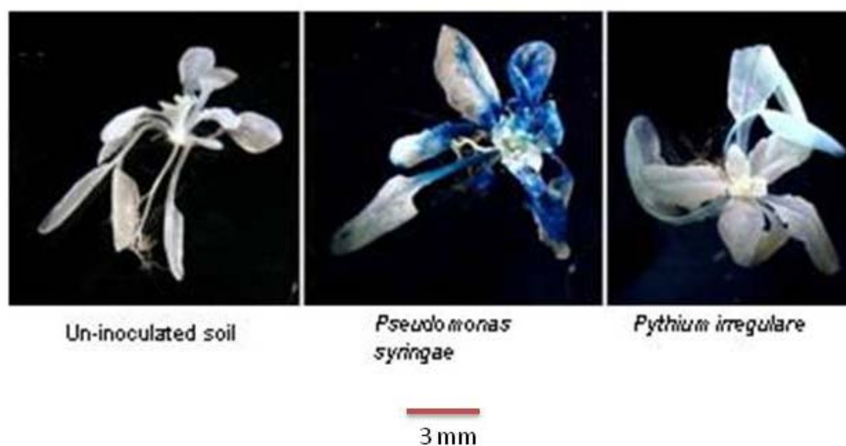


Figure 3-6. GUS expression in T3 rosette plants grown in the soil: **A**, a plant grown in un-inoculated soil showing no GUS expression; **B,** a plant grown in *Pseudomonas syringae* inoculated soil showing GUS expression in (veins, lamina, and petioles) both younger and older leaves; **C,** a plant grown in *Pythium irregulare* inoculated soil showing GUS expression in the main leaf veins. Plants were subjected to GUS staining at rosette stage. Other conditions are as mentioned in Figure 3-7 legend.

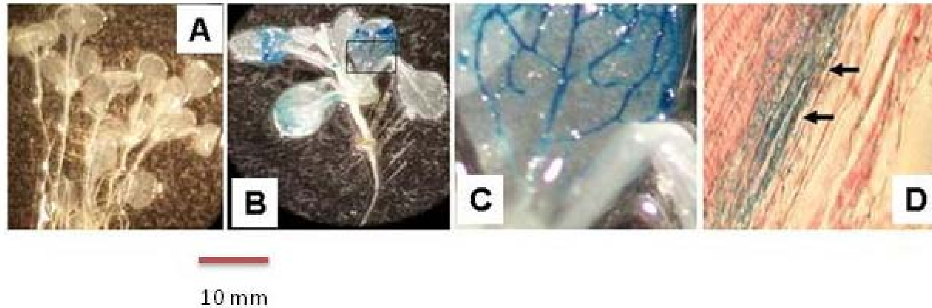


Figure 3-7: GUS expression in T_3 plants grown in MS-agar medium: **A**, seedlings grown in the absence of MeJa showing no GUS expression; **B**: a seedling grown in MeJa-supplemented medium showing GUS expression in leaf lamina and veins; **C**, an enlarged area of the rectangular block of “B”; **D**, a microtome section through a GUS stained vein showing deposition of blue GUS precipitates in the xylem tissue (black arrows). 3 to 5 independent T3 lines were tested and all of them showed similar pattern of GUS exprerssion.

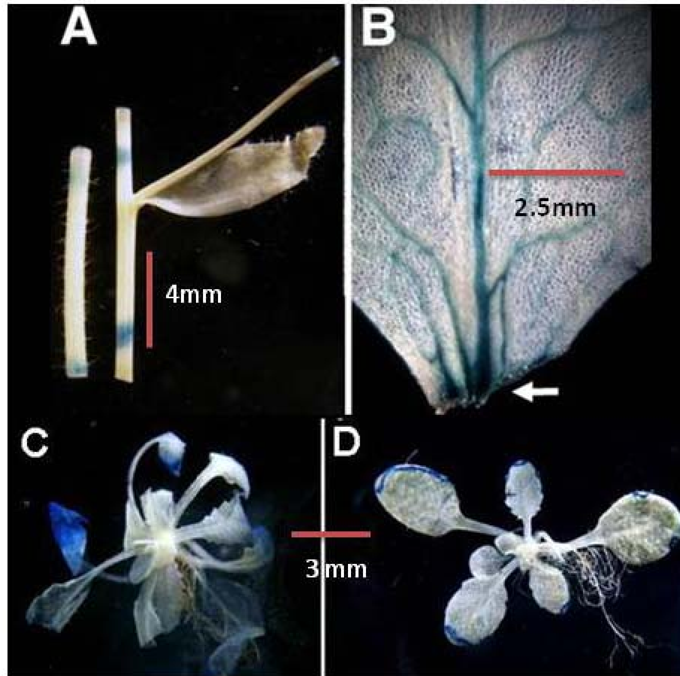


Figure 3-8. GUS expression in T3 plants induced by mechanical wounding: **A**: stem sections showing GUS expression in the form of rings above and below the cut ends; **B**, a leaf showing GUS expression at the cut site (white arrow) as well as in the veins, **C**, an in plant leaf sectining/cutting showing GUS expression at the cut sites; **D**, a plant showing GUS expression in the leaf-tips and margins cuased by natural tip/margin burnings. Three independent T3 lines were tested, and all of them exhibited similar pattern of GUS expression.

Table 3-1: Putative cis-regulatory sequences found in the 1.085 kb upstream region of the *PELPK1* gene used for the construction of promoter::GUS fusion construct

Motif	Sequence	Pos	Star	Score	Function
5UTR Pie-rich stretch	TTTCTTCTCT	1040	-	9	cis-acting element (CAE) conferring high transcription
AAGAA-motif	GAAAGAA	465	+	7	Unknown
ABRE	TACGTG	802	+	6	CAE involved in ABA response
		823	-	6	
AE-box	AGAAACAA	1049	-	8	Part of a module for light response
ARE	TGGTTT	130	+	6	Cis-regulatory element (CRE) for anaerobic induction
ATCT-motif	AATCTAATCT	236	-	9	Light response
CAAT-box	CAAT, CAATT, CAAAT	39, 638, 435, 744, 349, 737, 495, 855, 338, 675 448, 758, 388, 738, 565, 866		4-5	Promoter and enhancer elements
CATT-motif	GCATTC	185	-	6	Light response
CTAG-motif	ACTAGCAGAA	256	+	10	Unknown
G-Box	CACGTA	802	-	6	Light response
	CACGTA	823	+	6	
GCN4-motif	**TGTGTCA**	**747**	**+**	**7**	**CRE endosperm**
LTR	CCGAAA	84	-	6	CRE low temp.
		839	+	6	
MRE	AACCTAA	527	+	7	MYB for light
Skn-1-motif	**GTCAT**	**452**	**-**	**5**	**CRE endosperm**
		750	**+**	**5**	**CRE endosperm**
		648	**-**	**5**	**CRE endosperm**
TATA-box	TATA, TATAAA, TATAA TATATAA, taTATAAAgg TATTTAAA	90, 92, 109, 191, 346, 716, 723, 724, 731, 732, 734, 869, 875		4-9	Core promoter elements
TC-rich repeats	**GTTTTCTTAC**	**362**	**+**	**9**	**CAE defense and stress**
TCA-element	**CCATCTTTTT**	**944**	**+**	**9**	**CAE salicylic acid**
TCCACCT-motif	TCCACCT	277	-	7	unknown
TGA-element	AACGAC	503	-	6	Auxin-responsive
Unnamed-4	CTCC	428	-	4	Unknown
	CTCC	943	+	4	
Circadian	CAANNNNATC	23	-		Circadian control

Analysis of sub-cellular GFP localization

To investigate the sub-cellular localization of the PELPK1 protein, a translational construct was produced by the fusion of the coding region of the PELPK1 with the GFP reporter gene, and was expressed under the control of the 2xCaMV35S promoter (Figure 3-9). Because the PELPK1 contained a signal peptide at the N-terminal end of its coding sequence (Chapter 2), GFP was fused to the C-terminal of the PELPK1, with a peptide linker (Gly-Gly-Gly-Gly-Ala) (Chiu et al., 1996) separating the PELPK1 and GFP sequences. The T_3 plants harboring the above construct were examined at different stages of growth for localization of the PELPK1::GFP translational fusion protein.

Confocal laser-scanning microscopy showed two patterns of GFP localization in the transgenic tissues: (i) punctate aggregates within the aleurone of germinated seeds (Figure 3-10), and (ii) uniform expression in cell walls of other tissues including secondary walls of stem xylem tissues, mature siliques (Figure 3-11), and roots (not shown).

Figure 3-9. A schematic diagram of the *PELPK1*-coding sequence::green fluorescent protein (GFP) translational fusion construct. L = peptide linker. About ten T3 trasgenic lines were developed harboring the above construct.

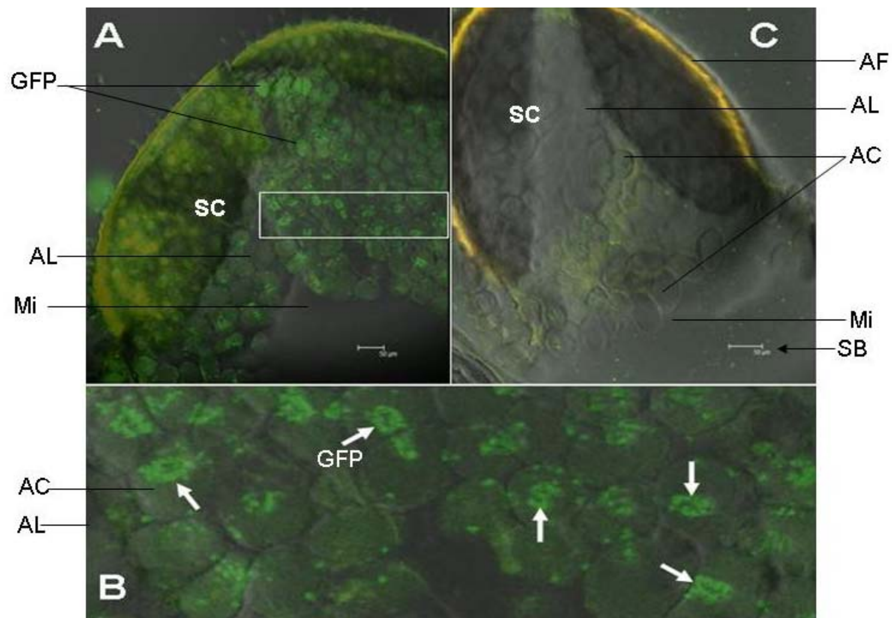


Figure 3-10: GFP expression in the seed coat of a germinated T3 seed harboring a CDS::GFP translational fusion construct as determined by confocal laser-scanning microscopy. **A**: a transgenic seed coat showing the super-imposition of fluorescence and DIC (differential Interference Contract) images, **B**: a magnified image of the rectangular block from "A" showing an aleurone layer with punctate aggregates of the fusion protein (infront of the white arrows) deposited into the aleurone cells, C: a WT seed coat showing an overlay of fluorescence and DIC images as in "A". (SC = seed coat, AF = autofluorescence, AL = aleurone layer, AC= aleurone cells, Mi = micropylar end, SB = scale bar (50 µm). About 10 transgenic lines were developed. At least 3 lines were tested for GFP expression.

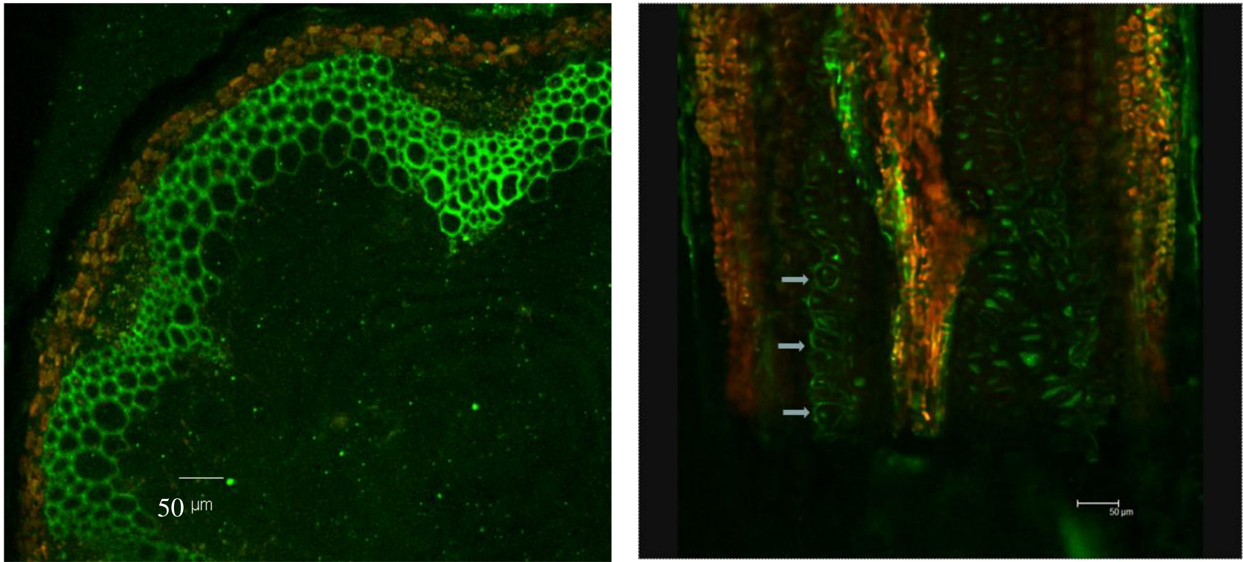


Figure 3-11: GFP emission in the cell walls of T3 transgenic tissues of *Arabidopsis* as determined by confocal laser-scanning microscopy. Figures are showing the super-imposition of fluorescence and autofluorescence images. Figure on the left is a stem cross section showing GFP expression in the cell walls of xylem tissues. Figure on the right is a longitudinal section of a silique showing GFP expression in the cell wall (in front of the white arrow heads). The horizontal scale bars in both photographs indicate 50µm. The tissues were also subjected to plasmolysis by 0.8M sorbitol.

4. DISCUSSION

GUS reporter expression

Histochemical analysis of GUS expression in the transgenic lines bearing a *PELPK1* promoter::GUS fusion construct suggests that *PELPK1* plays at least two apparently unrelated roles in *Arabidopsis* plants: (i) a presumably developmental role particularly during seed germination, and (ii) a general defense-related role.

A developmental role of *PELPK1* during normal development is suggested based on the following observations: (i) GUS reporter gene was spontaneously expressed only in the seed coat of germinating seed, particularly around the micropylar region of the seed coat through which radicle emerges during seed germination (Figure 3-2). Seed germination in *Arabidopsis* begins with imbibition and is completed when radicle protruded through seed coat (Begley 1997, Koornneef et al. 2002, Kucera et al. 2005), (ii) GUS reporter was expressed specifically in the aleurone layer of the seed coat (Figure 3-3), which is composed of a single layer of endosperm cells (Pritchard et al. 2002, Nambara and Marion-Poll 2003, Liu et al. 2005), that play important roles by providing nutrition to the germinating embryo (Vaughan et al. 1971, Corner 1976, Ruiz and Escale 1995, Nguyen et al. 2000; Muller et al., 2006). Computational analysis of the promoter sequence used in the development of promoter-reporter fusion construct showed that it contained two endosperm-specific regulatory elements, one of which was repeated three times in the promoter sequence used (Table 3-1). Thus, the spontaneous expression of the GUS reporter gene in the aleurone layer of germinating seeds corresponded well with the presence of highest number of endosperm-specific regulatory elements in the *PELPK1* promoter sequence tested (Figure 3-1).

From the above discussion, although it is proposed that the *PELPK1* might play a developmental role during germination of *Arabidopsis* seed, the exact mechanism of its involvement is unclear. It has been suggested that during seed germination, the embryo

produces gibberellins that trigger the expression of genes in the aleurone/ endosperm cells, which synthesize enzymes for the hydrolysis of starch and storage proteins to supply nutrition to germinating embryos (Debeaujon and Koornneef, 2000; Groot and Karssen, 1987; Groot et al., 1988). Since *PELPK1* has been annotated to encode a HRGP-family protein (TAIR) with no predicted catalytic activity (Cassab 1998), it is assumed that it may play a role during seed germination that is non-catalytic in nature. Because *PELPK1* is normally expressed in the aleurone layer around the micropylar region, where the aleurone cells become more active during seed germination to generate energy for germinating embryo, it might be possible that PELPK1 acts as a seed storage protein that undergoes hydrolysis by an unknown catalytic enzyme and provides nutrition to germinating embryo that indirectly helps cell wall loosening and embryo growth. Alternatively, it can also be possible that during radicle protrusion the germinating seeds generate reactive oxygen species (ROS) that may lead to oxidative cross-linking of PELPK1 in the cell wall to protect germinating embryo from pathogenic infection. However, this possibility appears to be less likely as protein cross-linking with the cell wall generally reduces cell wall extension and vegetative growth (Showalter 1993; Cassab 1998).

Experiments involving the inducibility of the *PELPK1* promoter showed that it is not induced by any of the common abiotic stress inducing factors tested including osmotic and salt stresses (data not presented). Computational analysis of the promoter sequence of the *PELPK1* gene used for constructing promoter::GUS fusion construct confirmed that it does not contain any regulatory elements that are responsive to the above abiotic stress factors (Table 3-1). However, further experiments showed that the GUS reporter was strongly induced by pathogen infection (Figure 3-6), defense chemicals (Figure 3-7), and mechanical wounding (Figure 3-8). Computational analysis of the *PELPK1* promoter sequence used confirmed that it contains regulatory elements

(scores: 9 out of 9) that are responsive to the above defense-related factors (Table 3-1). Induction of CWPs including extensins by the above mentioned factors was also reported earlier (Showalter 1993; Cassab 1998); Merkouropoulos et al., 1999; Reymond et al., 1998). It has been suggested that the CWPs such as extensins and PRPs make the cell wall impenetrable to pathogens by insolubilization and oxidative cross- linkages with the cell wall (Showalter 1993; Basavaraju et al., 2009). AGPs have been suggested to form gel plug against pathogen invasion (Narayanasamy, 2006).

Protein localization

Bioinformatics analysis demonstrated that the predicted *PELPK1*-encoded protein contained repetitive sequence motifs, is rich in Pro residues, contained a secretory pathway-directed signal peptide, and a transmembrane domain at the N-terminal end (Chapter 2). Based on the above characteristics, it was hypothesized that it might be associated with the cell wall in a similar manner as extensins and other CWPs (Cassab 1998, Showalter 1993). Bioinformatics analysis further predicted that the *PELPK1* encoded protein is transported to the extracellular region i.e. cell wall (Chapter 2). As has been discussed below, analysis of GFP emission in the transgenic plants bearing a PELPK1::GFP translational fusion construct under the control of a double CaMV35S promoter appears to be consistent with the above prediction and hypothesis (Figures 3-10 & 3-11).

The sub-cellular localization pattern of the translational fusion was not the same in all tissues. In germinated seeds, for example, punctate aggregates of fusion protein were observed inside cells of the aleurone layer (Figure 3-10). In contrast, in other tissues such as stem and silique, the fusion protein was predominantly localized in the cell wall (Figures 3-11). Plasmolysis of these tissues by high concentration of sorbitol (0.8M) did not change the location of GFP emission as determined confocal microscopy

mentioned above, suggesting that in these tissues the fusion protein was associated with the cell wall. From the above sub-cellular protein localization experiments, it is concluded, in partial agreement with the *PELPK1* bioinformatics predictions (Chapter 2), that the PELPK1::GFP translational fusion protein is deposited to the cell wall in addition to the seed coat aleurone cells.

5. **CONCLUSIONS**

It is concluded based on reporter gene experiments and translational fusions that *PELPK1* is involved in multiple processes in *Arabidopsis* plants, including: a developmental role that may indirectly help embryo growth presumably mediated by cell wall loosening during seed germination, and a general cell wall-based defense- and repair-related role in all tissues. Since the defense would normally be expected to involve hardening of the wall, it is unclear how *PELPK1* might act in both loosening and stiffening of cell walls in different contexts.

5. **REFERENCES**

Begley JD (1997) Seed germination and plant dormancy. Plant Cell 9: 1055-1066

Butenko, M. A., Patterson, S. E., Grini, P. E., Stenvik, G.-E., Amundsen, S. S., Mandal, A. & Aalen, R. B. (2003). Influrescence deficient in abscission controls Floral Organ Abscission in Arabidopsis and Identifies a Novel Family of Putative Ligands in Plants Plant Cell 15, 2296-2307.

Bustos, M M ., Begum, D., Kalkan, FA., Battraw, MJ., and Hall T.C. (1991).Positive and negative cis-acting DNA domains are required for spatial and temporal regulation of gene expression by a seed storage protein promoter. EMBO J. 1991 June; 10(6): 1469–1479.

Cassab, G.I. (1998). Plant cell wall proteins. Annual review of plant physiology and plant molecular biology. 49: 281-309

Chiu, W, Niwa, Y, Zeng, W, Hirano, T, Kobayashi, H, Sheen, J. (1996). Enginnered GFP as a vital reporter in plants. Curr. Biol. 6: 325-330

Corner, E.J.H. (1976) The Seeds of Dicotyledons. da Silva, E.A.A., Toorop, P.E., Nijsse, J., Bewley, J.D. and (Coffea arabica cv. Rubi) seed germination and cause cell death in the embryo. J Exp Bot 413: 1029–1038.

Disbrey, B.D. and Rack, J.H. (1970). Histological laboratory methods. pp. E & S Livingston, Edinburgh & London

Debeaujon, I., and Koornneef, M. (2000). Gibberellin Requirement for Arabidopsis Seed Germination Is Determined Both by Testa Characteristics and Embryonic Abscisic Acid. Plant Physiol. 122: 415–424.

Doyle JJ and Doyle JL (1987) A rapid DNA isolation procedure for small quatities of fresh leaf tissue. Phytochem Bull 19: 11-15.

Dunker, A. K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M., and Obradovic, Z. (2002). Intrinsic disorder and protein function. Biochemistry, 41:6573-6582.

Dunker, AK, Lawson, JD, Brown, CJ, Romero, P, Oh, JS, Oldfield, CJ, Campen, AM, Ratliff, CM, Hipps, KW, Ausio, J, Nissen, MS, Reeves, R, Kang, C, Kissinger, CR, Bailey, RW, Griswold, MD, Chiu, W, Garner EC, and Obradovic, Z. (2001). Intrinsically disordered protein, J. Mol. Graphics Modelling 19: 26–59.

Dyson, HJ, and Wright, PE (2005). Intrinsically unstructured proteins and their functions, Nat. Rev. Mol. Cell Biol. 6: 197–208.

Groot SPC, Karssen CM. (1987). Gibberellins regulate seed germination in tomato by endosperm weakening: a study with gibberellin-deficient mutants. Planta. 171:525–531.

Groot SPC, Kieliszewska-Rokicka B, Vermeer E, Karssen CM. (1988). Gibberellin-induced hydrolysis of endosperm cell walls in gibberellin-deficient tomato seeds prior to radicle protrusion. Planta. 174:500–504.

Huffaker, A., Pearce, G., and Ryan, CA. (2006). An endogenous peptide signal in Arabidopsis activates components if the innate immune response. PNAS. 103: 10098-10103

Jefferson, R.A., Kavanagh, T.A., and Bevan, M.W. (1987) GUS fusions: ß-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. EMBO J. 6, 3901-3907.

Koornneef M, Bentsink L, Hilhorst H (2002) Seed dormancy and germination. Curr Opin Plant Biol 5: 33–36

Kucera, B, Cohen, MA, and Leubner-Metzger, G, (2005). Plant hormone interactions during seed dormancy release and germination. Seed Sci. Res. 15:281-307

Lescot, M., Dephais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., Rouze, P. Rombauts. (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res. 30: 325-327.

Liu, P.-P., Koizuka, N., Homrichhausen, T.M., Hewitt, J.R., Martin, R.C. and Nonogaki, H. (2005) Large-scale screening of Arabidopsis enhancer-trap lines for seed germination-associated genes. Plant J 41: 936–944.

Martinez-Trujillo, M., Limones-Briones, V., Cabrera-Ponce, J.S., and Herrera-Estrella, L. (2004). Improving Transformation Efficiency of Arabidopsis thaliana by Modifying the Floral Dip. Plant Molecular Biology Reporter 22: 63–70

Matsubayashi Y (2003). Ligand-receptor pairs in plant peptide signaling. Cell Sci.116: 3863-70.

Matsubayashi, Y. and Sakagami, Y. (2006). Peptide Hormones in Plants. Annual Review of Plant Biol. 57:649-674

Mergaert, P., Nikovics, K., Kelemen, Z., Maunoury, N., Vaubert, D., Kondorosi, A. & Kondorosi, E. (2003). A Novel Family in Medicago truncatula Consisting of More Than 300 Nodule-Specific Genes Coding for Small, Secreted Polypeptides with Conserved Cysteine Motifs1 Plant Physiol. 132, 161-173.

Merkouropoulos, G., Barnett, D.C. and Shirsat, A.H. (1999) The Arabidopsis extensin gene is developmentally regulated, is induced by wounding, methyl jasmonate, abscisic and salicylic acid, and codes for a protein with unusual motifs. Planta, 208, 212–219.

Müller, K., Tintelnot, S., and Leubner-Metzger, J. (2006). Endosperm-limited Brassicaceae Seed Germination: Abscisic Acid Inhibits Embryo-induced Endosperm Weakening of Lepidium sativum (cress) and Endosperm Rupture of Cress and Arabidopsis thaliana. Plant Cell Physiol. 47: 864–877.

Muntz, K. (1998). Deposition of storage proteins. Plant Mol. Biol. 38, 77–99

Nambara, E. and Marion-Poll, A. (2003) ABA action and interactions in seeds. Trends Plant Sci 8: 213–217.

Narita NN, Moore S, Horiguchi G, Kubo M, Demura T, Fukuda H, Goodrich J, Tsukaya H. (2004). Overexpression of a novel small peptide ROTUNDIFOLIA4 decreases cell proliferation and alters leaf shape in Arabidopsis thaliana. Plant J. 38 :699-713

Narváez-Vásquez J, Pearce G, abd Ryan CA (2005). The plant cell wall matrix harbors a precursor of defense signaling peptides. Proc Natl Acad Sci. 102:12974-129777

Nguyen, H., Brown, R.C. and Lemmon, B.E. (2000) The specialized chalazal endosperm in Arabidopsis thaliana and Lepidium virginicum (Brassicaceae). Protoplasma 212: 99–110.

Nishiuchi, T, Hamada, T, Kodama, H, and Ibal, K (1997). Wounding changes the spatial expression pattern of the *Arabidopsis* plastid w-3 fatty acid desaturase gene (FAD7) through different signal transduction pathways. The Plant Cell 9: 1701-1712

Pearce, G. & Ryan, C. A. (2003). Systemic signaling in tomato plants for defense against herbivores. Isolation and characterization of three novel defense-signaling glycopeptide hormones coded in a single precursor gene. J. Biol. Chem. 278, 30044-30050

Pritchard, S.L., Charlton, W.L., Baker, A. and Graham, I.A. (2002) Germination and storage reserve mobilization are regulated independently in Arabidopsis. Plant J 31: 639–647.

