A Study of Seed Storage Protein Accumulation by Ectopic Expression in Arabidopsis

A Thesis Submitted to the College of Graduate Studies and Research
in Partial Fulfillment of the Requirements of the Degree of Doctor of Philosophy in the
Department of Plant Sciences, University of Saskatchewan, Saskatoon, SK

By

Gordon M. Gropp

ABSTRACT

Understanding the mechanisms plants utilize for seed storage protein (SSP) synthesis, transport and deposition have the potential rewards of enabling high yields of modified or foreign proteins. Hayashi et al. (1999) indicated that the machinery devoted to the synthesis of protein storage vacuoles in cotyledon cells can be induced in vegetative tissue by the constitutive expression of a pumpkin 2S albumin phosphinothricin-acetyl-transferase gene fusion (pumpkin 2S-PAT) resulting in the biogenesis of precursor-accumulating (PAC) vesicles in Arabidopsis leaves. This discovery was the impetus behind the work described which sought to examine this phenomenon further by ectopically evoking SSP trafficking and vesicle biogenesis machinery in leaves.

With the aim of elucidating the mechanisms necessary to evoke PAC vesicle biogenesis, a suite of constructs including the pumpkin 2S-PAT and analogous napin-PAT and napin-GFP variants were synthesized. Analysis of these transgenes in Arabidopsis revealed that the pumpkin 2S albumin has a capacity unique from napin peptides to result in fusion protein accumulation. Further, the truncated pumpkin 2S albumin peptide and the pumpkin 2S albumin C-terminus were found to direct deposition to vesicles; however, the C-terminus alone was not enough to direct deposition to vesicles unless combined with a significantly shortened napin peptide. An increased ER protein throughput was correlated to trafficking of the fusion protein by Golgi-independent mechanisms resulting in stable accumulation of the unprocessed protein whereas less ER throughput indicated passage through the Golgi-dependent pathway resulting in accumulation of a processed variant. At the level of gene expression, as examined by a microarray study, both inducible and constitutive ectopic expression of pumpkin 2S-PAT resulted in substantial perturbations of the endomembrane system affecting protein folding, flowering time and ER-associated biosynthetic functions which indicated that modulation of flowering time and photoperiodism are highly dependent on protein trafficking and vacuolar biogenesis mechanisms and that high ER protein throughput occurs at the expense of biosynthesis and cessation of ER functioning.

ACKNOWLEDGEMENTS

This thesis is the culmination of several years of labor that began with only an idea. The Hayashi et al. (1999) paper was the impetus for this idea. Not long after beginning work within the Seed Protein Group at AAFC, a large team interested in investigating the repression of seed storage protein expression in non-seed tissues, the Hayashi paper provided the impetus for examining ectopic expression more closely so as to discover the nuances of SSP sorting and trafficking. The idea was a slight digression of discovering SSP repression mechanisms in vegetative tissue, but it was thought that examining the perturbations caused by intentional ectopic expression in leaf tissue may provide insight into these mechanisms. I am eternally grateful to Dr. Derek Lydiate for embracing this idea and who championed me to become a Ph. D. graduate student within the AAFC organization.

There are several individuals I would like to acknowledge for their contributions to the completion of this work: namely, Mrs. Cathy Coutu for her expert assistance with Zeiss Apotome and confocal microscopy, Dr. Dwayne Hegedus for his valuable insight and encouragement, and my colleagues at AAFC whose camaraderie offered me inspiration. I sincerely thank the members of my committee for all their numerous suggestions which made my research and thesis more sound: Dr. Yuguang Bai, Dr. Kirstin Bett, Dr. Peta Bonham-Smith, Dr. Bruce Coulman, Dr. Gordon Gray and Dr. Sean Hemmingsen. I am especially grateful to Dr. Bonham-Smith whose devotion to my thesis allowed the document, and my own understanding of the science involved, to evolve considerably. Above all, I am eternally grateful to Dr. Kirstin Bett, who was, in addition to being a member of my committee, also my co-supervisor and whose efforts resulted in the completion of this work. Kirstin, thank you for your persistence and encouragement: Derek got me started and you got me finished. Thank you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation

AAFC	Agriculture and Agri-Food Canada
ANOVA	analysis of variance
AP	adaptor protein
aRNA	amplified RNA
AtELP	Arabidopsis thaliana epidermal growth factor-like protein
BiP	lumen binding protein
CaMV	cauliflower mosaic virus
ER	endoplasmic reticulum
CAB	chlorophyll a/b binding protein
CF	whole leaf protein extract (or crude) fraction
CRT	calreticulin
CCV	clathrin-coated vesicle
CM	cysteine motif
CmP	pumpkin 2S-PAT gene fusion
COPI	coat protein complex I
COPII	coat protein complex II
CPC	coat protein complex
CT	carboxy-terminus
ctVSD	carboxy-terminal vacuolar sorting determinant
dATP	deoxyadenosine triphosphate
dCTP	deoxycytosine triphosphate
dGTP	deoxyguanosine triphosphate
dNTP	deoxynucleoside triphosphate
dTTP	deoxythiamine triphosphate
DNA	deoxyribonucleic acid
DRP	dynamin-related proteins
DV	dense vesicle
EGF	epidermal growth factor-like receptor domain
EST	expressed sequence tag

FC	fold-change
FL	full length
GA	Golgi apparatus
GAP	GTPase activating protein
GCS	glucosidase
GDP	guanosine diphosphate
GEF	guanylate exchange factor
GFP	green fluorescent protein
GUS	β-glucuronidase
GRP	glucose regulated protein
GTP	guanosine triphosphate
HSP	heat shock protein
LB	Leuria-Bertani
Lea	Lewisa
LPVC	late prevacuolar compartment
LS	large subunit
LSP	leaderless secreted proteins
LV	lytic vacuole
mcKLH	mariculture keyhole limpet hemocyanin
mNap	codon modified Napin-2 gene
mRNA	messenger ribonucleic acid
MuDPIT	Multi-Dimensional Protein Identification Technology
NOS	nopaline synthase
ORF	open reading frame
OST	oligosaccharyltransferase
PA	protease-associated domain
PAC	precursor accumulating
PAT	phosphinothricin acetyl transferase
PCR	polymerase chain reaction
PDI	protein disulfide isomerase
pL4	probability loop region 4

pL5	probability loop region 5
PP	processed peptide
PPI	peptidyl-prolyl cis or trans isomerase
PPTR	phosphinothricin resistance
psVSD	physical structure vacuolar sorting determinant
PVC/MVB	prevacuolar compartment/multi-vesicular body
PM	plasma membrane
PSV	protein storage vacuole
Rab	a member of the Ras superfamily of monomeric G proteins
RAB	Rab GTPase protein
RE	recycling endosomes
RING	RING-H2 zinc finger domain
RMR	receptor homology-transmembrane-RING-H2 domain protein
RRE	arginine-arginine-glutamate amino acid triplet
rPCR	recombinant PCR
RT	room temperature
RT-PCR	reverse transcriptase polymerase chain reaction
SDM	site-directed mutagenesis
SDS-PAGE	sodium dodecylsulfate-polyacrylamide gel electrophoresis
SN	supernatant fraction
SNARE	soluble N-ethylmaleimide-sensitive factor attachment protein receptor
SOC	super optimal with catabolite repression
SP	signal peptide
SS	small subunit
SSP	seed storage protein
ssVSD	sequence-specific vacuolar sorting determinant
TGN	trans-Golgi network
TIP	tonoplast intrinsic protein
TMD	transmembrane domain
UGGT	glycoprotein glucosyltransferase
VF	vesicle fraction

VPE	vacuolar processing enzyme
VSD	vacuolar sorting determinant
VSR	vacuolar sorting receptor
WT	wild-type
X-GLUC	5-bromo-4-chloro-3-indolyl glucuronide

CHAPTER 1
INTRODUCTION

Seed development is a critical process in the life cycle of higher plants. The capacity to synthesize and accumulate large quantities of proteins is essential for early seedling growth and the success of the next sporophytic generation. In addition, seed storage proteins (SSP) represent the economic value of many dicot crops as they are vital nutritionally for humans and animals [1-3]. As such, the potential rewards of altering protein biosynthesis or enabling high yields of modified or foreign proteins through an understanding of the mechanisms underlying storage protein synthesis, transport and deposition are of keen interest.

The synthesis, trafficking and storage of seed proteins represent a major investment of assimilate and energy for which the plant endomembrane system is predominantly responsible [4]. Plant cells contain a variety of functionally distinct vacuolar compartments derived from the endomembrane system, each with unique attributes as to their contents and membrane constituents [5-7]. Seed storage proteins accumulate in specialized protein storage vacuoles (PSVs) that coexist in cells alongside lytic vacuoles (LVs) which are inherently different from each other. The co-existence of these and many other different types of vacuoles in a single cell represents a major challenge facing the sorting and trafficking machinery of the plant endomembrane system [8-11].

Although much has been ascertained regarding the sorting signals that are necessary for the trafficking of some cargo proteins to their correct destinations, we know little of the complex mechanisms that guide protein transport within the endomembrane system and the role of vacuole protein trafficking in development. Certainly, reverse and forward genetics approaches in Arabidopsis have assisted us greatly in elucidating these mechanisms, but they are hindered by gene redundancy or by being essential. On one hand, a high frequency of gene duplication and

functional overlap impose difficulties for studying loss-of-function mutations that often result in no observable phenotype [12]. Alternatively, the critical nature of some endomembrane trafficking components leads to lethality following mutation [13]. Researchers have begun to remedy these challenges through functional genomics approaches and chemical genomics, i.e. microscopy assisted by immunogold labeling and fluorescent protein markers and the use of small molecules aimed at creating perturbations in cellular processes, respectively [14]. These strategies permit *in vivo* observation of sorting behavior and real time assessment of cellular physiology both prior to and subsequent to perturbations of gene expression or protein—protein interactions [15-17].

Investigations into SSP trafficking and vacuolar biogenesis mechanisms as they occur in seeds face additional challenges imposed by handling difficulties for some microscopy procedures [18]. The findings of Hayashi et al. (1999) suggest that the machinery devoted to the synthesis of PSVs in cotyledon cells can be induced in vegetative tissue. The constitutive expression of a pumpkin albumin-phosphinothricin acetyl transferase gene fusion (pumpkin 2S-PAT) resulted in the biogenesis of precursor-accumulating (PAC) vesicles in Arabidopsis leaves [19]. Therefore, I sought to examine this phenomenon further by evoking SSP trafficking and vesicle biogenesis machinery ectopically in leaves. The utility of inducing these processes in leaves for the study of protein targeting behavior was three-fold: 1) the use of leaf tissue to analyze expression of SSP fusions offered a suitable rapid testing system with abundant experimental material, 2) fusion protein expression leading to the biogenesis of PAC-like vesicles might require the absence of endogenous SSP, and 3) the induction was expected to occur without the background of active genes normally found in cotyledon cells during seed development.

1.1 Objectives and Hypothesis

This thesis describes work conducted to 1) develop transgenic Arabidopsis plants harboring SSP gene fusions resulting in the accumulation of fusion protein in leaf tissue, 2) examine fusion protein localization by microscopy and molecular techniques, 3) compare various pumpkin 2S and napin albumin fusion constructs and elucidate which features of these albumins evoke accumulation behavior in leaf tissue, and 4) examine the effects of both induced and constitutive expression of the pumpkin 2S-PAT gene fusion in leaf tissue, and to identify genes involved in vesicle biogenesis, protein sorting and the trafficking machinery.

These objectives were undertaken to test the following hypotheses: 1) the expression of pumpkin and napin peptide gene sequences fused to phosphinothricin acetyl transferase (PAT) will illustrate that the capacity to evoke vesicle biogenesis and accumulation of albumin fusion proteins in vesicle structures is a general phenomenon of albumin proteins sharing common secondary structures, 2) the use of the PAT gene fused to albumin peptide gene sequences will permit the identification of plants capable of successfully sequestering albumin-PAT fusion proteins into vesicles by rendering them sensitive to phosphinothricin, 3) the fusion protein thus sequestered into vesicle structures will be protected from degradation and be stably accumulated, 4) the ectopic expression of albumin-PAT gene fusions resulting in the biogenesis of fusion protein accumulating vesicles in leaf tissue will facilitate the discovery of genes devoted to sorting and trafficking machinery of SSPs.

CHAPTER 2
LITERATURE REVIEW

2.1 Overview of the Plant Endomembrane System

The endomembrane system is thought to be the most versatile system in the plant cell. It must orchestrate the correct sorting and trafficking of a myriad of vacuolar cargo proteins during various developmental stages as well as facilitate endosomal traffic of membrane proteins and extracellular constituents appropriate to environmental cues. As such, the endomembrane system is in a constant state of flux as it navigates these various challenges. These tasks can generally be thought of as belonging to either: 1) synthesis, modification and trafficking of nascent proteins destined for secretion or subcellular compartmentalization, or 2) the post-Golgi regulation of endocytic traffic to and from the plasma membrane (PM) (Figure 2-1) [10].

The gateway for proteins of the synthetic route is the rough endoplasmic reticulum (ER) in which de novo synthesis occurs cotranslationally due to the presence of a hydrophobic N-terminal signal peptide (SP) that is cleaved from the growing peptide chain as it enters the luminal space [20]. Within the ER lumen, the nascent protein is glycosylated and folded, in a process mediated by chaperones, before proceeding further towards its destination [21]. From the ER several routes exist which may either involve passage through the Golgi complex or ER-derived vesicles that bypass the Golgi. Several factors unique to the proteins themselves are thought to govern the route chosen and include sorting signals, aggregation behavior and membrane association. There is also much evidence for plant species, tissue and developmental stage being contributing factors to the route employed [1, 6, 18, 22].

In plants, the Golgi apparatus (GA) is considered a highly variable and often amorphous organelle consisting of multiple stacks of membrane-bound flattened cisternae, the number of which depend on species, developmental stage, and functional state of the cell [23]. Despite

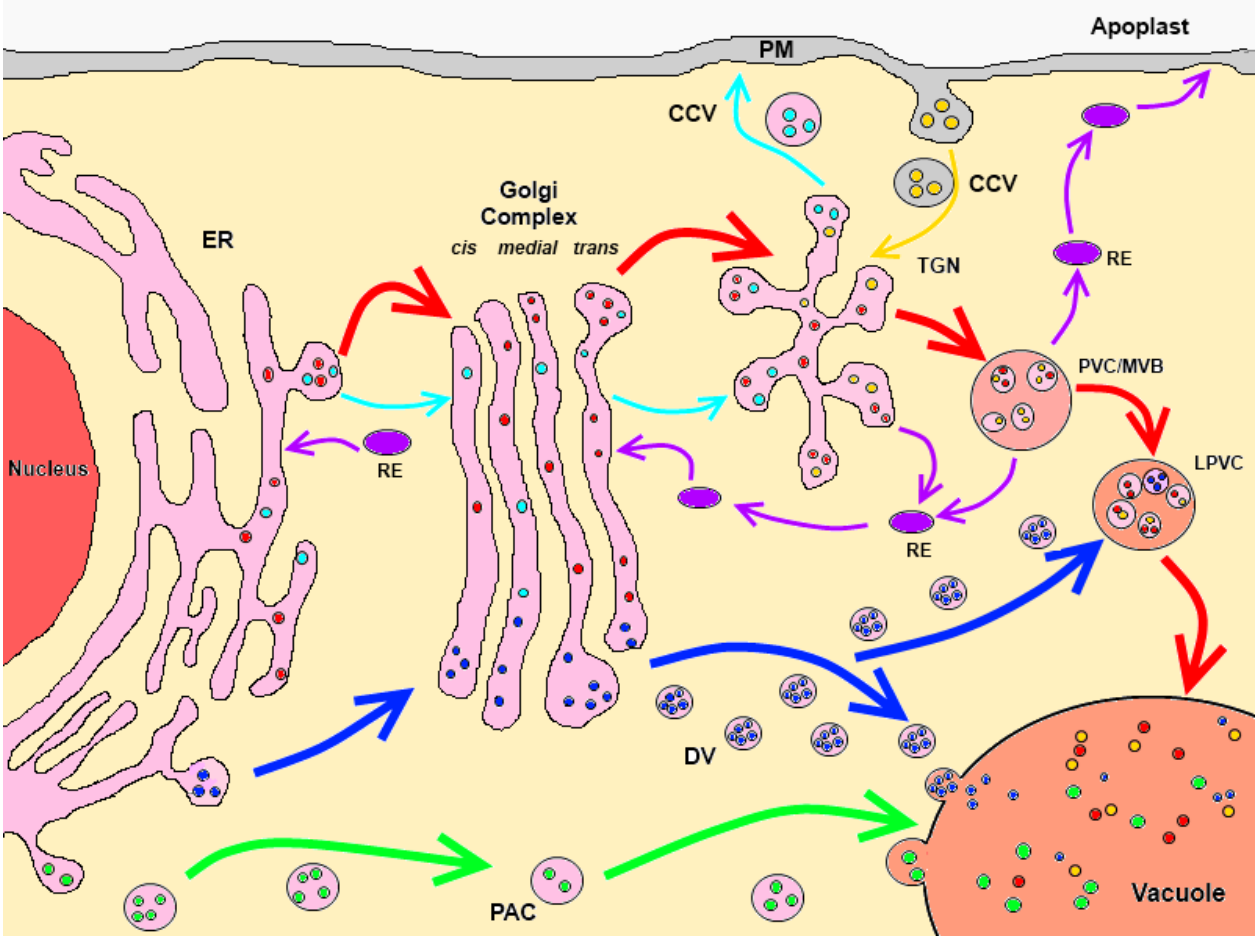


Figure 2-1. A simplified model of the plant endomembrane system showing biosynthetic and post-Golgi endocytic traffic. Most proteins, including 2S albumins destined for PSVs, are sorted by the clathrin-coated vesicle (CCV)-mediated pathway (red arrows) however, the 7S and 11S globulins have been known to follow Golgi-dependent aggregation-based sorting via dense vesicles (DV) (blue arrows) or may completely bypass the Golgi complex by a direct route using precursor-accumulating (PAC) vesicles as is the case in pumpkin seeds. The majority of proteins traverse by way of the Golgi to the trans-Golgi network (TGN). Proteins with no other targeting sequence aside from the signal peptide, are directed by default to the trans-Golgi network (TGN) and then to the plasma membrane (PM) for secretion. Extracellular constituents enter the cell by clathrin-mediated endocytosis and join the milieu of the TGN (yellow arrow). Once in the TGN, proteins destined for vacuoles are sorted to a prevacuolar compartment (PVC) (also called the multivesicular body, MVB) which undergo maturation to the late prevacuolar compartment (LPVC) comprised of mature content proteins and depleted of sorting receptors. The LPVC ultimately fuses with its vacuolar target. Underlying these processes is the retrograde transport of components essential to the sorting machinery that are recovered in recycling endosomes (RE) and directed back to earlier steps (purple arrows).

this incongruity, the GA has defined polarity from *cis* to *trans* proximity from the ER as identified by gradual changes in morphology and enzymatic properties [24, 25]. These changes in the GA from the *cis*- to *trans*-face correspond to a continuum of complex carbohydrate modifications necessary for the maturation of N-linked carbohydrate moieties conducted by glycosyltransferases and glycosidases [26]. As such, the localization of N-glycan modifying enzymes acting on cargo proteins represents their gradual passage through the GA [27].

The *trans*-Golgi network (TGN) has been proposed to act as a hub between the synthetic and endocytic routes because at this junction, newly synthesized proteins received from the Golgi complex and those cargo proteins endocytosed from the PM converge [14, 28]. Here it is thought that a multitude of sorting occurs whereby proteins are further directed to their destinations. Proteins flagged for secretion to the apoplast or integration as membrane constituents are sent to the PM and proteins having subcellular destinations such as the lytic vacuole (LV), protein storage vacuole (PSV) and plastids are believed to traverse through the prevacuolar compartment (PVC) or the multi-vesicular body (MVB), with components essential for continued sorting being transported in retrograde fashion via recycling endosomes (RE). There is not a clear distinction between the PVC and MVB with the terms being used interchangeably but the multi-vesicular nature of the organelle originates from its formation by fusion of multiple vesicles and its tendency to engulf membrane domains by autophagy for degradation in lysosomes [29-32]. However, there is a growing consensus that a distinction can be made between the PVC and the late PVC (LPVC) with the transition from early to late being related to aspects of PVC organelle maturation [33]. Prior to fusion of the LPVC with its vacuolar deposition target, it appears that the vesicle must attain fusion competence, during

which sorting proteins, such as vacuolar sorting receptors (VSRs) are recovered, and the vesicle is enriched sufficiently in soluble protein to attain a cargo threshold. Once this threshold is reached, further fusion to new vesicles originating from the Golgi/TGN no longer occurs, recovery of sorting proteins is complete and the composition of the LPVC matches its vacuolar destination, a prerequisite thought necessary for fusion [34].

2.2 Seed Storage Proteins and the Protein Storage Vacuole

To ensure the success of the next sporophytic generation, plants synthesize and accumulate vast stable quantities of seed proteins, aptly termed seed storage proteins (SSPs). A common feature of SSPs is their deposition during seed formation into discrete vacuolar compartments called protein storage vacuoles (PSVs) which during seed germination are transformed into lytic vacuoles resulting in PSV contents being mobilized as a nutritional resource for future growth [21, 35, 36]. This section will discuss in general terms the PSV and the different SSP families highlighting features unique to seed albumin proteins with which this thesis is primarily concerned.

2.2.1 SSP Classification

SSPs were first classified by Osborne (1924) based on their solubility characteristics: water soluble albumins, dilute saline soluble globulins, alcohol/water soluble prolamins, and dilute acid or alkali soluble glutelins [37]. Albumins and globulins are the major proteins in the seeds of dicots; whereas in monocots, prolamins comprise roughly 50% of SSP. This is true of the cereal grains except for oats and rice which instead have limited prolamin content (typically no more than 10%) and abundant glutelin [38-40]. Later, SSPs were classified according to their sedimentation coefficients by sucrose density gradient centrifugation: 2S albumins, 2S prolamins, and 7S or 11-12S globulins [41-43].

The 2S albumins have been thoroughly studied, due to their abundance in the Cruciferae, namely *Brassica napus* and *Arabidopsis thaliana*, in which they constitute approximately 20% of seed protein and have been called napin. Napin shares homology with 2S albumins from other dicots such as Brazil nut, pumpkin, walnut, cashew, sunflower or castor bean as well as cereal prolamins and is a member of the prolamin superfamily, many members of which have putative plant defense qualities suggested by their sequence identity with anti-nutritional α-amylase and

trypsin inhibitors (Figure 2-2) [3, 44-48]. Like all other SSP, 2S albumins are first synthesized as pre-proproteins that enter the biosynthetic secretory pathway of the endomembrane system through the ER as directed by an N-terminal SP (see section 2.3). Once in the ER, maturation of 2S albumins is unique from other SSP as it typically involves removal of proprotein regions to yield two subunits linked together by interchain disulfide bridges to create a heterodimer. In the case of napin, the heterodimer is comprised of a 4.5-kDa small subunit and a 10-kDa large subunit united by disulfide bonds that loop the long chain towards the short chain in anti-parallel fashion, specifically: Cys5-Cys59´, Cys18-Cys48´, Cys61-Cys104´, and likely Cys49-Cys96´ [49]. The location of cysteine residues within the albumin primary structure that permit this particular spatial arrangement of short and long chains is called the eight cysteine motif (8CM) and is highly conserved among albumins, having been observed in diverse members of the 2S family from castor bean, lupin, cotton and pumpkin. It is identified as a pattern of eight cysteine residues that permit four interchain disulfide bridges [3, 44, 49-53].

2.2.2 PSV Architecture and Biogenesis

The PSV is considered a vacuole unique to plants with specialized structure and function [54]. No other vacuole has been observed with a vacuole within a vacuole organization such that storage and lytic functions are partitioned. From microscopy studies of PSV ultra structure, it is known that PSVs possess three morphologically distinct regions: the matrix, the globoid and the crystalloid, although the crystalloid may not be present in some species [54]. In the case of the Cruciferae, the crystalloid is present and has been shown to be comprised of membrane-derived proteins and lipids, and phytic acid crystals in close association with 11S globulin hexamers. The 2S albumins and 7S globulins as well as other auxiliary storage proteins are stored together

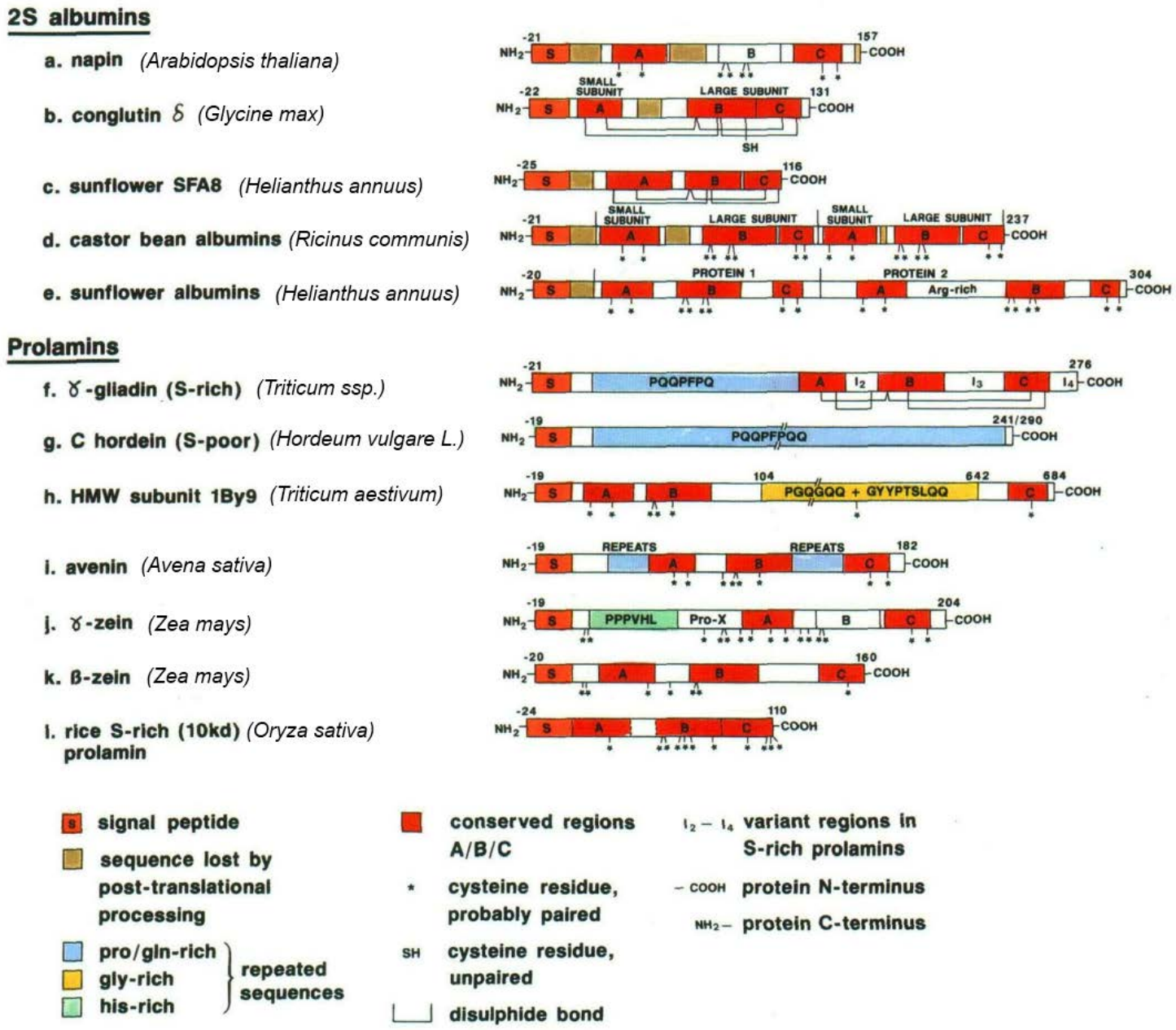


Figure 2-2. Schematic comparison of members of the seed protein prolamin superfamily, including 2S albumins of dicots and monocot prolamins indicating homologous regions A, B, and C which are represented in nearly every member except C hordein. Also, cysteine residues critical for common secondary and tertiary structural motifs are highly conserved. Adapted from Shewry, et al.,1995 [4].

in the adjoining matrix. For species in which no crystalloid is present, for example the PSV of legumes (including peas), all SSPs appear to be distributed evenly within the PSV [21]. The globoid has features common with lytic vacuoles and contain a wide array of enzymes capable of hydrolyzing macromolecules for nutrient mobilization during the early stages of germination.

The fact that the PSV is densely packed with both storage compounds and enzymes required for degradation calls into question the manner by which PSVs are made and the precarious nature of this co-localization [21]. One model postulates that the PSV is derived from a preexisting central vacuole which becomes successively subdivided and filled with storage proteins during SSP synthesis and accumulation. A more current model, and perhaps more widely accepted view, is that one large PSV is synthesized *de novo*, replaces the preexisting central vacuole by an autophagy-like process and then is partitioned into individual PSVs [6, 55]. Evidence for this resides in the fact that membrane composition differs between the PSV, tonoplast and the delimiting membrane of the globoid. PSV membranes characteristically possess the aquaporin alpha-tonoplast intrinsic protein (α-TIP) whereas the TIP variant associated with the central vacuole tonoplast is γ-TIP. However, γ-TIP has been identified in the membrane of the globoid suggesting a central vacuole origin of the PSV [54, 56, 57]. Regardless of PSV origin, evidence suggests that PSV filling is the result of autophagy-like internalizations of LPVCs or DVs derived from the Golgi-dependent pathway and Golgi-independent PAC vesicles (see Fig. 1-1) [18, 58-60].

2.3 SSP Sorting and Trafficking Mechanisms

Many biological functions in plant cells are mediated through vacuoles, which indicate multiple vacuoles must co-exist, each with their own unique assortment of membrane proteins, lipids and contents [6, 11]. This implies that separate mechanisms exist to ensure that correct sorting and trafficking of vacuolar contents occurs. Indeed two pathways have beeen identified; the synthetic pathway and the endocytic pathway, with the distinction being the involvement of the Golgi apparatus (see section 2.1). Much effort has been devoted to elucidating how correct sorting of proteins occurs but a clear understanding of these processes remains elusive. This is perhaps due to the study of multiple plant species and tissues to develop a singular model which explains vacuolar biogenesis and protein trafficking [6]. However, it has been well established that these processes are signal-mediated as proteins devoid of any known targeting signal other than the SP are ultimately secreted from the plant cell or they are sorted by an aggregation-based mechanism that has been observed in some species such as pumpkin [10, 22]. This section will discuss the current understanding as it pertains to SSP sorting and trafficking from the ER to the PSVs.

2.3.1 Vacuolar Sorting Determinants (VSDs)

As mentioned above, without a targeting signal newly synthesized proteins assembled and trafficked through the endomembrane system are destined for secretion to the apoplast. This implies that extracellular secretion is the default pathway and selection of a protein for a subcellular destination must rely on an active mechanism. Scientists in this field have identified vacuolar sorting determinants (VSDs) that reside within the proteins themselves and have been classified into three types: the sequence-specific VSD (ssVSD), the carboxy-terminal VSD (ctVSD) and the physical structure VSD (psVSD) (Table 2-1) [22, 61, 62]. The ssVSDs are typically within an N-terminal propeptide region and contain an NPIXL/NPIR amino acid

Table 2-1. Vacuolar Sorting Determinants of SSP Destined for the PSV

Protein	Organism	Sequence	Category	Location/Action[1]	Refs
Amaranthin	*Amaranthus hypochondriacus*	GNIFRGF	ssVSD	Internal Induces	[63]
Cruciferin	*Brassica napus*	ICSMR	ssVSD	Internal	Hegedus, Coutu personal com.
A3B4 subunit of glycinin	*Glycine max*	ICTMR (critical Ile-297)	ssVSD	Internal Required/Induces	[64]
2S albumin	*Ricinus communis*	STGEEVLRMPGDEN	ssVSD	Internal Required/Induces Interacts with BP-80	[65, 66]
Ricin (7S globulin)	*Ricinus communis*	SLLIRPVVPNFN	ssVSD	Internal Required/Induces Interacts with BP-80	[66, 67]
Amaranthin (11S globulin)	*Amaranthus hypochondriacus*	KISIA	ctVSD	C-terminal Induces	[63]
Cruciferin (11S globulin)	*Arabidopsis thaliana*	ASYGRPRVAAA	ctVSD	C-terminal Required/Induces	[68]
2S albumin	*Bertholletia excelsa*	IAGF	ctVSD	C-terminal Required/Induces Interacts with BP-80	[69]
2S albumin	*Cucurbita maxima*	KARNLPSMCGIRPQRCDF	ctVSD	C-terminal Induces Interacts with BP-80	[70]
α′-subunit of β-conglycinin (7S globulin)	*Glycine max*	PLSSILRAFY	ctVSD	C-terminal Required/Induces	[71]
β-subunit of β-conglycinin	*Glycine max*	PFPSILGALY	ctVSD	C-terminal Required/Induces	[72]
A1aB1b subunit of glycinin (11S globulin)	*Glycine max*	PQESQKRAVA	ctVSD	C-terminal Required/Induces	[64]
B-Phaseolin (7S globulin)	*Phaseolus vulgaris*	AFVY	ctVSD	C-terminal Required/Induces	[73]
B-type Legumin (11S globulin)	*Vicia faba*	Multiple elements	psVSD	Internal Induces	[74]
A1aB1b subunit of glycinin	*Glycine max*	Putative single element	psVSD	Internal Induces	[64]
A3B4 subunit of glycinin	*Glycine max*	Putative single element	psVSD	Internal Induces	[75]

[1] Required: Removal of the VSD from the SSP results in default secretion to the apoplast. Induces: Addition of the VSD to a secreted protein is sufficient for trafficking to the PSV. Modified from Vitale and Hinz, 2005 [22]

sequence motif but can also exist within non-propeptide or C-terminal propeptide regions and frequently possess an essential isoleucine or leucine amino acid residue. ssVSD have been known to associate with vacuolar sorting receptor (VSR) proteins during Golgi-dependent clathrin-mediated trafficking to the lytic vesicle (LV). The ctVSDs have no clearly identifiable consensus sequence other than a stretch of highly hydrophic residues of variable length, do not rely on essential isoleucine or leucine amino acid residues and do not function unless they remain at the C-terminus. The psVSDs have neither an identifiable consensus sequence nor other common feature and their sorting signal function may result from an entire domain within the protein that exerts its effects through secondary or tertiary polypeptide structure [22]. This is thought to be the mechanism underlying aggregation-based sorting of some SSP such as pumpkin 11S globulin or rice 11S glutelin that form by budding off from the ER membrane [59, 76, 77], or which may contribute to the formation of Golgi-derived DVs as occurs with pea legumin [18, 78-80]. The VSDs are found in SSPs both as singular or multiple entities that co-operatively form sorting receptor binding sites likely through protein folding [18, 64]. Evidence for co-operative effects of VSD, comes from the observation that the efficiency of sorting appears to depend on cumulative contributions of multiple determinants [18, 22]. Based on these studies, it has been found that a hierarchy of contribution of VSDs exists in which some contribute marginally to sorting whereas others are essential and their absence results in the SSP being secreted to the apoplast by the default pathway (“Required”, Table 2-1). Alternately, some VSDs can on their own, direct proteins to PSV (“Induces”, Table 2-1) and their examination for use in heterologous expression of recombinant proteins is an active area of research [81]. Much has been elucidated regarding sorting signals and their receptors; however the task has been daunting. Several questions still need answering. Namely, why do numerous proteins destined

for the same subcellular compartment possess different types of sorting determinants, or why do the majority of secreted proteins in the plant secretome have no signal peptide for entry into the ER (called leaderless secreted proteins, LPS) [82-84]? So far, the reasons remain obscure.

2.3.2 Targeting via VSRs and RMRs

Since finding sorting determinants within secreted proteins that function as sorting receptor binding sites, much effort has been devoted to the identification of these receptor proteins. Two separate membrane-protein receptor families have been identified: vacuolar sorting receptors (VSRs) and the receptor homology-transmembrane-RING H2 domain proteins (RMRs) [11, 18, 85-87]. Both are thought to mediate targeting by recruiting proteins of the synthetic pathway into transport vesicles destined for deposition into larger vacuoles.

VSRs. The VSRs are also known as the BP-80 family. BP-80, involved in the transport of SSP in pea cotyledon (*Pisum sativum*), was the first transmembrane protein to be isolated from clathrin-coated vesicles (CCVs). Since then ~80-kDa homologues have been identified in other species, namely PV72 from pumpkin and VSR1/AtELP (epidermal growth factor[EGF]-like protein) from Arabidopsis [88-90]. In Arabidopsis, seven members of AtELP exist possessing a conserved protease-associated domain known to interact with VSDs, especially the NPIR amino acid binding motifs required for sorting of proteins to the LV (Figure 2-3) [1, 11]. VSR proteins also possess three EGF repeats followed by a transmembrane domain and a short cytosolic tail containing a tyrosine motif, YMPL that is believed to be a binding site for the μA-adaptin subunit of an adaptor protein1 (AP-1) which recruits clathrin during formation of CCVs [32, 91, 92]. With certainty, clathrin-mediated sorting involving VSR1/AtELP in the endocytic pathway occurs, but it is currently unclear if VSR proteins employ clathrin during targeting in the synthetic route [32, 93-95]. Regardless, functional genetics studies in which VSR mutants misdirect SSP to the apoplast provide clear evidence for VSR involvement in sorting of SSP

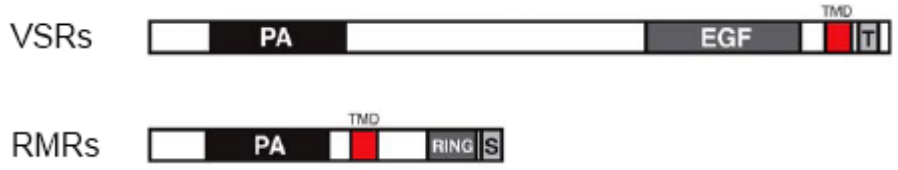


Figure 2-3. Comparison of VSR and RMR family of proteins showing shared secondary structure motifs. Both are transmembrane proteins possessing a protease-associated domain (PA) which for the VSR family is known to interact with the VSDs of target proteins. VSRs differ from RMRs on the luminal side of the transmembrane domain (TMD) by the size of the luminal region and by the presence of an epidermal growth factor-like receptor domain (EGF). The cytosolic regions of VSRs and RMRs differ considerably as VSR possesses a carboxy termini (T) that interacts with the μA-adaptin subunit of an AP-1 adaptor complex thought to be required for coat protein complex (CPC) formation, whereas RMRs have a larger cytosolic region that shares features with E3-ligases by virtue of its RING-H2 zinc finger (RING) domain and a serine rich domain (S). Adapted from Zouhar and Rojo, 2009 [11].

[96] and that vacuolar sorting receptor proteins-1, -3 and -4 (VSR1, VSR3 and VSR4) are predominantly responsible for the majority of sorting of the 12S globulin to the PSV, however the major VSR responsible for 2S albumin sorting has not been established [87, 97]. It is generally thought that VSRs first undergo ligand-receptor recognition and binding in the Golgi where accumulation and aggregation precedes the formation of DVs [1]. This is supported by localization studies which indicate that SSPs co-localize with VSRs at the edges of *trans*-Golgi cisternae and in DVs [95]. However, research conducted by Niemes et al., (2010), indicates that VSRs associate with their cargo proteins initially in the ER lumen and then begin to return to the ER via retromer-mediated recycling once they reach the TGN [98-100]. This implies that VSR-bound cargo traverses the synthetic pathway from the ER until reaching the TGN/PVC but localization studies have not yet identified VSR proteins in the *cis*- or *medial*-Golgi regions [11]

RMRs. The role of RMRs in SSP sorting is much less established. In Arabidopsis, these proteins are represented by a family of six genes. From this family the gene *RMR1* appears to be expressed ubiquitously in vegetative tissue independent of developmental stage of the plant [95] but recent evidence indicates that it, together with *RMR2*, are expressed at a greater level in seed tissue than in other tissues [87, 101]. VSR and RMR proteins share a common lumenal domain known to associate with VSDs (Figure 2-3) but RMRs differ from VSRs in that they possess a RING finger E3-ligase domain in their cytosolic tail. Furthermore, localization studies have shown that RMR proteins, unlike VSR proteins which dissociate from their ligands at acidic vacuolar pH, co-localize with SSPs in the *cis*- and *medial*-Golgi cisternae [95] and *in vitro* bind to ctVSD of SSPs independent of pH [86, 102]. Therefore, it is possible that SSP sorting is conducted through co-operative binding of protease-associated (PA) domains by VSRs and RMRs that target the carboxyl termini of VSDs through nonreversible binding. This mechanism suggests, together with the presence of the cytosolic RING finger domain of RMRs, that once bound with SSP cargo, E3 ligase function triggers auto-ubiquination to ensure targeting to the PVC and ultimately the PSV [11, 102]. This mechanism is supported by the finding of RMR protein in the PSV crystalloid [1, 103].

Aggregation-based sorting. The precise role of VSR and RMR proteins in SSP sorting has not yet been clearly established partly due to an inability to obtain genetic evidence for the involvement of these receptors. Both VSR and/or RMR single and multiple mutants show unimpaired or only partially impaired SSP sorting indicated by accumulation of SSP precursors, a hallmark of SSP missorting [87, 104, 105]. Researchers speculate that this may be due to functional redundancy among members of the VSR and RMR gene families or that this may be due to the presence of an aggregation-based sorting mechanism. This would imply that VSR, and

possibly RMR involvement is limited to scavenging SSP that escape sorting by a primary mechanism [87, 95, 106-108]. Evidence for an aggregation-based mechanism being the primary means for SSP sorting is supported by observations for rice glutelin in which an aggregation-based mechanism appears to be initiated in the ER, with glutelin mRNA localization into ER subdomains that contribute to ER-derived vesicles [77, 109].

2.3.3 Trafficking Machinery

Coupled with receptor-mediated recognition and binding of SSP is the sequestering of this cargo into transport vesicles, trafficking of these vesicles and the subsequent tethering and fusion to their target membranes. These processes involve a unique assortment of proteins particular to the type and destination of the cargo. This machinery cooperatively orchestrates the biogenesis and delivery of carrier vesicles as well as the retrograde transport of essential trafficking components. Coat protein complexes (CPC) are integral to formation of transport vesicles; Rab GTPases govern vesicle intracellular movement; and soluble N-ethylmaleimide-sensitive factor attachment protein receptors (SNAREs) mediate vesicle-target membrane fusion. All of these proteins are comprised of large families that are functionally diversified so that considerable division of labor exists for the completion of these tasks, illustrating the diverse nature of the endomembrane system in plants [110-113]. In addition, there are many other proteins involved including cytoskeleton associated proteins and molecular motor proteins, that make movement of transport vesicles possible, but these will not be discussed in this thesis [114, 115].

Coat Protein Complexes (CPC). The biogenesis of transport vesicles is achieved by the work of CPC that include: 1) coat protein complex II (COPII) that mediates anterograde traffic of transport vesicles from the ER to the *cis*-Golgi, 2) coat protein complex I (COPI) that mediates inter-Golgi traffic and retrograde traffic from the Golgi to the ER, and 3) clathrin-

mediated complex that is involved in post-Golgi traffic of proteins derived from both the synthetic and endocytic pathways [116-118]. Essential to CPC function are dyamin-related proteins (DRP) that promote the formation of discrete vesicles by scission between the donor and newly formed vesicle membranes through GTPase action [113, 119, 120]. Clathrin-mediated complexes are further dependant on adaptor protein complex I (AP-1), AP-2 and AP-3 which have been shown to interact with clathrin during complex formation of coated pits on donor membranes [121]. AP-1 and AP-2 are believed to be involved in traffic through the endocytic pathway as they have been observed during coated pit formation on the PM and TGN membranes [122, 123]; however, AP-3 is believed to assist with traffic from the TGN to the PSV/LV [117, 124]. The roles of the remaining members of adaptor protein complexes, AP-4 and AP-5 are yet to be clearly established but are believed to be involved in traffic involving the TGN and associated endosomes of the endocytic pathway [125, 126].

Rab GTPases. Rab GTPases are well represented in the Arabidopsis genome with 57 members organized into eight diversified clades (RAB-A to RAB-H) which are orthologous to Rab GTPases in mammals. Despite the observed similarities, the plant RAB-A clade (also called RAB11) is unique and has expanded and diversified substantially. The RAB-A clade is subdivided into six groups, RAB-A1 through RAB-A6. Because members of the RAB-A clade account for 26 of the total 57 Rab GTPases, it is thought that they are vital to and correlated with, the highly diversified functions in the trafficking machinery of land plants [127, 128]. Elucidating the roles of the various Rab GTPases in anterograde and retrograde trafficking events and identifying which of these are important for SSP trafficking has been challenging due to their large number and potential redundancy. Localization studies have allowed the function of the various Rab GTPases to be inferred: RabB and RabD localize to ER and Golgi compartments

[10, 129], RabE and RabH to the Golgi only [130, 131], RabG to vacuolar membranes and endosomes [132], RabF to the LPVC [133], and RabA and RabC localize to the TGN, endosomal compartments and the cell plate [10, 129]. Localization studies conducted by antibody labeling is not conclusive for a functional role, however, recently two Rab GTPases from the D clade, RAB-D1 and RAB-D2 were identified as having overlapping functions related to early stages of biosynthetic traffic between the ER and the *cis*-Golgi [134].

Despite shortcomings in our knowledge about the specific roles of these various Rab GTPases in plants, the mechanism behind Rab GTPase function has been well established. Rab GTPases are known to cycle between GDP-bound inactive and GTP-bound active states by the actions of guanylate exchange factors (GEFs) and GTPase activating proteins (GAPs). GEFs catalyze the exchange of bound GDP for GTP, where as GAPs stimulate the hydrolysis of bound GTP to bound GDP which results in cessation of Rab activity. [129]. Rab GTPases are typified by shared structural features and between 30 to 50% sequence identities (Figure 2-4). All Rab GTPase possess four regions involved in GTP nucleotide binding designated regions G1, G3, G4 and G5 which impart conformational changes during GTP-GDP cycling that is essential to their function [127]. In the active state, Rab-GTPases together with assembled effector proteins, orchestrate membrane recognition through lipid binding [135], assist in vesicle formation, tethering and fusion [136, 137], and recruit cytoskeletal molecular motor proteins for transport vesicle movement [129, 138]. A large number of Rab effectors are employed and characterizing their contribution to Rab GTPase function will be a long process [139, 140].

```
                                 G1                          G2                  G3          YRG          100
  AtRAB1b    (1) ---------MNPEYDYLFKLLLIGDSGVGKSCLLLRFSDDSYVESYISTIGVDFKIRTVEQDGKTIKLQIWDTAGQERFRTITSSYYRGAHGIIIVYDVT
 AtRABE1c    (1) --MAAPPARARADYDYLIKLLLIGDSGVGKSCLLLRFSDGSFTTSFITTIGIDFKIRTIELDGKRIKLQIWDTAGQERFRTITTAYYRGAMGILLVYDVT
  AtRABC1    (1) ----MGSSSGQPEFDYLFKVLLIGDSGVGKSSLLLSFTSNTFDDLSP-TIGVDFKVKYLTIGEKKLKLAIWDTAGQERFRTLTSSYYRGAQGIIMVYDVT
  tRABF2a    (1) -------MASSGNKNINAKLVLLGDVGAGKSSLVLRFVKDQFVEFQESTIGAAFFSQTLAVNDATVKFEIWDTAGQERYHSLAPMYYRGAAAIIVFDIT
 AtRABF2b    (1) -------MAAAGNKSINAKLVLLGDVGAGKSSLVLRFVKDQFVEFQESTIGAAFFSQTLAVNDATVKFEIWDTAGQERYHSLAPMYYRGAAAIIVFDVT
 AtRABG3e    (1) ---------MPSRRRTLLKVIILGDSGVGKTSLMNQYVNKKFSNQYKATIGADFLTKEVQFEDRLFTLQIWDTAGQERFQSLGVAFYRGADCCVLVYDVN
 AtRABH1d    (1) --------MASVSPLAKYKLVFLGDQSVGKTSIITRFMYDKFDTTYQATIGIDFLSKTMYLEDRTVRLQLWDTAGQERFRSLIPSYIRDSSVAVVVYDVA
 AtRABA2c    (1) -----MTHRVDQEYDYLFKIVLIGDSGVGKSNILSRFTRNEFCLESKSTIGVEFATRTTQVEGKTIKAQIWDTAGQERYRAITSAYYRGAVGALLVYDIT
 AtRabA4B    (1) MAGGGGYGGASGKVDYVFKVVLIGDSAVGKSQLLARFARDEFSMDSKATIGVEFQTRTLSIEQKSIKAQIWDTAGQERYRAVTSAYYRGAVGAMLVYDMT
 AtRABB1a    (1) -----------MSYAYRFKYIIIGDTGVGKSCLLLKFTDKRFQAVHDLTIGVEFGAKTITIDNKPIKLQIWDTAGQESFRSVTRSYYRGRAGTLLVYDIT
Consensus    (1)          A     YLFKLVLIGDSGVGKSSLLLRF  D F     STIGVDF SKTL VEDKTIKLQIWDTAGQERFRSLT AYYRGA GAILVYDVT

                101                                 G4                                  G5                          200
  AtRAB1b   (92) DEESFNNVK-QWLSEIDRYAS----DNVNKLLVGNKSDLTEN--RAIPYETAKAFADEIG-IPFMETSAKDATNVEQAFMAMSASIKERMASQPAG--NN
 AtRABE1c   (99) DESSFNNIR-NWIRNIEQHAS----DNVNKILVGNKADMDESK-RAVPTAKGQALADEYG-IKFFETSAKTNLNVEEVFFSIGRDIKQRLSDTDS----R
  AtRABC1   (96) RRDTFTNLSDIWAKEIDLYSTN---QDCIKMLVGNKVDKESE--RAVSKKEGIDFAREYG-CLFLECSAKTRVNVEQCFEELVLKILETPSLTAE-----
 AtRABF2a   (94) NQASFERAK-KWVQELQAQGN----PNMVMALAGNKADLLDA--RKVSAEEAEIYAQENS-LFFMETSAKTATNVKDIFYEIAKRLPRVQPAEN------
 AtRABF2b   (94) NQASFERAK-KWVQELQAQGN----PNMVMALAGNKSDLLDA--RKVTAEDAQTYAQENG-LFFMETSAKTATNVKEIFYEIARRLPRVQPTEN------
 AtRABG3e   (92) SAKSFEDLN-NWREEFLIQASPSDPENFPFVVIGNKIDVDGGSSRVVSEKKARAWCASKGNIPYYETSAKVGTNVEDAFLCITTNAMKSGEEEE------
 AtRABH1d   (93) NRLSFLNTS-KWIEEVRNERA----GDVIIVLVGNKTDLVEK--RQVSIEEGDSKGREYG-VMFIETSAKAGFNIKPLFRKIAAALPGMESYSN------
 AtRABA2c   (96) KRQTFDNVL-RWLRELRDHAD----SNIVIMMAGNKSDLNHL--RSVAEEDGQSLAEKEG-LSFLETSALEATNVEKAFQTILGEIYHIISKKALAAQEA
 AtRabA4B  (101) KRETFEHIP-RWLEELRAHAD----KNIVIILIGNKSDLEDQ--RAVPTEDAKEFAEKEG-LFFLETSALNATNVENSFNTLMTQIYNTVNKKNLASEGD
 AtRABB1a   (90) RRETFNHLA-SWLEEARQHAS----ENMTTMLIGNKCDLEDK--RTVSTEEGEQFAREHG-LIFMEASAKTAHNVEEAFVETAATIYKRIQDG---VVDE
Consensus  (101)    SFE L  KWL EL  HA      NMVIMLVGNKSDL D   R VS EEA  FA E G L FMETSAKTATNVED F  IA  I

                201                                234
  AtRAB1b  (182) -ARPPTVQIRGQPVAQ---------KNGCCST--
 AtRABE1c  (188) -AEPATIKISQTDQAAG--AGQATQKSACCGT--
  AtRABC1  (185) ----GSSGGKKNIFKQNPAQTTSTSSSYCCSS--
 AtRABF2a  (180) ----PTGMVLP----NG--PGATAVSSSCCA---
 AtRABF2b  (180) ----PTGMVLP----DR--AMDRAVSSSCCA---
 AtRABG3e  (185) MYLPDTIDVGTS--------N--PQRSTGCEC--
 AtRABH1d  (179) TKNEDMVDVNLKPTSNS---SQGDQQGGACSC--
 AtRABA2c  (188) AAANSAIPGQGTTINVD--DTSGGAKRACCSS--
 AtRabA4B  (193) SNNPGSLAGKKILIPGS--GQEIPAKTSTCCTSS
 AtRABB1a  (179) ANEPGITPG---PFGGK--DASSSQQRRGCCG--
Consensus  (201)     TI V                 SSCCS
```

Figure 2-4. Alignment of selected members of Arabidopsis Rab GTPase proteins indicating regions of conserved amino acid sequence to illustrate 30-55% sequence identity among Rab GTPases for which Rab11 homologs form the largest subgroup in Arabidopsis with 24 members. Rab GTPases typically possess four guanine nucleotide-binding domains (G1, G3, G4 and G5), an effector-binding domain (G2) and a highly conserved YRG domain of unknown function. Alignment conducted using AlignX module of Vector NTI Advance® 11.5.1; key: regions of greatest identity to least—yellow, blue, green

SNAREs. Fusion between transport vesicle and target vacuole membranes permitting deposition of cargo proteins into their final vacuolar destination, resides with SNARE protein functions. SNAREs, much like Rab GTPases, are a diverse group of protein exhibiting specialization and division of labor dependent on the type and destination of cargo. Membrane fusion is achieved through a collaborative effort of multiple subunits Qa-, Qb-, Qc-, and R-SNAREs which together create a functional SNARE complex. These subunits are distinguished from each other by differing amino acids in the core heptad region being either a glutamine residue (Q) or argenine residue (R) (hence the appropriation of "Q" vs. "R" nomenclature). SNAREs can be further categorized based on the membrane with which they associate namely, vesicle membranes (v-SNAREs) or target membranes (t-SNAREs) [141]. A common feature for a functional SNARE complex is the combination of one each of Qa-, Qb-, and Qc-SNAREs that associate with the vesicle membrane (and are therefore, v-SNAREs) and a single target membrane associated R-SNARE (which is therefore a t-SNARE). This combination of subunits is observed among other eukaryotes which illustrates the conservation of SNARE function. The functional diversity of SNARE complexes that gives rise to membrane specificity originates from the particular individual Qa-, Qb-, Qc- and R-SNAREs in the assembled complex [141]. A primary distinguishing feature of land plants is the absence of brevin-type R-SNAREs found in mammals, in favor of longin-type R-SNAREs which have become an expanded family of three groups based on their type of "longin domain" (LD): VAMP7, YKT6, and SEC22 [113]. The LD is a highly conserved structural motif essential to SNARE function [142]. The 12 members of the VAMP7 group are further subdivided based on large scale sequencing and annotation of plant genomes into VAMP71, VAMP72 and VAMP727 [143]. One particular R-SNARE within the VAMP72 subgroup, VAMP727 possesses an "acid loop"

within its LD that is required for Q-SNARE interaction, specifically the components Qa-SYP22/VAM3, Qb-TI11, and Qc-SYP51. Based on localization and mutant studies, this VAMP727/SYP22/VTI11/ SYP51 tetrad is responsible for targeting to, and biogenesis of PSVs by mediating PVC-PSV fusion [105, 142]. Contributing to the function of this tetrad are RabF members (also called Rab5) which mediate traffic to the PSV through association with the acidic LD of VAMP727 [144]. Interestingly, VAMP727 or its homologs are only found in seed plants. This illustrates the importance of this tetrad in SSP trafficking and suggests that the LD motif of VAMP727 was essential to the evolution of flowering plants [142, 144].

2.4 Processing Events of SSP Maturation

The following section provides a general description of the processing events to which nascent peptides are subjected and which culminate in the formation of mature SSP destined for delivery to PSV subcellular compartments. These processes begin in the ER and progress as passage through the Golgi complex and PVC take place.

