uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

uOttawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Jaimie A. Schnell
AUTEUR DE LA THÈSE / AUTHOR OF THESIS

Ph.D. (Biology)
GRADE / DEGREE

Department of Biology
FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

A transgenic plant system for the heterologous expression of bioproducts in soybean seed coats

TITRE DE LA THÈSE / TITLE OF THESIS

Douglas Johnson
DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

EXAMINATEURS (EXAMINATRICES) DE LA THÈSE / THESIS EXAMINERS

John Arnason

Brian Miki

Christiane Charest

Tim Xing

Allen Good

Gary W. Slater
Le Doyen de la Faculté des études supérieures et postdoctorales / Dean of the Faculty of Graduate and Postdoctoral Studies

A transgenic plant system for the heterologous expression of bioproducts in soybean seed coats

Jaimie A. Schnell

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
In partial fulfillment of the requirements
For the PhD degree in Biology

Ottawa-Carleton Institute of Biology
Faculty of Science
University of Ottawa

Your file Votre référence
ISBN: 978-0-494-51822-9
Our file Notre référence
ISBN: 978-0-494-51822-9

Abstract

A transgenic system for the heterologous expression of bioproducts in soybean (*Glycine max* (L.) Merrill) seed coats was developed. Expression was driven by the region upstream of the high expressing, seed coat-specific *Ep* gene, which encodes soybean peroxidase (SBP). Genes were modified to include sequences encoding the N- and C-terminal propeptides from SBP, which were predicted to target proteins to the vacuole.

To investigate the influence of targeting on foreign protein accumulation, a model system was developed. *Arabidopsis thaliana* (L.) Heynh. plants were transformed with genes encoding SBP-green fluorescent protein (GFP) fusions, targeted to the cytoplasm, apoplast, endoplasmic reticulum (ER), or vacuole with combinations of the SBP N- and C-terminal propeptides and an ER retention signal. The location of the targeted fusion proteins and the function of the SBP N- and C-terminal propeptides were confirmed. Average specific activity of all three of the targeted fusion proteins was significantly greater than the average specific activity of the non-targeted, cytoplasmic fusion protein. Immunoblotting revealed that the GFP portion of the fusion protein was being partially degraded, and that degradation was more severe during targeting to the vacuole and apoplast than the ER. These results suggest that targeting proteins to any of the three compartments investigated can significantly improve protein accumulation but that maximum stability may be found in the ER.

To compare the ability of soybean seed coats to express foreign proteins at levels comparable to the *Ep* gene, the *Ep* cDNA was heterologously expressed in a soybean cultivar that contains a deletion in the *Ep* gene, designated *epep*. Transgenic *Ep* plants produced SBP but activity levels were lower than those measured in the *EpEp* cultivar Harosoy 63. This suggests

that elements necessary for the high expression levels of the *Ep* locus were omitted in the construct.

To investigate the applicability of metabolically engineering soybean seed coats to divert metabolism towards the production of novel biochemicals, we introduced the polyhydroxybutyric acid (PHB) biosynthetic enzymes into soybean. PHB was produced at levels up to an average of 0.13% of seed coat DW. Although the levels of PHB are low, these results demonstrate that it is possible to metabolically engineer soybean seed coats.

Resumé

Un sytème transgénique a été développé pour l'expression hétérologue des bioproduits dans la couche externe des graines de soja (*Glycine max* (L.) Merrill). L'expression est menée par la région en amont de la région spécifique à haute expression du gène *Ep* de la couche externe, qui encode l'enzyme SBP (soybean peroxidase). Les gènes ont été modifiés pour inclure les séquences N-terminale et C-terminale de la propeptide SBP, qui devrait diriger les protéines vers la vacuole.

Pour examiner l'influence de direction sur l'accumulation de protéine étrangère, les plantes d'*Arabidopsis thaliana* (L.) Heynh. ont été transformées pour contenir les gènes nécessaires à la fusion SBP-protéine fluorescente verte (GFP), dirigé vers le cytoplasme, l'apoplaste, le réticulum endoplasmique ou la vacuole contenant les propeptides SBP N- et C-terminales et un signal de rétention pour le réticulum endoplasmique. L'emplacement des protéines SBP-GFP et la fonction des propeptides SBP N- et C- terminales ont été confirmés. L'activité spécifique moyenne des trois protéines de fusion ciblées était considérablement plus élevée que l'activité spécifique moyenne des protéines de fusion cytoplasmique non-ciblées. L'immunotransfert a démontré que la section GFP de la protéine de fusion a été partiellement dégradée et que la dégradation était plus prononcée dans la vacuole et l'apoplaste que dans le réticulum endoplasmique. Ces résultats suggèrent qu'en dirigeant les protéines vers l'une des trois régions examinées on pourrait considérablement améliorer l'accumulation de protéines mais que la région la plus stable serait le réticulum endoplasmique.

Pour comparer la capacité des couches externes de graine de soja à exprimer des protéines étrangères à un niveau comparable à celui du gène *Ep*, l'ADNc du gène *Ep* a été exprimé de façon hétérologue dans un cultivar de soja contenant l'omission du gène, nommé

epep. Les plantes transgéniques *Ep* ont produit le SBP mais à un niveau d'activité moins élevé que celui mesuré dans le cultivar *EpEp* Harosoy 63. Ceci suggère que les éléments nécessaires pour un niveau d'expression considérable de la région du gène *Ep* ont été omis durant le construit.

Pour explorer l'applicabilité de la manipulation métabolique des couches externes des graines de soja à dévier le métabolisme vers la production de produits biochimiques nouveaux, nous avons introduit des enzymes biosynthétiques d'acide polyhydroxybutyrique (PHB) dans les graines de soja. Le PHB a été ajouté à un niveau moyen de 0.13% poids sec aux couches externes des graines de soja. Bien que ces niveaux soient faible, ils démontrent qu'il est possible de manipuler le métabolisme des couches externes des graines de soja.

Table of Contents

List of Tables

List of Figures

Acknowledgements

I would like to sincerely thank my supervisor, Dr. Douglas Johnson, for providing me with the perfect mixture of guidance and independence. Over the last four years, his advice in research and beyond has been invaluable. I would also like to thank the other members of my committee, Dr. Brian Miki, Dr. John Arnason, Dr. Linda Bonen, and Dr. Federica Brandizzi, for the assistance they provided at all levels of my research project. I would additionally like to thank Dr. Federica Brandizzi for inviting me to do some research in her lab and for her training and advice in GFP imaging.

My graduate experience was greatly enriched by all members of the Johnson lab from the last four years and I would like to thank them all for creating a wonderful research environment and for their advice and support. I would also like to extend my appreciation to Martha Mullaly and Dr. David Currie for their help with statistical analyses. I would equally like to express my gratitude to Dr. Virginie Treyvaud-Amiguet for her help developing the methodology for the extraction and analysis of PHB. I am also grateful to my cousin Ivan Moïse for providing me with a French translation of the abstract to this thesis.

A big thank you also goes to my family, for always being willing to listen to me talk about science even when they do not understand me. In particular, I would like to thank my parents, Paul and Cathy, for being the foundation of everything that I am and everything that I have achieved. Finally, I must thank my husband, Joshua, for always inspiring and encouraging me and for being my center when I spin away.

I would also like to acknowledge financial support from NSERC in the form of a graduate scholarship as well as a Strategic Grant awarded to Dr. D.A. Johnson.

List of Abbreviations

3-HB	3-hydroxybutyric acid
ABTS	2,2'-azino-bis-(3-ethylbenzthiazoline-6-sulphonate)
Acetyl-CoA	Acetyl-Coenzyme A
AMP	Antimicrobial peptide
CTPP	C-terminal propeptide
ctVSS	C-terminal Vacuolar sorting signal
cv.	Cultivar
DPA	Days post anthesis
DW	Dry weight
Ep gene	Gene encoding soybean peroxidase
Ep 1.5 promoter	Region approximately 1.5 kb upstream of the *Ep* gene
ER	Endoplasmic reticulum
FW	Fresh weight
GFP	Green fluorescent protein
GM	Genetically modified
HRP	Horseradish peroxidase
nos	*Nopaline synthase*
NTPP	N-terminal propeptide
PAC vesicles	Precursor-accumulating vesicles
PHA	Polyhydroxyalkanoates
PHB	Polyhydroxybutyric acid
PSV	Protein storage vacuole
RMR protein	Receptor homology region transmembrane domain-Ring H2 motif proteins
SBP	Soybean peroxidase
ssVSS	Sequence-specific vacuolar sorting signal
TIP	Tonoplast intrinsin protein
UTR	Untranslated region
VSS	Vacuolar sorting signal

CHAPTER 1 INTRODUCTION

1.1 Overview

The goal of this research project was to develop a transgenic system for the heterologous expression of bioproducts in soybean (*Glycine max* (L.) Merrill) seed coats. Soybean seed coats have several advantages that make them an ideal tissue for such a system. First of all, a gene has been identified that is expressed at high levels primarily in soybean seed coats (Gijzen 1997). The presence of the *Ep* gene, encoding soybean peroxidase (SBP), is correlated with high levels of peroxidase activity in soybean seed coats and SBP has been calculated to represent at least 5% of seed coat total soluble protein (Gillikan and Graham 1991). The promoter that drives expression of this gene can be used to drive high levels of tissue-specific expression of the genes encoding the desired bioproducts. Since this promoter has not been fully characterized, we selected the region 1.5 kb upstream of the *Ep* gene to drive transgene expression in the soybean seed coat system, and herein this will be referred to as the *Ep* 1.5 promoter.

The second advantage of the soybean seed coat system is that several naturally occurring soybean cultivars, designated *epep*, have greatly reduced peroxidase levels in their seed coats (Gijzen et al. 1993). High levels of expression of foreign protein can often be toxic to the host organism. By using a low peroxidase cultivar as the host, we can be confidant that large quantities of bioproducts will have minimal toxicity since seed coats naturally harbour high levels of protein.

A third advantage is that SBP has an N-terminal and a C-terminal propeptide that are predicted to direct vacuolar targeting, which could be used to target the bioproducts to the vacuole in seed coats. This will again mimic the natural system in soybean seed coats, where SBP is likely sorted to vacuoles, which may also help to minimize any negative effects of high

protein levels. In addition, subcellular targeting has been shown to increase the yield of foreign proteins. Targeting the proteins to the vacuole in soybean seed coats may therefore be necessary to achieve expression levels similar to SBP. However, targeting to other subcellular compartments has also been shown to increase the yield of foreign proteins. Therefore, one aspect of this project was to compare the accumulation of foreign proteins targeted to different compartments of the plant secretory system in *Arabidopsis thaliana* (L.) Heynh, a plant model system.

Finally, during the processing of soybeans to produce oil and protein, the seed coat is removed. Therefore, the technology for harvesting seed coats already exists and no additional cost would be added to soybean processing.

For this research project, a model enzyme and biochemical pathway were selected, both of economic importance, DNA constructs were designed for their efficient expression and processing in soybean seed coats, and effective means of analyzing their level of expression were developed.