Remy, S, François, I, Cammue, BPA, Swennen, R, and Sági, L. (1997). Co-transformation as a potential tool to create multiple and durable disease

resistance in banana. ISHS Acta Horticulturae 461: International Symposium on Biotechnology of Tropical and Subtropical Species Part 2

Reymond P, Farmer EE (1998). Jasmonate and salicylate as global signals for defense gene expression. Curr Opin Plant Biol. 1998 Oct;1(5):404-11

Ruiz, Z.T. and Escale, M. (1995) Ultramicroscopic seed morphology of Cleome L. (Capparidaceae) in relation to their taxonomy and dispersal syndromes. Ernstia 5: 139–160.

Ryan, CA and Pearce G. (1998). Systemin: A polypeptide signal for plant defense genes. Annual Review Cell Dev. Biol. 14: 1-17

Ryan, C. A. & Pearce, G. (2004). in Plant Hormones: Biosynthesis, Signal Transduction, Action, ed. Davies, P. J. (Springer, Secaucus, NJ) pp. 654-670.

Salinas, J. and Sanchez-Serrano, JJ (2006). Arabidopsis protocols: 2nd Edition Humana Press, NJ, USA (2006)

Sieburth, L.E., and Meyerowitz, E.M. (1997). Molecular dissection of the AGAMOUS control region shows that cis elements for spatial regulation are located intragenically, Plant cell 9, 355-365.

Showalter AM. 1993. Structure and function of plant cell wall proteins. Plant Cell 5:9–23

Stratmann, J.W. (2003) Long distance run in the wound response - jasmonic acid is pulling ahead. Trends in Plant Science 8(6), 247-250.

Stratmann, J.and Ryan, C.A. (1997) Myelin basic protein kinase activity in tomato leaves is induced systemically by wounding and increases in response to systemin and oligosaccharide elicitors. Proc. Natl. Acad. Sci. USA 94, 11085 - 11089

Tylor, C.B. (1997). Promoter fusion analysis: An insufficient measure of gene expression. The Plant Cell. 9:273-275

Vaughan, J.G., Whitehouse, F.L.S. and Whitehouse, J.M. (1971) Seed structure and the taxonomy of the Cruciferae. Bot J Linn Soc 64: 383–409.

Weigel, D., Glazebrook, J. (2002). How to transform Arabidopsis. In: Arabidopsis: A laboratory manual. Cold Spring Harbor Laboratory Press, Cold Sping Harbor, NY.

Wright, P. E. and Dyson, H. J. (1999). Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. J. Mol. Biol., 293:321-33

Xie, M,, He, Y. and Gan, S. (2001). Bidirectionalization of polar promoters in plants. Nat Biotechnol. 19:677-679

Yoo, SY, Bomblies, K, Yoo, SK., Yang, JW, Choi, MS, Lee, JS, Weigel, D, and Ahn, JH (2005). The 35 S promoter used in a selectable marker gene of plant transformation vector affects the expression of the transgene. Planta. 221:532-530

Yucel, I, Xiao, Y, and Hutchenson, SW (1989). Influence of *Pseudomonas syringae* culture conditions on initiation of the hypersensitive response of cultured tobacco cells. Applied and Environ. Microbiol. 55: 1724-1729

Zimmermann, P, Hirch-Hoffman, M, Henning, L, and Gruissem, W. (2004). Genevestigator, Arabidopsis microarray database and analysis toolbox. Plant Physiol., 136:2621-2632

CHAPTER 4

MUTATIONAL ANALYSIS OF *PELPK1*

CHAPTER 4

MUTATIONAL ANALYSIS OF *PELPK1*

1. INTRODUCTION

In the functional characterization of genes of a known sequence, mutational analysis via reverse genetics approaches have been used frequently over the last several years. Thus, efficient reverse genetics has become an important aspect of functional genomics, particularly in the case of the model plant, *Arabidopsis,* whose genome has been completely sequenced (AGI 2000; Sessions et al., 2002). Among these approaches, sequence-indexed insertional mutagenesis (Feldmann, 1991; Jeon et al., 2000; Alonso, et al., 2003), gene silencing by antisense gene technology, RNA interference or RNAi (Chuang and Meyerowitz, 2000; Onouchi et al., 2000), and gene over-expression (Lloyd 2003) are most frequently used. The aim of these techniques has been to analyze phenotype of the progeny populations following alterations in the genes (Gilchrist and Haugh, 2010; An et al., 2005).

In the current investigation, the following reverse genetics approaches were applied to the functional characterization of the *PELPK1* gene: (i) sequence-indexed insertional mutagenesis, (ii) RNA interference (RNAi), and (iii) gene over-expression (OX).

Sequence-indexed insertional mutagenesis is a useful type of reverse genetics because it can reveal the function of a known DNA sequence by disrupting its expression. In most cases, the *Agrobacterium tumefaciens* T-DNA has been used as a mutagen (Krysan et al., 1999). The inserted T-DNA can not only potentially knock out the gene into which it is inserted, but it can also be used as a marker and as a tag to identify the site of insertion of the T-DNA into the gene (Pan et al., 2003). Because of its suitability for genetic analysis, as well as having a complete genome sequence

available, *Arabidopsis* has been used to creating insertional mutant populations (Azpiroz-Leehan and Feldmann, 1997; Bouche and Bouchez, 2001; Alonso et al., 2003). The University of Wisconsin has generated a population of 60,000 T-DNA-tagged lines (Krysan et al., 1999). In recent years, an additional 225,000 independent T-DNA insertion lines have been created by the SALK institute, which are referred to as SALK lines (Alonso et al., 2003). Syngenta Inc. has also generated numerous T-DNA insertion lines, which are referred to as SAIL (Syngenta Arabidopsis Insertion Library) lines (Sessions et al., 2002). In the present investigation, two SALK lines and one SAIL line containing putative T-DNA insertion in different sites of the PELPK1 (AT5G09530, TAIR) were used as a part of mutational analysis for functional characterization of this gene of interest.

Bioinformatics analysis (Chapter 2) clearly suggested that *PELPK1* (AT5G09530) was associated with a tandemly duplicated paralog, *PELPK2* (AT5G09520) that may be functionally redundant with *PELPK1*. Thus, for further functional characterization of *PELPK1* in *Arabidopsis* plants, RNAi and over-expression technologies were applied to simultaneously silence the two related genes, and to over-express the *PELPK1*. Because some mutants do not demonstrate distinct phenotypes under normal growing conditions (Bouchéa, and Bouchez, 2001), determination of the response of these mutants to common abiotic and biotic stresses was also another objective of this investigation. Thus, the overall objectives of this investigation were to characterize the effects of increasing or decreasing *PELPK* gene expression in *Arabidopsis*.

2. MATERIALS AND METHODS

Plant growth conditions

Conditions in MS-agar medium – Transgenic and WT *Arabidopsis* seeds were surface-sterilized and cold-incubated as described earlier (Chapter 3). They were sown either in circular (15 x 90 mm) or square (90 x 90 mm) petri-dishes containing ½ x MS medium (pH 5.7 by KOH) and 0.7% phytablend (MS-agar), with or without the presence of 15 - 50 mM sucrose. The medium was supplemented with additional treatment solutions as required including antibiotics or stress inducing agents.

Conditions in the soil – *Arabidopsis* seeds were sown in pots (140 x 150 mm) containing wet potting soil (Sunshine Mix, Sungro). The pots were incubated in the dark at 4°C for two days, and then transferred to the growth chamber as described earlier (Chapter 3). After 8-10 days, pots were thinned to keep a desired number of plants (10-12) in each pot. Plants were watered as required with caution to avoid over-watering.

Abiotic stress treatments

Seeds of mutant (T_3) and WT plants were surface-sterilized and cold-incubated as described above. They were sown in square petri-dishes containing MS-agar medium (pH 5.7 by KOH), supplemented with various concentrations of either sucrose (0, 50, 100, 150, and 200 mM), or sorbitol (0, 50, 100, 200 and 250 mM), or glucose (0, 50, 100, 150, and 200 mM) or NaCl (0, 50, 100, and 150 mM) or ABA (0, 0.5, 1, 5, and 10 µM). Experiments were conducted following randomized complete block design (RCB) with three replicated plates for each treatment including the control. Each plate was treated as a separate block. Plants were grown vertically with ~20 plants per plate.

Response to hormones

The experimental procedure was similar to the one as described above in case of abiotic stress treatments. The MS-agar medium was supplemented with different concentrations of either IAA (0, 10, 20, and 50 µM), or GA_3 (0, 0.1, 0.5, 1, 5, 10, and 20 µM). Experiments were performed using RCB design with three replications per treatment as mentioned above.

Response to the pathogen, *Pseudomonas syringae*

Experiments were conducted following Schreiber et al., (2008) with modifications. T_3 seeds of three mutant lines from each of RNAi and OX plants along with WT control were surface sterilized as described above. Five to eight seeds were distributed to each well of 96-well plates containing 200 µl liquid ½ X MS medium with 2 mM 2-(N-morpholino) ethanesulfonic acid (MES), pH 5.8, either in absence or in the presence of 20 µM MeJa or SA. The plates were cold incubated for 2 days at 4°C in the dark and then transferred to continuous light at 22°C. After germination, when the seedlings were four days old, each well was injected /inoculated with 10 µl *Pseudomonas syringae* suspension to give a final OD of 0.03 at 600nm. The plates were incubated as above with continuous agitation as in Schreiber et al., (2008). The seedlings were examined every day under microscope for more than 6 days.

Phenotypic and growth analyses

Analysis in the MS-agar medium **-** For screening plants in the MS-agar medium, mutant as well as parental control seeds were surface sterilized and cold incubated as described above. They were sown in square or circular petri-dishes containing MS-agar medium (pH 5.7 by KOH) supplemented with or without 15- 50 mM sucrose. Plants were analyzed for morphological characteristics, germination rates, and root elongation.

Analysis in the soil - For testing the plants under soil conditions, seeds of the above lines were sown in pots containing wet Sunshine Mix, cold incubated at 4°C for two days, and then pots were transferred to a growth chamber under environmental conditions described earlier (Chapter 3). After germination, thinning was done to keep 10-12 plants per pot. The experiments were performed using RCB design with three replicated pots for each mutant line, including parental control. Plants were watered as required with an equal volume of water to each pot. Plants were photographed, and analyzed for morphological characteristics (e.g. plant height and flowering rate). Plant height was measured from the base of the plant (at the soil surface) along the stem to the tip of the shoot apex. Flowering was recorded upon first appearance of flower bud. At maturity, above ground tissues were harvested, and subjected to oven drying at 70°C for overnight following harvesting of seeds. Then the shoot and seed dry weights and morphology of seeds were determined. Each experiment was repeated at least twice.

Extraction of RNA

Total RNA samples were extracted from leaf tissues (100 mg) of rosette plants grown as described above using RNeasy Plant Mini Kit (Qiagen). Samples were subjected to agarose gel electrophoresis for quality assurance. DNA was removed from the samples using DNA-free RNA kit (Ambion), and the samples were again subjected to agarose gel electrophoresis to ensure purity. RNA concentration was determined by NaNoDrop as described previously (Chapter 3).

Synthesis of cDNA

A reverse transcription reaction was carried out following Fermentas protocol to synthesize the first strand of cDNA using 2.5 µg of total RNA from each sample, and with either SuperScript III (Invitrogen) or RTAid H-MMLV (Fermentas), and Oligo(dT)$_{12\text{-}18}$

(Invitrogen), or oligo(dT)$_{18}$ (Fermantas) as the primer in a 20 μL reaction volume. The resulting cDNA preparations were subjected to reverse transcription polymerase chain reaction (RT-PCR) using actin-2 (At3g18780) primers (Fwd: 5'-GGTCGTACAACCGGTATTGTA-3'; Rvs: 5'-GATTCCTGGACCTGCCTAA-3'). These primers do not amplify the actin fragment from the genomic DNA as they target the exon/intron boundary of the actin gene. The PCR/RT-PCR cycling program was as described earlier (Chapter 3). Following RT-PCR, the products were subjected to agarose gel electrophoresis to ensure that cDNA was appropriately made (Figure 4-1).

Quantitative Real-Time PCR (qRT-PCR)

The cDNA samples were diluted with sterile dH_20, and 2.5 μL of the diluted samples were used as template for qRT-PCR experiments. The primers for qRT-PCR reactions were designed using either PrimerExpress3.0 (Applied Biosystem) or Primique (http://cgi-www.daimi.au.dk/cgi-chili/primique/front.py; Fredslund, 2007), to have a melting T_m of 60°C with an amplicon length of between 80 and 200 bp. To run primer validation experiments, a portion of cDNA transcribed from 5 μg of total RNA was diluted to 1/4 , 1/16, 1/64, 1/256, 1/1024, and 1/4096 with ddH_20. Primer efficiency (E) was calculated using the plots of C_t (threshold cycle) versus log (input) and the equation: $E=10^{-1}$/slope. The qRT-PCR experiments were performed with home-made 2x SYBR Green master mix in 10 μL reaction volume using 96 well plates (Axygen), and an ABI 7500 fast real-time PCR system (Applied Biosystem) following the manufacturer's instructions. Each PCR reaction contained 2.5 μL of diluted cDNA, 0.4 μM of each of the primers, and 5 μL SYBR Green I mix. The SYBR Green I mix contained: 20 mM Tris (pH

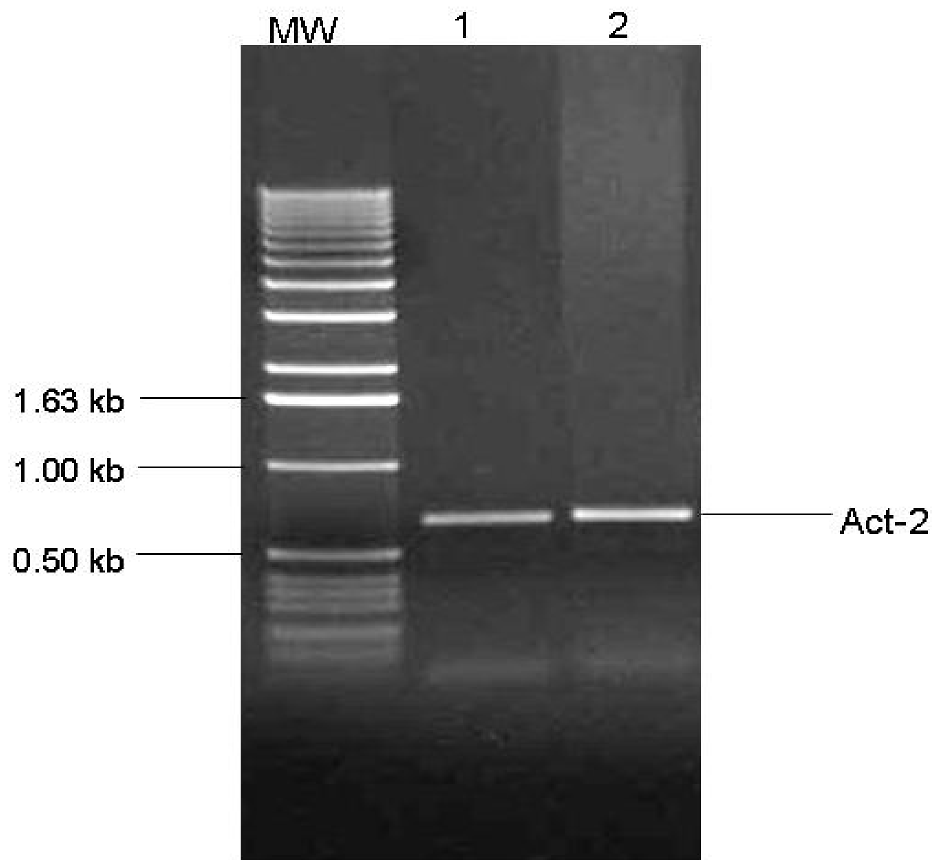


Figure 4-1: Agarose gel electrophoresis of RT-PCR products from T_3 transgenic lines using Act-2 primers (Act-663 Fwd + 1334 Rvs). MW: 1 kb molecular weight marker, Lane -1: cDNA from a RNAi line, Lane-2: cDNA from an OX line. Three lindependent T3 lines from each mutant were tested for RT-PCR.

8.3), 50 mM KCl, 3 mM $MgCl_2$, 0.8% glycerol, 0.01% Tween 20, 2% dimethyl sulfoxide (DMSO), 200 μM dNTPs, 1x ROX, 0.25x SYBR Green I, 0.03 units/μL Platinum Taq DNA polymerase. The PCR reaction was run according to the following program: initial denaturation for 2 min at 95°C followed by 35 cycles consisting of 94°C for 15 sec and 60°C for 1 min with a single fluorescence measurement. The GOI/paralog and the reference gene (UBQ10; At4g05320; RG) -specific primers designed to run qRT-PCR to determine transcript abundance of WT as well as other *Arabidopsis* mutant lines are listed in Table 4-1.

Table 4-1: Primer pairs used for transcript analysis by qRT-PCR

Primer ID	Gene ID	Primer sequence
Primer pair I	*PELPK1*	Fwd: 5’-CGAAGAGAAGGCCCAGCTACT-3’ Rvs: 5’-CAGACATAAGTGCATCTCTGCAAA-3’
Primer pair II	*PELPK2*	Fwd : 5’-ATGACATTGAAGAAGAGTT-3’ Rvs : 5’-GTATGAGTGTAATAACAAT-3’
Primer pair III	*PELPK1*	Fwd: 5’-AAAGGTACCGGAGATTCAG-3’ Rvs: 5’-CTCAGGCTTTGGAATCTC-3’
UBQ10	RG	Fwd: 5’-GGCCTTGTATAATCCCTGATGAA-3’ Rvs: 5-AGAAGTTCGACTTGTCATTAGAAAGAAA-3'

Gene-specific PCR efficiency was used to calculate the transcript expression of *PELPK1* and *PELPK2* relative to the expression of the reference gene, RG (Pfaffl, 2001). The qRT-PCR data showing C_t values between 10 and 30 were used for transcript analysis. To calculate transcript expression of *PELPK1* in mutant (MUT) and WT plants, the following equations (Real-Time PCR System Chemistry Guide) were used:

$$\text{Mean } C_t \text{ PELPK1 (WT)} - \text{Mean } C_t \text{ RG (WT)} = \text{d } C_t \text{ PELPK1 (WT)}$$

$$\text{Mean } C_t \text{ PELPK1 (MUT)} - \text{Mean } C_t \text{ RG (MUT)} = \text{d } C_t \text{ PELPK1 (MUT)}$$

$d\ C_t$ PELPK1 (MUT) – $d\ C_t$ PELPK1 (WT) = $dd\ C_t$ PELPK1 (MUT)

$2^{-(ddCt\ GOI\ MUT)}$ = fold difference

Identification of homozygous T-DNA insertion lines

Seeds of the *Arabidopsis* T-DNA insertion lines, SALK_007409 (SALK-1), SALK _002771 (SALK-2), and SAIL_517_G06 (SAIL), and their parental control, CS6000, were obtained from ABRC (*Arabidopsis* Biological Resource Center, Ohio, USA). They were sown in soil under greenhouse conditions as described earlier, and seeds were collected separately from each mutant line. Plants raised from these seeds were subjected to PCR to identify homozygous T-DNA insertion lines using primers designed following the procedure described in http://signal.salk.edu/tdnaprimers.2.html. The primer sequences used were as follows:

<u>SALK lines</u>

LBa1 (T-DNA left border-specific): 5'-TGGTTCACGTAGTGGGCCATCG-3' (Fwd)

LBb1 (T-DNA left border-specific): 5'-CGTGGACCGCTTGCTGCAACT-3' (Fwd)

SALK_007409:

LP (gene-specific): 5'-TGTTTGAGGCGGTGAGTAATC-3' (Fwd)

RP (gene-specific): 5'-AAGGTACCGGAGATCCAGAAG-3' (Rvs)

SALK_002771:

LP (gene specific): 5'-TTGACGTGTTCATGTGTTTGG-3' (Fwd)

RP (gene specific): 5'- GCCCAAGGTTCCAGAAATTAC-3' (Rvs)

<u>SAIL line</u>

SAIL_517_G06:

LB1 (Left border-specific): 5'-GCCTTTTCAGAAATGGATAAATAGCCTTGCTTC-3 (Fwd)

LP (gene specific): 5;-GATTCCAAAGCCTGAGGAAAC-3' (Fwd)

RP (gene specific): 5'-AACAAGAAGAAGATAGCCGTCG-3' (Rvs)

Genomic DNA was isolated from the leaves of independent rosette plants of the above T-DNA insertion lines as well as from the parental control plants grown in the soil as described earlier. PCR was performed involving at least 10 independent plants from each mutant line. The primer combinations used were: LB + RP and LP + RP following the protocol as described in http://signal.salk.edu/tdnaprimers.2.html.

The approximate sites of T-DNA insertion into *PELPK1* (AT5G09530) of knock out (KO) mutant lines were located by aligning insertion flanking sequence of the T-DNA with the *PELPK1* sequence using SeqMan, DNASTAR (http://www.dnastar.com/forms.aspx?), and are diagrammatically presented in Figure 4-2. In the SALK_007409 line used, it was putatively inserted at the position of 1102nd bp down stream of the translational start site of a 1113-bp CDS, in SALK_002771 line, it was inserted at the position of 50th bp up-stream of a 79-bp 5'UTR from the translational start site, and in the SAIL_517_G06 line, theT-DNA was inserted at the position of 964th bp from the translational start site. The PCR confirmation of homozygous T-DNA insertion is shown in Figure 4-3.

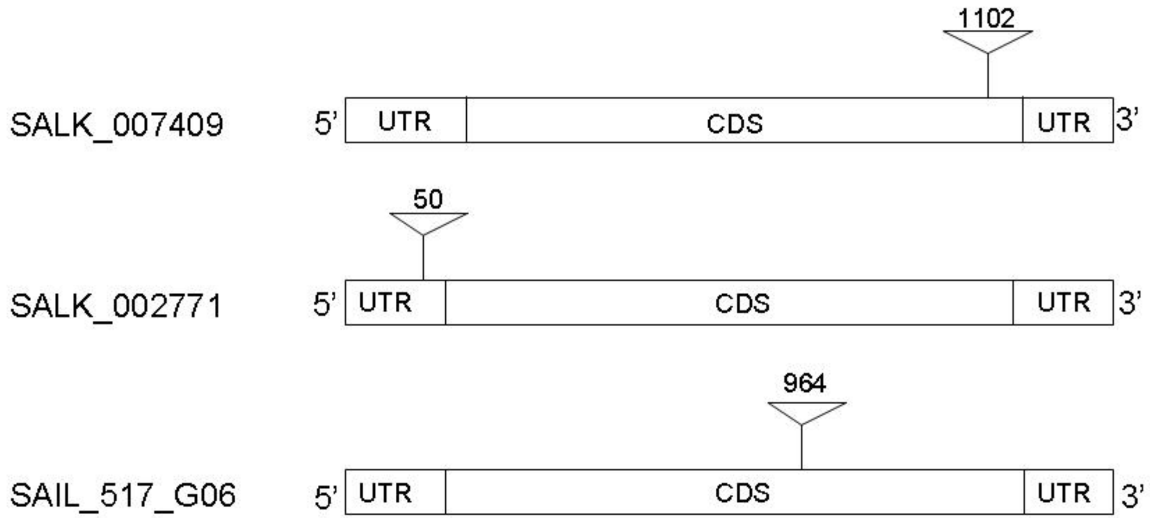


Figure 4-2: Schematic diagram of *PELPK1* showing approximate location of T-DNA insertion sites. *PELPK1* contained a single open reading frame with no intron present. The insertion sites are shown by triangles. Numbers on the triangles indicate base pairs. UTR: untranslated region, CDS: coding sequence

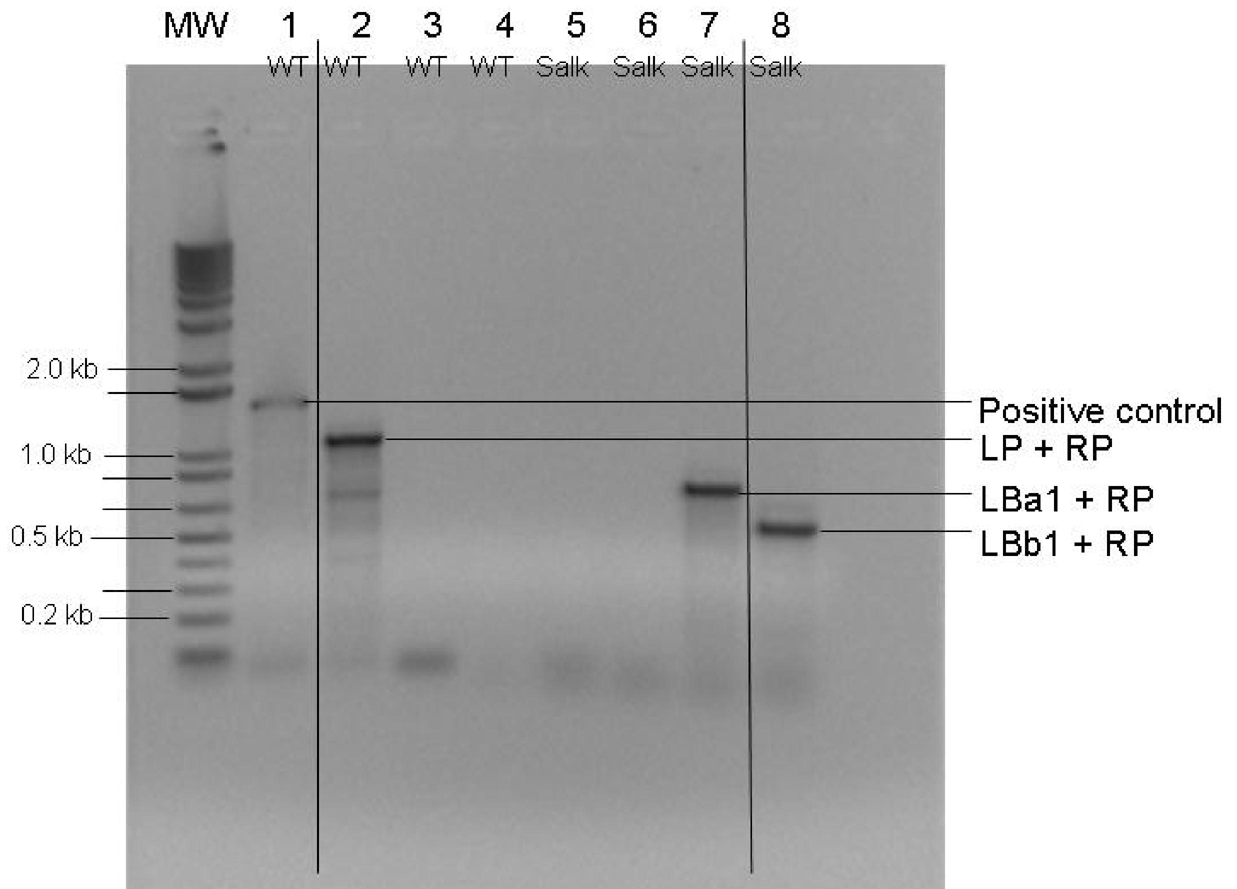


Figure 4-3: Confirmation of homozygous T-DNA insertion in a SALK line. Lane 1: *PELPK1*-specific primers, Lane 2: LP+RP, Lane-3: LBa1+RP, Lane-4: LBb1+RP, Lane-5:GOI-specific primers (as in lane-1), Lane-6: LP+RP, Lane-7: LBa1+RP, Lane 8: LBb1+RP (LBa1/b1: T-DNA left border-specific primers; LP: *PELPK1*-specific left primer, RP: *PELPK1*-specific right primer). Three independent SALK and SAIL lines were tested and all of them gave similar PCR products.