2.4.1 Chaperones and Foldases

The role of the ER as being critical for correct folding and assembly of proteins has been well established since the discovery of molecular chaperones and their associated machinery. These mechanisms oversee the correct folding and assembly of newly synthesized secretory proteins as well as orchestrate the disposal of misfolded proteins [145, 146]. The importance of these mechanisms in stress response has also been well documented [147].

Entry into the endomembrane system through the ER gateway is initiated by delivery of the nascent polypeptide to a translocation pore (also called a translocon) by a ribosome [148]. As the peptide enters the ER lumen the SP is removed by a signal peptidase located on the luminal surface of the membrane and the polypeptide is received by the first of many chaperone proteins, the lumen binding protein (BiP) [149]. Through association with hydrophobic regions of nascent polypeptides, the BiP molecular chaperone acts as a protein folding mediator by stabilizing intermediate folded states thereby preventing aggregation that could result in misfolding [146, 150]. BiP has also been shown to participate in the assembly of SSP oligomers which is a necessary prerequisite for exit from the ER [151, 152]. BiP is a well characterized ER chaperone known to belong to the HSP70 family and its role in plant stress response has been well documented with much evidence indicating that stress adaptation toward impaired or increased demand in protein folding results in elevated BiP expression [153, 154].

To assist BiP in the correct folding regimen are foldases, enzymes which catalyze rate-limiting steps. Two important examples include protein disulfide isomerase (PDI) and peptidyl-prolyl *cis* or *trans* isomerise (PPI) (immunophilins), that catalyze the formation of disulfide bridges and peptide bonds, respectively. Both enzymes are known to catalyze bond formation adjacent to proline (Pro) residues [147]. It is thought that molecular chaperones BiP and GRP94 (Glucose Regulated Protein or endoplasmin), a member of the HSP90 family, participate in this process. Binding studies indicate that BiP has a temporary affinity for disulfide bond transition state intermediates whereas GRP94 preferentially binds the resulting oxidized disulfide bridges for several minutes [155]. Based on these findings it would seem likely that these chaperones together initiate disulfide bridge formation and "hold in place" constructed bridges to facilitate completed folding of the entire molecule. Once completed, correct folding is ensured by a quality control assessment by calnexin and calreticulin (CRT) machinery which target misfolded proteins for degradation [156, 157].

2.4.2 Glycosylation

Intricately tied to correct folding by chaperones and foldases, glycosylation of cargo proteins by glycosidase and glycosyltransferase enzymes contributes to stability, function and recognition of the nascent proteins. Addition of carbohydrate moieties occurs as either N-glycans, linked to proteins via an amide bond of target asparagine (Asn) residues, or O-linked, attachment via the hydroxyl functional groups of serine (Ser), threonine (Thr), hydroxylysine (Hyl) or hydroxyproline (Hyp) residues [158]. Little is known about O-glycosylation except that O-glycans are assembled one monosaccharide at a time onto the folded protein in the ER and GA and that O-glycan moieties are important for protein folding, stability and cellular recognition[159]. This differs from N-glycosylation which begins in the ER. Soon after the nascent polypeptide enters the ER lumen through the translocation pore, a preformed

$Glc_3Man_9GlcNAc_2$ glycan is transferred from the lipid isoprenyl carrier dolichol to Asn residues of the consensus sequence Asn—X—Thr (where X ≠ proline) of the peptide chain by the heterooligomeric oligosaccharyltransferase (OST) complex [158, 160]. Following this step, numerous removals and additions of carbohydrate groups carried out by glucosidase I (GCSI), glucosidase II (GCSII), a mannosidase and a glycoprotein glucosyltransferase (UGGT) to generate high-mannose N-glycan, $Man_8GlcNAc_2$. Cycles of GCSII and UGGT action continue until proper folding and glycosylation of the proteins are recognized, at which time the proteins are exported from the ER to the GA where conversion from high-mannose N-glycans to complex-type N-glycans occurs. Once in the *cis*-Golgi, the concerted action of mannosidase and N-acetyl-glucosaminyltransferase enzymes removes mannose residues and adds N-acetylglucosamine residues [161]. Although these enzymes have been shown to be ER residents, it is believed that the majority of modifications resulting in conversion to complex-type N-glycans occurs in the GA such that the resulting complex-type N-glyan structures identify the route of cargo proteins through the endomembrane system [22, 24, 27].

The plant glycosylation machinery appears to be highly conserved with most variation attributed to species, developmental stage, and functional state of the cell [23]. In this regard, the 7S globulins of dicots, monocots and gymnosperms are frequently glycosylated predominantly as N-linked glycans, whereas there is little evidence for glycosylation of 2S albumins and the 11S globulins are only rarely glycosylated [4, 21, 162-164]. Further modification of the glycan structure, beyond the synthesis of the initial complex-type N-glycan, results in plant-specific modification typically involving α-1,3-fucosylation and β-1,2-xylosylation and often results in the generation of the Lewis (Le^a) epitope, a specific Gal-β-1,3-(Fuc-α-1,4)-GlcNAc glycan structure important for cell–cell recognition and cell adhesion processes [158, 165]. The Le^a

eptiope is seen abundantly in the seeds of dicots and to a lesser extent in their vegetative tissues whereas monocots have very limited occurrence and only in seeds [166, 167]. Where further modification of the glycan structure does not involve the inclusion of the Le^a epitope, these alterations occur post-Golgi and result in paucimannose-type N-glycans. Because these structures are observed only in vacuolar proteins it is thought that they result from exoglycosidase activity within the PVC or in the vacuole itself [168]. These differences illustrate the complexity of glycan profiles in plants.

2.4.3 Proteolytic processing

In plants, many proteins, including SSP, exit the ER as proproteins having peptide regions that are removed by proteolytic processing enzymes. These cleavable peptide regions often contain C- or N-terminal sorting signals and are thought to be involved in stabilization and folding of the newly synthesized protein [169]. Removal of these peptide regions contributes to the maturation processes of the protein and occurs during transit to their vacuolar destination or after their deposition by specific vacuolar proteases [62]. The contribution of the cysteine vacuolar processing enzymes (VPEs) and the aspartic protease A1 to protein maturation has been shown for several seed proteins including 2S albumins and 11S and 7S globulins from Arabidopsis, pumpkin and castor beans and 11S glutelin from rice [170-173]. For example, the initial translation product of Arabidopsis 2S albumin is a ~21 kDa precursor possessing three propeptide regions: an N-terminal processed peptide that follows the SP, an internal processed peptide, and a C-terminal processed peptide [3, 174, 175]. Aside from removal of the SP in the ER lumen by the signal peptidase of the translocation pore (see section 2.4.1), removal of these propeptides begins in the MVB in which predominantly β-VPE and δ-VPE cysteine proteases as well as aspartic protease A1 and carboxypeptidases are involved [68, 176-183]. Interestingly cysteine VPE and aspartic protease have differing pH optima. Cysteine VPE proteases have

optimal activity at pH 5 to 6 whereas the pH optimum of aspartic proteases is pH 3 to 4. This suggests that protein maturation by removal of propeptide regions occurs gradually as the MVB organelle acidifies during its evolution from PVC to LPVC (see Figure 1-1). This propeptide excision together with phosphorylation renders these proteins PSV-ready. For example, conformational changes in immature 11S cruciferin globulin resulting from selective phosphorylation induces the conversion of proprotein trimers into mature SSP hexamers that possess decreased solubility, reduced susceptibility to further proteolysis and a compact structure compatible for tight packing into PSVs [152, 184-186].

CHAPTER 3
MATERIALS AND METHODS

3.1 Informatics

3.1.1 Sequence Analysis

Selected dicot 2S albumin protein gene accessions were acquired from the National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov/) and from The Arabidopsis Information Resource (TAIR, http://www.Arabidopsis.org/): namely, *Anacardium occidentale* (cashew, AY081853), *Juglans nigra* (black walnut, AY102930), *Arachis hypogaea* (peanut, AY581853), *Arabidopsis thaliana* (water cress, BT002073-At4g27150), *Cucurbita maxima* (pumpkin, Q39649.1), *Gossypium hirsutum* (cotton, AAA33049.1), *Brassica napus* (rape, P01090), *Sesamum indicum* (sesame, DQ256292), *Fagopyrum esculentum* (common buckwheat, DQ304682), *Bertholletia excelsa* (Brazil nut, M80400), *Juglans regia* (English walnut, U66866), *Glycine max* (soybean, U71194), *Ricinus communis* (castor bean, X54158). The 2S albumin sequence alignments were conducted using Vector NTI® version 11.5.1 informatics software employing AlignX® module. Hydropathy plot primary structure predictions were conducted on pumpkin 2S (Q39649.1) and Napin-2 (P01090) proalbumin precursor sequences employing the ExPASy Bioinformatics Resource Portal (http://web.expasy.org/protscale/) using the hydropathy index developed by Kyte and Doolittle [187]. Sequence annotation and functional analysis of these same proteins was reported as shown by the UniProtKB/Swiss-Prot Protein Knowledge Base provided by UniProt (http://www.uniprot.org/).

3.1.2 Secondary Structure Predictions

Prediction of the secondary structure of pumpkin 2S (Q39649.1) and Napin-2 (P01090) proalbumin precursor protein sequence according to the probability of helix and flexible loop

regions was determined by PredictProtein protein sequence analysis server (https://www.predictprotein.org/) and reported using DeltaGraph v6 graphical analysis software (© 2010, Red Rock Software, Inc.).

3.2 Construct Design

3.2.1 PCR-based Cloning

The full-length napin open reading frame was amplified from construct mNap(FL) (Clone No. 695, kindly provided by Dr. Dwayne Hegedus, AAFC, Saskatoon). DNA to be used as template for PCR was isolated using Wizard® Plus SV Minipreps DNA Purification System (Promega, Madison, USA). PCRs were 35 µL reactions prepared in MicroAmp® 0.2 mL capacity Reaction Tube with Cap (Applied Biosystems®, Burlington, Canada) and were comprised of 3.5 µL 10 X High Fidelity PCR Buffer (60 mM Tris-SO_4 (pH 8.9), 18 mM ammonium sulfate, final concentration), 1 µL 10 mM dNTPs (0.28 mM each of dATP, dCTP, dGTP and dTTP final concentration prepared from 100 mM dNTP stocks, Invitrogen), 1.2 µL 50 mM $MgSO_4$ (2 mM final concentration), 0.2 µM forward and reverse primers (Table 3.1, primers M1 and M5), ~100 ηg of *mNap(FL)-GFP* template DNA, 1.0 U High-Fidelity Platinum® *Taq* (Invitrogen, Burlington, Canada) and the balance to 35 µL of autoclaved, distilled water. Reactions were conducted using a 96-well aluminum block GeneAmp® PCR System 9700 thermal cycler (Applied Biosystems®, Burlington, Canada) employing the following protocol: initial denaturation of 94 ºC for 2 minutes, 35 cycles of denature at 94 ºC for 30 seconds, anneal at 55°C for 30 seconds, extend at 68 ºC for 1 minute per kb of PCR product, and final extension at 68 ºC for 7 minutes prior to a hold at 4 ºC. Finished reactions were combined with 1/10th volume 10 X DNA gel loading buffer (0.21% Bromophenol Blue [w/v][Sigma, Oakville, Canada], 0.21% Xylene Cyanol FF [w/v][Sigma, Oakville, Canada], 0.2 M EDTA pH 8.0 [Sigma, Oakville, Canada] and 50% molecular biology grade glycerol [Sigma,

Oakville, Canada] in autoclaved, distilled water) and resolved at 100 V through 0.8% agarose (Sigma, Oakville, Canada) including 1:10,000 GelRed™ nucleic acid stain (Biotium, Hayward, USA) employing a Bio-Rad Mini Sub-Cell GT horizontal agarose gel electrophoresis system (Bio-Rad, Mississauga, Canada). Gel images were acquired using a Bio-Rad Molecular Imager® Gel Doc™ XR+ System with Image Lab™ Software (Bio-Rad, Mississauga, Canada). When appropriate, PCR-generated fragments were excised using a sterile No. 22 scalpel blade (Fisher Scientific, Burlington, Canada) and DNA purified using the Qiagen MinElute Gel Extraction Kit or the Qiagen QIAquick Gel Extraction Kit (Qiagen, Toronto, Canada) according to the manufacturer's instructions. The full-length amplicon was subcloned into *E. coli* vector pBKS (Stratagene, La Jolla, California) by directional cloning methodology employing Invitrogen PstI and XbaI restriction endonucleases and T4 DNA ligase (Invitrogen, Burlington, Canada) according to the manufacturer's instructions. Fidelity of the amplicon was confirmed by sequencing using pBKS annealing M13 forward and reverse primers (Table 3.1, primers U1 and U2) as conducted by the National Research Council DNA Sequencing Services (Saskatoon, Canada). The

confirmed open reading frame was integrated by directional cloning, facilitated by XbaI, SmaI (Invitrogen, Burlington, Canada) and EcoICRI (Promega, Madison, USA) restriction endonucleases, into the pZP121 plant transformation vector to be driven by the CaMV 35S promoter and followed by the NOS terminator [188] (see Table 4.1).

3.2.2 Recombinant PCR for Synthesis of Gene Fusions

Recombinant PCR was employed to facilitate synthesis of the various length napin-PAT fusion constructs. This involved design and use of two sets of primers for each construct suitable for the creation of napin and PAT amplicons sharing a 20 bp overlap (Table 3.1, primers M1-

Table 3-1 Primer Inventory

Key	Name	Sequence (5´-3´)	Construct	ORF	Primer Partner	Procedure[1]
M1	BnNapM-F1	CCT CTA GAC ATG GCA AAC AAA C	mN(FL), M(4)P, M(5)P, M(FL)P, M(B)GFP-CT	mNap(FL)	M2-M5, P2	PCR-DC, rPCR, GT
M2	BnNapM-FR2	CTG GGC TCA TCC TTT GCT GAG GTC C	M(4)P	mNap(pL4)-PAT	M1	rPCR-DC
M3	BnNapM-FR3	CTG GGC TCA TCA CAC AGA GAG GCT C	M(5)P	mNap(PL5)-PAT	M1	rPCR-DC
M4	BnNapM-FR4	CTG GGC TCA TAT AAG AAG GAC CTG G	M(FL)P	mNap(FL)-PAT	M1	rPCR-DC
M5	BnNapM-R5	TTC TGC AGT TAA TAA GAA GGA C	mN(FL)	mNap(FL)	M1	PCR-DC, GT
A1	BnNapA-F1	CCG GAT CCC ATG GCG AAC AAG C	aN(FL), A(4)P, A(5)P, A(FL)P	NapA(FL)	A2-A5, P2	PCR-DC, rPCR
A2	BnNapA-FR2	CTG GGC TCA TTC TCT GCT GTG GAC C	A(4)P	NapA(pL4)-PAT	A1	rPCR-DC
A3	BnNapA-FR3	TAA CGC AAA GGG GCT CTT CCT GGT G	A(5)P	NapA(pL5)-PAT	A1	rPCR-DC
A4	BnNapA-FR4	CTG GGC TCA TGT AGG AGG GCC C	A(FL)P	NapA(FL)-PAT	A1	rPCR-DC
A5	BnNapA-R5	TTC TGC AGC TAG TAG GAG GGC C	aN(FL)	NapA(FL)	A1	PCR-DC
P1	PAT-R2	TTC TGC AGT CAA ATC TCG GTG A	all albumin-PAT fusions	PAT	P2-9, M1, A1, C1	rPCR-DC, GT,RTPCR
P2	PAT-FF1	ACA GCA GAG AAT GAG CCC AGA ACG A	A(4)P	NapA(pL4)-PAT	P1	rPCR-DC
P3	PAT-FF3	GGA AGA GCC CCT TTG CGT TAT GAG C	A(5)P	NapA(pL5)-PAT	P1	rPCR-DC
P4	PAT-FF4	GCC CTC CTA CAT GAG CCC AGA ACG A	A(FL)P, M(FL)P	NapA(FL)-PAT, mNap(FL)-PAT	P1	rPCR-DC
P5	PAT-FF5	TCA GCA AAG GAT GAG CCC AGA ACG A	M(4)P	mNap(pL4)-PAT	P1	rPCR-DC
P6	PAT-FF6	TCT CTG TGT GAT GAG CCC AGA ACG A	M(5)P	PAT	P1	rPCR-DC
P7	PAT-R9		all albumin-PAT fusions	PAT	n/a	SEQ
P8	PAT-R10	TTC TCG AGT CAA ATC TCG GTG A	all albumin-PAT fusions	PAT	P9	GT, RTPCR
P9	PAT-F11	ATG AGC CCA GAA CGA CGC CC	all albumin-PAT fusions	PAT	P8	GT, RTPCR
C1	Cm-F1	GAT CCA TGG CCA GAC TCA CAA GCA	CmP, CmGFP, CmPCT, Cm(FL)	Cm(Δ79)-PAT, Cm(Δ79)-GFP, Cm(Δ79)-PAT-	C2, P1, CT1	rPCR-DC, GT,RTPCR

[1] Procedure: PCR-DC, PCR and directional cloning; rPCR-DC, recombinant PCR and directional cloning; SDM, site-directed mutagenesis; RTPCR, reverse transcriptase PCR; GT, transgenic plant or bacterial clone genotyping by PCR; SEQ, sequencing

				CT, Cm(FL)		
C2	Cm-R2	CTA GAC TCA GAA GTC GCA TCG CTG	Cm(FL)	Cm(FL)	C1	PCR-DC
C3	CmPJ-R				C1, G5	GT, RTPCR
CT1	Cm(79)P-FR3	AGT TCC TAG CAA TCT CGG TGA CGG G	CmPCT	Cm(Δ79)-PAT-CT	C1	rPCR-DC
CT2	CmCT-NOS-FF1	CAC CGA GAT TGC TAG GAA CTT GCC T	CmPCT	Cm(Δ79)-PAT-CT	CT5	rPCR-DC
CT3	NapM-FR6	AGT TCC TAG CAG GGA TCC CGG TAG A	M(B)GFP-CT	mNap(B)-GFP-CT	M1	rPCR-DC
CT4	CT:NOS-FF2	CGG GAT CCC TGC TAG GAA CTT GCC T	M(B)GFP-CT	mNap(B)-GFP-CT	CT5	rPCR-DC
CT5	CT:NOS-R3	CGA ATT CGA TCT AGT AAC ATA GAT G	CmPCT, M(B)GFP-CT	Cm(Δ79)-PAT-CT, mNap(B)-GFP-CT	CT2, CT4, M1	rPCR-DC
GA-1	GA-NapM-F8	TCC CTA AGT GCC GTA AAG AGT TCC	all GA constructs	all Nap2-PAT	GA-2	GT, RTPCR
GA-2	GA-PAT-R14	CCA AGG TCC AGC GTA AGC GAT TCC	all GA constructs	all Nap2-PAT	GA-1	GT, RTPCR
D1	742(X)SDM-F	GGA AGA TCT ACC CTC TAG CCT CTC TGT GTG TG	N(X)GFP	mNap(X)-GFP	D-2	SDM
D2	742(X)SDM-R	CAC ACA CAG AGA GGC TAG AGG GTA GAT CTT CC	N(X)GFP	mNap(X)-GFP	D-1	SDM
D3	749(A)SDM-F	GGA AGA TCT ACC TCC TGA TCT GGA CCT TCT TG	N(A)GFP	mNap(A)-GFP	D-4	SDM
D4	749(A)SDM-R	CAA GAA GGT CCA GAT CAG GAG GTA GAT CTT CC	N(A)GFP	mNap(A)-GFP	D-3	SDM
D5	750(B)SDM-F	TCT ACC GGG ATC CCT TAG TGT AGA AAA GAG	N(B)GFP	mNap(B)-GFP	D-6	SDM
D6	750(B)SDM-R	CTC TTT ACA CTA AGG GAT CCC GGT AGA	N(B)GFP	mNap(B)-GFP	D-5	SDM
D7	751(S)SDM-F	AGG CCT CCT CTT CTT TAA CAA TGT TGT AAC	N(S)GFP	mNap(S)-GFP	D-8	SDM
D8	751(S)SDM-R	GTT ACA ACA TTG TTA AAG AAG AGG AGG CCT	N(S)GFP	mNap(S)-GFP	D-7	SDM
G1	GFP5-F1	AGT GGA GAG GGT GAA GGT GA	all GFP fusions	GFP	G2	GT
G2	GFP5-R2	AAA GGG CAG ATT GTG TGG AC	all GFP fusions	GFP	G1, T1	GT, SEQ
CPT1	CmPTOP-1R	GTG GCG GCC GCA GAT TTA GGT GA	CmP:pHTOP	3´-Cm(Δ79)-PAT	CPT2	GT, SEQ
CPT2	CmPTOP-2F	ACT GGC ATG ACG TGG GTT TCT GG	CmP:pHTOP	Cm(Δ79)-PAT	CPT1	GT, SEQ
CPT3	CmPTOP-3R	TCG CCC TTG CCT GTT TTC CTC CAC	CmP:pHTOP	Cm(Δ79)-PAT	C1	GT, SEQ
CPT4	1358F	GGG AAC CGG AGT TCC CTT	CmP:pHTOP	Cm(Δ79)-PAT	CPT5,	GT, SEQ
CPT5	2511R	AAT TGC CCG GCT TTC TTG TAA CG	CmP:pHTOP	Cm(Δ79)-PAT	CPT3,CPT4	GT, SEQ
G3	GUS-F10	GGG CAG GCC AGC GTA TCG	all pHTOP.A constructs	GUS	G4	GT, RT-PCR, SEQ

G4	GUS-R11	CTT CAC CCG GTT GCC AGA GG	all pHTOP.A constructs	GUS	G3	GT, RT-PCR, SEQ
G5	GUS-R12	CAC TTT TCC CGG CAA TAA CAT ACG	all pHTOP.A constructs	GUS	G3, CAB	GT, RT-PCR
T1	35S-F3	CAA TCC CAC TAT CCT TCG CAA GAC CC	all binary vectors	35S promoter	T3	GT, SEQ
T2	35S-F4	ATC TAC CCG AGC AAT AAT CT	all binary vectors	35S promoter	T3	GT, SEQ
T3	35S-R5	GAG CCA CCT TCC TTT TCC ACT A	all binary vectors	35S promoter	T1, T2	GT, SEQ
CA1	CAB1-F1	GCA AGC TTA TGT CTA GTT GGT TTT ACT CAG	all CAB promoter	CAB1 promoter	CA2	PCR-DC
CA2	CAB1-R2	GCT CTA GAT TGA GGT TGA GTA GTG CAG CAC	all CAB promoter	CAB1 promoter	CA1	PCR-DC
CA3	CAB1-F3	GCC AAT CCA TGA AAC GCA CCT A	CAB:GUS	CAB1 promoter	G4	GT
U1	M13-F	GTA AAA CGA CGG CCA	pUC19-based vectors	n/a	U2	GT, SEQ
U2	M13-R	CAG GAA ACA GCT ATG AC	pUC19-based vectors	n/a	U1	GT, SEQ
CNTR1	ACT2-qPCR-F	CTG TTG ACT ACG AGC AGG AGA TGG A	n/a	ACT2	CNTR2	RTPCR
CNTR2	ACT2-qPCR-R	GAC TTC TGG GCA TCT GAA TCT CTC A	n/a	ACT2	CNTR1	RTPCR
CNTR3	ACT2Q-F	GTC GTA CAA CCG GTA TTG TG	n/a	ACT2	CNTR4	RTPCR
CNTR4	ACT2Q-R	GAG CTG GTC TTT GAG GTT TC	n/a	ACT2	CNTR3	RTPCR
CNTR5	18S-rDNA-F	GGT GGT AAC GGG TGA CGG AGA A	n/a	18S rDNA	CNTR6	RTPCR
CNTR6	18S-rDNA-R	AAG AAC GGC CAT GCA CCA CCA C	n/a	18S rDNA	CNTR5	RTPCR
CNTR7	AtACT3-F	ATG GCC GAT GGT GAG GAC ATT C	n/a	ACT3	CNTR8	RTPCR
CNTR8	AtACT3-R	GGT GCG ACC ACC TTG ATC TTC	n/a	ACT3	CNTR7	RTPCR

M5). These fragments were used in a second round of PCR amplification as template DNA for the complete gene fusion open reading frame (ORF). For the first round reactions, plant transformation vector pGSA1252 and the mNap(FL)-GFP (Clone No. 771) open reading frame (kindly provided by Dr. Dwayne Hegedus, AAFC, Saskatoon) were used as template DNA for PAT and napin amplifications, respectively. Although they were not used in plants, as an alternative to the modified Napin-2 open reading frame, gene fusion constructs with the *Brassica napus napA* gene (Accession J02798) were also created. For these, in house expressed sequence tag (EST) *B. napus* clone EL908 was used as template DNA. PCRs, agarose gel electrophoresis and DNA fragment purifications were conducted as described above (see section 3.2.1) with the following exceptions: template DNAs as described above were used, primers were designed to facilitate directional cloning of synthetic gene fusions (Table 3.1, primers A1-A5), and directional subcloning into vector pBKS was carried out using the restriction endonucleases BamHI, PstI, and XbaI (Invitrogen, Burlington, Canada) according to the manufacturer's instructions. Correct gene fusions were then integrated, by directional cloning of the open reading frames driven by the CaMV 35S promoter or the chlorophyll a/b binding protein (CAB) promoter followed by the NOS terminator, into the pZP121 plant transformation vector [188] (see Tables 4.1 and 4.2).

To assess the potential of the C-terminus of the pumpkin 2S albumin as a sorting determinant, three open reading frames were created in which the coding region of the pumpkin 18 amino acid C-terminus was fused in frame to the C-termini of the open reading frames for pumpkin 2S(Δ79)-PAT, mNapin2(pL4)-PAT and mNapin2(BamHI)-GFP to create open reading frames Cm(Δ79)-PAT-CT, mNap(pL4)-PAT-CT and mNap(B)-GFP-CT. These were prepared by the recombinant PCR methodology described above except that the Cm(Δ79) peptide

fragment and the pumpkin 18 amino acid CT region were derived from GeneArt synthesized constructs as described in section 3.1.4 below and primers were designed to facilitate creation of gene fusions (Table 3.1, primers C1, CT1 – CT5)

3.2.3 Site-directed Mutagenesis for Synthesis of Napin-GFP Fusions.

Four codon modified Napin-2 (mNap) constructs possessing green fluorescent protein (GFP) insertions at sites XhoI, AccI, BamHI, and StuI within the napin open reading frame (ORF) were kindly provided by Dr. Dwayne Hegedus (Clone No. 742, 749, 750 and 751, respectively). By exploiting genetic code redundancy, the mNap synthetic sequence differed from the original Napin-2 sequence only in that the codon usage was changed to facilitate RNA interference repression. For my purpose, these constructs were further modified using site directed mutagenesis (SDM) to introduce stop codons at the end of the GFP coding region. A fifth construct, received from Dr. Hegedus which consisted of GFP fused in frame to the carboxyl terminus of the mNap ORF was used without modification (Clone No. 771). Isolation of plasmid DNA was conducted as in section 3.2.1 and SDM was facilitated by the use of the QuikChange® Site-Directed

Mutagenesis Kit (Stratagene, Mississauga, Canada) according to the manufacturer's instructions with the following additions: 10 ηg and 50 ηg template DNA originating from clones 742, 749, 750 and 751 as described above were used, primers were designed to introduce a stop codon into the synthetic gene fusions (Table 3.1, primers D1-D8), PCR was conducted using 12 cycles with a 68 °C 5 minute extension time, and 3 μL of DpnI treated reaction product was used to transform *E. coli* XL1-Blue supercompetent cells. DNA from each of five clones produced from each SDM reaction were isolated and sequenced to confirm introduction of the stop mutation. Once the open reading frames were confirmed, they were introduced into the plant

transformation vector pMDC32 using the Gateway® LR Clonase® Enzyme Mix (Invitrogen, Burlington, Canada) according to the manufacturer's instructions.

3.2.4 Construct Synthesis by GeneArt

Gene fusion constructs comprised of open reading frames custom synthesized by GeneArt® Gene Synthesis services included those denoted "GA" in Tables 4.1 and 4.2 as well as a pumpkin 18 amino acid carboxyl terminal-NOS terminator fusion that was used in the three CT gene fusions (see 3.2.1). This procedure involved *in silico* design of gene fusion open reading frames and the submission of that sequence to GeneArt® (www.geneart.com). These open reading frame sequences were optimized according to the GeneArt® sequence optimization algorithm which selects for preferred codon usage for expression in *Arabidopsis thaliana* and for optimal GC content to ensure mRNA secondary structure stability. Once the gene fusions were received, their fidelity was confirmed by sequencing by National Research Council Sequencing Services (Saskatoon, Canada) and integrated by directional cloning into plant transformation vector pZP121 for CaMV 35S driven expression.

To confirm that codon usage had no effect on sorting behavior, the napin-PAT fusion constructs were also synthesized by GeneArt employing a codon optimization algorithm for expression in *Arabidopsis thaliana* and also created using the *Brassica napus napA* open reading frame (GeneBank: J02798.1) by rPCR. However, this latter group was not pursued beyond construct synthesis when evidence provided by the GeneArt napin-PAT fusion constructs indicated that codon usage did not affect sorting behavior.

3.3 Plant Material and Growing Conditions

Experimental plant material employed for this work was exclusively *Arabidopsis thaliana* var. Columbia. Plants were sown in Redi-Earth B soilless mix (W. R. Grace & Co., New York) in 2" wells of 36-well flats and grown under growth chamber or green house

conditions employing a 16-hour light period at 20˚C followed by an 8-hour dark period at 17˚C. All Purpose 20-20-20 fertilizer at 0.3% (w/v) (Early's Farm and Garden Centre, Saskatoon) was applied as needed.

3.4 Generation of Transgenic *Arabidopsis thaliana*

3.4.1 *Agrobacterium*-mediated Transformation

Gene fusion constructs validated by sequencing were used to transform *Agrobacterium tumefaciens* GV3101 pMP90 by combining ~50 ηg of construct DNA with 50 µL of electrocompetent cells in 10% glycerol in a 1.0 mm gap electroporation cuvette which received an electric pulse (2.5 $KVcm^{-1}$ field strength, 25 µF capacitance, 600 Ω resistance) using a Gene Pulser® Xcell Electroporation System (Bio-Rad, Mississauga, Canada). Immediately after applying the current, 1.0 mL LB or SOC broth (Appendix 1 Compositions of Microbiological and Plant Media) was added and cells were allowed to recover at 28 ˚C for 2 hours before 10 µL and 100 µL were cultured at 28 ˚C for 48 hours on LB medium (Appendix 1 containing antibiotic selection agents (Sigma, Oakville, Canada) containing 10 µg/mL rifampicin to select for strain GV3101, 25 µg/mL gentamycin sulfate to select for the pMP90 helper plasmid, and either 34 µg/mL chloramphenicol for pZP121 binary vector or 30 µg/mL kanamycin sulfate for pMDC32 binary vector. Five well-separated clones grown on selective media were chosen and confirmed by colony PCR that involved combining a small portion of bacterial colony in 30 µL autoclaved, distilled water with the remaining colony being used to inoculate a replica plate. The colony slurry for each clone was boiled for 5 minutes in a boiling water bath followed by 5 minutes on ice and 5 minutes centrifugation at 20,000 g using an Eppendorf model 5417R microcentrifuge (Eppendorf, Mississauga, Canada). Supernatants were combined with 0.2 µM forward and reverse gene-specific primers (Table 3.1), 3.5 µL 10X PCR buffer, 1.5 mM $MgCl_2$, 0.2 mM dNTPs and 1 U of *Taq* DNA polymerase (Invitrogen, Burlington, Canada) to a total

volume of 35 µL. PCRs were optimized according to primer annealing temperature specifications and the regimen for thermal cycler protocol. Analysis of reaction products was as described in section 3.2.1 above. Positive clones were subcultured by inoculating 2.0 mL LB broth containing rifampicin (10 µg/mL) and kanamycin (30 µg/mL) and incubated at 28˚for two days. From these seed cultures, 10 µL were used to inoculate 200 mL LB broth containing the same concentration of antibiotics. These 200 mL cultures were incubated with 200 rpm shaking at 28 ˚C until an OD_{600} of 0.8 was reached using a BioMate 3S UV-Visible spectrophotometer (Thermo Scientific, Burlington, Canada). From these cultures, 1.0 µL was used to perform a second colony PCR as described above to confirm the presence of the intact transgene. Once fidelity of the transgene was confirmed by PCR, these 200 mL cultures were used to transform Arabidopsis plants that were approximately 6 weeks of age and had been trimmed of developed sliques. Arabidopsis plants were typically grown 6-7 plants per 2" pot and had their primary bolts cut to encourage lateral shoots and increased flowering at least once prior to transformation. Floral dip transformation [189] involved centrifugation of 200 mL *Agrobacterium* cultures at 5000 rpm for 20 min using a Sorvall Model RC-6+ centrifuge with a Fiberlite® F14-6x250y Rotor GSA rotor (Sorvall, Burlington, Canada). *Agrobacterium* cell pellets were re-suspended into a slurry of 0.5% Silwet® L-77 (Loveland Industries, Greenley, Colorado), 5% table sugar (w/v) and tap water to 1 L total volume. Arabidopsis plants were inverted in this slurry for 30 seconds and then exposed to 15 mm Hg vacuum applied to a desiccation chamber (Thermo Scientific, Burlington, Canada) using a Savant model GP110 Gel Pump vacuum (Fisher Scientific, Burlington, Canada) for 1 minute before being returned to the green house to set seed.

3.4.2 Recovery and Establishment of Transgenic Arabidopsis Populations

Seeds harvested from Arabidopsis plants transformed by the floral-dip method described above were surface sterilized using 30% bleach (v/v), suspended in sterile 0.5% Bacto® agar (BD Biosciences, Mississauga, Ontario) and vernalized at 4 ˚C for 5 days. Following this period, seed-agar slurries were poured evenly onto ½ MS-S plates (Appendix 1) containing specific selectable marker antibiotics. For constructs in which transgenes were integrated into the Arabidopsis genome using the pZP121 plant transformation vector, 50 µg/mL gentamycin sulfate (Sigma Cat. No. G1264) was used; whereas for the pMDC32 plant transformation vector, 15 µg/mL hygromycin B plus 25 µg/mL kanamycin sulfate (Sigma, Oakville, Canada) were used. Plates were allowed to dry, wrapped with a strip of Parafilm® M laboratory film (Fisher Scientific, Burlington, Canada), and incubated for 7 to 10 days at 20 ˚C, 16-hour day growth conditions. Germinated seedlings were transferred by sterile technique to new plates containing the same medium and allowed to grow until roots were established at which time they were transferred to water-saturated Redi-Earth B soilless mix in 2" pots, and covered tightly with plastic wrap to create a high humidity environment. Over the course of several days, perforations in the plastic were made to gradually acclimatize the growing plants to greenhouse humidity. Cling wrap covers were ultimately removed and these plants were allowed to grow to maturity and set seed. To confirm that these plants were T_1 transgenic, small 5 mm diameter discs of rosette leaf tissue from each plant was aseptically transferred to a sterile microcentrifuge tube and homogenized with 0.5 mL Rapid Plant DNA Extraction Buffer (Appendix 2 Composition of Extraction Solutions and Buffers) using a micropestle (Eppendorf, Mississauga, Canada) before centrifugation at 20,000 g for 5 minutes using a Eppendorf model 5417R microcentrifuge (Eppendorf, Mississauga, Canada). From this, 300 µL was transferred to a new sterile microcentrifuge tube, combined with 300 µL room temperature (RT) isopropanol (Sigma,

Oakville Canada) and mixed vigorously using a VWR Analog Vortex Mixer (VWR-Canlab, Mississauga, Canada). Contents were allowed to incubate at RT for 5 minutes before centrifugation at 20,000 g for 5 minutes. Isopropanol-buffer supernatants were discarded and the pellets were washed with RT 70% ethanol prepared from autoclaved distilled water and 95% ethanol (Commercial Alcohols, Brampton, Canada). Ethanol-pellet mixtures were agitated for 30 s using a vortex mixer and subjected to 20,000 g centrifugation for 5 minutes before supernatants were discarded and pellets were dried using a SpeedVac Concentrator (Thermo Scientific, Burlington, Canada) for 2 minutes. White powdery DNA clinging to the taper of microcentrifuge tube was collected in 30 μL autoclaved distilled water and used as DNA template in PCRs using gene-specific primers as described for colony PCR (see 3.4.1). The seed from PCR confirmed transgene positive T_1 plants was harvested and cataloged.

3.5 Monitoring Expression in Transgenic Progeny

3.5.1 PCR and RT-PCR Methodology

Transgenic T_1 Arabidopsis established as described in section 3.4.2 were advanced to T_2, T_3 and T_4 generation through sowing seed of confirmed PCR-positive parents and monitoring their progeny by PCR as described in section 3.4.2 and RT-PCR methodology. To facilitate mRNA extraction, 3-4 rosette leaves (approximately 500 mg tissue) were taken from each plant of 4 to 5 weeks of age, ground to a fine powder using liquid N_2 and a porcelain mortar and pestle, and aliquotted to ~100 mg per microcentrifuge tube. mRNA was extracted from ~100 mg rosette leaf tissue with Qiagen's RNeasy Plant Mini Kit (Qiagen, Toronto, Canada) and purified from contaminating DNA by two incubations per sample with DNase using RNase-Free DNase Set (Qiagen, Toronto, Canada) according to the manufacturer's instruction for on-column and subsequent post-column DNase treatment. Following extraction and DNase treatment, mRNA samples were checked for quantity and quality using a NanoDrop spectrophotometer (Thermo

Scientific, Burlington, Canada) and 0.8% agarose gel electrophoresis as described in section 3.2.1. RT-PCR was conducted using Invitrogen's SuperScript III One-Step RT-PCR System according to the manufacturer's instructions (Invitrogen, Burlington, Canada). For this analysis, 1.0 μg mRNA and open reading frame-specific primers (Table 3.1) were used. Each set of RT-PCR amplifications included one additional reaction per mRNA sample to monitor for DNA contamination, where the SuperScript RT/ High Fidelity Platinum® *Taq* enzyme mix was replaced with High Fidelity Platinum® *Taq* enzyme. Reaction products were resolved using 0.8% agarose gel electrophoresis and documented as described in section 3.2.1.

3.5.2 Western Blot Analysis

Western blot analysis for examination of leaf protein contents for T_2, T_3 and T_4 transgenic plants involved rosette leaf tissue, harvested from plants of 4-5 weeks of age and homogenized using liquid N_2 as described in section 3.5.1. For each plant ~100 mg ground tissue was combined with 500 μL of 4 ˚C Thiourea-Urea Protein Extraction Buffer (Appendix 2) and mixed by vigorous agitation using a micropestle. Homogenates were subjected to centrifugation at 8,600 g using an Eppendorf 5417 R (Eppendorf, Mississauga, Canada) refrigerated microcentrifuge held at 4 ˚C for 20 minutes. For each sample, 200 μL supernatant was transferred to a new sterile microcentrifuge tube and assayed for total protein content using a Bradford Protein Assay Kit (Bio-Rad, Mississauga, Canada) and employing bovine serum albumin as a protein standard (Sigma, Canada). Subsequent to protein determination of the extracts, 10 μg of protein was combined with 6 μL of 5X Laemmli Sample Buffer (Appendix 2) and autoclaved distilled water to 30 μL total volume. Contents were then incubated in a boiling water bath for 5 minutes and centrifuged at RT for 5 minutes at 20,800 g before loading into individual wells of a 4% acryalmide stacking/15% acrylamide resolving gel and resolved at 150 volts for approximately 2 hours using Bio-Rad's Mini-PROTEIN Electrophoresis system (Bio-

Rad, Mississauga, Canada) along with 7.5 μL Fermentas PageRuler™ Prestained Protein Ladder (10 – 170 kDa) (Thermo Scientific, Burlington, Canada). Following electrophoresis, resolving gels were transferred to Immun-Blot® PVDF Membrane (Bio-Rad, Mississauga, Canada) using a Mini Trans-Blot® Electrophoretic Transfer Cell apparatus (Bio-Rad, Mississauga, Canada) or silver stained using Invitrogen's SilverQuest™ Silver Staining Kit according to the manufacturer's instructions (Invitrogen, Burlington, Canada). Routine immunodetection of PAT and GFP fusion proteins was carried out with rabbit polyclonal anti-PAT antiserum (Sigma, Oakville, Canada) and rabbit polyclonal anti-GFP antiserum (Invitrogen, Burlington, Canada) at dilutions 1:5,000 to 1:10,000 in 5% skim milk (w/v) in TBST buffer (see Appendix 2). Chemiluminescent detection of bound primary antiserum was made possible with a goat anti-rabbit-HRP conjugate antibody (Bio-Rad, Mississauga, Canada) used at 1:25,000 dilution in 5% skim milk (w/v) in TBST buffer, Millipore's Immobilon™ Western Chemiluminescent HRP Substrate (Millipore, Billerica, Massachusetts) and Kodak® BioMax® light x-ray film (Bio-Rad, Mississauga, Canada) used according to the manufacturer's instructions.

3.5.3 Assay for PAT Activity—Resistance to PPT

Seeds harvested from PAT gene fusion transgenic Arabidopsis were examined for resistance to phosphinothricin phenotype (PPT^R) by first surface sterilizing approximately 50 seeds with 30% (v/v) bleach in 0.5% Tween 20 (Fisher Scientific, Burlington, Canada), suspending them in 0.1% Bacto® agar (BD Biosciences, Mississauga, Ontario) and vernalization at 4˚C for 5 days. Vernalized seeds were then even distributed across 100 mm diameter petri dishes (Fisher Scientific, Burlington, Canada) containing ½ MS medium (Appendix 1) and 0, 20, or 100 μg/mL DL-phosphinothricin (Sigma, Oakville, Canada). Plates were allowed to air dry in a sterile flow hood, sealed with Parafilm® laboratory film, incubated at 20 ˚C with 16 hour daylight conditions and monitored for several days. Presence of PPT^R phenotype among

transgenic seedlings was evident by 14 days at which time the plates were photographed using a Canon Rebel XS digital camera (Canon, Canada).

3.5.4 Tobacco Transient Expression and Microscopy of GFP Transgenic Plants

Prior to infiltration, an *Agrobacterium* GV3101 pMP90 culture harboring the appropriate plant transformation binary vector was initiated by inoculation of 2 mL LB broth containing 10 μg/mL rifampicin, 25 μg/mL gentamycin and antibiotics for bacterial selection of the harbored plant transformation vector. This culture was incubated at 28 ˚C with agitation for approximately 30 hours until OD_{600} was between 0.8 and 1.0. From this, 1 mL was transferred to a microcentrifuge tube and centrifuged at 2200 g for 5 minutes at RT. The *Agrobacterium* pellet was washed twice with 0.5 mL infiltration buffer (Appendix 1) before re-suspension in infiltration buffer sufficient to obtain OD_{600} between 0.8 and 1.0. To conduct infiltration of tobacco leaf tissue, this bacterial suspension was transferred to a 1.0 mL sterile Luer lock-style syringe (Fisher Scientific, Burlington, Canada) and pressed to the underside of a young tobacco leaf. By applying steady pressure, the *Agrobacterium* suspension was forced into the leaf mesophyl layers through open stomata. Following infiltration, the area of the affected leaf was delineated with an indelible pen. Following 48 hours incubation in green house conditions, tobacco plants were subjected to 16 hours darkness before examination by microscopy.

Microscopic examination of GFP transgene expressing tissue was carried out with a Zeiss Axio Imager.Z1 Apotome microscope equipped with an X-Cite series UV light source, a 232 power supply, and an AxioCam HRm digital camera. Images were acquired using a 40X water immersion objective configured with MTP2004 Configuration Utility to accommodate GFP fluorescence (450-490 nm excitation, 495 nm beam split, 500-550 nm emission) and analyzed with AxioVision 4.8 software (Zeiss, Toronto, Canada).

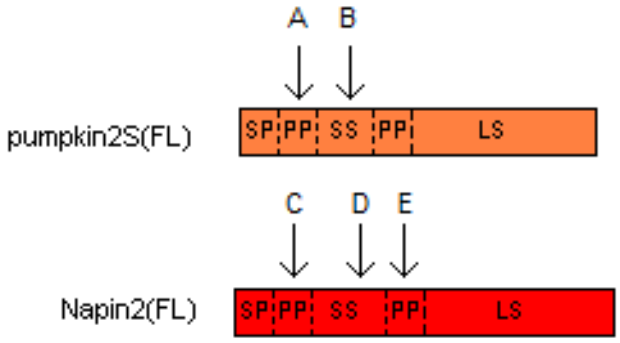
3.6 Anti-peptide Antibody Development

3.6.1 Peptide Design and Synthesis

Antisera raised against peptide regions were developed for use in the characterization of leaf-accumulating fusion protein. A suite of antibodies directed against the small subunit and processed peptide regions of pumpkin 2S and napin albumins were designed to permit identification of internal peptide processing of fusion proteins (Figure 3-1). To facilitate identification of PAC-like vesicles and vegetative PSVs, two peptide regions somewhat unique to vacuolar membrane protein α-tonoplast intrinsic protein (α-TIP) were identified by TIP protein sequence alignments and synthesized (regions “a” and “b” in Figure 3-2). Peptides were custom synthesized by the McGill Peptide Facility (Sheldon Biotechnology Centre, McGill University, Canada). The α-TIP peptide region 250-259 (region “b”) was re-synthesized by EZBiolabs Custom Antibody Synthesis Service (EZBiolabs Inc., Carmel, Indiana USA).

3.6.2 Development of Antiserum

Custom synthesized peptides were obtained from the Sheldon Biotechnology Centre as lyophilized peptides and were suspended at 4 mg/mL in Imject® EDC Conjugation Buffer and used immediately for conjugation to mariculture keyhole limpet hemocyanin (mcKLH) using the Pierce Imject® Immunogen EDC Kit with mcKLH and BSA (Thermo Scientific, Burlington, Canada) according to the manufacturer’s instructions. Polyclonal antibodies reactive towards KLH-conjugated peptides were raised by immunization of New Zealand White rabbits by intramuscular injection with 0.3 mg of KLH-peptide in 0.5 mL of Freund’s complete adjuvant, followed by subsequent immunizations using Freund’s incomplete adjuvant 2 and 6 weeks later as conducted by the Western College of Veterinary Medicine (University of Saskatchewan, Saskatoon, Canada).



Antibody Name	Peptide Location	Size (aa)	Region	Antibody Target
A-CmPP	A	12	23-34	Pumpkin 2S albumin processed peptide
A-CmSS	B	12	51-62	Pumpkin 2S albumin small subunit
A-NapPP1	C	16	22-37	Napin processed peptide
A-NapSS	D	15	46-60	Napin small subunit
A-NapPP2	E	15	77-91	Napin processed peptide

Figure 3-1. The suite of custom anti-peptide antibodies created for western blot detection of pumpkin 2S and napin albumin potentially subjected to peptide processing. Key: aa, amino acid; SP, signal peptide; PP, processed peptide; SS, small subunit; LS, large subunit

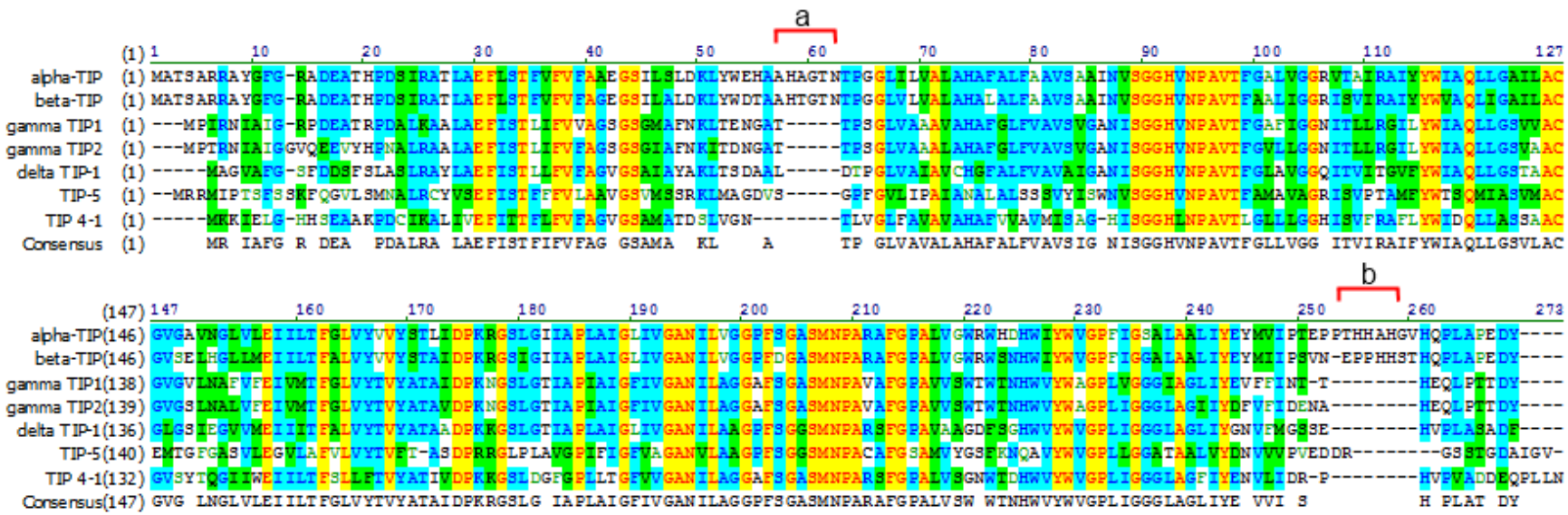


Figure 3-2. Protein sequence alignment of all known Arabidopsis tonoplast intrinsic proteins (TIP) employing AlignX module of Vector NTI Advance® 11.5.1 showing regions “a” and “b”; key: regions of greatest identity to least—yellow, blue, green.

3.7 Vesicle Enrichment

3.7.1 Differential and Gradient Ultracentrifugation

Isolation of vesicle fractions was accomplished following the method of Shimada et al. (2002) [70] which involved combining ~100 mg of liquid N_2 ground rosette leaf tissue with 500 µL of Vesicle Extraction Buffer A (Appendix 2) and mixing vigorously with a micropestle. Homogenates were subjected to centrifugation at 3000 g and 4˚C for 20 minutes. From this, 250 µL of supernatants were transferred to a new microcentrifuge tube and centrifuged at 8,000 g for 20 minutes at 4˚C which resulted in a small delicate vesicle fraction pellet. The supernatants from this step were carefully removed and retained whereas the pellets were suspended in 50 µL Vesicle Suspension Buffer (Appendix 2). These vesicle suspensions were routinely quantified for protein content by Bradford protein determination assay and analyzed by western blot (see section 3.5.2) or used in vesicle isolation by antibody capture (see section 3.7.2). When required, further purification was achieved by layering the vesicle suspensions on 28% Percoll® self-forming gradient solution (GE Healthcare Life Sciences, Baie d'Urfe, Canada) in Beckman-Coulter UltraClear™ tubes fitted with 16 mm titanium caps and ultracentrifuged at 40,000 g for 35 minutes at 4˚C using a Beckman Coulter L8-70M Refrigerated Ultracentrifuge and Type 80 Ti fixed angle rotor (Beckman-Coulter, Mississauga, Canada). Following ultracentrifugation, 0.5 µL aliquots were removed from the gradients from top to bottom and analyzed by western blot. The density of each aliquot was estimated based on the migration of Density Marker Beads (GE Healthcare Life Sciences, Baie d'Urfe, Canada).

3.7.2 Antibody Capture

To facilitate vesicle isolation through antibody binding to vacuolar membrane associated proteins, Dynabeads® M-450 Tosylactivated beads (Invitrogen, Burlington, Canada) were conjugated to 200 µg rabbit polyclonal AtRabA4b antibody (Sigma, Oakville, Canada) or rabbit

polyclonal anti-αTIP antibody (design and synthesis described in section 3.6). Coupling reactions were conducted according to the manufacturer's instructions whereby 5 μg of antibody protein per 10^7 beads was combined in Antibody Coupling Buffer (Appendix 2) and incubated for 24 hours at 37 ˚C with constant gentle agitation by tube inversion using a Mini LabRoller™ rotator (Sigma, Oakville, Canada). Following incubation, Dynabead-antibody conjugates were recovered using a DynaMag™ Spin Magnet (Invitrogen, Burlington, Canada) and washed twice with 1 mL Antibody Capture Buffer (Appendix 2) before storage in 0.02% sodium azide (Sigma, Oakville, Canada) at 4˚C. For capture of vesicles, 10 μg protein of vesicle fractions (obtained as described in section 3.7.1) were combined with 25 μL Dynabead-antibody conjugates in 225 μL Antibody Capture Buffer and incubated for 20 minutes at 4 ˚C before washing four times using 1 mL Antibody Capture Buffer. Bound vesicles were eluted with 50 μL of 50 mM glycine (pH 2.8). Eluate was split evenly and used for western blot and silver staining SDS-PAGE analysis as described in section 3.5.2. It was noted, based on quantity of proteins recovered in the final eluates that this protocol required optimization of coupling conditions and incubation times, temperatures and/or buffer compositions, as well as use of Dynabeads more suited to recovery of protein complexes, such as Dynabeads® M-270 (Invitrogen, Burlington, Canada) which are now available.

3.8 Gene Activation by Dexamethasone Induction

3.8.1 Construct synthesis and Generation of Double-transgenic Arabidopsis

For the purpose of studying the effects of inducible expression of albumin-PAT gene fusions, a suite of constructs were made in which the pumpkin 2S(Δ79)-PAT, mNap(pL4)-PAT and PAT (no SP) open reading frames were integrated into the pHTOP binary vector, the target of the dexamethasone activated LhGR-N transcription factor fusion (see Figure 4-25). Development of double transgenic lines reactive to dexamethasone treatment required use of

these pHTOP constructs as well as the empty pHTOP vector in *Agrobacterium*-mediated transformation of LhGR-N(4C-S5) Arabidopsis plants. The pHTOP-A construct and the LhGR-N(4C-S5) Arabidopsis plant line were kindly provided by Dr. Ian Moore, Department of Plant Sciences, University of Oxford (South Parks Road, Oxford, OX1 3RB, UK). Detailed aspects of directional cloning facilitated by standard molecular biology techniques and *Agrobacterium*-mediated transformation of Arabidopsis have been described previously (see sections 3.2 and 3.3).

3.8.2 Dexamethasone Treatment and Induction Monitoring

The double transformed lines created as described in section 3.8.1 were allowed to set seed, and this harvested T_1 seed was screened by selection on ½ MS medium (Appendix 1) containing 15 μg/mL hygromycin B plus 30 μg/mL kanamycin sulfate. Surviving T_2 progeny were advanced to the T_3 generation, confirmed positive by PCR and examined for response to dexamethasone. Dexamethasone induction of seedlings was achieved by cultivation on ½ MS medium containing 10 μM dexamethasone plus 20 μg/mL PPT. Monitoring and documentation of a PPT^R phenotype was conducted as described in section 3.5.3. For dexamethasone induction of mature plants, Arabidopsis seed from selected plants (Table 3.2) were sown in Redi-Earth B soilless mix (W. R. Grace & Co.-Conn., New York) in 2" wells of 36-well inserts (one plant per well) placed into trays without drainage holes and grown under 20 ˚C, 16 hour daylight conditions. Plants of 4 week age were watered with 300 mL 20 μM dexamethasone in tap water on day 0, 2, 4 and 6 by flooding the tray. On each day 0 through 7, prior to watering, one rosette leaf (~1 cm × 2.5 cm in size) from each of ten randomly chosen plants were harvested. Three of these leaves were used for GUS staining analysis while the remaining seven were pooled for western analysis and mRNA extraction. mRNA extraction and RT-PCR methodology using

transgene or GUS reporter gene specific primers and western blot analysis using anti-PAT antiserum was conducted as described in section 3.5.1 and 3.5.2, respectively.