In the following sections of the Introduction, information pertinent to this research project will be discussed. First, an outline of the current soybean market as well as the directions that Canadian soybean growers are investigating to improve the market value of this crop will be presented, as well as how this research relates to these directions. Following this, the structure of soybean seed coats will be described and the *Ep* gene and its promoter that provide seed coat specific expression will be discussed. The current understanding of how proteins are targeted to different subcellular compartments as well as the nature of the subcellular compartments and how this may relate to protein yield and stability will also be described.

Finally, in order to test the seed coat as a system for the accumulation of bioproducts, soybean plants were engineered to express SBP and three enzymes that form the biosynthetic pathway for the biodegradable plastic polyhydroxybutyric acid (PHB). The final two sections of the Introduction will provide relevant information about these two bioproducts, including a brief discussion of their economic importance.

1.2 Soybean market

In the 2007/08 marketing year, approximately 220 million metric tons of soybean seeds were produced worldwide and sold at an average price of $358/ton as whole seeds (US farm price; USDA), $338/ton as soybean meal (Decatur, average wholesale 48% protein; USDA), and $1025/ton as vegetable oil (Decatur, average wholesale tank crude; USDA) (http://www.fas.usda.gov/default.asp)[1]. In Canada alone, 2.7 million metric tons of soybeans were produced, of which 1.37 million metric tons were exported, mainly to the USA (http://www.fas.usda.gov/default.asp). Canada is the world's seventh largest producer of soybeans and is the fifth largest exporter of soybeans.

Soybean seed coats or hulls are removed during seed processing and are sold at an average price of $160/ton (http://ww.mnsoy.com). They are typically added back to soybean meal after oil extraction in order to increase the fibre content for use as animal feed. Since seed coats typically represent up to 10% of the seed by weight, there is available worldwide 22 million metric tons of a low value product. The large quantity but low value of soybean seed coats presents an opportunity for increasing the value of soybean crops through engineering expression of valuable bioproducts

[1] Prices are in US dollars and masses are in metric tons (1 ton = 1000 kg = 2204.6 lb)

Currently, several alternative markets for soybeans are being considered with the aim of increasing the value of soybeans. Soy20/20, a Canadian company that seeks to promote and improve soybean market opportunities, suggests several alternative markets for soybean, including the production of SBP, biodiesel, soy flour, and polyurethane foam (http://www.soy2020.ca/about.php). Each of these areas takes advantage of the many useful properties of soybean, such as its health benefits, the presence of commercially-valuable enzymes, as well as the oil composition.

Another approach to developing alternative markets for soybean is applying biotechnology to add value to soybeans through the introduction of novel traits. In Canada, 7.0 million hectares are planted with genetically modified (GM) canola, corn, and soybean crops, and worldwide there are 58.6 million hectares of GM soybean, primarily engineered for herbicide resistance (James 2007). This demonstrates the successful adoption of GM crops by farmers worldwide and suggests that GM technology could be extended to value-added traits.

Transgenic technologies provide an opportunity for increasing the market of crops that are currently limited by the area available for cultivation. Through the development of systems using high expressing, tissue-specific promoters, such as the *Ep* promoter, we now have the ability to turn agricultural waste products, such as the soybean seed coat, into high value production systems, creating a range of products that can benefit many sectors of society, including agriculture, industry, and medicine. In the following section, I will discuss in more detail the nature of soybean seed coats, the host tissue we have selected to pursue this goal, and the molecular biology necessary to exploit it for the production of foreign proteins.

1.3 Soybean seed coats

1.3.1 Soybean seed coat structure

Seed coats originate from integuments that surround the ovary prior to fertilization and are thus of maternal origin. They form a protective outer covering around both the endosperm and embryo that collectively constitute the seed. Seed coats generally consist of only a few cell layers, many of which are crushed at maturity due to the expansion of the cotyledons (Moïse et al. 2005). However, some cell layers undergo distinct differentiation processes that may result in the thickening of one or more cell walls, cell elongation, or the accumulation of substances such as pigments or mucilage (Moïse et al. 2005). The soybean seed coat has two distinct cell layers, the outermost palisade cell layer (macrosclereids) and then the hourglass cell layer (osteosclereids), followed by a multicellular layer of crushed parenchyma (Miller et al. 1999). The innermost layer of the soybean seed coat is composed of aleurone cells, which originate from the outermost layer of the endosperm, and are, therefore, not of maternal origin (Miller et al. 1999). This structural arrangement appears conserved throughout the legumes (Moïse et al. 2005), although only a few species have been investigated in detail.

As will be discussed below, SBP is primarily located within the hourglass cell layer. This environment appears to be particularly stable, since storage of soybean seeds for several years has minimal effect on seed coat peroxidase activity. It has been hypothesized that the hourglass cell layer may be a repository for proteins that are perhaps required only at the time of imbibition (Gijzen et al. 1999). The stability of proteins within the hourglass cell layer is another advantage of using the seed coat as a production system for foreign proteins.

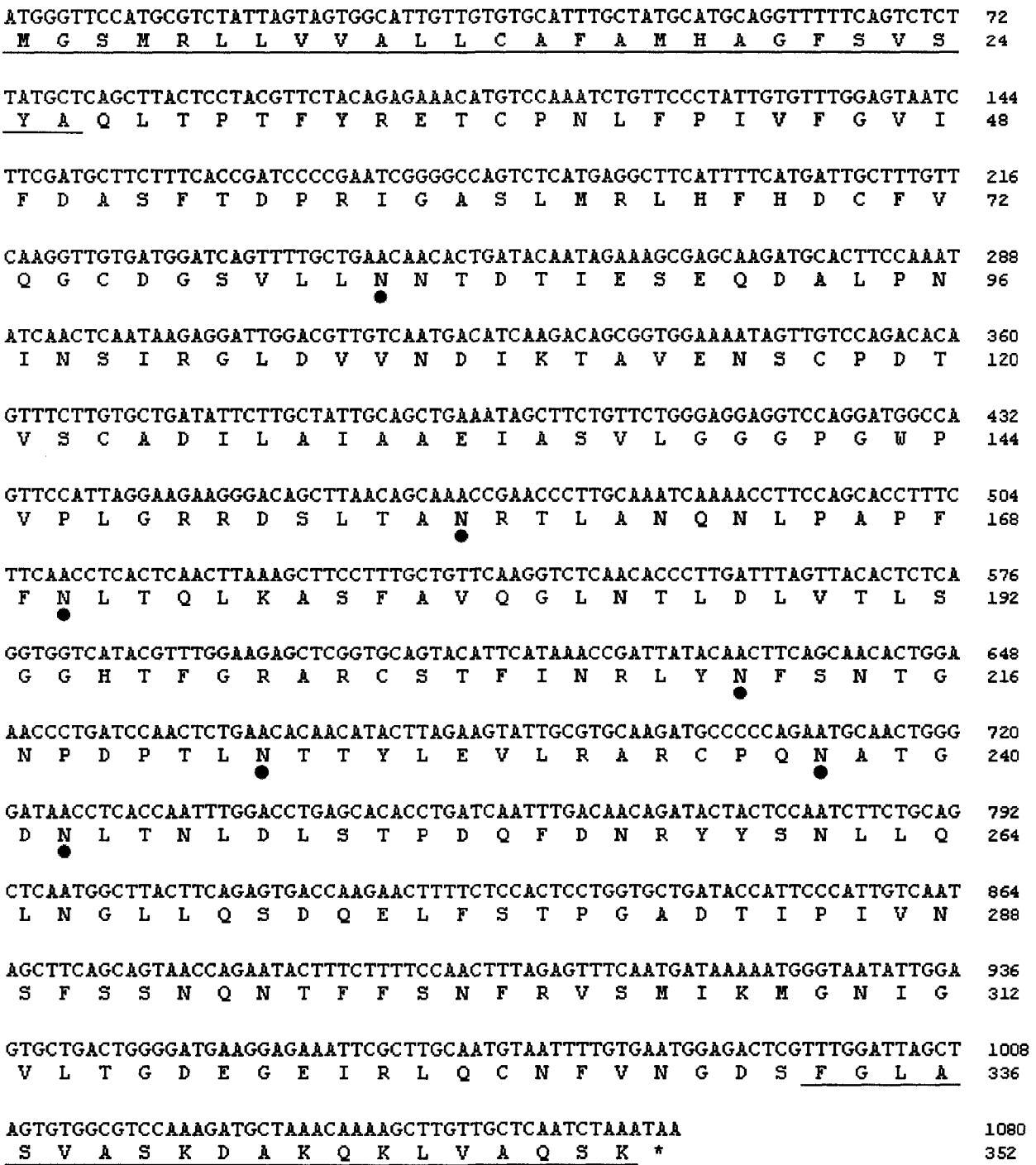
1.3.2 Molecular biology of soybean peroxidase

In soybean, the presence of the dominant *Ep* gene, encoding SBP, is associated with high seed coat peroxidase activity while *epep* genotypes have low peroxidase activity in the seed coats. Activity in seed coats from *EpEp* cultivars is typically 100-fold higher compared to *epep* cultivars (Gijzen et al. 1993). In mature, dry seed coats, SBP represents at least 5% of total soluble protein (Gillikan and Graham 1991). Total peroxidase activity in roots and leaves are similar between *EpEp* and *epep* cultivars, although a peroxidase with a similar pI to SBP is absent in the roots of the *epep* cultivars (Gijzen et al. 1993). Based on chloronaphthol staining of mature *EpEp* seed coats, peroxidase activity is localized in the hourglass cells of the subepidermis and the placental cells of the hilum (Gijzen et al. 1993).

The primarily seed coat-specific, high level expression of the *Ep* gene makes its promoter an ideal candidate for creating an expression system in soybean seed coats. It may also be possible in future studies to combine domains responsible for seed coat specific expression and the high levels of expression with other expression enhancer elements while removing root specific expression domains, in order to further enhance high level, tissue-specific expression in the soybean seed coat system.

The *Ep* gene was isolated and characterized (Gijzen 1997). The cDNA and the predicted amino acid coding sequence are shown in Figure 1.1. The *ep* allele has a deletion of 87 bp beginning 9 bp upstream from the translation start site and including the first 78 bp of the open reading frame (Gijzen 1997). Both the TATA box and the putative transcription initiation sites are unaffected (M. Gijzen, unpublished). Northern analysis of *epep* soybean seed coats indicates that peroxidase transcript abundance is greatly reduced compared to *EpEp* seed coats (Gijzen 1997). There is one downstream, in-frame start codon in the *ep* transcript that would produce a

Figure 1.1 The cDNA (L78163) and predicted amino acid (AAL40127) coding sequence of soybean (*Glycine max*) peroxidase encoded by the *Ep* gene. The N-terminal and C-terminal propeptides are underlined. Potential N-glycosylation sites are indicated by the symbol •.



protein with all the necessary active site residues, which could account for the low levels of peroxidase activity that are observable in *epep* soybean cultivars. However, it is perhaps more likely that these low peroxidase levels can be attributed to other uncharacterized peroxidase isoenzymes present in soybean seed coats (Gillikan and Graham 1991; Gijzen et al. 1999).