Plasmid construction

Construction of RNAi plasmids- For the construction of RNAi vector constructs, standard molecular biology (Sambrook et al., 1989) and Gateway Recombination Technology (Invitrogen; Helliwell and Waterhouse 2003) were used. Conserved regions of the deduced amino acid sequences of the two genes, *PELPK1* (AT5G09530) and its putative paralog, *PELPK2* (AT5G09520), obtained by alignment (shown below) were used to design a pair of common forward and reverse RNAi primers by 6-frame translation of the proteins using BCM Search Launcher (Human Genome Sequencing Center, One Baylor Plaza, Houston, TX; http://searchlauncher.bcm.tmc.edu/multi-align/multi-align.html) as shown below:

```
At5g09530_protein      MALMKKSLSAALLSSPLLIICLIALLADPFSVGARRLLEDPKPEIPKLPELPKFEVPKLP
At5g09520_protein      --MTLKKSFSASLLSPFLIICLIALLSVPVSVGARRLLEEPKPEIPTFPELPKPEMPKLP
                         :  *.  :* * **:*********: *.*********:******.:***** *:****
At5g09530_protein      EFPKPELPKLPEFPKPELPKIPEIPKPELPKVPEIPKPEETKLPDIPKLELPKFPEIPKP
At5g09520_protein      EFPKLELPKLPEIPKPEMPKLPEIQKPELPTFPELP----------------KMPEFPKF
                       **** *******:****:**:*** *****..**:*                *:**:**
At5g09530_protein      ELPKMPEIPKPELPKVPEIQKPELPKMPEIPKPELPKFPEIPKPDLPKFPENSKPEVPKL
At5g09520_protein      DFPKLPELPKPEETKVPAFTMPKFPGSP--------------------------------
                       ::**:**:**** .*** :  *::*  *
At5g09530_protein      METEKPEAPKVPEIPKPELPKLPEVPKLEAPKVPEIQKPELPKMPELPKMPEIQKPELPK
At5g09520_protein      ------------------------------------------------------------

At5g09530_protein      LPEVPKLEAPKVPEIQKPELPKMPELPKMPEIQKPELPKMPEIQKPELPKVPEVPKPELP
At5g09520_protein      ------------------------------------------------------------

At5g09530_protein      TVPEVPKSEAPKFPEIPKPELPKIPEVPKPELPKVPEITKPAVPEIPKPELPTMPQLPKL
At5g09520_protein      ------------------------------------------------------------

At5g09530_protein      PEFPKVPGTP
At5g09520_protein      ----------
```

<u>Gene-specific forward primer</u>

PELPK1 (At5g09530): 6-frame translation

```
DNA: CTCTGCTGCTCTTCTCTCATCACCACTTCTGATCATATGTCTTATCGCATT
 +3:   L  L  L  F  S  H  H  H  F  *  S  Y  V  L  S  H  C
 +2:  S  A  A  L  L  S  S  P  L  L  I  I  C  L  I  A  L
 +1: L  C  C  S  S  L  I  T  T  S  D  H  M  S  Y  R  I

DNA: GCTCGCTGATCCGTTTTCAGTCGGTGCTCGCCGGTTATTGGAGGATCCTAA
 +3:   S  L  I  R  F  Q  S  V  L  A  G  Y  W  R  I  L  N
 +2:  L  A  D  P  F  S  V  G  A  R  R  L  L  E  D  P  K
```

```
+1: A  R  *  S  V  F  S  R  C  S  P  V  I  G  G  S  *
```

PELPK1-specific (Fwd): *5'-TTCTGATCATATGTCTTAT*CGCATTGCTC-3'

PELPK2 (At5g09520): 6-frame translation

```
DNA: ATGACATTGAAGAAGAGTTTCTCTGCTTCTCTACTTTCACCATTTCTGATC
 +3:   D  I  E  E  E  F  L  C  F  S  T  F  T  I  S  D  H
 +2:  *  H  *  R  R  V  S  L  L  L  Y  F  H  H  F  *  S
 +1: M  T  L  K  K  S  F  S  A  S  L  L  S  P  F  L  I

DNA: ATATGTCTTATTGCATTGCTTTCTGTTCCGGTATCCGTTGGAGCTCGCCGG
 +3:   M  S  Y  C  I  A  F  C  S  G  I  R  W  S  S  P  V
 +2:  Y  V  L  L  H  C  F  L  F  R  Y  P  L  E  L  A  G
 +1: I  C  L  I  A  L  L  S  V  P  V  S  V  G  A  R  R
```

PELPK2-specific (Fwd): 5'-*TTCTGATCATATGTCTTAT*TGCATTGCTT-3'

Alignment of the forward primers:
PELPK1-specific (Fwd) *5'-TTCTGATCATATGTCTTAT*CGCATTGCTC-3'
PELPK2-specific (Fwd): 5'-*TTCTGATCATATGTCTTAT*TGCATTGCTT-3'

Gene-specific reverse primer

PELPK1 (At5g09530): 6-frame translation

```
DNA: GCCGGAGTTCCCTAAACCAGAGTTGCCCAAGTTACCCGAATTTCCAAAGCC
 +3:   R  S  S  L  N  Q  S  C  P  S  Y  P  N  F  Q  S  L
 +2:  P  E  F  P  K  P  E  L  P  K  L  P  E  F  P  K  P
 +1: A  G  V  P  *  T  R  V  A  Q  V  T  R  I  S  K  A
```

PELPK1-specific (Rvs): 5'-*AGAGTTGCCCAAGTTACC*CGAATTT-3'

PELPK2 (At5g09520): 6-frame translation

```
DNA: CCAGAGATGCCAAAGTTACCCGAATTTCCGAAGCTAGAGTTGCCTAAGTTA
 +3:   R  D  A  K  V  T  R  I  S  E  A  R  V  A  *  V  T
 +2:  Q  R  C  Q  S  Y  P  N  F  R  S  *  S  C  L  S  Y
 +1: P  E  M  P  K  L  P  E  F  P  K  L  E  L  P  K  L

DNA: CCTGAGATTCCAAAACCAGAGATGCCTAAGTTACCGGAGATTCAGAAGCCT
 +3:   *  D  S  K  T  R  D  A  *  V  T  G  D  S  E  A  *
 +2:  L  R  F  Q  N  Q  R  C  L  S  Y  R  R  F  R  S  L
 +1: P  E  I  P  K  P  E  M  P  K  L  P  E  I  Q  K  P
```

PELPK2-specific (Rvs): 5'-*AGAGTTGCCTAAGTTACC*TGAGATT-3'

Alignment of the two reverse primers:
PELPK1-specific (Rvs): 5'-*AGAGTTGCC*C*AAGTTACC*CGAATTT-3'
PELPK2-specific (Rvs): 5'-*AGAGTTGCC*T*AAGTTACC*TGAGATT-3

Final gene-specific (Fwd): 5'-TCTGATCATATGTCTTAT-3' (18 bases)
Final gene-specific (Rvs): 5'-GGTAACTTAGGCAACTCT-3' (18 bases)

For Gateway cloning (Invitrogen), the above gene-specific forward and reverse primers were added, respectively to the end of the Gateway forward and reverse primers, with some modifications as shown below:

Original Gateway® forward primer (attB1):

5´-GGGGACAAGTTTGTACAAAAAAGCAGGCTNN(gene-specific sequence)-3´

Modified Gateway forward primer with gene-specific Fwd primer attached:

5´-GGGGACAAGTTTGTACAAAAAAGCAGGCTT*TCTGATCATATGTCTTAT*-3´ (48 bases)

Original Gateway® reverse primer (attB2):

5´-GGGGACCACTTTGTACAAGAAAGCTGGGTN(gene-specific sequence)-3´

Modified Gateway reverse primer with gene-specific Rvs primer attached:

5´-GGGGACCACTTTGTACAAGAAAGCTGGGT*GGTAACTTAGGCAACTCT*-3' (47 bases)

The above Gateway recombination primers (GPs) with attB1 and attB2 sites incorporated were then used to conduct PCR with WT genomic DNA. The resulting PCR fragment of ~ 218 kb was separated by agarose gel (Figure 4-4A) and purified by gel purification kit as described earlier. This fragment was cloned into a donor vector (pDONR222) following the BP cloning protocol (Invitrogen) to generate an entry vector, which was then used to transform *E. coli* (DH5α) competent cells. Positive clones, selected in the presence of 50 μg/ml kanamycin, were used to isolate plasmids. The positive entry-vector plasmid was subjected to sequencing using a vector-specific forward primer (5'-d(GTAAAACGACGGCCAGT)-3'). The manufacturer's BigDye terminator cycle sequencing reagent, and an AB13730 sequencer (Applied Biosystem) were used for sequencing experiments.

B

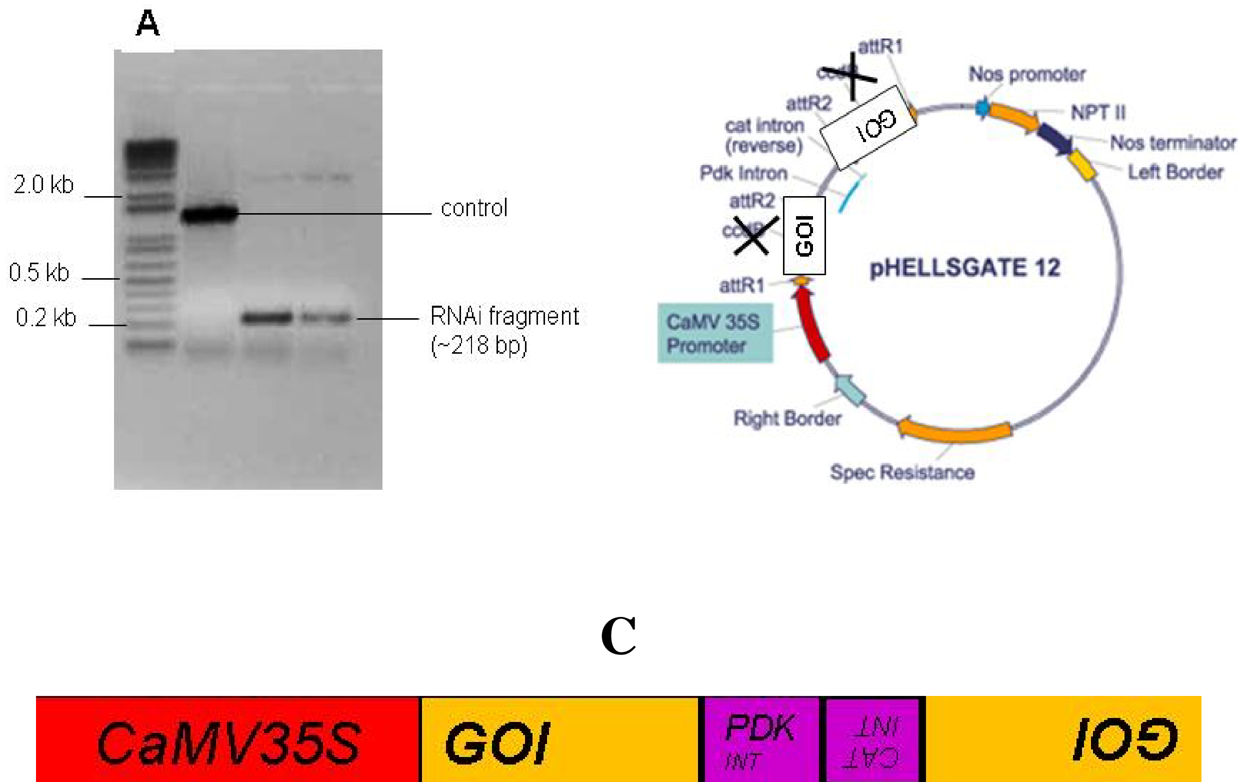


Figure 4-4: Gateway recombination cloning. **A**: the RNAi fragment (~218 bp) amplified from WT genomic DNA by PCR using gateway recombination primers, **B**: the map of RNAi constitutive plant transformation vector showing the insertion of RNAi fragment (GOI = genes of interest: *PELPK1+PELPK2*) in reverse orientation by replacing the ccdB gene (vector map from Helliwell and Waterhouse 2003), **C**: a schematic diagram of the RNAi construct. PDK = Pyruvate dehydrogenase kinase. About ten T3 trasgenic lines were developed harboring the above construct.

Afterward, an LR reaction was performed to transfer the gene from the entry clone (pDONR222) to the destination vector, pHellsgate12 (constitutive). Following the *E. coli* (DH5α) transformation with this construct, the positive clones were selected in the presence of 100 μg/ml spectinomycin. Plasmids isolated from the above positive clones were tested by PCR to confirm the presence of the gene insert, and then subjected to restriction digestion using *Xba*1 or *Xho*1/*Kpn*1 for pHellsgate12 to confirm that the insert was properly positioned in the construct. The resulting RNAi construct of pHellsgate12 (Figure 4-4B) was then used to transform *Agrobacterum tumifaciens* (GV3101) using freeze-thaw method as described earlier (Chapter 3). The transformants were selected in the presence of 50 μg/ml kanamycin, 25 μg/ml gentamycin, and 15 μg/ml rifamycin as before, and used for *Arabidopsis* transformation as described below.

***Construction of over-expression plasmid* –** Genomic DNA isolated from WT *Arabidopsis* plants was used to conduct PCR using a pair of primers designed to amplify the coding region of the *PELPK1* (At5g09530) and its 3'UTR. The primers were designed by incorporating *Nco*I and *Bst*EII sites, respectively in the forward and the reverse primers as follows: 5'-TACCATGGCACTAATGAAGAAGA-3' (Fwd); 5'-GGTGACCTCATTGTATAGCTTTTG-3' (Rvs). The PCR product (1.442 kb) was then cloned into a TA cloning vector (pCRII-TOPO), or a Blunt-cloning vector (pCR-BluntII-TOPO) following Invitrogen protocol. These constructs were used to transform Invitrogen-supplied chemically competent *E. coli* (DH5α) cells. The positive clones identified on LB medium containing 50μg/ml ampicillin, were tested by colony PCR to confirm the presence of the insert. Pure *E. coli* cell cultures were then prepared from the transformed cells, and used to isolate plasmids. The above plasmid constructs as well as the plant transformation binary vector, pCAMBIA1303, were digested with *Nco*I plus *Bst*EII. The *PELPK1*-insert and the empty pCAMBIA1303 vector were separated by

agarose gel, and purified by gel purification kit as described earlier. Following ligation reactions between the above vector and the insert, the resulting plasmid construct (OX construct, Figure 4-5) was transformed into *E. coli* competent cells as above. The positive transformants selected in the presence of 50 μg/ml kanamycin were used to prepare pure cultures to isolate plasmids. These OX- plasmids were used to transform *Agrobacterium* competent cells as described above.

Arabidopsis transformation and selection of transgenic lines

WT *Arabidopsis* plants were separately transformed using the RNAi (constitutive) and OX-constructs following floral dip method as described earlier. T_1 plants harboring the above constructs were selected on Phytablend plates containing ½ x MS medium, 30 mM sucrose, and either 100 μg/ml Kan for RNAi- or 50μg/ml Hyg for OX-lines. These transgenic lines were grown up to the T_3 generation by selecting them on their respective antibiotic-containing media.

Genomic DNA was isolated separately from the T_3 plants of both RNAi and OX lines as described earlier (Chapter 3) and subjected to PCR to confirm the presence of transgene constructs (Figure 4-6). In the case of the RNAi lines, the PCR was conducted using vector-specific (Kan^R gene-specific) primers as follows: 5'-CGCTCAGAAGAACTCGTCAAGAA-3' (Fwd), 5'-TTTGTCAAGACCGACCTCTCC-3' (Rvs. In the case of OX-lines, PCR was conducted using vector-specific forward and *PELPK1*-specific reverse primers as follows: pCAMBIA1303 (MCS-F): 5'-TGTGGAATTGTGAGCGGATA-3' (Fwd) or pCAMBIA1303 (seqF): 5'-CAACCACGTCTTCAAAGCAA-3' (Fwd); *PELPK1*-specific: GGTGACCTCATTGTATAGCTTTTG-3' (Rvs).

Figure 4-5. A schematic diagram of an over-expression construct. CDS = coding sequence of the *PELPK1*, UTR = untranslated region. More than ten T3 trasgenic lines were developed harboring the above construct.

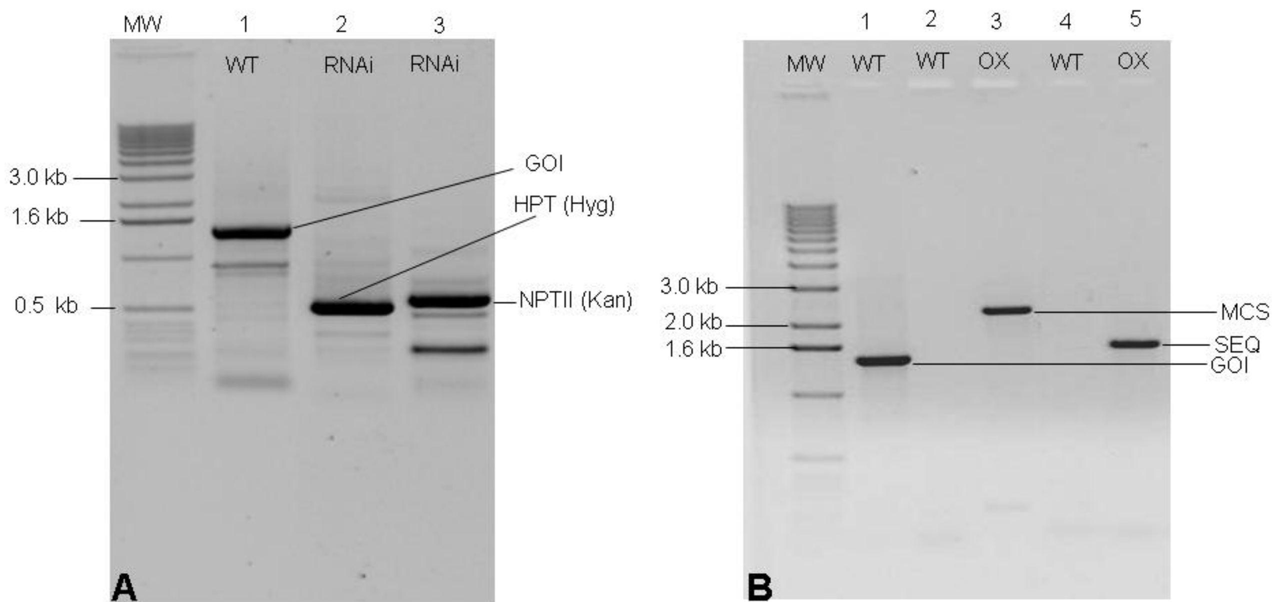


Figure 4-6: Confirmation of insert constructs in the T3 transgenic lines by PCR. Genomic DNA was extracted from rosette leaves. **A**: RNAi constructs; MW: 1 kb ladder, Lane-1: WT (positive control), Lane-2: pOpOff2 (Vector-specific primers), Lane-3: pHellsgate12 (Vector-specific primers). **B**: OX-construct; MW: 1 kb ladder, Lane-1: GOI-specific primers (positive control), Line-2 and 3: Vector-specific Fwd (MCS) and GOI-specific Rvs primers, Lane-4 and 5: Vector-specific Fwd (SEG) and *PELPK1*-specific Rvs primers. At least ten T3 lines from each construct were developed and confirmed the presence of inserts by PCR.

Table 4-2: Primer pairs used for PCR amplification of *PELPK1* and *PELPK2* genes from WT genomic DNA

Primer ID	Primer sequence
PELPK1-specific	Fwd: 5'-ATGGCACTAATGAAGAAGA-3' Rvs: 5'-TCATTGTATAGCTTTTG-3'
PELPK2-specofic	Fwd: 5'-ATGACATTGAAGAAGAGTT-3' Rvs: 5'-GTATGAGTGTAATAACAAT-3'

Table 4-3: Primers used for PCR amplification of genes from cloned vectors

Primer ID	Gene ID	Primer sequence
Primer pair I	*PELPK1*	Fwd: 5'-CGAAGAGAAGGCCCAGCTACT-3' Rvs: 5'-CAGACATAAGTGCATCTCTGCAAA-3'
Primer pair II	*PELPK1*	Fwd: 5'-TGCTGCTCTTCTCTCATCAC-3' Rws: 5'-CTGCTGTTCTCTCATCAC-3'
Primer pair III	*PELPK1*	Fwd: 5'-AAAGGTACCGGAGATTCAG-3' Rvs: 5'-CTCAGGCTTTGGAATCTC-3'
Primer pair IV	*PELPK2*	Fwd: 5'-GTTCCGGTATCCGTTGGAG-3' Rvs: 5'-CAAGGATCAAGGATGAGATGG-3'
Primer pair V	*PELPK2*	Fwd : 5'-ATGACATTGAAGAAGAGTT-3' Rvs : 5'-GTATGAGTGTAATAACAAT-3'

3. RESULTS

Analysis of T-DNA insertion mutant lines

To evaluate the phenotypic effect of loss of function of *PELPK1*, three sequence-indexed insertion mutant lines (SALK_007409, SALK_002771, and SAIL_517_G06) obtained from ABRC were identified with possible T-DNA insertions in or near *PELPK*1. Analysis of the reported insertion flanking sequence showed that in SALK_007409, the insertion was near the 1102nd nucleotide of the 1113-nucleotide coding sequence (CDS), in SALK_002771 line, the insertion was near the 50th nucleotide of the 79-nucleotide 5'UTR, and that in SAIL_517_G06 line, the insertion was near the 964th nucleotide of the CDS (Figure 4-2). At least ten plants from each mutant line were confirmed by PCR to be homozygous for the insertion, prior to further characterization (Figure 4-3).

The progeny populations of the above homozygous T-DNA insertion mutant lines were examined at different stages of growth to detect phenotypic changes. Compared to that of the parental control, none of these insertional mutant lines showed any detectable morphological differences when grown under normal conditions either in MS-agar medium (Figure 4-7) or in the soil (not shown).

Because some conditional mutants show a phenotype only when exposed to particular treatments (Krysan et al., 1996, Hirsch et al., 1998; Bouchéa, and Bouchez, 2001), these insertion lines were also subjected to stress-inducing conditions: NaCl (150 mM), sucrose (200 mM), mannitol (250 mM), and ABA (20 μM) in petri-dishes. However, even under these extreme stress conditions, no visible phenotype was detected that could distinguish insertion mutants from parental controls (not shown).

The above homozygous T-DNA insertion lines were then subjected to qRT-PCR using primer pair I (Table 4-1) designed using 3'UTR sequence of the *PELPK1* to

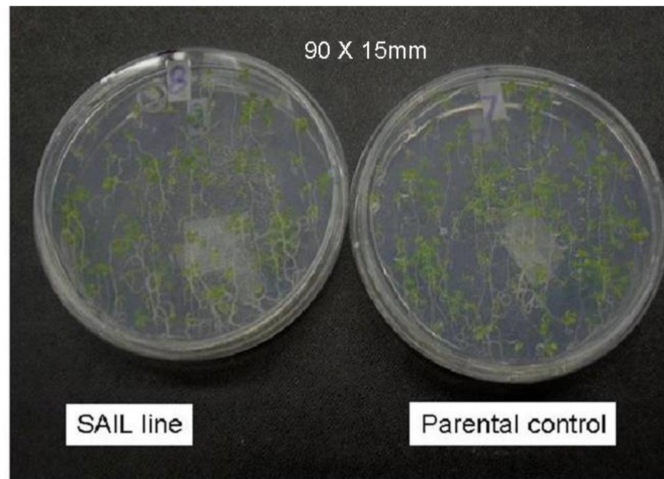


Figure 4-7: Seedlings of a homozygous SAIL line and the parental control showing germination rate, root and shoot growths, and morphological characteristics. The medium contained ½ x MS and 30 mM sucrose. Similar test was also conducted with SALK lines. The size of the round plates = 90 x 15 mm.

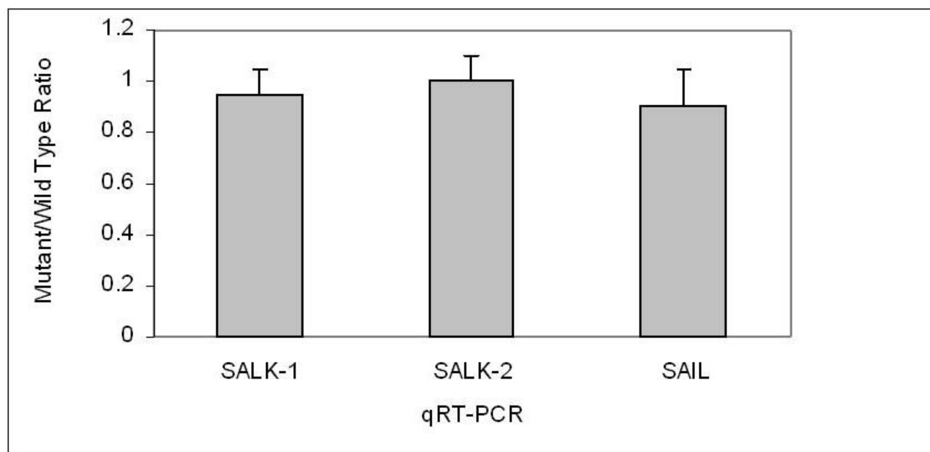


Figure 4-8: Quantitative real-time PCR (qRT-PCR) analysis for transcript levels in three T-DNA insertion lines along with WT plants grown under normal conditions in the green house. The primers (primer pair I, Table 4-1) used were designed using 3'UTR of the *PELPK1* gene. Error bars represent standard deviations (SD). RNA was isolated from leaf tissues at rosette stage

determine *PELPK1* transcript abundance. None of these insertion mutant lines showed transcript level that was different from the parental control (Figure 4-8). There might be two or possibly three reasons for the lack of loss of function phenotype and reduction in transcript abundance in these mutant lines as compared to WT plants as follows: (i) it has been shown previously that a T-DNA insertion into a gene can lead to many different outcomes including no null mutation (referred to as gene knock-about) even in the presence of an insertion that is localized in the middle of the coding region (Krysan et al., 1999), and (ii) the *PELPK1* contains an adjacent, putative paralog (*PELPK2*) that can compensate for the function of the *PELPK1*, and (iii) TAIL-PCR used to index TDNA may not be precise. Therefore, the following alternative mutagenesis methodologies were applied for functional characterization of *PELPK1*.