β-Glucuronidase (GUS) activity was assessed using a GUS assay whereby intact rosette leaves were individually placed in wells of a 12-well cell culture plate (BD Falcon™, BD Biosciences, Mississauga, Ontario) and immersed in 200 µL X-GLUC stain comprised of 0.5 mg/mL 5-Bromo-4-chloro-3-indolyl β-D-glucuronide cyclohexylammonium salt (X-GLUC, Sigma, Oakville, Canada) in X-GLUC Buffer (Appendix 2). Leaf tissue was infiltrated with X-GLUC staining solution by application of a 15 mm Hg vacuum for 60 minutes. Following infiltration, culture plates were kept dark by wrapping in aluminum foil and incubated at 37 ˚C for 24 to 72 hours. To visualize GUS staining more clearly, staining solution was removed by aspiration and replaced with 200 µL 95% ethanol and refrigerated overnight. Ethanol washing and refrigeration was repeated for several days if necessary to clear the tissue of chlorophyll pigmentation. GUS staining results were documented using a Canon Rebel XS digital camera.

3.8.3 Microarray Target Preparation and Array Hybridization

Each flat of 36 transgenic plants was treated as a single biological replicate, from which seven rosette leaves (4-5 weeks of age) were harvested and pooled daily. These pooled leaf samples were used for total RNA extraction and used in duplicate, triplicate and quadruplicate for amplified RNA (aRNA) synthesis reactions (technical replicates). To synthesize aRNA Affymetrix GeneChip® 3‘-IVT Express Kit (Affymetrix, Santa Clara, California) was employed, whereby 100 ηg of total RNA samples were reverse transcribed into double-stranded cDNA followed by *in vitro* biotin labeling according to the manufacturer’s instructions and the recommendations for use of AGRONOMICS1 transcriptome profiling arrays (http://www.agron-omics.eu/). For the pilot study, total RNA samples for 0, 1, 2, and 3 day inductions were used in duplicate for synthesis and hybridization of aRNA, whereas for the large-scale study, total RNA

Table 3-2. T_3 generation transgenic Arabidopsis parents used in large-scale microarray study[1]

T_1 CmP:pHTOP-3	T_1 CmP:pHTOP-4	T_1 CmP:pHTOP-5	T_1 CmP:pHTOP-6	T_1 CmP:pHTOP-7	T_1 PAT:PHTOP-3	T_1 CmP
3.2.4	4.3.1	5.1.1	6.1.1	7.2.1	3.1.1	14.1.3
3.2.6	4.3.2	5.1.2	6.1.2	7.2.3		18.1.2
3.2.7	4.3.3	5.1.3	6.1.4	7.2.6		

[1] Arabidopsis parent lines are annotated such that their predecessor is described according to the succession $T_1.T_2.T_3$; for example plant CmP:pHTOP-3.2.4 indicates T_3 generation plant 4 which originated from T_2 generation plant 2, which in turn originated from T_1 generation plant 3. Column headings indicate the construct used to generate transgenic plant lines.

samples for day 0, 1-, 2-, and 3-day incubations were used in triplicate for T_4 generation CmP:pHTOP-3.2.4, CmP:pHTOP-3.2.6 and PAT:pHTOP-3.1.1 plants and in quadruplicate for T_4 generation CmP-14.1.3, CmP-18.1.2 and wild-type plants. From the yield of aRNA, 15 µg was fragmented at 94 ˚C for 35 minutes in Fragmentation Buffer and hybridized for 16 hours at 45 ˚C to AGRONOMICS1 custom arrays (Affymetrix Cat. No. 520654) employing the Affymetrix GeneChip® Hybridization Wash and Stain Kit and the Hybridization Controls Kit (Affymetrix, Santa Clara, California). Arrays were washed in an Affymetrix Fluidics Station 450 using the FS450_0004 protocol and scanned with an Affymetrix GeneChip® Scanner 3000. Because a limited number of hybridizations could be conducted per day, a hybridization regimen was utilized whereby eight hybridizations to AGRONOMICS1 arrays were conducted daily such that each group of triplicate hybridizations were conducted together and quadruplicate replications were split over two days.

3.8.4 Data Analysis Employing GeneSpring GX12

The gene expression resulting from dexamethasone induction was monitored by GeneSpring GX12 software (www.agilent.com) employing Atdschipb520654 library files designed for AGRONOMICS1 custom array. Dexamethasone-responsive genes were identified based on both significance (ANOVA p-value < 0.05) and fold change (≥ 2). For the pilot study, pair-wise comparisons using a T-test to identify significant ($p<0.05$) fold changes were made between the non-induced day 0 control and day 1, 4 and 7. For the large-scale microarray study, pair-wise comparisons of gene expression profiles were made between the PAT:pHTOP-3.1.1 control and the expression profiles of for each time, day 0, 1, 2 and 3. In this way, the expression profiles of CmP:pHTOP-3.2.4 and CmP:pHTOP-3.2.6 were weighted against the corresponding dexamethasone-responsive genes attributed to PAT and GUS reporter expression alone. To be considered valid, candidate genes had to pass two criteria: 1) candidate genes would be

disqualified if they were also represented in their day 0 comparison to PAT:pHTOP-3.1.1, i.e. if they showed a change in gene expression (either an increase or decrease) prior to administering dexamethasone, and 2) candidate genes would be disqualified if their gene expression fold-changes were not in the top 50 observed at each time point day 1 through day 3 as they were considered of less consequence. Annotation of candidate genes was done according to that used by The Arabidopsis Information Resource (www.Arabidopsis.org).

CHAPTER 4
RESULTS

4.1 Ectopic Expression of Pumpkin 2S and Napin Albumin Variants

The goal of this thesis research was to express albumin construct variants in leaf tissue that would result in PAC-like vesicle biogenesis and accumulation and thereby emulate the work of Hayashi et al**.** [19] for the purpose of understanding these processes. At the outset it was important to reproduce the findings of Hayashi and thereby confirm that generation of PAC-like vesicles in the leaves of Arabidopsis were not anomalous artifacts seen only in their laboratory. In conjunction with this, analogous napin gene fusion variants were created with the goal of determining if SSP from *B. napus* could also be employed to direct protein targeting and evoke vesicle biogenesis. It was anticipated that the use of varying lengths of napin in gene fusion constructs, would permit discovery of essential sorting signals within this protein. To facilitate construct design, the pumpkin 2S albumin and *B. napus* napin were examined using bioinformatics tools to assess which features are shared among these and other similar albumins and which features might be essential to evoke PAC-like vesicle biogenesis as a consequence of their ectopic expression. The knowledge gleaned from this informatics analysis was then used in the creation of analogous napin variants that were compared to the effects of the recreated pumpkin 2S-PAT developed by Hayashi et al. [19] in transgenic Arabidopsis plants.

4.1.1 Informatics Analysis of 2S Albumin SSP

The informatics analysis began with a dissection of the pumpkin and Arabidopsis 2S albumin primary structures in relation to other SSP of the 2S albumin family. Furthermore, a comparison of the predicted secondary structures of the pumpkin 2S albumin and napin was conducted.

Albumin sequence alignments. Alignment of sequences of 2S albumins from several sources revealed that only marginal sequence identity (~25%) exists among dicot seed proteins (Figure 4-1); however, the conservation of cysteine residues (highlighted blue or yellow) occurring in the albumin family illustrates the importance of these amino acids for proper protein folding and disulfide bonding between the small and large chain peptides [44, 45]. Although not highly conserved, there are a number of short regions having conservative amino acids within the family (highlighted blue); namely, "MAKL" and "A--YRT" which occur in the first 21-26 residues and typifies the signal peptide (Figure 4-2a and b). Three modestly conserved regions are likely to contribute to interchain disulfide bonding: consensus sequence positions 50-70, in which two cysteine residues are found, the region surrounding the cysteine doublet, "QCC_EL_Q" at location 108-122, and cysteine residues near the carboxyl terminal at positions 163-172 (green bars, Figure 4-3). These findings compare favorably with previous assessments [44].

Pumpkin 2S and napin secondary structure predictions. Despite only marginal sequence identity among the various 2S albumins shown in Figure 4-1, a comparison of the predicted secondary structures between pumpkin 2S and *B. napus* albumins using PredictProtein online software (www.predictprotein.org) reveals considerable shared secondary structural features. This software assesses the probability of each amino acid of the sequence to participate in alpha helical or flexible loop regions. According to this probability algorithm, each albumin consists of five alpha helical regions separated by five internal flexible loop regions (Figures 4-4, 4-5 and 4-6). These findings suggest that pumpkin 2S and napin albumins share secondary and tertiary structural features, which gives credence to probable common protein sorting and

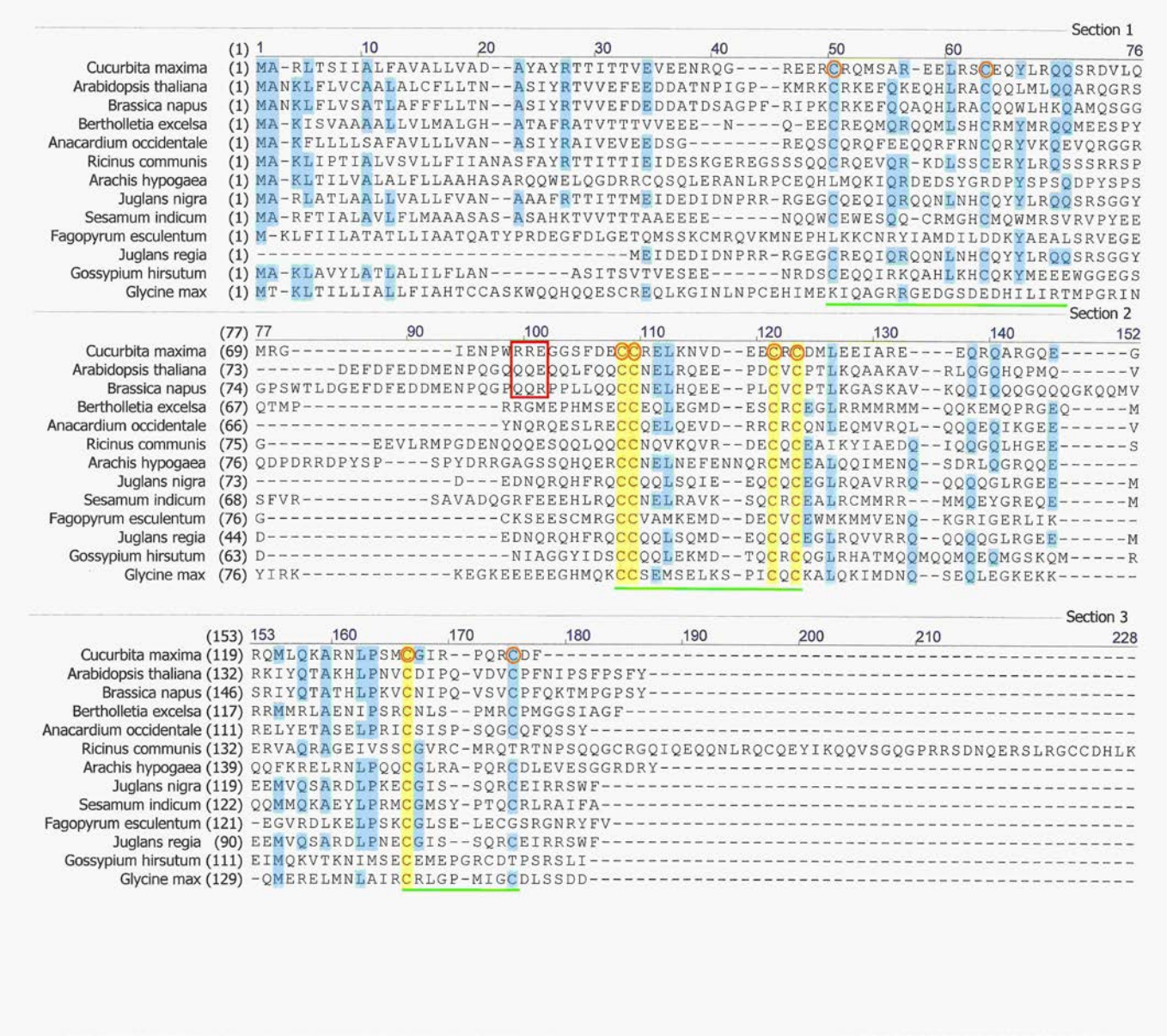


Figure 4-1. Sequence alignment of dicot albumin proteins employing the AlignX module of Vector NTI Advance® 11.5.1 software. Amino acids highlighted in blue represent amino acids identical in nearly all albumins examined and those highlighted in yellow identify those having unanimous identity. Green bars indicate moderately conserved regions possessing cysteine residues likely to be involved in interchain disulfide bonding (orange circles). The red box highlights the "RRE" amino acid triplet of pumpkin 2S albumin thought to be essential for evoking PAC vesicle biogenesis and the potentially analogous regions in *A. thaliana* and *B. napus*.

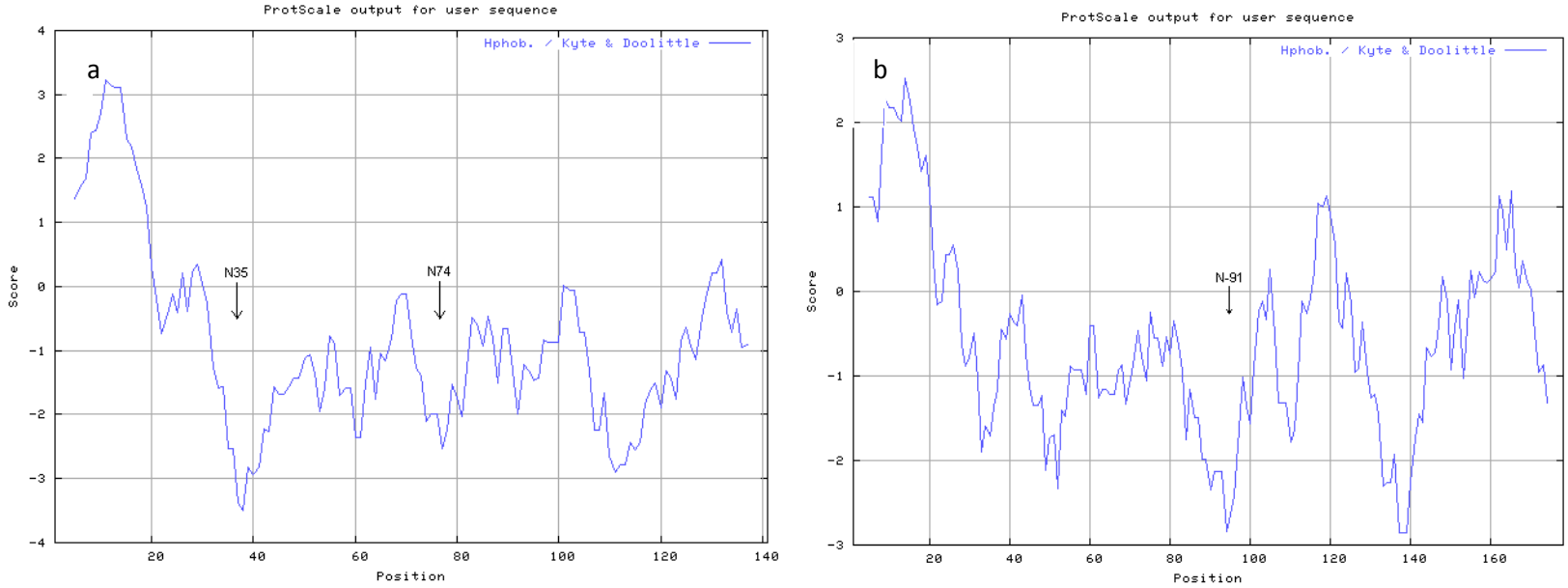


Figure 4-2 ProtScale amino acid hydropathy plots of pumpkin 2S pro-albumin (a) and *B. napus* Napin-2 pro-albumin (b) sequence according to the amino acid hydropathy index created by Kyte and Doolittle [182]. Shown are the signal peptide regions exemplified by a high positive hydropathy score typical of membrane spanning peptide regions occurring within the first 21 amino acid residues and the location of asparagine residues (N35, N74 and N91) that are targets for C-terminal cleavage by vacuolar processing enzyme (VPE)-cysteine protease.

Pumpkin 2S (Q39649)

Feature	Position	Length	Description	Graphical View (N-terminal to C-terminal)
Signal peptide	1 – 22	22		
Propeptide	23 – 35	13		
Chain	36 – ?		2S albumin small chain	
Chain	75 – 141	67	2S albumin large chain	

Pumpkin 2S albumin predicted disulfide bonding

Disulfide bond	43 ↔ 97	Interchain (between small and large chains)	
Disulfide bond	55 ↔ 86	Interchain (between small and large chains)	
Disulfide bond	87 ↔ 132		
Disulfide bond	99 ↔ 139		

Figure 4-3. Molecular sequence analysis indicating the peptide regions (green bars) and the predicted disulfide bonding of pumpkin 2S albumin (Q39649) from *Cucurbita maxima* (Winter squash) as reported by UniProtKB/Swiss-Prot protein sequence resource. The peptide regions of Napin-2 (P01090) from *B. napus* (Rape) are also shown for comparison.

trafficking mechanisms in the synthetic pathway of the endomembrane systems for these two species

.**Implications for construct design.** Based on their secondary structure commonalities, it was possible to formulate a strategy for the development of napin constructs analogous to the pumpkin 2S albumin-PAT fusion employed by Hayashi et al. (1999) [19]. The details of the synthesis of these constructs is outlined in Section 3.2 but the strategy employed in their design, based on the findings of the informatics analyses, is discussed here. The entire suite of constructs is summarized in Figure 4-7.

The Hayashi group described the use of the first 79 amino acid residues of the pumpkin 2S albumin coding region fused in frame to phosphinothricin acetyltransferase (PAT). Their use of PAT was to provide a selectable marker for their transgene and the expression of their albumin-PAT fusion open reading frame was driven by the CaMV 35S promoter in the pMAT037 binary vector [19]. To expedite production of a suitable replica for my research, two gene fusion open reading frames were synthesized (GeneArt) and integrated into pPZP121 binary vector [188] down-stream of the CaMV 35S promoter: namely, pumpkin2S(Δ79)-PAT and pumpkin2S(Δ79)-GFP. The first of these was intended to emulate the work of Hayashi et al. [19] while the second GFP (green fluorescent protein) fusion was intended to facilitate fluorescence microscopy in the study of localization behavior of the accumulating fusion protein.

The analogous napin-PAT fusion constructs were created by recombinant PCR (rPCR) in which a modified Napin-2 open reading frame was truncated at critical locations (Figure 4-6b). The Hayashi group surmised that the amino acid triplet arginine-arginine-glutamate (RRE) at location 77-79 is involved in binding to a sorting receptor and was essential for the creation of the observed novel vegetative vesicles [19, 190]. This triplet occurs distal to asparagine residue,

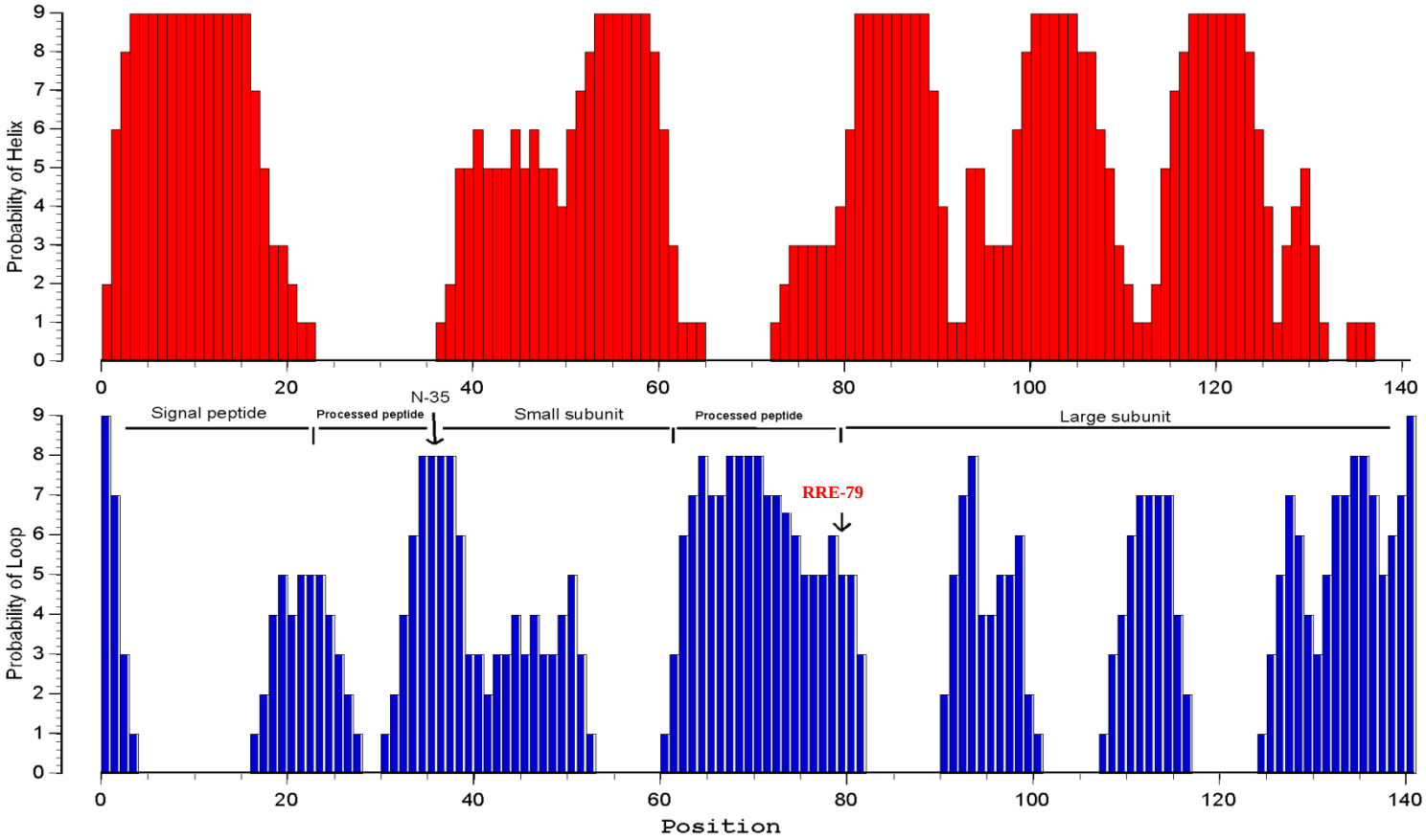


Figure 4-4. Secondary structure prediction of pumpkin 2S pro-albumin according to the online PredictProtein protein sequence analysis tool. Indicated are the signal peptide, small and large subunit and processed peptide regions as well as asparagine residues at position 35 and 79 thought to delineate the C-terminus of processed peptides in seed storage proteins. The “RRE” triplet claimed to be essential for PAC vesicle formation by Hayashi et al. (1999) is highlighted in red.

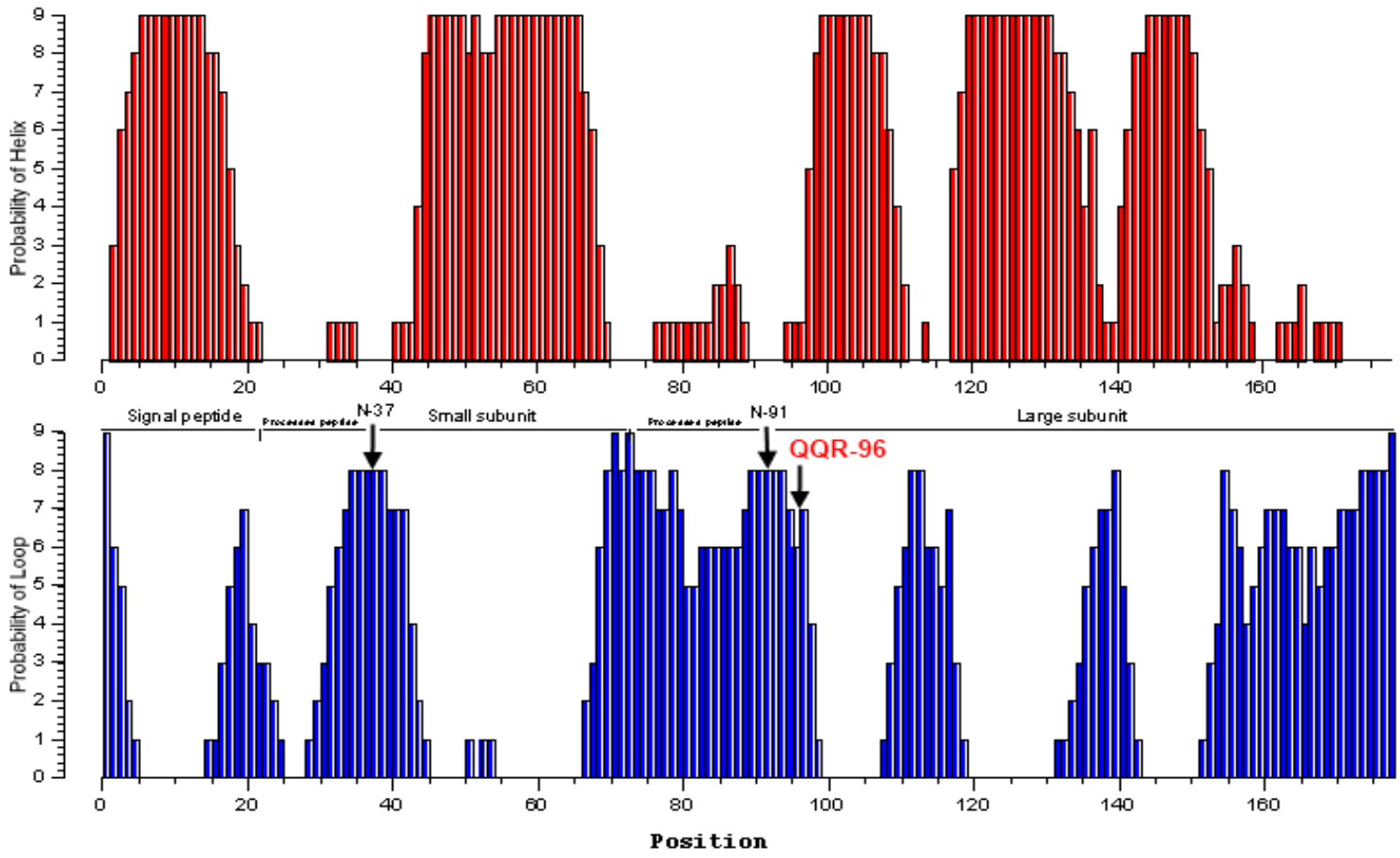


Figure 4-5. Secondary structure prediction of Napin-2 pro-albumin according to the online PredictProtein protein sequence analysis tool. Indicated are the signal peptide, small and large subunit and processed peptide regions as well as asparagine residues at position 37 and 91 thought to delineate the C-terminus of processed peptides in seed storage proteins. Shown in red is the "QQR" at position 96 thought to be analogous to the pumpkin 2S albumin "RRE" amino acid triplet.

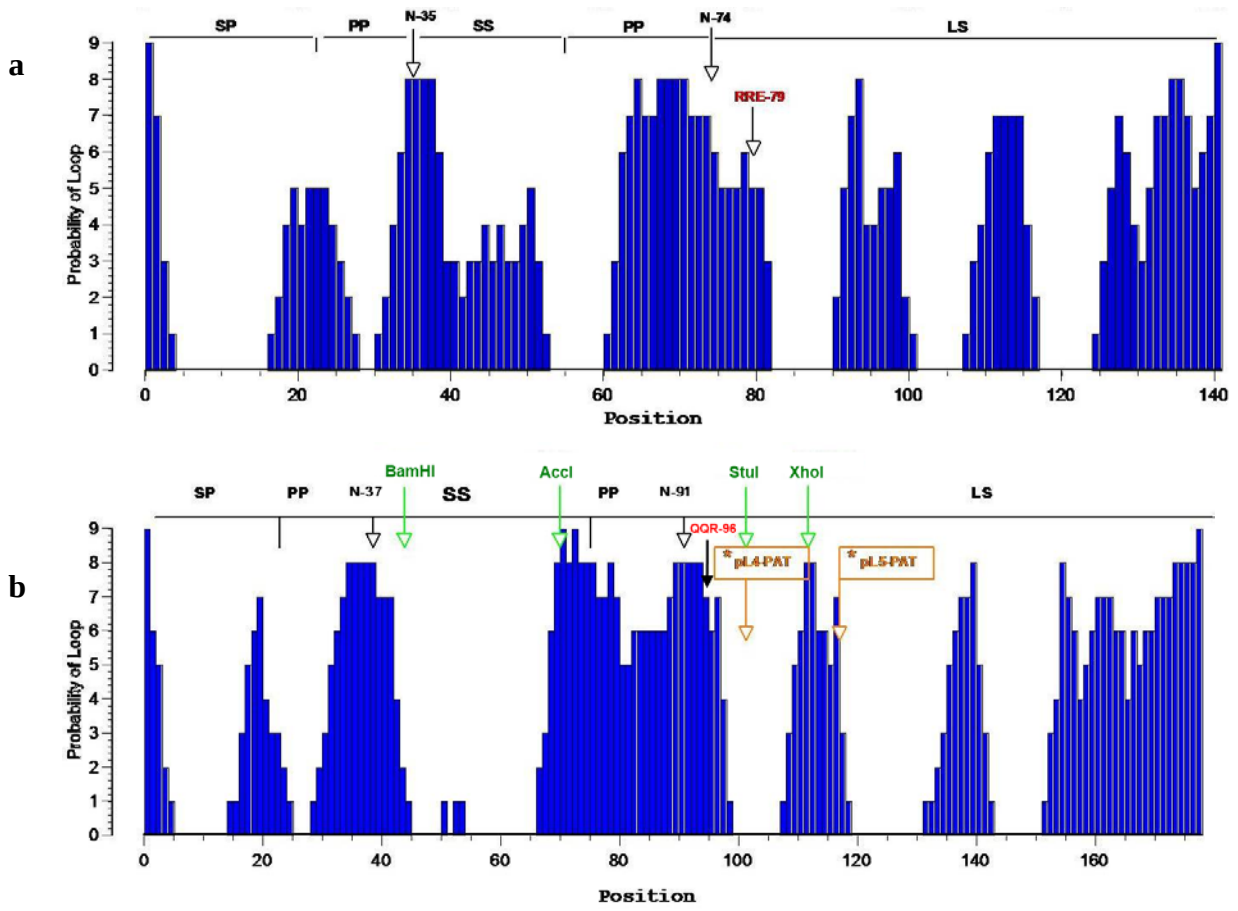


Figure 4-6. Comparison of pumpkin 2S (a) and Napin-2 (b) pro-albumin protein sequences according to their flexible loop regions as determined by PredictProtein secondary structure prediction tool. Key**:** SP Signal Peptide, PP Processed Peptide, SS Small Subunit, LS Large Subunit. The critical "REE" triplet of the pumpkin 2S albumin is highlighted in red. Green labeled restriction endonuclease sites BamHI, AccI, StuI and XhoI(E~L) and orange labeled *pL4-PAT and *pL5-PAT indicate the truncation sites used in the napin-GFP and napin-PAT fusions, respectively.

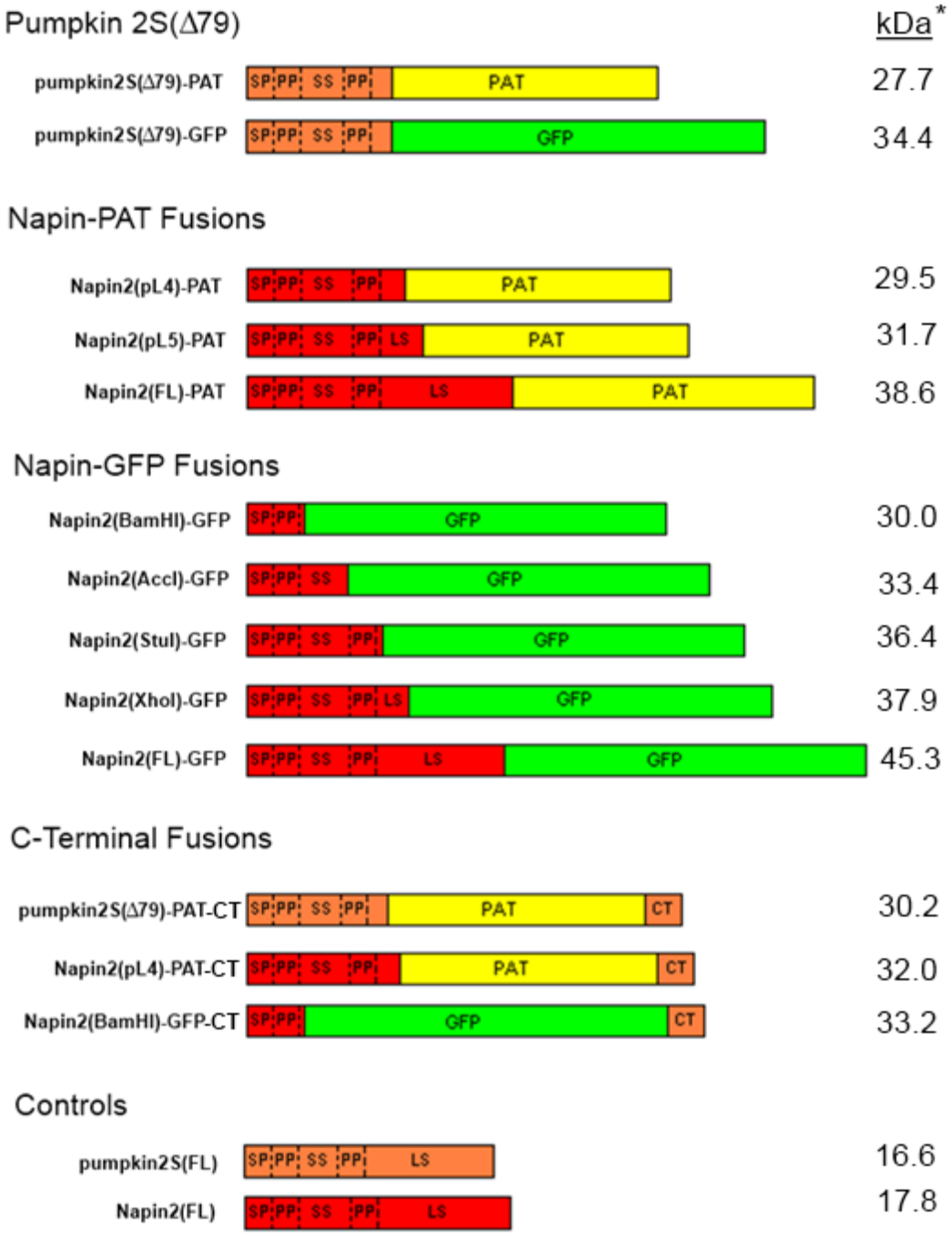


Figure 4-7 . A graphical representation of open reading frames employed for the discovery of PAC vesicle sorting determinants. Key**:** SP Signal Peptide, PP Processed Peptide, SS Small Subunit, LS Large Subunit, CT Carboxyl Terminus, PAT Phosphinothricin Acetyl Transferase, GFP Green Fluorescent Protein. * kDa size expected of the fusion protein following removal of the signal peptide (SP)

N^{74}, that is critical for processed peptide cleavage of the pumpkin pro-albumin and near the carboxyl end of probability loop region #4 (Figure 4-6a). A comparable critical asparagine was identified in *B. napus* napin, N-91 that delineates the end of its internal processed peptide. Although an RRE triplet does not occur in the napus sequence following N-91, it is followed by a short segment of amino acids rich in amino group side-chains bracketed by proline residues, "PQGPQQRPP". This highly polar stretch is even more prominent in Arabidopsis napin as "PQGQQQEQQL" and a similar short polar segment is observed in this region of castor bean (*Ricin communis*) albumin (Figure 4-1). This suggests the peptide segment is an exposed region that may participate as a sorting receptor target as surmised by Hayashi et al. (1999). By deduction then, truncations of napin distal to this critical asparagine, N-91 but retaining this basic segment would create partial napin peptides analogous to pumpkin 2S(Δ79). As such, truncations removing sequence distal to probability loops #4 and #5 were chosen, hence the origin of napin fusion constructs pL4-PAT and pL5-PAT, respectively (Figure 4-6b). To account for any effects towards sorting behavior attributed to either PAT or GFP, the full length pumpkin 2S albumin and full-length Napin-2 sequences alone without fusion to either PAT or GFP were included as controls.

The napin-GFP fusion constructs were initially acquired from the lab of Dr. Dwayne Hegedus as GFP inserts of codon modified Napin-2. For my study, they were further modified by site directed mutagenesis (SDM) to facilitate napin truncations at four locations as delineated by restriction sites BamHI, AccI, StuI and XhoI (see Figure 4-6b). Napin truncations at StuI and XhoI sites closely approximate those at locations pL4 and pL5 respectively that were fused to PAT, and would facilitate microscopic observation of localization behavior. The napin truncations at AccI and BamHI were included to assess the contribution towards protein sorting

of the pro-napin internal processed peptide (region 71-94) and the napin small subunit sequence, respectively.

4.1.2 Validation of Transgenic Arabidopsis plants

This section details the molecular analyses conducted on transgenic Arabidopsis harboring one of the transgenes described above, which to facilitate leaf ectopic expression, were driven by the CaMV 35S promoter in either pPZP121 or pMDC32 plant transformation binary vectors (see Fig. 4.7). As indicated in Section 3, the light inducible chlorophyll a/b binding protein 1 (*CAB1*) promoter was also employed in selected constructs, with the intention to assess the contribution of expression levels on sorting behavior. First generation transgenic plants obtained by gentamycin selection and confirmed by PCR were developed to the T_2 and T_3 generations for most constructs (Tables 4-1a and 4-1b). An exception to this were constructs A(4)P, A(5)P, and A(FL)P, which were created using the *napA* coding region of the napin gene from *B. napus* as an alternative to the codon modified mNAP, and constructs M(5)P and M(FL)P which were not successfully created by recombinant PCR. Ultimately, these five were abandoned in favor of the codon optimized GA-M4P, GA-M5P and GA-MFLP constructs. In addition, successive generations of plants harboring transgenes driven by the CAB1 promoter were not pursued due to poor survivability of T_1 plants except for constructs CAB:M(4)P and the CAB:GUS control which were both of consistently poor health in the T_2 generation. A summary of the constructs made and the resulting stable transgenic plants of T_1, T_2 or T_3 generation developed for this study are shown in Tables 4-a1 and 4-1b.

CmP. Pumpkin 2S(Δ79)-PAT open reading frame in T_1 generation CmP transgenic plants was confirmed (Figure 4-8a). Protein extracts acquired from the rosette leaves of numerous confirmed positive T_1 plants indicated as many as four peptides reactive to anti-PAT

Table 4-1a. Constructs Employing the CaMV 35S Promoter in Arabidopsis

Construct Name	ORF	Synthesized by	Binary Vector	No. Plants of Generation A.t. T_1	A.t. T_2	A.t. T_3
CmP	Cm(Δ79)-PAT	GA	pPZP121	10	24	32
CmGFP	Cm(Δ79)-GFP	GA	pPZP121	6	31	
M(4)P	mNap(pL4)-PAT	rPCR	pPZP121	14	42	
M(5)P	mNap(pL5)-PAT	rPCR	pPZP121	0		
M(FL)P	mNap(FL)-PAT	rPCR	pPZP121	0		
GA-N4P	Nap2(pL4)-PAT	GA	pPZP121	10	11	
GA-N5P	Nap2(pL5)-PAT	GA	pPZP121	10	11	
GA-NFLP	Nap2(FL)-PAT	GA	pPZP121	13	10	
A(4)P	NapA(pL4)-PAT	rPCR	pPZP121	0		
A(5)P	NapA(pL5)-PAT	rPCR	pPZP121	0		
A(FL)P	NapA(FL)-PAT	rPCR	pPZP121	0		
N(B)GFP	mNap(B)-GFP	rPCR/SDM	pMDC32	11	17	
N(A)GFP	mNap(A)-GFP	rPCR/SDM	pMDC32	0		
N(S)GFP	mNap(S)-GFP	rPCR/SDM	pMDC32	0		
N(X)GFP	mNap(X)-GFP	rPCR/SDM	pMDC32	6	7	
N(FL)GFP	mNap(FL)-GFP	rPCR	pMDC32	3		
CmPCT	Cm(Δ79)-PAT-CT	rPCR	pPZP121	3	20	
M4PCT	mNap(pL4)-PAT-CT	rPCR	pPZP121	10	28	
N(B)GFPCT	mNap(B)-GFP-CT	rPCR	pPZP121	9	22	
Cm(FL)	Cm(FL)	GA	pPZP121	11	18	
Nap(FL)	mNap(FL)	PCR	pPZP121	18	22	

Table 4-1b. Constructs Employing the *CAB1* Promoter in Arabidopsis

Construct Name	ORF	Synthesized by	Binary Vector	A.t. T^1
CAB:CmP	Cm(Δ79)-PAT	GA	pZP121	
CAB:M(4)P	mNap(pL4)-PAT	rPCR	pZP121	11
CAB:CmPCT	Cm(Δ79)-PAT-CT	rPCR	pZP121	
CAB:M4PCT	mNap(pL4)-PAT-CT	rPCR	pZP121	
CAB:N(B)GFPCT	mNap(B)-GFP-CT	rPCR	pZP121	
CAB:GUS Cntrl	GUS	ligation	pZP121	2

Key: ORF, Open Reading Frame; A.t., *A. thaliana*; Cm, *C. maxima* 2S albumin; (Δ79), Truncation at Residue 79; PAT, Phosphinothricin Acetyltransferase; GA, GeneArt; GFP, Green Fluorescent Protein; mNap, modified *B.napus* Napin-2; (pL4), Loop Region 4 Truncation; (pL5), Loop Region 5 Truncation; (FL), Full Length Peptide; rPCR, Recombinant PCR; Nap2, *B.napus* Napin-2; NapA, *B.napus* napA; (B), (A), (S), (X), Truncations at BamHI, AccI, StuI and XhoI sites; SDM, Site Directed Mutagenesis; CT, 18 Residue Carboxy Terminus of *C. maxima* 2S albumin; GUS, beta-glucuronidase.

antibody (Figure 4-8b - d). Two of these peptides compare favorably in size to predicted molecular mass of the pumpkin 2S(Δ79)-PAT fusion protein devoid of signal peptide (27.7 kDa) and the fusion protein devoid of both the signal peptide and the adjoining processed peptide region (26.1 kDa). These two peptides were frequently observed but the additional smaller peptides of 20 kDa and 12 kDa were not observed in T_1 or T_2 generation plants except CmP-21.2 (Figure 4d). This suggests a varying degree of processing of the fusion protein among subjects that may be correlated with differing expression levels between plants or other factors such as leaf size, age or position and/or environmental factors. The accumulation behavior of these plants in T_2 and subsequent generations was explored more rigorously and these findings are described in section 4.2.

CmGFP. T_2 generation CmGFP transgenic plants were confirmed by PCR (Figure 4-9a). The silver stained gel and corresponding western blot analysis of pooled rosette leaf samples for each of these positive plants are indicated in Figure 4-9b. A number of peptides reactive to anti-GFP antibody that do not occur in the wild-type or the negative control are evident; however two peptides occur consistently for each pooled sample and can be seen clearly for sample 6.x (Figure 4-9). One peptide corresponds in size to the predicted value of pumpkin 2S(Δ79)-GFP fusion protein following release of the signal peptide (34.4 kDa). The other peptide of ~22 kDa was not predicted, however Tamura et al., (2003) also observed an unpredicted anti-GFP reactive peptide when a secreted GFP targeted to vacuoles was expressed in Arabidopsis plants. These authors cite a half-life of vacuolar GFP due to a cysteine proteinase active under acidic vacuolar conditions [190].

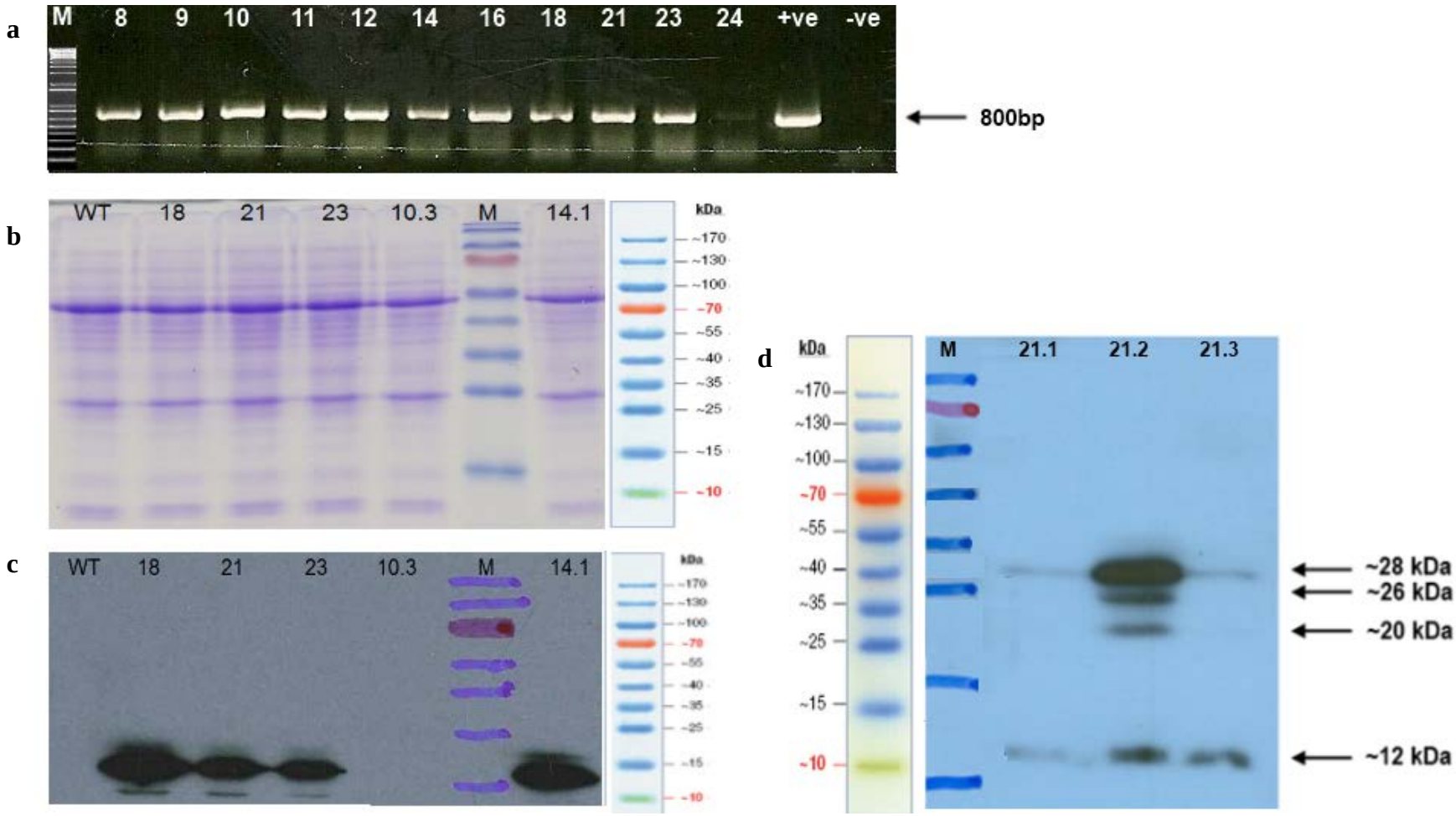


Figure 4-8. Molecular analysis of Arabidopsis transgenic plants harboring the CmP transgene as indicated by the 800bp PCR amplicon generated by the use of Cm-F1 and PAT-R2 primers (a), Coomassie stained 15% acrylamide SDS-PAGE gels of rosette leaf protein extracts (b) and their corresponding western blot employing anti-PAT antibody (c and d). Key: M, 1 kb plus ladder (a) or PageRuler™ Prestained Protein Ladder—Fermentas (b-d); +ve, CmP:pZP121 construct positive control; -ve, water negative control; WT, Arabidopsis (var. Columbus) wild type.

M(4)P. Several T_1 generation Arabidopsis plants harboring the mNap(pL4)-PAT open reading frame were isolated and confirmed by PCR (Figure 4-10a) and their progeny were advanced to the T_2 generation. Although M(4)P transgenic plants showed partial resistance to the herbicide phosphinothricin similar to CmP transgenic plants (detailed in section 4.2.1), it was not possible to observe accumulation of the fusion protein in leaf protein extracts among several dozen candidates by western blot except for three T_2 plants, namely 13.3, 16.1 and 16.2 (Figure 4-10b). Protein extracts of plant 13.3 and 16.1 possess peptides of approximately 30 kDa, 27 kDa and 22 kDa that are reminiscent of the peptides observed for plant CmP-21.2 (Figure 8d). Similar to those of CmP-21.2, these peptides correlate to the predicted size of the mNap(pL4)-PAT fusion protein devoid of the signal peptide (29.5 kDa) and the fusion protein devoid of signal peptide and the 17 amino acid processed peptide (27.5 kDa). However, for the protein extract of plant 16.2 only a single peptide of ~30 kDa was clearly observed. This again suggests varying degrees of processing likely dependent on differing expression levels between plants.

GA-N(4)P, GA-N(5)P and GA-N(FL)P. As mentioned previously in section 3.2, these three constructs were utilized to confirm that codon usage had no bearing on sorting behavior. Figure 4-11a indicates the results of RT-PCR analysis of pooled T_2 generation rosette leaf samples. Amplified product for RT reaction (columns labeled ‘a’) in excess of PCR controls (columns labeled ‘b’) confirmed transgene expression in samples GA-N(4)P: 1.x, 3.x and 6.x, GA-N(5)P: 3.x and 6.x, and GA-N(FL)P: 1.x. The predicted sizes of GA-N(4)P, GA-N(5)P and GA-N(FL)P fusion proteins devoid of their SP were 29.5 kDa, 31.7 kDa and 38.6 kDa, respectively. Western blot analysis confirmed that accumulation of GA-N(5)P fusion protein had not occurred at detectable levels as no anti-PAT reactive peptides were observed in samples GA-N(5)P-3.x and 6.x; however two peptides, ~30 kDa and ~38 kDa, were observed for each of

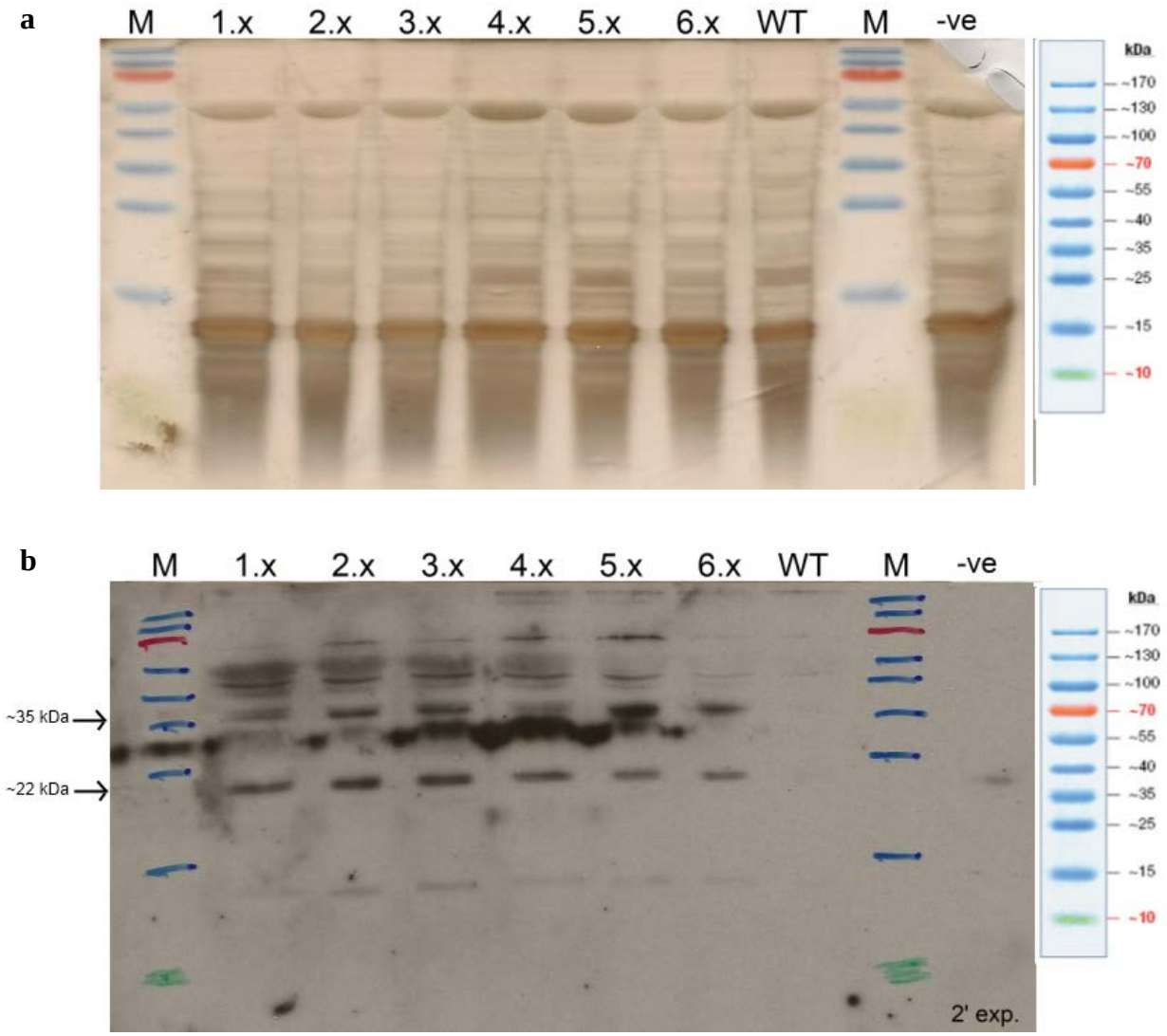


Figure 4-9. (a) Silver-stained SDS-PAGE gel and (b) western blot analysis of protein extracts of pooled rosette leaves of T_2 plants tested for accumulation of CmGFP fusion protein employing rabbit anti-GFP antibody detected by anti-Rabbit-HRP conjugate. Key: M and inset, PageRuler™ Prestained Protein Ladder—Fermentas; WT, Arabidopsis (var. Columbus) wild type, -ve, CmP-14.1 leaf protein extract; 1.x, 2.x, etc. indicates rosette leaf tissue pooled from T_2 generation plants originating from T_1 parent lines 1, 2, etc.