The *Ep* transcript can be detected in seed coats and roots but not in leaf, flower, embryo, or pod tissue extracts by northern blotting (Gijzen 1997). Based on *in situ* hybridization, *Ep* expression can be detected as early as 6 days post anthesis (DPA) in the thin-walled parenchyma of the outer integument but as the hourglass cells begin development at 12 DPA, expression is restricted to this layer (Gijzen et al. 1999). Clearly the *Ep* promoter initiates expression early during seed coat development, an aspect that is likely related to the production of such high levels of SBP and that can be exploited by using the *Ep* promoter in the creation of the soybean seed coat system. By using *epep* cultivars, toxic effects of foreign protein overexpression can be minimized since the seed coat can naturally host a single protein as 5% of total soluble protein. This will be an additional advantage in the creation of the soybean seed coat system.

The *Ep* gene encodes a 352 amino acid protein that is processed at maturity to 306 residues following removal of a 26 amino acid N-terminal propeptide (NTPP) and a 20 amino acid C-terminal propeptide (CTPP) (Welinder et al. 2004; Gijzen 1997). The NTPP is characteristic of signals that direct import of soluble proteins into the ER (Gijzen 1997). SBP is a heterogenous glycoprotein with seven glycosylation sites, some of which are only partially glycosylated (Welinder et al. 2004; Gray and Montgomery 2006). The glycosylation of SBP indicates that it passes through the plant secretory system, where glycosylation occurs, further supporting the function of the NTPP as a targeting signal. CTPPs similar to that found in SBP have been identified in a number of peroxidases from other plant species and are thought to function as

vacuolar sorting signals. An alignment of the CTPP of soybean peroxidase and several other peroxidases also possessing CTPPs is shown in Figure 1.2. The subcellular location of SBP has not yet been determined, but the enzyme does appear to be intracellular (Gijzen et al. 1993; Gijzen 1997). There is evidence that some peroxidase CTPPs can function as vacuolar sorting signals. The CTPP in horseradish peroxidase (HRP) C1a was found to be necessary and sufficient for vacuolar localization of HRP C1a in tobacco BY-2 cells (Matsui et al. 2003). Furthermore, 6 of 9 *Arabidopsis* peroxidases possessing CTPPs were found in vegetative vacuoles during an analysis of the *Arabidopsis* vacuolar proteome (Carter et al. 2004).

The SBP N- and C-terminal propeptides can be used to target foreign proteins to the same compartment as SBP. This may be necessary to achieve high yield and stability of foreign proteins in the soybean seed coat system. One objective of this research project is to determine the function of the SBP propeptides in protein targeting and to examine how they can impact the level of expression of a foreign protein. The following section will discuss the nature of protein targeting signals and the mechanism by which proteins are targeted to cellular compartments.

1.4 Targeting of soluble proteins in plant cells

1.4.1 Advantages to targeting soluble proteins in plant cells

A number of studies have demonstrated that targeting foreign proteins to specific plant cell compartments can increase the yield of the foreign protein (Hood et al. 2003; Ramírez et al. 2002; Yang et al. 2005). In transgenic maize, targeting to the cell wall in embryos resulted in the highest yields compared to cytoplasmic and ER-targeted laccase (Hood et al. 2003). Yang et al. (2005) found that production of a synthetic analogue of a spider dragline silk protein in *Arabidopsis* was maximized by seed-specific expression and ER-targeting, as compared to

Figure 1.2 ClustalW alignment of C-terminal propeptides from *Glycine max* soybean peroxidase (SBP) (T05723), *Phaseolus vulgaris* L. FBP1 (AAC37427), *Hordeum vulgare* BP1 (CAA37464), and *Lupinus albus* LEP1 (AAO13837). Black boxes represent identical residues in at least two sequences and gray boxes represent similar residues (based on a BLOSUM62 matrix) in at least two sequences.

constitutive expression and targeting to the vacuole, apoplast, or cytoplasm. The yield of a single chain Fv antibody fragment in transgenic tobacco plants was significantly higher when the fragment was targeted to the ER as opposed to the apoplast and vacuole (Ramírez et al, 2002). Targeting bioproducts to a subcellular compartment in soybean seed coats is one approach that was taken to increase yield in the soybean seed coat system.

Some of the potential advantages to targeting foreign proteins in plant systems include: 1) higher levels of protein accumulation (Yoshida et al. 2004); 2) passage through the endoplasmic reticulum (ER) and Golgi apparatus where many post-translational modifications that may be crucial to protein stability and activity occur (Faye et al. 2005); and 3) sequestration of foreign proteins, which may minimize toxic effects inherent to the protein or that result from the high protein levels necessary for a useful production system (Murray et al. 2002).

1.4.2 The plant secretory system

The plant secretory system is composed of the ER, Golgi apparatus, and vacuoles, including lytic vacuoles (LVs) and protein storage vacuoles (PSVs), as well as a number of smaller intermediate compartments (Figure 1.3). It is a complex system that ensures the correct folding, processing and delivery of a large number of proteins.

The different compartments of the secretory system present a range of environments that can impact the stability and activity of foreign proteins. The cytoplasm has a neutral pH (Kurkdjian and Guern 1989; Gao et al. 2004). In contrast, the apoplast in plant tissues is typically between pH 5 and 6.5, although the pH can range as low as 4 and as high as 7, and this compartment may contain high levels of osmotically active solutes and inorganic electrolytes, particularly potassium salts (Sakurai 1998; Grignon and Sentenac 1991). The ER lumen in mammalian cells is pH 7.1, which is similar to the cytoplasm, but the redox state of the ER

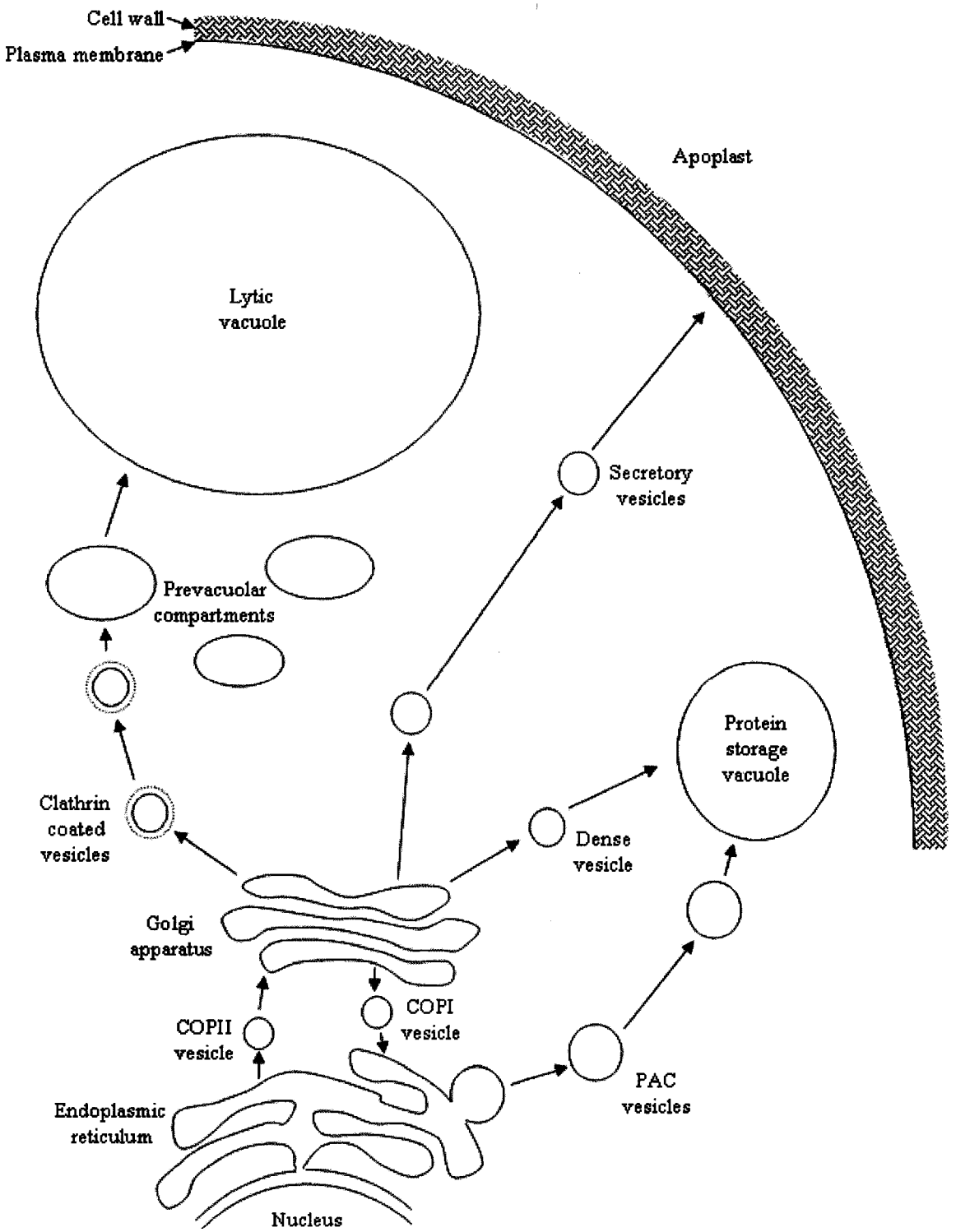


Figure 1.3 A model of the plant secretory system, demonstrating the possible trafficking pathways of soluble proteins (Drawn by J. Schnell). Abbreviations: PAC vesicles, precursor accumulating vesicles.

lumen varies considerably from that of the cytoplasm, having a high ratio of oxidized to reduced glutathione, which promotes disulfide bond formation (Vitale and Denecke 1999). Studies in plants indicate that the ER lumen has a similar nature (Shimoni and Galili 1996; Dickinson et al. 1987). LVs present an acidic environment with a pH in the range of 5 to 6.5 (Kurkdjian and Guern 1989). These vacuoles contain a dilute mixture of metabolites and salts that are actively transported into this compartment against their electrochemical gradient, as well as a variety of acid hydrolases (Taiz 1992; Boller and Kende 1979). In contrast, PSVs have a pH that is closer to neutral during seed maturation and then become acidic during seed germination (Bethke et al. 1998; He et al. 2007; Okamoto et al. 1999).

1.4.3 Transport in the plant secretory system

Within plants, targeting information is generally contained within a protein's amino acid sequence. A similar system exists in yeast, but in mammalian cells the modified oligosaccharide mannose-6-phosphate may also function as a vacuolar targeting signal (Kornfeld and Mellman 1989).

The ER is the entry point to the plant secretory system. Proteins enter the ER cotranslationally through a translocon pore (Vitale and Denecke 1999). In the ER lumen are a variety of molecular chaperones, such as the chaperonins and heat shock protein family members, such as BiP, that aid in proper protein folding. In addition, there are enzymes such as protein disulfide isomerase, which aid in the formation of disulfide bridges, and glucosidases and glucosyl transferases, which are involved in the glycosylation of proteins, as discussed in Section 1.4.5 (Vitale and Denecke 1999). The ER also has the additional function of quality control, whereby incorrectly folded proteins are retained within the ER and targeted for destruction. This

function involves the lectins calnexin and calreticulin as well as another glucosyl transferase (Vitale and Denecke 1999).