Analysis of RNA interference mutant lines

An RNAi construct was produced that is capable of silencing both *PELPK1* and its putative paralog, *PELPK2* simultaneously. The 171-nucleotide targeting region within the RNAi construct, excluding the Gateway Recombination Sequence, shared 100% identity with *PELPK1* and 95% identity with *PELPK2* (refer to Materials and Methods).

Following confirmation of stable transformation of *Arabidopsis* with the RNAi construct (Figure 4-6A), the T_3 plants were subjected to transcript analysis by qRT-PCR. Because the *PELPK1* and *PELPK2* genes have close sequence similarity (Chapter 2), primers were tested against cloned templates of both the genes. Primers used in cloning are listed in Table 4-2.

Each cloned gene was subjected to PCR using three pairs of primers designed to anneal to the *PELPK1*, and two pairs of primers designed to anneal to the *PELPK2* sequence (Table 4-3).

All primer pairs were tested for specificity by using cloned *PELPK1* or cloned *PELPK2* as templates in separate PCR reactions. Primer pairs I, II, IV, and V amplified both the *PELPK1* and *PELPK2* and therefore cannot be considered specific for either gene. On the other hand, primer pair III (Table 4-3) which was designed manually after aligning the two genes, did not amplify the *PELPK2* from the cloned vector but did amplify the cloned *PELPK1*. However, it produced a ladder of amplicons from the cloned *PELPK1*. When these primers were aligned against their own gene sequence i.e. *PELPK1* sequence (http://blast.ncbi.nlm.nih.gov/Blast.cgi) they showed partial annealing in more than one place in the *PELPK1* sequence (data not shown). These observations suggest that the reason for the unexpected PCR product amplification by this pair of primers was due to the presence of multiple, repetitive, and conserved sequences in the *PELPK1*. Based on the above experiments, it was concluded that primer pair III could specifically amplify *PELPK1*, and none of the primer pairs could specifically amplify the *PELPK2*; i.e. all other primer pairs except pair III amplified both *PELPK1* and *PELPK2*.

Primer pairs I – III were used in subsequent qRT-PCR experiments using RNA from three independent T_3 RNAi lines as template (Table 4-1). Results presented in Figure 4-9 showed the transcript abundance in these RNAi lines, as determined by qRT-PCR. Primer pair III, which was shown to be *PELPK1*-specific, showed almost complete reduction in *PELPK1* transcript in the RNAi lines. Primer pairs I and II, which amplify both the *PELPK1* and *PELPK2*, showed between 20 and 40% of the transcript abundance of WT plants (Figure 4-9). Based on these results and the high sequence identity between the two genes and the RNAi construct, it seems likely that expression of both *PELPK2* and *PELPK1* were significantly reduced in the RNAi lines.

To determine the effect of reduced *PELPK1*/*PELPK2* expression on plant growth, T_3 RNAi lines were observed on both MS-agar medium and on soil. Seedling morphology and germination rate, defined by the presence of an emerged radicle of at

least 1mm, did not differ between WT and RNAi plants in MS-agar medium, supplemented with 0 - 50 mM sucrose (not shown). However, when the RNAi seeds were sown in MS-agar medium supplemented with 100 mM sucrose, the percentage of germination recorded 3, 5, and 6 days after sowing (DAS) was significantly smaller in RNAi lines ($p<0.01$) than in WT (Figure 4-10). However, the percentage of germinated seeds was similar to that of WT seeds at 7 DAS and after (Figure 4-10).

Root growth of RNAi seedlings was also compared to WT in MS-agar medium supplemented with four different sucrose concentrations: 0, 50, 100, and 150 mM (Figure 4-11). Plants photographed at 12 DAS showed that although supplementation of 50 mM sucrose significantly enhanced root elongation in both RNAi and WT seedlings as compared to no sucrose addition, supplementation of 100 or 150 mM sucrose significantly slowed down the growth of RNAi roots as compared to the growth of WT roots (Figure 4-11). At 150 mM sucrose supplemented medium the growth of RNAi seedlings was almost completely stopped whereas the growth of WT seedlings was only slowed down. Quantitative determination of root length is presented in Figure 4-11C. Analysis of these data by a two-way ANOVA followed by a post-hoc t-test with Bonferroni corrections showed that the root length of RNAi lines particularly at 100 and 150 mM sucrose was significantly slowed down compared to that of the WT roots ($p<0.01$).

The root growth of RNAi seedlings was further tested in MS-agar medium supplemented with 100 mM sucrose (Figure 4-12). The quantitative root elongation data recorded 7 and 10 DAS are presented in Figure 4-12D. Statistical analysis of these data by a two-way analysis of variance (ANOVA) showed that the mean root length of the RNAi mutant was significantly different from that of the WT at both 7 and 10 DAS ($p<0.01$).

Analysis of public microarray data indicated that *PELPK1* transcripts are induced by a variety of abiotic stresses and other treatments (Chapter 2). The RNAi mutant lines were therefore subjected to the MS-agar medium supplemented with the following: sorbitol (0 -250 mM), glucose (0 - 200 mM), NaCl (0 -150 mM), ABA (0 -10 μM), gibberellic acid, GA_3 (0 -20 μM) or IAA (0 -50 μM). However, in none of these treatments was the growth rate or morphology of RNAi and WT plants significantly different (data not presented).

The phenotype and growth characteristics of the RNAi lines were further tested under controlled environmental conditions as described earlier by growing them in the soil up to maturity (Figure 4-1). When sown and germinated on soil, the emergence of the RNAi plants from the soil was delayed, their size was smaller, and color was slightly darker (Figure 4-13A), and their early vegetative growth was significantly reduced (Figure 4-13B) compared to that of the WT plants. Quantitative determination of plant height (height of inflorescence shoot) indicated that the shoot length of RNAi lines recorded 25 and 30 DAS was significantly shorter than that of WT plants (Figure 4-14). Statistical analysis of the above plant height data by a two-factor ANOVA showed that the mean plant height between WT and RNAi plants recorded at all time intervals (25, 30, and 40 DAS) in the soil was significantly different ($p<0.01$).

The flowering of RNAi plants recorded 21, 23, 25 and 30 DAS was also significantly delayed ($p<0.01$) compared to that of WT plants (Figures 4-15, 4-13B). The calculated time required for flowering of 50% RNAi plants was ~27 DAS as compared to ~25 DAS for WT plants. However, the dry weights of shoots and mature seeds per plant and the morphology of seeds, determined at maturity following the harvest, were not significantly different from that of the WT plants (data not presented).

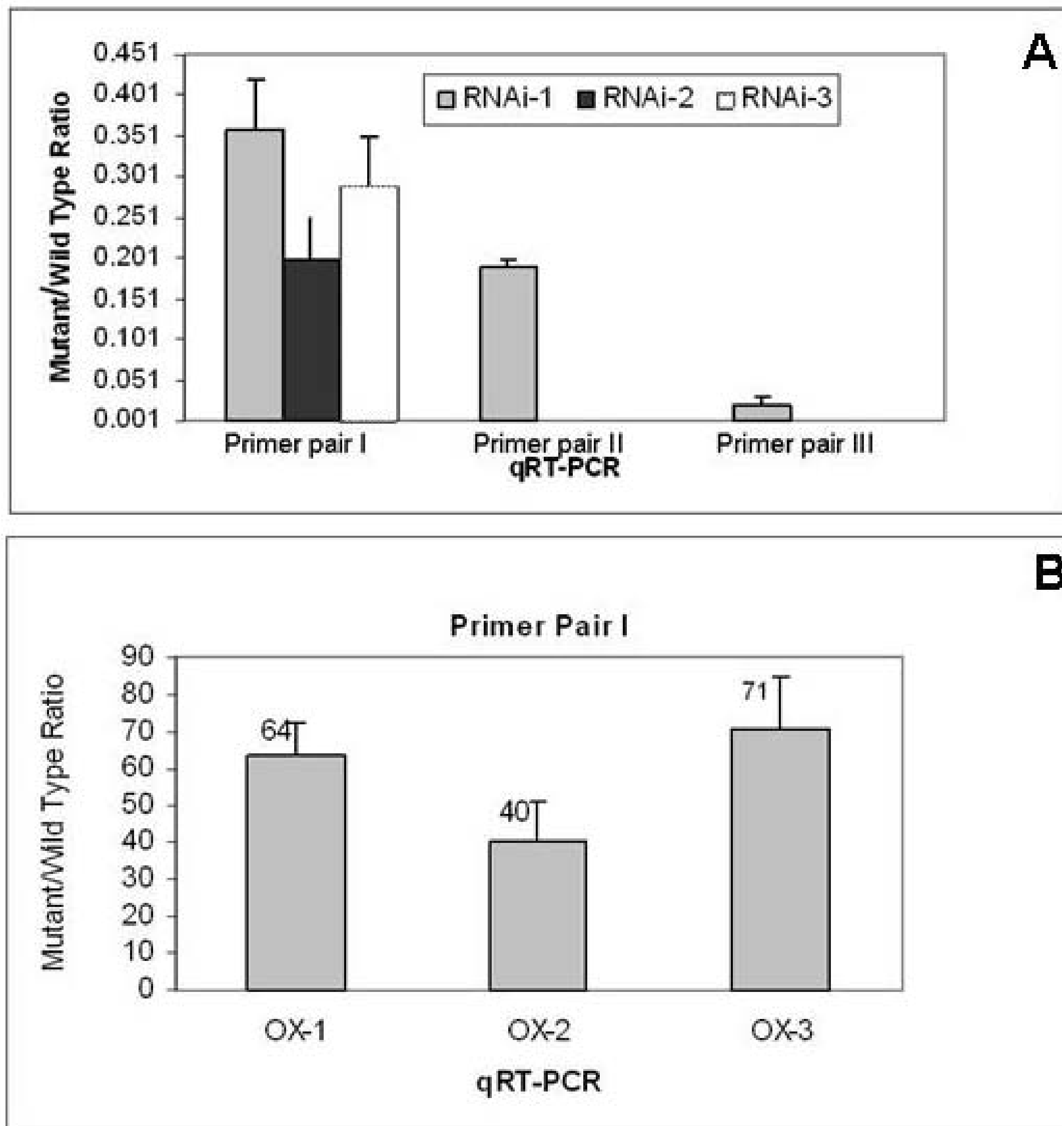


Figure 4-9: qRT-PCR analysis of the transcript abundance of mutants and WT plants. **A**: RNAi lines, **B**: OX-lines. Error bars represent SD. Plants were grown under normal conditions in the green house. Leaf tissues for RNA isolation were collected from young rosette plants. Three independent mutant lines were used from each construct.

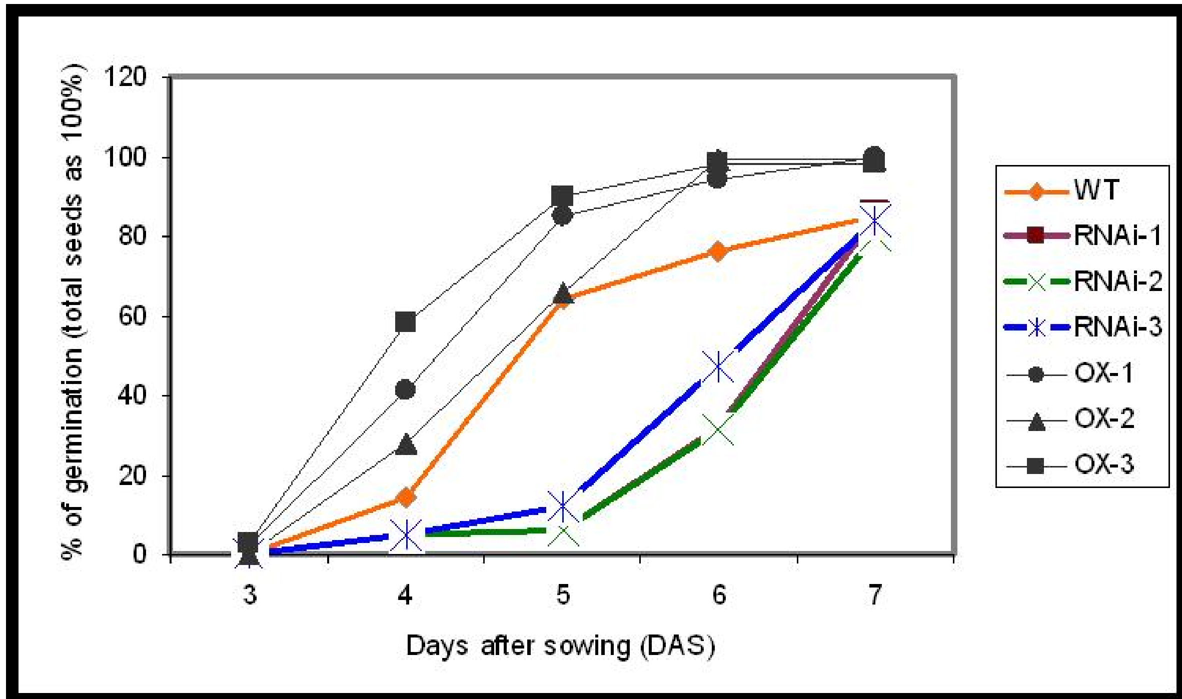
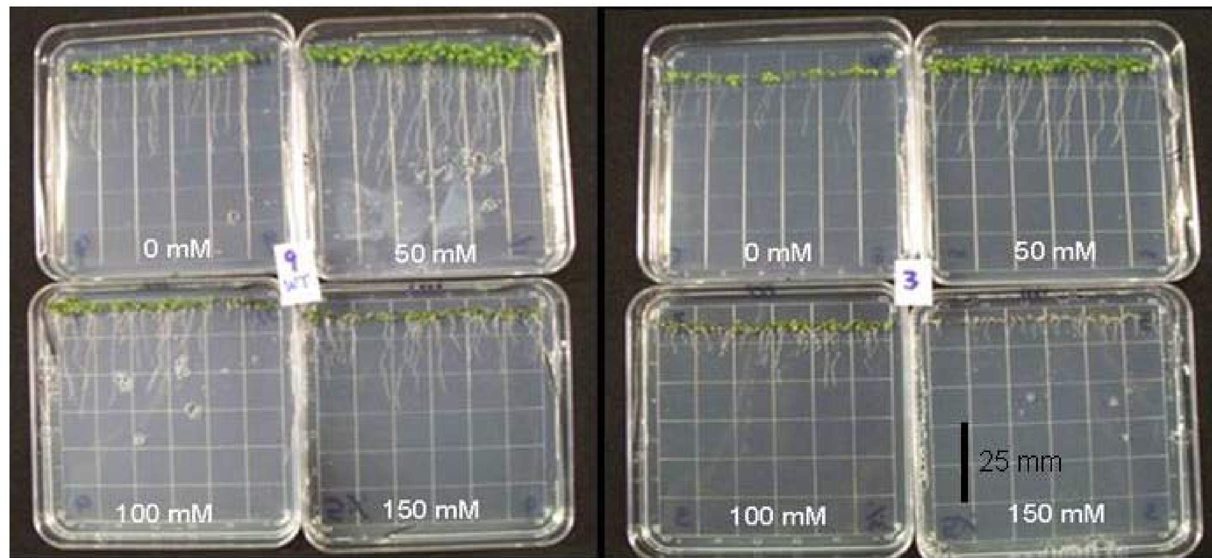


Figure 4-10: Percentage of seed germination in petri-dishes containing MS-agar medium supplemented with 100 mM sucrose. The data were recorded 3, 4, 5, 6 and 7 days after sowing (DAS). Seeds with 1.0 -1.5 mm radicle length were counted for determining germination rate. WT: wild type, OX: over-expression lines, RNAi: RNA interference lines. The experiment was performed using randomized complete block design. There were 5 replicated plates for each line containing 20 seedlings per plate. Each plate was treated as a separate block. Standard deviation ranges from 0 to <10%. The percentage of germination of mutant lines was significant different from that of the WT at 4, 5, and 6 DAS ($p<0.05$).



A B

C

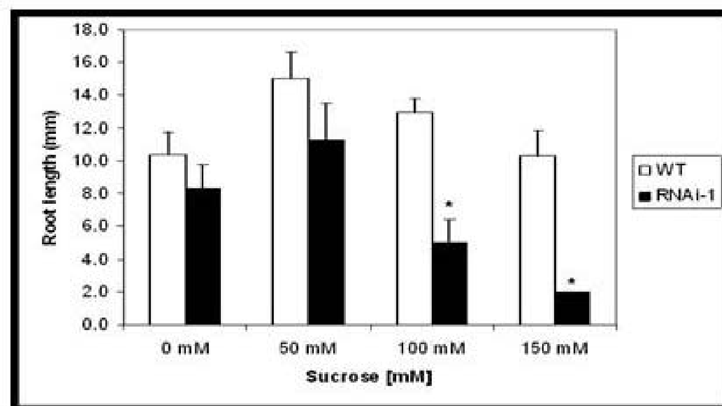


Figure 4-11. Root elongation assay of RNAi and WT seedlings in ½ x MS medium supplemented with various sucrose concentrations. Pictures were taken 12 DAS. There were 3 transgenic lines for the RNAi construct but results of one of the lines are presented as other lines produced similar results. Each treatment was replicated 3 times with ~ 20 seeds per replicated plate. The experiment was conducted using RCB design. Each plate was considered as one block. **A**: WT seedlings, **B**: RNAi seedlings; **C**: Root length of seedlings recorded 9 DAS. *indicates significantly different from WT ($p<0.05$). Error bars indicate SD. The size of the square plates = 90 x 15 mm

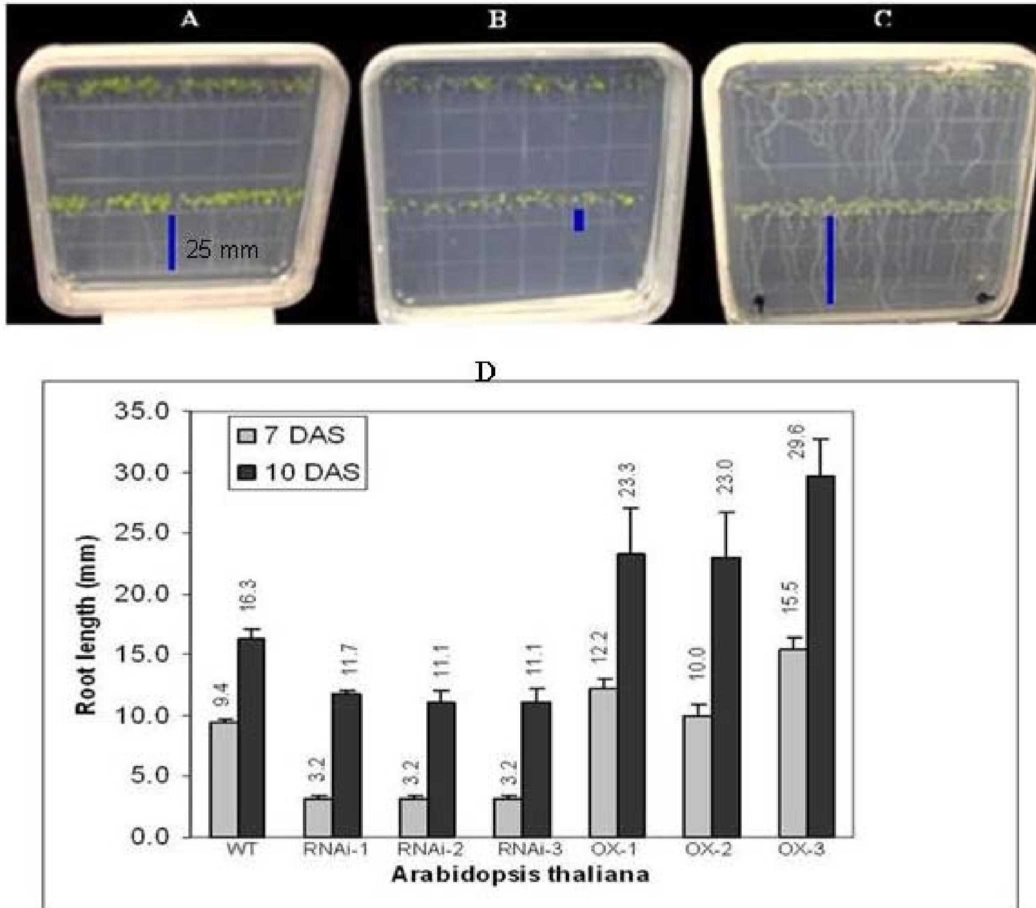


Figure 4-12. Root growth of mutant (RNAi and OX) and WT seedlings in ½ x MS-agar medium supplemented with 100 mM sucrose. Photograph was taken 13 DAS. **A**: WT, **B**: RNAi, and **C**: OX lines. **D**: Quantitative root length data recorded 7 and 10 DAS. Seedlings with >2 mm radicle length were used for measuring root length. The experiment was conducted using RCB design with 3 replicated plates for each line. Root lengths of the mutant (RNAi and OX) lines were significantly different from that of the WT at both recording times ($p<0.05$). Error bars indicate SD. The size of the square plates = 90 x 15 mm.

Analysis of over-expression mutant lines

To determine the effects of up-regulation of the *PELPK1* on transcript abundance and phenotype of the transgenic plants, over-expression (OX) lines were developed using a constitutive CaMV35S promoter (Benfey, Chua 1990; Benfey et al., 1990; Bert et al., 1999; Jackson et al. 2002, Llyod 2003). Following confirmation of the insertion of transgene construct into the OX-lines by PCR (Figure 4-6B), three independent T_3 lines were subjected transcript analysis by qRT-PCR using primer pair I (Table 4-1). Results presented in Figure 4-9B demonstrated that the transcript level in these lines was increased from 40 to 70 fold compared to WT plants, suggesting that *PELPK1* was significantly up-regulated in OX-lines.

The T_3 OX lines were then subjected to phenotypic and growth analyses by growing them in MS-agar medium as well as in the soil as described above. Plants grown in MS medium supplemented with 100 mM sucrose exhibited significantly faster ($p<0.01$) seed germination (Figure 4-10), and higher ($p<0.01$) root elongation (Figure 4-12) as compared to the WT plants. The responses of OX- lines in MS medium supplemented with other abiotic stress inducing factors or hormones described above were not significantly different from that of the WT plants (data not presented).

The phenotype of the above OX lines was further examined under soil conditions in a growth room as described above. In the soil, the emergence of the OX-plants was faster, their size was larger, their color was slightly lighter (Figure 4-13A), and their early vegetative growth was also faster (Figure 4-13C) compared to that of the WT plants. Quantitative determination of plant height as described above and presented in Figure 4-14 indicated that the shoot length of these plants recorded 25, 30, and 40 DAS was much longer than that of the WT plants. Statistical analysis of these data by a two-factor ANAOVA as described above showed that the mean plant height between WT and OX-lines was significantly different ($p<0.01$) at all recording times (25, 30 and 40 DAS).

The flowering of OX plants recorded as mentioned above was also significantly earlier ($p<0.01$) than that of the WT plants (Figures 4-15). The calculated time required for flowering of 50% OX plants was ~23 DAS as compared to ~25 DAS for WT plants. However, the dry weights of shoot and mature seeds per plant, determined as mentioned above were not significantly different from that of the WT plants (data not presented). The morphology of dry and mature seeds of the mutant and WT plants also did not significantly differ (not shown).

Furthermore, in order to understand whether RNAi or OX modified the anatomy of these seeds, microtomic analysis was carried out as described earlier (Chapter 3). The aleurone cells of RNAi seeds appear to be relatively thicker and OX seeds appear to be thinner as compared to that of WT aleurone cells (not presented). However, whether these differences in seed coat aleurone layers contributed to the differences in the seed germination or growth of mutant vs WT plants are not known.

Because expression analysis (Chapter 3) showed that *PELPK1* is induced by pathogen infection, the responses of the *PELPK1*-mutagenized (RNAi and OX) seedlings to the bacterial pathogen, *Pseudomonas syringae* were studied along with the WT seedlings as control either in absence or in presence of the defense hormones, MeJa or SA. No differences in the response of these mutants to this pathogen were observed as compared to WT plants under any of the above two conditions (not presented).