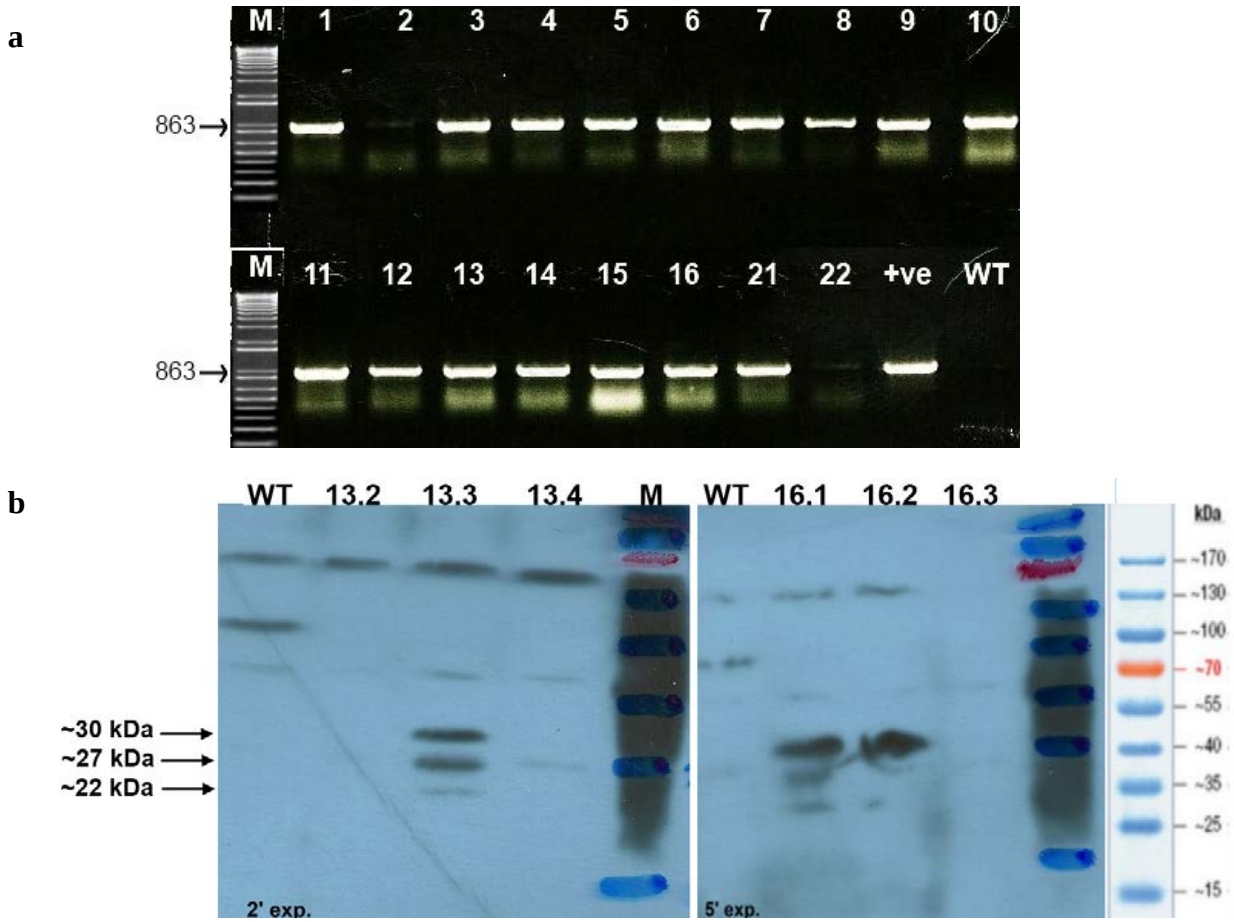


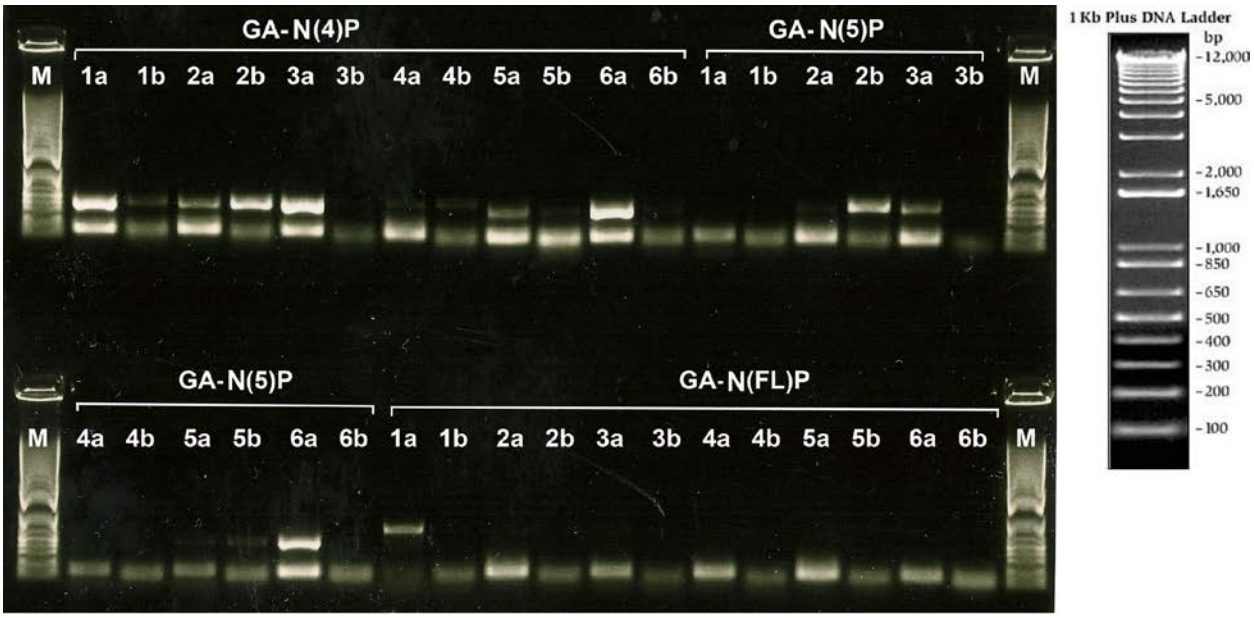
Figure 4-10. a) PCR analysis of T_1 generation Arabidopsis transgenic plants harboring the M(4)P transgene as indicated by the 863 bp amplicon produced by the use of NapM-F1 and PAT-R2 primers. Key: M, 1 kb plus ladder; +ve, M(4)P-pZP121-2A construct positive control; b) western blot analysis of protein extracts of T_2 plants tested for accumulation of M(4)P fusion protein employing rabbit anti-PAT antibody detected by anti-Rabbit-HRP conjugate. Key: M and inset, PageRuler™ Prestained Protein Ladder—Fermentas; WT, Arabidopsis (var. Columbus) wild type.

samples GA-N(4)P-1.x and 3.x and GA-N(FL)P-1.x (Figure 4-11b). In the case of samples GA-N(4)P-1.x and 3.x the ~30 kDa peptide predominated and correlated to the expected size of 29.5 kDa; however for sample GA-N(FL)P-1.x the larger peptide, which correlated to the expected size of 38.6 kDa, is of greater concentration. Because this 38.6 kDa band is observed also in the adjoining wild-type (WT) lane it is possible that this peptide band was due to nonspecific binding of the antibody reagents. Regardless, these findings confirmed that the napin-PAT transgenes did not elicit accumulation behavior as was observed for the CmP and that this is independent of codon usage.

N(B)GFP, N(A)GFP, N(S)GFP, N(X)GFP and N(FL)GFP. For this group of constructs, which consisted of a series of truncations of modified napin fused to GFP, difficulty was encountered recovering T_1 plants. PCR-confirmed positive plants were obtained for only constructs N(B)GFP, N(X)GFP and N(FL)GFP and none were obtained for N(A)GFP and N(S)GFP (Figure 4-12). Analysis of protein extracts from these plants and their T_2 progeny resulted in no observed accumulation. These challenges necessitated testing this group of constructs by tobacco transient expression assay, which is discussed in section 4.2.6.

CmPCT, M4PCT and N(B)GFPCT. It was known from the work of Tamura et al. (2003) [185], that the 18 amino acid C-terminus of the pumpkin 2S albumin is a ctVSD sufficient to direct a GFP fusion protein to the vacuole of Arabidopsis protoplast suspension cultured cells. In similar fashion, the 18 aa C-terminus of pumpkin 2S albumin was appended to open reading frames pumpkin 2S(Δ79)-PAT, mNap(pL4)-PAT, and mNap(BamHI)-GFP to assess what contribution this ctVSD might have as a sorting determinant in vegetative tissue of stably transformed Arabidopsis. Recovery of T_1 plants surviving on selection medium for constructs CmPCT and M4PCT proved troublesome until it was discovered that their T_0 seed

a



b

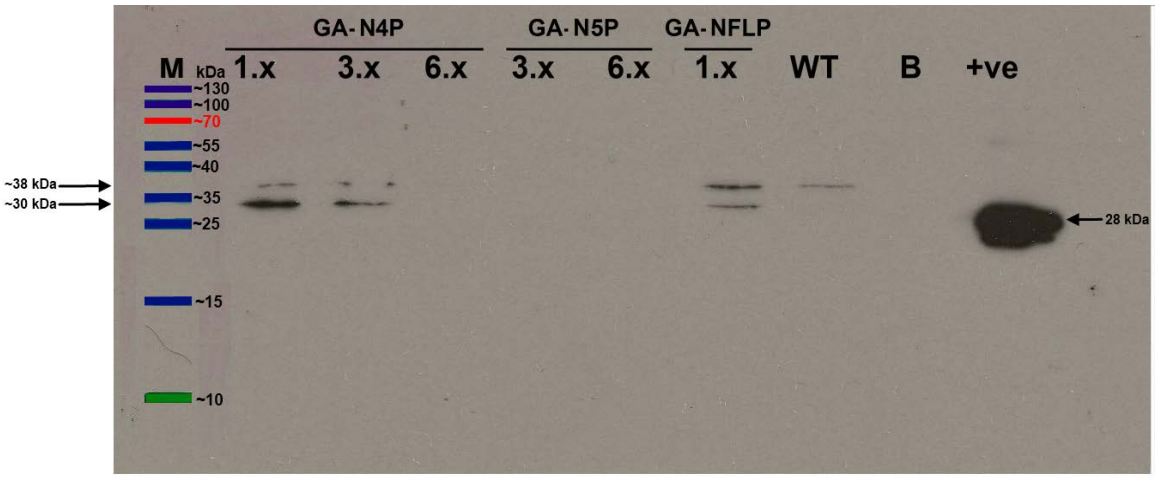


Figure 4-11. a) RT-PCR analysis of T_2 generation Arabidopsis transgenic plants harboring the GA-N(4)P, GA-N(5)P and GA-N(FL)P transgenes as indicated by the 395 bp,452 bp and 641 bp amplicons, respectively, produced by the use of GA-NapN-F8 and GA-PAT-R14 primers. Samples are leaf explants pooled from six T_2 plants derived from their T_1 parents 1 through 6. Columns labeled 'a' show the RT-PCR and columns labeled 'b' indicates their respective PCR controls. Key: M, 1 kb plus ladder, b) western blot analysis of protein extracts of pooled rosette leaves of T_2 plants tested for accumulation of napin-PAT fusion protein employing rabbit anti-PAT antibody detected by anti-Rabbit-HRP conjugate; Key: M and inset, PageRuler™ Prestained Protein Ladder—Fermentas; WT, Arabidopsis (var. Columbus) wild type; B, blank (empty) lane; +ve, CmP 14.1.3 positive control.

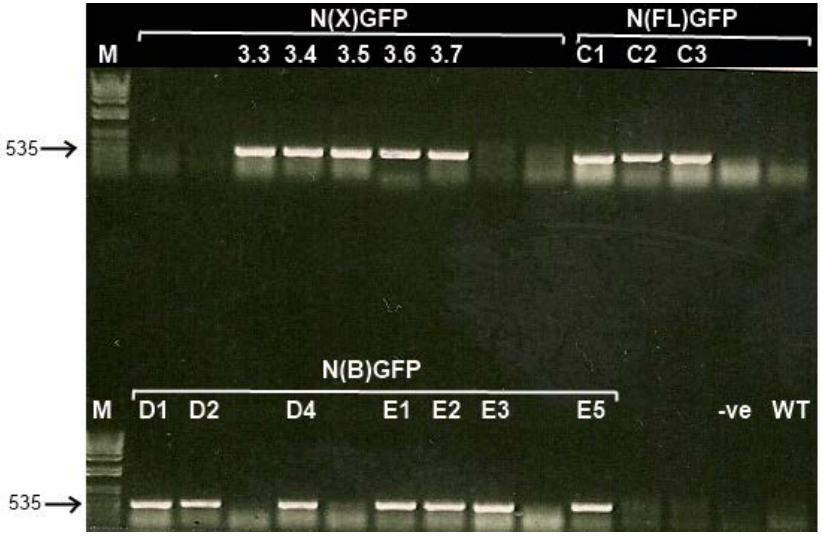


Figure 4-12. PCR analysis of T_2 generation Arabidopsis transgenic plants harboring the N(X)GFP, N(FL)GFP, or N(B)GFP transgenes as indicated by the 535 bp amplicon produced by the use of GFP-F1 and GFP-R2 primers. Key: M, 1 kb plus ladder; -ve, water negative control; WT, Arabidopsis wild-type negative control.

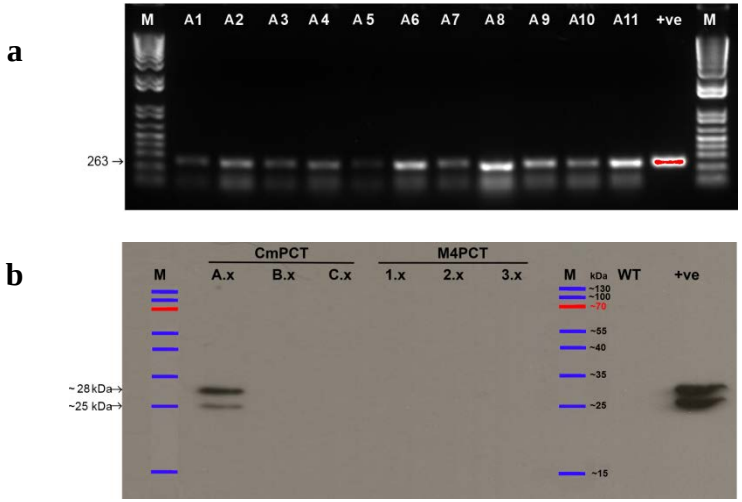


Figure 4-13. PCR analysis of T_2 generation CmPCT transgenic Arabidopsis as indicated by the 256 bp amplicon produced by the use of Cm-F1 and CmPJ-R primers. Key: M, 1kb plus ladder; +ve, CmPCT:pZP121construct positive control, b) western blot analysis of protein extracts of pooled rosette leaves of T_2 plants tested for accumulation of PAT fusion protein employing rabbit anti-PAT antibody detected by anti-Rabbit-HRP conjugate; Key: M and inset, PageRuler™ Prestained Protein Ladder—Fermentas; WT, Arabidopsis wild type; +ve, CmP 18.1.2 positive control.

could be encouraged to germinate following a long 10-14 day 4˚C, vernalization. As shown in Figures 4-13a and b, PCR-confirmed T_2 generation transgenic plants accumulated pumpkin 2S(Δ79)-PAT fusion protein in samples CmPCT-A.x[2], whereas no accumulation of mNap(pL4)-PAT fusion protein was observed for the M4PCT transgenic plants. In sample CmPCT-A.x two peptides were evident as was observed for the CmP-18.1.2 positive controland which had been seen previously in CmP transgenic plant samples (Fig. 4-8b and c). The anti-GFP western blot analysis of T_2 generation N(B)GFPCT plants also revealed an accumulation of two peptides but of approximate sizes 35 kDa and 25 kDa (Figures 4-14b). The detection of these peptides was reminiscent of construct CmGFP and findings of Tamura et al. (2003) [190] in which vacuole targeted GFP was subjected to processing or cleavage. The size of the 35 kDa peptide compares favorably with the calculated values of mNap(BamHI)-GFP devoid of its signal peptide (33.1 kDa) whereas the smaller 25 kDa peptide approximates the 27 kDa-GFP processed variant observed by Tamura et al. (2003) [190]. Further evidence for the expression and accumulation of the mNap(BamHI)-GFP fusion protein was obtained by microscopy of these plants following a 48 h dark treatment (see section 4.2.6).

Cm(FL) and mN(FL). Negative control constructs comprised of full length pumpkin 2S albumin or full length modified Napin2 open reading frames were used to observe the sorting behavior of native albumins in leaf tissue and to account for any effects towards sorting behavior attributed to either PAT or GFP (Figures 4-15 and 4-16). Despite RT-PCR confirming expression of the mNap(FL) open reading frame (Figure 4-16b), accumulation of the native albumins could not be detected by western blot analysis employing antiserum directed against pumpkin 2S and napin albumin small subunit and processed peptide.

[2] Note: ".x" denotes a grouping of pooled rosette leaf samples from T_2 generation transgenic plants

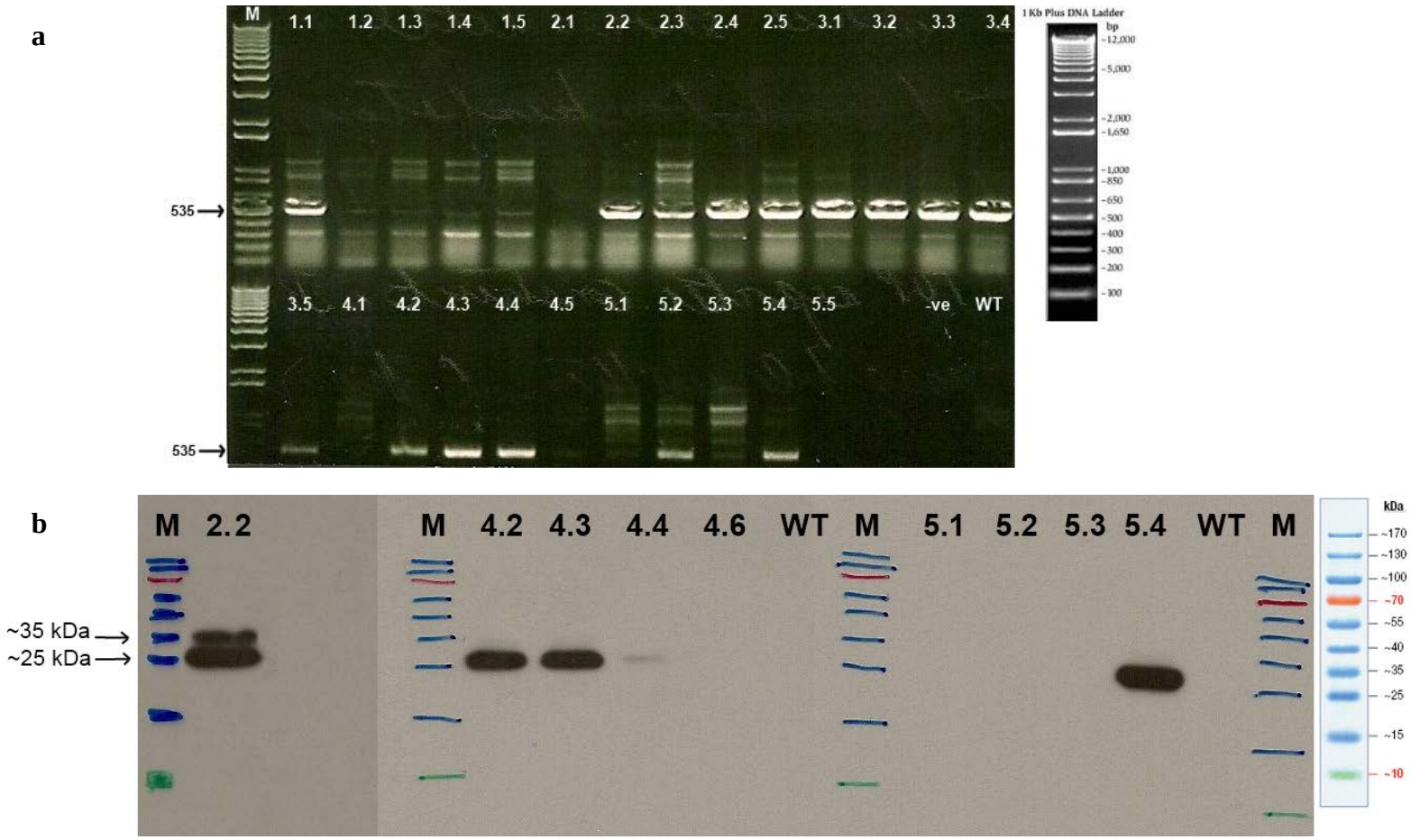


Figure 4-14. a) PCR analysis of T_2 generation Arabidopsis transgenic plants harboring the N(B)GFPCT transgene as indicated by the 535 bp amplicon produced by the use of GFP-F1 and GFp=R2 primers. Key: M, 1 kb plus ladder; -ve, water negative control; WT, Arabidopsis wild-type negative control; b) western blot analysis of protein extracts of rosette leaves of T_2 plants tested for accumulation of N(B)GFPCT fusion protein employing rabbit anti-GFP antibody detected by anti-Rabbit-HRP conjugate. Key: M and inset, PageRuler™ Prestained Protein Ladder—Fermentas; WT, Arabidopsis wild type.

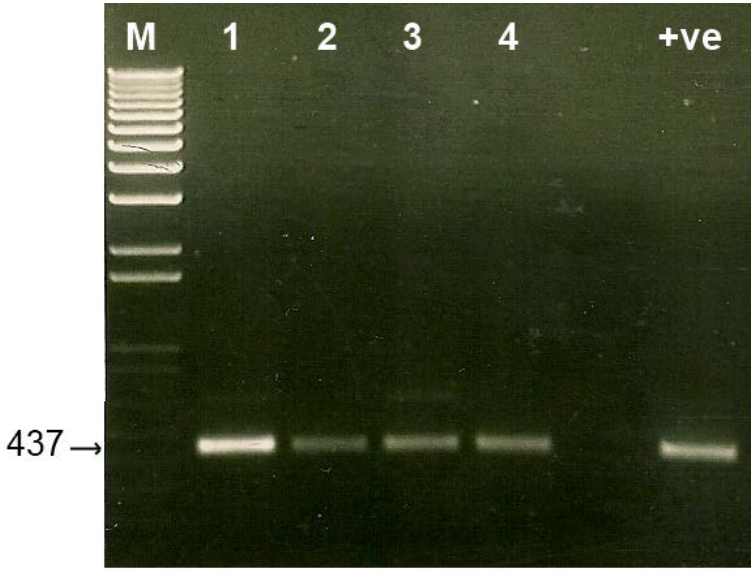


Figure 4-15. PCR analysis of T_1 generation Arabidopsis transgenic plants harboring the Cm(FL) transgene as indicated by the 437 bp amplicon produced by the use of CmFL-F1 and CmFL-R2 primers. Key: M, 1 kb plus ladder; +ve, CmFL:pZP121 construct positive control.

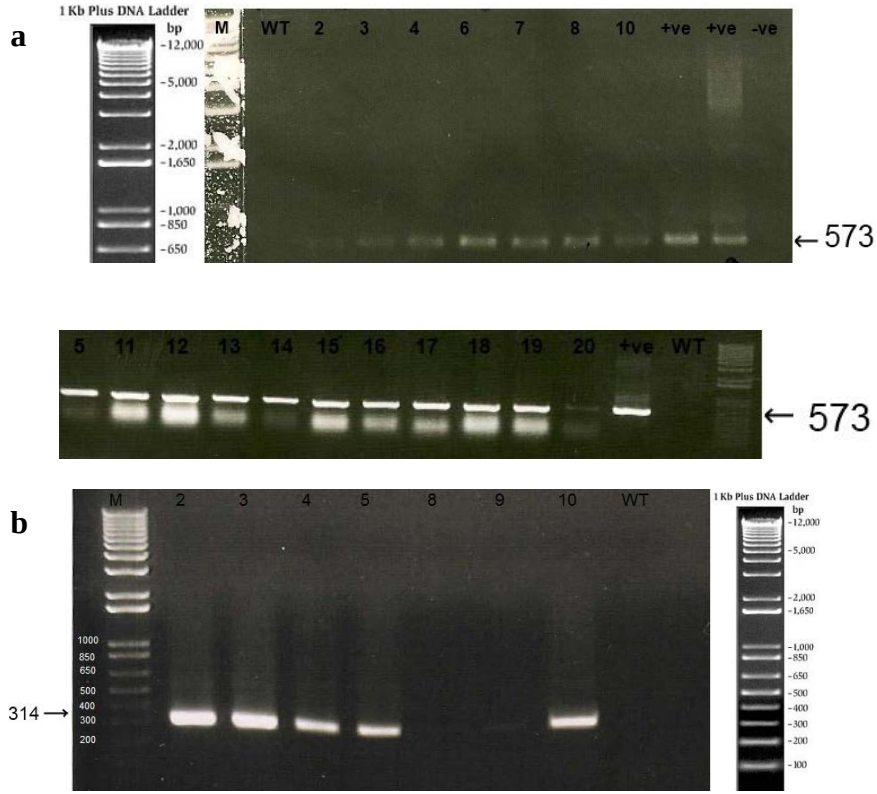


Figure 4-16. a) PCR analysis of T_1 generation Arabidopsis transgenic plants harboring the mNap(FL) open reading frame as indicated by the 573 bp amplicon produced by the use of NapM-F1 and NapM-R5 primers; b) RT-PCR analysis indicating expression of mNap(FL) open reading frame as evidenced by the 314 bp amplicon generated by the use of primers NapM-F1 and NapM-FR2.

4.2 Functionality and Targeting of Fusion Proteins

Detection of appropriate transcripts and fusion protein gene products of stable transgenic Arabidopsis plants through RT-PCR and western blot analysis implied successful expression. However, to understand sorting behavior and targeting mechanisms, further characterization aimed at assessing correct fusion protein folding and functionality was necessary. For the PAT gene fusion constructs, this assessment focused on five facets: 1) evaluation of PAT enzyme activity by detection of resistance to phosphinothricin (PPT), 2) observing the accumulation behavior in leaf tissue samples in subsequent generations by western blot analysis, 3) determination of fusion protein localization by PAC-like vesicle isolation facilitated by differential centrifugation, 4) antibody marker assisted characterization of PAC-like vesicles and 5) isolation of PAC-like vesicles by antibody-capture. These studies were complemented by microscopy of Arabidopsis leaf tissues expressing GFP-fusion transgenes.

4.2.1 Assessment of PPT Resistance of CmP and M(4)P Transgenic Arabidopsis

The T_1 generation CmP transgenic Arabidopsis showed partial resistance to PPT concomitant with pumpkin 2S(Δ79)-PAT fusion protein accumulation (section 4.1.2, Figures 4-8c and 4-17). This compares favorably with findings of Hayashi et al. (1999) who attributed only partial resistance toward PPT to the incomplete compartmentalization of the fusion gene product into vesicles. Arabidopsis harboring the analogous M(4)P transgene were also found to possess partial resistance to PPT even though the mNapin2(pL4)-PAT fusion protein failed to appreciably accumulate in leaf tissue (section 4.1.2 and Figure 4-10). The greater number of surviving seedlings at 20 μg/mL PPT for M(4)P transgene would even suggest greater PPT^R than the CmP transgene. Certainly, this observation indicated that proper transgene expression, protein folding and functionality occurred for both CmP and M(4)P transgenes but the M(4)P

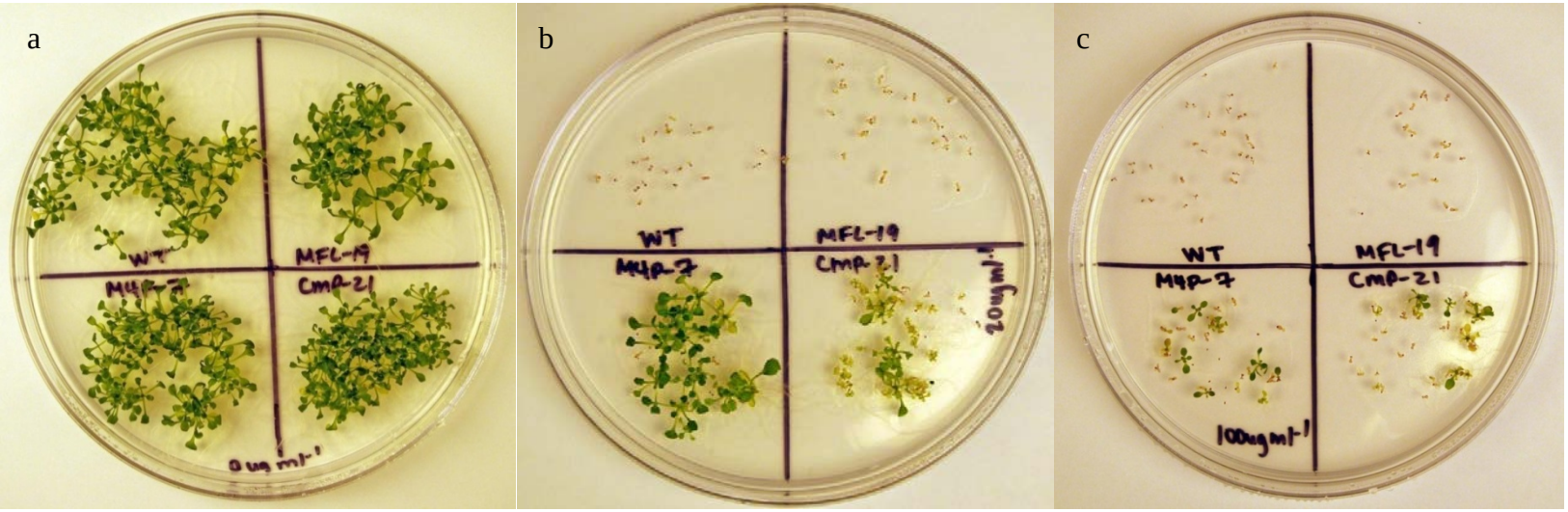


Figure 4-17. Comparison of wild-type (WT) to T_1 transgenic Arabidopsis plants harboring CmP, M(4)P or M(FL) transgenes on ½ MS medium containing a) 0, b) 20 and c) 100 µg/mL PPT.

transgeene had a limited capacity to evoke vesicle biogenesis and accumulate the mNap(pL4)-PAT fusion protein.

4.2.2 Assessment of Accumulation Behavior of CmP Transgenic Arabidopsis

To evaluate the consistency of fusion protein accumulation in leaf tissue in T_2 and subsequent generations, the progeny of T_1 CmP-14 and 18 plants were developed to the T_4 generation. From T_3 and T_4 plants, leaves of varying locations and age were examined for pumpkin 2S(Δ79)-PAT fusion protein (Figures 4-18a and b). From this, it was evident that accumulation did not occur consistently among T_4 plants derived from the same T_3 parent. This inconsistency was also observed for T_3 plants originating from the same T_2 parent. This suggested that the capacity of leaf tissue to accumulate the fusion protein might be due to the segregation of one or multiple copies of the transgene and it was thought that the inconsistency from generation to generation would diminish with selection of successive generations. Because this trend continued into the T_4 generation, it would seem that accumulation in leaf tissue was not a consistently heritable phenotype in these transgenic lines. Furthermore, inconsistent accumulation behavior was observed for differing rosette leaves of a single plant. With certainty however, fusion protein accumulation, when it had occurred, did so in leaves of size ~1 cm × 2.5 cm of transgenic plants whereas it was never observed in younger rosette leaves of less size or the alternate leaves of the stem. To determine if aspects of processing or proteolysis had occurred giving rise to the observed inconsistency of accumulation behavior, custom antibodies reactive towards the small subunit and processed peptide regions of the pumpkin 2S albumin were developed (Figure 3-1). Unfortunately the quality of these antibodies were poor but in spite of this, the antiserum targeting pumpkin 2S albumin processed peptide (A-CmPP) reacted with a ~26 kDa peptide of the CmP-14.1 leaf protein extract and a ~20 kDa peptide was identified with

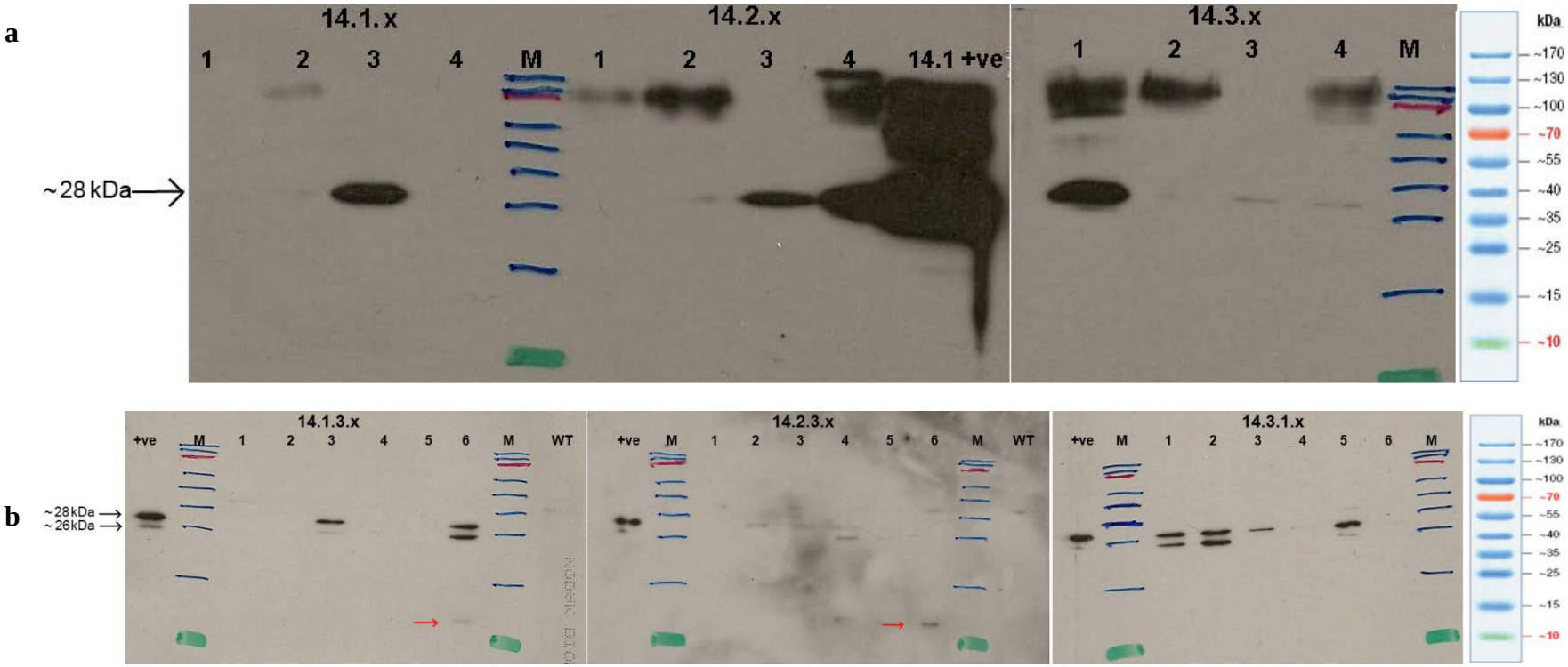


Figure 4-18. Survey of T_3 and T_4 generation Arabidopsis CmP-14 tested for accumulation of CmP fusion protein employing rabbit anti-PAT antibody detected by anti-Rabbit-HRP conjugate. a) Western blot analysis of leaf protein extracts from rosette leaves harvested from T_3 generation plants followed by b) western blot analysis of leaf protein extracts from six T_4 generation plants originated from their positive T_3 parents in a) Key: M, PageRuler™ Prestained Protein Ladder—Fermentas; WT, Arabidopsis wild type; +ve, CmP 14.1 positive leaf protein extract.

the antiserum targeting the pumpkin 2S small subunit (A-CmSS) (Figures 4-19a and b). These ~26 and ~20 kDa peptides have been observed previously in anti-PAT western blots (see Figure 4-8) which provides further evidence for processing of the pumpkin 2S(Δ79)-PAT fusion protein indicative of passage through the Golgi-dependant pathway.

4.2.3 PAC-like Vesicle Isolation from CmP Transgenic Arabidopsis

Shimada et al. (2002) [70] described a procedure to isolate and identify PAC vesicles employing Percoll® and sucrose-density gradient purification of vesicle fractions coupled with western blot analysis. Using their methodology, it was shown that the pumpkin 2S(Δ79)-PAT fusion protein isolated from T_2 generation CmP transgenic Arabidopsis was located in the vesicle-enriched fractions (Figure 4-20). A portion of the fusion protein was present in the supernatant fraction as a peptide doublet but the majority of fusion protein was observed in fractions of least density, F1 to F3, and trailing off substantially in fractions of increasing density F4 to F7. This result is typical of PAC vesicles, which are known to be considerably buoyant [70]. The supernatant fraction contained two peptides, ~28 kDa and ~26 kDa, reminiscent of those previously observed in CmP transgenic rosette leaves in which the smaller peptide occurred as a minor species (section 4.1.2, Figure 4-8d); however in this analysis the smaller processed variant was the dominant species in the supernatant fraction. In the buoyant vesicle fractions only a single ~28 kDa peptide was recovered which correlated in size to the intact pumpkin 2S(Δ79)-PAT devoid of its signal peptide (28.4 kDa). This suggested that the sequestering of the fusion protein in PAC-like vesicles rendered it protected from processing and that the lower molecular mass processed variant observed in the supernatant fraction indicated incomplete compartmentalization into PAC-like vesicles and that a minor portion of the fusion protein had been sorted to the PVC/MVB where peptide processing is known to occur (see section 2.4.3). The compartmentalization of the intact fusion protein was further confirmed by

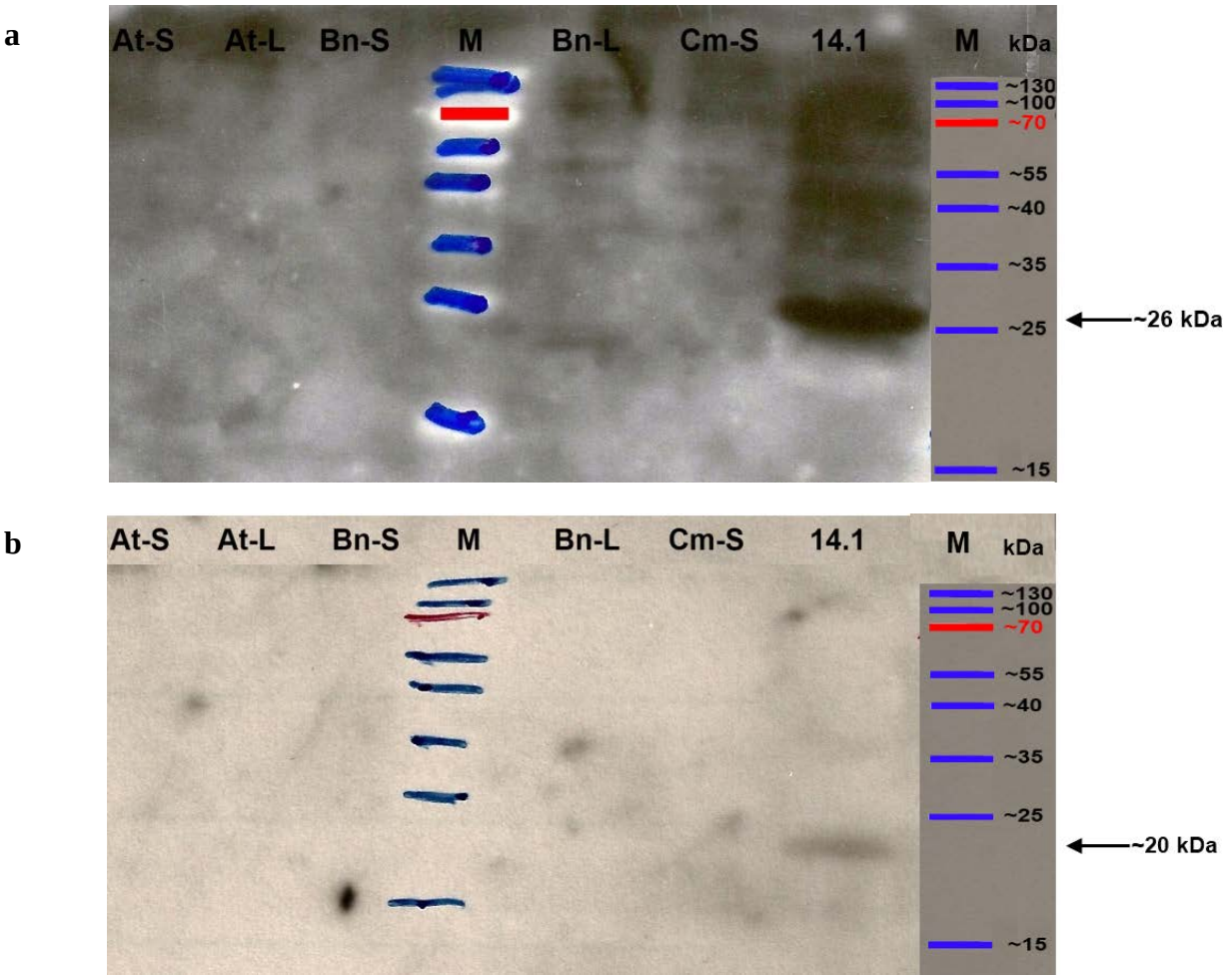


Figure 4-19. Western blot analysis of seed and leaf protein samples extracted from Arabidopsis, *B. napus* and *Cucurbita maxima* (pumpkin) employing antibodies A-CmPP (a) and A-CmSS (b). Key: M, PageRuler™ Prestained Protein Ladder—Fermentas; At-S, *A. thaliana* seed protein; At-L, *A.thaliana* leaf protein; Bn-S, *B. napus* seed protein; Bn-L, *B. napus* leaf protein; Cm-S, *C. maxima* seed protein; 14.1, CmP-14.1 positive control).

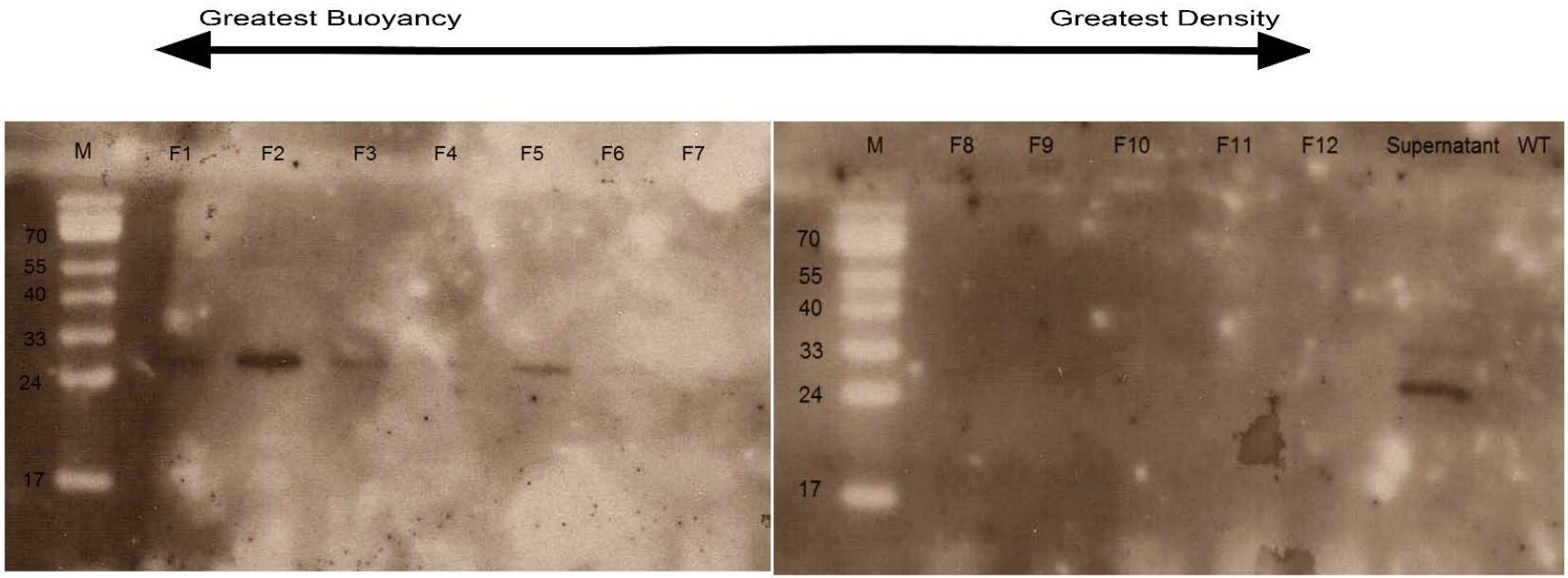


Figure 4-20. Western blot of rosette leaf vesicle fractions isolated from transgenic Arabidopsis CmP-14.1 using anti-PAT antibody. Fractions F1 through F12 were recovered top to bottom following layering of the sample on 28% Percoll® gradient followed by centrifugation at 40,000 g for 35 min, ie. Fraction 1 possessed the greatest buoyancy and Fraction 12 possessed the least buoyancy (greatest density). Intact vesicles, due to their buoyancy, would be contained in top layers of the gradient. Key: M, PageRuler™ Prestained Protein Ladder—Fermentas; WT, Arabidopsis wild type.

examining T_4 generation CmP-18 rosette leaf tissue subjected to a subsequent round of vesicle enrichment by differential centrifugation (Figure 4-21). Although, the yield of intact fusion protein in the vesicle fraction was low in relation to supernatant fractions (SN) and whole leaf protein extract fractions (CF), a subsequent round of vesicle purification (VF2) resulted in the elimination of the smaller ~20 kDa processed variant, whereas it was observed along with the intact ~28 kDa pumpkin 2S(Δ79)-PAT fusion protein in the SN, CF and VF1 fractions. This finding supports the notion that sequestration and stable accumulation of the fusion protein in PAC-like vesicles protects the protein, a claim also made by Hayashi et al. (1999)[19].

4.2.4 Antibody Marker Assisted Characterization of PAC-like Vesicles

To facilitate localization and identification of vesicles that were the recipients of fusion protein deposition, antibodies designed to be used in conjunction with anti-PAT were acquired including antiserum toward the PSV membrane-associated α-tonoplast intrinsic protein (α-TIP (Figure 3-2) [9, 191] and antiserum toward AtRabA4b, a member of the Rab11 family of Rab GTPase proteins that associate with vesicles and regulate trafficking to distinct compartments (see section 2.3.3 and Figure 2-4) [129]. Figure 4-22 shows western blot analysis employing anti-PAT, anti-αTIP and anti-AtRabA4b antiserum of vesicle enriched fractions of rosette leaves of T_2 generation CmP and M(4)P transgenic Arabidopsis plants partitioned by 28% Percoll® gradient centrifugation and collected from least to most dense (Fractions 1 through 12). Although no accumulation was detected for any fraction for the M(4)P transgene, the majority of the anti-PAT reactive peptides occurred in the most buoyant fractions 1-3 of leaf samples from CmP 14, 16 and 18 plants and tapered off quickly in more dense fractions (Figures 4-22a-c, note fractions 4 through 8 are not shown as no reactive peptides were detected). The intact ~28 kDa pumpkin 2S(Δ79)-PAT fusion protein was the dominant species with the smaller ~26 kDa processed

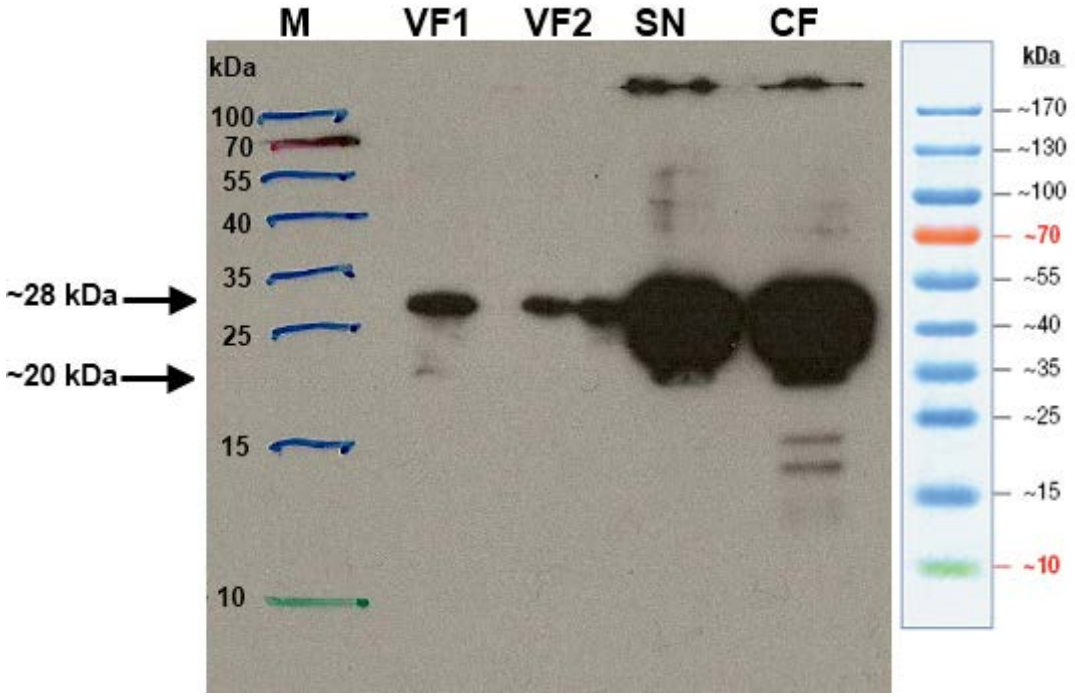


Figure 4-21. Western blot of rosette leaf protein fractions isolated from T_4 generation CmP-18.1.2.3 transgenic Arabidopsis using anti-PAT antibody. One round of vesicle enrichment employing the differential centrifugation methodology of Shimada et al. (2003) [69] yielded fraction VF1 which when subjected to a second round of enrichment yielded fractions VF2 and a remaining supernatant fraction SN. These fractions were compared to whole leaf protein extract, CF. Key: M, PageRuler™ Prestained Protein Ladder—Fermentas; WT, Arabidopsis (var. Columbus) wild type; +ve, CmP 14.1 positive leaf protein extract.

variant appearing only in the densest fraction (Fraction 9, Figure 4-22d). In the fraction of greatest density (Fraction 9) of CmP 18 leaf tissue the anti-αTIP and anti-AtRabA4b antiserum reacted with proteins in excess of 100 kDa. This did not correlate with the expected size of α-TIP (268 aa) of ~30 kDa or the expected size of AtRabA4b (224 aa) of ~24 kD which suggested that targets of these antisera occurred as large sedimented aggregates. The behavior of the membrane α-TIP protein to form aggregates and migrate aberrantly in SDS-PAGE has been observed by others and can be remedied by not boiling the protein in Laemmli buffer or by the use of Tricine-SDS gel electrophoresis. Regardless, that they occur together within the same fraction as the ~26 kDa pumpkin 2S(Δ79)-PAT processed variant, implied an association between proteins.

4.2.5 Antibody Marker Assisted Isolation of PAC-like Vesicles

The preliminary evidence for an association between the pumpkin 2S(Δ79)-PAT fusion protein and vacuolar membrane targets anti-αTIP and anti-AtRabA4b, led to the development of a strategy aimed at obtaining a highly enriched vesicle/vacuole fraction. It was thought that employing the discriminating capacity of protein associations to recover PAC-like vesicles and their accumulated fusion protein contents was an alternate and superior strategy over isolating vesicles indiscriminately by differential centrifug-ation and looking for protein associations following their disruption. Furthermore, PAC-like vesicles isolated by this manner would exist in pure form and could be subjected to MudPIT analysis for characterization of component PAC-like vesicle proteins. Therefore, the α-TIP and AtRabA4b antiserum were employed in an antibody capture strategy whereby the antibodies were conjugated to Dynal® Dynabead M-450 magnetic beads. Employing anti-αTIP-conjugated beads in this manner resulted in a vesicle fraction from the same T_3 generation CmP-18.1.2 leaf samples showing anti-PAT reactivity (Figure 4-23a). Despite release of antibody from conjugation beads which caused gel overloading and considerable background of high molecular weight protein reactive

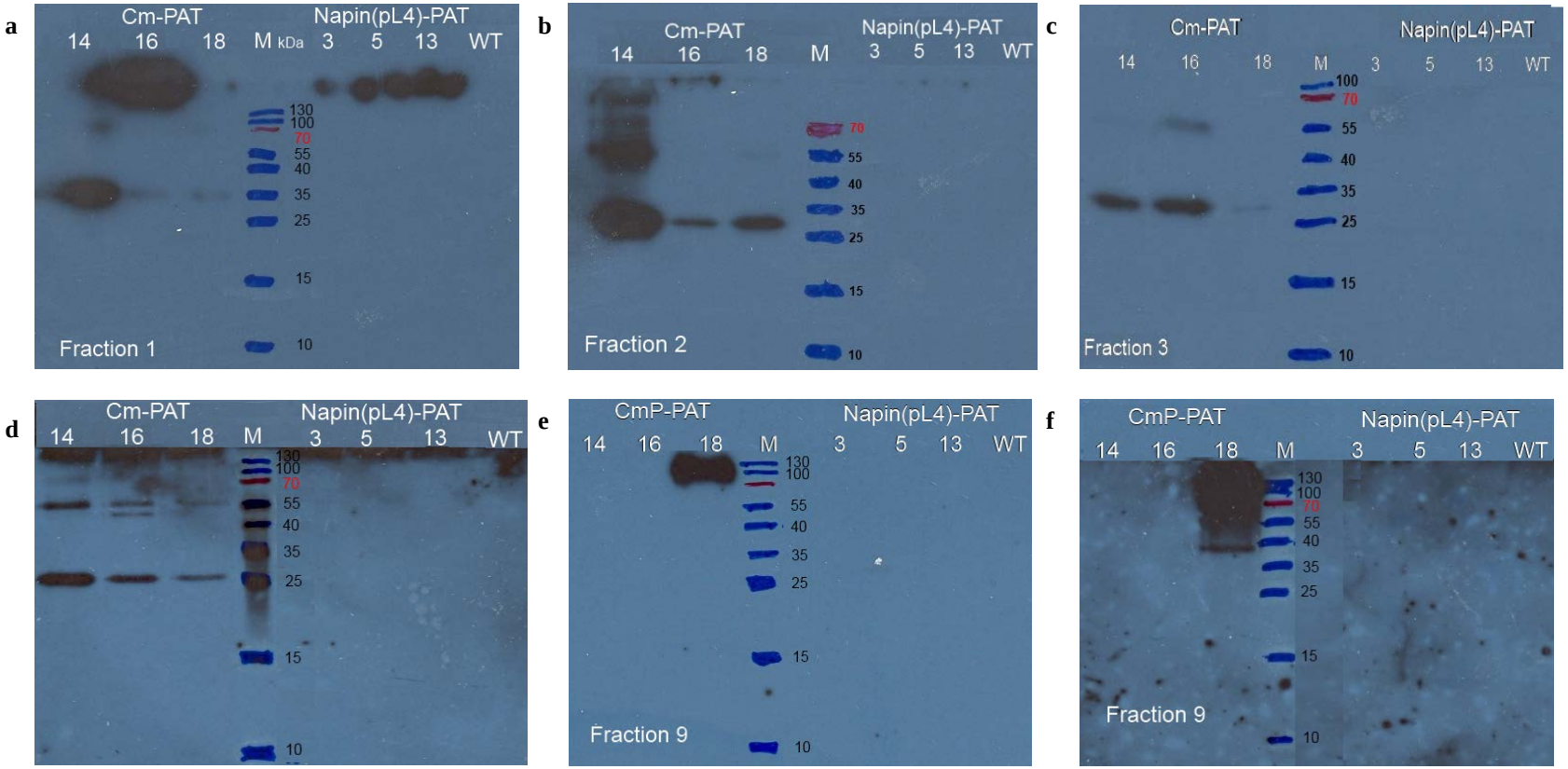


Figure 4-22. Western blot analysis of vesicle enriched rosette leaf protein extracts fractionated by 28% Percoll® ultracentrifugation (40,000 g, 35 min). Recovered fractions were tested using (a-d) anti-PAT, (e) anti-αTIP and (f) anti-AtRabA4b antibodies. Key: M, PageRuler™ Prestained Protein Ladder—Fermentas; WT, Arabidopsis wild type.

with the anti-rabbit secondary antibody (anti-PAT signal in excess of 55 kDa), an anti-PAT reactive peptide was observed which correlated in size to a processed pumpkin 2S(Δ79)-PAT variant (~20 kDa). Enrichment employing anti-AtRabA4b-conjugated beads resulted in seven of eight T_3 generation CmP-14 leaf extracts indicating the presence of the anti-PAT peptide doublet with each being of equal intensity or the smaller processed variant being more plentiful than the other (Figure 4-23b). It was clear from the silver stained gels that copious non-specific binding by the antibody-bead conjugates recovered huge populations of proteins in the final eluate. Steps were made to improve specificity including modification of the elution regimen by using a 50 mM glycine elution buffer, addition of 5% skim milk as a blocking agent to increase binding stringency and using vesicle fractions enriched by differential centrifugation as a starting point. These strategies did substantially reduce the release of conjugated antibody and non-specific binding of proteins in the final eluate but the increased binding stringency came at the expense of detectable anti-PAT reactive peptides. It was thought that further optimization of the technique involving the use of different magnetic beads and a mouse antiserum directed against a peptide region of PAT would remedy these limitations. Consequently the MudPIT analysis planned for the purified vesicle fractions was abandoned.