Proteins that are targeted to the plant secretory system have an N-terminal signal peptide that directs co-translational import into the ER (Vitale and Hinz 2005). This signal peptide has a conserved sequence composed of a central hydrophobic region rich in Ala and Leu flanked by a short, positively charged N-terminus and a polar C-terminus (Emanuelsson and von Heijne 2001). This N-terminal signal peptide is removed co-translationally by signal peptidase, located in the ER membrane (Vitale and Denecke 1999). In the absence of additional targeting information, proteins with an N-terminal signal peptide are secreted (Denecke et al. 1990). The NTPP of soybean peroxidase is expected to function as a signal peptide directing import into the ER. This signal can therefore be used to target foreign proteins to the plant secretory system.

1.4.4 Traffic between endoplasmic reticulum and Golgi apparatus

Most soluble proteins in the ER are transported to the Golgi apparatus. Proteins exit the ER in COPII vesicles by bulk flow (Crofts et al. 1999; Phillipson et al. 2001). Since this transport is non-specific, ER resident proteins, known as reticuloplasmins, are exported as well and there is, therefore, a second system in place for retrieval of reticuloplasmins. Retrograde traffic of the reticuloplasmins from Golgi to ER occurs via COPI vesicles (Pimpl et al. 2003; Ritzenthaler et al. 2002). Reticuloplasmins typically have ER retention signals, containing the conserved K/HDEL sequence, that are recognized by sorting receptors such as ERD2 that are localized at the Golgi apparatus and which sort the reticuloplasmins into the COPI vesicles (Jürgens 2004).

In order to retain a foreign protein within the ER, it is only necessary to add four additional amino acids to form the K/HDEL signal. However, the K/HDEL signal is not

removed, unlike the other protein signals. This could impact protein function, for example, by interfering with the formation of proper tertiary structure and is one disadvantage of targeting foreign proteins to the ER.

1.4.5 Glycosylation in the plant secretory system

Secretory proteins may be glycosylated at specific Asn sites that correspond to the consensus sequence Asn-X-Ser/Thr, where X is any amino acid but proline (Faye et al. 2005). Such modifications are essential to proper protein function as they affect protein folding, solubility, immunogenicity, activity, and interactions with other proteins, as well as protecting proteins from degradation and thermal denaturation (Faye et al. 2005). This, therefore, is another important aspect to consider when determining whether or not to target foreign proteins to the secretory system in plant cells.

The first stage in the glycosylation of soluble proteins occurs in the ER and involves the transfer of the oligosaccharide precursor $Glc_3Man_9GlcNAc_2$ from a dolichol lipid carrier to Asn residues (Rayon et al. 1998). This precursor is modified in the ER through the action of glucosidases I and II, which remove three terminal glucose units. There may also be other modifications that occur in the ER, although less is known about these processes. In the Golgi apparatus, a number of enzymes found throughout the *cis*, medial, and *trans*-Golgi network provide further modifications, leading to the formation of complex-type *N*-glycans. Finally, other modifications may occur to the glycan side-chains in the other compartments of the plant secretory system, such as the vacuole (Rayon et al. 1998). In addition to *N*-glycosylation, plant proteins may also undergo *O*-glycosylation. This occurs in the Golgi apparatus, but as of yet, little is known about this process in plants (Faye et al. 2005).

1.4.6 The diversity of plant vacuoles

Plants are unique in containing more than one type of vacuole, and these vacuoles may at times coexist within a single cell (Di Sansebastiano et al. 1998; Paris et al. 1996; Park et al. 2004). Lytic vacuoles (LVs), which are acidic and contain hydrolytic enzymes, are functionally equivalent to animal lysosomes and yeast vacuoles. They function to maintain turgor pressure and protoplasmic homeostasis, as well as to store metabolic products, sequester xenobiotics, and digest cytoplasmic constituents (Marty 1999). LVs can be identified by the presence of various hydrolytic enzymes, such as aleurain and sporamin, as well as γ-TIP[2] (tonoplast intrinsic protein) (Jauh et al. 1999).

Another type of vacuole that is unique to plants is the protein storage vacuole (PSV). PSVs are found mostly in storage tissues such as seeds and tubers and their main function is the long-term storage of proteins, such as seed storage proteins (Vitale and Hinz 2005). This function may make PSVs the ideal cellular compartment for the storage of foreign proteins, although no studies to date have compared the yield of proteins targeted to this compartment with other compartments (Jiang and Sun 2002). These vacuoles are marked by the presence of α-TIP and δ-TIP or α-TIP, δ-TIP, and γ-TIP (Jauh et al. 1999).

PSVs are complex organelles that may contain internal compartments with specialized functions. Compartments, known as globoid cavities, have been identified within PSVs. These compartments contain phytic acid, as well as proteins that are characteristic of LVs, possibly in preparation for mobilization of seed stores during germination (Jiang et al. 2001). This suggests that PSVs are able to compartmentalize storage and hydrolytic functions. Also found in the PSVs

[2] TIP isoforms have traditionally been used as markers for different types of vacuoles (Jauh et al. 1998; 1999). A recent study by Hunter et al. (2007) found that the TIP isoforms are differentially distributed based on tissue and developmental stage. This suggests that they may not be effective markers of vacuole type and so caution should be used in interpreting studies in which they have been used in this fashion.

of some plant species is the crystalloid compartment, which is derived from the internalization of a small organelle labelled with another TIP isoform known as dark intrinsic protein (Jiang et al. 2000).

Storage-type vacuoles, named delta-vacuoles, have also been found in vegetative tissues. These vacuoles are derived from lytic vacuoles and are identified by the presence of vegetative storage proteins and δ-TIP and sometimes also γ-TIP (Jauh et al. 1998; 1999).

The vacuoles characterized above have been identified in only a very small subset of plant species and tissue types. It is unknown what types of vacuoles may be found in soybean seed coats. The presence of a putative vacuolar sorting signal in SBP and the fact that it is stable in seeds stored over several years might suggest that a storage type vacuole exists in soybean seed coats to which foreign proteins can be directed with the SBP propeptides.

1.4.7 Vacuolar sorting signals

Three main categories of vacuolar sorting signals (VSSs) for soluble proteins exist. Sequence-specific VSSs (ssVSSs) can be located anywhere within a protein, although they are most often found at the N-terminus, following the N-terminal signal peptide that directs transport into the ER (Koide et al. 1997; Matsuoka and Neuhaus 1999). ssVSSs typically contain the conserved amino acid sequence [N/L]-[P/I/L]-[I/P]-[R/N/S], with an Ile or Leu residue being most critical, although there may be additional determinants that are necessary for vacuolar targeting (Joliffe et al. 2005; Matsuoka and Nehaus 1999; Vitale and Hinz 2005). This type of signal has been identified in the vacuolar thiol protease aleurain and the sweet potato storage protein sporamin (Holwerda et al. 1992; Matsuoka and Nakamura 1991).

In comparison, C-terminal VSS (ctVSSs) tend to have a large number of hydrophobic residues, but there appears to be no further conservation in sequence (Joliffe et al. 2005;

Matsuoka and Nehaus 1999; Vitale and Hinz 2005). As the name indicates, their position at the C-terminus of a protein is important for their function (Joliffe et al. 2005; Matsuoka and Nehaus 1999; Vitale and Hinz 2005). These ctVSSs have been identified in tobacco chitinase A, barley lectin, and common bean phaseolin (Neuhaus et al. 1994; Bednarek et al. 1990; Bednarek and Raikhel 1991; Dombrowski et al. 1993; Frigerio et al. 1998).

SBP has a C-terminal propeptide that is suspected to function as a vacuolar sorting signal, as discussed in Section 1.3.2. Its position at the C-terminus and the lack of the conserved amino acid sequence typically found in ssVSSs suggest that it could be classified as a ctVSS.

The third type of vacuolar signal, the physical structure VSS (psVSS), is less well-defined than the other two sorting signals. These signals are found within the polypeptide chain and form signal patches as a result of three-dimensional folding (Matsuoka and Neuhaus 1999). These types of signals have been identified in seed storage proteins such as the B-type legumin of *Vicia faba*, phytohemaglutinin of *Phaseolus vulgaris*, and phytepsin of *Hordeum vulgare* (Vitale and Hinz 2005).

Targeting in some proteins can be even more complex. Soybean glycinin, an 11S globulin, appears to contain both an ssVSS and a ctVSS, and there is evidence for a third sorting signal (Maruyama et al. 2006). Similarly, soybean β-conglycinin, a 7S globulin, has multiple sorting determinants, with a C-terminal signal that is responsible for transport to the PSV matrix, and a second, undefined signal that is responsible for transport to a globoid-equivalent compartment in *Arabidopsis* seeds (Nishizawa et al. 2003). However, the C-terminal signal appears to contain both a ctVSS and an ssVSS (Nishizawa et al. 2006).

1.4.8 Traffic to lytic vacuoles

Evidence suggests that the transport of soluble proteins to vacuoles occurs via different pathways. Wortmannin, a specific inhibitor of mammalian phophastidylinositol 3-kinase, inhibits transport of the ctVSS of barley lectin but not the ssVSS of sporamin, suggesting that the VSSs are differently transported (Matsuoka et al. 1995). Also, in pea cotyledons, storage proteins are highly concentrated in the *cis*-cisternae of the Golgi apparatus and virtually absent in the *trans*-network where the lytic vacuole receptor BP-80 is more concentrated, suggesting a spatial separation of the two sorting pathways in plant cells (Hillmer et al. 2001).

Proteins destined for LVs are transported in clathrin-coated vesicles from the Golgi apparatus to a prevacuolar compartment and then into the LV (Joliffe et al. 2005). Proteins targeted to LVs typically contain an ssVSS. A family of receptors, including pea BP-80 and *Arabidopsis* AtELP, have been identified that bind ssVSS and are associated with clathrin coated vesicles (Kirsch et al. 1994; Ahmed et al. 1997; Hinz et al. 1999). BP-80 binds to the proaleurain ssVSS and AtELP binds to the ssVSSs of barley and *Arabidopsis* proaleurains and sweet potato sporamin; however, they do not bind to the ctVSSs of barley lectin or tobacco chitinase (Kirsch et al. 1994; Ahmed et al. 2000). These receptors are typically found in the Golgi apparatus and prevacuolar compartments and they contain a motif that can bind to the adaptor complex of the clathrin coat in clathrin coated vesicles (Paris et al. 1997; Ahmed et al. 1997; Li et al. 2002; Happel et al. 2004).

1.4.9 Traffic to protein storage vacuoles

Soluble proteins targeted to PSVs may traffic through the Golgi apparatus to PSVs or they may follow a direct ER to PSV pathway. Proteins containing a ctVSS are typically sorted to PSVs. Storage proteins are transported from the Golgi apparatus to the PSV in dense vesicles

(Hohl et al. 1996). In developing pea cotyledons, one of the major storage proteins, prolegumin, was associated with dense vesicle fractions, as was the PSV marker α-TIP, supporting the role of dense vesicles in transport to PSVs (Hinz et al. 1999).