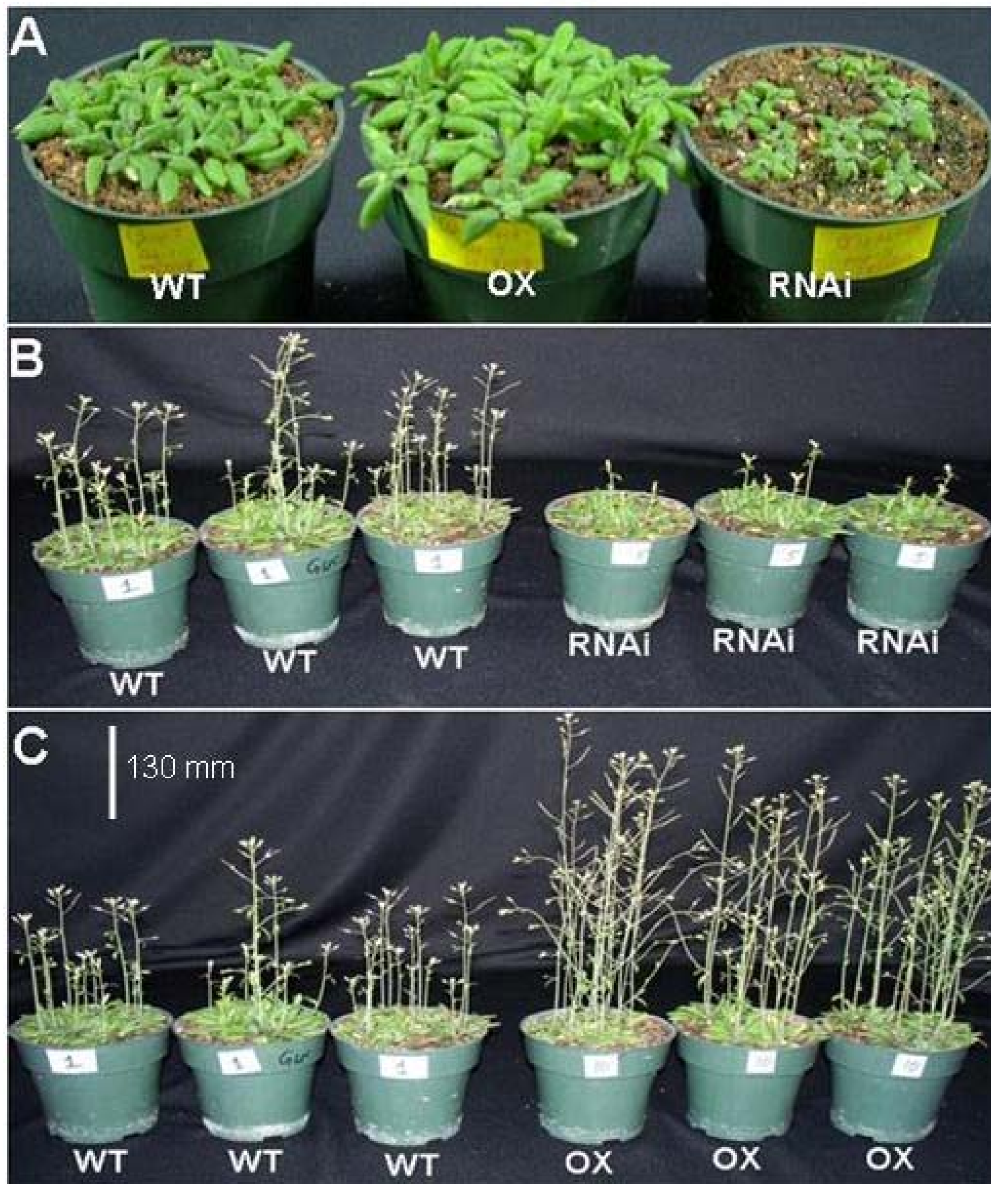


Figure 4-13: Plant morphology and growth characteristics of T_3 mutant (RNAi and OX) lines and WT plants in the soil. **A**: Phenotype of the plants 15 DAS; **B & C**: Phenotype of the plants 28 DAS (from a separate experiment). Experiments were performed using a RCB design. There were 3 replicated pots for each of three lines from each construct (RNAi and OX) tested including WT. Each pot containing 12 plants was treated as a separate block. Plants were watered as required with equal volume of water to each pot. The size of the pots = 180 x 130 mm

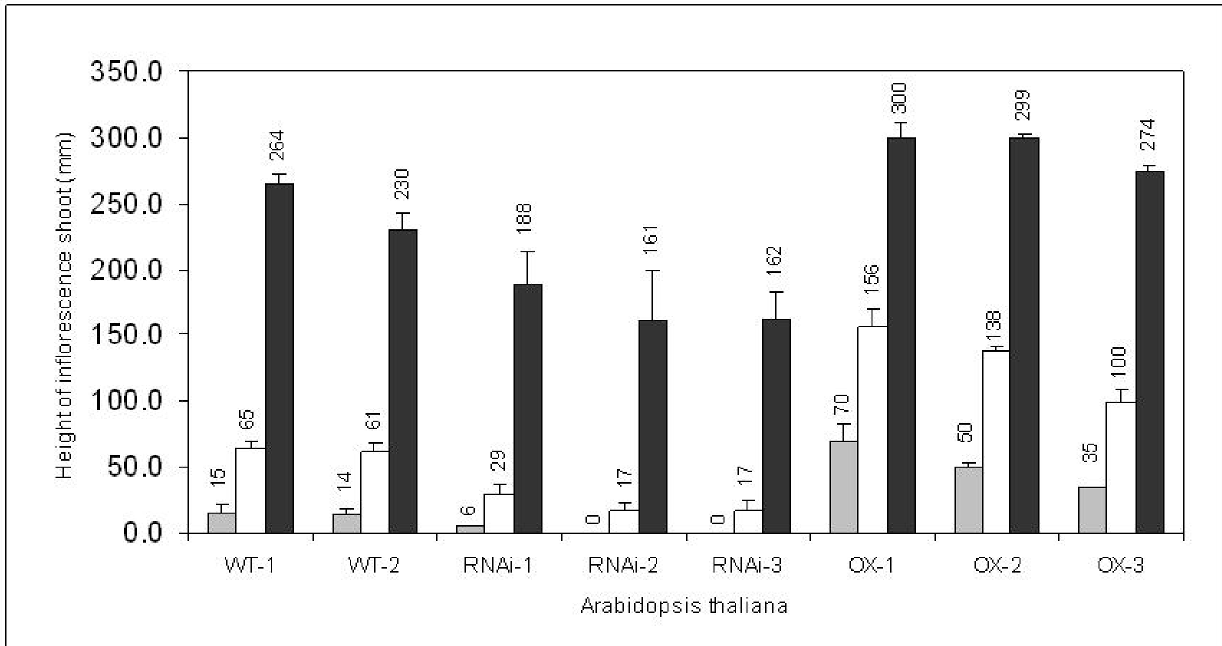


Figure 4-14. Height of inflorescence shoots recorded from the plants photographed in Figure 4-13 (**B & C**) at 25 DAS (gray bars), 30 DAS (white bars), and 40 DAS (black bars). WT: wild type, RNAi: RNA interference lines, OX: over expression lines, DAS: days after sowing. Measurements were taken from the base of the rosette to the tip of the shoot apex. There were three replications for each line with ~12 plants per replication. The shoot lengths of the mutant (RNAi and OX) lines were significantly different from that of the WT at 25 and 30 DAS ($p<0.01$). Error bars indicate SD.Other conditions are as described in Figure 4-13.

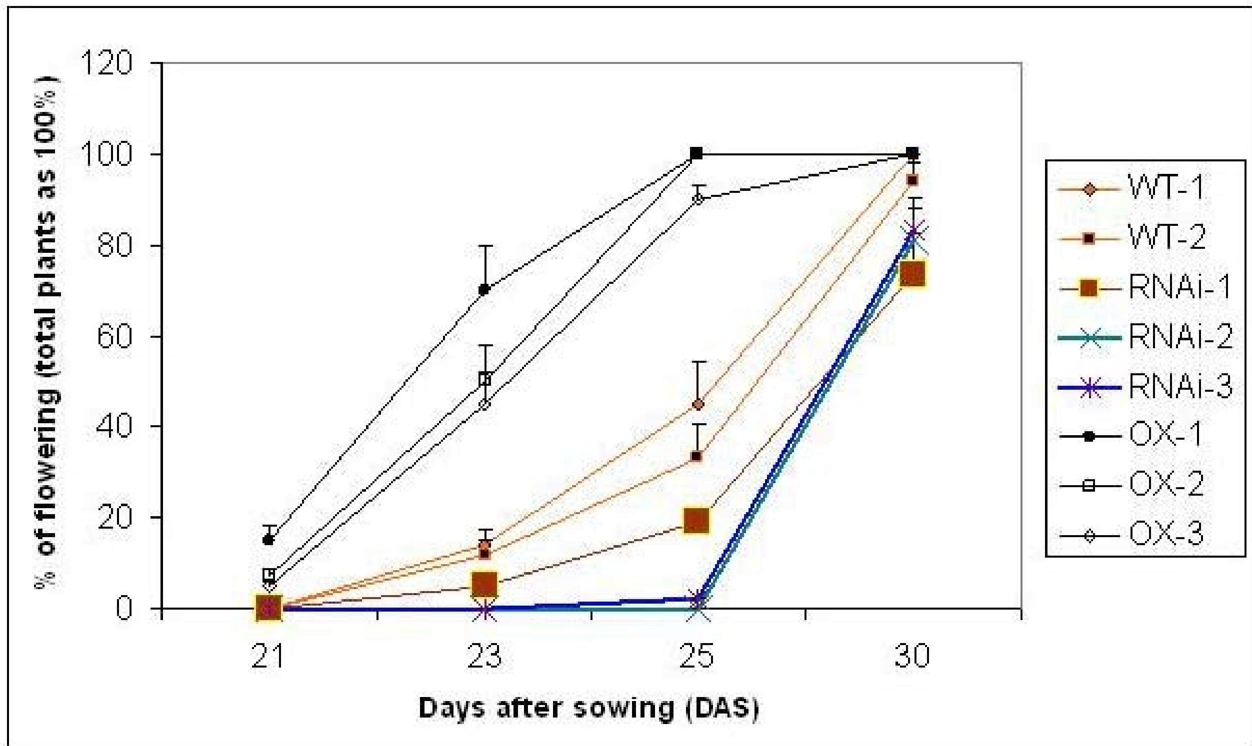


Figure 4-15. Percentage of flowering of mutant (RNAi and OX) and WT plants recorded 21, 23, 25, and 30 DAS. Plants showing flower buds (refer to Figure 4-13 B&C) were counted for determining flowering percentage using total number of plant as 100%. WT: wild type, RNAi: RNA interference lines, OX: over expression lines. The percentage of flowering of mutants was significantly different from that of the WT at 25 DAS ($p<0.05$). Other conditions are as described in Figure 4-13.

4. DISCUSSION

Sequence-indexed T-DNA insertion lines, which were confirmed to be homozygous for insertions within or near *PELPK1* (Figure 4-2), failed to show any phenotypic differences from WT and did not appear to have a reduction in the total *PELPK1* + *PELPK2* transcript abundance that could be detected by qRT-PCR (Figure 4-8). It can be inferred that genetic redundancy (Jander, and Barth, 2007) or no null mutation including gene knock about (Krysan et al., 1999), or a combination of both might be responsible for the absence of any detectable effect of the insertions. Because of tight genetic linkages between tandemly duplicated genes (Jander and Barth, 2007), no attempt was made to obtain double homozygous knockout lines for both *PELPK1* and *PELPK2*. Instead, RNAi and over-expression methodologies (Helliwell et al., 2002; Helliwell and Waterhouse 2003; Wielopolska et al., 2005; Lloyd 2003) were applied.

Transcript expression analysis of the RNAi lines using the primer pairs mentioned above (Table 4-1) showed that the abundance of detectable transcript was reduced from 65 - 80% of WT (Figure 4-9). Primer pair III, which has been shown to be *PELPK1*-specific, showed a strong reduction in target amplification (Figure 4-9), thus transcript expression of at least the *PELPK1* was greatly reduced in the RNAi lines. It is likely that the expression of the *PELPK2* was also reduced, but this cannot be concluded definitively, due to the high sequence similarity of the *PELPK1* and *PELPK2*. Nevertheless, it can be reasonably assumed that simultaneous silencing of both *PELPK1* and *PELPK2* genes occurred in *Arabidopsis* plants because: (i) the RNAi gene fragment was amplified from WT genomic DNA using the conserved sequences of both the genes, and contained respectively 100% and 95% sequence identity with the *PELPK1* and *PELPK2* gene, and (ii) clear phenotypes were observed in the RNAi lines, but not in the KO mutant lines, which inhibited at most one of the two paralogous genes.

Seed germination and root elongation

The mutant RNAi plants grown in MS-agar medium supplemented with 0-50 mM sucrose did not show significant differences in the rate of seed germination (not shown) or extent of root elongation (Figure 4-11) as compared to WT plants. However, when they were grown in the same medium but supplemented with 100 mM sucrose, seed germination and root elongation were significantly slowed down as compared to that of the WT plants (Figure 4-10; 4-11; 4-12). The calculated time required for 50% seed germination was >1.5 days longer in RNAi plants compared to WT plants. Similarly the root length of RNAi plants on 100mM sucrose supplemented MS-agar medium was respectively 2.9x and 1.5x shorter than WT plants when measured 7 and 10 DAS, respectively (Figure 4-12). These observations suggest that RNAi made the plants less efficient in utilizing sucrose for their root growth than WT plants. This modified response of RNAi plants to slightly elevated level of sucrose does not appear to be due to an increased susceptibility of these plants to osmotic stress because the response of these mutants to other abiotic stresses including high concentrations of mannitol (250 mM) did not significantly differ from that of the WT plants (data not presented). It might be possible that RNAi negatively modified sucrose metabolic and/or signaling pathway resulting in the reduction of cell wall extensibility, and as a consequence at only slightly elevated level of sucrose (100 mM), these mutant plants responded negatively with delayed seed germination and reduced root elongation when compared to WT plants.

In contrast, the OX-plants sown in similar MS-agar medium supplemented with similar sucrose concentration (100 mM) showed relatively faster seed germination (Figure 4-10) and significantly higher root elongation (Figures 4-12) as compared to WT plants. The calculated time required for 50% seed germination was >1 day earlier in OX-plants compared to WT plants. Similarly, the root length of OX-plants was respectively 1.4x and ~1.6x longer than WT plants when determined 7 and 10 DAS. Furthermore,

even in the presence of 200 mM sucrose, a concentration which may cause slight osmotic stress in *Arabidopsis* plants (Neales 1968) the root growth of OX-plants was significantly higher than that of WT plants. However, the response of OX-roots to other abiotic stress-inducing factors such as high concentrations of mannitol (250 mM), that may cause osmotic stress, or NaCl (150 mM), that may cause salt stress or ABA (10 µM) were not significantly different from that of the WT plants (data not presented). The above results suggest that the OX-plants might have effectively metabolized sucrose into more useable form(s) and/or up-regulated the genes involved in the complex sucrose metabolic/signaling pathways and consequently exhibited enhanced growth.

The results presented above based on mutational analysis suggest that *PELPK1* is somehow linked to sucrose utilization in *Arabidopsis* plants. A slow-down of sucrose utilization in RNAi mutants particularly in the presence of slightly elevated level in the medium can, not only slow-down cell wall loosening causing slow growth but also can elevate endogenous level in the plants. Although the exact mechanism of sucrose involvement in the functional activity of *PELPK1* is unclear, it has previously been suggested that elevated level of sucrose can interfere with nutrient uptake and affect root elongation (Hammond and White 2008).

It might be possible that in *Arabidopsis* plants, the *PELPK1* acts as a precursor for peptide hormone signal for activating the genes involved in sucrose metabolic and/or signaling pathway. Thus, in RNAi mutants as a result of RNAi of the *PELPK1* the above pathway is repressed and as a consequence these mutants may become unable to utilize even slightly elevated level of externally applied sucrose, causing a slow-down of seed germination and root elongation as compared to WT plants. In contrast, in OX mutants an OX of *PELPK1* enhanced the above pathway causing more efficient utilization of sucrose and thereby accelerated seed germination and root growth, in the presence of elevated level of externally added sucrose.

Shoot growth and flowering

The vegetative growth of shoot was slower and the flowering was delayed in RNAi plants in the soil compared to that of WT plants (Figures 4-13; 4-14; 4-15). The height of inflorescence shoot determined 25, 30, and 40 DAS was respectively 12.5x, 3x and 1.5x shorter in RNAi plants compared to that of WT plants (Figure 4-14). RNAi also caused a delay of flowering for 2 days when expressed on 50% population basis (Figures 4-15). Furthermore, visual observation of RNAi plants in the early stages of vegetative growth showed that although exhibited normal growth, these plants were slow to emerge from the soil, smaller in size, and had slightly darker leaves (Figure 4-13A) as compared to WT plants. The above symptoms observed in the soil-grown RNAi plants might be linked to perturbed sucrose metabolism and/or signaling in RNAi plants as discussed above.

In contrast, OX of *PELPK1* significantly enhanced shoot growth and flowering in OX-plants in the soil as compared to WT plants (Figures 4-13, 4-14, 4-15). The height of inflorescence shoot determined 25, 30, and 40 DAS was respectively 3.6x, 2.1x and 1.2x longer in OX-plants compared to that of WT plants (Figure 4-14). The flowering of OX plants recorded as mentioned above was also 2 days earlier than that of WT plants (Figures 4-15). In addition, visual observation of these plants as described above showed that their emergence from the soil was faster and their size was larger (Figure 4-13A & C) as compared to WT plants. Although the above growth characteristics of OX-plants in the soil appear to suggest that gibberellins might be responsible for these phenotypes, studies involving the responses of OX and RNAi plants to exogenously applied GA_3 did not support this assumption as neither of these mutants differentially responded to GA_3 compared to WT plants.

Finally, since functional expression of the phenotype of *PELPK1* in RNAi versus OX plants grown under similar conditions in the soil was opposite to each other: RNAi

significantly slowed down the shoot growth while OX enhanced the growth, it might be possible that RNAi repressed cell wall loosening while OX enhanced the same. Although it is not clear what mechanism or factor is involved in causing these differential effects on cell wall loosening, differential modifications of sucrose metabolic or signaling pathway in RNAi versus OX plants might be responsible.

5. CONCLUSIONS

It is concluded based on mutational analysis that the *PELPK1* function in *Arabidopsis* plants is associated with sucrose-related processes of growth or signaling. This conclusion is made based on the observations that the RNAi and OX mutant plants differentially responded to a level of exogenously added sucrose in the MS-agar medium as compared to the WT plants. It is further concluded based on soil experiments that RNAi and OX differentially modified the cell wall loosening in the mutant plants as compared to WT plants. Although it is not clear how modification of sucrose metabolism and/or signaling affects cell wall loosening and plant growth in *Arabidopsis* plants, it has previously been suggested that elevated level of endogenous sucrose in plants can negatively affect nutrient uptake and consequently plant growth.

5. REFERENCES

Allen, R.S. Li, J, Stahle, MI., Dubroue, AI, Gubler, F, and Millar, AA. (2007). Genetic analysis reveals functional redundancy and the major target genes of the Arabidopsis miR159 family. PNAS 104: 16371–16376

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., Ecker, J.R. (2003) "Genome-wide insertional mutagenesis of Arabidopsis thaliana" Science 301: 653-657

An, G, Jeong, D-H, Jung, K-H, and Lee, S. (2005). Reverse genetic approaches for functional genomics of rice. Plant Molecular Biology 59:111–123

Anonymous (1999). Phosphorus Deficiency Symptoms in Some Crops. Better Crops. 83: 20-23.

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

AzpirozLeehan, R. & Feldmann, K.A. (1997) T-DNA insertion and mutagenesis in Arabidopsis: going back and forth. Trends Genet. 13, 152 – 156

Benfey, PN, Chua, N-H (1990). The Cauliflower Mosaic Virus 35S Promoter: Combinatorial Regulation of Transcription in Plants. Science 250: 959-966

Benfey, P. N., Ren, L., and Chua, N. H. (1990) Combinatorial and synergistic properties of CaMV 35S enhancer sub-domains. EMBO J. 9, 1685–1696.

Bert J. van der Zaal, Leon W. Neuteboom, Johan E. Pinas, Agnes N. Chardonnens, Henk Schat, Jos A.C. Verkleij, and Paul J.J. Hooykaas (1999). Over-expression of a Novel Arabidopsis Gene Related to Putative Zinc-Transporter Genes from Animals Can Lead to Enhanced Zinc Resistance and Accumulation. Plant Physiol. 119: 1047–1056

Bouchéa, N and Bouchez, D (2001). Arabidopsis gene knockout: phenotypes wanted. Current Opinion in Plant Biology, 4: 111-117

Chuang, C.F. and Meyerowitz, E.M. 2000. Specific and heritable genetic interference by double stranded RNA in Arabidopsis thaliana. Proc. Natl. Acad. Sci. U.S.A. 97: 4985–4990.

Feldmann, K.A.1991. T-DNA insertion mutagenesis in Arabidopsis: mutational spectrum. Plant J. 1: 71–82.

Fredslund, J. (2007). Primique: automatic design of specific PCR primers for each sequence in a family. BMC Bioinformatics 8:369

Gilchrist, E and Haughn, G (2010). Reverse genetics techniques: engineering loss and gain of gene function in plants. Briefing functional genomics 9, 103-110

Hammond, JP, and White, PJ (2008). Sucrose transport in the phloem: integrating root responses to phosphorus starvation. Journal of Experimental Botany. 59: 93–109, 2008

Helliwell, C.A., Wesley, S.V., Wielopolska, A.J. and Waterhouse, P.M. (2002) High-throughput vectors for efficient gene silencing in plants. Funct. Plant Biol. 29, 1217–1255.

Helliwell, C. and Waterhouse, P. (2003) Constructs and methods for high-throughput gene silencing in plants. Methods, 30, 289–295.

Hirsch, R.E., Lewis, B.D., Spalding, E.P., and Sussman, M.R. (1998). A role for AKT1 potassium channel in plant nutrition. Science 280: 918–921.

Jackson RG, Kowalczyk M, Li Y, Higgins G, Ross J, Sandberg G, Bowles DJ (2002). Over-expression of an Arabidopsis gene encoding a glucosyltransferase of indole-3-acetic acid: phenotypic characterisation of transgenic lines. Plant J. 2002 Nov;32(4):573-83

Jander, G and Barth, C (2007). Tandem gene arrays: a challenge for functional genomics. Trends in Plant Science, 12: 203-210

Jeon, J.S., Lee, S., Jung, K.H., Jun, S.H., Jeong, D.-H., Lee, J., Kim, C., Jang, S., Yang, K., Nam, J., An, K., Han, M.J., Sung, R.J., Choi, H.S., Yu, J.H., Choi, J.H., Cho, S.Y., Cha, S.S., Kim, S.I., An, G. (2000) "T-DNA insertional mutagenesis for functional genomics in rice" Plant J. 22: 561-570

Krysan, P.J., Young, J.C., Tax, F., and Sussman, M.R. (1996). Identification of transferred DNA insertions within Arabidopsis genes involved in signal transduction and ion transport. Proc. Natl. Acad. Sci. USA 93, 8145–8150.

Krysan, P.J., Young, J.C., Tax, F., and Sussman, M.R. (1999). T-DNA as an insertional mutagen in Arabidopsis. The Plant Cell 11: 2283–2290,

Lloyd, A. (2003) Vector Construction for gene Over-expression as a tool to elucidate gene function. Methods in Molecular Biology, vol. 236: Plant Functional

Genomics: Methods and Protocols, Edited by: E. Grotewold © Humana Press, Inc., Totowa, NJ

Narváez-Vásquez J, Pearce G, abd Ryan CA (2005). The plant cell wall matrix harbors a precursor of defense signaling peptides. Proc Natl Acad Sci. 102:12974-129777

Neales, TF (1968). The nutritional requirements of excised roots of three genotypes of Arabidopsis thaliana. New Phytol. 67, 159-165.

Onouchi, H, Igeño, MI, Périlleux, C, Graves, K and Coupland, G. (2000). Mutagenesis of Plants Over-expressing CONSTANS demonstrates novel interactions among Arabidopsis flowering-time genes. The Plant Cell, 12, 885–900

Pan X., Liu H., Clarke J., Jones J., Bevan M. and Stein L. (2003); ATIDB: Arabidopsis thaliana insertion database; Nucleic Acids Research 31(4): 1245-1251

Pfaffl, M.W. (2001). A new mathematical model for relative quantification in real time RT-PCR. Nucleic Acid Research. 29: 2002-2007.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Schreiber, K, Ckurshumova, W, Peek, J, and Desveaux, D (2008). A high-throughput chemical screen for resistance to *Pseudomonas syringae* in *Arabidopsis*. The Plant Journal 54: 522–531

Sessions, A, Burkea, E, Prestinga, G, Auxb, G, McElverb, J, Pattonb, D, Dietrichb, B, Hoa, P, Bacwadena, J, Koa, C, Clarkea, CD, Cottona, D, Bullisa, D, Snella, J, Miguela, T, Hutchisona, D, Kimmerly, B, Mitzelc, T, Katagiria, F, Glazebrooka, J, Lawb, M and Goffa, SA. (2002). A High-Throughput Arabidopsis Reverse Genetics System. The Plant Cell, 14: 2985-2994.

Wielopolska, A., Townley, H., Moore, I., Waterhouse, P and Helliwell, (2005). A high-throughput inducible RNAi vector for plants. Plant Biotechnology Journal. 3: 583–590

CHAPTER 5

PROTEOMIC ANALYSIS OF *ARABIDOPSIS* SEED COAT/WHOLE SEED

CHAPTER 5

PROTEOMIC ANALYSIS OF *ARABIDOPSIS* SEED COAT/WHOLE SEED

1. INTRODUCTION

Because the *PELPK1* transcript appears to be normally expressed during seed germination, based on promoter::reporter analysis (Chapter 3), proteomic analysis was carried out to see whether the *PELPK1* could be detected among the proteins of a germinating seed. The process of seed germination in *Arabidopsis* begins with the uptake of water by the dry seed through imbibition and end with the protrusion of radicle through seed coat (Bewley 1997, Koornneef et al. 2002, Kucera et al. 2005). The seed coat of this plant was shown to be composed of, excluding the outer mucilage layer, two cell layers surrounding the embryo: the testa and the endosperm (Figure 5-1; also refer to Müller et al. 2006 and http://www.seedbiology.de).

The endosperm, also called aleurone layer, is predominantly composed of a single layer of cells (Pritchard et al. 2002, Nambara and Marion-Poll 2003, Liu et al. 2005). In a mature seed, the testa is composed of non-living cells, but the aleurone cells are living. The testa provides protection to embryo, whereas the aleurone cells provide nutrition to the embryo during seed germination and perhaps also early seedling growth (Vaughan et al. 1971, Corner 1976, Ruiz and Escale 1995, Nguyen et al. 2000; Muller et al., 2006). Interestingly, however, seeds of some related genera such as *Sinapis alba*, *Raphanus sativus*, and *Brassica* sp are devoid of endosperm/aleurone layer (Schafer and Peachy 1984, Barged and Schafer 1986, Schafer et al. 2001).

Because *PELPK1* is specifically expressed in the aleurone layer (Chapter 3), which remains attached with the seed coat even after the separation of the seed coat from the embryo following germination (Muller et al., 2006), proteomic analysis was carried out on both the seed coat as well as whole germinated seed of *Arabidopsis thaliana*. The

objectives were (i) to identify the PELPK1 protein in the seed coat and/or in the germinated whole seed, and (ii) to separate the seed-coat proteomic profile from that of whole seed. This is the first report that made an attempt to separate the seed coat proteins from that of whole seed in *Arabidopsis thaliana*.