4.2.6 Examination of Leaf Ultrastucture of GFP-transgenic Arabidopsis

To augment the findings obtained with the PAT gene fusion constructs, fusion protein localization at the cellular level by use of an in-house Zeiss Apitome microscope was conducted for the GFP gene fusion constructs employing transient expression in tobacco and stable transgenic Arabidopsis plants.

Tobacco transient expression assay of GFP fusion constructs. As described previously, PCR confirmed positive transgenic Arabidopsis plants were obtained only for constructs N(B)GFP, N(X)GFP and N(FL)GFP and none were obtained for N(A)GFP and N(S)GFP (see Figure 4-7).

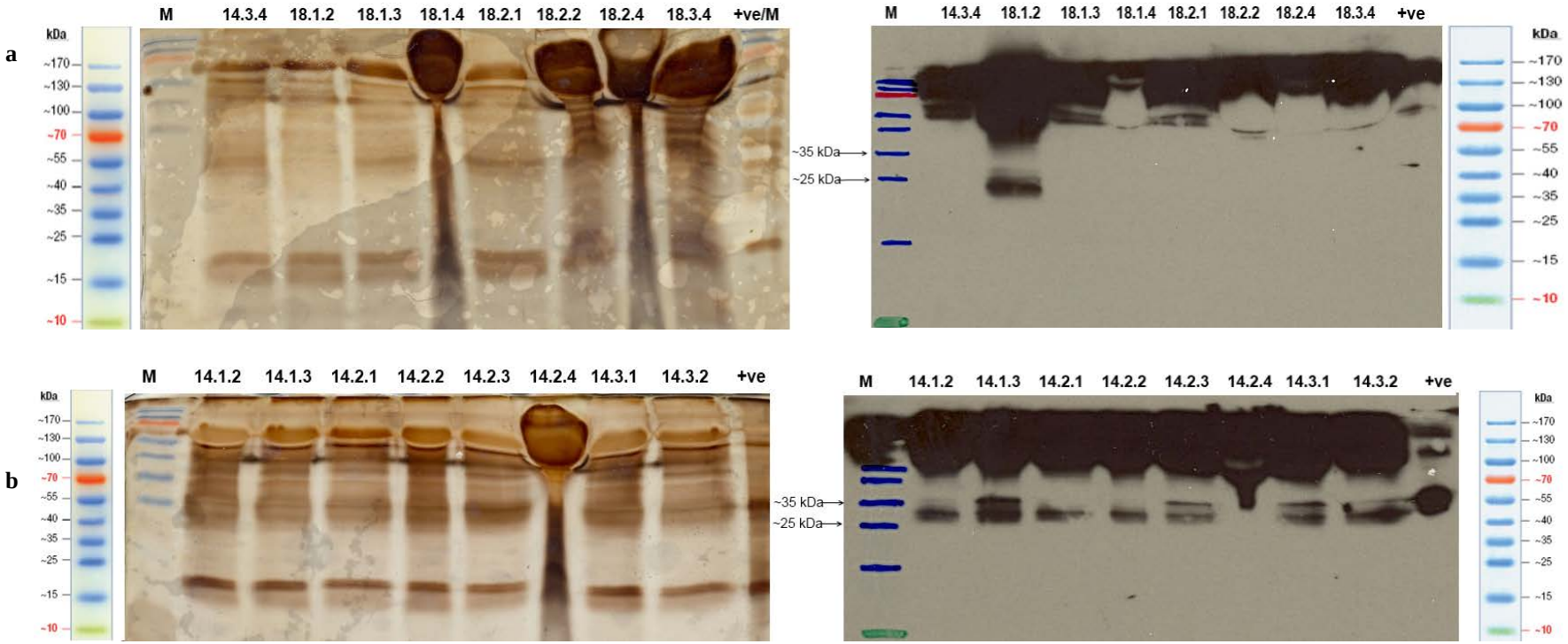
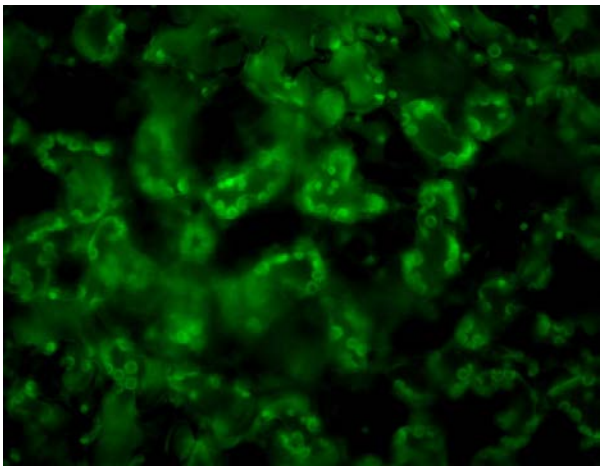


Figure 4-23. Western blot of vesicle content using anti-PAT antiserum following antibody capture of vesicles employing (a) anti-αTIP-conjugated beads and (b) anti-AtRabA4b-conjugated beads on rosette leaf protein extracts of T_3 generation CmP-14 and CmP-18 plants. Key: M, PageRuler™ Prestained Protein Ladder—Fermentas; +ve, CmP 14.1 positive leaf protein extract.

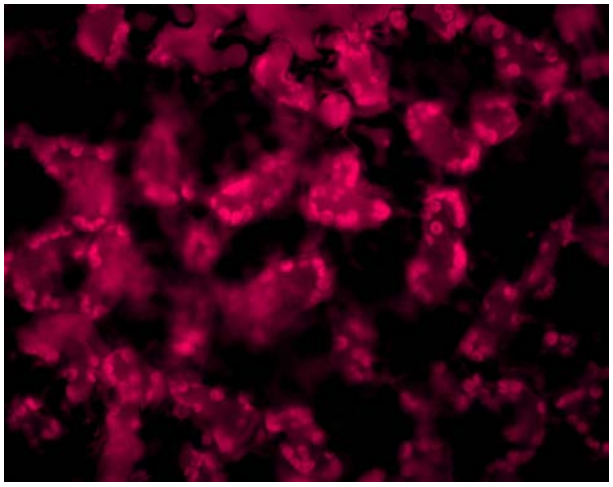
Furthermore, western blot analysis of protein extracts from rosette leaves of N(B)GFP, N(X)GFP and N(FL)GFP plants and their T_2 progeny using anti-GFP antibody resulted in no observed accumulation (see section 4.1.2). To confirm that the various GFP fusion constructs were capable of expressing the expected fusion proteins, these constructs plus N(B)GFPCT and CmGFP were tested in a tobacco transient expression assay. Following infiltration of tobacco leaves and dark treatment for 48 h, expression of constructs N(X)GFP, N(B)GFPCT, and the CmGFP resulted in leaf cells possessing punctate fluorescent globules which migrated rapidly by cytoplasmic streaming. Although nothing was observed for the transient expression of constructs N(B)GFP, N(A)GFP, N(FL)GFP, and N(S)GFP, this result confirmed that at least the N(X)GFP construct was functional for plant transgene expression and proper protein folding of the fusion protein had occurred, indicating some other factor prevented the accumulation of fusion protein in stable transgenic Arabidopsis.

Microscopic Examination of Napin-GFP Transgenic Arabidopsis. Among the various Napin-GFP fusion protein constructs created and used to transform Arabidopsis, only the mNap(B)-GFP-CT open reading frame resulted in a stable accumulation of GFP in leaf tissue sufficient for visualizing by microscopy (Figure 4-24). Panels 4-24 a) and b) show GFP and chlorophyll channels, respectively, of leaf spongy mesophyll following 48 hour dark treatment of T_3 generation transgenic Arabidopsis expressing the mNap(B)-GFP-CT open reading frame. Merging GFP and chlorophyll channels showed small punctate vesicle structures.

a



b



c

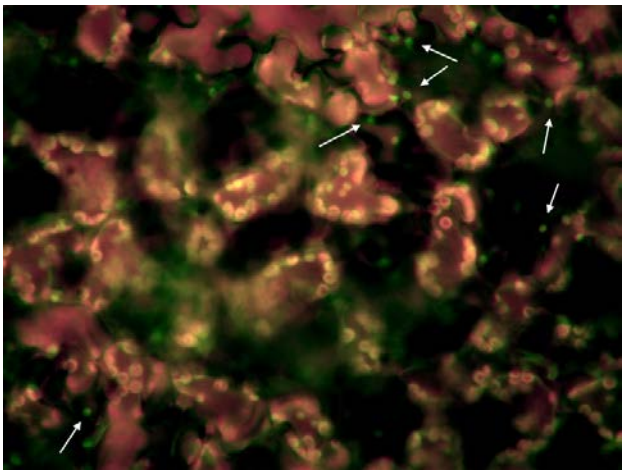


Figure 4-24. Leaf mesophyll cells of 48 hour dark-treated T_3 generation N(B)GFPCT-4.2 transgenic Arabidopsis using 40X water immersion lens of a Zeiss Apitome microscope showing the (a) GFP, (b) chlorophyll and (c) GFP + chlorophyll channels. The combined channel (c) indicates small punctate vesicular structures (white arrows).

4.3 SSP Processing and Trafficking in Leaf Tissue—Microarray Study

The development of Arabidopsis plants capable of ectopic expression of transgenes that resulted in vesicle biogenesis and accumulation of fusion protein was central to profiling gene expression necessary for SSP trafficking mechanisms. The use of leaf tissue as experimental material together with the use of an inducible transgene expression system was thought to simplify the study of the genes involved by potentially evoking gene expression and recruiting essential components not typically active in vegetative tissues. The intricate dexamethasone inducible expression system developed by Ian Moore (Craft et al., 2005) [192, 193] which allows for stringent regulation of "on" and "off" states was chosen (Figure 4-25). This system controlled expression of SSP fusion constructs precisely and allowed identification of genes activated both at the onset of fusion protein synthesis and during fusion protein accumulation.

4.3.1 Testing Dexamethasone Inducible Constructs

The pHTOP dexamethasone inducible constructs were used to transform the "activator" LhGR-N plant line harboring the constitutively expressed LhGR-N transcription factor (Figure 4-25). The double transformed lines so created were screened by PCR, advanced to the T_2 generation and tested for induction by dexamethasone. Initially, there was some uncertainty as to how best to administer dexamethasone to facilitate the study of fusion protein accumulation behavior. Moore et al. (2005) reported dexamethasone induction on seedlings cultivated on ½ MS containing 10 μM dexamethasone or subterranean irrigation of 4 week-old plants with 20 μM dexamethasone [192, 193]. To explore this, T_2 generation CmP:pHTOP seedlings were sown on ½ MS media with or without 10 μg dexamethasone and 20 μg/mL PPT (Figure 4-26). Because none of the seedlings survived exposure to PPT, it was thought that

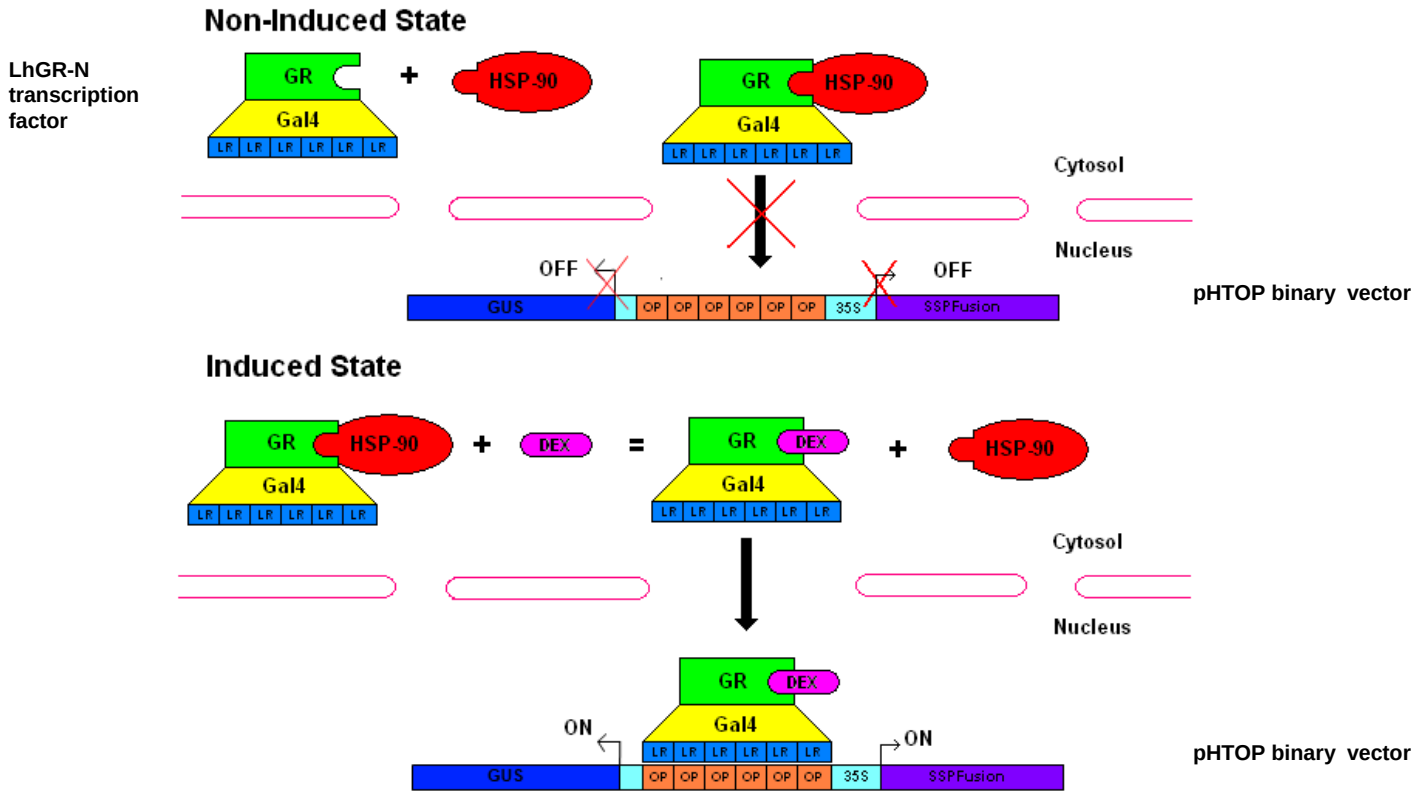


Figure 4-25. The dexamethasone inducible expression system developed by Moore et al. [187, 188]. Shown above is the effect of the glucocorticoid DNA binding/activation fusion protein (LhGR-N) on the inducible expression transgene. In the non-induced state (dexamethasone absent), the constitutively expressed LhGR-N transcription factor is prevented from entering the nucleus by its association with heat-shock protein 90 (HSP-90). In the induced state (application of dexamethasone), the glucocorticoid receptor (GR) preferential binds to dexamethasone. The exchange of HSP-90 for dexamethasone permits entry of the LhGR-N complex into the nucleus where it binds to its 6X *lac* operator target to initiate gene expression. Key: Gal4, *S. cerevisieae Gal4* transcription activation domain; OP, *E. coli lac* operon promoter element; LR, *E. coli lac* repressor protein; DEX, dexamethasone.

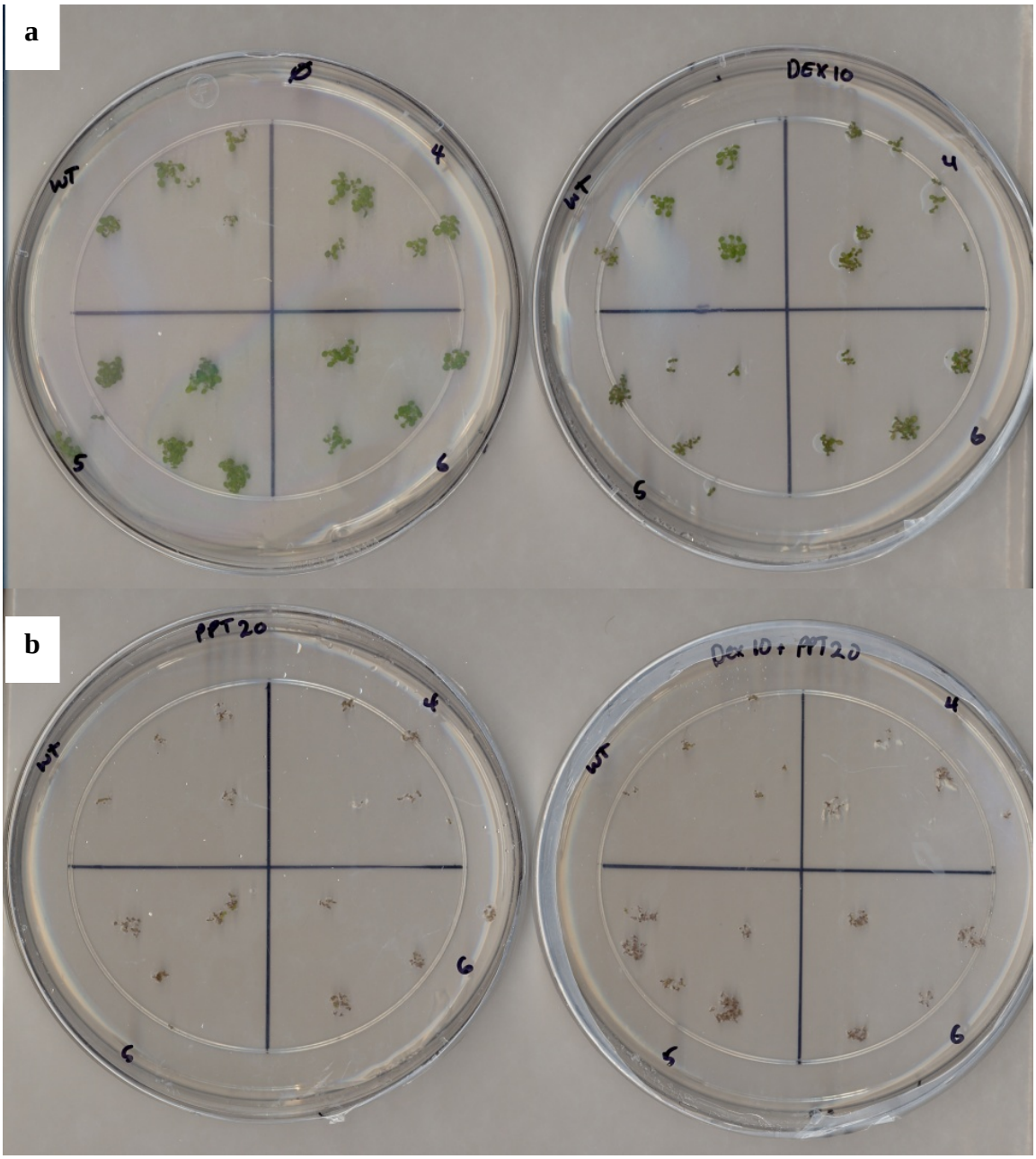


Figure 4-26. Cultivation of T_2 generation Arabidopsis harboring the CmP:pHTOP and the pLhGR-N activator transgenes on ½ MS medium containing (a) 0 or 10 µg dexamethasone and (b) 0 or 10µg dexamethasone plus 20 µg/mL PPT.

administering dexamethasone in this manner could not sufficiently induce the CmP transgene in either degree or timeliness necessary for a gene activation study and consequently, dexamethasone induction of seedlings was not pursued further. As alternate strategies, two other methods were devised which employed ½ MS broth containing 20 μM dexamethasone: 1) liquid culture of whole-plants removed from soil, partially immersed in broth and placed upright in racks, and 2) liquid culture of leaf explants. Neither of these liquid culture methods was utilized in the final dexamethasone inductions employed for the microarray because it was deemed necessary to minimize gene effects resulting from atypical growing conditions; however these techniques served as a convenient way to monitor the effects of dexamethasone induction and track the development of T_3 generation transgenic plants possessing consistent induction behavior. Pooled rosette leaves (~1 cm × 2.5 cm in size) from induced whole plants were analyzed by RT-PCR for CmP and GUS reporter transgene expression (Figure 4-27). Transgene expression was indicated when use of CmP and GUS specific primers (columns “a” and “b”, respectively) in RT-PCR amplifications resulted in bands of 246 bp and 159 bps, respectively, without product generated in DNA contamination controls using Platinum Taq® and GUS specific primers (column “c”). Expression of the CmP transgene was observed for all samples except pooled sample 2.4; however, concomitant expression of GUS was indicated only for plant samples of T_1 lineage 3, 4, 5 and 6. PCR analysis of T_3 generation plants derived from parent CmP:pHTOP-3.2 confirmed presence of the transgene in seven of eight analyzed (863 bp amplicon, Figure 28a). Analysis of these T_3 pants by western blot of vesicle enriched fractions from rosette leaves incubated for 3 days in liquid culture revealed numerous anti-PAT reactive peptides (Figure 4-28b).

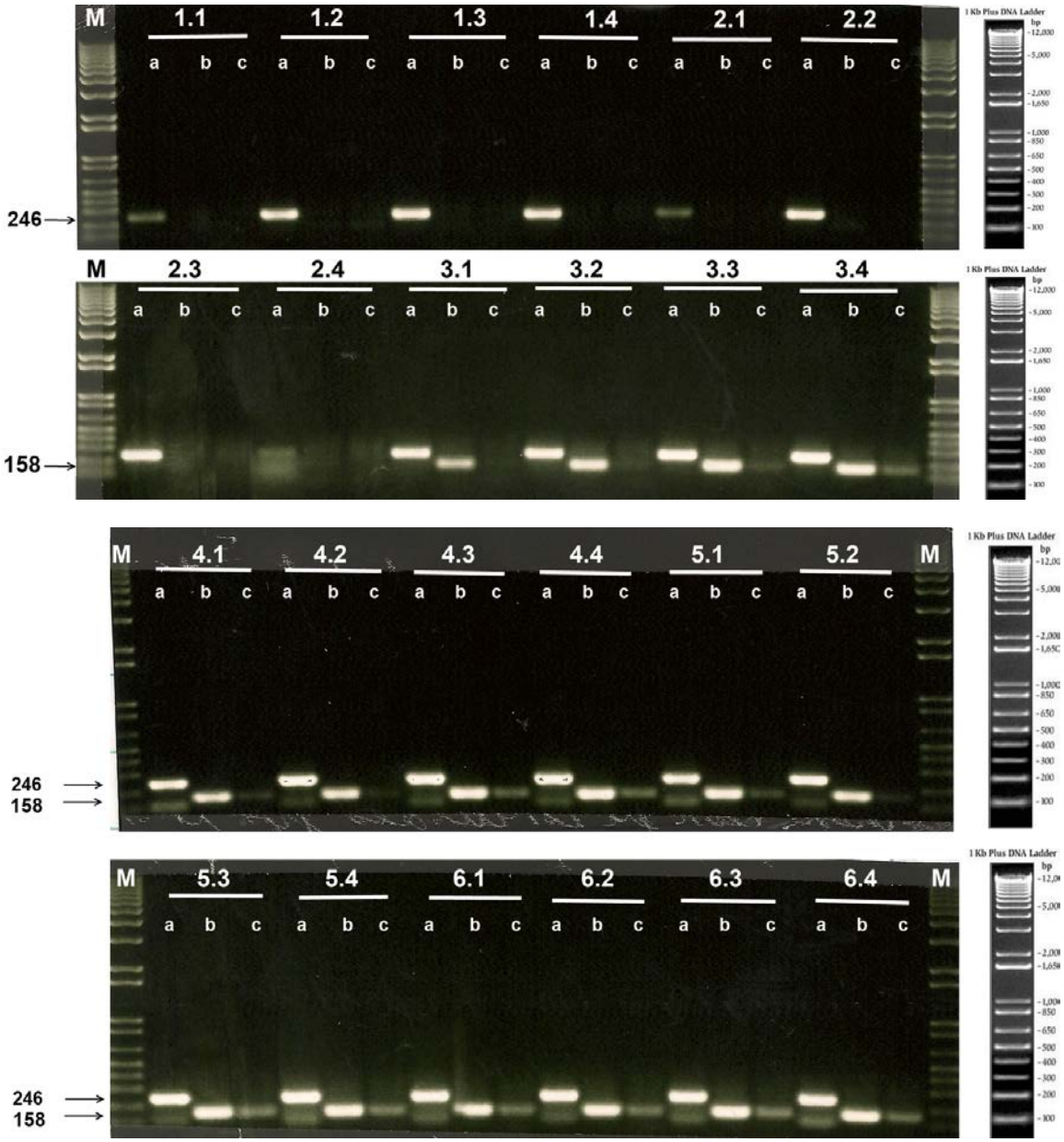


Figure 4-27. RT-PCR expression analysis of T_2 generation CmP:pHTOP transgenic Arabidopsis plants following induction by partial immersion of the plant in ½ MS broth + 20 μM dexamethasone. Expression of the CmP transgene is indicated by the 246 bp amplicon from the use of Cm-F1 and CmPJ-R primers in RT-PCR reactions labeled "a" and expression of the GUS reporter transgene (158 bp) by use of GUS-F10 and GUS-R12 primers in RT-PCR reaction labeled "b" whereas columns labeled "c" are their respective GUS-F10/GUS-R12 PCR controls. Samples consist of total RNA extracted from pooled rosette leaves from individual T_2 generation plants. Key: M, 1 kb plus DNA ladder® (Invitrogen).

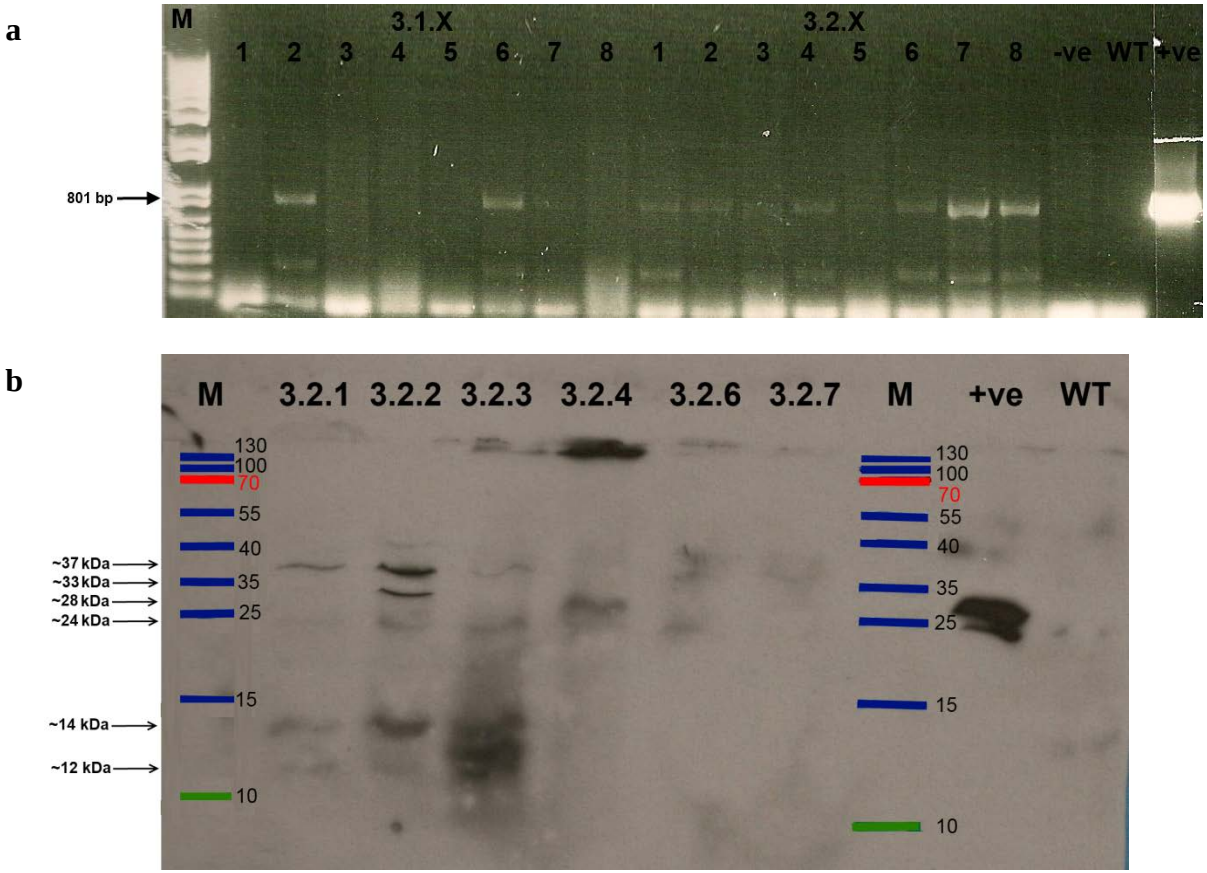


Figure 4-28. a) PCR analysis of T_3 generation transgenic Arabidopsis harboring the CmP:pHTOP transgene as indicated by the 801 bp amplicons generated by employing primers C1 and P1; b) western blot analysis of T_3 generation CmP:pHTOP-3.2 transgenic Arabidopsis induced by liquid culture. Samples are comprised of four rosette leaves from each plant cultured in ½ MS broth + 20 µM dexamethasone for three days followed by vesicle enrichment extraction and western analysis employing anti-PAT antiserum. Key: M, PageRuler™ Prestained Protein Ladder—Fermentas; WT, Arabidopsis wild type; +ve, CmP 14.1.3 positive control.

Although the protein yields from these vesicle fractions of dexamethasone induced leaf explants were low, the ~ 28 kDa intact pumpkin2S(Δ79)-PAT fusion was indicated in sample 3.2.4, the ~26 kDa processed variant was indicated in all samples except 3.2.7, and in samples 3.2.1, 3.2.2 and 3.2.3 lower molecular mass processed variants of ~12 kDa and ~14 kDa were observed (see also Figures 4-8 and 4-18b, sample 14.2.3.6). Samples 3.2.1, 3.2.2 and 3.2.3 also possessed an unexpected anti-PAT reactive peptide of ~37 kDa, and for sample 3.2.2 an ~33 kDa peptide as well. It is unclear, but a possible explanation for the presence of these peptides might be due to glycosylation events resulting from liquid culture induction by dexathesone. Regardless, because sample 3.2.4 was the only plant showing evidence for accumulation of the ~28 kDa intact fusion protein, T_4 generation CmP:pHTOP-3.2.4 plants were chosen for a pilot microarray study.

4.3.2 Pilot Microarray Experiment

Because of the scope of the final microarray experiment in terms of time and resources, it was thought that in addition to the preliminary work described above, a pilot microarray experiment was necessary. For this purpose 4-week old T_4 generation CmP:pHTOP-3.2.4 plants were subjected to induction according to the method of Craft et al. (2005) [192] (described in section 3.8). Harvesting of rosette leaves over the time course of dexamethasone application and monitoring the induction by GUS reporter staining revealed successful but varying transgene expression (Figures 4-29). Pooled samples of leaves of the same day, extracted for vesicle fractions and analyzed by western blot employing anti-PAT antiserum revealed a dose-dependent accumulation of an ~12 kDa peptide (Figure 4-30). Although the intact fusion protein was anticipated, the presence of the dose-dependent accumulation of the previously observed ~12 kDa peptide processed variant was encouraging. These same tissue samples for days 0, 1, 4 and 7 were extracted for total RNA, used for amplified RNA (aRNA) synthesis and labeling for hybridization to Affymetrix custom AGRONOMICS1 transcriptome profiling arrays [194]. The

gene expression resulting from the induction having greater than two-fold change (FC>2) for comparisons between day 1, 4 and 7 and the non-induced control, day 0 is reported in Figures 4-31 to 4-33. The results of this analysis identified several candidate genes that appeared to respond to dexamethasone in a dose-dependent fashion over the 7-day time course; however, several genes having a significant increase in expression were not represented at all three times. From this pilot study, the importance of suitable controls and the value of genes having decreased expression were realized. Furthermore, the pilot study indicated that identification of genes observed after 1-2 days induction are the most likely candidates to be involved in vesicle biogenesis and accumulation as the fusion protein had begun to accumulate in vesicles by day three and that the greatest fold-changes were observed for day 1, in particular a gene encoding C1A cysteine-type peptidase (At2g27420). Because At2g27420 had its greatest expression on day 1 and did not diminish appreciably by day 7, it was thought that beyond day 2, a prolonged exposure of the fusion protein to this protease resulted in no accumulation of the fusion protein except for the ~12 kDa processed variant.

4.3.3 Large-scale Microarray Study

Rationale. Based on the findings of the pilot microarray experiment, it was understood that: 1) despite evidence for accumulation of a process variant instead of the intact pumpkin2S(Δ79)-PAT fusion protein, the dexamethasone inducible system would allow for the discovery of genes necessary for seed protein trafficking machinery, 2) expression of candidate genes would be detectable during the early stages of induction, from day 1 to day 3, 3) accurate identification of these candidate genes would require a series of carefully designed controls, 4) a sufficient quantity of biological replicates would need to be employed and, 5) a sampling regimen designed to minimize environmental effects would be necessary. To not introduce additional potential variation, it was thought necessary to conduct the experiment within the

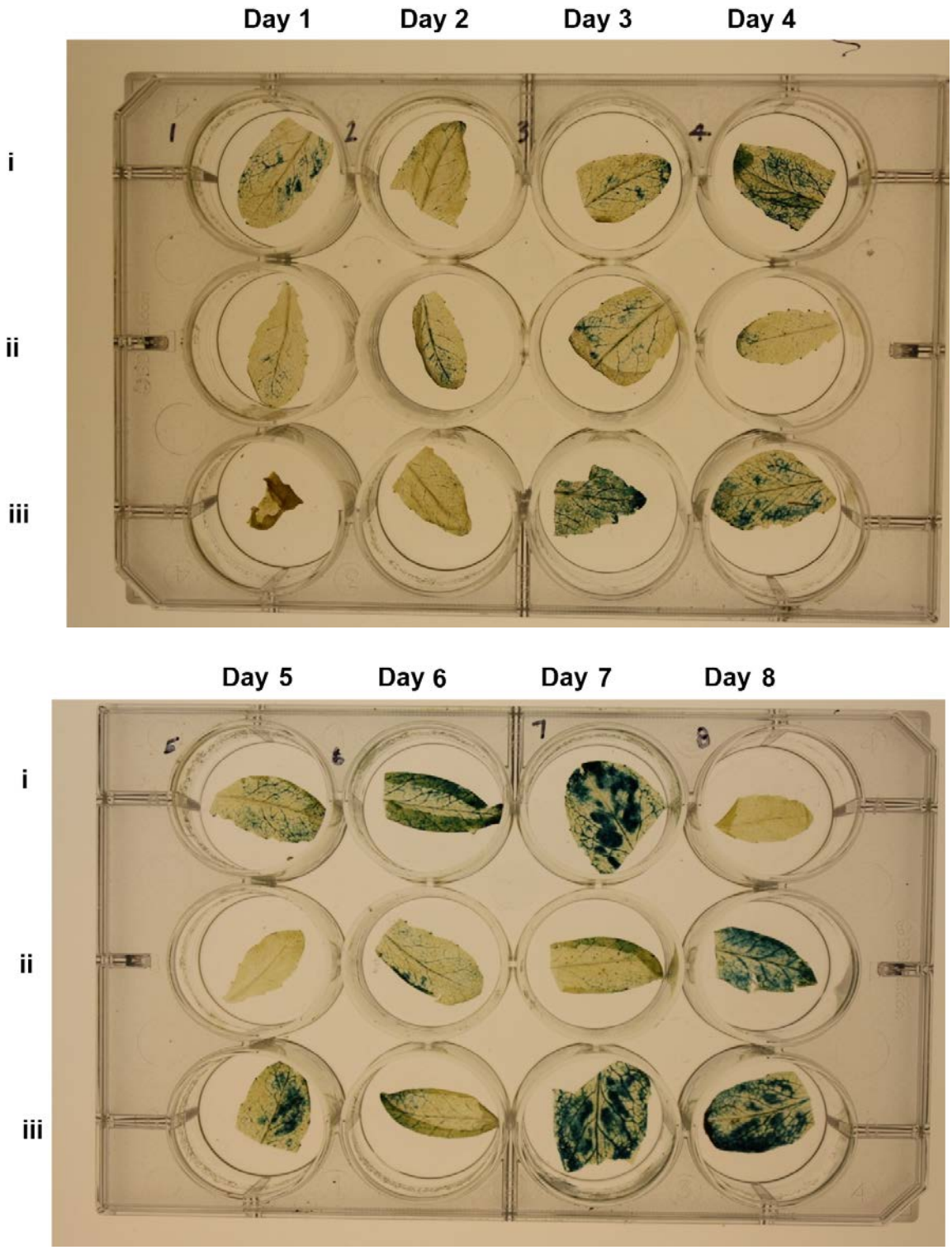


Figure 4-29. GUS staining analysis of T_4 generation CmP:pHTOP-3.2.4 plants induced by subterranean irrigation in 300 mL tap water containing 20 µM dexamethasone on day 0, 3 and 6. For each of days 1 through 8, leaf samples were taken from each of three individual plants (i, ii, iii) and incubated in GUS staining reagent for 24 h at 37˚C (shown in rows 1 to 3 for each "day" column).

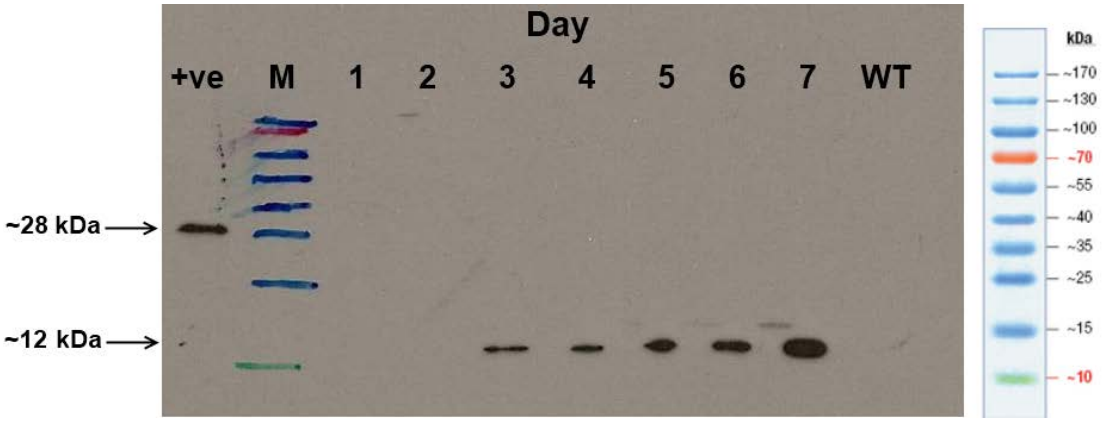


Figure 4-30. Western blot analysis using anti-PAT antibody of T_4 generation CmP:pHTOP-3.2.4.x plants induced by subterranean irrigation employing 300 mL tap water containing 20 μM dexamethasone on day 0, 3 and 6. For each of days 1 through 7, leaf samples were taken from each of seven individual plants, pooled, subjected to vesicle enrichment extraction and analyzed by western blot using anti-PAT antiserum. Key: M, PageRuler™ Prestained Protein Ladder—Fermentas; WT, Arabidopsis wild type; +ve, CmP 14.1.3 positive control.

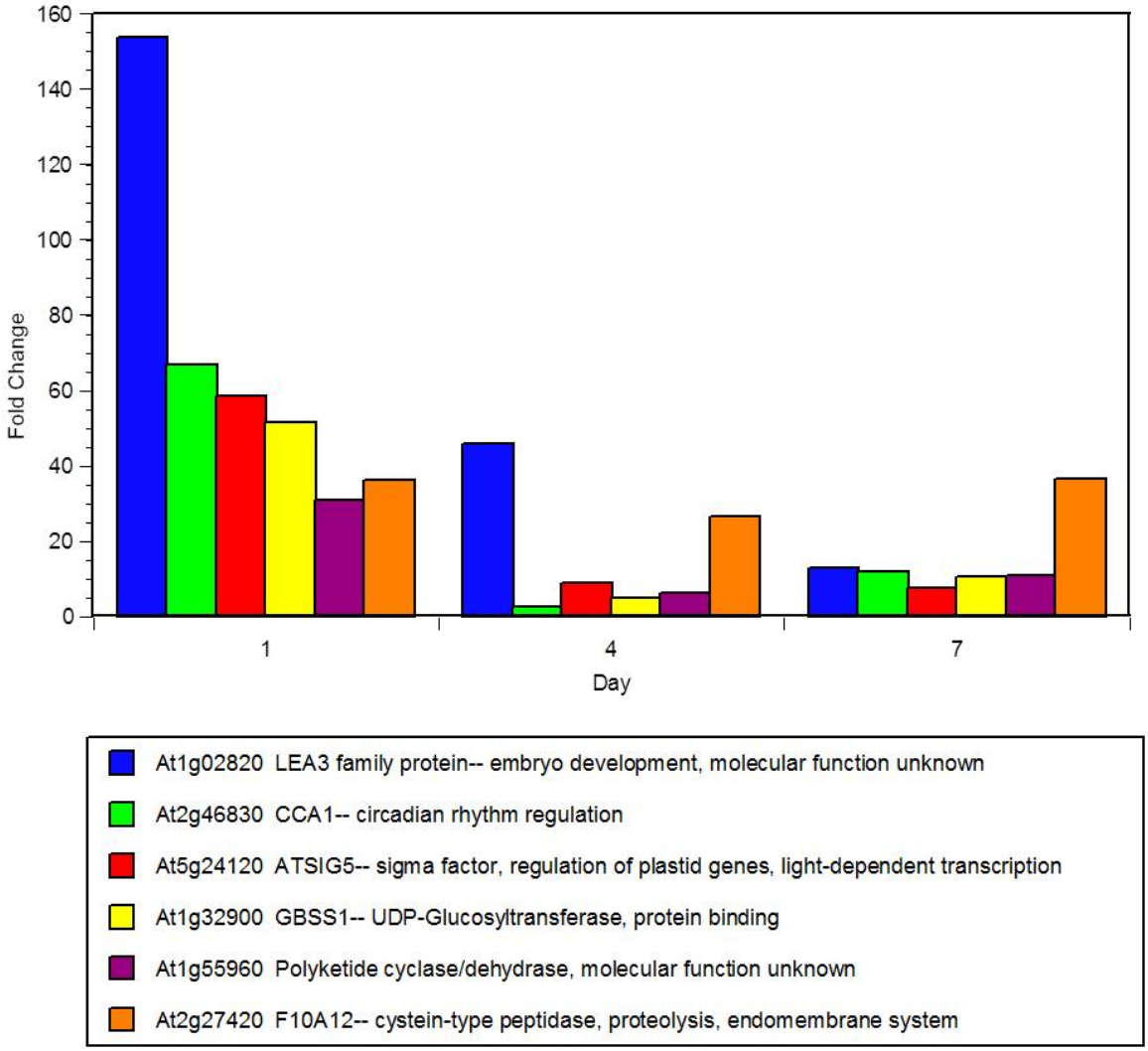


Figure 4-31. Pilot Microarray—Progression of gene expression of genes having the greatest fold change following 1 day of 20 μM dexamethasone induction.

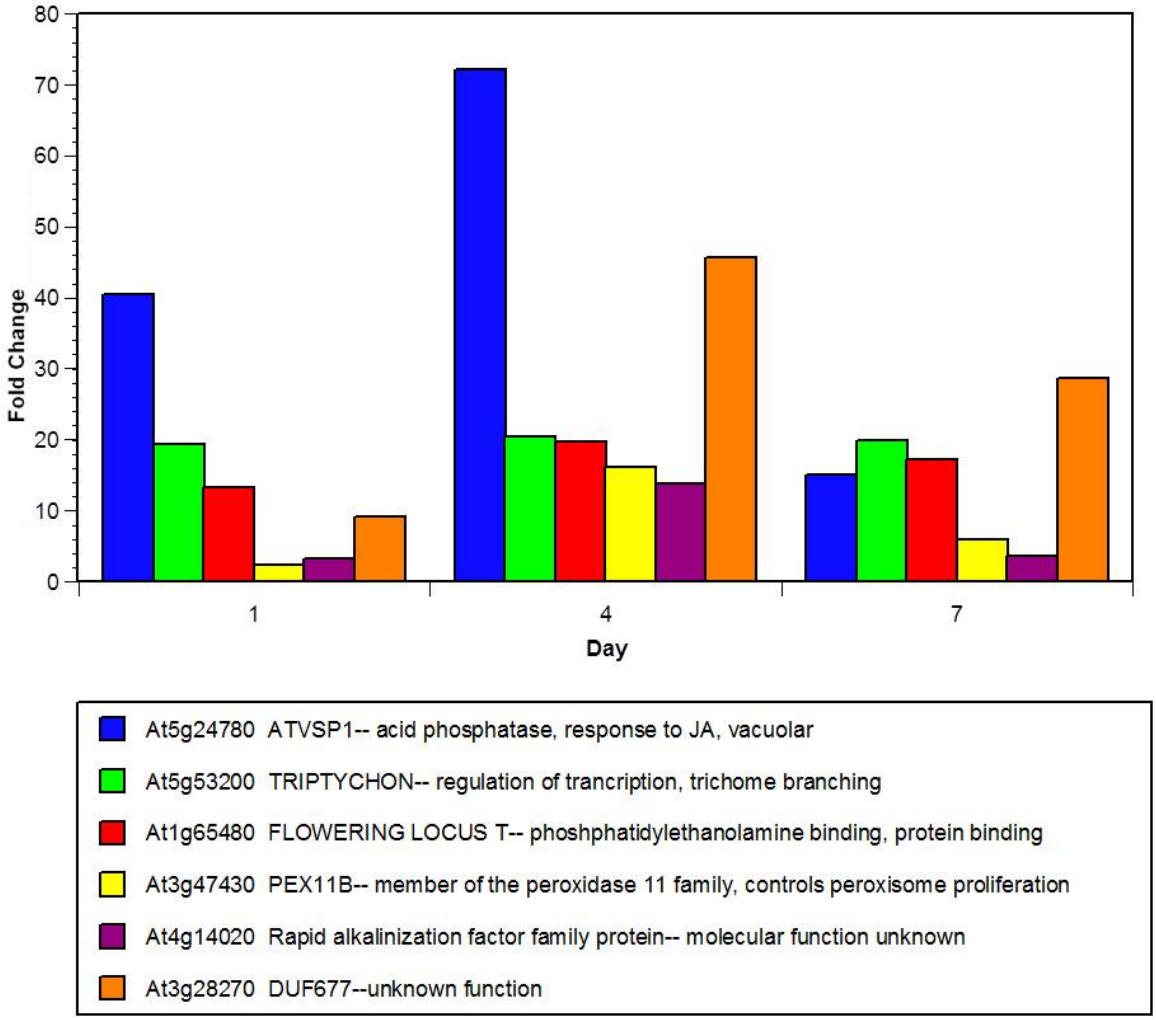


Figure 4-32. Pilot Microarray—Progression of gene expression of genes having the greatest fold change following 4 days of 20 μM dexamethasone induction.

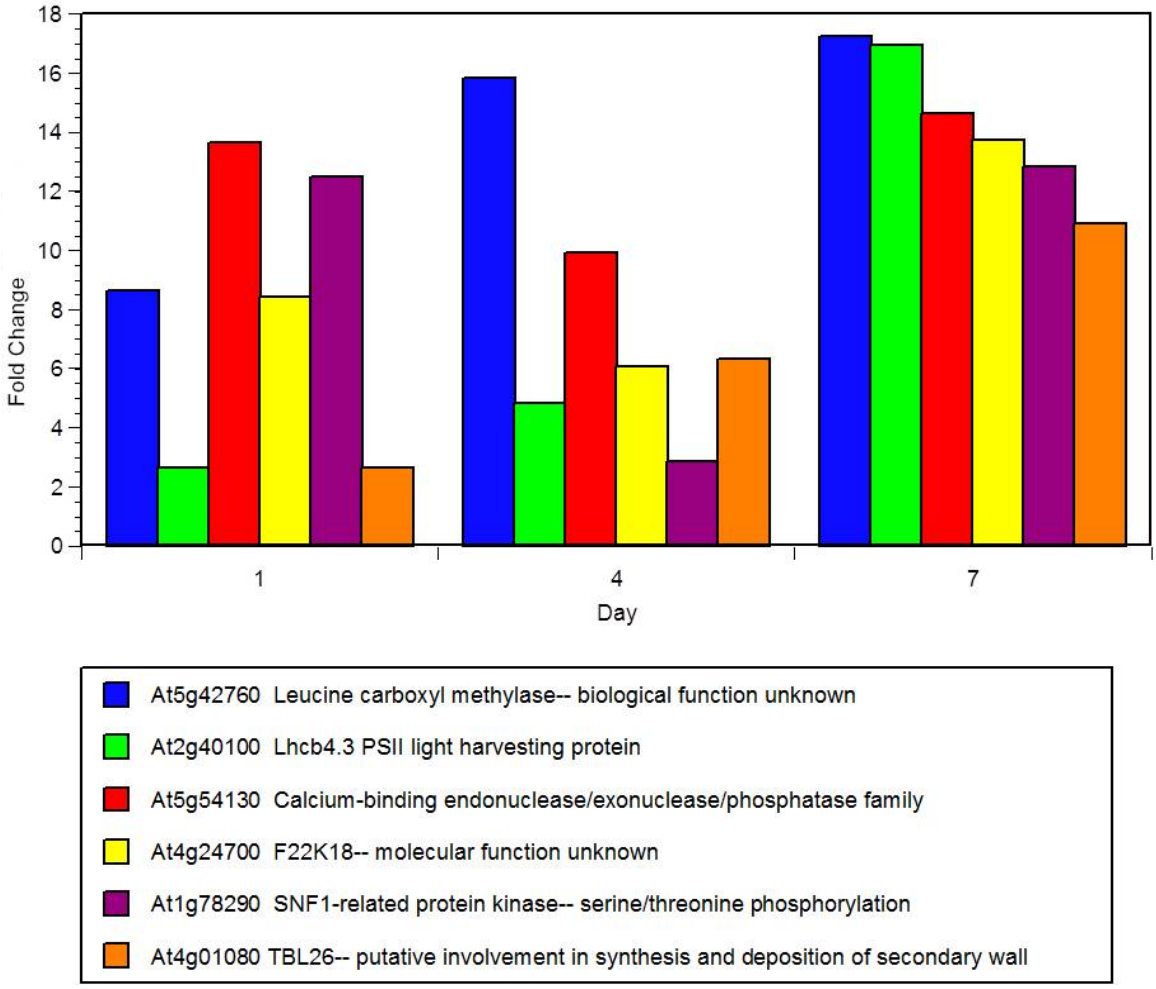


Figure 4-33. Pilot Microarray—Progression of gene expression of genes having the greatest fold change following 7 days of 20 μM dexamethasone induction.

limitations of one lot of AGRONOMICS1 custom microarrays, i.e. no more that 50 hybridizations, the maximum lot size. To accommodate this, the rationale for choice of control constructs for the large-scale study was reworked. As discovered by continued examination of the CmP transgenic plants, the capacity for the pumpkin 2S(Δ79)-PAT gene fusion to evoke accumulation behavior and vesicle biogenesis was not consistent even within a population of T_4 generation plants (see section 4.2). The issue of inconsistent stable accumulation was not an impediment to the microarray study; however, the aim was to employ plants that stably accumulated the fusion protein at least some of the time such that the genes and proteins involved in those instances of accumulation could be discovered. Therefore, the mNap(pL4)-PAT open reading frame found unlikely to evoke accumulation behavior which was to serve as a useful negative control was omitted. Instead several groups of T_4 generation CmP:pHTOP plants were compared among themselves to identify potential correlations between their observed accumulation behavior and concomitant gene expression. In this way, CmP:pHTOP plants indicating differing accumulation behavior, processing or vesicle genesis as detected by western blot analysis could be compared by gene expression profiling. Also, the inducible PAT (no SP) control plants were used as controls instead of those harboring the empty vector as it was thought that gene activation attributed to expression of non-secreted PAT could comprehensively account for the effect of simultaneous GUS reporter expression, presence of active PAT enzyme in the cytosol, as well as the effects of actual dexamethasone application on the plant. Furthermore, gene discovery using inducible expression was augmented by the inclusion of plants harboring the CaMV 35S driven pumpkin 2S(Δ79)-PAT open reading frame, namely T_4 generation CmP-14.1.3 and 18.1.2 , and wild-type Arabidopsis. This was done so that gene effects of constitutive

transgene expression and its resulting long-term accumulation of fusion protein could be compared to the effects of inducible expression.

Dexamethasone treatments and monitoring. As conducted for the pilot study, employing dexamethasone induction used by Craft et al. (2005) [192], 4-week old T_4 generation plants (detailed in section 3.8 and Table 3-1), plus wild-type Arabidopsis (Columbia ectotype) were grown and watered with 300 mL 20 μM dexamethasone on days 0, 2, 4 and 6. Monitoring over the time course of dexamethasone application by GUS staining indicated induction had occurred but not as strongly as was observed previously in the pilot study (Figure 4-34). In fact, despite positive GUS staining, few samples from T_4 generation plants indicated GUS transcript and accumulation of a ~12 kDa anti-PAT reactive peptide, which in this induction appeared to positively correspond with the frequency of dexamethasone application on days 0, 2, 4 and 6 (Figure 4-35a-d).

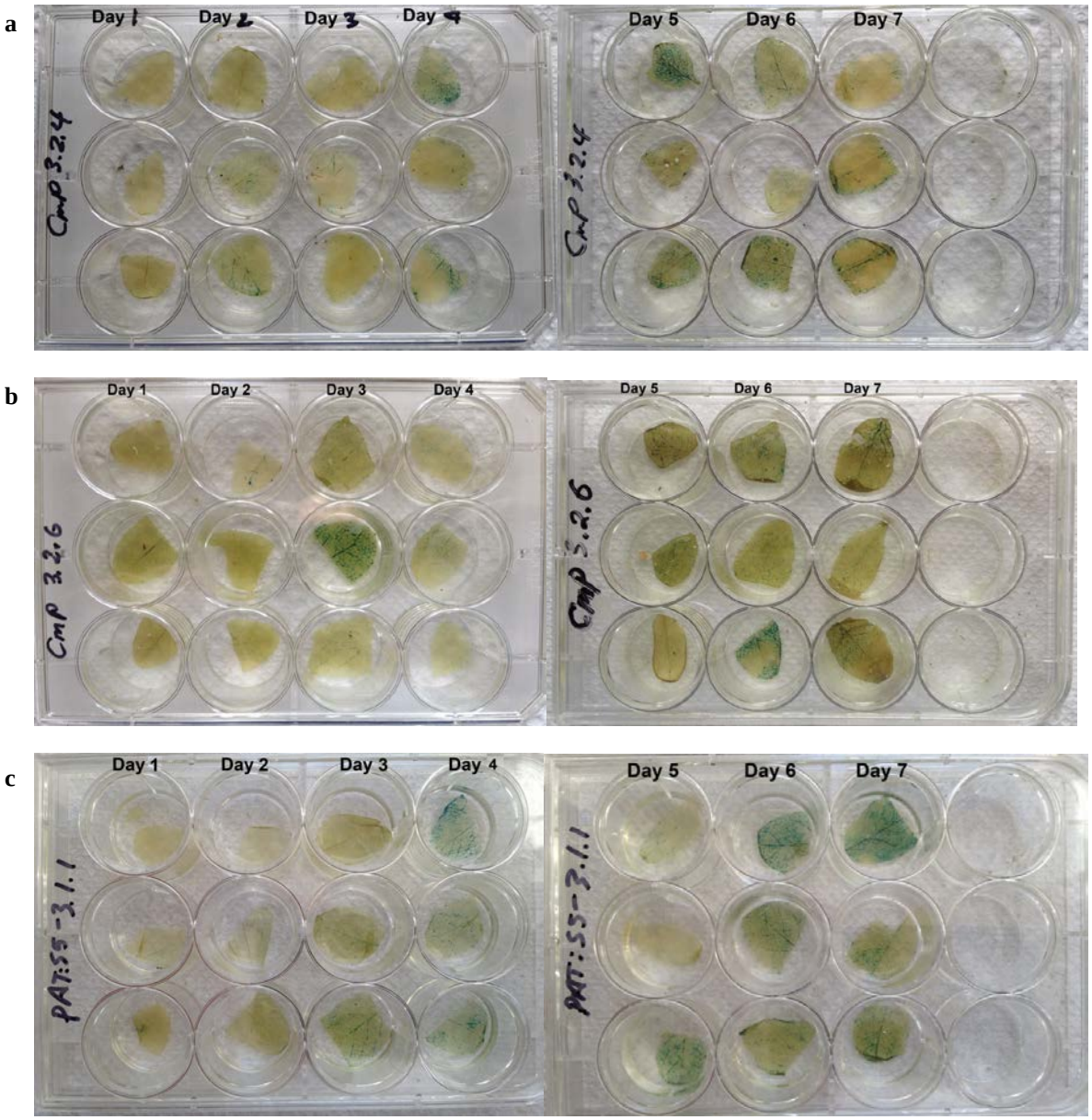


Figure 4-34. GUS staining of (a) CmP:pHTOP-3.2.4, (b) CmP:pHTOP-3.2.6, and (c) PAT:pHTOP-3.1.1 following 1 to 7 days dexamethasone induction by watering plants with 300 mL 20 μM dexamethasone on day 0, 2, 4, and 6.