A different type of vesicle appears to be involved in the direct ER to PSV pathway. In maturing pumpkin seeds, precursor-accumulating (PAC) vesicles have been identified that emerge from the ER (Hara-Nishimura et al. 1998). PAC vesicles are larger than dense vesicles, and, in addition to an electron dense core, they are surrounded by an electron-translucent layer. The electron-translucent layer is associated with complex glycans, suggesting that glycoproteins originating from the Golgi are also incorporated into PAC vesicles. Some of these PAC vesicles also have smaller electron-translucent vesicles within them (Hara-Nishimura et al. 1998). Similarly, a novel vesicle has been identified in rice endosperm that forms from storage protein aggregates in the ER and traffics directly to PSVs (Takahashi et al. 2005). Also, in developing soybean cotyledons, protein bodies are formed from the ER and traffic directly to PSVs (Mori et al. 2004).

The mechanism of transport of proteins to PSVs via dense vesicles and PAC vesicles has not yet been fully elucidated. The presence of a ctVSS suggests that transport may require a receptor that specifically recognizes this signal. Furthermore, transport of the *Phaseolus vulgaris* storage protein phaseolin is saturable, which supports a receptor-mediated pathway (Holkeri and Vitale 2001). However, there is also evidence for an aggregation-based mechanism of protein sorting to PSVs. In developing pea cotyledon, aggregates of prolegumin and vicilin are visible in the Golgi and prolegumin is also more aggregated in dense vesicles than in the Golgi apparatus (Hillmer et al. 2001; Hinz et al. 1999). Furthermore, while prolegumin shows tight binding to membranes, mature legumin does not (Hinz et al. 1997; Wenzel et al. 2005). Also, phaseolin has

a tendency to form aggregates and a high affinity for membranes that is dependent on its ctVSS (Castelli and Vitale 2005).

It has been suggested based on morphological comparisons that the mechanism of transport of soluble proteins to the PSV may be similar to that for protein secretion in immature secretory granules found in mammalian endocrine and exocrine gland cells (Robinson and Hinz. 1999; Hillmer et al. 2001; Hinz et al. 1999). Based on this model, it has been suggested that storage protein aggregates form in close association with the Golgi membrane and this triggers budding of dense vesicles (Joliffe et al. 2005, Robinson et al. 2005). The aggregation would be mediated by transmembrane nucleation factors.

The receptor homology region-transmembrane domain-RING-H2 motif (RMR) proteins have been proposed as potential receptor/nucleation factor candidates for transport to PSVs. These receptors were identified based on homology of the lumenal domain of the receptor to a region present in the BP-80 family of receptors that is involved in ligand binding (Jiang et al. 2000; Cao et al. 2000). RMR proteins have an N-terminal signal peptide, followed by the region homologous to the ligand binding domain found in the BP-80 family of proteins, then a transmembrane region containing a RING-H2 domain, and finally a serine-rich region (Jiang et al. 2000). The RMR proteins have been found to bind specifically to the ctVSSs of tobacco chitinase but not to the ssVSS of proaleurain (Park et al. 2007). An *Arabidopsis* RMR protein, AtRMR1, co-immunoprecipitates with phaseolin when coexpressed in *Arabidopsis* protoplasts, but this interaction does not occur when the phaseolin ctVSS has been deleted (Park et al. 2005). Deletion of either the lumenal domain or the cytosolic tail of AtRMR1 caused it to be retained within the Golgi apparatus. Coexpression with phaseolin resulted in phaseolin also being

retained in the Golgi apparatus but did not affect sporamin and an aleurain-like protease that are sorted to the LV (Park et al. 2005).

These RMR proteins traffic through the Golgi apparatus to structures in the interior prevacuolar compartments that ultimately fuse with PSVs, depositing the RMR-labelled structures which then form the PSV crystalloid (Jiang et al. 2000). In *Arabidopsis* seeds, a distinct crystalloid does not form, but RMR proteins are still found within a crystalloid-like fraction (Gillespie et al. 2005). However, these RMR proteins do not appear to be recycled back to the Golgi apparatus, which is not consistent with their function as a receptor or nucleation factor (Jiang et al. 2000).

1.4.10 Limitations of the model for transport to plant vacuoles

The above results have suggested a model for vacuolar transport in which proteins containing ssVSSs are transported to lytic vacuoles via clathrin-coated vesicles mediated by the BP-80 family of receptors while storage proteins and other proteins containing ctVSSs are transported to PSVs via dense vesicles and PAC vesicles mediated either by receptors and/or an aggregation-based mechanism (Joliffe et al. 2005). A number of additional studies have suggested that this model may be inaccurate or at least incomplete.

Vacuolar sorting signals identified in some seed storage proteins appear to be ssVSSs, such as in the castor bean storage proteins 2S albumin and ricin, where a Leu or Ile residue is critical for accurate sorting, and these ssVSSs can interact with proteins that show high sequence similarity to members of the BP-80/AtELP family of receptors (Brown et al. 2003; Frigerio et al. 2001; Joliffe et al. 2005).

An *Arabidopsis* line in which the *AtELP* gene is disrupted by an insertion secretes storage proteins instead of transporting them to PSVs and a further analysis of AtELP demonstrated that

this receptor could bind to the ctVSS of 12S globulin (Shimada et al. 2003). Further analysis demonstrated that *AtELP* downregulation results in missorting of proteins bearing both ssVSSs and ctVSSs (Craddock et al. 2008). One member of the BP-80 family of receptors, pumpkin PV72, was initially found associated with PAC vesicle membranes and can bind the ctVSS of pro2S albumin in a calcium-dependent manner, suggesting it might function as a receptor for storage proteins (Shimada et al. 2002; Watanabe et al. 2002). However, expression of the lumenal domain of PV72 fused to the ER-retention signal HDEL, resulted in the retention of AtALEU, a lytic vacuole protein, in the ER (Watanabe et al. 2004). Futhermore, an analysis of the *Arabidopsis* vegetative vacuole proteome identified many vacuolar proteins, only some of which had either an ssVSS or a ctVSS, suggesting that there may be other additional pathways that have not yet been identified (Carter et al. 2004).

1.4.11 Targeting foreign proteins in soybean seed coats

A better understanding of protein targeting in plants is essential for creating an effective protein expression system in soybean seed coats. It is important to understand how the SBP N- and C-terminal propeptides may function and how this will impact the processing, storage, and, ultimately, function of the foreign proteins produced in the seed coats. For example, if the NTPP directs entry of foreign proteins into the plant secretory system, as expected, then these proteins will likely be glycosylated. If the foreign protein is a glycoprotein, this is crucial, since the activity of many proteins depends upon such modifications. However, in some cases, if the protein is not normally glycosylated, then this could negatively impact protein function. This may be the case for many bacterial proteins, since bacteria do not typically glycosylate their proteins. Initially, foreign proteins will be targeted using the SBP propeptides, which presumably will target them to the vacuole. However, there may be advantages to targeting the foreign

proteins to other compartments. This will be investigated in the model system *A. thaliana* using a reporter protein that consists of a fusion between SBP and green fluorescent protein (GFP) combined with different combinations of the SBP propeptides and the ER retention signal HDEL. Knowledge from this study can be used to further improve the soybean seed coat system by identifying advantages to targeting proteins to different compartments of the plant secretory system.

1.5 Soybean peroxidase as a model enzyme

SBP is naturally produced at high levels in *EpEp* soybean cultivars but is essentially absent in *epep* cultivars, which we are using as the host for the soybean seed coat expression system. One protein that we attempted to produce in soybean seed coats is SBP. This experiment has two purposes. First, SBP is an enzyme that has a number of useful industrial, environmental, and medicinal applications. The soybean seed coat system could be used to create an even greater source of this enzyme for use in these applications, further improving the value of soybean seed coats. Furthermore, current molecular technologies can be applied to enhance SBP activity. Future studies can focus on expression of such enhanced enzymes in soybean seed coats. Second, a comparison of transgenic *epep* soybean plants expressing recombinant SBP to wild-type *EpEp* soybean plants that naturally express SBP will indicate whether or not we have successfully designed our transgene to achieve the levels of expression that can normally be produced by the endogenous *Ep* promoter. The following sections will describe the useful properties and potential applications of SBP.

1.5.1 Properties of soybean peroxidase

SBP belongs to the class III family of plant peroxidases (Hiraga et al. 2001). SBP catalyzes an oxidoreduction reaction between hydrogen peroxidase and various reductants as seen in the following general reaction, where A represents the reductant: $AH_2 + H_2O_2 \rightarrow A + 2H_2O$ (Hiraga et al. 2001). The optimal pH for SBP is 6.0 but activity remains high between pH 2.2 and 8.0 (Geng et al. 2001). At pH 6.0, activity is highest at 80°C, three times higher than at room temperature, and more than 80% of this activity is maintained between 60 and 85°C. Activity remains relatively high in different organic solvents, with 80% of activity maintained in a 20% solution of acetone, acetylnitrile, and ethanol and slightly less than 80% for 20% methanol (Geng et al. 2001). SBP has a high thermal and conformational stability compared to HRP C due to a stronger heme-apoprotein affinity (Kamal and Behere 2002; Kamal and Behere 2008). The k_{cat} and k_{cat}/K_M values for SBP, using 2,2'-azino-bis-(3-ethylbenzthiazoline-6-sulphonate) (ABTS) as the substrate, are superior to HRP C, indicating that SBP is a more potent peroxidase (Kamal and Behere 2003).

1.5.2 Industrial applications of soybean peroxidase

Peroxidases have a number of potential industrial uses including the decolorization of synthetic dyes, organic and polymer synthesis, bioremediation of phenolic compounds, delignification in the pulp and paper industry, and the creation of biosensors. They can also be used in analysis and diagnostic kits and enzyme immunoassays (Regalado et al. 2004).

SBP shows good catalytic activity with the phenolic substrates eugenol, caffeic acid, and ferulic acid (Schmitz et al. 1997). It is also effective at removing phenolic compounds from contaminated soil and water systems (Caza et al. 1999; Geng et al. 2004). Its proficiency at removing the chromophore from the stilbene dye Direct Yellow 11 and the methane dye Basazol

46L, two commonly used paper dyes, also demonstrates its usefulness to the pulp and paper industry (Knutson et al. 2005). Furthermore, compared to HRP C, a chemiluminescent enzyme-linked immunosorbent assay with SBP as the conjugate was more sensitive and provided a wider linear range (Sakharov et al. 2006). It also may be possible through directed evolution to improve upon these properties.

1.6 Polyhydroxybutyric acid as a model for metabolic engineering of soybean seed coats

Metabolic engineering is the process of diverting cellular metabolism towards the production of new metabolites or to enhance or minimize production of already existing metabolites. This is usually accomplished through modification of existing enzymes in the metabolic pathway of interest or the introduction of novel enzymes. This technology has been used to alter the yield of primary metabolites, including starch, cellulose, amino acids, polyamines, and fatty acids, and secondary metabolites such as alkaloids, terpenoids, flavonoids, and lignin, as well as to introduce novel metabolites, of which the most classic example is the production of the biodegradable plastic PHB (Capell and Christou 2004).