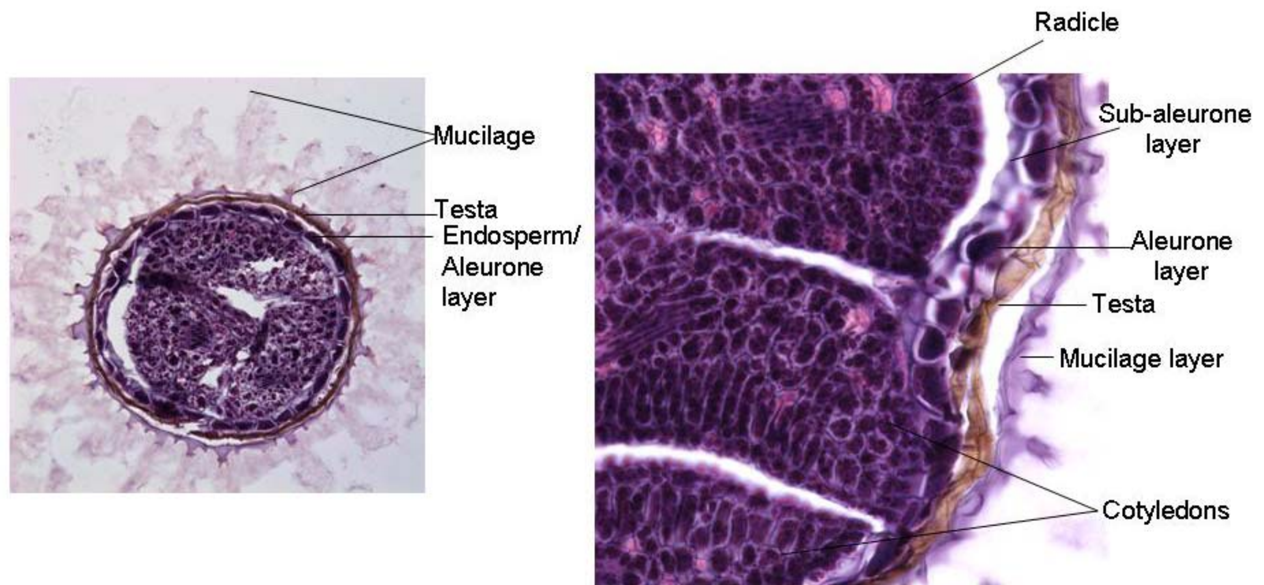


Figure 5-1: Microtome sectioning through mature *Arabidopsis* seeds showing among others aleurone and sub-aleurone layers of the seed coat. The seed coat of a mature *Arabidopsis* seed is composed of an outer mucilage layer, followed by a non-living testa, and an innrer aleurone layer composed of a single layer of cells surrounding the embryo.

2. MATERIALS AND METHODS

Separation of seed coat

Wild-type *Arabidopsis* seeds (5 -10g) were surface sterilized, and cold incubated in the dark for two days as described earlier, and then spread over ½ x MS medium (pH 5.7 by KOH) containing 50 mM sucrose, and 0.7% phytablend. After germination, when the shoots were about ~ 0.5 -1 cm long, seed coats were collected from the surface of the medium by repeated rinsing with sterile dH_20. The resulting seed coats were collected by centrifugation at 5000xg for 5 min, air dried at room temperature, and then frozen in liquid nitrogen.

Extraction of total proteins

The seed coats or germinated seeds (~0.35g) were ground into fine powder by using mortar and pestle in liquid nitrogen, and then transferred to three 2-mL tubes. Total proteins were extracted following Harder et al., (1999) with modifications. The powder in each tube was suspended by adding 1 ml of lyses buffer containing 7M urea, 2M theorem, 4% (w/v) 3-[(3-cholamidopropryl)dimethylammonio]-1-propanesulfonate (CHAPS), 1% (v/v) ampholyte, 18 mM Tris-HCl, and 14 mM Trisha base (pH 8.5). The samples were mixed vigorously, and then added with 53 U/ml DNase I, 5 U/ml RNase A, and 0.2% (v/v) Triton X-100. After 10 min incubation at 4°C, 14 mM dithiothreitol (DTT) was added to each sample and the protein extracts were stirred for 20 min at 4°C, then centrifuged (35000xg, 10 min) at 4°C. The supernatants were transferred to new tubes and subjected to a second centrifugation as above. The final supernatant in each tube corresponding to the total protein extract was subjected to desalting (Thermo Scientific), followed by lyophilization over night. Each sample was then dissolved in 300 µL of rehydration buffer (RE) containing RE (BioRAD): TBP (tyributyl phosphate) in the ratio of

100:1. Protein concentration was determined by Bradford (1976) using bovine serum albumin as a standard.

One-dimensional gel electrophoresis (1D PAGE)

The proteomic analysis was carried out following Pocsfalvia et al., (2006) with modifications. 75 μg total protein of each sample was boiled at 70°C for 10 min and loaded on to 10% polyacrylamide gel. Electrophoresis was performed using Hoefer MiniVE electrophoresis apparatus, at 90V until the dye front reached the bottom edge of the gel using Tris–glycine–SDS buffer system (25 mM Tris, 192 mM glycine and 0.1% SDS).

For in-gel digestion of proteins, gel lanes were cut manually into 1 mm slices and placed separately into eppendorf tubes. Reduction, alkylation and hydrolysis of proteins from gel bands were carried out following Jensen et al. (1999) with modifications. In brief, gel pieces were first washed with 150 μL of HPLC-grade water (Fisher Scientific), and then dehydrated two times with 50 μL of 100% acetonitrile (ACN). Following reduction by adding 30μL of 10 mM DTT/0.1 M NH_4HCO_3 at 56°C for 30 min, they were dehydrated, and alkylated by adding 30 μL of 55 mM iodoacetamide/0.1 M NH_4HCO_3 at room temperature for 20 min in the dark. The samples were then rinsed with 200 μL of 0.1 M NH_4HCO_3 and dehydrated with ACN, and dried in a SpeedVac for 10 min. Each sample was then added with 20 μL of trypsin solution containing 0.02 μg/ μL Trypsin Gold (Promega) in 40 mM NH_4HCO_3/10% ACN, and incubated on ice for 1 h followed by at 37°C overnight. Each sample was then added with 3 μL of 2% formic acid to stop the digestion reactions. Finally, the peptide extract from each tube was transferred to a fresh tube, followed by two more extractions of peptides with 15 μL each of 50%

ACN/0.1% formic acid. The total extract (~ 55 µL) was mixed well and stored at –20°C until used.

Two-dimensional gel electrophoresis (2D PAGE)

2D gel electrophoresis of protein extracts was carried out following Cao et al., (2008) with modifications. 17 cm IPG strips of pH 4-7 linear (Bio-Rad) were used for first dimensional separation of proteins. A total of 100 µg of protein in 125µL rehydration buffer (8 M urea, 2% CHAPS, 40 mM DTT, 0.2% Bio-Lyte, and 2 mM TBP) were used to hydrate the IPG strips in a res-welling tray for 8-10 hrs at room temperature. Isoelectric focusing was conducted using a Bio-Rad PROTEAN IEF unit under a maximum field strength of 600 V/cm and a 50 mA limit per strip. The IPG strips were initially held at 250 V for 15 min to remove charged ion contaminants, followed by a linear voltage ramping step to achieve 4000 V in 2 h. Following that the IEF proceeded at a constant 4000 V for 20,000 V h without exceeding the limit of 50 mA/ strip. A 500 V holding step was included at the end of focusing to prevent protein diffusion and minimize over-focused artifacts, until a manual disruption of the run.

After focusing, the IPG strips were equilibrated for 20 min at room temperature in 50 mM Tris–HCl (pH 8.8) containing 6 M urea, 20% glycerol, 2%(w/v) SDS, and 2% (w/v) DTT, followed by another 20 min in the same Tris–HCl buffer, but with 2.5%(w/ v) iodoacetamide used in place of 2% (w/v) DTT, to solubilize, reduce, and alkylate the proteins for SDS binding. 2D gel electrophoresis was performed on 10% polyacrylamide gels (8.2 cm x 8.2 cm, 1 mm thickness) using a Mini PROTEAN 3 system (Bio-Rad) at a constant voltage of 150 V for 70 min. Gels were stained using a Colloidal Blue Staining Kit (Invitrogen Co., Carlsbad, CA), destained in sterile dH_2O overnight, and then scanned using a GS-800 Calibrated Densitometer (Bio-Rad).

Twenty eight protein spots including the ones showing pI values and molecular weights closer to the GOI (pI ~ 5.87; mw ~ 41.6 kDa) were excised manually from the gel and processed for in-gel digestion of peptides as described above.

Liquid chromatography/Tandem mass spectrometry (LC-MS/MS)

LC-MS/MS analysis of the trypsin-digested peptide mixtures was performed using an Agilent 1100 LC/MSD Trap XCT (Agilent Technologies) as described earlier (Jiang et al., 2007). An auto-sampler was used to inject 20-40 µL of each sample onto the first of two C-18 columns. This short 5 µm enrichment column Zorbax 300SB-C18, 5 µm, 5x0.3 mm, served to trap and concentrate the samples. After that, the sample was eluted onto the next C-18 column (Zorbax 300SB-C18, 5 µm, 150 x 0.3 mm), which was used in conjunction with a solvent gradient to separate the peptides. The peptide-separation gradient started at 85% solvent A (0.1% FA in dH_2O) and ended at 55% solvent B (0.1% FA, 5% dH_2O in ACN) over a 42 min span. This was followed by a 10 min period of 90% solvent B to cleanse the columns before returning to 97% solvent A for the next sample. The ion trap mass spectrometer collected information by first running an MS 300–2200 m/z scan and followed that with an MS/MS analysis of the most intense ions. In addition to the most intense ion for each scan, the software was set to exclude this ion after two spectra and gather MS/MS information on the next most intense ion(s). Raw spectral data were processed into Mascot Generic File (.mgf) format using the default method in the ChemStation Data Analysis module. The MS/MS ion search was performed using MASCOT (http://www.matrixscience.com) searching the NCBInr database and taking green plant as the taxonomy. The parameters for searching were: significance threshold ($p< 0.05$), maximum number of hits (auto), scoring (standard), expect cut-off and sub-sets (0), enzyme (trypsin), fixed modifications (carbamidomethyl), allow up to one missed cleavages, peptide tolerance of ± 1.2 Da,

peptide charge of 1+ 2+ 3+, monoisotopic, and ESITRAP instrument. Only significant hits, as defined by the Mascot probability analysis ($p < 0.05$), were accepted. The mass spectrometer sequenced the peptides by generating fragments through collision-induced dissociation followed by measurement of the masses of the fragments. The resulting MS/MS spectrum was compared with the database of known proteins and ranked by scoring.

3. RESULTS AND DISCUSSION

Analysis of *Arabidopsis* seed coat protein profile

Promoter activity analysis as well as sub-cellular protein localization studies (Chapter 3) showed that *PELPK1* is spontaneously expressed in the seed coat of germinated seed, specifically in the aleurone layer. Bioinformatics analysis (Chapter 2) predicted *PELPK1* to encode a protein of calculated molecular weight 41.6 and a pI value of ~5.9. Thus, in order to identify PELPK1 protein from the seed coat proteomic profile, total proteins extracted from the seed coat of germinated seeds were subjected to 2D gel electrophoresis (Figure 5-2) followed by LC-MS/MS (Table 5-1).

Colloidal blue staining of a 2D gel showed a number of protein spots in a pH gradient of 4-7 IEF/12% gel (Figure 5-2). Twenty eight protein spots including the ones suspected to be PELPK1 protein based on molecular weights and pI values were excised manually from the 2D gel and subjected to LC-MS/MS following trypsinization and extraction of the peptides. Out of 28, 16 spots that provided the significant protein hits ($p<0.05$) are listed in Table 5-1 with the identities of the proteins.

As shown in Table 5-1, my analysis did not identify the PELPK1 (AT5G09530) or its paralog, PELPK2 (AT5G09520) protein. Almost all of the proteins identified from the seed coat proteomic profile belonged to the family of seed storage/nutrient reservoir proteins. These include legume-type globulins such as α and ß subunits of the 12S cruciferins. The major AGI loci involved in making up the seed coat proteomic profile in the present study include: AT4G28520 (CRU3), AT5G44120 (CRA1), AT1G03880 (CRB/CRU2) and AT1G03890 (CFP). These loci are arranged according to their number of hits/isoforms as follows: AT4G28520 (5 isoforms; spot ID: 12, 18, 19, 20, 21, 27, 28) =

AT5G44120 (5 isoforms; spot ID: 5, 10, 11, 21)> AT1G03880 (3 isoforms, spot ID: 4, 22, 25)> AT1G03890 (2 isoforms spot ID: 5, 8, 13, 18).

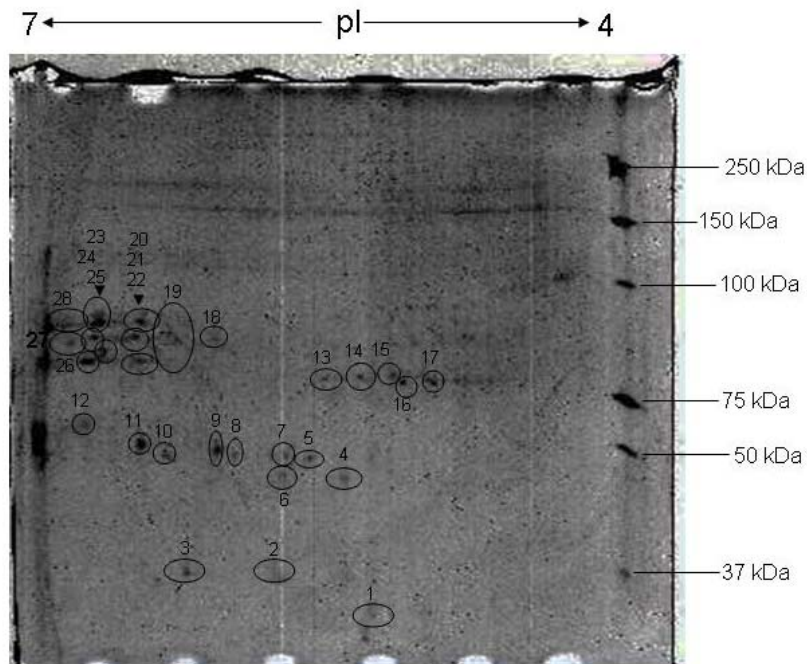


Figure 5-2: A two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) of proteins extracted from seed coats of germinated WT *Arabidopsis* seeds. The immobilized pH gradient (IPG) strip used was 17 cm; pH ranges from 4 -7 (PELPK1 MW = 41.6 kDa; pI = 5.87). Protein spots in the gel were identified using numbers, and collected manually using blunt sterile pipette tips. Experiment was repeated at least twice with similar results and the results of one experiment are presented.

.

Table 5 -1. Identity of seed coat proteins of *Arabidopsis* germinated seed as determined by 2D gel electrophoresis followed by LC-MS/MS (refer to Figure 5-2 for identity of the protein spots on 2D gel). Protein scores >48 indicate identity and extensive homology ($p<0.05$). pI (calculated pI value), Score (protein score), Mass (nominal mass), Pep (number of peptide sequence), Mat (number of peptide match), % Cov (percentage of sequence coverage), SC, seed coat

Spot	NCBI #	AGI locus	Putative protein name	pI	Score	Mass	Pep	Mat	% Cov
SC-4	62321455	AT1G03880	Putative cruciferin 12S seed storage protein	7.9	63	19.9	3	3	23
	166678	AT1G03880	12S storage protein CRB	6.8	60	50.6	3	3	9
	15219582	AT1G03880	CRU2 (Cruciferin 2)/nutrient reservoir	6.5	60	50.5	3	3	9
SC-5	166676	AT5G44120	12S storage protein CRA1	7.7	70	52.6	2	2	8
	15241422	AT5G44120	CRA1 (Cruciferina)/nutrient reservoir	7.7	70	52.6	2	2	8
	30694455	AT5G44120	CRA1 (Cruciferina)/ nutrient reservoir	6.6	70	41.0	2	2	11
	62319667	AT5G44120	Legumin-like protein	9.8	70	13.1	2	2	34
	30694452	AT5G44120	CRA1 (Cruciferina)/ nutrient reservoir	9.0	69	31.6	2	2	14
	15219584	AT1G03890	Cupin family protein (CFP)	5.5	62	49.6	2	2	6
	51970110	AT1G03890.	Putative cruciferin 12S seed storage protein	5.2	62	7.9	2	2	43
SC-8	15219584	AT1G03890	Cupin family protein	5.5	114	49.6	3	3	9
SC-10	166676	AT5G44120	12S storage protein CRA1	7.7	58	52.6	3	5	10
	15241422	AT5G44120	CRA1/nutrient reservoir	7.7	58	52.6	3	5	10
	30694455	AT5G44120	CRA1/nutrient reservoir	6.6	58	41.0	3	5	13
	62319667	AT5G44120	Legumin-like protein	9.8	58	13.1	3	5	40
	30694452	AT5G44120	CRA1 (Cruciferina)/nutrient reservoir	9.0	65	31.6	3	5	16
SC-11	166676	AT5G44120	12S storage protein	7.7	53	52.6	2	2	8
	15241422	AT5G44120	CRA1 (CRUCIFERINA)/nutrient reservoir	7.7	53	52.6	2	2	8

	30694455	AT5G44120	CRA1 (CRUCIFERINA)/nutrient reservoir	6.6	53	41.0	2	2	11
	62319667	AT5G44120	Legumin-like protein	9.8	53	13.1	2	2	35
	30694452	AT5G44120	CRA1 (CRUCIFERINA)/nutrient reservoir	9.0	51	31.6	2	2	14
SC-12	15235321	AT4G28520	CRU3 (CRUCIFERIN 3)/nutrient reservoir	6.5	116	58.2	5	7	11
	19699273	AT4G28520	F20O9_210	6.6	116	58.2	5	7	11
	30688006	AT4G28520	CRU3 (CRUCIFERIN 3)/nutrient reservoir	6.7	116	50.0	5	7	13
SC-13	15219584	AT1G03890	Cupin family protein	5.5	100	49.6	3	3	12
SC-18	15235321	AT4G28520	CRU3 (CRUCIFERIN 3)/Nutrient reservoir	6.5	52	58.2	3	3	9
	19699273	AT4G28520	F20O9_210	6.6	52	58.2	3	3	9
	30688001	AT4G28520	CRU3 (CRUCIFERIN 3)/ nutrient reservoir	6.1	52	43.7	3	3	13
	30688006	AT1G03890	CRU3 (CRUCIFERIN 3); nutrient reservoir	6.7	52	50.0	3	3	11
	145334157	AT4G28520	CRU3 (CRUCIFERIN 3); nutrient reservoir	6.5	52	43.9	3	3	13
SC-19	15235321	AT4G28520	CRU3 (CRUCIFERIN 3)/Nutrient reservoir	6.5	165	58.2	6	7	19
	19699273	AT4G28520	F20O9_210 .	6.6	165	58.2	6	7	19
	30688001	AT4G28520	CRU3 (CRUCIFERIN 3)/nutrient reservoir	6.1	165	43.7	6	7	25
	145334157	AT4G28520	CRU3 (CRUCIFERIN 3)/nutrient reservoir	6.5	165	44.0	6	7	25
SC-20	15235321	AT4G28520	CRU3 (CRUCIFERIN 3)/Nutrient reservoir	6.5	105	58.2	5	5	15
	30688001	AT4G28520	CRU3 (CRUCIFERIN 3)/nutrient reservoir	6.1	105	43.7	5	5	21
	145334157	AT4G28520	CRU3 (CRUCIFERIN 3)/nutrient reservoir	6.5	105	43.9	5	5	20
SC-21	166676	AT5G44120	CRA1 (CRUCIFERINA)/nutrient reservoir	7.7	89	52.6	5	5	10
	15241422	AT5G44120	CRA1 (CRUCIFERINA)/ nutrient reservoir	7.7	89	52.6	4	5	10
	30694452	AT5G44120	CRA1 (CRUCIFERINA)/ nutrient reservoir	9.0	89	31.6	4	5	17
	30694455	AT5G44120	CRA1 (CRUCIFERINA)/ nutrient reservoir	6.6	89	41.0	4	5	13

SC-22	166678	AT1G03880	12S storage protein CRB	6.8	144	50.6	5	7	18
	15219582	AT1G03880	CRU2 (CRUCIFERIN 2); nutrient reservoir	6.5	144	50.5	5	7	18
	166676	AT5G44120	CRA1 (CRUCIFERINA)/nutrient reservoir	7.7	57	52.6	3	4	12
	15241422	AT5G44120	CRA1 (CRUCIFERINA)/ nutrient reservoir	7.7	57	52.6	3	4	12
	30694455	AT5G44120	CRA1 (CRUCIFERINA)/ nutrient reservoir	6.6	57	41.0	3	4	15
	30694452	AT5G44120	CRA1 (CRUCIFERINA)/ nutrient reservoir	9.0	55	31.6	3	4	20
SC-23	15235321	AT4G28520	CRU3 (CRUCIFERIN 3); nutrient reservoir	6.5	184	58.2	6	7	10
	30688001	AT4G28520	CRU3 (CRUCIFERIN 3); nutrient reservoir	6.1	184	43.7	6	7	25
	145334157	AT4G28520	CRU3 (CRUCIFERIN 3); nutrient reservoir	6.5	184	43.9	6	7	25
	166676	AT5G44120	12S storage protein CRA1	7.7	48	52.6	2	2	9
	15241422	AT5G44120	CRA1 (CRUCIFERINA)/ nutrient reservoir	7.7	48	52.6	2	2	9
	30694455	AT5G44120	CRA1 (CRUCIFERINA)/ nutrient reservoir	6.6	48	41.0	2	2	12
	30694452	AT5G44120	CRA1 (CRUCIFERINA)/ nutrient reservoir	9.0	46	31.6	2	2	15
SC-25	166678	AT1G03880	12S storage protein CRB	6.8	78	50.6	3	3	10
	15219582	AT1G03880	CRU2 (CRUCIFERIN 2); nutrient reservoir	6.5	78	50.5	3	3	10
	166676	AT5G44120	12S storage protein CRA1	7.7	76	52.6	3	3	9
	15241422	AT5G44120	CRA1 (CRUCIFERINA)/ nutrient reservoir	7.7	76	52.6	3	3	9
SC-27	15235321	AT4G28520	CRU3 (CRUCIFERIN 3)/ nutrient reservoir	6.5	78	58.2	3	3	8
	19699273	AT4G28520	CRU3	6.6	78	58.2	3	3	8
	30688001	AT4G28520	CRU3 (CRUCIFERIN 3); nutrient reservoir	6.1	78	43.7	3	3	11
	30688006	AT4G28520	CRU3 (CRUCIFERIN 3)/ nutrient reservoir	6.7	78	50.0	3	3	10
	145334157	AT4G28520	CRU3 (CRUCIFERIN 3); nutrient reservoir	6.5	78	43.9	3	3	11
SC-28	15235321	AT4G28520	CRU3 (CRUCIFERIN 3)/ nutrient reservoir	6.5	105	58.2	5	5	17

30688001	AT4G28520	CRU3 (CRUCIFERIN 3); nutrient reservoir	6.1	105	43.7	5	5	22
145334157	AT4G28520	CRU3 (CRUCIFERIN 3); nutrient reservoir	6.5	105	43.9	5	5	22

Table 5 -2. Identity of whole seed proteins of Arabidopsis germinated seed as determined by shotgun electrophoresis and LC-MS/MS. Other parameters are as in Table 5-1.

NCBI Acc#	AGI locus	Putative protein name	pI	Score	Mass	Pep	Mat	% Cov
Seed storage/nutrient reservoir proteins								
166678	AT1G03880	12S storage protein CRB	7.9	374	51.0	8	11	31
62321455	AT1G03880	Putative cruciferin 12S seed storage protein	7.9	315	19.9	7	11	75
15219582	AT1G03880	CRU2/nutrient reservoir	6.5	338	50.9	7	8	23
15226403	AT2G28490	Cupin family protein	5.8	118	55.9	2	2	5
15219584	AT1G03890	Cupin family protein	5.5	397	49.9	11	15	35
	AT3G04120	Cupin family protein	7.7	61	37.1	3	3	10
166676	AT5G44120	12S storage protein CRA1	7.7	313	52.9	8	15	26
1345841	AT5G44120	Cruciferin BnC2/ subunit alpha	8.7	86	54.5	2	2	6
15235321	AT4G28520	CRU3/nutrient reservoir	6.5	570	58.2	12	26	35
1076408	AT4G28520	Seed storage protein beta-chain 3 fragment	4.2	86	2.1	2	2	52
30688006	AT4G28520	CRU3 (CRUCIFERIN 3); nutrient reservoir	6.7	240	50.3	6	8	19
18403467	AT3G22640	PAP85/nutrient reservoir	6.6	278	55.0	7	9	19
15236995	AT4G27150	2S seed storage protein 2/2S albumin storage protein	6.8	57	19.9	2	2	15
15237012	AT4G27160	AT2S3/lipid binding /nutrient reservoir	8.0	55	19.3	2	2	15
15237014	AT4G27170	2S seed storage protein 4//NWMU2-2S albumin-4	7.4	65	19.7	2	2	14

1076411	??	Seed storage protein beta-chain 6 fragment	4.5	56	1.1	2	1	100
Other proteins								
15222848	AT1G13440	G-3-P dehydrogenase C2/NAD or NADH binding	6.6	177	37.0	5	7	28
166706	AT3G04120	Cytosolic G-3-P dehydrogenase	6.3	217	37.1	6	8	26
15235730	AT4G37870	Phosphoenol pyruvate carboxylase-1 /ATP binding/purine nucleotide binding	6.6	118	73.9	5	5	10
62321275	AT4G37870	Phosphoenolpyruvate carboxykinase-like protein	6.2	72	51.2	4	4	14*
7769871	AT1G53240	Malate dehydrogenase	8.5	123	37.2	5	5	24
15237551	AT5G03860	Malate synthase	8.0	77	64.3	2	2	4
2497857	AT3G15020	Malate dehydrogenase, mitochondrial	8.4	106	35.9	4	4	17
15219721	AT1G04410	Malate dehydrogenase, cytosolic (putative)	6.1	61	35.9	2	2	9
15219721	AT5G44120	Malate dehydrogenase, cytosolic	6.1	52	35.9	1	1	3
15232845	AT3G02090	Mitochondrial processing peptidase ß subunit	6.3	60	59.2	2	2	8
17939849	AT5G08670	Mitochondrial F1 ATP synthase ß subunit	6.5	77	63.6	3	3	10
1402914	AT1G71695	Peroxidase 12 precursor	8.6	76	39.9	2	2	9
5237615	AT5G64120	Peroxidase, putative	8.6	59	35.4	2	2	13
14517542	AT1G47128	Cysteine proteinase precursor-like protein/dehydration stress-responsive gene	5.3	82	52.2	3	3	7
18422289	AT5G43060	Cysteine proteinase/tiol protease (putative)	5.7	58	52.4	2	2	7
15242210	AT5G15090	VDAC3/ voltage-gated anion channel	7.9	112	29.2	2	2	8
15240765	AT5G67500	VDAC2/ voltage-gated anion channel	8.9	83	29.6	2	2	12
1354272	AT1G11910.	Aspartic proteinase	5.3	58	53.0	2	2	7

2160151	AT1G62290	Aspartic protease (gb-X77260)	8.0	92	47.4	2	2	7
15236863	AT4G18375	KH domain-containing protein	8.1	53	66.6	2	2	3
15225471	AT2G45410	LBD19 (Lob domain-containing protein-19)	7.7	49	21.3	2	2	8
12322163	AT1G54870	Dormancy related protein, putative	5.9	181	31.2	7	7	40
553043	AT3G21720	Isocitrate lyase	6.3	260	50.7	5	5	24
15237059	AT4G20360	ATRABE1B (Arabidopsis RAB GTPase HOMOLOG E1B)/ GTP binding/GTPase/ translation elongation factor	5.8	64	51.9	2	3	5
15227987	AT2G36530	LOS2/copper binding/ phosphopyruvate hydratase	5.54	118	48.0	5	5	28
15228198	AT3G16420	PBP1 (Pyk10-binding protein1)/ copper binding	5.5	108	32.1	4	5	23
3242075	AT4G13940	S-adenosyl-L-homocysteine hydrolase	5.9	94	52.0	4	4	17
15240793	AT5G07440	GDH2 (Glutamate dehydrogenase 2)/ATP binding/Glutamate dehydrogenase [NAD(P)+]/ Oxidoreductase	6.0	89	45.0	4	4	14
4587542	AT1G53940	Lipase/Acylhydrolase/GDSL-motif family/ ESTs gb-T45815, gb-T45130 and Gb-Z38046	6.1	132	48.2	5	5	20
3776021	AT1G72730	RNA helicase	5.2	58	50.2	2	2	6
18421009	AT5G26280	Meprin and TRAF homology domain-containing protein/MATH domain-containing protein	8.5	51	39.4	2	2	8
15225798	AT2G33150	Peroxisomal 3-ketocyl-COA thiolase-3 /acetyl-CoA C-acyltransferase 1	8.6	51	49.0	4	4	18
18423187	AT5G50700	HSD1 binding/catalytic/ oxido-reductase	5.9	75	39.5	2	2	13
15231715	AT3G52930	Fructose-bisphosphate aldolase (putative)	6.0	60	38.9	3	4	20

15233518	AT4G04460	Aspartyl protease family protein	6.8	105	56.3	3	4	9
5228198	AT3G16420	PBP1 (PYK10-binding protein 1)/copper binding	5.5	52	32.1	2	2	13
1592686	AT3G27660	Oleosin type4	6.9	56	20.3	2	2	18
Unknown								
9758672	AT5G45690	Unnamed protein product	5.9	69	28.8	1	1	5
18405086	AT1G54860	unknown protein	6.1	59	21.9	2	2	12
15220526	AT1G05510	unknown protein	6.2	79	27.4	3	3	14

In *Arabidopsis*, although the most prominent storage proteins in the mature seeds were shown to be the 12S-cruciferin subunits (Gallardo et al., 2001), their exact site of storage in the seed has not been clearly known. Studies suggested that the precursors of these storage proteins are synthesized at the endoplasmic reticulum, and the mature proteins accumulate inside protein storage vacuoles, PSVs (Muntz, 1998; Robinson and Hinz, 1999; Holkeri and Vitale, 2001; Jiang et al., 2001). Plants accumulate these storage proteins in the PSVs of their seeds for use during seed germination as a source of nitrogen (Otegui et al. 2006).