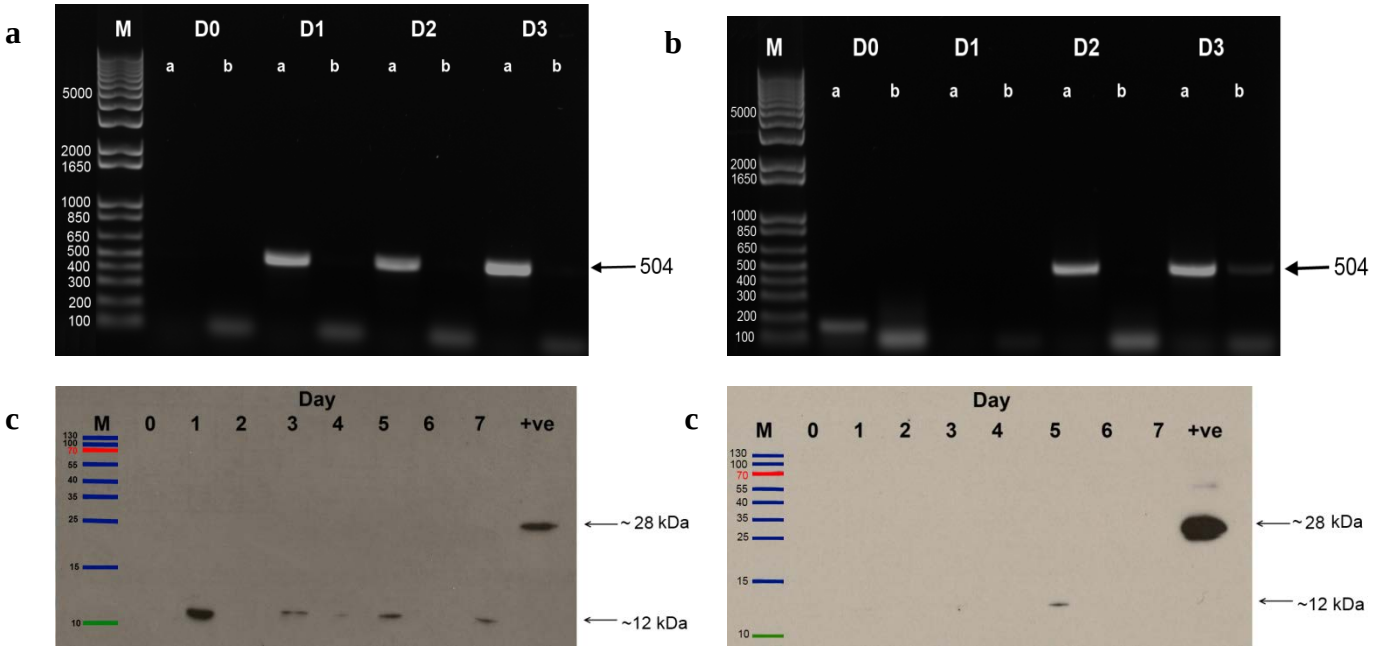


Figure 4-35 CmP transgene expression analysis following subterranean irrigation employing 300 mL tap water containing 20 μM dexamethasone on day 0, 3, 4 and 6. For each of days 1 through 7, leaf samples were taken from each of seven individual plants, pooled, subjected to vesicle enrichment extraction and analyzed by RT-PCR analysis for the detection of GUS transcript in CmP:pHTOP-3.2.4 (a) and CmP:pHTOP-3.2.6 plants (c) as indicated by the 504 bp amplicon from the use of GUS-F10 and GUS-R11 primers in RT-PCR reactions labeled “a” whereas columns labeled “b” are their respective GUS-F10/GUS-R11 PCR controls. Also shown is the western blot analysis using anti-PAT antiserum of T_4 generation CmP:pHTOP-3.2.4 (b) and CmP:pHTOP-3.2.6 (d). Key: M, PageRuler™ Prestained Protein Ladder—Fermentas; +ve, CmP 14.1.3 positive control.

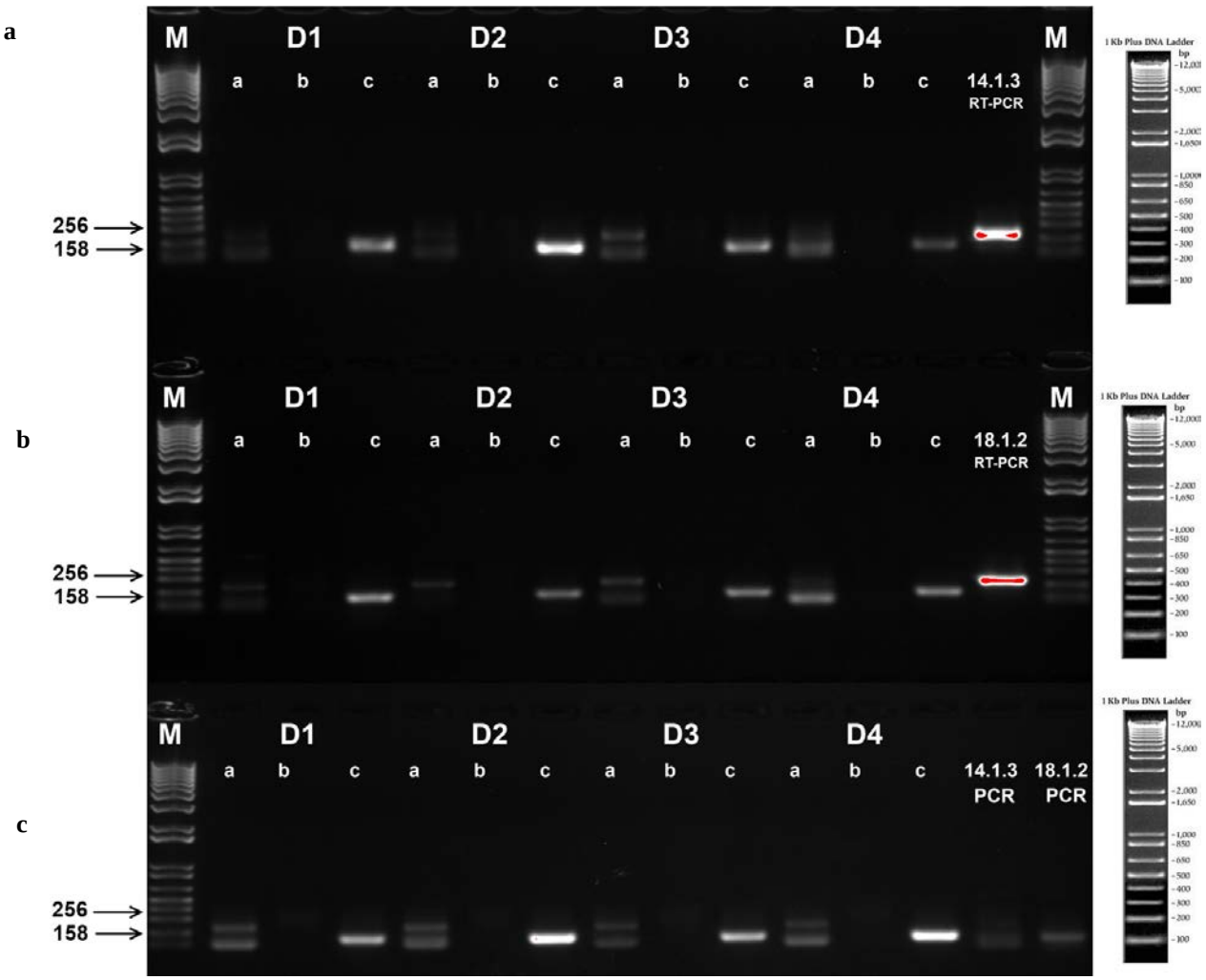


Figure 4-36. RT-PCR analysis of (a) CmP:pHTOP-3.2.4, (b) CmP:pHTOP-3.2.6 and (c) CmP:pHTOP-3.2.7 plants following dexamethasone induction for 1 to 4 days (D1 to D4). Expression of the CmP transgene is indicated by the 246 bp amplicon from the use of Cm-F1 and CmPJ-R primers in RT-PCR reactions labeled "a", whereas columns labeled "b" are their respective Cm-F1/CmPJ-R PCR controls. Expression of the GUS transgene is indicated by the 158 bp amplicon from the use of GUS-F10 and GUS-R12 primers in RT-PCR reactions labeled "c". For comparison, RT-PCR analysis of CmP-14.1.3 and CmP-18.1.2 (see also Figure 4-38); Key: M, 1 kb plus DNA ladder® (Invitrogen).

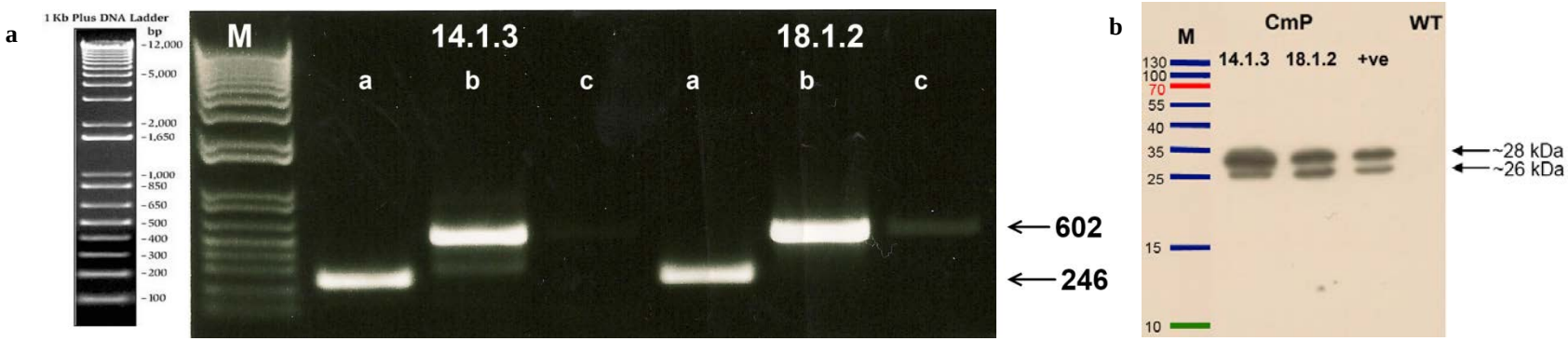


Figure 4-37. RT-PCR analysis of CmP-14.1.3 and CmP-18.1.2 (a). Expression of the CmP transgene is indicated by the 246 bp amplicon from the use of Cm-F1 and CmPJ-R primers in RT-PCR reactions labeled "a", and the 602 bp amplicon from the use of PAT-F12 and PAT-R13 primers in columns labeled "b" whereas columns labeled "c" are their respective PAT-F12/PAT-R13 PCR controls. Key: M, 1 kb plus DNA ladder® (Invitrogen); western blot analysis using anti-PAT antiserum of leaf samples from T_4 generation CmP 14.1.3 and CmP 18.1.2 (b). Key: M, PageRuler™ Prestained Protein Ladder—Fermentas; +ve, CmP 14.1.3 positive control.

Western blot analysis indicated accumulation of the ~12 kDa peptide in rosette leaves of CmP:pHTOP-3.2.4 plants, whereas little accumulation was observed in rosette leaves of CmP:pHTOP-3.2.6 (Figure 35b and d). RT-PCR analysis confirmed similar levels of transgene induction in samples from plants CmP:pHTOP-3.2.4, CmP:pHTOP-3.2.6 and CmP:pHTOP-3.2.7 (Figure 4-36) and that expression of the pumpkin 2S(Δ79)-PAT transgene resulting from dexamethasone induction was not as great as for CaMV 35S promoter driven expression in plants CmP-14.1.3 and CmP-18.1.2 (Figures 4-37). These findings indicated that comparison of CmP:pHTOP-3.2.4 and CmP:pHTOP-3.2.6 plants in the microarray study would facilitate identifying genes involved in fusion protein accumulation and that inclusion of CmP-14.1.3 and CmP-18.1.2 plants in the microarray study would permit the examination of gene activation in scenarios of both induced low-level expression and long-term constitutive expression of the transgenes.

Microarray data analysis. Gene expression resulting from dexamethasone induction was monitored by GeneSpring GX11 software employing Atdschipb520654 library files. To identify genes involved in the sorting and trafficking machinery that had resulted from dexamethasone induction a series of pair-wise comparisons were necessary. For each time, day 0 to 3, the expression profiles of CmP:pHTOP-3.2.4 and CmP:pHTOP-3.2.6 were weighted against the expression profiles of the corresponding PAT:pHTOP-3.1.1 control. For example, 3.2.4 day 0 vs. PAT day 0, 3.2.4 day 1 vs. PAT day 1, etc. To be considered valid, candidate genes had to pass two criteria: 1) they would be disqualified if they were also represented in their day 0 comparison to the PAT:pHTOP-3.1.1 control, i.e. if they showed a differential gene expression (either an increase or decrease) prior to administering dexamethasone, and 2) if gene expression changes were not represented in the top 50 fold-change for each time point day 1 through

day 3 as a lesser change in gene expression could be argued to be of little consequence. These stringent conditions were intended to provide a conservative assessment of potential effected genes. Consequently, only a few candidate genes passed these stringent criteria (Tables 4-3 and 4-4). Induction of CmP:pHTOP-3.2.4 and CmP:pHTOP-3.2.6 indicated that perturbations of the biochemical pathways involving the endoplasmic reticulum were frequent, which suggests involvement of the endomembrane system, as expected. Pie charts in Figures 38a and b summarize these effects according to biological function. These perturbations were found to be more dramatic in the assessment of CmP-14.1.3 and CmP-18.1.2 verses their wild-type control. Tables 4-5, 4-6, and 4-7[3] list candidate genes having the greatest fold change and pie charts Figures 4-39a and 4-39b summarize these affects according to biological function. As indicated, a number of these genes are typically involved in protein folding and ER-mediated biotic and abiotic stress response. This implies that the endomembrane system was under considerable demand due to long term constitutive expression of the pumpkin 2S(Δ79)-PAT transgene. Of note, the greatest reduction observed was for gene At3g09260 (Table 4-7), a major constituent of ER bodies. This indicated that as a result of transgene induction, the fusion protein was not sequestered into ER bodies as part of an unfolded protein response, but rather was being sequestered in different structures derived from the endomembrane system. This suggested that fusion protein accumulation and vesicle biogenesis had occurred in these plants and that many of the candidate genes identified from the microarray study were involved in these processes.

[3] The top 25 candidate genes identified resulting in decreased gene expression for CmP-14.1.3 and CmP-18.1.2 were identical except for marginal differences in fold-change (FC) values. Consequently, only the top 25 FC reductions for CmP-14.1.3 are reported.

Table 4-3. Top 7 greatest-fold change in gene expression of dexamethasone induced CmP:pHTOP-3-2-4 following pair-wise comparison to PAT:pHTOP-3.1.1 control

	Probe Set ID	[324 d1] vs [PAT d1]	[324 d2] vs [PAT d2]	[324-d3] vs [PAT-d3]	Gene Name	Location	Biological Function	Description
1	At2g29460_PM_at	13.3	7.4	29.6	ATGSTU4, GLUTATHIONE S-TRANSFERASE 22, GLUTATHIONE S-TRANSFERASE TAU 4, GST22, GSTU4	cytoplasm	response to cyclopentenone, toxin catabolic process	Encodes glutathione transferase belonging to the tau class of GSTs.
2	At4g23150_PM_at	9.0	6.5	20.8	CRK7, CYSTEINE-RICH RLK (RECEPTOR-LIKE PROTEIN KINASE) 7	endomembrane system	phosphorylation, protein phosphorylation	Encodes a cysteine-rich receptor-like protein kinase.
3	At1g21240_PM_at	6.7	6.2	24.9	WAK3, WALL ASSOCIATED KINASE 3	extracellular region	protein phosphorylation	encodes a wall-associated kinase
4	At1g51890_PM_at	6.3	6.7	28.1	Not yet established	Not yet established	defense response to bacterium, defense response to fungus, negative regulation of programmed cell death, protein phosphorylation, protein targeting to membrane, regulation of plant-type hypersensitive response, salicylic acid mediated signaling pathway	Leucine-rich repeat protein kinase family protein; FUNCTIONS IN: kinase activity; INVOLVED IN: protein amino acid phosphorylation; LOCATED IN: endomembrane system
5	At1g09350_PM_at	-43.0	-7.8	-210.1	ATGOLS3, GALACTINOL SYNTHASE 3, GOLS3	Not yet established	cellular component	carbohydrate biosynthetic process, response to cold, response to oxidative stress
6	At5g37300_PM_at	-6.5	-25.5	-27.5	WSD1	endoplasmic reticulum, nucleus	ovule development, wax biosynthetic process	Encodes a bifunctional enzyme, wax ester synthase (WS) and diacylglycerol acyltransferase (DGAT). In vitro assay indicated a ratio of 10.9 between its WS and DGAT activities. Both mutant and in vivo expression/analysis in yeast studies indicated a role in wax biosynthesis
7	At5g07010_PM_at	-30.1	-7.2	-22.2	ARABIDOPSIS THALIANA SULFOTRANSFERASE 2A, ATST2A, ST2A, SULFOTRANSFERASE 2A	cellular_component chloroplast	jasmonic acid biosynthetic process, jasmonic acid mediated signaling pathway, jasmonic acid metabolic process, response to jasmonic acid stimulus, response to wounding	Encodes a sulfotransferase that acts specifically on 11- and 12-hydroxyjasmonic acid. Transcript levels for this enzyme are increased by treatments with jasmonic acid (JA), 12-hydroxyJA, JA-isoleucine, and 12-oxyphytodienoic acid (a JA precursor).

Table 4-4. Top 20 greatest fold change in gene expression of dexamethasone induced CmP:pHTOP-3-2-6 following pair-wise comparison to PAT:pHTOP-3.1.1 control

	Probe Set ID	[326 d1] vs [PAT d1]	[326 d2] vs [PAT d2]	[326 d3] vs [PAT d3]	Gene Name	Location	Biological Function	Description
1	At4g19430_PM_at	4.4	19.2	18.9	Not yet established	mitochondrion	unknown	unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: chloroplast).
2	At5g48570_PM_at	3.5	25.8	15.9	ATFKBP65, FKBP65, ROF2	membrane, nucleus, vacuole	N-terminal protein myristoylation, cellular heat acclimation, peptidyl-proline modification, protein folding, protein peptidyl-prolyl isomerization, response to endoplasmic reticulum stress, response to heat, response to high light intensity, response to hydrogen peroxide	Encodes one of the 36 carboxylate clamp (CC)-tetratricopeptide repeat (TPR) proteins (Prasad 2010, Pubmed ID: 20856808) with potential to interact with Hsp90/Hsp70 as co-chaperones.
3	At5g51440_PM_at	3.3	11.6	15.5	Not yet established	mitochondrion	protein folding, response to endoplasmic reticulum stress, response to heat, response to high light intensity, response to hydrogen peroxide	HSP20-like chaperones superfamily protein; CONTAINS InterPro DOMAIN/s: Heat shock protein Hsp20 (InterPro:IPR002068), HSP20-like chaperone (InterPro:IPR008978); BEST Arabidopsis thaliana protein match is: mitochondrion-localized small heat shock protein 23.6 (TAIR:AT4G25200.1
4	At5g20250_PM_at	6.8	9.4	14.4	DARK INDUCIBLE 10, DIN10, RAFFINOSE SYNTHASE 6, RS6	chloroplast, plasmodesma	metabolic process, response to cold, response to karrikin, response to oxidative stress, aging	Encodes a member of glycosyl hydrolase family 36. Expression is induced within 3 hours of dark treatment, in senescing leaves and treatment with exogenous photosynthesis inhibitor.
5	At1g59860_PM_s_at	3.0	32.7	12.5	Not yet established	cytoplasm	protein folding, protein oligomerization, response to heat, response to osmotic stress, response to salt stress	HSP20-like chaperones superfamily protein; CONTAINS InterPro DOMAIN/s: Heat shock protein Hsp20 (InterPro:IPR002068), HSP20-like chaperone (InterPro:IPR008978); BEST Arabidopsis thaliana protein match is: HSP20-like chaperones superfamily protein (TAIR:AT1G07400.1
6	At5g64510_PM_at	3.6	6.5	11.3	TIN1, TUNICAMYCIN INDUCED 1	chloroplast, endoplasmic reticulum	ER-nucleus signaling pathway, cellular response to biotic stimulus, pollen development, protein folding, response to endoplasmic reticulum stress, response to heat, response to high light intensity, response to hydrogen peroxide	Encodes Tunicamycin Induced 1(TIN1), a plant-speci-c ER stress-inducible protein. TIN1 mutation affects pollen surface morphology. Transcriptionally induced by treatment with the N-linked glyclsylation inhibitor tunicamycin.
7	At5g58770_PM_at	4.2	11.2	10.6	ATCPT4, CIS-PRENYLTRANSFERASE 4, CPT4	chloroplast	dolichol biosynthetic process, ubiquinone biosynthetic process	Undecaprenyl pyrophosphate synthetase family protein; FUNCTIONS IN: dehydrodolichyl diphosphate synthase activity; INVOLVED IN: dolichol biosynthetic process; LOCATED IN: chloroplast;
8	At1g61566_PM_s_at	4.7	5.1	9.5	RALF-LIKE 9, RALFL9	apoplast, extracellular region	cell-cell signaling	Member of a diversely expressed predicted peptide family showing sequence similarity to tobacco Rapid Alkalinization Factor (RALF)
9	At4g21870_PM_at	7.1	8.2	9.4	Not yet established	cellular component, cytoplasm	response to heat, response to high light intensity, response to hydrogen peroxide	HSP20-like chaperones superfamily protein; CONTAINS InterPro DOMAIN/s: Heat shock protein Hsp20
10	At4g12400_PM_at	2.3	10.4	7.5	HOP3	Not yet established	heat acclimation, protein folding, response to endoplasmic reticulum stress, response to heat, response to high light intensity, response to hydrogen peroxide, response to stress	Encodes one of the 36 carboxylate clamp (CC)-tetratricopeptide repeat (TPR) proteins with potential to interact with Hsp90/Hsp70 as co-chaperones.

	Probe Set ID	[326 d1] vs [PAT d1]	[326 d2] vs [PAT d2]	[326 d3] vs [PAT d3]	Gene Name	Location	Biological Function	Description
11	At1g55850_PM_at	3.5	11.1	7.3	ATCSLE1, CELLULOSE SYNTHASE LIKE E1, CSLE1	Golgi apparatus, endoplasmic reticulum, membrane, plasma membrane	cellulose biosynthetic process, para-aminobenzoic acid metabolic process, plant-type cell wall biogenesis, polysaccharide biosynthetic process	encodes a protein similar to cellulose synthase
12	At1g65480_PM_at	12.1	9.0	7.3	FLOWERING LOCUS T, FT	cytoplasm, nucleus	photoperiodism, flowering, positive regulation of flower development, regulation of flower development, regulation of stomatal movement	FT, together with LFY, promotes flowering and is antagonistic with its homologous gene, TERMINAL FLOWER1 (TFL1). FT is expressed in leaves and is induced by long day treatment
13	At4g24400_PM_at	4.5	6.5	5.8	ATCIPK8, CBL-INTERACTING PROTEIN KINASE 8	plasma membrane	protein autophosphorylation, protein phosphorylation, response to glucose stimulus, response to nitrate, root development, signal transduction	Encodes a CBL (calcineurin B-like calcium sensor proteins) -interacting serine/threonine protein kinase. Regulates the low-affinity phase of the primary nitrate response
14	At1g09350_PM_at	-8.7	-11.5	-52.3	ATGOLS3, GALACTINOL SYNTHASE 3, GOLS3	cellular component, nucleus	carbohydrate biosynthetic process, response to cold, response to oxidative stress	Predicted to encode a galactinol synthase
15	At3g50970_PM_at	-4.9	-5.7	-24.1	LOW TEMPERATURE-INDUCED 30, LTI30, XERO2	membrane	cold acclimation, defense response to fungus, response to abscisic acid stimulus, response to cold, response to stress, response to water deprivation, response to water stimulus	Belongs to the dehydrin protein family, which contains highly conserved stretches of 7-17 residues that are repetitively scattered in their sequences, the K-, S-, Y- and lysine rich segments. LTI29 and LTI30 double overexpressors confer freeze tolerance. Located in membranes..
16	At3g54400_PM_at	-3.4	-8.7	-9.1	Not yet established	apoplast, cell wall, chloroplast, extracellular region, plant-type cell wall	cysteine biosynthetic process, proteolysis	Eukaryotic aspartyl protease family protein; FUNCTIONS IN: aspartic-type endopeptidase activity; INVOLVED IN: proteolysis; LOCATED IN: apoplast, cell wall, chloroplast, plant-type cell wall;
17	At4g02850_PM_at	-3.7	-14.1	-8.8	Not yet established	endomembrane system	biosynthetic process, circadian rhythm, cytokinin mediated signaling pathway	phenazine biosynthesis PhzC/PhzF family protein; FUNCTIONS IN: catalytic activity; INVOLVED IN: biosynthetic process; LOCATED IN: endomembrane system;
18	At1g06350_PM_s_at	-3.7	-11.1	-8.1	Not yet established	mitochondrion	lipid metabolic process, oxidation-reduction process	Fatty acid desaturase family protein; FUNCTIONS IN: oxidoreductase activity, oxidoreductase activity, acting on paired donors, with oxidation of a pair of donors resulting in the reduction of molecular oxygen to two molecules of water; INVOLVED IN: oxidation reduction, lipid metabolic process;
19	At1g14250_PM_s_at	-13.5	-17.5	-7.1	Not yet established	plasma membrane, vacuole	Not yet established	GDA1/CD39 nucleoside phosphatase family protein; FUNCTIONS IN: hydrolase activity; INVOLVED IN: biological_process unknown; LOCATED IN: vacuole
20	At2g39030_PM_at	-4.9	-7.6	-5.8	N-ACETYLTRANSFERASE ACTIVITY 1, NATA1	cellular_component, cytoplasm	response to jasmonic acid stimulus	Encodes a protein that acts as an ornithine N-delta-acetyltransferase, leading to the formation of N-delta-actetylornithine. This compound is likely used in plant defense and levels of it are increased in Arabidopsis plants in response to MeJA and ABA

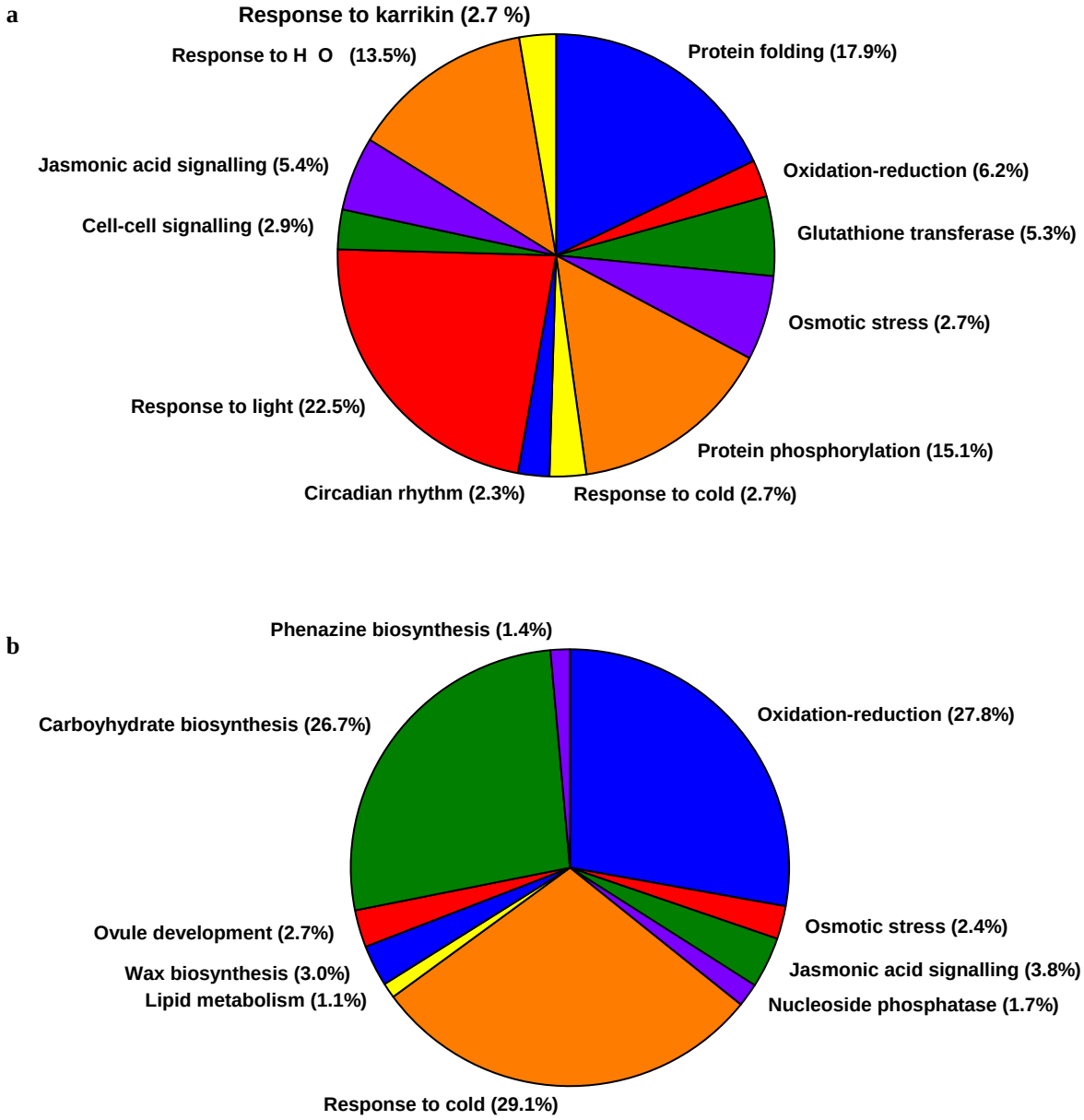


Figure 4-38. Top (a) up-regulated and (b) down-regulated candidate genes according to biological function as a result of dexamethasone induced expression of CmP transgene of CmP:pHTOP-3.2.4 and -3.2.6.

Table 4-5. Top 25 greatest-fold increase in gene expression of CmP-14.1.3 following pairwise comparison to WT control

	Probe Set ID	FC ([14] vs [WT])	Gene Name	Location	Biological Process	Description
1	At5g23240_PM_at	148.7		chloroplast	protein folding	DNAJ heat shock N-terminal domain-containing protein; FUNCTIONS IN: unfolded protein binding, heat shock protein binding; INVOLVED IN: protein folding; LOCATED IN: chloroplast; EXPRESSED IN: 21 plant structures; EXPRESSED DURING: 11 growth stages; CONTAINS InterPro DOMAIN/s: Molecular chaperone, heat shock protein, Hsp40, DnaJ (InterPro:IPR015609), Heat shock protein DnaJ, N-terminal (InterPro:IPR001623), Heat shock protein DnaJ (InterPro:IPR003095); BEST Arabidopsis thaliana protein match is: DNAJ heat shock N-terminal domain-containing protein (TAIR:AT2G42750.1)
2	At1g07050_PM_at	83.4		cellular component	Not yet established	CCT motif family protein; CONTAINS InterPro DOMAIN/s: CCT domain (InterPro:IPR010402); BEST Arabidopsis thaliana protein match is: CCT motif family protein (TAIR:AT4G25990.1)
3	At5g42900_PM_at	64.4	COLD REGULATED GENE 27, COR27	Not yet established	response to cold, response to karrikin	cold regulated gene 27 (COR27); BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT4G33980.1)
4	At4g33980_PM_at	44.2		Not yet established	response to karrikin	FUNCTIONS IN: molecular_function unknown; INVOLVED IN: response to karrikin; LOCATED IN: cellular_component unknown; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 15 growth stages; BEST Arabidopsis thaliana protein match is: cold regulated gene 27 (TAIR:AT5G42900.3).
5	At4g16146_PM_at	43.0		cellular component	Not yet established	cAMP-regulated phosphoprotein 19-related protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: cAMP-regulated phosphoprotein/endosulphine conserved region (InterPro:IPR006760); BEST Arabidopsis thaliana protein match is: cAMP-regulated phosphoprotein 19-related protein (TAIR:AT1G69510.2)
6	At2g40080_PM_at	39.3	EARLY FLOWERING 4, ELF4	nucleus	entrainment of circadian clock, photoperiodism, photoperiodism, flowering, positive regulation of circadian rhythm, red or far-red light signaling pathway, regulation of flower development, response to karrikin, response to red light	Encodes a novel nuclear 111 amino-acid phytochrome-regulated component of a negative feedback loop involving the circadian clock central oscillator components CCA1 and LHY. ELF4 is necessary for light-induced expression of both CCA1 and LHY, and conversely, CCA1 and LHY act negatively on light-induced ELF4 expression. ELF4 promotes clock accuracy and is required for sustained rhythms in the absence of daily light/dark cycles. It is involved in the phyB-mediated constant red light induced seedling de-etiolation process and may function to coregulate the expression of a subset of phyB-regulated genes
7	At5g60100_PM_at	28.4	APRR3, PRR3, PSEUDO-RESPONSE REGULATOR 3	intracellular, nucleus	circadian rhythm, negative regulation of protein binding, regulation of circadian rhythm, regulation of transcription, DNA-dependent, two-component signal transduction system (phosphorelay)	Encodes pseudo-response regulator 3 (APRR3/PRR3). PRR3 transcript levels vary in a circadian pattern with peak expression at dusk under long and short day conditions. PRR3 affects the period of the circadian clock and seedlings with reduced levels of PRR3 have shorter periods, based on transcriptional assays of clock-regulated genes. PRR3 is expressed in the vasculature of cotyledons and leaves where it may help stabilize the TOC1 protein by preventing interactions between TOC1 and the F-box protein ZTL.
8	At2g39920_PM_at	27.5		cellular component	dephosphorylation, response to cadmium ion	HAD superfamily, subfamily IIIB acid phosphatase ; FUNCTIONS IN: acid phosphatase activity; INVOLVED IN: response to cadmium ion; LOCATED IN: cellular_component unknown; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Acid phosphatase (Class B) (InterPro:IPR005519); BEST Arabidopsis thaliana protein match is: HAD superfamily, subfamily IIIB acid phosphatase (TAIR:AT4G29260.1)

	Probe Set ID	FC ([14] vs [WT])	Gene Name	Location	Biological Process	Description
9	At2g21660_PM_at	27.3	"COLD, CIRCADIAN RHYTHM, AND RNA	chloroplast, cytoplasm, cytosol, intracellular,	DNA duplex unwinding, RNA secondary structure unwinding, circadian rhythm,	Encodes a small glycine-rich RNA binding protein that is part of a negative-feedback loop through which AtGRP7 regulates the circadian oscillations of its own transcript. Gene

			BINDING 2", ATGRP7, CCR2, GLYCINE RICH PROTEIN 7, GLYCINE-RICH RNA-BINDING PROTEIN 7, GR-RBP7, GRP7	nucleus, peroxisome, plasmodesma	innate immune response, mRNA export from nucleus, regulation of stomatal movement, response to cadmium ion, response to cold, response to osmotic stress, response to salt stress, response to zinc ion, vegetative to reproductive phase transition of meristem	expression is induced by cold. GRP7 appears to promote stomatal opening and reduce tolerance under salt and dehydration stress conditions, but, promotes stomatal closing and thereby increases stress tolerance under conditions of cold tolerance. Loss of function mutations have increased susceptibility to pathogens suggesting a role in mediating innate immune response
10	At5g62360_PM_at	27.0		cellular component	biological_process, metabolic process, negative regulation of catalytic activity	Plant invertase/pectin methylesterase inhibitor superfamily protein; FUNCTIONS IN: enzyme inhibitor activity, pectinesterase inhibitor activity, pectinesterase activity; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Pectinesterase inhibitor (InterPro:IPR006501); BEST Arabidopsis thaliana protein match is: Plant invertase/pectin methylesterase inhibitor superfamily protein (TAIR:AT5G62350.1)
11	At2g15890_PM_at	26.3	MATERNAL EFFECT EMBRYO ARREST 14, MEE14	chloroplast	defense response to fungus, embryo development ending in seed dormancy	maternal effect embryo arrest 14 (MEE14); FUNCTIONS IN: molecular_function unknown; INVOLVED IN: defense response to fungus, embryo development ending in seed dormancy; LOCATED IN: chloroplast; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages
12	At3g46640_PM_s_at	24.9	LUX, LUX ARRHYTHMO, PCL1, PHYTOCLOCK 1	Not yet established	circadian rhythm, regulation of transcription, DNA-dependent	Encodes a myb family transcription factor with a single Myb DNA-binding domain (type SHAQKYF) that is unique to plants and is essential for circadian rhythms, specifically for transcriptional regulation within the circadian clock. LUX is required for normal rhythmic expression of multiple clock outputs in both constant light and darkness. It is coregulated with TOC1 and seems to be repressed by CCA1 and LHY by direct binding of these proteins to the evening element in the LUX promoter.
13	At2g33830_PM_at	24.8		cellular component	Not yet established	Dormancy/auxin associated family protein; CONTAINS InterPro DOMAIN/s: Dormancyauxin associated (InterPro:IPR008406); BEST Arabidopsis thaliana protein match is: dormancy-associated protein-like 1 (TAIR:AT1G28330.1)
14	At4g04330_PM_at	22.8	ATRBCX1, HOMOLOGUE OF CYANOBACTERIAL RBCX 1, RBCX1	chloroplast thylakoid	chaperone-mediated protein folding, response to cold, response to salt stress, response to water deprivation	Encodes a chloroplast thylakoid localized RbcX protein that acts as a chaperone in the folding of Rubisco.
15	At4g25100_PM_at	21.1	FE SUPEROXIDE DISMUTASE 1, FSD1	chloroplast, plasmodesma	circadian rhythm, oxidation-reduction process, removal of superoxide radicals, response to cadmium ion, response to copper ion, response to oxidative stress	Fe-superoxide dismutase
16	At3g48360_PM_at	20.2	ATBT2, BT2, BTB AND TAZ DOMAIN PROTEIN 2	cytoplasm, nucleus	abscisic acid mediated signaling pathway, auxin mediated signaling pathway, circadian rhythm, embryo sac development, pollen development, positive regulation of telomerase activity, regulation of response to stress, regulation of transcription, DNA-dependent, response to carbohydrate stimulus, response to cold, response to hydrogen peroxide, response to jasmonic acid stimulus, response to nitrate, response to salicylic acid stimulus, response to salt stress, response to wounding, sugar mediated signaling	encodes a protein (BT2) that is an essential component of the TAC1-mediated telomerase activation pathway. Acts redundantly with BT3 and BT1 during female gametophyte development and with BT3 during male gametophyte development. BT2 also mediates multiple responses to nutrients, stresses, and hormones.
17	At1g17665_PM_at	19.2		endomembrane system	Not yet established	FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; EXPRESSED IN: 15 plant structures; EXPRESSED DURING: 11 growth stages
	Probe Set ID	FC ([14] vs [WT])	Gene Name	Location	Biological Process	Description

18	At1g56300_PM_at	18.5		Not yet established	protein folding	Chaperone DnaJ-domain superfamily protein; FUNCTIONS IN: heat shock protein binding; INVOLVED IN: protein folding, response to cyclopentenone; LOCATED IN: cellular_component unknown; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Molecular chaperone, heat shock protein, Hsp40, DnaJ (InterPro:IPR015609), Heat shock protein DnaJ, N-terminal (InterPro:IPR001623), Heat shock protein DnaJ, conserved site (InterPro:IPR018253); BEST Arabidopsis thaliana protein match is: Chaperone DnaJ-domain superfamily protein (TAIR:AT1G71000.1)
19	At5g47330_PM_at	18.3		endomembrane system	metabolic process	alpha/beta-Hydrolases superfamily protein; FUNCTIONS IN: palmitoyl-(protein) hydrolase activity; INVOLVED IN: response to salt stress; LOCATED IN: endomembrane system; EXPRESSED IN: 12 plant structures; EXPRESSED DURING: LP.06 six leaves visible, LP.04 four leaves visible, 4 anthesis, petal differentiation and expansion stage; CONTAINS InterPro DOMAIN/s: Palmitoyl protein thioesterase (InterPro:IPR002472); BEST Arabidopsis thaliana protein match is: alpha/beta-Hydrolases superfamily protein (TAIR:AT5G47340.1)
20	At5g48250_PM_at	17.4		plasma membrane	regulation of transcription, DNA-dependent	B-box type zinc finger protein with CCT domain; FUNCTIONS IN: sequence-specific DNA binding transcription factor activity, zinc ion binding; LOCATED IN: plasma membrane; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: CCT domain (InterPro:IPR010402), Zinc finger, B-box (InterPro:IPR000315); BEST Arabidopsis thaliana protein match is: CONSTANS-like 9 (TAIR:AT3G07650.4)
21	At4g15700_PM_at	17.2		endomembrane system	cell redox homeostasis, oxidation-reduction process	Thioredoxin superfamily protein; FUNCTIONS IN: electron carrier activity, arsenate reductase (glutaredoxin) activity, protein disulfide oxidoreductase activity; INVOLVED IN: cell redox homeostasis; LOCATED IN: endomembrane system; EXPRESSED IN: 16 plant structures; EXPRESSED DURING: 9 growth stages; CONTAINS InterPro DOMAIN/s: Glutaredoxin-like, plant II (InterPro:IPR011905), Thioredoxin fold (InterPro:IPR012335), Glutaredoxin (InterPro:IPR002109), Glutaredoxin subgroup (InterPro:IPR014025), Thioredoxin-like fold (InterPro:IPR012336); BEST Arabidopsis thaliana protein match is: Thioredoxin superfamily protein (TAIR:AT4G15690.1)
22	At5g24470_PM_at	16.7	APRR5, PRR5, PSEUDO-RESPONSE REGULATOR 5	intracellular, nucleus	circadian rhythm, negative regulation of transcription, DNA-dependent, nuclear import, photomorphogenesis, regulation of transcription, DNA-dependent, response to far red light, response to red light, two-component signal transduction system (phosphorelay)	Encodes a pseudo-response regulator whose mutation affects various circadian-associated biological events such as flowering time in the long-day photoperiod conditions, red light sensitivity of seedlings during early photomorphogenesis, and the period of free-running rhythms of certain clock-controlled genes including CCA1 and APRR1/TOC1 in constant white light. Acts as transcriptional repressor of CCA1 and LHY.
23	At4g34950_PM_at	16.7		Not yet established	Not yet established	Major facilitator superfamily protein; CONTAINS InterPro DOMAIN/s: Nodulin-like (InterPro:IPR010658), Major facilitator superfamily, general substrate transporter (InterPro:IPR016196); BEST Arabidopsis thaliana protein match is: Major facilitator superfamily protein (TAIR:AT2G16660.1)
24	At5g20250_PM_at	15.8	DARK INDUCIBLE 10, DIN10, RAFFINOSE SYNTHASE 6, RS6	chloroplast, plasmodesma	metabolic process, response to cold, response to karrikin, response to oxidative stress, aging	encodes a member of glycosyl hydrolase family 36. Expression is induced within 3 hours of dark treatment, in senescing leaves and treatment with exogenous photosynthesis inhibitor. Induction of gene expression was suppressed in excised leaves supplied with sugar. The authors suggest that the gene's expression pattern is responding to the level of sugar in the cell.
25	At4g15680_PM_at	15.1		endomembrane system	cell redox homeostasis, oxidation-reduction process	Thioredoxin superfamily protein; FUNCTIONS IN: electron carrier activity, arsenate reductase (glutaredoxin) activity, protein disulfide oxidoreductase activity; INVOLVED IN: cell redox homeostasis; LOCATED IN: endomembrane system; EXPRESSED IN: 14 plant structures; EXPRESSED DURING: 8 growth stages; CONTAINS InterPro DOMAIN/s: Glutaredoxin-like, plant II (InterPro:IPR011905), Thioredoxin fold (InterPro:IPR012335), Glutaredoxin (InterPro:IPR002109), Glutaredoxin subgroup (InterPro:IPR014025), Thioredoxin-like fold (InterPro:IPR012336); BEST Arabidopsis thaliana protein match is: Thioredoxin superfamily protein (TAIR:AT4G15670.1)

Table 4-6. Greatest-fold increase in gene expression of CmP-18.1.2 following pairwise comparison to WT control (top 25)

Probe Set ID	FC ([18] vs [WT])	Gene Name	Location	Biological Process	Description
At5g23240_PM_at	148.03119		chloroplast	protein folding	DNAJ heat shock N-terminal domain-containing protein; FUNCTIONS IN: unfolded protein binding, heat shock protein binding; INVOLVED IN: protein folding; LOCATED IN: chloroplast; BEST Arabidopsis thaliana protein match is: DNAJ heat shock N-terminal domain-containing protein (TAIR:AT2G42750.1
At1g07050_PM_at	93.36311		cellular component		CCT motif family protein; CONTAINS InterPro DOMAIN/s: CCT domain (InterPro:IPR010402); BEST Arabidopsis thaliana protein match is: CCT motif family protein (TAIR:AT4G25990.1)
At5g42900_PM_at	52.785534	COLD REGULATED GENE 27, COR27		response to cold, response to karrikin	cold regulated gene 27 (COR27); BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT4G33980.1)
At2g40080_PM_at	45.95312	EARLY FLOWERING 4, ELF4	nucleus	entrainment of circadian clock, photoperiodism, photoperiodism, flowering, positive regulation of circadian rhythm, red or far-red light signaling pathway, regulation of flower development, response to karrikin, response to red light	Encodes a novel nuclear 111 amino-acid phytochrome-regulated component of a negative feedback loop involving the circadian clock central oscillator components CCA1 and LHY. ELF4 is necessary for light-induced expression of both CCA1 and LHY, and conversely, CCA1 and LHY act negatively on light-induced ELF4 expression. ELF4 promotes clock accuracy and is required for sustained rhythms in the absence of daily light/dark cycles. It is involved in the phyB-mediated constant red light induced seedling de-etiolation process and may function to coregulate the expression of a subset of phyB-regulated genes
At4g16146_PM_at	37.987194		cellular component		cAMP-regulated phosphoprotein 19-related protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; CONTAINS InterPro DOMAIN/s: cAMP-regulated phosphoprotein/endosulphine conserved region (InterPro:IPR006760); BEST Arabidopsis thaliana protein match is: cAMP-regulated phosphoprotein 19-related protein (TAIR:AT1G69510.2);
At5g60100_PM_at	36.85168	APRR3, PRR3, PSEUDO-RESPONSE REGULATOR 3	intracellular, nucleus	circadian rhythm, negative regulation of protein binding, regulation of circadian rhythm, regulation of transcription, DNA-dependent, two-component signal transduction system (phosphorelay)	Encodes pseudo-response regulator 3 (APRR3/PRR3). PRR3 transcript levels vary in a circadian pattern with peak expression at dusk under long and short day conditions. PRR3 affects the period of the circadian clock and seedlings with reduced levels of PRR3 have shorter periods, based on transcriptional assays of clock-regulated genes. PRR3 is expressed in the vasculature of cotyledons and leaves where it may help stabilize the TOC1 protein by preventing interactions between TOC1 and the F-box protein ZTL.
At4g33980_PM_at	34.313267			response to karrikin	FUNCTIONS IN: molecular_function unknown; INVOLVED IN: response to karrikin; LOCATED IN: cellular_component unknown; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 15 growth stages; BEST Arabidopsis thaliana protein match is: cold regulated gene 27 (TAIR:AT5G42900.3).
At4g25100_PM_at	29.358429	FE SUPEROXIDE DISMUTASE 1, FSD1	chloroplast, plasmodesma	circadian rhythm, oxidation-reduction process, removal of superoxide radicals, response to cadmium ion, response to copper ion, response to oxidative stress	Fe-superoxide dismutase

Probe Set ID	FC ([18] vs [WT])	Gene Name	Location	Biological Process	Description
At3g48360_PM_at	28.364069	ATBT2, BT2, BTB AND TAZ DOMAIN PROTEIN 2	cytoplasm, nucleus	abscisic acid mediated signaling pathway, auxin mediated signaling pathway, circadian rhythm, embryo sac development, pollen development, positive regulation of telomerase activity, regulation of response to stress, regulation of transcription, DNA-dependent, response to abscisic acid stimulus, response to auxin stimulus, response to carbohydrate stimulus, response to cold, response to hydrogen peroxide, response to jasmonic acid stimulus, response to nitrate, response to salicylic acid stimulus, response to salt stress, response to wounding, sugar mediated signaling pathway	encodes a protein (BT2) that is an essential component of the TAC1-mediated telomerase activation pathway. Acts redundantly with BT3 and BT1 during female gametophyte development and with BT3 during male gametophyte development. BT2 also mediates multiple responses to nutrients, stresses, and hormones.
At3g46640_PM_s_at	25.463598	LUX, LUX ARRHYTHMO, PCL1, PHYTOCLOCK 1		circadian rhythm, regulation of transcription, DNA-dependent	Encodes a myb family transcription factor with a single Myb DNA-binding domain (type SHAQKYF) that is unique to plants and is essential for circadian rhythms, specifically for transcriptional regulation within the circadian clock. LUX is required for normal rhythmic expression of multiple clock outputs in both constant light and darkness. It is coregulated with TOC1 and seems to be repressed by CCA1 and LHY by direct binding of these proteins to the evening element in the LUX promoter.
At2g21660_PM_at	19.981934	"COLD, CIRCADIAN RHYTHM, AND RNA BINDING 2", ATGRP7, CCR2, GLYCINE RICH PROTEIN 7, GLYCINE-RICH RNA-BINDING PROTEIN 7, GR-RBP7, GRP7	chloroplast, cytoplasm, cytosol, intracellular, nucleus, peroxisome, plasmodesma	DNA duplex unwinding, RNA secondary structure unwinding, circadian rhythm, innate immune response, mRNA export from nucleus, regulation of stomatal movement, response to cadmium ion, response to cold, response to osmotic stress, response to salt stress, response to zinc ion, vegetative to reproductive phase transition of meristem	Encodes a small glycine-rich RNA binding protein that is part of a negative-feedback loop through which AtGRP7 regulates the circadian oscillations of its own transcript. Gene expression is induced by cold. GRP7 appears to promote stomatal opening and reduce tolerance under salt and dehydration stress conditions, but, promotes stomatal closing and thereby increases stress tolerance under conditions of cold tolerance. Loss of function mutations have increased susceptibility to pathogens suggesting a role in mediating innate immune response. Mutants are also late flowering in a non-photoperiodic manner and are responsive to vernalization suggesting an interaction with the autonomous flowering pathway. There is a reduction of mRNA export from the nucleus in grp7 mutants. GRP7:GFP fusion proteins can be found in the cytosol and nucleus. A substrate of the type III effector HopU1 (mono-ADP-ribosyltransferase).
At1g71000_PM_at	17.685808		cytoplasm, nucleus	protein folding, response to heat, response to high light intensity, response to hydrogen peroxide	Chaperone DnaJ-domain superfamily protein; FUNCTIONS IN: heat shock protein binding; INVOLVED IN: protein folding; LOCATED IN: cellular_component unknown;
At3g07650_PM_at	16.495216	COL9, CONSTANS-LIKE 9	intracellular, nucleus	circadian rhythm, negative regulation of long-day photoperiodism, flowering, regulation of transcription, DNA-dependent	This gene belongs to the CO (CONSTANS) gene family. This gene family is divided in three subgroups: groups III, to which COL9 belongs, is characterised by one B-box (supposed to regulate protein-protein interactions) and a second diverged zinc finger. COL9 downregulates expression of CO (CONSTANS) as well as FT and SOC1 which are known regulatory targets of CO.