Soybean plants that produce the enzymes necessary for the production of PHB can be created in order to investigate diversion of metabolism within the seed coat. PHB is one type of polyhydroxyalkanoate (PHA) that has been successfully produced in a number of plant species. It presents a renewable, environmentally-friendly alternative to petroleum-derived plastics. Current plant production systems are limited in their capacity to produce enough PHB per plant weight to be economically competitive with the petroleum industry. While this may not be possible to achieve in soybean seed coats, especially since the mass of the target tissue is relatively small compared to systems in which PHB is expressed throughout the plant,

production of PHB in soybean seed coats will still provide a good system for determining the extent to which carbon flow can be diverted towards the production of novel biochemicals. If PHB can be successfully produced in seed coats, this will indicate that some carbon can be diverted without affecting seed health. In the following sections, I will describe the synthetic pathway for PHB and how this has been transferred to plant systems, as well as some of the limitations to expressing PHB in plants.

1.6.1 Polyhydroxybutyric acid synthesis in bacteria

A number of different bacteria, such as *Ralstonia eutropha*, *Rhodospirillum rubrum*, *Halobacterium mediterranei*, *Rhizobium melilote*, *Methylocystis parvus*, *Methylosinus trichosporium* and members of the *Pseudomonas*, *Rhodococcus*, *Corynebacterium*, and *Nocardia* genera, produce PHAs as carbon and energy stores (Anderson and Dawes 1990). Over 100 different monomers can be incorporated into PHAs. As a result, the physical properties of the polymers range from stiff and brittle to quite soft and malleable, leading to the creation of a range of products including plastics, elastomers, rubbers, and glues (Poirier 1999; Steinbüchel and Valentin 1995). PHAs can be completely degraded to carbon dioxide and water through the action of PHA depolymerases that are secreted by many bacteria, and therefore have great potential as biodegradable polymers that could be environmentally-friendly substitutes for petroleum-based polymers (Jendrossek et al. 1996). A cradle-to-gate life cycle assessment of PHB produced by bacteria reported that it is superior in all life cycle categories to polypropylene and also superior to polyethylene in all categories except for acidification and eutrophication (Harding et al. 2007).

The best characterized PHA to date is PHB, which is synthesized from acetyl-Coenzyme A (acetyl-CoA) through the action of three enzymes (Figure 1.4) (Schubert et al. 1988; Slater et

al. 1988). First, two molecules of acetyl-CoA are condensed by the enzyme 3-ketothiolase to form acetoacetyl-CoA, which is then reduced to 3-hydroxybutyryl-CoA through the action of acetoacetyl-CoA reductase. PHB synthase then polymerizes the 3-hydroxybutyryl-CoA to PHB. This pathway was first characterized in *Ralstonia eutropha* (formerly *Alcaligenes eutrophus*), and it was found that the genes encoding the three enzymes, *phbA*, *phbB*, and *phbC*, respectively, are all encoded within a single operon (Peoples and Sinskey 1989a; 1989b).

1.6.2 Synthesis of polyhydroxyalkanoates in *Arabidopsis*

The potential of PHAs as an alternative to petroleum-based polymers has largely been limited by the high production costs associated with bacterial fermentation systems (Poirier 1999). Over the last fifteen years, the production of PHAs in plant systems has seen a lot of progress (Table 1.1). PHB was first produced in the model plant *A. thaliana*, through the introduction of the *phbB* and *phbC* genes, achieving levels of 20 to 100 μg/g fresh weight (FW) (0.002% to 0.01% FW); however, production of PHB had a negative impact on plant health (Poirier et al. 1992a; 1992b). These levels were increased to 14% of plant DW by the introduction of all three of the PHB biosynthetic genes modified by the addition of chloroplast targeting sequences (Nawrath et al. 1994), which further underlines the importance of organellar targeting. In addition to increasing the level of PHB production, targeting to the chloroplast had the additional benefit of reducing the negative impact on plant health.

The greatest level of PHB production in any plant system to date was achieved by including all three genes of the PHB biosynthetic pathway, with their chloroplast targeting signals, in a single construct, thereby allowing for a single transformation event (Bohmert et al. 2000). Using this approach, plants were identified that accumulated PHB up to 40% DW. This level of PHB production severely affected plant health, stunting growth and decreasing fertility

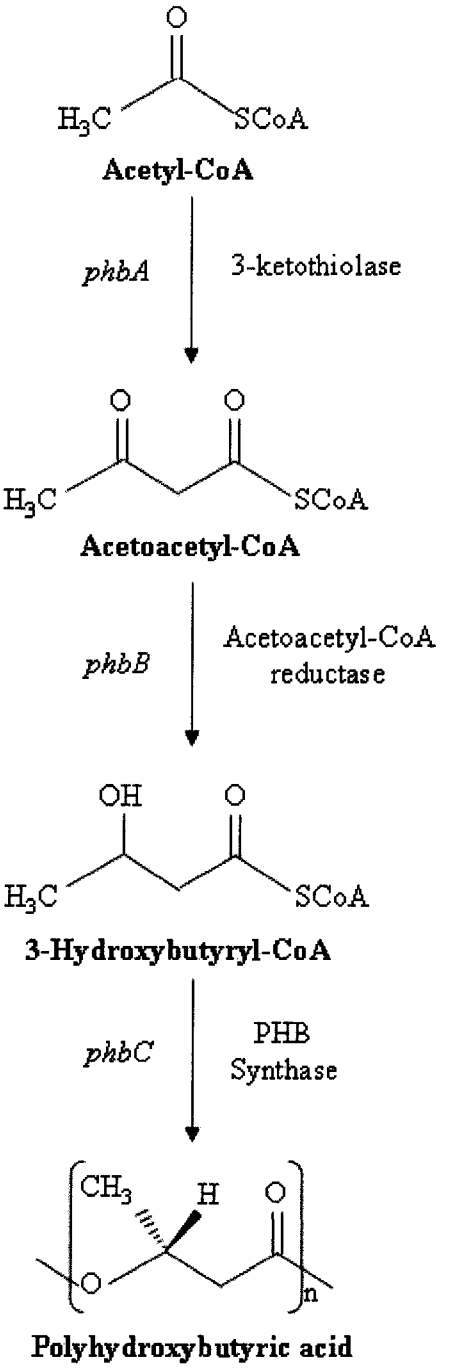


Figure 1.4 The polyhydroxybutyric acid biosynthesis pathway from *Ralstonia eutropha*. The enzyme and corresponding gene for each reaction are indicated. Adapted from Arai et al. (2004).

Table 1.1 Summary of studies in which polyhydroxyalkanoates are expressed in plants. Abbreviations: FW, fresh weight; DW, dry weight; PHA, polyhydroxyalkanoate; PHB, polyhydroxybutyrate; PHBV, poly(3-hydroxybutyrate-*co*-3-hydroxyvalerate); MCL-PHAs, medium chain length PHAs.

PHA	Genes	Plant Species	Location	Levels	Phenotypic Effects	References
PHB	*phbB*, *phbC* (Introduced separately)	*Arabidopsis*	Cytoplasm, vacuole, nucleus	0.002% – 0.01% FW[A]	Reduced growth and fertility	Poirier et al. 1992a, 1992b
PHB	*phbA*, *phbB*, *phbC* (Introduced separately)	*Arabidopsis*	Chloroplast	Up to 1% FW[A] (14% DW)	No effect	Nawrath et al. 1994
PHB	*phbA*, *phbB*, *phbC*	*Arabidopsis*	Chloroplast	Up to 40% DW (4% FW)	Reduced growth and fertility	Bohmert et al. 2000
PHB	*phbB*, $phaC_{AC}$	Tobacco	Cytoplasm	Up to 0.001% FW[A]	No effects	Nakashita et al. 1999
PHB	*phbA*, *phbB*, *phbC* (Bacteria-type operon)	Tobacco	Chloroplast	0.0002% to 0.0008% DW[B]	No effects	Nakashita et al. 2001
PHB	*phbA*, *phbB*, *phbC* (Bacteria-type operon with modified regulatory sequences)	Tobacco	Chloroplast	0.005 to 0.016% DW[B]	Embryogenesis and seed germination affected	Arai et al. 2004
PHB	*phbA*, *phbB*, *phbC* (Bacteria-type operon with modified regulatory sequences)	Tobacco	Chloroplast	Up to 1.7% DW	Reduced growth, male sterility	Lössl et al. 2003
PHB	*phbA*, *phbB*, *phbC* (Bacteria-type operon with inducible promoter)	Tobacco	Chloroplast	Up to 0.1383% DW[B]	Leaf bleaching	Lössl et al. 2005
PHB	*phbA*, *phbB*, *phbC*	*Brassica napus*	Leukoplasts	Up to 7.7% seed FW	No effects	Valentin et al. 1999; Houmiel et al. 1999
PHB	*phbA*, *phbB*, *phbC*	Black Mexican Sweet maize suspension cultures	Peroxisome	Up to 0.2% FW[A]	No effects	Hahn et al. 1999
PHB	*phbA*, *phbB*, *phbC*	Flax (*Linum usitatissimum* L.)	Chloroplast	Up to 0.005% FW[A]	Chlorosis, reduced growth, senescence	Wróbel et al. 2004

(Continued)

Table 1.1 (Continued)

PHA	Genes	Plant Species	Location	Levels	Phenotypic Effects	References
PHBV	*bktB*, *phbB*, *phbC*, threonine deaminase	*Arabidopsis*, *Brassica napus*	Chloroplast/ Leukoplast (*Brassica* seeds)	Less than 3% DW in *Arabidopsis* leaves or *Brassica napus* seed	Some sterility, reduced fertiliy, low vigor, and death in *Brassica*	Valentin et al. 1999; Slater et al. 1999
MCL-PHAs	*phaC1*	*Arabidopsis*	Peroxisome	Up to 0.4% DW[A]	No effects	Mittendorf et al. 1998
PHB	*phbB*, *phbC*	Cotton (*Gossypium hirsutum* L.)	Cytoplasm	0.003 %to 0.344% DW of fibres[A]	No effects	John and Keller 1996
PHB	Synthetic *phbA*-*phbB* fusion, *phbC*	*Arabidopsis*	Chloroplast	Up to 6.4% DW	Chlorosis, reduced growth	Kourtz et al. 2005
	phbA, *phbB*, *phbC* (inducible promoter)	*Arabidopsis*	Chloroplast	Up to 14.3% DW in young leaves	Chlorosis, reduced growth	Kourtz et al. 2007
PHB	*phbA*, *phbB*, *phbC*	Sugarcane (*Saccharum* spp. Hybrinds)	Cytoplasm, mitochondria, and chloroplasts	Trace, undetectable, and up to 1.88% DW, respectively	No effect	Petrasovits et al. 2007; Purnell et al. 2007
PHB, PHBV	*phbA*/*bktB*, *phbB*, *phbC*	Alfalfa (*Medicago sativa* L.)	Chloroplast	0.0025% to 0.18 DW[A]	No effect	Saruul et al. 2002
PHB	*phbA*, *phbB*, *phbC* (inducible or somatically activated promoters)	Tobacco	Chloroplast	Up to 0.0375% FW (0.32% DW)[A]	Chlorosis, changes in leaf morphology, number of trichomes	Bohmert et al. 2002
PHB	*phbA*, *phbB*, *phbC* (inducible or somatically activated promoters)	Tobacco	Chloroplast	Up to 0.0011% FW (0.0087% DW)[A]	No effect	Bohmert et al. 2002
PHB	*phbA*, *phbB*, hybrid *Pseudomonas oleovoran* / *Zoogloea ramigera* PHA synthase	Switchgrass (*Panicum virgatum* L.)	Chloroplasts	Up to 3.72% DW in leaf tissues and 1.23% in whole tillers	Some slight reduction in growth, chlorosis	Somleva et al. 2008