In *Arabidopsis*, the aleurone layer that plays a central role in seed germination and breaking of dormancy has been reported to respond to a number of signaling molecules including nitric oxide, and gibberellins which induce the formation of numerous PSVs in the aleurone cells (Bethke et al., 2007). Based on the above information it is suggested that the seed storage proteins identified in the seed coat of germinated seeds listed above were deposited into the PSVs of the aleurone cells. The above suggestion is in agreement with the report showing that the testa, which is the other cell layer of mature *Arabidopsis* seed coat (Figure 5-1), is composed on non-living cells (Muller et al., 2006). Thus, as far as my knowledge goes, this is the first experiment that separated the seed storage proteins from the PSVs of the seed coat aleurone layer of *Arabidopsis.*

Analysis of whole seed proteomic profile

Total proteins extracted from germinated whole seeds were subjected to 1D gel electrophoresis followed by LC-MS/MS (Table 5-2). This approach also failed to identify the *PELPK1* protein in the *Arabidopsis* whole seed proteome (refer to Table 5-2). However, a proteomic analysis of *Arabidopsis* root by multidimensional reverse-phase chromatography with on-line tandem mass spectrometry (MudPIT) has previously

detected the PELPK1 (Basu et al., 2006). It might be possible that being an extracellular soluble protein (Basu et al., 2006), the PELPK1 was lost during seed coat collection from germinated seeds by repeated rinsing. The other possibility is that PELPK1 constitutes a minor component in the seed/seed coat proteome, and may require more sensitive detection methodology such as MudPIT for identification.

Further analysis of the germinated whole seed proteomic profile (Table 5-2) identified 16 seed storage proteins involving 10 AGI loci including AT1G03880 (3 isoform), AT2G28490, AT1G03890, AT3G04120, AT5G44120 (2 isoforms), AT4G28520 (3 isoforms), AT3G22640, AT4G27150, AT4G27160, and AT4G27170. The additional seed storage proteins belonging to the AGI loci of AT2G28490, AT3G04120, AT3G22640, AT4G27150, AT4G27160, and AT4G27170, which were not identified in the PSVs of seed coat aleurone cells (Table 5-1) might have deposited to the cells of embryo containing PSVs (Otegui et al., 2006).

4. CONCLUSIONS

It appears that this is the first experiment that separately identified seed storage proteins and their corresponding AGI loci of the PSVs of seed coat aleurone cells. It is concluded that the lack of success in identifying the PELPK1 in the seed coat or in the whole seed proteome might either be due to its presence in insignificant quantity in these tissues, or due its loss during the process of seed coat/protein isolation. A more sensitive detection technology such as MudPIT can be tested to determine if it is possible to identify the PELPK1 in the above tissues of *Arabidopsis thaliana*.

5. REFERENCES

Basu, U, Francis, JL, Whittal, RM, Stephens, JL, Wang, Y, Zaiane, OS, Goebel, R, Muench, DG, Good, AG, and Taylor, GJ (2006). Extracellular Proteomes of Arabidopsis Thaliana and Brassica Napus Roots: Analysis and Comparison by MudPIT and LC-MS/MS. Plant and Soil, 286: 357-376

Belle, A., Tanay, A., Bitincka, A., Shamir, R., and O'Shea, K.K. (2006). Quantification of protein half-lives in the budding yeast proteome. PNAS 103: 13004–13009

Bergfeld, R. and Schopfer, P. (1986) Differentiation of a functional aleurone layer within the seed coat of Sinapis alba L. Ann Bot 57: 25–33.

Bethke, PC, Libourel, IGL, Aoyama, N, Chung, YY, Still, DW, and Jones, RL. (2007). The Arabidopsis aleurone layer responds to nitric oxide, gibberellin, and abscisic acid and is sufficient and necessary for seed dormancy. Plant Physiol. 143:1173-1188

Bewley JD (1997) Seed germination and plant dormancy. Plant Cell 9: 1055-1066[

Bradford M (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein using the principle of protein dye binding. Anal Biochem 72, 248-254

Cao, T., Srivastava, S., Rahman, MH, Kav. N., Hotte, N., Deyholos, MK., and Strelkov, SE (2008). Proteome-level changes in the roots of Brassica napus as a result of Plasmodiophora brassicae infection Plant Science 174: 97–115

Corner, E.J.H. (1976) The Seeds of Dicotyledons. da Silva, E.A.A., Toorop, P.E., Nijsse, J., Bewley, J.D. and (Coffea arabica cv. Rubi) seed germination and cause cell death in the embryo. J Exp Bot 413: 1029–1038.

Dhingraa, V, Gupta, M., Andacht, T. and Fu, ZF (2005). New frontiers in proteomics research: A perspective. International Journal of Pharmaceutics 299: 1–18

Gallardo, K, Job, C., Groot, SPC, Puype, M. Demol, H. Vandekerckhove, J, and Job, D. (2001), Proteomic Analysis of Arabidopsis Seed Germination and Priming. Plant Physiol, 126: 835-848

Harder A, Wildgruber R, Nawrocki A, Fey SJ, Larsen PM, Görg A (1999) Comparison of yeast cell protein solubilization procedures for two-dimensional electrophoresis. Electrophoresis 20, 826-829

Holkeri, H., and Vitale, A. (2001). Vacuolar sorting determinants within a plant storage protein trimer act cumulatively. Traffic 2: 737–741.

Jensen ON, Wilm M, Shevchenko A, Mann M. 1999. Sample preparation methods for mass spectrometric mapping directly from 2-DE gels. Methods in Molecular Biology 112, 513–5

Jiang, L., Phillips, T.E., Hamm, C.A., Drozdowicz, Y.M., Rea, P.A., Maeshima, M., Rogers, S.W., and Rogers, J.C. (2001). The protein storage vacuole: A unique compound organelle. J. Cell Biol. 155, 991–1002.

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

Liu, P.-P., Koizuka, N., Homrichhausen, T.M., Hewitt, J.R., Martin, R.C. and Nonogaki, H. (2005) Large-scale screening of Arabidopsis enhancer-trap lines for seed germination-associated genes. Plant J 41: 936–944.

Koornneef M, Bentsink L, Hilhorst H (2002) Seed dormancy and germination. Curr Opin Plant Biol 5: 33–36

Kucera, B, Cohen, MA, and Leubner-Metzger, G, (2005). Plant hormone interactions during seed dormancy release and germination. Seed Sci. Res. 15:281-307

Maton A, Hopkins J, McLaughlin CW, Johnson S, Warner MQ, LaHart D, Wright JD (1993). Human Biology and Health. Englewood Cliffs, Michigan, USA: Prentice Hall.

Müller, K., Tintelnot, S., and Leubner-Metzger, J. (2006). Endosperm-limited Brassicaceae Seed Germination: Abscisic Acid Inhibits Embryo-induced Endosperm Weakening of Lepidium sativum (cress) and Endosperm Rupture of Cress and Arabidopsis thaliana. Plant Cell Physiol. 47: 864–877.

Muntz, K. (1998). Deposition of storage proteins. Plant Mol. Biol. 38, 77–99

Nambara, E. and Marion-Poll, A. (2003) ABA action and interactions in seeds. Trends Plant Sci 8: 213–217.

Nguyen, H., Brown, R.C. and Lemmon, B.E. (2000) The specialized chalazal endosperm in Arabidopsis thaliana and Lepidium virginicum (Brassicaceae). Protoplasma 212: 99–110.

Otegui, MS, Herder, R, Schulze, J, Jung, R and Staehelind, LA (2006). The proteolytic processing of seed storage poroteins in Arabidopsis embryo cells starts in the multivesicular bodies. The Plant Cell: 18, 2567–2581

Pocsfalvia, G., Cuccurulloa, M., Schlossera, G., Cacacea, G., Sicilianoa, RA., Mazzeoa, MP., Scaccoc, S., Coccoc, T., Gnonic, A., Malornia, A. and Papac, S. (2006). Shotgun proteomics for the characterization of subunit composition of mitochondrial complex Biochimica et Biophysica Acta (BBA) 1757: 9-10

Pritchard, S.L., Charlton, W.L., Baker, A. and Graham, I.A. (2002) Germination and storage reserve mobilization are regulated independently in Arabidopsis. Plant J 31: 639–647.

Robinson, D.G., and Hinz, G. (1999). Golgi-mediated transport of seed storage proteins. Seed Sci. Res. 9, 267–283.

Rogers, S., Girolami, M., Kolch, W., Waters, K.M., Liu, T., Thrall, KM., and Wiley, HS (2008). Investigating the correspondence between transcriptomic and proteomic expression profiles using coupled cluster models. Bioinformatics 24 (24): 2894–2900.

Ruiz, Z.T. and Escale, M. (1995) Ultramicroscopic seed morphology of Cleome L. (Capparidaceae) in relation to their taxonomy and dispersal syndromes. Ernstia 5: 139–160.

Schopfer, P. and Plachy, C. (1984) Control of seed germination by abscisic acid. II. Effect on embryo water uptake in Brassica napus L. Plant Physiol 76: 155–160.

Schopfer, P., Plachy, C. and Frahry, G. (2001) Release of reactive oxygen intermediates (superoxide radicals, hydrogen peroxide, and hydroxyl radicals) and peroxidase in germinating radish seeds controlled by light, gibberellin, and abscisic acid. Plant Physiol 125: 1591–1602.

Steven, B (2003). The major world of microRNAs. http://www.nature.com/horizon/rna/background/micrornas.html.

Vaughan, J.G., Whitehouse, F.L.S. and Whitehouse, J.M. (1971) Seed structure and the taxonomy of the Cruciferae. Bot J Linn Soc 64: 383–409.

CHAPTER 6

GENERAL DISCUSSION AND FUTURE RESEARCH

CHAPTER 6

GENERAL DISCUSSION AND FUTURE RESEARCH

1. GENERAL DISCUSSION

The cell wall plays a vital role in plants. One of the main groups of cell wall structural proteins (CWPs) is the hydroxyproline-rich glycoproteins (HRGPs) which include three major gene families: extensins, proline-rich proteins (PRPs), and arabinogalactan proteins (AGPs). Because of my interest in CWPs, I have selected the current project with an objective to characterize a novel *Arabidopsis* gene, *AT5G09530*. I have named this gene as *PELPK1* because of its unique sequence motif, PELPK, which is repeated many times in the protein sequence of *AT5G09530* (Chapter 2). This gene has been variously annotated to encode either an extensin-like protein, or a PRP, a HRGP-family protein containing Pro-rich extensin-like domains, or a periaxin-like protein (reviewed in Chapter 2). Most recently it was named *PRP10* by a software program for classification of HRGPs (Showalter et al., 2010). However, PELPK1 does not contain the repeated Pro-Pro motif typical of PRPs. Based on my *in silico* analysis (Chapter 2), I have instead classified PELPK1 and all PELPK1-like proteins as a distinct type of HRGP (Chapter 2).

I found that PELPK1 has many similarities with other HRGPs, but lacks some of the features that define typical extensins and PRPs. Computational analysis of PELPK1 structure indicates that it is likely an intrinsically disordered protein (IDP) that lacks the regular structure found in most extensins. Specifically, PELPK1 (i) contains repetitive motifs like that of extensins and PRPs, but the structure of the repeated motif of PELPK1 (PELPK) is completely different from that of extensins (SPPPP), and PRPs (PP), (ii) PELPK1 is rich in Pro residues like that of extensins and PRPs, but the distribution of this residue in PELPK1 is completely different from that of extensins and PRPs; for

example, in the case of PELPK1, the Pro residues are distributed all over the protein including sequence motifs, whereas in extensins and PRPs, they are clustered in the sequence motifs only, and (iii) contains predicted O-linked glycosylation sites similar to that of extensins and PRPs, but the number of sites is much less in PELPK1 (7 sites) compared to extensins (18 sites), and PRPs (11 sites); refer to Chapter 2. Furthermore, in PELPK1, the PELPK motif is repeated 34 times, whereas it is totally absent in extensins and PRPs. Similarly, in extensins, the pentapeptide motif (SPPPP) is repeated more than 16 times, whereas, this motif is not at all conserved in PELPK1. Likewise, in PRPs, the PP repeat is repeated about 28 times, whereas it is completely lacking in PELPK1 (Chapter 2). Based on these observations I have also hypothesized that PELPK1 is a cell wall associated protein like that of other HRGPs but its function may be different from that of typical extensins and PRPs.

PELPK1 is predicted to be targeted to the secretory system (Chapter 2), and can thus be localized to the cell wall. This prediction is partially consistent with my observations that a translational fusion of PELPK1 to the N-terminus of a GFP reporter gene can be detected in the cell wall and also in the aleurone cells (Chapter 3). The above observations suggest that the *PELPK1* is not only functioning in the cell wall but also acting in the endosperm. Based on these observations, it can be hypothesized either (i) that the PELPK1 is undergoing hydrolysis during seed germination and providing nutrition to germinating embryo, or (ii) that like other HRGPs, it is forming oxidative cross-linkages with the cell wall mediated via reactive oxygen species (ROS) that may be generated during radicle penetration through the seed coat perhaps to protect germinating embryo from pathogenic infection, or (iii) that it is acting as a peptide signaling molecule to activate other genes linked to seed germination. The second possibility appears to be less likely as PELPK1 cross-linking with the cell wall may hinder

cell wall expansion and consequently can slow-down the growth of embryo (Showalter 1993, Cassab 1998), which appears to be not the case.

Thus, the above predicted functions of the *PELPK1* in seed germination appear to be different from that of other CWPs including extensins. Since both the promoter::GUS fusion as well as CDS::GFP transitional fusion lines showed that the *PELPK1* is normally expressed in the seed coat during germination, in addition to cell wall of other tissues (Chapter 3), it is assumed that the *PELPK1* expression in the seed coat aleurone layer is not an artifact. By analyzing my results, I hypothesize that the *PELPK1* is somehow linked either to the growth of embryo presumably through cell wall extension or to embryo defense mediated via oxidative cross-linking with the cell wall, or acting as a signaling molecule to trigger other genes associated with seed germination in addition to general cell wall-based defense and repair related roles (Chapter 3). As mentioned above, the second possibility is less likely as it can negatively impact embryo growth.

PELPK1 transcripts are reportedly expressed in a range of tissues both during normal development and in response to external stimuli, especially biotic stresses and hormones (Chapter 2). Microarray-based expression analyses by other researchers showed that *PELPK1* is expressed most abundantly in root, seed, hypocotyl and radicle, and during flowering and silique stages; and in response especially to biotic stresses such as disease infection particularly by *Pseudomonas syringae*. These microarray data are largely consistent with my experiments conducted using transgenic plants harboring a fusion of the *PELPK1* promoter with the GUS reporter gene (Chapter 3). I detected the reporter in the seed-coat of germinated seeds, enriched specifically in the endosperm/ aleurone layer at the micropylar end of the seed (through which the radicle penetrates). This observation is consistent with the conservation of a number of endosperm-specific cis-regulatory elements in the upstream intergenic sequence

included in this GUS construct (Chapter 3). However, I did not detect any GUS expression in developing seed or in mature seed prior to germination. This is inconsistent with a recently reported microarray expression profile of *PELPK1* which showed highest *PELPK1* expression in the seed coat of maturing green seed including endosperm/aleurone layer, with more intense expression in the cell layers outside of aleurone layer (Chapter 2). This suggests that PELPK1 function is more likely linked to seed germination and perhaps growth of the embryo.

As has been reported for many CWPs and for *PELPK1* in particular, expression of transcripts can be induced by biotic factors and defence hormones. The *PELPK1* upstream genomic region contains conserved cis-elements for pathogen-inducibility, and indeed I observed strong induction of reporter gene expression when plants were exposed to *Pseudomonas syringae* (Chapter 3). However, although mechanical wounding also produced GUS reporter expression in transgenic plants in the present study, no wound-inducible cis-elements were detected in the upstream DNA sequence (Chapter 3). Furthermore, microarray analysis by Seki et al., (2002) showed that *PELPK1* was induced by abiotic factors, such as osmotic and salt stresses and ABA (Chapter 2), whereas I did not detect *PELPK1* promoter::reporter induction by abiotic stress factors (Chapter 3). The absence of reporter gene induction by osmotic and salt stresses in the present study, and the lack of conserved, wounding-associated cis-elements, suggests that some necessary cis-elements may be located outside of the upstream intergenic region of *PELPK1* (Tylor 1997). It might be possible that a complete set of regulatory elements was not included in the promoter fragment used in the promoter::GUS fusion construct. Nevertheless, results of both microarray (Chapter 2) and reporter gene experiments (Chapter 3) clearly show that *PELPK1* is, at least at the transcript level, induced by pathogens and wounding.

Thus, it can be concluded that the *PELPK1* plays a role in the cell wall (and possibly in the seed aleurone) during normal seed germination, and in tissues responding to pathogen attack. During both germination and pathogen responses, the cell wall is presumably undergoing dynamic changes. PELPK1 may stabilize the cell wall during remodeling to allow loosening during germination (or stiffening to promote rupture), and may likewise help reinforce the cell wall in response to pathogen attack. The expression of PELPK1::GFP in the aleurone is less clear; PELPK1 may help supply nutrients that indirectly help cell wall extension, or act as a peptide signaling to activate other genes or the observed expression is also indicative of ongoing localization to the cell wall (Chapter 3).

Further studies were carried out using plants with perturbed levels of *PELPK1* expression (Chapter 4). Analysis of sequence-indexed insertion mutants of *PELPK1* did not show any phenotype or change in the transcript level as compared to the parental control, suggesting either that the *PELPK2* perhaps acted redundantly (Allen et al., 2007) or that no null mutation was induced by the T-DNA in the mutagenized lines tested (Krysan et al., 1999), or that TAIL-PCR used to index T-DNA was not precise. Analysis of RNAi mutant plants in which both *PELPK1* and *PELPK2* genes were silenced showed that their transcript level as determined by qRT-PCR was significantly down-regulated, and their response to sucrose as determined by seed germination and root elongation *in-vitro* was modified with increased susceptibility to a slightly elevated level as compared to WT plants. In contrast, the transcript level in OX plants as determined above was dramatically up-regulated and their germination and root elongation responded positively to levels of sucrose at which the WT plants responded negatively (Chapter 4). Based on these observations, I suggest that the function of *PELPK1* in *Arabidopsis* plants can also be linked to sucrose metabolism, which may be directly or indirectly linked to its inferred activity in the cell wall. From these observations, I further

hypothesize that the *PELPK1* is mediating the expression of genes involved in sucrose metabolic and/or a signaling pathway, perhaps by generating some kind of signal such as peptide signal (R¨ohrig et al., 2002; 2004; Matsubayashi, 2003; Matsubayashi and Sakagami, 2006).

A study suggested that a tomato leaf HRGP family protein acted as a precursor for peptide hormone signals against wounding, and defense chemicals (Narvaez-Vasquez et al., 2005). Peptide hormones have roles in cell to cell communication and defence (Matsubayashi and Sakagami, 2006; Huffaker et al., 2006; Ryan and Pearce 2004). Although the structure of the tomato peptide hormone HRGP is different from PELPK1, the possibility remains that PELPK1 is acting as a signaling molecule.

Other reports suggested that CWPs including extensins perform structural as well as defence-related roles in plants (Showalter 1993; Cassab 1998). In particular with extensins, these reports suggested that extensins produce oxidative cross-linkages with the cell wall and become insolubilized, and make the cell wall harder to pathogen penetration (Showalter 1993; Cassab 1998; Wei and Shirsat 2006).

The present research shows that PELPK1 is cell wall associated and is induced by biotic factors, defense chemicals, and wounding comparable to that of extensins and perhaps to other HPGPs including PRPs. However, it appears that the mechanism of its function in the cell wall is different from that of extensins as over-expression of *PELPK1* did not make the plants less susceptible or RNAi did not make the plants more susceptible to disease infection as compared to WT plants (Chapter 4).

Furthermore, the constitutive over-expression of *PELPK1* makes the plants more responsive, whereas under-expression by RNAi makes them less responsive to exogenously applied sucrose with regard to growth as compared to WT plants (Chapter 4). RNAi plants grown in MS medium with or without 50 mM sucrose did not show significant differences in seed germination or root elongation from the WT plants.

However, when they were grown in 100 mM sucrose, a concentration that normally does not have any significant adverse effect on WT plants, the seed germination was significantly delayed and root elongation was significantly slowed down as compared to that of the WT plants (Chapter 4). These observations indicate that the RNAi plants have become sensitive to a concentration of sucrose which had no adverse effects on WT plants. This modification in the response of RNAi lines to sucrose appears to be not due to increased susceptibility of these plants to osmotic stress as the responses of these mutants to different concentrations of mannitol, sorbitol, glucose, salt (NaCl) or growth hormones (IAA, GA, ABA) were not significantly different from that of the WT plants (Chapter 4).

It appears that in RNAi plants, either sucrose metabolism or signaling or transport is somehow modified. As a result, at a slightly elevated level of sucrose, these plants responded negatively to seed germination and root elongation processes as compared to WT plants. In a recent review (Hammond and White 2008), it has been reported from a number of studies that increased sucrose uptake or accumulation in plants can lead to inhibition of nutrient uptake particularly phosphorus causing nutrient starvation in plants. This review further added that low nutrient, particularly the phosphorus, availability in plants is preceded by increased carbohydrate accumulation in the shoot and phloem tissues; inhibition of sucrose biosynthesis and/or translocation in plants reversed nutrient starvation; nutrient deficient plants accumulate increased sugars and starch in their leaves; the photosynthate transport in plants takes in the form of sucrose; and in *Arabidopsis* phosphorus starvation results in the inhibition of primary root growth, increase in root-shoot ratio, and accumulation of anthocyanin pigments in plants (Hammond and White 2008; Mura et al. 2005; Sanchez-Calderon et al., 2005; Ticconi et al. 2004). Based on the above information, it can be assumed that the increased sensitivity of the RNAi plants, as indicated particularly by the reduced root

elongation at a slightly elevated level of sucrose (Chaper 4), can perhaps be related to the interference of nutrient uptake in these plants from the MS-agar medium by sucrose.

Furthermore, analysis of the growth characteristics of RNAi plants in the soil also showed that they emerged from the soil slowly, produced relatively smaller plants, and their leaves can be slightly darker, vegetative growth was slower, and flowering was delayed compared to that of WT plants (Chapter 4). The above abnormal growth symptoms in RNAi plants in the soil can be interpreted as follows. It might be possible that in RNAi plants the sucrose metabolic pathway is repressed, and as a consequence (i) the hydrolysis of stored carbohydrates in the seed endosperm/aleurone cells was also slowed down during seed germination, causing slow cell wall extension and slow emergence of the seedlings from the soil, and (ii) the break-down of photosynthates in the above ground shoots was also slowed down resulting in higher accumulation of carbohydrates inside the plants thus giving slightly dark-green coloration of leaves and also interfering with the nutrient uptake from the soil causing slow cell wall extension and slow vegetative growth (Anonymous 1999). From the above discussion it can be hypothesized that RNAi mutagenesis of *PELPK1* not only modified sucrose metabolic or signaling pathway but also negatively impacted nutrient uptake, cell wall extension, and consequently vegetative growth of *Arabidopsis* plants.

The above hypothesis was further tested by analyzing the growth characteristics of *PELPK1* OX-lines. Transcript expression analysis by qRT-PCR showed dramatic upregulation of *PELPK1* in OX plants as compared to WT. Phenotypic analysis of these plants grown either in MS-agar medium or in the soil also showed detectable morphological differences as compared to WT plants. The OX- plants grown in MS-agar medium germinated faster and produced much longer root in the presence of 100 mM sucrose compared to WT plants (Chapter 4). Even in the presence of 200 mM sucrose, a concentration which generally causes osmotic stress in WT *Arabidopsis (*Neales,

1968), the root growth of OX-plants was significantly higher than that of WT plants (Chapter 4). However, their responses to other abiotic stresses such as high concentrations of mannitol, NaCl, or ABA were similar to WT plants (Chapter 4), suggesting that the OX-plants might have effectively metabolized sucrose to more useable forms to enhance cell wall extension and plant growth. Based on the above observations, I hypothesize that *PELPK1* OX up-regulated the complex sucrose metabolic and/or signaling pathway in OX-plants causing faster break-down of sucrose and consequently induced faster seed germination and enhanced cell wall extension and root growth compared to WT plants (Hammond and White, 2008).