Probe Set ID	FC ([18] vs [WT])	Gene Name	Location	Biological Process	Description
At2g21130_PM_at	15.287721		cytoplasm, cytosol, plasma membrane	circadian rhythm, protein folding, response to abscisic acid stimulus, response to cold, response to water deprivation	Cyclophilin-like peptidyl-prolyl cis-trans isomerase family protein; FUNCTIONS IN: peptidyl-prolyl cis-trans isomerase activity; INVOLVED IN: protein folding; LOCATED IN: plasma membrane;
At1g56300_PM_at	15.203608			protein folding	Chaperone DnaJ-domain superfamily protein; FUNCTIONS IN: heat shock protein binding; INVOLVED IN: protein folding, response to cyclopentenone; LOCATED IN: cellular_component unknown; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Molecular chaperone, heat shock protein, Hsp40, DnaJ (InterPro:IPR015609), Heat shock protein DnaJ, N-terminal (InterPro:IPR001623), Heat shock protein DnaJ, conserved site (InterPro:IPR018253); BEST Arabidopsis thaliana protein match is: Chaperone DnaJ-domain superfamily protein (TAIR:AT1G71000.1)
At2g40350_PM_s_at	14.670263		nucleus	heat acclimation, regulation of transcription, DNA-dependent	encodes a member of the DREB subfamily A-2 of ERF/AP2 transcription factor family. The protein contains one AP2 domain. There are eight members in this subfamily including DREB2A AND DREB2B that are involved in response to drought
At2g33830_PM_at	14.611758		cellular component		Dormancy/auxin associated family protein; CONTAINS InterPro DOMAIN/s: Dormancyauxin associated (InterPro:IPR008406); BEST Arabidopsis thaliana protein match is: dormancy-associated protein-like 1 (TAIR:AT1G28330.1)
At4g04330_PM_at	14.434133	ATRBCX1, HOMOLOGUE OF CYANOBACTERIAL RBCX 1, RBCX1	chloroplast thylakoid	chaperone-mediated protein folding, response to cold, response to salt stress, response to water deprivation	Encodes a chloroplast thylakoid localized RbcX protein that acts as a chaperone in the folding of Rubisco.
At5g48250_PM_at	14.161593		plasma membrane	regulation of transcription, DNA-dependent	B-box type zinc finger protein with CCT domain; FUNCTIONS IN: sequence-specific DNA binding transcription factor activity, zinc ion binding; LOCATED IN: plasma membrane; EXPRESSED IN: 23 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: CCT domain (InterPro:IPR010402), Zinc finger, B-box (InterPro:IPR000315); BEST Arabidopsis thaliana protein match is: CONSTANS-like 9 (TAIR:AT3G07650.4)
At1g17665_PM_at	14.157685		endomembrane system		FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; EXPRESSED IN: 15 plant structures; EXPRESSED DURING: 11 growth stages;
At2g15890_PM_at	13.979204	MATERNAL EFFECT EMBRYO ARREST 14, MEE14	chloroplast	defense response to fungus, embryo development ending in seed dormancy	maternal effect embryo arrest 14 (MEE14); FUNCTIONS IN: molecular_function unknown; INVOLVED IN: defense response to fungus, embryo development ending in seed dormancy; LOCATED IN: chloroplast; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages
At4g34950_PM_at	13.624301				Major facilitator superfamily protein; CONTAINS InterPro DOMAIN/s: Nodulin-like (InterPro:IPR010658), Major facilitator superfamily, general substrate transporter (InterPro:IPR016196); BEST A.thaliana protein match is: Major facilitator superfamily protein (TAIR:AT2G16660.1)

Probe Set ID	FC ([18] vs [WT])	Gene Name	Location	Biological Process	Description
At2g39920_PM_at	13.52464		cellular component	dephosphorylation, response to cadmium ion	HAD superfamily, subfamily IIIB acid phosphatase ; FUNCTIONS IN: acid phosphatase activity; INVOLVED IN: response to cadmium ion; LOCATED IN: cellular_component unknown; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Acid phosphatase (Class B) (InterPro:IPR005519); BEST Arabidopsis thaliana protein match is: HAD superfamily, subfamily IIIB acid phosphatase (TAIR:AT4G29260.1)
At1g03020_PM_at	12.432245		cellular component	anther development, cell redox homeostasis	
At2g38465_PM_at	11.977656		plasma membrane		

Table 4-7. Top 25 greatest-fold reduction in gene expression of CmP-14.1.3 following pairwise comparison to WT control

	Probe Set ID	FC ([14] vs [WT])	Gene Name	Location	Biological Process	Description
1	At3g09260_PM_at	-113.0	BGLU23, LEB, LONG ER BODY, PSR3.1, PYK10	ER body, membrane, nucleus, peroxisome, plasmodesma, vacuole	ER body organization, cellular response to cold, metabolic process, negative regulation of defense response, response to osmotic stress, response to salt stress, response to symbiotic fungus	Encodes beta-glucosidase.The major constituent of ER bodies. One of the most abundant proteins in Arabidopsis seedlings. Exist in an soluble (inactive) and non-soluble (active) form, most probably formed in a polymerization process. Involved in the mutualistic interaction between Arabidopsis and the endophytic fungus Piriformospora indica.
2	At5g48490_PM_at	-69.4	Not yet established	endomembrane system	lipid transport	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein; FUNCTIONS IN: lipid binding; INVOLVED IN: lipid transport; LOCATED IN: endomembrane system; EXPRESSED IN: 20 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Bifunctional inhibitor/plant lipid transfer protein/seed storage (InterPro:IPR016140), Plant lipid transfer protein/seed storage/trypsin-alpha amylase inhibitor (InterPro:IPR003612); BEST Arabidopsis thaliana protein match is: Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein (TAIR:AT5G48485.1).
3	At2g39310_PM_at	-37.9	JACALIN-RELATED LECTIN 22, JAL22	cellular component	Not yet established	jacalin-related lectin 22 (JAL22); FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; EXPRESSED IN: 13 plant structures; EXPRESSED DURING: 6 growth stages; CONTAINS InterPro DOMAIN/s: Mannose-binding lectin (InterPro:IPR001229); BEST Arabidopsis thaliana protein match is: jacalin-related lectin 23 (TAIR:AT2G39330.1)
4	At2g42840_PM_at	-36.2	PDF1, PROTODERMAL FACTOR 1	extracellular region	Not yet established	Encodes a putative extracellular proline-rich protein is exclusively expressed in the L1 layer of vegetative, inflorescence and floral meristems and the protoderm of organ primordia
5	At3g02380_PM_at	-31.6	ATCOL2, COL2, CONSTANS-LIKE 2	intracellular	regulation of flower development, regulation of transcription, DNA-dependent	homologous to the flowering-time gene CONSTANS (CO) encoding zinc-finger proteins
6	At5g45670_PM_at	-29.8	Not yet established	endomembrane system	lipid metabolic process, metabolic process	GDSL-like Lipase/Acylhydrolase superfamily protein; FUNCTIONS IN: hydrolase activity, acting on ester bonds, carboxylesterase activity; INVOLVED IN: lipid metabolic process; LOCATED IN: endomembrane system; EXPRESSED IN: 19 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Lipase, GDSL (InterPro:IPR001087); BEST Arabidopsis thaliana protein match is: GDSL-like Lipase/Acylhydrolase superfamily protein (TAIR:AT4G18970.1)
7	At4g29020_PM_at	-29.7	Not yet established	endomembrane system	Not yet established	glycine-rich protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages.
8	At2g46830_PM_at	-27.9	ATCCA1, CCA1, CIRCADIAN CLOCK ASSOCIATED 1	nucleus	circadian rhythm, long-day photoperiodism, flowering, negative regulation of circadian rhythm, negative regulation of transcription, DNA-dependent, positive regulation of transcription, DNA-dependent, regulation of protein homodimerization activity, regulation of transcription, DNA-dependent, response to abscisic acid stimulus, response to auxin stimulus, response to cadmium ion, response to cold, response to ethylene stimulus, response to gibberellin stimulus, jasmonic acid stimulus, response to organic nitrogen, response to salicylic acid stimulus, response to salt stress	Encodes a transcriptional repressor that performs overlapping functions with LHY in a regulatory feedback loop that is closely associated with the circadian oscillator of Arabidopsis. Binds to the evening element in the promoter of TOC1 and represses TOC1 transcription. CCA1 and LHY colocalize in the nucleus and form heterodimers in vivo. CCA1 and LHY function synergistically in regulating circadian rhythms of Arabidopsis.

	Probe Set ID	FC ([14] vs [WT])	Gene Name	Location	Biological Process	Description
9	At4g02290_PM_at	-26.1	ATGH9B13, GH9B13, GLYCOSYL HYDROLASE 9B13	endomembrane system	carbohydrate metabolic process, metabolic process	glycosyl hydrolase 9B13 (GH9B13); FUNCTIONS IN: hydrolase activity, hydrolyzing O-glycosyl compounds, catalytic activity; INVOLVED IN: carbohydrate metabolic process; LOCATED IN: endomembrane system; EXPRESSED IN: 20 plant structures; EXPRESSED DURING: 9 growth stages; CONTAINS InterPro DOMAIN/s: Six-hairpin glycosidase (InterPro:IPR012341), Glycoside hydrolase, family 9, active site (InterPro:IPR018221), Six-hairpin glycosidase-like (InterPro:IPR008928), Glycoside hydrolase, family 9 (InterPro:IPR001701); BEST Arabidopsis thaliana protein match is: cellulase 2 (TAIR:AT1G02800.1).
10	At3g16450_PM_at	-24.9	JACALIN-RELATED LECTIN 33, JAL33	nucleus, plasmodesma	response to cold, response to zinc ion	Mannose-binding lectin superfamily protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: response to zinc ion, response to cold; LOCATED IN: nucleus; EXPRESSED IN: 6 plant structures; EXPRESSED DURING: LP.04 four leaves visible, seedling growth; CONTAINS InterPro DOMAIN/s: Mannose-binding lectin (InterPro:IPR001229); BEST Arabidopsis thaliana protein match is: myrosinase-binding protein-like protein-300B (TAIR:AT3G16440.1).
11	At3g08770_PM_at	-23.9	Not yet established	endomembrane system	lipid transport, response to water deprivation	Predicted to encode a PR (pathogenesis-related) protein. Belongs to the lipid transfer protein (PR-14) family with the following members: At2g38540/LTP1, At2g38530/LTP2, At5g59320/LTP3, At5g59310/LTP4, At3g51600/LTP5, At3g08770/LTP6, At2g15050/LTP7, At2g18370/LTP8, At2g15325/LTP9, At5g01870/LTP10, At4g33355/LTP11, At3g51590/LTP12, At5g44265/LTP13, At5g62065/LTP14, At4g08530/LTP15.
12	At4g29030_PM_at	-20.2	Not yet established	endomembrane system	Not yet established	Putative membrane lipoprotein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; EXPRESSED IN: 15 plant structures; EXPRESSED DURING: 7 growth stages).
13	At4g22490_PM_at	-20.1	Not yet established	endomembrane system	lipid transport	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein; FUNCTIONS IN: lipid binding; INVOLVED IN: lipid transport; LOCATED IN: endomembrane system; EXPRESSED IN: 22 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Bifunctional inhibitor/plant lipid transfer protein/seed storage (InterPro:IPR016140), Plant lipid transfer protein/seed storage/trypsin-alpha amylase inhibitor (InterPro:IPR003612), Plant lipid transfer protein/hydrophobic protein, helical domain (InterPro:IPR013770); BEST Arabidopsis thaliana protein match is: azelaic acid induced 1 (TAIR:AT4G12470.1).
14	At2g33850_PM_at	-17.7	Not yet established	endomembrane system	Not yet established	unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; EXPRESSED IN: 17 plant structures; EXPRESSED DURING: 9 growth stages; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT1G28400.1); Has 3053 Blast hits to 2119 proteins in 133 species: Archae - 6; Bacteria - 52; Metazoa - 135; Fungi - 96; Plants - 73; Viruses - 2; Other Eukaryotes - 2689 (source: NCBI BLink).
15	At5g13930_PM_at	-17.1	ATCHS, CHALCONE SYNTHASE, CHS, TRANSPARENT TESTA 4, TT4	endoplasmic reticulum, nucleus, plant-type vacuole membrane	flavonoid biosynthetic process	Encodes chalcone synthase (CHS), a key enzyme involved in the biosynthesis of flavonoids. Required for the accumulation of purple anthocyanins in leaves and stems. Also involved in the regulation of auxin transport and the modulation of root gravitropism

	Probe Set ID	FC ([14] vs [WT])	Gene Name	Location	Biological Process	Description
16	At4g22870_PM_s_at	-16.6	Not yet established	cellular component	oxidation-reduction process	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein; FUNCTIONS IN: oxidoreductase activity; LOCATED IN: cellular_component unknown; CONTAINS InterPro DOMAIN/s: Oxoglutarate/iron-dependent oxygenase (InterPro:IPR005123); BEST Arabidopsis thaliana protein match is: leucoanthocyanidin dioxygenase (TAIR:AT4G22880.2).
17	At3g04290_PM_at	-16.3	ATLTL1, LI-TOLERANT LIPASE 1, LTL1	endomembrane system	hyperosmotic salinity response, lipid metabolic process, metabolic process, response to lithium ion, response to salicylic acid stimulus	Li-tolerant lipase 1 (LTL1); FUNCTIONS IN: hydrolase activity, acting on ester bonds, carboxylesterase activity; INVOLVED IN: lipid metabolic process; LOCATED IN: endomembrane system; EXPRESSED IN: 19 plant structures; EXPRESSED DURING: 12 growth stages; CONTAINS InterPro DOMAIN/s: Lipase, GDSL (InterPro:IPR001087), Esterase, SGNH hydrolase-type (InterPro:IPR013830); BEST Arabidopsis thaliana protein match is: GDSL-like Lipase/Acylhydrolase superfamily protein (TAIR:AT5G33370.1).
18	At1g01060_PM_at	-15.4	LATE ELONGATED HYPOCOTYL, LATE ELONGATED HYPOCOTYL 1, LHY, LHY1	Not yet established	LHY encodes a myb-related putative transcription factor involved in circadian rhythm along with another myb transcription factor CCA1	circadian rhythm, long-day photoperiodism, flowering, negative regulation of circadian rhythm, negative regulation of sequence-specific DNA binding transcription factor activity, regulation of circadian rhythm, regulation of transcription, DNA-dependent, response to abscisic acid stimulus, response to auxin stimulus, response to cadmium ion, response to cold, response to ethylene stimulus, response to gibberellin stimulus, response to jasmonic acid stimulus, response to salicylic acid stimulus, response to salt stress
19	At1g73330_PM_at	-13.7	ATDR4, DR4, DROUGHT-REPRESSED 4	endomembrane system	negative regulation of peptidase activity, response to water deprivation	encodes a plant-specific protease inhibitor-like protein whose transcript level in root disappears in response to progressive drought stress. The decrease in transcript level is independent from abscisic acid level.
20	At1g06350_PM_s_at	-13.6	Not yet established	Not yet established	lipid metabolic process, oxidation-reduction process	Fatty acid desaturase family protein; FUNCTIONS IN: oxidoreductase activity, oxidoreductase activity, acting on paired donors, with oxidation of a pair of donors resulting in the reduction of molecular oxygen to two molecules of water; INVOLVED IN: oxidation reduction, lipid metabolic process; CONTAINS InterPro DOMAIN/s: Fatty acid desaturase, type 1, core (InterPro:IPR015876), Fatty acid desaturase, type 1 (InterPro:IPR005804); BEST Arabidopsis thaliana protein match is: Fatty acid desaturase family protein (TAIR:AT1G06360.1).
21	At5g42800_PM_at	-11.0	DFR, DIHYDROFLAVONOL 4-REDUCTASE, M318, TT3	extrinsic to endoplasmic reticulum membrane	anthocyanin-containing compound biosynthetic process, oxidation-reduction process	dihydroflavonol reductase. Catalyzes the conversion of dihydroquercetin to leucocyanidin in the biosynthesis of anthocyanins. Not expressed in roots (qRT-PCR).
22	At1g07180_PM_at	-10.0	ALTERNATIVE NAD(P)H DEHYDROGENASE 1, ARABIDOPSIS THALIANA INTERNAL NON-PHOSPHORYLATING NAD (P) H DEHYDROGENASE, ATNDI1, NDA1	intrinsic to mitochondrial inner membrane, mitochondrion	oxidation-reduction process	Internal NAD(P)H dehydrogenase in mitochondria. The predicted protein sequence has high homology with other designated NAD(P)H DHs from microorganisms; the capacity for matrix NAD(P)H oxidation via the rotenone-insensitive pathway is significantly reduced in the Atndi1 mutant plant line; the in vitro translation product of AtNDI1 is imported into isolated mitochondria and located on the inside of the inner membrane.

	Probe Set ID	FC ([14] vs [WT])	Gene Name	Location	Biological Process	Description
23	At1g62510_PM_at	-8.3	Not yet established	endomembrane system	lipid transport	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein; FUNCTIONS IN: lipid binding; INVOLVED IN: lipid transport; LOCATED IN: endomembrane system; EXPRESSED IN: shoot, cotyledon, leaf whorl, leaf; EXPRESSED DURING: LP.04 four leaves visible, LP.02 two leaves visible; CONTAINS InterPro DOMAIN/s: Bifunctional inhibitor/plant lipid transfer protein/seed storage (InterPro:IPR016140), Plant lipid transfer protein/seed storage/trypsin-alpha amylase inhibitor (InterPro:IPR003612), Plant lipid transfer protein/hydrophobic protein, helical domain (InterPro:IPR013770); BEST Arabidopsis thaliana protein match is: Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein (TAIR:AT4G12490.1)
24	At3g28220_PM_at	-7.8	Not yet established	chloroplast envelope, vacuole	Not yet established	TRAF-like family protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: response to salt stress; LOCATED IN: vacuole, chloroplast envelope; EXPRESSED IN: 19 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: TRAF-like (InterPro:IPR008974), MATH (InterPro:IPR002083); BEST Arabidopsis thaliana protein match is: TRAF-like family protein (TAIR:AT1G58270.1).
25	At1g32900_PM_at	-7.7	GBSS1, GRANULE BOUND STARCH SYNTHASE 1	chloroplast	biosynthetic process, glucan biosynthetic process, metabolic process	UDP-Glycosyltransferase superfamily protein; FUNCTIONS IN: protein binding, transferase activity, transferring glycosyl groups; INVOLVED IN: biosynthetic process, glucan biosynthetic process; LOCATED IN: chloroplast; EXPRESSED IN: 21 plant structures; EXPRESSED DURING: 13 growth stages; CONTAINS InterPro DOMAIN/s: Glycogen/starch synthases, ADP-glucose type (InterPro:IPR011835), Starch synthase, catalytic domain (InterPro:IPR013534), Glycosyl transferase, group 1 (InterPro:IPR001296); BEST Arabidopsis thaliana protein match is: starch synthase 2 (TAIR:AT3G01180.1)

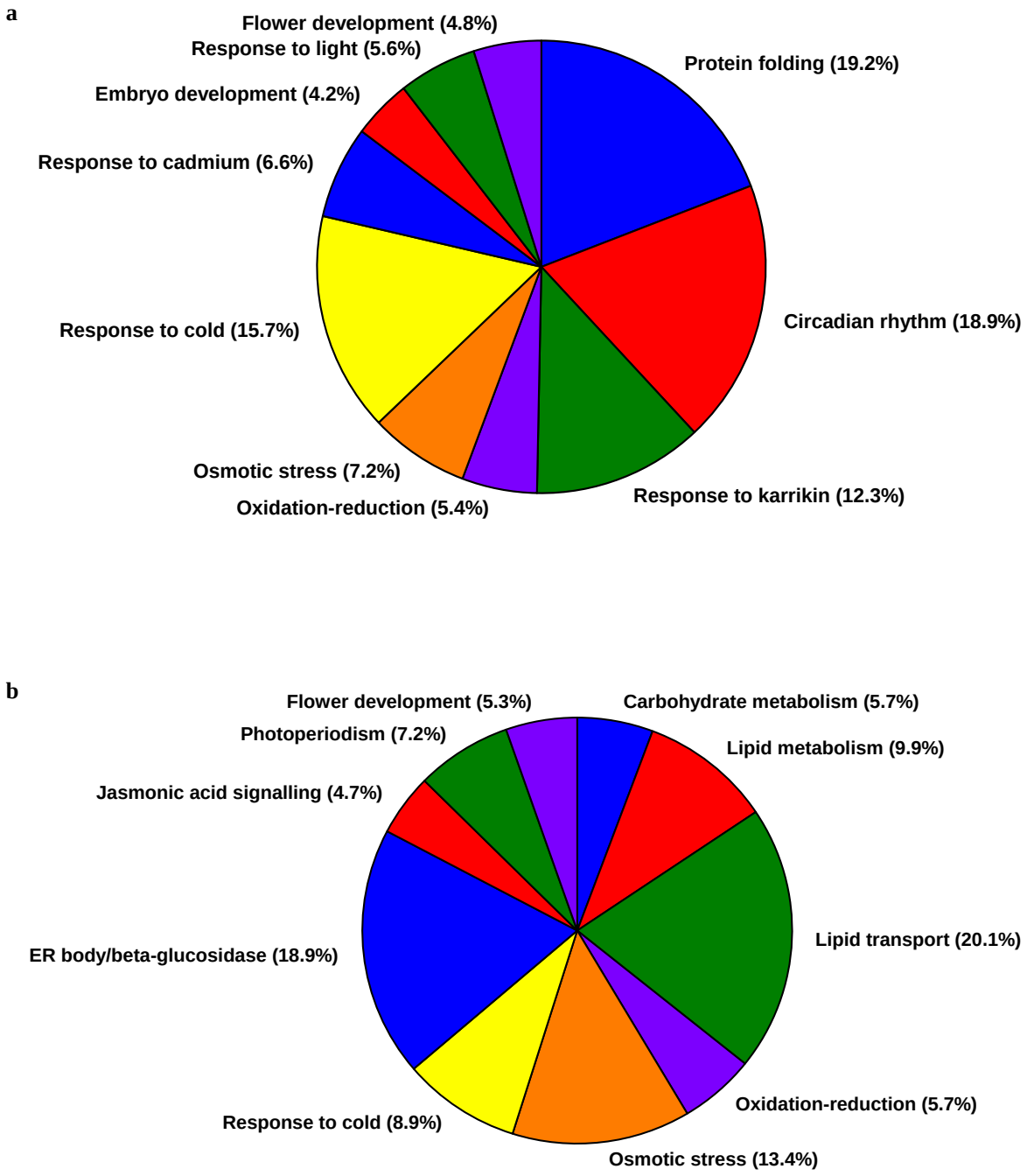


Figure 4-39. Top 10 (a) up-regulated and (b) down-regulated candidate genes according to biological function as a result of constitutive expression of CmP transgene.

CHAPTER 5
DISCUSSION

5.1 Pumpkin 2S vs. Analogous Napin Fusions—Elucidation of Sorting Signals

The goal of this work was to study the SSP sorting and trafficking mechanisms in Arabidopsis by using ectopic expression in leaf tissue. The crux of this strategy was the generation of transgenic Arabidopsis plants capable of fusion protein accumulation and vesicle biogenesis resulting from ectopic expression of SSP fusion transgenes. It was known that this can be readily achieved using a pumpkin 2S(Δ79)-PAT transgene previously shown by Hayashi et al, (1999) to result in the formation of leaf resident PAC vesicles which housed the pumpkin 2S(Δ79)-PAT fusion protein [19]. In an effort to advance knowledge of SSP mechanisms in *Brassica* ssp, a number of napin variant constructs were expressed ectopically in Arabidopsis leaves.

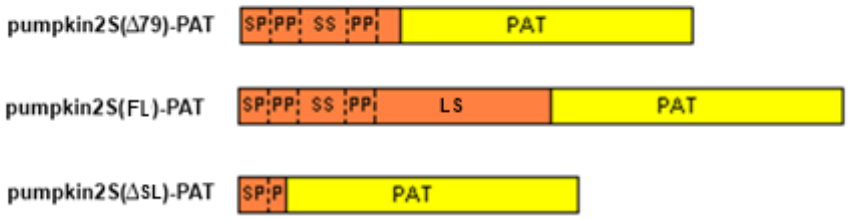
Assessment of transgenic plants developed for elucidation of sorting signals revealed that accumulation of pumpkin 2S(Δ79)-PAT and pumpkin 2S(Δ79)-GFP fusion proteins had occurred readily whereas the addition of the pumpkin 18 amino acid carboxy terminus to the mNap(B)-GFP open reading frame was required before appreciable accumulation was observed. However addition of the pumpkin C-terminus to either pumpkin 2S(Δ79)-PAT or mNap(pL4)-PAT open reading frames did not result in increased accumulation. It was thought, based on the secondary structure analysis, that mNap(pL4)-PAT would be analogous to pumpkin 2S(Δ79)-PAT M(4)P, but despite their structural similarity, accumulation of mNap(pL4)-PAT fusion protein was rare, occurring at low level in only three T_2 generation transgenic plants among a population of forty-two. In the event that some critical element had not been incorporated, open reading frames comprised of longer and full-length napin peptide, Nap2(pL5)-PAT and Nap2(FL)-PAT were employed and these too resulted in no accumulation (despite transgene

expression). This indicated the following[4]: 1) the pumpkin 18 amino acid C-terminus did not affect the sorting of fusion protein to vesicles in leaf tissue unless combined with a very short napin segment, the mNap(B) truncation; 2) the inclusion of napin peptide beyond probability loop region 4 (pL4) in the Nap2(pL5)-PAT and Nap2(FL)-PAT open reading frames did not improve the incidence of accumulation and therefore did not contribute to vesicle biogenesis and accumulation behavior in leaf tissue, 3) the region of napin beyond the BamHI site up to the pL4 truncation did not evoke and perhaps even prevented accumulation in leaf tissue similar to that observed for the pumpkin 2S(Δ79)-PAT, pumpkin 2S(Δ79)-PAT-CT and mNap(B)-GFP-CT open reading frames.; and 4) despite the low incidence and yield of mNap(pL4)-PAT compared to the pumpkin 2S(Δ79)-PAT, a common pathway and sorting mechanism for these two fusion proteins is suggested by accumulation of both the intact and processed variants of the fusion proteins. Had no difficulty been encountered in the development of transgenic Arabidopsis expressing the mNap(A)-GFP or mNap(S)-GFP open reading frames, perhaps more conclusions could have been drawn regarding which regions between the BamHI site and the pL4 were responsible for this effect. Regardless, although it is difficult to extrapolate the findings obtained with tobacco transient expression, constructs CmGFP, N(B)GFP-CT and N(X)GFP had similar sorting behavior which indicated that the pumpkin 2S(Δ79) peptide, the pumpkin 2S 18 amino acid C-terminal vacuolar sorting domain, and the analogous mNapin(pL4) peptide employed similar sorting mechanisms (see section 4.2.6).

Based on these findings, it is conceivable that elements within the pumpkin 2S(Δ79) region were sufficient for the observed sorting and accumulation behavior of CmP plants independent of the effects of the pumpkin 2S C-terminal vacuolar sorting domain. This supports the findings of Hayashi et al. (1999) who surmised that the RRE amino acid triplet of the

[4] I refer the reader to Figure 4-6 for graphical representation to assist with this discussion.

pumpkin albumin internal processed peptide was likely a binding site for pumpkin vacuolar protein sorting receptor PV72/82 [19]. Because they did not observe accumulation and vesicle biogenesis for PAT fusion constructs employing a shorter pumpkin albumin region (Figure 5-1, construct pumpkin 2S(ΔSL)-PAT), it is possible that the accumulation observed for CmP was instead due to the omission of some necessary element within the large subunit of the pro2S albumin. However the failure of the full-length pumpkin 2S-PAT (Figure 5-1, construct 2SP), which possessed all the potential sorting elements of the albumin, to accumulate fusion protein refutes this notion. Neither these researchers, nor anyone else to my knowledge, have since sought to unravel this paradox. However, based on my additional findings, the capacity of the CmP to evoke vesicle biogenesis and fusion protein accumulation in Arabidopsis leaves relies on both the RRE amino acid triplet of pumpkin pro2S albumin and the absence of the analogous region from napin. The failure of the napin-PAT or napin-GFP constructs to result in accumulation, except where only a very short region of the mNapin peptide is present (as in construct Nap(B)GFPCT), supports this suggestion. This then begs the question: why does the napin region comparable to the pumpkin 2S RRE triplet not elicit the same sorting behavior? PV72 and PV82 are thought to be VSR proteins of pumpkin that possess binding affinity for the pumpkin proprotein precursor 2S albumin. Shimada et al. (2002) examined this interaction in detail and demonstrated that PV72/82 binds *in vitro* to the internal processed peptide region and *in vivo* to both the internal processed peptide and the C-terminal region [70, 90, 97]. The evidence for the pumpkin 2S(Δ79) peptide to possess a unique capacity to evoke vesicle biogenesis and accumulation in Arabidopsis leaf tissue lies in that there was fusion protein accumulation in the analogous M(4)P plant leaf tissue, albeit very minor and for which no evidence of accumulation was observed for the re-synthesized napin-PAT fusion constructs



Open Reading Frame	PPT Resistance	PAC Vesicles
pumpkin 2S(FL)-PAT	+	-
pumpkin 2S(Δ79)-PAT	-	+
pumpkin 2S(ΔSL)-PAT	++	-

Figure 5-1. Pumpkin 2S albumin-PAT gene fusion constructs employed by Hayashi et al. (1999) and their phenotype with respect to PPT^R and capacity to induce the biogenesis of PAC-like vesicles. Adapted from Hayashi et al., 1999 [19].

GA-N4P, GA-N5P or GA-NFL. Because the analogous region in napin appears to prevent fusion protein accumulation to any significant capacity indicates that this region is involved in sorting behavior. Based on the bioinformatics analysis of the primary and secondary protein structures of the two albumins, it was recognized that an RRE triplet does not occur in the napus sequence, but it has instead a short segment of basic amino acids rich in amino group side-chains bracketed by proline residues, "PQGPQQRPP" (see Figure 4-1). The pumpkin 2S RRE triplet follows proline-tryptophan residues in "PWRREGGS" (see Figure 4-1) and is thought to be an exposed sorting receptor binding site [70, 90, 97]. Both triplets possess a +1 net positive charge at physiological pH 7. During passage through the endomembrane system, these triplet regions would be exposed to a gradual acidification but would retain their +1 net positive charge as the ER lumen is thought to be pH 7.1, the cis-Golgi pH 6.5, the MVB pH 6.2, and secretory vesicles between pH 5.0 and pH 6.0[5][195]. This suggests that the analogous region of napin sequence is chemically similar to the RRE of pumpkin 2S and could therefore also act as a sorting receptor binding site. The pumpkin VSR protein PV72 shares 74% sequence identity at the protein level with its orthogolous Arabidopsis VSR1. However, it is possible that the RRE region of pumpkin 2S does not possess the same binding affinity for VSR1 as does the "PQGPQQRPP" region of napin due to differences in binding site-receptor recognition. If true, this would explain the differing capacity to evoke accumulation of fusion protein but the question of lesser or greater affinity for VSR1 remains. Which would be likely: a lesser binding affinity of VSR1 for pumpkin RRE than the analogous region of napin or a greater affinity? Certainly, a lesser binding affinity of the pumpkin RRE for VSR1 would result in the fusion protein having a poorly recognized VSD and consequently a buildup of a portion of the fusion protein would occur

[5] The glutamate carboxyl side chain has pK_R=4.07 and therefore would not incur an additional +1 charge while residing within the endomembrane system.

within the ER. Conversely, a greater binding affinity might result in poor cargo-VSR uncoupling and ultimate depletion of VSR1 proteins due to prevention of retromer-mediated recycling [98-100]. Either way, a buildup of fusion protein residing in the ER gives credence to the notion that accumulation of pumpkin 2S(Δ79)-PAT fusion protein in PAC-like vesicles in leaves might have occurred because sorting was stalled due to an overload of the protein sorting machinery of the plant. Regardless, the question of why PAC-vesicle biogenesis and fusion protein accumulation was evoked by the pumpkin 2S(Δ79)-PAT open reading frame was proved to be more complex than anticipated, and this question was set aside to study aspects of sorting behavior and identifying genes of Arabidopsis involved in these processes.

5.2 Accumulation of Pumpkin 2S(Δ79)-PAT Fusion Protein

Analyses of transgenic plants confirmed the presence and ectopic expression of transgenes and indicated which SSP elements evoked accumulation of the resulting fusion protein. The question of why accumulation was or was not observed for the various transgenes is a complex one and is likely related to the capacity of the plant to tolerate SSP fusion protein in leaves. Certainly, elimination of the fusion protein by sorting to the proteosome might also have occurred due to misfolding or missorting. To understand these processes more clearly, it was necessary to demonstrate in which subcellular location the accumulating fusion protein resided and if it had been correctly folded as evidenced by functionality.

In the presence of phosphinothricin (PPT), wild-type Arabidopsis plants will germinate but fail to develop further unless the toxin is degraded by PAT enzyme as it enters the cytosol [196]. Hayashi et al. (1999) reported that the 2S(79)-PAT construct, that evoked the biogenesis of novel PAC-like vesicles, showed partial resistance to PPT despite accumulation of PAT [19]. Assessing resistance to phosphinothricin (PPT^R) proved to be a useful indicator of correct folding and protein functionality for the CmP and M(4)P transgenes. Despite a lack of evidence for appreciable accumulation of mNap(pL4)-PAT fusion protein, in M(4)P transgenic Arabidopsis plants had a PPT^R phenotype comparable to the CmP plants. This apparent conundrum indicated the reasons for accumulation and vesicle biogenesis were not related to poor expression of the M(4)P transgene or misfolding of the mNap(pL4)-PAT fusion protein, but were subtle and instead related to aspects of protein trafficking. If it can be surmised that only partial resistance to PPT observed for the CmP transgenic plants was due to incomplete compartmentalization, then it also seems plausible that partial resistance to PPT observed for M(4)P plants may have occurred for the same reason. The product of the PAT gene, phosphinothricin acetyl transferase, is thought to catalyze PPT deactivation as the toxin enters

the cytosol. PAT catalyzes acetylation of the free NH_2 group of PPT which otherwise impairs the function of glutamine synthetase and causes an interruption in the ammonia assimilation cycle [197]. In Arabidopsis, there are seven genes for glutamine synthetase, six of which encode a cytosolic enzyme whereas the remaining gene encodes a chloroplast enzyme. This suggests that site of PPT deactivation is the cytosol where the majority of glumatine synthetase enzyme resides. Considering this, PPT resistance would be acquired by plants in which at least a portion of pumpkin 2S(Δ79)-PAT and mNap(pL4)-PAT fusion proteins were liberated from the endomembrane system into the cytosol. There is an inclination to assume that the PAT fusion proteins were liberated into the cytosol where PPT is known to deactivate glutamine synthetase, but resistance to PPT would also occur if the fusion protein was liberated into the extracellular apoplast and detoxification of PPT occurred prior to its entry into the cell. Because PPT resistance appeared more robust for seedlings of M(4)P than CmP, trafficking of mNap(pL4)-PAT fusion protein to the apoplast may have occurred to a greater extent than for with the M(4)P than pumpkin 2S(Δ79)-PAT (see PPT^R at 20 μg/mL, Figure 4-17). It is known that without any vacuolar sorting determinant, proteins possessing the signal peptide enter the ER and by default, are secreted to the apoplast (see Figure 2.1). This would suggest that the mNap(pL4)-PAT open reading frame lacked a sorting determinant suitable for recognition by sorting receptor proteins. Certainly, because both CmP and M(4)P transgenes resulted in a PPT^R phenotype and that appreciable pumpkin 2S(Δ79)-PAT fusion protein but not mNap(pL4)-PAT was discovered in leaves, sorting of the fusion proteins using exclusive trafficking to the apoplast was not indicated and that multiple potential routes were employed.

Evidence for trafficking pumpkin 2S(Δ79)-PAT fusion protein by multiple routes was also suggested from the accumulation in CmP transgenic plants of intact pumpkin 2S(Δ79)-PAT

fusion protein as well as multiple processed peptides (section 4.1.2). Isolation of PAC-like vesicles by density gradient centrifugation indicated that the anti-PAT reactive protein recovered from the vesicle fractions were exclusively the intact fusion protein, whereas the processed variant was observed in the supernatant (see section 4.2.3, Figure 4-21 and 4-22). Vesicle transport of the fusion protein as an intact precursor was further suggested by the tobacco transient expression assay in which fluorescent punctuate structures were seen travelling rapidly by cytoplasmic streaming for constructs N(B)GFPCT, CmGFP and N(X)GFP (see section 4.2.6). Furthermore, western blot data agreed with the findings of Tamura, et al. (2003) [190] who had shown GFP to be susceptible to a vacuolar cysteine protease which eliminated GFP fluorescence indicating that only the unprocessed full length GFP fusion proteins transported in vesicles had been visualized. Accumulation of the GFP fusion protein did not occur consistently among T_3 generation plants derived from the same T_2 parent (Figure 4-14), which correlated with the conspicuous absence of GFP fluorescence in leaves of Napin-GFP stable transgenic plants (Figure 4-24). The reasons for the apparent inconsistency of fusion protein accumulation in GFP gene fusion transgenic plants were not clear but might be attributed to the same factors responsible for the inconsistency observed with the PAT gene fusion transgenic plants. Certainly the tendency for GFP to be degraded rapidly in the acidic vacuolar milieu by a cysteine protease is known to impair microscopic observation of GFP fluorescence in plants and would have reduced accumulation of intact fusion protein [190].

Accumulation of peptides reactive with anti-PAT antiserum not only of size correlating to the intact fusion protein, but also correlating to peptides from which internal propeptide regions had been excised, suggested that some of the expressed fusion protein had been subjected to peptide processing by vacuolar processing enzymes (VPEs), known to reside in the MVB (see

section 2.4.3) [96, 183]. Trafficking of the pumpkin 2S(Δ79)-PAT fusion protein to the MVB strongly indicated that the Golgi-dependent pathway was employed at least in instances where processed variants had been observed. However, the accumulation of a processed variant was not observed by Hayashi et al. (1999) who reported the accumulation of only the intact ~28 kDa pumpkin 2S(Δ79)-PAT fusion protein in PAC-like vesicles. This difference might be attributed to the use of green cotyledons from 14-day seedlings by Hayashi et al. (1999)[19], whereas 4-5 week old rosette leaves were used for my analyses. Two likely scenarios for the absence of the processed variant in younger tissue are 1) trafficking of the fusion protein during the early stages of accumulation employ a different sorting strategy than in older mature leaf tissue, or 2) the fusion protein directed to PAC-like vesicles is modified as its residence in PAC-like vesicles continues beyond 14 days. In the first scenario, the sorting strategy and the ultimately the tendency to accumulate processed variants, could have been correlated to the strength of the 35S promoter. Studies of the cauliflower mosaic virus reveal that viral replication, driven by the 35S promoter, occurs rapidly subsequent to leaf infection by an aphid vector [198]. Because aphids feed on developed leaves rather than small seedlings, 35S promoter expression likely occurs with more vigor in leaves than in seedlings. This suggested that incidence of accumulated process fusion protein variants in rosette leaves was the result of greater 35S expression in leaves used in this study than the seedling tissue used in the study by Hayashi study et al. (1999)[19]. Alternately for scenario 2, a residence time in PAC vesicles beyond 14 days may have resulted in development of processed variants. There has been considerable speculation as to the role of PAC vesicles in SSP trafficking. Some researchers theorize PAC vesicles originate due to an overwhelming demand on the endomembrane system caused by rapid SSP synthesis during seed development and PAC vesicles themselves act like PVCs in which cross-talk with the Golgi

apparatus delivers VPEs and other components essential for SSP maturation and deposition into target PSVs [18, 60]. That PAC-like vesicles function as PVCs and deposit their contents into PSVs or even evolve into them is supported by the use of the anti-αTIP and anti-AtRabA4b antibodies. In conjunction with differential centrifugation and western blot, these antibodies indicated that for CmP-18 samples the anti-αTIP and anti-AtRabA4b reactive proteins did not occur in the buoyant PAC-like vesicle fractions but instead occurred as a high molecular mass aggregation in the most dense vesicle fraction in which the processed fusion protein was also recovered (Figure 4-24d-f). Although hindered by non-specific binding of antibody-conjugated beads, the antibody-capture strategy employing the αTIP antibody resulted in the recovery of the processed fusion protein whereas use of the AtRabA4b antibody resulted in recovery of both the intact and processed forms. Although this antibody capture technique appeared to have successfully isolated membrane compartments that were the sites of fusion protein deposition, background effects attributed to the release of anti-αTIP antibody from M-450 beads were problematic. Furthermore, it was clear from silver stained gels that copious non-specific binding by the antibody-bead conjugates recovered huge populations of proteins in the final eluate. There were a number of possible reasons for this, including non-specific binding by the beads themselves, poor display of the antibody on these particular beads which are designed for isolation of large membrane bound targets, or even that the storage vesicles in leaf tissue that contained the pumpkin 2S(Δ79)-PAT fusion protein did not possess sufficient αTIP or Rab GTPase membrane associated targets. Curiously, a recent paper [199] reported the first use of affinity purification of trans-Golgi network vesicles employing antibody-coated beads with binding specificity towards the TGN binding SYP61 SNARE protein. The authors claim to remedy non-specific binding by pretreatment of vesicle extracts with unconjugated beads [199].

αTIP is an integral transmembrane protein implicated in PSV membrane composition that is thought to be directed to PSV membranes by a Golgi-independent mechanism [9, 200] and Rab GTPases direct vesicle transport to vacuolar targets through membrane association (see section 2.3.3). Therefore, based on these findings, it seems plausible that the anti-αTIP antibody permitted the recovery of ectopic PSVs containing processed fusion protein and that the anti-AtRabA4b antibody permitted capture of PSVs and PAC-like vesicles containing unprocessed fusion protein.

In addition to the presence of a processed variant fusion protein, the accumulation behavior for the CmP transgene product also differed from the findings of Hayashi et al. (1999)[19] in that they did not report an inconsistent accumulation among members of T_3 generation transgenic plants. That this inconsistency was due to continued segregation of transgenes was unlikely as the development of successive generations originating from accumulating parents failed to increase the frequency of accumulating progeny. Fusion protein accumulation, when it was observed, had only ever been indicated in rosette leaves of the same approximate age and size and never in cauline leaves of the stem. Furthermore, the inconsistency of fusion protein accumulation observed between leaves of a single plant indicated accumulation was instead correlated with leaf size, age, or position and/or other environmental factors. One possible explanation for apparent variability might be that accumulation behavior was attributed to differing expression levels between plant leaves and correlated with transgene copy number increased through endoreduplication. The CaMV 35S promoter driving transgene expression in this research is widely accepted as a constitutive expression promoter. Strong expression is consistently observed in the roots, cotyledons, leaves, and flower tissues of Arabidopsis [201], however, studies in tobacco indicate greatest expression levels in actively growing young tissues

or during the S phase of dividing protoplasts that seems to be correlated with DNA replication [202, 203]. Despite widespread use of the CaMV 35S promoter to drive transgene expression in numerous Arabidopsis tissues (except in seeds for which CaMV 35S expression does not appear to be constitutive), no correlation between expression levels and S phase DNA replication in Arabidopsis has been reported. Endoreduplication is a curious phenomenon in which an unpairing between cell division and nuclear genome replication occurs such that the cessation of mitosis at the end of the S phase of the cell cycle results in polyploidy and is thought to arise from modulation of cell cycle kinase (CDK) activity. Although endoreduplication occurs in certain insect and mammalian cell types, in plants, it appears to be a developmental mechanism coupled to cellular differentiation. The role of endoreduplication in plants is unclear but there is evidence indicating that it is a mechanism used by plants to buffer the plant genome against DNA damage and to mediate stress response by improving the plant's capacity to adapt to environmental conditions [204]. In Arabidopsis, genome content in leaf cells and trichomes increased through endoreduplication were estimated by microspectrofluorometry to be as high as 16C and 64C, respectively (where C equals the haploid chromosome content) [205], and appears to be governed by the cyclin A2 (CYCA2) gene family [206, 207].

5.3 Early vs. Late Gene Effects of Ectopic CmP Transgene Expression

Ectopic expression of SSP gene fusions could have considerable impact on gene expression in leaf tissue where metabolic machinery is normally devoted to photosynthesis. It was expected that the accumulation of SSP in leaves would oppose normal leaf physiology and that the plant would seek avenues for its speedy removal. Therefore, much of the gene activity caused by SSP expression was anticipated to be correlated with actions taken by leaf cells to remedy perturbations to the endomembrane system. In addition to this, for the ectopic expression of the CmP transgene, gene activity devoted to synthesis, sorting, trafficking and accumulation of the fusion protein into PAC-like vesicles was expected. Why plants accumulated pumpkin 2S(Δ79)-PAT fusion protein more readily than the other transgenic proteins in this study has proven to be complex (see section 4.1 and 4.2). Continued elucidation of sorting signals to discover these reasons was possible but was not pursued further in favor of discovering *how* these processes occur. The apparent unique capacity of the CmP transgene to result in accumulation in leaves presented an opportunity to study SSP sorting and the vesicle biogenesis machinery at the gene level.

Gene expression activity for these elaborate and multistage processes was expected to be complex, dynamic and dependant on numerous potential developmental and environmental factors affecting the plant. As nascent SSPs traverse through the various folding, modifying and sorting check points of the endomembrane system, the required gene activity would be expected to change. The dexamethasone inducible expression system permitted precise initiation of the CmP transgene which allowed for the identification of genes activated at the onset of fusion protein synthesis followed by those activated at later stages as the plant responded to the challenge of accommodating fusion protein deposition.

To facilitate the microarray study, a suite of constructs were made in which several open reading frames were integrated into the pHTOP binary vector, the target of the dexamethasone activated LhGR-N transcription factor fusion (Figure 4-25). These open reading frames included pumpkin 2S(Δ79)-PAT, mNap(pL4)-PAT, PAT(no SP), and the empty vector control. The rationale for these open reading frames being that pumpkin 2S(Δ79)-PAT was capable of evoking gene expression necessary for accumulation and vesicle genesis, whereas mNap(pL4)-PAT could not. It was thought that the differences in gene expression between pumpkin 2S(Δ79)-PAT and mNap(pL4)-PAT would facilitate identification of genes induced by the unique elements of the pumpkin 2S(Δ79)-PAT and would account for any gene effects attributed to PAT enzyme routed through the secretory pathway of the endomembrane system. The PAT(no SP) open reading frame, which did not possess a signal peptide, accounted for the gene effects attributed to the presence of active PAT enzyme in the cytosol. This was important in the event that incomplete compartmentalization or liberation of the pumpkin 2S(Δ79)-PAT fusion protein from vesicles occurred. Together with the empty vector, the PAT open reading frame also served to account for gene activation caused by the actual dexamethasone application and concomitant expression of the GUS reporter.

Several double transformed Arabidopsis had been developed to the T_3 generation and included CmP, M(4)P, PAT(no SP) and empty vector. However, because the large-scale microarray analysis needed to be conducted within finite constraints of available reagents, AGRONOMICS1 gene chips and time required for hybridizations and data analysis, most of these were not used. It was known previously that the tendency to accumulate fusion protein was variable and this necessitated choosing T_4 plants based on the findings of a preliminary western blot analysis. Therefore, two groups of CmP:pHTOP transgenic T_4 plants from T_3 generation

parents were selected, namely CmP-3.2.4 and -3.2.6. This choice was based on western blot analysis of T_3 generation plants in which plant CmP-3.2.4 was the only one showing evidence of the intact unprocessed ~28 kDa peptide, a hallmark of Golgi-independent sorting and accumulation of fusion protein of the into vesicles (Figures 4-20 and 4-28). Because CmP-3.2.6 indicated accumulation of only the ~26 kDa processed variant which is indicative of Golgi-mediated delivery to the MVB, it was thought that subtle differences in gene expression attributed to use of the Golgi-dependent vs. Golgi-independent pathways might be detected. The T_4 generation PAT:pHTOP-3.1.1 plants were employed as controls by "dynamic pairing" whereby the gene expression of CmP-3.2.4 and CmP-3.2.6 could be weighed against PAT-3.1.1 at each time point. Many of the most dramatic changes in gene expression observed for the CmP transgenic plants were observed also for the PAT control and were nullified by this pairing. Exclusion of genes expressed during PAT:pHTOP transgene activation may have resulted in the loss of gene candidates important for protein synthesis and trafficking, but a means to correct for the substantial gene effects anticipated for PAT and GUS reporter expression as well as dexamethasone watering was necessary. Normalizing the data in this way was intended to identify candidate genes specific to fusion protein trafficking and vesicle biogenesis.

Candidate up-regulated genes identified for CmP-3.2.4 included one glutathione transferase (At2g29460) and three kinases (At4g23150, At1g21240 and At1g51890) involved in response to biotic or abiotic stress with one specifically localized to the endomembrane system. In comparison, up-regulated genes identified for CmP-3.2.6 included a number of proteins (six of the thirteen up-regulated candidates) which have ER resident functions or are known to associate with heat-shock-protein (HSP) chaperones involved in correct protein folding (At5g48570, At5g51440, At1g59860, At5g64510, At4g21870 and At4g12400). One other up-regulated

candidate included a Golgi-associated enzyme involved in cellulose (At1g55850) biosynthesis. Although dolichol biosynthesis in the chloroplast (At5g58770), has not been correlated with an ER residency, dolichol is known to be required for correct glycosylation of proteins for passage through the endomembrane system. This indicated that for CmP-3.2.6 plants dexamethasone activation resulted in a greater throughput of pumpkin 2S (Δ79)-PAT fusion protein by way of the Golgi-dependent pathway and the gene activation correlated with ER stress and protein folding mechanisms indicated that a buildup of fusion protein in the ER may have occurred. Conversely for CmP-3.2.4 plants, a high throughput of pumpkin 2S (Δ79)-PAT fusion protein, which would pose challenges for the ER, was not indicated. This suggested that the Golgi-independent pathway, which alleviates stress on the ER by shuttling ER accumulating protein to PAC vesicles [108], was more active in these plants than the Golgi-dependent pathway. This correlates well with having seen by western blot analysis the propensity for CmP-3.2.4 plants to accumulate the intact ~28 kDa pumpkin 2S (Δ79)-PAT fusion protein, whereas for CmP-3.2.6 the ~26 kDa processed variant had accumulated (Figure 4-28). However, despite these findings, the intact ~28 kDa or the ~26 kDa processed variant was not observed in western blot analysis following dexamethasone induction of these plants in this study (Figure 4-35).

Other up-regulated genes of known function in CmP-3.2.6 included responses to abiotic stress (At5g20250), protein phosphorylation (At4g24400), cell to cell signaling (At1g61566), and the *FLOWERING TIME* (*FT*) locus (At1g65480). Up-regulation of the FT locus is especially interesting as other researchers have speculated on a correlation between flowering time/photoperiodism and endomembrane trafficking [208]. Sohn et al. (2007) [209]reported a mutation that interrupted the function of *TERMINAL FLOWER1* (*TFL1*), a gene with opposing function to FT, resulted in impaired trafficking to a leaf PSV-like organelle and co-localization

with an adaptin AP-3 complex involved in vacuolar biogenesis [209]. Furthermore, Ebine et al. (2012) recently have shown that promotion of flowering is dependent on the vacuolar SNARE *SYP22* which if impaired, results in delayed flowering due to effects on *FLOWERING LOCUS C* (*FLC*) expression [208]. Together, these researchers speculate that the putative vegetative PSV-like organelle harbors factors critical for the correct transition from vegetative to reproductive phases. Although up-regulation of FT would suggest a reduced time to flowering due to its opposing role to TFL1, no differences in flowering time due to dexamethasone induction of CmP-3.2.4 and -3.2.6 plants were noted.

Dexamethasone induction appeared to have resulted in down-regulation of genes devoted to biosynthetic processes. In CmP-3.2.4 plants, genes involved in biosynthesis of carbohydrate (At1g09350), wax (At5g37300), and jasmonic acid (At5g07010) synthesis were reduced. Similarly, the down-regulated genes observed in CmP-3.2.6 included numerous genes involved in biosynthetic processes: galactinol synthase (At1g09350), cysteine biosynthesis (At3g54400), phenazine biosynthesis (At4g02850), lipid metabolism (At1g06350) and jasmonic acid response (At2g39030). In addition to galactinol synthase (At1g09350) which is known to be involved in response to cold, cold acclimation gene *LOW TEMPERATURE INDUCED* (*LTI30*) (At3g50970) was also down-regulated.

Much more dramatic alterations in gene expression were observed for CaMV 35S constitutive expression of the CmP transgene in CmP-14.1.3 and CmP-18.1.2 plants. Compared to wild-type controls, increased expression of genes implicated in protein folding and/or circadian rhythm was observed: nine of 20 up-regulated genes in CmP-14.1.3 and 13 of 23 up-regulated genes in CmP-18.1.2. As expected from their accumulation behavior observed previously by western blot analysis, the gene expression profiles of CmP-14.1.3 and 18.1.2 were

similar with the majority of up-regulated genes being shared. In fact for each, the greatest fold-change increase was observed for a DnaJ domain containing protein (At5g23240). The J-domain proteins are a large group of proteins well represented in Arabidopsis that function as co-chaperones with heat shock protein HSP70, DnaK, and a nucleotide exchange factor GrpE to form a molecular chaperone complex with multiple functions including promotion and isomerization of disulfide bonds through protein disulfide isomerase activity (see section 2.4.1) [210-212]. Curiously, the J-domain has been recently implicated in modulation of flowering time through transcriptional regulation and integration of flowering signals necessary for the transition from vegetative to reproductive development [213]. Increased expression of another flowering gene, *EARLY FLOWERING 4 (ELF4)* (At2g40080) was identified in both CmP-14.1.3 and 18.1.2 where approximately a 40-fold increase was observed. ELF4 is known, through its association with GIGANTEA (GI), to promote the expression of *CONSTANS1* (*CO*) for which a direct positive correlation between the amount of CO protein and *FT* transcript serves to strongly influence flowering time [214]. Interestingly, concomitant with increased gene expression of *ELF4*, a reduction in genes having opposing function, *CIRCADIAN CLOCK ASSOCIATED1* (*CCA1*)(At2g46830) and *LATE ELONGATED HYPOCOTYL* (*LHY*), were observed to have a ~28-fold and ~15-fold reduction, respectively. These findings appear to confirm the association between flowering regulation and vacuolar trafficking and suggests that a high throughput of pumpkin 2S (Δ79)-PAT fusion protein in the endomembrane system caused a ramping-up of machinery devoted to flowering time. This is most likely an effect of greatly increased expression of *DnaJ* (~148 fold change) in response to the demands of fusion protein folding and sorting caused by constitutive CaMV 35S expression of the CmP transgene. In this scenario, it is understandable that high *DnaJ* expression would have these ramifications. Despite these

findings, although a detailed examination of flowering time of these transgenic plants compared to their controls was not conducted, no differences were noted.

Similar to what was observed for dexamethasone induction, 14 of the top 25 decreased genes for both CmP-14.1.3 and 18.1.2 are involved in ER function essential for carbohydrate and lipid biosynthesis or transport. Interestingly, the greatest reduction in gene activity was indicated for an ER-associated beta-glucosidase, *BGLU23* (At3g09260), that is a major constituent of ER bodies: structures that are hallmarks of ER stress prior to cell senescence in response to wounding [215]. This would suggest expression of the CmP transgene resulted in a ramping-up of machinery necessary for high protein through-put at the expense of genes devoted to biosynthesis and cessation of ER functioning.

In summary, the microarray study identified genes essential for SSP trafficking and vesicle biogenesis. Dexamethasone induction of the CmP transgene resulted in changes in gene expression attributed to short duration fusion protein synthesis and trafficking. By using a dynamic pairing method, much of the effect on gene expression was found to be attributed to dexamethasone watering and/or passage of the PAT enzyme through the endomembrane system and consequently a relatively small, but targeted numbers of genes were identified. Conversely, gene affects resulting from CaMV 35S promoter driven expression were likely due to longer duration accumulation of fusion proteins. Despite the differences between low-level dexamethasone induced fusion protein synthesis and higher level accumulation resulting from constitutive CaMV 35S expression, common trends were observed. Both indicated perturbations of the endomembrane system affecting genes implicated in protein folding, flowering time and ER-associated biosynthetic function. Certainly, an increased demand on endomembrane function and ER protein throughput was indicated, as was modulation of flowering time and

photoperiodism being highly dependent on protein trafficking and vacuolar biogenesis mechanisms. By its nature, microarray experimentation is exploratory and serves to identify candidate genes for further study. Often the value of microarray data is not truly realized until the function of these candidate genes is elucidated through functional genomics.

CHAPTER 6
CONCLUSIONS AND FUTURE RESEARCH

Arabidopsis plants successfully expressing SSP fusion transgenes were produced and genes potentially involved in vesicle biogenesis and SSP sorting and trafficking machinery were identified. The capacity to sequester substantial quantities of unprocessed fusion protein in PAC-like vesicles was found to be a unique quality of the pumpkin(Δ79) peptide region and not a general phenomenon as expected from shared secondary structural features. The different accumulation behavior observed between the pumpkin 2S (Δ79)-PAT and mNap(pL4)-PAT fusion proteins was most likely due to a differing binding affinity of the pumpkin 2S peptide's RRE amino acid triplet for VSR1 sorting protein compared to the similar region of napin. The notion that VSR1 possesses some affinity towards this region in both pumpkin and napin peptides was supported by PPT^R assays in that both CmP and M(4)P showed partial resistance to PPT. This illustrated that passage of the fusion protein to the PVC/MVB via the Golgi-mediated pathway for ultimate secretion to the apoplast had occurred and this targeting also resulted in the accumulation of processed variants that were absent in purified vesicle fractions. Thus, both Golgi-dependent and independent pathways were active resulting in fusion protein secretion and processing, and accumulation in PAC-like vesicles, respectively. For the napin gene fusions, it appeared that use of the Golgi-mediated pathway dominated because accumulation of anti-PAT reactive protein in leaf tissue was rare yet transgenic plants exhibited little sensitivity to PPT.

The consequences of high fusion protein throughput in the ER was exemplified by the microarray study of CaMV 35S expression lines in which gene activity devoted to fusion protein folding and the creation of disulfide bonds was up regulated largely at the expense of genes involved in ER-mediated biosynthesis and cessation of ER function. Use of a dynamic pairing

method indicated that many of the effects on gene expression were attributed to dexamethasone watering and/or passage of the PAT enzyme through the endomembrane system. Consequently a relatively small, but targeted number of genes were identified. Although the short term expression of the CmP transgene facilitated by dexamethasone induction resulted in the accumulation of only a ~12 kDa process variant, the gene expression changes suggested that the Golgi-independent pathway dominated for CmP-3.2.4 plants whereas the Golgi-dependent pathway was employed by CmP-3.2.6 plants. Both induced and constitutive expression of the CmP transgene indicated that gene activities involved in flowering time and fusion protein throughput in the endomembrane system increased in tandem which confirmed that a strong correlation exists between flowering regulation and endocytic vacuolar trafficking.