[A]Converted from mg/g to % weight based on the equivalency 10 mg/g = 1 % weight
[B]Converted from ppm to % weight based on the equivalency 10,000 pmm = 1% weight

(Bohmert et al. 2000). Kourtz et al. (2007) placed the three PHB biosynthetic enzymes under control of an inducible promoter to delay PHB synthesis and improve plant health, but negative plant effects often still developed following induction of the system. Through the addition of a fourth gene encoding a threonine deaminase, a copolymer with both 3-hydroxybutyrate and 3-hydroxyvalerate units at levels of 0.2 to 0.8% DW was produced in *Arabidopsis* (Valentin et al. 1999; Slater et al. 1999). *Arabidopsis* plants expressing a PHA synthase targeted to the peroxisome were able to produce medium-chain length PHAs at levels of 4 mg/g DW (0.4% DW) (Mittendorf et al. 1998). The medium-chain length PHAs were derived from 3-hydroxyacyl-CoA that are produced as intermediates of β-oxidation of alkanoic acids.

1.6.3 Synthesis of polyhydroxyalkanoates in other plant species

A number of other studies have focused on the production of PHAs in tobacco. Expression of *phbB* and *phbC* genes in tobacco yielded PHB at less than 10 μg/g FW (0.001% FW) (Nakashita et al. 1999). Similarly, plastid transformation of tobacco plants with all three PHB biosynthetic enzymes in a bacteria-type operon yielded low levels of PHB of less than 8 ppm DW (0.0008% DW) (Nakashita et al. 2001). By modifying the bacteria-type operon to include plant regulatory sequences, PHB production could be increased to 50 to 160 ppm DW (0.005% to 0.016% DW) (Arai et al. 2004). Some of the deleterious effects of PHB production on plant health were found to be limited by inducible expression of the PHB bacteria-type operon in transplastomic tobacco plants, such that PHB levels of 1383 ppm DW (0.1383% DW) were produced without affecting flowering and fertility (Lössl et al. 2005).

PHAs have been produced in several other plant species, including Black Mexican Sweet maize suspension cultures (Hahn et al. 1999), *Brassica napus* seeds (Valentin et al. 1999; Slater et al. 1999; Houmiel et al. 1999), flax (Wróbel et al. 2004), cotton (John and Keller 1996;

Chowdhury and John 1998), sugarcane (Petrasovits et al. 2007), potato (Bohmert et al. 2002), alfalfa (Saruul et al. 2002), and switchgrass (Somleva et al. 2008).

In some of the above examples, the goal was not to produce a source of PHB for harvesting but to create novel biocomposite products with improved characteristics. In the case of flax, the tensile strength was improved, which would benefit the environmentally-friendly automobile and construction industries (Wróbel et al. 2004). In cotton, the thermal properties were altered, resulting in improved insulation, increased heat capacity, and lower thermal properties (John and Keller 1996; Chowdhury and John 1998).

1.6.4 Compartmentation of the polyhydroxybutyric acid biosynthetic pathway

The acetyl-CoA necessary for the production of PHB is found in a number of plant compartments where it is involved in different metabolic pathways. In the cytoplasm, acetyl-CoA is necessary for the mevalonate pathway, in the plastids it is involved in the synthesis of fatty acids, and in the peroxisome it is produced from fatty acid β-oxidation. PHB biosynthetic enzymes targeted to the chloroplast and peroxisome have resulted in increased levels of PHB, possibly as a result of the presence of larger acetyl-CoA pools compared to the cytoplasmic pools (Hahn et al. 1999; Nawrath et al. 1994). PHB has been found in plant vacuoles and the nucleus when the untargeted PHB biosynthetic enzymes were expressed (Poirier et al. 1992a; 1992b). Similarly, when the PHB biosynthetic enzymes were targeted to the peroxisome, PHB was occasionally found within the vacuole (Mittendorf et al. 1998). This phenomenon has led to the suggestion that PHB granules may be able to pass through the membranes of these organelles (Mittendorf et al. 1998; Poirier et al. 1992a; 1992b). Alternatively, the PHB granules may enter the nucleus during cell division (Poirier et al. 1992a; 1992b).

No study to date has targeted the PHB biosynthetic enzymes to the vacuole. Plant vacuoles serve as storage compartments, and many metabolites can be found in high concentrations in this compartment (Marty 1999). For instance, malate concentrations are greater than 5-fold higher in the vacuole compared to the chloroplast stroma and greater than 8-fold compared to the cytosol in spinach leaves (Winter et al. 1994). It is possible that vacuoles could contain a pool of acetyl-CoA that could be redirected towards the production of PHB; however, there are few studies that have examined the subcellular distribution of acetyl-CoA, and especially not acetyl-CoA concentrations in the vacuole.

1.6.5 Challenges to the production of polyhydroxyalkanoates in plants

Several challenges remain for the efficient production of PHAs in plants. One of the major challenges is producing enough PHA in plants to be cost-competitive with petroleum-derived products. It has been estimated that PHAs must accumulate to at least 15% of DW in order to achieve this (Conrad 2005). Although these levels have been produced in the model plant *A. thaliana* (Bohmert et al. 2000), they have not yet been produced within a crop species.

The second challenge will be minimizing negative effects on the plant host. Plants expressing high levels of PHAs often show reduced growth and fertility, indicative of stress either directly from the presence of PHA or indirectly as a result of the diversion of plant metabolism (Bohmert et al. 2000; Arai et al. 2004; Lössl et al. 2003; Wróbel et al. 2004; Slater et al. 1999; Mittendorf et al. 1998; Mittendorf et al. 1999). Bohmert et al. (2002) demonstrated that in tobacco and potato, constitutive expression of the *phbA* gene was detrimental to plant health. Nawrath et al. (1994) found that targeting the PHB biosynthetic enzymes to the chloroplast helped reduced negative phenotypic effects from PHB production. Also, use of the stem-specific

14-3-3 promoter in flax eliminated the negative effects produced as a result of constitutive expression (Wróbel et al. 2004).

Metabolic profiling of PHA producing plants may give insight into the source of plant stresses resulting in the abnormal phenotypes. This approach was taken with *Arabidopsis* plants accumulating PHB in the chloroplast, using gas chromatography-mass spectrometry (GC-MS) (Bohmert et al. 2000). Although no changes were observed in the levels of the fatty acids examined, decreases were found in the levels of fumarate and isocitrate, possibly indicating a reduction in the activity of the tricarboxylic acid cycle (Bohmert et al. 2000). There were also increases in the levels of osmoprotectants suggested to be indicative of stress, particularly osmotic stress (Bohmert et al. 2000). This study clearly demonstrates that the production of PHB in plants can cause multiple changes in plant metabolism that must be addressed in order to balance PHB production and plant health.

A further challenge for the production of biodegradable plastics in plants is the production of PHAs other than PHB. The PHB biosynthetic pathway has been preferentially expressed in plants because of its simplicity. However, PHB is not an ideal polymer because it is too stiff and brittle for many applications (Poirier 1999). Some alternatives to PHB that have been produced in plants are the copolymer poly(3-hydroxybutyrate-*co*-3-hydroxyvalerate) and medium chain length PHAs (Valentin et al. 1999; Slater et al. 1999; Mittendorf et al. 1998; Mittendorf et al. 1999). However, these alternatives are subject to many of the same limitations described above for PHB and often have more complex genetic pathways, making their introduction into plants more difficult.

The goal of producing PHB in soybean seed coats was to investigate the redirection of carbon flow towards the production of a novel biochemical. PHB was chosen since the pathway

is relatively simple and has been successfully introduced into plant systems in past studies. The SBP N- and C-terminal propeptides were added to each of the three biosynthetic enzymes, ideally targeting them to vacuoles in hourglass cells. Since vacuoles function as storage compartments for excess metabolites, this compartment may be an ideal location for introducing novel biochemical pathways.

1.7 Hypotheses and objectives

For this research project, several aspects of the soybean seed coat expression system were investigated. In particular, the following four hypotheses were tested.

Hypothesis 1: The SBP N-terminal propeptide is a targeting signal that can be used to target proteins to the plant secretory system and the SBP C-terminal propeptide is a vacuolar sorting signal that can be used to target proteins in the secretory system to plant vacuoles.

Objective 1: Determine the subcellular location of a reporter protein expressed with no targeting signals, with the SBP N-terminal propeptide, with the N-terminal propeptide and an ER-retention signal, and with both the SBP N- and C-terminal propeptides in the model plant *Arabidopsis thaliana*.

Hypothesis 2: The yield of a foreign protein is greater if targeted to the vacuole compared to the yield of non-targeted, ER-retained, and apoplast-targeted proteins.

Objective 2: Compare the level of activity of a reporter enzyme targeted to the cytoplasm, endoplasmic reticulum, apoplast, and vacuole.

Hypothesis 3: Transformation of an *epep* soybean cultivar with the *Ep* cDNA driven by the *Ep* 1.5 promoter yields a transgenic soybean with seed coat levels of SBP comparable to *Ep* soybean cultivars.

Objective 3: Compare the peroxidase specific activity in seed coats of *epep* plants transformed with the *Ep* gene, wild-type *Ep* plants, and wild-type *epep* plants.

Hypothesis 4: Introduction of a novel biochemical pathway in soybean seed coat vacuoles redirects carbon flow towards the production of a novel biochemical.

Objective 4: Introduce the three genes encoding the biosynthetic pathway for the biodegradable plastic PHB into soybean plants, driven by the seed coat specific *Ep* 1.5 promoter and targeted with the SBP N- and C-terminal propeptides, and measure the yield of PHB in seed coats.

CHAPTER 2 MATERIALS AND METHODS

2.1 Source of chemicals and materials

The chemicals and materials used in this project are listed below in alphabetical order, showing manufacturer and catalogue number.