The above hypothesis was further tested under soil conditions and found to be consistent with *in vitro* observations. The emergence of OX-seedlings from the soil was faster, their size was larger, color was slightly lighter-green, early vegetative growth was significantly enhanced, and flowering was earlier compared to that of the WT plants (Chapter 4). These observations suggest that the OX-plants might have effectively metabolized not only the seed storage carbohydrates that causes enhanced embryo growth and early soil emergence of seedlings but also the products of photosynthesis (i.e. carbohydrates) in the above ground shoot, causing enhanced vegetative growth mediated by enhanced cell wall extension and as well as enhanced nutrient uptake as a secondary effect.

Regarding the physiological role of sucrose in plants, it has been suggested that sucrose degradation is necessary for nutrient uptake (Gordon et al., 1999) and elevated level of sucrose either endogenous or added to the growth media can interfere with nutrient uptake causing reduced cell wall elongation and vegetative growth (Hammond and White 2008; Vance et al., 2003; Anonymous 1999). Sucrose was shown to bind to the 93 kDa subunit of sucrose synthase, an essential component in sucrose metabolic pathway (Rohrig et al., 2002; 2004).

Thus, based on mutational analysis (Chapter 4), it is suggested that *PELPK1* is associated with cell wall extension perhaps mediated via sucrose metabolism and/or signaling. On the other hand, based on expression analysis (Chapter 3), it is suggested that the *PELPK1* is linked to cell wall based defense-, and repair-related functions. From the above discussion, it appears that in the absence of pathogenic infection, the *PELPK1* might be acting on cell wall loosening, and in the presence of pathogenic infection, it might be acting on the hardening of the cell wall.

However, it is not clear how *PELPK1* is performing these two opposite functions under two different circumstances. Analyses of bioinformatics (Chapter 2), as well as spatial expression pattern and sub-cellular protein localization (Chapter 3) suggest that the *PELPK1* encodes a cell wall- associated non-enzymatic intrinsically disordered protein. Other cell wall non-enzymatic structural proteins such as extensins and PRPs have been reported to cause cell wall hardening in response to pathogen attack by forming oxidative cross-linkages followed by insolubilization (Showalter 1993; Cassab 1998). However, in the present study analysis of the responses of RNAi and OX plants to *Pseudomonas syringae* infection did not show any significant differences as compared to WT plants (Chapter 4) suggesting that the mechanism of PELPK1-mediated defense-responses in *Arabidopsis i*s not similar to that of extensins or PRPs.

To establish connections between these two apparently unrelated functions of *PELPK1*, I propose that PELPK1 might be acting as a signaling molecule to perform multiple functions in *Arabidopsis* plants. Peptide signaling molecules have been identified in plants with important roles including cell-to-cell communication and defense. Majority of them have been reported to process in the endoplasmic reticulum, including the removal of the N-terminal signal sequence and sometimes glycosylation and finally secreted to the cell wall matrix (Matsubayashi and Sakagami, 2006). Plant peptide signals, which are also comparable to that of animals and yeasts, have been reported to

derive from larger precursor proteins, and are shown to be receptor-mediated (Ryan and Pearce 2004). For example, systemin a small peptide is functioning as a long-distance signal to activate chemical defenses against herbivores. It was the first plant hormone proven to be a peptide. Systemin identified in tomato leaves induces the production of protein defense compound called protease inhibitors. It was found to be an 18-amino acid peptide processed from the C-terminus of a 200-amino acid precursor, which is called prosystemin (McGurl et al., 1992). Furthermore, it has been reported that the tomato leaf polyprotein, the precursor of three hydroxyproline-rich glycopeptide defense signals, referred to as LeHypSys I, II, and III, was synthesized in phloem parenchyma cells in response to wounding, systemin, and methyl jasmonate, and the precursor protein was sequestered in the cell wall matrix (Narvaez-Vasquez et al., 2005). Based on the above findings, the authors suggested that the plant cell wall can play an active role in defenses by using its associated proteins as a precursor for peptide signaling to activate defense-related genes. Another report stated that ENOD40, an early nodulin gene putatively encoded two small peptide signals of 12 and 18 amino acid residues. These peptides have been reported to bind to the 93 kDa subunit of sucrose synthase, an essential component in sucrose metabolism (Rohrig et al., 2004). Sucrose degradation has been shown to be a key step in nitrogen fixation, and was a pre-requisite for normal nodule development (Gordon et al., 1999).

In conclusion, PELPK1 is a unique cell wall associated intrinsically disordered hydrophilic protein that is involved in multiple functions including cell wall extension and general defense presumably by producing peptide signals in *Arabidopsis* plants.

2. FUTURE RESEARCH

My computational analysis revealed that the *PELPK1* encodes an intrinsically disordered protein, IDP (Chapter 2). The functions of IDPs have been reported to reflect their structural characteristics such as high pliability and adaptability, and extended conformational state (Tompa et al., 2006). These characteristics appear to give IDPs the ability to contribute plants during various stress conditions. For instance, some IDPs have been reported to act as chaperones against abiotic stresses by preventing protein aggregation and enzyme inactivation (Kovacs et al., 2008). My promoter activity analysis showed that the *PELPK1* is induced by biotic factors, defense chemicals and wounding (Chapter 3). However, the exact mechanism by which the *PELPK1* performs these defense- and repair-related functions is unclear, although subcellular protein localization studies showed that PELPK1 is normally secreted to the cell wall. Thus, further research is necessary to delineate the IDP characteristics of PELPK1.

In order to accomplish the above objective, the following experiments will be performed: (i) introduce order-forming amino acid residues (Cys and Trp) by replacing highly disorder-forming residues (Pro and Lys) in selected motifs using the technique of site-directed mutagenesis, (ii) express the mutagenized protein in a heterologous system and non-mutagenized protein as a control, and (iii) analyze the protein structure by using techniques such as NMR spectroscopy.

Site-directed mutagenesis in *PELPK1*

Analysis of the protein disorder in PELPK1 and all PELPK1-like proteins in different species (Chapter 2) shows that they all are typically rich in Pro and Lys residues. Although Pro is a hydrophobic residue, normally preferred for protein folding, its cyclic side-chain has been reported to hinder protein folding and thus contributes to intrinsic disorder in proteins. On the other hand, Lys is a positively charged, highly

hydrophilic (pI ~10), and typically disorder-forming residue (Tompa 2002; Dunker et al., 2001). Thus, in order to introduce site-directed mutagenesis in the *PELPK1* gene, I would like to replace Pro and Lys residues in selected motifs by order-forming residues such as Cys, Tyr, or Trp. The experiments will be performed using QuikChange Site-Directed Mutagenesis Kit (Stratagene). The following is an example of the experimental procedure that I will follow for site-directed mutagenesis in *PELPK1* gene.

For introducing mismatches in the selected protein motif (refer to the underlined area in the deduced protein sequence of *PELPK1* shown below), the 43rd amino acid residue, Pro (bases 127th -128th -129th = **CCG)** will be changed to Cys (TGC), and the 47th amino acid, **Lys** (bases 139th -140th -141st = AAA) will be changed to Trp (TGG) as shown below.

PELPK1 (At5g09530): deduced protein sequence (TAIR)

Motif selected to mutagenize is underlined and the residues to be mutagenized are high-lighted

```
  1 MALMKKSLSA ALLSSPLLII CLIALLADPF SVGARRLLED PKPEIPKLPE
 51 LPKFEVPKLP EFPKPELPKL PEFPKPELPK IPEIPKPELP KVPEIPKPEE
101 TKLPDIPKLE LPKFPEIPKP ELPKMPEIPK PELPKVPEIQ KPELPKMPEI
151 PKPELPKFPE IPKPDLPKFP ENSKPEVPKL METEKPEAPK VPEIPKPELP
201 KLPEVPKLEA PKVPEIQKPE LPKMPELPKM PEIQKPELPK LPEVPKLEAP
251 KVPEIQKPEL PKMPELPKMP EIQKPELPKM PEIQKPELPK VPEVPKPELP
301 TVPEVPKSEA PKFPEIPKPE LPKIPEVPKP ELPKVPEITK PAVPEIPKPE
351 LPTMPQLPKL PEFPKVPGTP
```

***PELPK1*: coding sequence**

Complementary strand showing the codons to be substituted in the selected motif (Pro = CCG, Lys = AAA)

```
ATGGCACTAATGAAGAAGAGTCTCTCTGCTGCTCTTCTCTCATCACCACTTCTGATCATATGTCTTATCGC
ATTGCTCGCTGATCCGTTTTCAGTCGGTGCTCGCCGGTTATTGGAGGATCCTAAACCGGAGATACCAAAA
TTGCCTGAGCTACCTAAATTCGAAGTTCCCAAGTTGCCGGAGTTCCCTAAACCAGAGTTGCCCAAGTTACC
CGAATTTCCAAAGCCTGAGTTGCCAAAGATCCCGGAGATTCCAAAGCCAGAGTTACCAAAGGTACCGGAGA
TTCCAAAGCCTGAGGAAACTAAACTGCCAGATATTCCCAAGCTTGAATTGCCCAAGTTTCCGGAAATTCCA
AAACCTGAGCTCCCAAAGATGCCAGAGATTCCAAAACCTGAGTTACCAAAGGTACCGGAGATTCAGAAGCC
CGAGTTACCAAAAATGCCGGAGATTCCAAAGCCTGAATTACCAAAGTTTCCAGAGATTCCAAAGCCTGATT
TGCCAAAGTTTCCAGAGAATTCAAAGCCTGAGGTGCCTAAGCTAATGGAGACTGAAAAGCCTGAGGCTCCT
AAGGTGCCAGAGATTCCAAAGCCTGAGTTGCCAAAGTTGCCAGAAGTTCCCAAGCTTGAGGCTCCTAAGGT
ACCAGAGATCCAGAAGCCGGAGTTGCCCAAAATGCCGGAGTTACCTAAGATGCCGGAGATTCAGAAACCTG
AGTTGCCAAAGTTGCCAGAAGTTCCCAAGCTTGAGGCTCCTAAGGTACCGGAGATCCAGAAGCCGGAGTTG
```

CCCAAAATGCCGGAGTTACCTAAGATGCCGGAGATTCAGAAACCCGAGTTGCCCAAGATGCCGGAGATTCA
GAAGCCTGAGTTGCCGAAGGTGCCAGAGGTTCCAAAGCCCGAATTGCCAACGGTTCCAGAGGTTCCAAAGT
CTGAGGCTCCTAAGTTTCCAGAGATTCCAAAGCCTGAACTGCCGAAGATTCCAGAAGTTCCAAAACCTGAA
CTGCCCAAGGTTCCAGAAATTACAAAACCTGCAGTTCCAGAGATTCCAAAGCCAGAGCTACCGACGATGCC
TCAACTTCCCAAGTTGCCGGAATTCCCAAAAGTTCCCGGAACTCCTTAA

Forward strand showing the codons to be substituted (Pro = CGG, Lys = TTT)

TTAAGGAGTTCCGGGAACTTTTGGGAATTCCGGCAACTTGGGAAGTTGAGGCATCGTCGGTAG
CTCTGGCTTTGGAATCTCTGGAACTGCAGGTTTTGTAATTTCTGGAACCTTGGGCAGTTCAGGT
TTTGGAACTTCTGGAATCTTCGGCAGTTCAGGCTTTGGAATCTCTGGAAACTTAGGAGCCTCA
GACTTTGGAACCTCTGGAACCGTTGGCAATTCGGGCTTTGGAACCTCTGGCACCTTCGGCAAC
TCAGGCTTCTGAATCTCCGGCATCTTGGGCAACTCGGGTTTCTGAATCTCCGGCATCTTAGGTA
ACTCCGGCATTTTGGGCAACTCCGGCTTCTGGATCTCCGGTACCTTAGGAGCCTCAAGCTTGG
GAACTTCTGGCAACTTTGGCAACTCAGGTTTCTGAATCTCCGGCATCTTAGGTAACTCCGGCAT
TTTGGGCAACTCCGGCTTCTGGATCTCTGGTACCTTAGGAGCCTCAAGCTTGGGAACTTCTGGC
AACTTTGGCAACTCAGGCTTTGGAATCTCTGGCACCTTAGGAGCCTCAGGCTTTTCAGTCTCCA
TTAGCTTAGGCACCTCAGGCTTTGAATTCTCTGGAAACTTTGGCAAATCAGGCTTTGGAATCTC
TGGAAACTTTGGTAATTCAGGCTTTGGAATCTCCGGCATTTTTGGTAACTCGGGCTTCTGAATC
TCCGGTACCTTTGGTAACTCAGGTTTTGGAATCTCTGGCATCTTTGGGAGCTCAGGTTTTGGAA
TTTCCGGAAACTTGGGCAATTCAAGCTTGGGAATATCTGGCAGTTTAGTTTCCTCAGGCTTTGG
AATCTCCGGTACCTTTGGTAACTCTGGCTTTGGAATCTCCGGGATCTTTGGCAACTCAGGCTTT
GGAAATTCGGGTAACTTGGGCAACTCTGGTTTAGGGAACTCCGGCAACTTGGGAACTTCGAAT
TTAGGTAGCTCA*GGCAA**TTT**TGGTATCTCCGGTTTAG*GATCCTCCAATAACCGGCGAGCACCG
ACTGAAAACGGATCAGCGAGCAATGCGATAAGACATATGATCAGAAGTGGTGATGAGAGAA
GAGCAGCAGAGAGACTCTTCTTCATTAGTGCCAT

Designing primers for site-directed mutagenesis:

(CTAAACCGGAGATACCAAAATTGC => CTAAATGCGAGATACCATGGTTGC)

Regular left primer (primer-A): 5'-ATGGCACTAATGAAGAAGAGTCTC-3'
Regular right primer (primer- B): 5'-GCAATTTTGGTATCTCCGGTTTAG-3'
Mutagenized right primer (primer-b): 5-'GCAACCATGGTATCTCGCATTTAG-3
Product size: 145 bp

Regular left primer (Primer-C): 5'-CTAAA**CCG**GAGATACCA**AAA**TTGC-3'
Mutagenized left primer (primer c): 5'-CTAAA**TGC**GAGATACCA**TGG**TTGC-3'
Regular right primer (primer-D): 5'-TTAAGGAGTTCCGGGAACTTTTGG-3'
Product size: 992
Blue = Lys, magenta = Pro

Annealing verification (Primer3 Version 4)

```
  1 ATGGCACTAATGAAGAAGAGTCTCTCTGCTGCTCTTCTCTCATCACCACTTCTGATCATA
    >>>>>>>>>>>>>>>>>>>>>>>> primer-A

 61 TGTCTTATCGCATTGCTCGCTGATCCGTTTTCAGTCGGTGCTCGCCGGTTATTGGAGGAT

121 CCTAAACCGGAGATACCAAAATTGCCTGAGCTACCTAAATTCGAAGTTCCCAAGTTGCCG
     <<<<<<<<<<<<<<<<<<<<<<<< primer-B

181 GAGTTCCCTAAACCAGAGTTGCCCAAGTTACCCGAATTTCCAAAGCCTGAGTTGCCAAAG
```

```
 241 ATCCCGGAGATTCCAAAGCCAGAGTTACCAAAGGTACCGGAGATTCCAAAGCCTGAGGAA

 301 ACTAAACTGCCAGATATTCCCAAGCTTGAATTGCCCAAGTTTCCGGAAATTCCAAAACCT

 361 GAGCTCCCAAAGATGCCAGAGATTCCAAAACCTGAGTTACCAAAGGTACCGGAGATTCAG

 421 AAGCCCGAGTTACCAAAAATGCCGGAGATTCCAAAGCCTGAATTACCAAAGTTTCCAGAG

 481 ATTCCAAAGCCTGATTTGCCAAAGTTTCCAGAGAATTCAAAGCCTGAGGTGCCTAAGCTA

 541 ATGGAGACTGAAAAGCCTGAGGCTCCTAAGGTGCCAGAGATTCCAAAGCCTGAGTTGCCA

 601 AAGTTGCCAGAAGTTCCCAAGCTTGAGGCTCCTAAGGTACCAGAGATCCAGAAGCCGGAG

 661 TTGCCCAAAATGCCGGAGTTACCTAAGATGCCGGAGATTCAGAAACCTGAGTTGCCAAAG

 721 TTGCCAGAAGTTCCCAAGCTTGAGGCTCCTAAGGTACCGGAGATCCAGAAGCCGGAGTTG

 781 CCCAAAATGCCGGAGTTACCTAAGATGCCGGAGATTCAGAAACCCGAGTTGCCCAAGATG

 841 CCGGAGATTCAGAAGCCTGAGTTGCCGAAGGTGCCAGAGGTTCCAAAGCCCGAATTGCCA

 901 ACGGTTCCAGAGGTTCCAAAGTCTGAGGCTCCTAAGTTTCCAGAGATTCCAAAGCCTGAA

 961 CTGCCGAAGATTCCAGAAGTTCCAAAACCTGAACTGCCCAAGGTTCCAGAAATTACAAAA

1021 CCTGCAGTTCCAGAGATTCCAAAGCCAGAGCTACCGACGATGCCTCAACTTCCCAAGTTG

1081 CCGGAATTCCCAAAAGTTCCCGGAACTCCTTAA


   1 ATGGCACTAATGAAGAAGAGTCTCTCTGCTGCTCTTCTCTCATCACCACTTCTGATCATA

  61 TGTCTTATCGCATTGCTCGCTGATCCGTTTTCAGTCGGTGCTCGCCGGTTATTGGAGGAT

 121 CCTAAACCGGAGATACCAAAATTGCCTGAGCTACCTAAATTCGAAGTTCCCAAGTTGCCG
       >>>>>>>>>>>>>>>>>>>>>>>> primer-C

 181 GAGTTCCCTAAACCAGAGTTGCCCAAGTTACCCGAATTTCCAAAGCCTGAGTTGCCAAAG

 241 ATCCCGGAGATTCCAAAGCCAGAGTTACCAAAGGTACCGGAGATTCCAAAGCCTGAGGAA

 301 ACTAAACTGCCAGATATTCCCAAGCTTGAATTGCCCAAGTTTCCGGAAATTCCAAAACCT

 361 GAGCTCCCAAAGATGCCAGAGATTCCAAAACCTGAGTTACCAAAGGTACCGGAGATTCAG

 421 AAGCCCGAGTTACCAAAAATGCCGGAGATTCCAAAGCCTGAATTACCAAAGTTTCCAGAG

 481 ATTCCAAAGCCTGATTTGCCAAAGTTTCCAGAGAATTCAAAGCCTGAGGTGCCTAAGCTA

 541 ATGGAGACTGAAAAGCCTGAGGCTCCTAAGGTGCCAGAGATTCCAAAGCCTGAGTTGCCA

 601 AAGTTGCCAGAAGTTCCCAAGCTTGAGGCTCCTAAGGTACCAGAGATCCAGAAGCCGGAG

 661 TTGCCCAAAATGCCGGAGTTACCTAAGATGCCGGAGATTCAGAAACCTGAGTTGCCAAAG
```

```
 721 TTGCCAGAAGTTCCCAAGCTTGAGGCTCCTAAGGTACCGGAGATCCAGAAGCCGGAGTTG

 781 CCCAAAATGCCGGAGTTACCTAAGATGCCGGAGATTCAGAAACCCGAGTTGCCCAAGATG

 841 CCGGAGATTCAGAAGCCTGAGTTGCCGAAGGTGCCAGAGGTTCCAAAGCCCGAATTGCCA

 901 ACGGTTCCAGAGGTTCCAAAGTCTGAGGCTCCTAAGTTTCCAGAGATTCCAAAGCCTGAA

 961 CTGCCGAAGATTCCAGAAGTTCCAAAACCTGAACTGCCCAAGGTTCCAGAAATTACAAAA

1021 CCTGCAGTTCCAGAGATTCCAAAGCCAGAGCTACCGACGATGCCTCAACTTCCCAAGTTG

1081 CCGGAATTCCCAAAAGTTCCCGGAACTCCTTAA
             <<<<<<<<<<<<<<<<<<<<<<<<< primer-D
```

Heterologous expression of *PELPK1*

Preparation of mutagenized PCR products

Arabidopsis genomic DNA will be extracted as described earlier (Chapter 3). PCR will be conducted using genomic DNA and the primers designed and listed above. The primer pair, A + B, will provide a non-mutagenized (native) PCR product of size 145 bp, whereas, the primer pair, A + b, will provide a similar-sized PCR product (145 bp) but with mutations in one of the two strands (complementary strand). Similarly, the primer pair, C + D will provide a non-mutagenized (native) PCR product of size, 992 bp, however, the primer pair, c + D) will provide a similar-sized PCR product (992 bp) but with mutations in the same loci in forward strand. Following gel purification of the native and mutagenized products as described earlier (chapter 3), the two native products and the two mutagenized products will be mixed together separately with equal proportion. They will then be further used separately to generate native and mutagenized PCR products of the PELPK1 gene as described above using CDS-specific forward (primer A) and reverse (primer D) primers. The two products will be purified as described above and used for cloning in expression vectors as described below.

Construction of expression vector

For the preparation of expression plasmid, pET28a vector (Novogen) will be used. The above *PELPK1* coding sequence will be amplified by using primers which will introduce *NcoI* restriction site at 5' end and a polyhistidine tag (His6) attached to an *EcoRI* restriction site at the 3' end. The resulting fragment will be cloned into the above expression vector, and the construct will be used for transformation of *E. coli* cells (DH5α). Following isolation of the above plasmid from the DH5α *E. coli* cells, the plasmid will be transformed into *E. coli* strain BL21(DE3) and used for expression analysis.

Time course experiment will be performed by growing the above *E. coli* strain in liquid LB medium containing 50 mg/ml Kanamycin and inducing the expression system by adding 1mM IPTG. Samples will be collected at different time intervals up to 24 hr post-induction. Cells will be collected by centrifugation and used for protein extraction using PBS buffer. The resulting proteins will be used to run 15% SDS page.

The native and the mutagenized proteins will be further processed as required for structural determination using techniques such as NMR spectroscopy

3. REFERENCES

Allen, R.S. Li, J, Stahle, MI., Dubroue, AI, Gubler, F, and Millar, AA. (2007). Genetic analysis reveals functional redundancy and the major target genes of the Arabidopsis miR159 family. PNAS 104: 16371–16376

Anonymous (1999). Phosphorus Deficiency Symptoms in Some Crops. Better Crops. 83: 20-23.

Cassab, G.I. (1998). Plant cell wall proteins. Annual review of plant physiology and plant molecular biology. 49: 281-309

Dunker, A. K., Brown, CJ, Lawson, J D. Iakoucheva, LM, and Obradovic, Z. (2002). Intrinsic disorder and protein function. Biochemistry, 41:6573-6582.

Dyson HJ, Wright PE (2005) Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol 6: 197–20

Gordon AJ, Minchin FR, James CL, Komina O. 1999. Sucrose synthase in legume nodules is essential for nitrogen fixation. Plant Physiol. 120:867–78

Hammond, JP, and White, PJ (2008). Sucrose transport in the phloem: integrating root responses to phosphorus starvation. Journal of Experimental Botany. 59: 93–109, 2008

Huffaker, A., Pearce, G., and Ryan, CA. (2006). An endogenous peptide signal in Arabidopsis activates components if the innate immune response. PNAS. 103: 10098-10103

Kovacs, D, Kalmar, E, Torok, Z and Tompa, P (2008). Chaperone activity of ERD10 and ERD14, two disordered stress-related plant Proteins. Plant Physiol. 147: 381–390

Krysan, P.J., Young, J.C., Tax, F., and Sussman, M.R. (1999). T-DNA as an insertional mutagen in Arabidopsis. The Plant Cell 11: 2283–2290,

McGurl, B, Orozco-Cardenas, ML, Pearce, G, and Ryan, CA (1994). Overexpression of the prosystemin gene in transgenic tomato plants generates a systemic signal that constitutively induces proteinase inhibitor synthesis. Proc. Natl. Acad. Sci. USA 91: 9799-9802.

Matsubayashi Y (2003). Ligand-receptor pairs in plant peptide signaling. Cell Sci.116: 3863-70.

Matsubayashi, Y. and Sakagami, Y. (2006). Peptide Hormones in Plants. Annual Review of Plant Biol. 57:649-674

Miura K, Rus A, and Sharkhuu A (2005). The Arabidopsis SUMO E3 ligase SIZ1 controls phosphate deficiency responses. Proceedings of the National Academy of Sciences, USA 102:7760–7765.

Narváez-Vásquez J, Pearce G, and Ryan CA (2005). The plant cell wall matrix harbors a precursor of defense signaling peptides. Proc Natl Acad Sci. 102:12974-129777

Neales, TF (1968). The nutritional requirements of excised roots of three genotypes of Arabidopsis thaliana. New Phytol. 67: 159-165

Ryan, C. A. & Pearce, G. (2004). in Plant Hormones: Biosynthesis, Signal Transduction, Action, ed. Davies, P. J. (Springer, Secaucus, NJ) pp. 654-670.

R¨ohrig H, Schmidt J, Miklashevichs E, Schell J, John M. 2002. Soybean ENOD40 encodes two peptides that bind to sucrose synthase. Proc. Natl. Acad. Sci. USA 99:1915–20

R¨ohrig H, John M, Schmidt J. 2004. Modification of soybean sucrose synthase by Sthiolation with ENOD40 peptide A. Biochem. Biophys. Res. Commun. 325:864–70
Sa´nchez-Caldero´n L, Lo´pez-Bucio J, Chaco´n-Lo´pez A, Cruz-Ramı´rez A, Nieto-Jacobo F, Dubrovsky JG, Herrera-Estrella L (2005). Phosphate starvation induces a determinate developmental program in the roots of Arabidopsis thaliana. Plant and Cell Physiology 46: 174–184.
Seki M, Narusaka M, Ishida J, Nanjo T, Fujita M, Oono Y, Kamiya A, Nakajima M, Enju A, Sakurai T et al. (2002) Monitoring the expression profiles of 7000 Arabidopsis genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. Plant J 31: 279-292
Showalter AM. 1993. Structure and function of plant cell wall proteins. Plant Cell 5:9–23
Showalter, AM, Keppler, B, Lichtenberg, J, Gu, D, and Welch, LR (2010). A bioinformatics approach to the identification, classification, and analysis of hydroxyproline-rich glycoproteins. Plant Physiology 153:485-513
Ticconi CA, Delatorre CA, Abel S. (2001). Attenuation of phosphate starvation responses by phosphite in Arabidopsis. Plant Physiology 127, 963–972
Tompa P, Dosztanyi Z, Simon I (2006) Prevalent structural disorder in E. coli and S. cerevisiae proteomes. J Proteome Res 5: 1996–2000
Tompa P (2002) Intrinsically unstructured proteins. Trends Biochem Sci 27:527–533
Tylor, C.B. (1997). Promoter fusion analysis: An insufficient measure of gene expression. The Plant Cell. 9:273-275
Vance CP, Uhde-Stone C, Allan DL. 2003. Phosphorus acquisition and use: critical adaptations by plants for securing a nonrenewable resource. New Phytologist 157, 423–447.
Wei, G. and Shirsat, A.H., (2006). Extensin over-expression in Arabidopsis limits pathogen invasiveness. Molecular Plant Pathology. 6: 579–592