To assess the contribution of VSR1 binding and channeling of nascent proteins through the Golgi-independent route, the chemical genomics tools brefeldin (BFA) and wortmannin could be used to great advantage. Exposure to BFA results in impaired recycling endosomal retrograde traffic from the Golgi to the ER by its action on ARF-GTPase, whereas wortmannin blocks retrograde traffic from the PVC/MVB by inhibition of phosphoinositol-3-kinase. Examination of the effects that these chemicals have on PAC-like vesicle biogenesis and fusion protein trafficking, together with an elaborate study of VSR1 protein-protein interactions involving mutagenized receptor protein variants or VSR deletion mutants, would permit elucidation of sorting and trafficking mechanisms. Deglycosylation analysis aimed at identifying carbohydrate modifications indicative of passage through Golgi subdomains would augment this work. Although GFP readily indicated vacuolar compartmentalization, due to its susceptibility to vacuolar cysteine protease cleavage, its use was limited for *in situ* tracking of fusion protein. Instead, tracking of various fluorescent protein fusions through the endomembrane system by

real time microscopy, perhaps together with dark period incubations, would provide interesting findings. More stable fluorescent proteins such as yellow fluorescent protein (YFP) or red fluorescent protein (RFP) are now available and would be ideal for a continuation of this work.

This microarray work presented a number of candidate genes which through loss-of-function and protein-protein interaction studies might be found to have wide ranging implications for endomembrane system function. To pursue these avenues, the CaMV 35S constitutive expression and dexamethasone inducible expression plant lines developed in this work will prove indispensible as they will permit further examination and confirmation of gene activity through continued microarray and quantitative real-time PCR experimentation.

As a final note, the observation that CmP fusion protein was unlikely to accumulate except in leaf tissues of a certain size and age was an unexpected conundrum. If in fact accumulation of fusion protein was correlated with endoreduplication, the CmP transgenic Arabidopsis plants would be useful for further examination of this phenomenon; however, the correlation between transgene expression levels, genome content and accumulation behavior would first need to be confirmed.

LITERATURE CITED

1. Robinson, D.G., P. Oliviusson, and G. Hinz, *Protein sorting to the storage vacuoles of plants: a critical appraisal.* Traffic, 2005. **6**(8): p. 615-25.
2. Gutierrez, L., G. Conejero, M. Castelain, S. Guenin, J.L. Verdeil, B. Thomasset, and O. Van Wuytswinkel, *Identification of new gene expression regulators specifically expressed during plant seed maturation.* J Exp Bot, 2006. **57**(9): p. 1919-32.
3. Wanasundara, J.P.D., *Proteins of Brassicaceae oilseeds and their potential as a plant protein source.* Crit Rev Food Sci Nutr, 2011. **51**(7): p. 635-677.
4. Shewry, P.R., J.A. Napier, and A.S. Tatham, *Seed storage proteins: structures and biosynthesis.* Plant Cell, 1995. **7**(7): p. 945-56.
5. Robinson, D.G. and J.C. Rogers, *Vacuolar compartments in plants*. Ann Plant Rev. Vol. 5. 2000, Sheffield.
6. Frigerio, L., G. Hinz, and D. Robinson, *Multiple vacuoles in plant cells: rule or exception?* Traffic, 2008. **9**: p. 1564 - 1570.
7. Rogers, J.C., *Multiple Vacuoles in Plant Cells.* Plant Physiol, 2008. **146**(3): p. 1024-1025.
8. Paris, N., C.M. Stanley, R.L. Jones, and J.C. Rogers, *Plant cells contain two functionally distinct vacuolar compartments.* Cell, 1996. **85**(4): p. 563-72.
9. Park, M., S.J. Kim, A. Vitale, and I. Hwang, *Identification of the protein storage vacuole and protein targeting to the vacuole in leaf cells of three plant species.* Plant Physiol, 2004. **134**(2): p. 625-39.
10. Bassham, D.C., F. Brandizzi, M.S. Otegui, and A.A. Sanderfoot, *The secretory system of Arabidopsis.* Arabidopsis Book, 2008. **6**: p. e0116.
11. Zouhar, J. and E. Rojo, *Plant vacuoles: where did they come from and where are they heading?* Curr Opin Plant Biol, 2009. **12**(6): p. 677-84.
12. Sanmartin, M., A. Ordonez, E.J. Sohn, S. Robert, J.J. Sanchez-Serrano, M.A. Surpin, N.V. Raikhel, and E. Rojo, *Divergent functions of VTI12 and VTI11 in trafficking to storage and lytic vacuoles in Arabidopsis.* Proc Natl Acad Sci U S A, 2007. **104**(9): p. 3645-50.
13. Rojo, E., C.S. Gillmor, V. Kovaleva, C.R. Somerville, and N.V. Raikhel, *VACUOLELESS1 is an essential gene required for vacuole formation and morphogenesis in Arabidopsis.* Dev Cell, 2001. **1**(2): p. 303-10.
14. Viotti, C., J. Bubeck, Y.D. Stierhof, M. Krebs, M. Langhans, W. van den Berg, W. van Dongen, S. Richter, N. Geldner, J. Takano, G. Jurgens, S.C. de Vries, D.G. Robinson, and K. Schumacher, *Endocytic and secretory traffic in Arabidopsis merge in the trans-Golgi network/early endosome, an independent and highly dynamic organelle.* Plant Cell, 2010. **22**(4): p. 1344-57.
15. Hicks, G.R. and N.V. Raikhel, *Opportunities and challenges in plant chemical biology.* Nat Chem Biol, 2009. **5**(5): p. 268-72.
16. Hicks, G.R. and N.V. Raikhel, *Advances in dissecting endomembrane trafficking with small molecules.* Curr Opin Plant Biol, 2010. **13**(6): p. 706-13.
17. Hicks, G.R. and N.V. Raikhel, *Small molecules present large opportunities in plant biology.* Annu Rev Plant Biol, 2012. **63**: p. 261-82.

18. Ibl, V. and E. Stoger, *The formation, function and fate of protein storage compartments in seeds.* Protoplasma, 2012. **249**(2): p. 379-92.
19. Hayashi, M., K. Toriyama, M. Kondo, I. Hara-Nishimura, and M. Nishimura, *Accumulation of a fusion protein containing 2S albumin induces novel vesicles in vegetative cells of Arabidopsis.* Plant Cell Physiol, 1999. **40**(3): p. 263-72.
20. Sparkes, I.A., L. Frigerio, N. Tolley, and C. Hawes, *The plant endoplasmic reticulum: a cell-wide web.* Biochem J, 2009. **423**(2): p. 145-55.
21. Herman, E.M. and B.A. Larkins, *Protein storage bodies and vacuoles.* Plant Cell, 1999. **11**(4): p. 601-14.
22. Vitale, A. and G. Hinz, *Sorting of proteins to storage vacuoles: how many mechanisms?* Trends Plant Sci, 2005. **10**(7): p. 316-23.
23. Andreeva, A.V., M.A. Kutuzov, D.E. Evans, and C.R. Hawes, *The structure and function of the Golgi apparatus: a hundred years of questions.* J Exp Bot, 1998. **49**(325): p. 1281-1291.
24. Moore, P.J., K.M. Swords, M.A. Lynch, and L.A. Staehelin, *Spatial organization of the assembly pathways of glycoproteins and complex polysaccharides in the Golgi apparatus of plants.* J Cell Biol, 1991. **112**(4): p. 589-602.
25. Fitchette, A.C., M. Cabanes-Macheteau, L. Marvin, B. Martin, B. Satiat-Jeunemaitre, V. Gomord, K. Crooks, P. Lerouge, L. Faye, and C. Hawes, *Biosynthesis and immunolocalization of Lewis a-containing N-glycans in the plant cell.* Plant Physiol, 1999. **121**(2): p. 333-44.
26. Saint-Jore-Dupas, C., V. Gomord, and N. Paris, *Protein localization in the plant Golgi apparatus and the trans-Golgi network.* Cell Mol Life Sci, 2004. **61**(2): p. 159-71.
27. Saint-Jore-Dupas, C., A. Nebenführ, A. Boulaflous, M.-L. Follet-Gueye, C. Plasson, C. Hawes, A. Driouich, L. Faye, and V. Gomord, *Plant N-Glycan Processing Enzymes Employ Different Targeting Mechanisms for Their Spatial Arrangement along the Secretory Pathway.* Plant Cell Online, 2006. **18**(11): p. 3182-3200.
28. Scheuring, D., C. Viotti, F. Kruger, F. Kunzl, S. Sturm, J. Bubeck, S. Hillmer, L. Frigerio, D.G. Robinson, P. Pimpl, and K. Schumacher, *Multivesicular bodies mature from the trans-Golgi network/early endosome in Arabidopsis.* Plant Cell, 2011. **23**(9): p. 3463-81.
29. Katzmann, D.J., G. Odorizzi, and S.D. Emr, *Receptor downregulation and multivesicular-body sorting.* Nat Rev Mol Cell Biol, 2002. **3**(12): p. 893-905.
30. Otegui, M.S., R. Herder, J. Schulze, R. Jung, and L.A. Staehelin, *The proteolytic processing of seed storage proteins in Arabidopsis embryo cells starts in the multivesicular bodies.* Plant Cell, 2006. **18**(10): p. 2567-81.
31. Richter, S., U. Voss, and G. Jurgens, *Post-Golgi traffic in plants.* Traffic, 2009. **10**(7): p. 819-28.
32. Park, M. and G. Jurgens, *Membrane traffic and fusion at post-Golgi compartments.* Front Plant Sci, 2012. **2**: p. 111.
33. Foresti, O., D.C. Gershlick, F. Bottanelli, E. Hummel, C. Hawes, and J. Denecke, *A recycling-defective vacuolar sorting receptor reveals an intermediate compartment situated between prevacuoles and vacuoles in tobacco.* Plant Cell, 2010. **22**(12): p. 3992-4008.

34. De Marcos Lousa, C., D.C. Gershlick, and J. Denecke, *Mechanisms and concepts paving the way towards a complete transport cycle of plant vacuolar sorting receptors.* Plant Cell, 2012. **24**(5): p. 1714-32.
35. Zheng, H. and L.A. Staehelin, *Protein Storage Vacuoles Are Transformed into Lytic Vacuoles in Root Meristematic Cells of Germinating Seedlings by Multiple, Cell Type-Specific Mechanisms.* Plant Physiol, 2011. **155**(4): p. 2023-2035.
36. Tan-Wilson, A.L. and K.A. Wilson, *Mobilization of seed protein reserves.* Physiol Planta, 2012. **145**(1): p. 140-153.
37. Osborne, T.B., *The vegetable proteins*. Monographs on biochemistry. 1909, London,: New York etc. Longmans, Green. xiii, 125 p.
38. Shewry, P.R. and N.G. Halford, *Cereal seed storage proteins: structures, properties and role in grain utilization.* J Exp Bot, 2002. **53**(370): p. 947-58.
39. Mandal, R.K. and S. Mandal, *Seed storage proteins and approaches for improvement of their nutritional quality by genetic engineering.* Curr Sci, 2000. **79, NO. 5, 10** (5): p. 576-589.
40. Li, C., M. Li, J.M. Dunwell, and Y.M. Zhang, *Gene duplication and an accelerated evolutionary rate in 11S globulin genes are associated with higher protein synthesis in dicots as compared to monocots.* BMC Evol Biol, 2012. **12**: p. 15.
41. Youle, R.J. and A.H. Huang, *Occurrence of low molecular weight and high cysteine containig albumin storage proteins in oilseeds of diverse species.* American Journal of Botany, 1981. **68**(1): p. 44-48.
42. Templeman, T.S., A.E. Demaggio, and D.A. Stetler, *Biochemistry of fern spore germination: globulin storage proteins in matteuccia struthiopteris L.* Plant Physiol, 1987. **85**(2): p. 343-9.
43. Templeman, T.S., D.B. Stein, and A.E. DeMaggio, *A fern spore storage protein is genetically similar to the 1.7 seed storage protein of Brassica napus.* Biochemical genetics Oct, 1988. **26**(9-10): p. 595-603.
44. Krebbers, E., L. Herdies, A. De Clercq, J. Seurinck, J. Leemans, J. Van Damme, M. Segura, G. Gheysen, M. Van Montagu, and J. Vandekerckhove, *Determination of the processing sites of an Arabidopsis 2S albumin and characterization of the complete gene family.* Plant Physiol, 1988. **87**(4): p. 859-66.
45. Agnieszka, B., B. Byczynska, and J. Barciszweski, *The biosynthesis, structure and properties of napin - the storage protein from rape seeds.* J Plant Physiol, 1999. **154**: p. 417-125.
46. de Souza Cândido, E., M.F.S. Pinto, P.B. Pelegrini, T.B. Lima, O.N. Silva, R. Pogue, M.F. Grossi-de-Sá, and O.L. Franco, *Plant storage proteins with antimicrobial activity: novel insights into plant defense mechanisms.* FASEB J, 2011. **25**(10): p. 3290-3305.
47. Mandal, S., P. Kundu, B. Roy, and R.K. Mandal, *Precursor of the inactive 2S seed storage protein from the Indian mustard Brassica juncea is a novel trypsin inhibitor. Characterization, post-translational processing studies, and transgenic expression to develop insect-resistant plants.* J Biol Chem, Oct, 2002. **277**(40): p. 37161-37168.
48. Barciszewski, J., S. Maciej, and T. Haertlé, *Minireview: Analysis of rape seed napin structure and potential roles of the storage protein.* J Prot Chem, 2000. **19**(4): p. 249-254.
49. Pantoja-Uceda, D., O. Palomares, M. Bruix, M. Villalba, R. Rodríguez, M. Rico, and J. Santoro, *Solution structure and stability against digestion of rproBnIb, a recombinant 2S*

albumin from rapeseed: Relationship to its allergenic properties, Biochemistry, 2004. **43**(51): p. 16036-16045.
50. Sharief, F.S. and S.S. Li, *Amino acid sequence of small and large subunits of seed storage protein from Ricinus communis.* J Biol Chem, 1982. **257**(24): p. 14753-9.
51. Lilley, G.G. and A.S. Inglis, *Amino acid sequence of conglutin δ, a sulfur-rich seed protein of Lupinus angustifolius L.: Sequence homology with the C-III α-amylase inhibitor from wheat.* FEBS Lett, 1986. **195**(1–2): p. 235-241.
52. Galau, G.A., H.Y. Wang, and D.W. Hughes, *Cotton Mat5-A (C164) gene and Mat5-D cDNAs encoding methionine-rich 2S albumin storage proteins.* Plant Physiol, 1992. **99**(2): p. 779-82.
53. Hara-Nishimura, I., K. Takeuchi, K. Inoue, and M. Nishimura, *Vesicle transport and processing of the precursor to 2S albumin in pumpkin.* Plant J, 1993. **4**: p. 793-800.
54. Jiang, L., T.E. Phillips, C.A. Hamm, Y.M. Drozdowicz, P.A. Rea, M. Maeshima, S.W. Rogers, and J.C. Rogers, *The protein storage vacuole: a unique compound organelle.* J Cell Biol, 2001. **155**(6): p. 991-1002.
55. Hoh, B., G. Hinz, B.K. Jeong, and D.G. Robinson, *Protein storage vacuoles form de novo during pea cotyledon development.* J Cell Sci, 1995. **108 (Pt 1)**: p. 299-310.
56. Hofte, H., L. Hubbard, J. Reizer, D. Ludevid, E. Herman, and M. Chrispeels, *Vegetative and seed-specific forms of tonoplast intrinsic protein in the vacuolar membrane of Arabidopsis thaliana.* Plant Physiol, 1992. **99**: p. 561 - 570.
57. Jauh, G.-Y., T. Phillips, and J. Rogers, *Tonoplast intrinsic protein isoforms as markers for vacuolar functions.* Plant Cell, 1999. **11**: p. 1867 - 1882.
58. Jurgens, G., *Membrane trafficking in plants.* Annu Rev Cell Dev Biol, 2004. **20**: p. 481-504.
59. Hara-Nishimura, I.I., T. Shimada, K. Hatano, Y. Takeuchi, and M. Nishimura, *Transport of storage proteins to protein storage vacuoles is mediated by large precursor-accumulating vesicles.* Plant Cell, 1998. **10**(5): p. 825-36.
60. Jolliffe, N.A., C.P. Craddock, and L. Frigerio, *Pathways for protein transport to seed storage vacuoles.* Biochem Soc Trans, 2005. **33**(Pt 5): p. 1016-8.
61. Vitale, A. and N.V. Raikhel, *What do proteins need to reach different vacuoles?* Trends Plant Sci, 1999. **4**(4): p. 149-155.
62. Matsuoka, K. and J.-M. Neuhaus, *Cis-elements of protein transport to the plant vacuoles.* J Exp Bot, 1999. **50**(331): p. 165-174.
63. Petruccelli, S., M.I. Molina, F.J. Lareu, and A. Circosta, *Two short sequences from amaranth 11S globulin are sufficient to target green fluorescent protein and beta-glucuronidase to vacuoles in Arabidopsis cells.* Plant Physiol Biochem, 2007. **45**(6-7): p. 400-9.
64. Maruyama, N., L.C. Mun, M. Tatsuhara, M. Sawada, M. Ishimoto, and S. Utsumi, *Multiple vacuolar sorting determinants exist in soybean 11S globulin.* Plant Cell, 2006. **18**(5): p. 1253-73.
65. Brown, J.C., N.A. Jolliffe, L. Frigerio, and L.M. Roberts, *Sequence-specific, Golgi-dependent vacuolar targeting of castor bean 2S albumin.* Plant J, 2003. **36**(5): p. 711-719.
66. Jolliffe, N.A., J.C. Brown, U. Neumann, M. Vicré, A. Bachi, C. Hawes, A. Ceriotti, L.M. Roberts, and L. Frigerio, *Transport of ricin and 2S albumin precursors to the storage*

vacuoles of Ricinus communis endosperm involves the Golgi and VSR-like receptors. Plant J, 2004. **39**(6): p. 821-833.
67. Frigerio, L., N.A. Jolliffe, A. Di Cola, D.H. Felipe, N. Paris, J.M. Neuhaus, J.M. Lord, A. Ceriotti, and L.M. Roberts, *The internal propeptide of the ricin precursor carries a sequence-specific determinant for vacuolar sorting.* Plant Physiol, 2001. **126**(1): p. 167-75.
68. Shimada, T., K. Yamada, M. Kataoka, S. Nakaune, Y. Koumoto, M. Kuroyanagi, S. Tabata, T. Kato, K. Shinozaki, M. Seki, M. Kobayashi, M. Kondo, M. Nishimura, and I. Hara-Nishimura, *Vacuolar processing enzymes are essential for proper processing of seed storage proteins in Arabidopsis thaliana.* J Biol Chem, 2003. **278**(34): p. 32292-9.
69. Saalbach, G., M. Rosso, and U. Schumann, *The vacuolar targeting signal of the 2S albumin from Brazil nut resides at the C terminus and involves the C-terminal propeptide as an essential element.* Plant Physiol, 1996. **112**(3): p. 975-85.
70. Shimada, T., E. Watanabe, K. Tamura, Y. Hayashi, M. Nishimura, and I. Hara-Nishimura, *A vacuolar sorting receptor PV72 on the membrane of vesicles that accumulate precursors of seed storage proteins (PAC vesicles).* Plant Cell Physiol, 2002. **43**(10): p. 1086-95.
71. Nishizawa, K., N. Maruyama, R. Satoh, Y. Fuchikami, T. Higasa, and S. Utsumi, *A C-terminal sequence of soybean beta-conglycinin alpha' subunit acts as a vacuolar sorting determinant in seed cells.* Plant J, 2003. **34**(5): p. 647-59.
72. Nishizawa, K., N. Maruyama, R. Satoh, T. Higasa, and S. Utsumi, *A vacuolar sorting determinant of soybean β-conglycinin β subunit resides in a C-terminal sequence.* Plant Sci, 2004. **167**(4): p. 937-947.
73. Frigerio, L., O. Foresti, D.H. Felipe, J.-M. Neuhaus, and A. Vitale, *The C-terminal tetrapeptide of phaseolin is sufficient to target green fluorescent protein to the vacuole.* J Plant Physiol, 2001. **158**(4): p. 499-503.
74. Saalbach, G., R. Jung, G. Kunze, I. Saalbach, K. Adler, and K. Muntz, *Different legumin protein domains act as vacuolar targeting signals.* Plant Cell, 1991. **3**(7): p. 695-708.
75. Mori, T., Y. Saruta, T. Fukuda, K. Prak, M. Ishimoto, N. Maruyama, and S. Utsumi, *Vacuolar sorting behaviors of 11S globulins in plant cells.* Biosci Biotechnol Biochem, 2009. **73**(1): p. 53-60.
76. Herman, E. and M. Schmidt, *Endoplasmic reticulum to vacuole trafficking of endoplasmic reticulum bodies provides an alternate pathway for protein transfer to the vacuole.* Plant Physiol, 2004. **136**(3): p. 3440-6.
77. Takahashi, H., Y. Saito, T. Kitagawa, S. Morita, T. Masumura, and K. Tanaka, *A novel vesicle derived directly from endoplasmic reticulum is involved in the transport of vacuolar storage proteins in rice endosperm.* Plant Cell Physiol, 2005. **46**(1): p. 245-9.
78. Hohl, I., D.G. Robinson, M.J. Chrispeels, and G. Hinz, *Transport of storage proteins to the vacuole is mediated by vesicles without a clathrin coat.* J Cell Sci, 1996. **109** (Pt 10): p. 2539-50.
79. Müntz, K., *Deposition of storage proteins.* Plant Mol Biol, 1998. **38**(1): p. 77-99.
80. Wenzel, D., G. Schauermann, A.v. Lüpke, and G. Hinz, *The cargo in vacuolar storage protein transport vesicles is stratified.* Traffic, 2005. **6**(1): p. 45-55.
81. Benchabane, M., C. Goulet, D. Rivard, L. Faye, V. Gomord, and D. Michaud, *Preventing unintended proteolysis in plant protein biofactories.* Plant Biotech J, 2008. **6**(7): p. 633-648.

82. Carter, C., S. Pan, J. Zouhar, E.L. Avila, T. Girke, and N.V. Raikhel, *The vegetative vacuole proteome of Arabidopsis thaliana reveals predicted and unexpected proteins.* Plant Cell, 2004. **16**(12): p. 3285-303.
83. Agrawal, G.K., N.-S. Jwa, M.-H. Lebrun, D. Job, and R. Rakwal, *Plant secretome: Unlocking secrets of the secreted proteins.* PROTEOMICS, 2010. **10**(4): p. 799-827.
84. Ding, Y., J. Wang, J. Wang, Y.-D. Stierhof, D.G. Robinson, and L. Jiang, *Unconventional protein secretion.* Trends in Plant Sci, 2012. **17**(10): p. 606-615.
85. Park, M., D. Lee, G.J. Lee, and I. Hwang, *AtRMR1 functions as a cargo receptor for protein trafficking to the protein storage vacuole.* J Cell Biol, 2005. **170**(5): p. 757-67.
86. Park, J.H., M. Oufattole, and J.C. Rogers, *Golgi-mediated vacuolar sorting in plant cells: RMR proteins are sorting receptors for the protein aggregation/membrane internalization pathway.* Plant Sci, 2007. **172**(4): p. 728-745.
87. Zouhar, J., A. Muñoz, and E. Rojo, *Functional specialization within the vacuolar sorting receptor family: VSR1, VSR3 and VSR4 sort vacuolar storage cargo in seeds and vegetative tissues.* Plant J, 2010. **64**(4): p. 577-588.
88. Kirsch, T., N. Paris, J.M. Butler, L. Beevers, and J.C. Rogers, *Purification and initial characterization of a potential plant vacuolar targeting receptor.* Proc Natl Acad Sci U S A, 1994. **91**(8): p. 3403-7.
89. Paris, N., S.W. Rogers, L. Jiang, T. Kirsch, L. Beevers, T.E. Phillips, and J.C. Rogers, *Molecular cloning and further characterization of a probable plant vacuolar sorting receptor.* Plant Physiol, 1997. **115**(1): p. 29-39.
90. Shimada, T., M. Kuroyanagi, M. Nishimura, and I. Hara-Nishimura, *A pumpkin 72-kDa membrane protein of precursor-accumulating vesicles has characteristics of a vacuolar sorting receptor.* Plant Cell Physiol, 1997. **38**(12): p. 1414-20.
91. Sanderfoot, A.A., S.U. Ahmed, D. Marty-Mazars, I. Rapoport, T. Kirchhausen, F. Marty, and N.V. Raikhel, *A putative vacuolar cargo receptor partially colocalizes with AtPEP12p on a prevacuolar compartment in Arabidopsis roots.* Proc Natl Acad Sci U S A, 1998. **95**(17): p. 9920-5.
92. Ahmed, S.U., E. Rojo, V. Kovaleva, S. Venkataraman, J.E. Dombrowski, K. Matsuoka, and N.V. Raikhel, *The plant vacuolar sorting receptor AtELP is involved in transport of NH(2)-terminal propeptide-containing vacuolar proteins in Arabidopsis thaliana.* J Cell Biol, 2000. **149**(7): p. 1335-44.
93. Zhao, Y., A. Yan, J.A. Feijo, M. Furutani, T. Takenawa, I. Hwang, Y. Fu, and Z. Yang, *Phosphoinositides regulate clathrin-dependent endocytosis at the tip of pollen tubes in Arabidopsis and tobacco.* Plant Cell, 2010. **22**(12): p. 4031-44.
94. Kitakura, S., S. Vanneste, S. Robert, C. Lofke, T. Teichmann, H. Tanaka, and J. Friml, *Clathrin mediates endocytosis and polar distribution of PIN auxin transporters in Arabidopsis.* Plant Cell, 2011. **23**(5): p. 1920-31.
95. Hinz, G., S. Colanesi, S. Hillmer, J.C. Rogers, and D.G. Robinson, *Localization of vacuolar transport receptors and cargo proteins in the golgi apparatus of developing Arabidopsis embryos.* Traffic, 2007. **8**(10): p. 1452-1464.
96. Shimada, T., K. Fuji, K. Tamura, M. Kondo, M. Nishimura, and I. Hara-Nishimura, *Vacuolar sorting receptor for seed storage proteins in Arabidopsis thaliana.* Proc Natl Acad Sci U S A, 2003. **100**(26): p. 16095-100.

97. Watanabe, E., T. Shimada, M. Kuroyanagi, M. Nishimura, and I. Hara-Nishimura, *Calcium-mediated association of a putative vacuolar sorting receptor PV72 with a propeptide of 2S albumin.* J Biol Chem, 2002. **277**(10): p. 8708-15.
98. Niemes, S., M. Labs, D. Scheuring, F. Krueger, M. Langhans, B. Jesenofsky, D.G. Robinson, and P. Pimpl, *Sorting of plant vacuolar proteins is initiated in the ER.* Plant J, 2010. **62**(4): p. 601-614.
99. Shimada, T., Y. Koumoto, L. Li, M. Yamazaki, M. Kondo, M. Nishimura, and I. Hara-Nishimura, *AtVPS29, a putative component of a retromer complex, is required for the efficient sorting of seed storage proteins.* Plant Cell Physiol, 2006. **47**(9): p. 1187-94.
100. Niemes, S., M. Langhans, C. Viotti, D. Scheuring, M. San Wan Yan, L. Jiang, S. Hillmer, D.G. Robinson, and P. Pimpl, *Retromer recycles vacuolar sorting receptors from the trans-Golgi network.* Plant J, 2010. **61**(1): p. 107-121.
101. Manfield, I.W., C.H. Jen, J.W. Pinney, I. Michalopoulos, J.R. Bradford, P.M. Gilmartin, and D.R. Westhead, *Arabidopsis Co-expression Tool (ACT): web server tools for microarray-based gene expression analysis.* Nucl Acids Res, 2006. **34**(Web Server issue): p. W504-9.
102. Wang, H., J.C. Rogers, and L. Jiang, *Plant RMR proteins: unique vacuolar sorting receptors that couple ligand sorting with membrane internalization.* FEBS J, 2011. **278**(1): p. 59-68.
103. Jiang, L., T.E. Phillips, S.W. Rogers, and J.C. Rogers, *Biogenesis of the protein storage vacuole crystalloid.* J Cell Biol, 2000. **150**(4): p. 755-70.
104. Li, L., T. Shimada, H. Takahashi, H. Ueda, Y. Fukao, M. Kondo, M. Nishimura, and I. Hara-Nishimura, *MAIGO2 is involved in exit of seed storage proteins from the endoplasmic reticulum in Arabidopsis thaliana.* Plant Cell, 2006. **18**(12): p. 3535-47.
105. Ebine, K., Y. Okatani, T. Uemura, T. Goh, K. Shoda, M. Niihama, M.T. Morita, C. Spitzer, M.S. Otegui, A. Nakano, and T. Ueda, *A SNARE complex unique to seed plants is required for protein storage vacuole biogenesis and seed development of Arabidopsis thaliana.* Plant Cell, 2008. **20**(11): p. 3006-21.
106. Craddock, C.P., P.R. Hunter, E. Szakacs, G. Hinz, D.G. Robinson, and L. Frigerio, *Lack of a vacuolar sorting receptor leads to non-specific missorting of soluble vacuolar proteins in Arabidopsis seeds.* Traffic, 2008. **9**(3): p. 408-416.
107. Rojo, E. and J. Denecke, *What is moving in the secretory pathway of plants?* Plant Physiol, 2008. **147**(4): p. 1493-503.
108. Hara-Nishimura, I., R. Matsushima, T. Shimada, and M. Nishimura, *Diversity and formation of endoplasmic reticulum-derived compartments in plants. Are these compartments specific to plant cells?* Plant Physiol, 2004. **136**(3): p. 3435-9.
109. Crofts, A.J., H. Washida, T.W. Okita, M. Satoh, M. Ogawa, T. Kumamaru, and H. Satoh, *The role of mRNA and protein sorting in seed storage protein synthesis, transport, and deposition.* Biochem Cell Biol, 2005. **83**(6): p. 728-37.
110. Woollard, A.A. and I. Moore, *The functions of Rab GTPases in plant membrane traffic.* Curr Opin Plant Biol, 2008. **11**(6): p. 610-9.
111. Saito, C. and T. Ueda, *Chapter 4: functions of RAB and SNARE proteins in plant life.* Int Rev Cell Mol Biol, 2009. **274**: p. 183-233.
112. Chong, Y.T., S.K. Gidda, C. Sanford, J. Parkinson, R.T. Mullen, and D.R. Goring, *Characterization of the Arabidopsis thaliana exocyst complex gene families by*

phylogenetic, expression profiling, and subcellular localization studies. New Phytol, 2010. **185**(2): p. 401-19.
113. Fujimoto, M. and T. Ueda, *Conserved and plant-unique mechanisms regulating plant post-Golgi traffic.* Front Plant Sci, 2012. **3**: p. 197.
114. Reddy, A.S., *Molecular motors and their functions in plants.* Int Rev Cytol, 2001. **204**: p. 97-178.
115. Hussey, P.J., T. Ketelaar, and M.J. Deeks, *Control of the actin cytoskeleton in plant cell growth.* Annu Rev Plant Biol, 2006. **57**: p. 109-25.
116. McMahon, H.T. and I.G. Mills, *COP and clathrin-coated vesicle budding: different pathways, common approaches.* Curr Opin Cell Biol, 2004. **16**(4): p. 379-91.
117. Zwiewka, M., E. Feraru, B. Moller, I. Hwang, M.I. Feraru, J. Kleine-Vehn, D. Weijers, and J. Friml, *The AP-3 adaptor complex is required for vacuolar function in Arabidopsis.* Cell Res, 2011. **21**(12): p. 1711-22.
118. Lee, M.C., E.A. Miller, J. Goldberg, L. Orci, and R. Schekman, *Bi-directional protein transport between the ER and Golgi.* Annu Rev Cell Dev Biol, 2004. **20**: p. 87-123.
119. Fujimoto, M., S. Arimura, T. Ueda, H. Takanashi, Y. Hayashi, A. Nakano, and N. Tsutsumi, *Arabidopsis dynamin-related proteins DRP2B and DRP1A participate together in clathrin-coated vesicle formation during endocytosis.* Proc Natl Acad Sci U S A, 2010. **107**(13): p. 6094-9.
120. Heymann, J.A. and J.E. Hinshaw, *Dynamins at a glance.* J Cell Sci, 2009. **122**(Pt 19): p. 3427-31.
121. Bonifacino, J.S. and L.M. Traub, *Signals for sorting of transmembrane proteins to endosomes and lysosomes.* Annu Rev Biochem, 2003. **72**: p. 395-447.
122. Robinson, M.S., D.A. Sahlender, and S.D. Foster, *Rapid inactivation of proteins by rapamycin-induced rerouting to mitochondria.* Dev Cell, 2010. **18**(2): p. 324-31.
123. Jackson, L.P., B.T. Kelly, A.J. McCoy, T. Gaffry, L.C. James, B.M. Collins, S. Honing, P.R. Evans, and D.J. Owen, *A large-scale conformational change couples membrane recruitment to cargo binding in the AP2 clathrin adaptor complex.* Cell, 2010. **141**(7): p. 1220-9.
124. Feraru, E., T. Paciorek, M.I. Feraru, M. Zwiewka, R. De Groodt, R. De Rycke, J. Kleine-Vehn, and J. Friml, *The AP-3 beta adaptin mediates the biogenesis and function of lytic vacuoles in Arabidopsis.* Plant Cell, 2010. **22**(8): p. 2812-24.
125. Burgos, P.V., G.A. Mardones, A.L. Rojas, L.L. daSilva, Y. Prabhu, J.H. Hurley, and J.S. Bonifacino, *Sorting of the Alzheimer's disease amyloid precursor protein mediated by the AP-4 complex.* Dev Cell, 2010. **18**(3): p. 425-36.
126. Hirst, J., L.D. Barlow, G.C. Francisco, D.A. Sahlender, M.N. Seaman, J.B. Dacks, and M.S. Robinson, *The fifth adaptor protein complex.* PLoS Biol, 2011. **9**(10): p. e1001170.
127. Agarwal, P., M.K. Reddy, S.K. Sopory, and P. Agarwal, *Plant Rabs: characterization, functional diversity, and role in stress tolerance.* Plant Mol Biol Rep, 2009. **27**(4): p. 417-430.
128. Rutherford, S. and I. Moore, *The Arabidopsis Rab GTPase family: another enigma variation.* Curr Opin Plant Biol, 2002. **5**(6): p. 518-28.
129. Nielsen, E., A.Y. Cheung, and T. Ueda, *The regulatory RAB and ARF GTPases for vesicular trafficking.* Plant Physiol, 2008. **147**(4): p. 1516-1526.
130. Speth, E.B., L. Imboden, P. Hauck, and S.Y. He, *Subcellular localization and functional analysis of the Arabidopsis GTPase RabE.* Plant Physiol, 2009. **149**(4): p. 1824-37.

131. Latijnhouwers, M., T. Gillespie, P. Boevink, V. Kriechbaumer, C. Hawes, and C.M. Carvalho, *Localization and domain characterization of Arabidopsis golgin candidates.* J Exp Bot, 2007. **58**(15-16): p. 4373-86.
132. Bottanelli, F., O. Foresti, S. Hanton, and J. Denecke, *Vacuolar transport in tobacco leaf epidermis cells involves a single route for soluble cargo and multiple routes for membrane cargo.* Plant Cell Online, 2011. **23**(8): p. 3007-3025.
133. Bottanelli, F., D.C. Gershlick, and J. Denecke, *Evidence for sequential action of Rab5 and Rab7 GTPases in prevacuolar organelle partitioning.* Traffic, 2012. **13**(2): p. 338-354.
134. Pinheiro, H., M. Samalova, N. Geldner, J. Chory, A. Martinez, and I. Moore, *Genetic evidence that the higher plant Rab-D1 and Rab-D2 GTPases exhibit distinct but overlapping interactions in the early secretory pathway.* J Cell Sci, 2009. **122**(20): p. 3749-3758.
135. Novick, P. and P. Brennwald, *Friends and family: the role of the Rab GTPases in vesicular traffic.* Cell, 1993. **75**(4): p. 597-601.
136. Allan, B.B., B.D. Moyer, and W.E. Balch, *Rab1 recruitment of p115 into a cis-SNARE complex: programming budding COPII vesicles for fusion.* Science, 2000. **289**(5478): p. 444-8.
137. Moyer, B.D., B.B. Allan, and W.E. Balch, *Rab1 interaction with a GM130 effector complex regulates COPII vesicle cis--Golgi tethering.* Traffic, 2001. **2**(4): p. 268-76.
138. Hoepfner, S., F. Severin, A. Cabezas, B. Habermann, A. Runge, D. Gillooly, H. Stenmark, and M. Zerial, *Modulation of receptor recycling and degradation by the endosomal kinesin KIF16B.* Cell, 2005. **121**(3): p. 437-50.
139. Yu, I.M. and F.M. Hughson, *Tethering factors as organizers of intracellular vesicular traffic.* Annu Rev Cell Dev Biol, 2010. **26**: p. 137-56.
140. Grosshans, B.L., D. Ortiz, and P. Novick, *Rabs and their effectors: achieving specificity in membrane traffic.* Proc Natl Acad Sci U S A, 2006. **103**(32): p. 11821-7.
141. Kim, S.J. and F. Brandizzi, *News and Views into the SNARE Complexity in Arabidopsis.* Front Plant Sci, 2012. **3**: p. 28.
142. Vedovato, M., V. Rossi, J.B. Dacks, and F. Filippini, *Comparative analysis of plant genomes allows the definition of the "Phytolongins": a novel non-SNARE longin domain protein family.* BMC Genomics, 2009. **10**: p. 510.
143. Sanderfoot, A., *Increases in the number of SNARE genes parallels the rise of multicellularity among the green plants.* Plant Physiol, 2007. **144**(1): p. 6-17.
144. Ebine, K., N. Miyakawa, M. Fujimoto, T. Uemura, A. Nakano, and T. Ueda, *Endosomal trafficking pathway regulated by ARA6, a RAB5 GTPase unique to plants.* Small GTPases, 2012. **3**(1): p. 23-27.
145. Vitale, A. and J. Denecke, *The endoplasmic reticulum-gateway of the secretory pathway.* Plant Cell, 1999. **11**(4): p. 615-28.
146. Pimpl, P., J.P. Taylor, C. Snowden, S. Hillmer, D.G. Robinson, and J. Denecke, *Golgi-mediated vacuolar sorting of the endoplasmic reticulum chaperone BiP may play an active role in quality control within the secretory pathway.* Plant Cell, 2006. **18**(1): p. 198-211.
147. Gupta, D. and N. Tuteja, *Chaperones and foldases in endoplasmic reticulum stress signaling in plants.* Plant Sig Behav, 2011. **6**(2): p. 232-6.

148. Hamman, B.D., L.M. Hendershot, and A.E. Johnson, *BiP maintains the permeability barrier of the ER membrane by sealing the lumenal end of the translocon pore before and early in translocation.* Cell, 1998. **92**(6): p. 747-58.
149. VITALE, A., A. CERIOTTI, and J. DENECKE, *The role of the endoplasmic reticulum in protein synthesis, modification and intracellular transport.* J Exp Bot, 1993. **44**(9): p. 1417-1444.
150. Gething, M.J., *Role and regulation of the ER chaperone BiP.* Semin Cell Dev Biol, 1999. **10**(5): p. 465-72.
151. Vitale, A., A. Bielli, and A. Ceriotti, *The Binding Protein Associates with Monomeric Phaseolin.* Plant Physiol, 1995. **107**(4): p. 1411-1418.
152. Holkeri, H. and A. Vitale, *Vacuolar sorting determinants within a plant storage protein trimer act cumulatively.* Traffic, 2001. **2**(10): p. 737-41.
153. Alvim, F.C., S.M. Carolino, J.C. Cascardo, C.C. Nunes, C.A. Martinez, W.C. Otoni, and E.P. Fontes, *Enhanced accumulation of BiP in transgenic plants confers tolerance to water stress.* Plant Physiol, 2001. **126**(3): p. 1042-54.
154. Reis, P.A., G.L. Rosado, L.A. Silva, L.C. Oliveira, L.B. Oliveira, M.D. Costa, F.C. Alvim, and E.P. Fontes, *The binding protein BiP attenuates stress-induced cell death in soybean via modulation of the N-rich protein-mediated signaling pathway.* Plant Physiol, 2011. **157**(4): p. 1853-65.
155. Melnick, J., J.L. Dul, and Y. Argon, *Sequential interaction of the chaperones BiP and GRP94 with immunoglobulin chains in the endoplasmic reticulum.* Nature, 1994. **370**(6488): p. 373-5.
156. Crofts, A.J. and J. Denecke, *Calreticulin and calnexin in plants.* Trends in Plant Sci, 1998. **3**(10): p. 396-399.
157. Williams, D.B., *Beyond lectins: the calnexin/calreticulin chaperone system of the endoplasmic reticulum.* J Cell Sci, 2006. **119**(Pt 4): p. 615-23.
158. Gomord, V., A.C. Fitchette, L. Menu-Bouaouiche, C. Saint-Jore-Dupas, C. Plasson, D. Michaud, and L. Faye, *Plant-specific glycosylation patterns in the context of therapeutic protein production.* Plant Biotechnol J, 2010. **8**(5): p. 564-87.
159. Wopereis, S., D.J. Lefeber, É. Morava, and R.A. Wevers, *Mechanisms in protein O-glycan biosynthesis and clinical and molecular aspects of protein O-glycan biosynthesis defects: A review.* Clin Chem, 2006. **52**(4): p. 574-600.
160. Ruiz-Canada, C., D.J. Kelleher, and R. Gilmore, *Cotranslational and posttranslational N-glycosylation of polypeptides by distinct mammalian OST isoforms.* Cell, 2009. **136**(2): p. 272-283.
161. Strasser, R., J. Schoberer, C. Jin, J. Glössl, L. Mach, and H. Steinkellner, *Molecular cloning and characterization of Arabidopsis thaliana Golgi α-mannosidase II, a key enzyme in the formation of complex N-glycans in plants.* Plant J, 2006. **45**(5): p. 789-803.
162. Duranti, M., C. Horstmann, J. Gilroy, and R.R.D. Croy, *The molecular basis for N-glycosylation in the 11S globulin (legumin) of lupin seed.* J Prot Chem, 1995. **14**(2): p. 107-110.
163. Boisson, M., V. Gomord, C. Audran, N. Berger, B. Dubreucq, F. Granier, P. Lerouge, L. Faye, M. Caboche, and L. Lepiniec, *Arabidopsis glucosidase I mutants reveal a critical role of N-glycan trimming in seed development.* EMBO J, 2001. **20**(5): p. 1010-9.

164. Sehgal, P., O. Kumar, M. Kameswararao, J. Ravindran, M. Khan, S. Sharma, R. Vijayaraghavan, and G.B.K.S. Prasad, *Differential toxicity profile of ricin isoforms correlates with their glycosylation levels.* Toxicol, 2011. **282**(1–2): p. 56-67.
165. Strasser, R., J.S. Bondili, U. Vavra, J. Schoberer, B. Svoboda, J. Glossl, R. Leonard, J. Stadlmann, F. Altmann, H. Steinkellner, and L. Mach, *A unique beta1,3-galactosyltransferase is indispensable for the biosynthesis of N-glycans containing Lewis a structures in Arabidopsis thaliana.* Plant Cell, 2007. **19**(7): p. 2278-92.
166. Leonard, R., G. Costa, E. Darrambide, S. Lhernould, P. Fleurat-Lessard, M. Carlue, V. Gomord, L. Faye, and A. Maftah, *The presence of Lewis a epitopes in Arabidopsis thaliana glycoconjugates depends on an active alpha4-fucosyltransferase gene.* Glycobiology, 2002. **12**(5): p. 299-306.
167. Leonard, R., D. Kolarich, K. Paschinger, F. Altmann, and I.B. Wilson, *A genetic and structural analysis of the N-glycosylation capabilities.* Plant Mol Biol, 2004. **55**(5): p. 631-44.
168. Strasser, R., J.S. Bondili, J. Schoberer, B. Svoboda, E. Liebminger, J. Glossl, F. Altmann, H. Steinkellner, and L. Mach, *Enzymatic properties and subcellular localization of Arabidopsis beta-N-acetylhexosaminidases.* Plant Physiol, 2007. **145**(1): p. 5-16.
169. Faye, L., A. Boulaflous, M. Benchabane, V. Gomord, and D. Michaud, *Protein modifications in the plant secretory pathway: current status and practical implications in molecular pharming.* Vaccine, 2005. **23**(15): p. 1770-8.
170. Hara-Nishimura, I. and M. Nishimura, *Proglobulin processing enzyme in vacuoles isolated from developing pumpkin cotyledons.* Plant Physiol, 1987. **85**(2): p. 440-445.
171. Hara-Nishimura, I., K. Inoue, and M. Nishimura, *A unique vacuolar processing enzyme responsible for conversion of several proprotein precursors into the mature forms.* FEBS Lett, 1991. **294**(1-2): p. 89-93.
172. Hara-Nishimura, I., Y. Takeuchi, and M. Nishimura, *Molecular characterization of a vacuolar processing enzyme related to a putative cysteine proteinase of Schistosoma mansoni.* Plant Cell, 1993. **5**(11): p. 1651-9.
173. Yamada, K., T. Shimada, M. Kondo, M. Nishimura, and I. Hara-Nishimura, *Multiple functional proteins are produced by cleaving Asn-Gln bonds of a single precursor by vacuolar processing enzyme.* J Biol Chem, 1999. **274**(4): p. 2563-70.
174. Gruis, D.F., D.A. Selinger, J.M. Curran, and R. Jung, *Redundant proteolytic mechanisms process seed storage proteins in the absence of seed-type members of the vacuolar processing enzyme family of cysteine proteases.* Plant Cell, 2002. **14**(11): p. 2863-82.
175. Gruis, D., J. Schulze, and R. Jung, *Storage protein accumulation in the absence of the vacuolar processing enzyme family of cysteine proteases.* Plant Cell, 2004. **16**(1): p. 270-290.
176. D'Hondt, K., D. Bosch, J. Van Damme, M. Goethals, J. Vandekerckhove, and E. Krebbers, *An aspartic proteinase present in seeds cleaves Arabidopsis 2 S albumin precursors in vitro.* J Biol Chem, 1993. **268**(28): p. 20884-91.
177. D'Hondt, K., J. Van Damme, C. Van Den Bossche, S. Leejeerajumnean, R. De Rycke, J. Derksen, J. Vandekerckhove, and E. Krebbers, *Studies of the role of the propeptides of the Arabidopsis thaliana 2S albumin.* Plant Physiol, 1993. **102**(2): p. 425-33.
178. Kinoshita, T., M. Nishimura, and I. Hara-Nishimura, *Homologues of a vacuolar processing enzyme that are expressed in different organs in Arabidopsis thaliana.* Plant Mol Biol, 1995. **29**(1): p. 81-89.

179. Kinoshita, T., M. Nishimura, and I. Hara-Nishimura, *The sequence and expression of the γ-VPE gene, one member of a family of three genes for vacuolar processing enzymes in Arabidopsis thaliana.* Plant and Cell Physiol, 1995. **36**(8): p. 1555-1562.
180. Murén, E., B. Ek, and L. Rask, *Processing of the 2S storage protein pronapin in Brassica napus and in transformed tobacco.* Eur J Biochem, 1995. **227**(1-2): p. 316-321.
181. Muren, E. and L. Rask, *Processing in vitro of pronapin, the 2S storage-protein precursor of Brassica napus produced in a baculovirus expression system.* Planta, 1996. **200**(4): p. 373-379.
182. Muren, E., B. Ek, I. Bjork, and L. Rask, *Structural comparison of the precursor and the mature form of napin, the 2S storage protein in Brassica napus.* Eur J Biochem, 1996. **242**(2): p. 214-9.
183. Hara-Nishimura, I., T. Kinoshita, N. Hiraiwa, and M. Nishimura, *Vacuolar processing enzymes in protein-storage vacuoles and lytic vacuoles.* J Plant Physiol, 1998. **152**(6): p. 668-674.
184. Jung, R., M.P. Scott, Y.W. Nam, T.W. Beaman, R. Bassuner, I. Saalbach, K. Muntz, and N.C. Nielsen, *The role of proteolysis in the processing and assembly of 11S seed globulins.* Plant Cell, 1998. **10**(3): p. 343-57.
185. Wan, L., A.R.S. Ross, J. Yang, D.D. Hegedus, and A.R. Kermode, *Phosphorylation of the 12 S globulin cruciferin in wild-type and abi1-1 mutant Arabidopsis thaliana (thale cress) seeds.* Biochem J, 2007. **404**(2): p. 247-256.
186. Kumamaru, T., Y. Uemura, Y. Inoue, Y. Takemoto, S. Uddin Siddiqui, M. Ogawa, I. Hara-Nishimura, and H. Satoh, *Vacuolar processing enzyme plays an essential role in the crystalline structure of glutelin in rice seed.* Plant Cell Physiol, 2010. **51**(1): p. 38-46.
187. Kyte, J. and R.F. Doolittle, *A simple method for displaying the hydropathic character of a protein.* J Mol Biol, 1982. **157**(1): p. 105-32.
188. Hajdukiewicz, P., Z. Svab, and P. Maliga, *The small, versatile pPZP family of Agrobacterium binary vectors for plant transformation.* Plant Mol Biol, 1994. **25**(6): p. 989-94.
189. Bent, A., *Arabidopsis thaliana floral dip transformation method*. 2006. Methods Mol Biol, 343:87-103.
190. Tamura, K., T. Shimada, E. Ono, Y. Tanaka, A. Nagatani, S.-i. Higashi, M. Watanabe, M. Nishimura, and I. Hara-Nishimura, *Why green fluorescent fusion proteins have not been observed in the vacuoles of higher plants.* Plant J, 2003. **35**(4): p. 545-555.
191. Jauh, G.Y., T.E. Phillips, and J.C. Rogers, *Tonoplast intrinsic protein isoforms as markers for vacuolar functions.* Plant Cell, 1999. **11**(10): p. 1867-82.
192. Craft, J., M. Samalova, C. Baroux, H. Townley, A. Martinez, I. Jepson, M. Tsiantis, and I. Moore, *New pOp/LhG4 vectors for stringent glucocorticoid-dependent transgene expression in Arabidopsis.* Plant J, 2005. **41**(6): p. 899-918.
193. Moore, I., M. Samalova, and S. Kurup, *Transactivated and chemically inducible gene expression in plants.* Plant J, 2006. **45**(4): p. 651-83.
194. Rehrauer, H., C. Aquino, W. Gruissem, S.R. Henz, P. Hilson, S. Laubinger, N. Naouar, A. Patrignani, S. Rombauts, H. Shu, Y. Van de Peer, M. Vuylsteke, D. Weigel, G. Zeller, and L. Hennig, *AGRONOMICS1: A new resource for Arabidopsis transcriptome profiling.* Plant Physiol, 2010. **152**(2): p. 487-499.
195. Shen, J., Y. Zeng, X. Zhuang, L. Sun, X. Yao, P. Pimpl, and L. Jiang, *Organelle pH in the Arabidopsis endomembrane system.* Molecular Plant, 2013. **6**(5): p. 1419-1437.

196. Rathore, K.S., V.K. Chowdhury, and T.K. Hodges, *Use of bar as a selectable marker gene and for the production of herbicide-resistant rice plants from protoplasts.* Plant Mol Biol, 1993. **21**(5): p. 871-84.
197. Skokut, T.A., C.P. Wolk, J. Thomas, J.C. Meeks, and P.W. Shaffer, *Initial organic products of assimilation of [N]ammonium and [N]nitrate by tobacco cells cultured on different sources of nitrogen.* Plant Physiol, 1978. **62**(2): p. 299-304.
198. Hoh, F., M. Uzest, M. Drucker, C. Plisson-Chastang, P. Bron, S. Blanc, and C. Dumas, *Structural insights into the molecular mechanisms of cauliflower mosaic virus transmission by its insect vector.* J Virol, 2010. **84**(9): p. 4706-4713.
199. Drakakaki, G., W. van de Ven, S. Pan, Y. Miao, J. Wang, N.F. Keinath, B. Weatherly, L. Jiang, K. Schumacher, G. Hicks, and N. Raikhel, *Isolation and proteomic analysis of the SYP61 compartment reveal its role in exocytic trafficking in Arabidopsis.* Cell Res, 2012. **22**(2): p. 413-24.
200. Rivera-Serrano, E.E., M.F. Rodriguez-Welsh, G.R. Hicks, and M. Rojas-Pierce, *A small molecule inhibitor partitions two distinct pathways for trafficking of tonoplast intrinsic proteins in Arabidopsis.* PLoS ONE, 2012. **7**(9): p. e44735.
201. Holtorf, S., K. Apel, and H. Bohlmann, *Comparison of different constitutive and inducible promoters for the overexpression of transgenes in Arabidopsis thaliana.* Plant Mol Biol, 1995. **29**(4): p. 637-646.
202. Nagata, T., K. Okada, T. Kawazu, and I. Takebe, *Cauliflower mosaic virus 35 S promoter directs S phase specific expression in plant cells.* Mol Gen Genet, 1987. **207**(2-3): p. 242-244.
203. Williamson, J.D., M.E. Hirsch-Wyncott, B.A. Larkins, and S.B. Gelvin, *Differential accumulation of a transcript driven by the CaMV 35S promoter in transgenic tobacco.* Plant Physiol, 1989. **90**(4): p. 1570-6.
204. Cookson, S.J., A. Radziejwoski, and C. Granier, *Cell and leaf size plasticity in Arabidopsis: what is the role of endoreduplication?* Plant Cell Environ, 2006. **29**(7): p. 1273-1283.
205. Melaragno, J.E., B. Mehrotra, and A.W. Coleman, *Relationship between endopolyploidy and cell size in epidermal tissue of Arabidopsis.* Plant Cell Online, 1993. **5**(11): p. 1661-1668.
206. Imai, K.K., Y. Ohashi, T. Tsuge, T. Yoshizumi, M. Matsui, A. Oka, and T. Aoyama, *The A-type cyclin CYCA2;3 is a key regulator of ploidy levels in Arabidopsis endoreduplication.* Plant Cell Online, 2006. **18**(2): p. 382-396.
207. Yoshizumi, T., Y. Tsumoto, T. Takiguchi, N. Nagata, Y.Y. Yamamoto, M. Kawashima, T. Ichikawa, M. Nakazawa, N. Yamamoto, and M. Matsui, *INCREASED LEVEL OF POLYPLOIDY1, a conserved repressor of CYCLINA2 transcription, controls endoreduplication in Arabidopsis.* Plant Cell Online, 2006. **18**(10): p. 2452-2468.
208. Ebine, K., T. Uemura, A. Nakano, and T. Ueda, *Flowering time modulation by a vacuolar SNARE via FLOWERING LOCUS C in Arabidopsis thaliana.* PLoS One, 2012. **7**(7): p. e42239.
209. Sohn, E.J., M. Rojas-Pierce, S. Pan, C. Carter, A. Serrano-Mislata, F. Madueño, E. Rojo, M. Surpin, and N.V. Raikhel, *The shoot meristem identity gene TFL1 is involved in flower development and trafficking to the protein storage vacuole.* Proc Natl Acad Sci U S A, 2007. **104**(47): p. 18801-18806.

210. Miernyk, J.A., *The J-domain proteins of Arabidopsis thaliana: an unexpectedly large and diverse family of chaperones.* Cell Stress Chap, 2001. **6**(3): p. 209-18.
211. Yamamoto, M., D. Maruyama, T. Endo, and S.-i. Nishikawa, *Arabidopsis thaliana has a set of J proteins in the endoplasmic reticulum that are conserved from yeast to animals and plants.* Plant Cell Physiol, 2008. **49**(10): p. 1547-1562.
212. V. Rajan, V. and P. D'Silva, *Arabidopsis thaliana J-class heat shock proteins: cellular stress sensors.* Funct Integr Genomics, 2009. **9**(4): p. 433-446.
213. Shen, L., Y.G.G. Kang, L. Liu, and H. Yu, *The J-domain protein J3 mediates the integration of flowering signals in Arabidopsis.* Plant Cell, 2011. **23**(2): p. 499-514.
214. Song, Y.H., S. Ito, and T. Imaizumi, *Flowering time regulation: photoperiod- and temperature-sensing in leaves.* Trends in Plant Sci, 2013. **18**(10): p. 575-583.
215. Ogasawara, K., K. Yamada, J.T. Christeller, M. Kondo, N. Hatsugai, I. Hara-Nishimura, and M. Nishimura, *Constitutive and inducible ER bodies of Arabidopsis thaliana accumulate distinct beta-glucosidases.* Plant Cell Physiol, 2009. **50**(3): p. 480-8.