2.1.1 Source of chemicals

Acetone (BDH, ACS 009); acrylamide (Sangon Ltd. Canada, A 001); agar (Fisher, A360-500); agarose (Invitrogen, 15510-027); ammonium persulfate (Fisher, BP179-100); ampicillin sodium salt (Fisher, BP1760-5); 2,2'-azino-bis (3-ethylbenzthiazoline-6-sulfonic acid) (Sigma, A9941 or Pierce, 34026); benzoic acid (Fisher, A65-500); boric acid (H_3BO_3) (BDH, ACS 141); 5-bromo-4-chloro-3-indolyl phosphate p-toluidine salt (BCIP) (Fisher, BP1610-100); bromphenol blue (BioRad, 161-0404); calcium chloride ($CaCl_2$) (BDH, ACS 186); calcium nitrate tetrahydrate ($Ca(NO_3)_2 \cdot 4H_2O$) (Fisher, C109-500); cellulase Onozuka R-10 from *Trichoderma viride* (Serva, 16419); chloroform (Fisher, C298-500); cupric sulphate, anhydrous ($CuSO_4$) (Fisher, C495); dimethyl formamide (DMF) (ACS); dimethyl sulfoxide (DMSO) (OmniSolv, B90140); ethanol 99% (Commercial Alcohold Ltd., P210-EA95); ethidium bromide (PlusOne, 17-1328-01); ethylenediaminetetraacetic acid disodium salt (EDTA) (Boehringer Mannheim, 808 288); extract of yeast (BDH 1.03753); ferrous sulphate heptahydrate ($FeSO_4 \cdot 7H_2O$) (Fisher, S93247); ficoll (Sigma, F-4375); glycerol (Fisher, G33-1); glycine (Fisher, BP381-1); HEPES (Boehringer Mannheim, 737 151); hydrochloric acid (HCl) (Fisher, UN1789); hydrogen peroxide 30% (H_2O_2) (Sigma, H-1009-100ML); hygromycin B (Calbiochem, 400051); javex (bleach) (Lavo); kanamycin sulfate (Gibco, 11815-024); maceroenzyme R-10 from *Rhizopus* sp. (Serva, 28302); magnesium chloride ($MgCl_2$) (BDH, B10149); magnesium sulphate heptahydrate ($MgSO_4 \cdot 7H_2O$) (BDH, ACS 486); manganese sulphate monohydrate ($MnSO_4 \cdot H_2O$) (Fisher, M120-500);

D-mannitol (Fisher, M120-500); 2-mercaptoethanol (Kodak, 117-0208); methanol (Fisher, A412-4); N,N'-methylenebisacryladmide (Acros Organics, 16 3801000); N-methyl N-trimethylsilyltrifluoroacetamide (MSTFA) (Thermo Scientific, 48913); 4-morpholineethanesulfonic acid (MES) (Sigma, M3671-50G); Murashige and Skoog basal salt mixture (MS salt) (Sigma, M5524); non-fat dry milk (Local supermarket); peptone from casein (EMD Chemicals, 1.07213.2500); poly[(R)-3-hydroxybutyric acid] (Fluka, 81329); ponceau S concentrate (Sigma, P7767); potassium acetate (EM Science, B10 350-34); potassium chloride (BDH, B10198); potassium dihydrogen phosphate (KH_2PO_4) (Fisher, P-28513); potassium nitrate (KNO_3) (Sigma, P6030-1KG); 2-propanol (Fisher, A416-500); Silwet L-77 (Lehle Seeds, VIS-02); sodium azide, (Fluka, 71289); sodium chloride (EMD Chemicals, SX0420-5); sodium dodecyl sulfate (SDS) (BDH, B44244); sodium molybdate ($Na_2MoO_4 \cdot 2H_2O$) (BDH, ACS 822); sodium phosphate (Na_2HPO_4) (JT Baker Chemical Co., 3828); sucrose (Fisher, S5-3); TEMED (Fisher, BP150-20); timentin (Smithkline Beecham, DIN 01916939); trichloroacetic Acid (TCA) (Acros Organics, 152135000); ultrapure Tris (Invitrogen, 15504-020); zinc chloride ($ZnCl_2$) (Sigma, 208086-100G).

2.1.2 Sources of materials

1302-tCZ-Prx plasmid (Dr. L. Gudynaitę-Savitch, University of Ottawa, Ottawa, ON); *Agrobacterium tumefaciens* strain GV3101 competent cells (Dr. B. Miki, AAFC, Ottawa, ON, prepared according to Hellens et al. 2000); Coomassie (Bradford) Protein Assay Kit (Pierce, 1856209); DNase I, Amplification Grade (Invitrogen, 18068-015); DNeasy plant mini kit (50) (Qiagen, 69104); dNTPs (UBI, DNTP-01); *entCUP4-GUS-nos* in pTZ19R (Dr. B. Miki, AAFC, Ottawa, ON, as referenced in Malik et al. 2002); Expand Hi Fidelity PCR system (Roche Diagnostics, 1732650); GenElute Plasmid Miniprep Kit (Sigma, PLN350-1KT); GeneRuler™

100 bp DNA ladder (Fermentas, SM0241); GeneRuler™ 1kb DNA ladder (Fermentas, SM0311); Goat anti-rabit IgG (H+L)-AP conjugate (BioRad, 170-6518); High Purity Taq DNA polymerase, recombinant, 10x Reaction Buffer, and 20 mM $MgSO_4$ (UBI, HPTaq-01); Klenow fragment (Fermentas, EP0051); Mark 12™ unstained standard (Invitrogen, LC5677); NBT/BCIP Reagent Kit (Molecular Probes, N6547); OneStep RT-PCR kit (QIAGEN, 210212); pCAMBIA1300 and other pCAMBIA plasmids (CAMBIA, www.cambia.org); pCAMBIA1300-tCUP#1 (Dr. L. Gudynaitę-Savitch, University of Ottawa, Ottawa, ON); peroxidase from soybean (Sigma, P-1432); pPRX-1 plasmid (Dr. M. Gijzen, AAFC, London, ON); pPRX-3 (Dr. M. Gijzen, London, AAFC, London, ON); 5'PrxSoy03 (Dr. D.A. Johnson, University of Ottawa, Ottawa, ON); plant protease inhibitor cocktail (Sigma, P9599-1ML); Precision Plus Protein™ All Blue Standards (BioRad, 161-0373); Precision Plus Protein™ Unstained Standards (BioRad, 161-0363); primers (Sigma or IDT); PRX_CAMBIA1302 (S. Han, University of Ottawa, Ottawa, ON); PVDF membrane (Millipore, IPVH00010); QIAEX II Gel Extraction Kit (Qiagen, 20021); QIAquick Gel Extraction Kit (Qiagen, 28704); QIAquick PCR Purification Kit (Qiagen, 28106); rabbit polyclonal anti-ascorbate oxidase (Agrisera, AS06 184); rabbit polyclonal anti-BiP antibody (at-95) (Santa Cruz Biotechnologies, SC-33757); rabbit polyclonal anti-cFBPase (cytosolic fructose-1,6 bisphosphatase) antibody (Agrisera, AS04 043); rabbit polyclonal anti-GFP (FL) antibody (Santa Cruz Biotechnologies, SC-8324); rabbit polyclonal anti-SBP antibody (M.C. Romero, AAFC, London, ON); rabbit polyclonal anti-V-ATPase (H^+-ATPase) antibody (Agrisera, AS07 213); rabbit polyclonal anti-Xet (xyloglucan endotransglucosylase) (Agrisera, AS05 080); recombinant GFP protein (Roche, 11 814 524 001); REDExtract-N-Amp™ Plant PCR kit (Sigma, XNAP-1KT); restriction enzymes (Fermentas or Invitrogen); RNeasy Plant Mini Kit (Qiagen, 74904); soybean seeds from cvs. Jack, Harosoy 63, Kunitz, X5, Fayette, and

Williams (Dr. D. Brown, AAFC, London, ON); Subcloning Efficiency DH5α chemically competent *E. coli* cells (Invitrogen, 18265-017); T4 DNA Ligase (Invitrogen, 15224-017).

2.2 Construction of plasmids

2.2.1 General methodologies

Standard molecular techniques were used for plasmid construction (Sambrook et al. 1989). *Escherichia coli* DH5α (Invitrogen) was grown in LB broth (10 g/L peptone from casein, 5g/L extract of yeast, 10 g/L NaCl, pH ~7.0) or on LB plates (LB with 15 g/L agar) with either kanamycin or ampicillin at 50 μg/ml as required. Plasmid DNA was isolated from *E. coli* using the GenElute Plasmid Miniprep Kit. For PCR amplification, the Expand Hi Fidelity PCR system was used. A typical amplification was performed in 50 μl containing 1xPCR buffer, 1.5 mM $MgCl_2$, 0.2 mM dNTPs, 0.5 to 0.7 μM of each primer, 2.6 U of Expand Hi Fidelity Enzyme Mix, and between 0.1 and 250 ng of template. PCR cycling was as follows: initial DNA denaturation at 94°C for 2 min followed by 30 cycles of 15 s at 94°C, 30 s at the annealing temperature of the primer pair, and then 30 s to 1 min at 72°C, depending on the size of the product, plus an additional 5 s each cycle, followed by a final elongation at 72°C for 10 min. PCR products were isolated with the QIAquick PCR Purification kit.

A typical restriction digest was performed in 10 μl and contained 1x or 2x reaction buffer (Fermentas; Invitrogen) as required by the restriction enzyme, 5 U of the restriction enzyme, and between 0.5 and 2 μg of DNA. Digests were incubated at 37°C for 30 min to 1 h, with the exception of *Sma*I digests which were incubated at 30°C for 1 h.

Restriction fragments were blunt-ended with the Klenow fragment of *E. coli* DNA polymerase I when required. A typical reaction was performed in 30 μl and contained 1x React2

buffer (Invitrogen), 16.6 μM dNTPs, 0.5 U of the Klenow fragment, and between 0.5 and 1 μg of DNA. A typical ligation was performed in 10 μl and contained 1x DNA ligase reaction buffer (Invitrogen), 0.5 U of T4 DNA ligase, 2.5 to 25 ng of vector and 7.5 to 75 ng of insert. Ligations were typically incubated overnight at room temperature except for blunt ended ligations which were incubated at 14°C overnight.

PCR products and restriction digests were resolved by electrophoresis in agarose gels containing approximately 0.5 μl/ml ethidium bromide. Fragments were isolated from gels using the QIAEX II Gel Extraction Kit or the QIAquick Gel Extraction kit.

2.2.2 Construction of soybean peroxidase-green fluorescent protein plasmids

Four constructs were designed containing a gene fusion between the *Ep* gene encoding SBP and the *mGFP5* gene encoding GFP. Expression of the fusion gene is driven by the *entCUP4* promoter (Malik et al. 2002) and the gene is followed by the *nopaline synthase* (*nos*) transcriptional terminator. The four different constructs contained either no targeting signals, or sequence encoding the SBP N-terminal propeptide alone, the SBP N-terminal propeptide plus a C-terminal ER retention signal, or both of the SBP propeptides. These constructs are predicted to produce an SBP-GFP fusion protein that will be localized in the cytoplasm, apoplast, ER, and vacuole, respectively. These constructs will be transformed into *A. thaliana* plants and the plants will be analyzed to determine the location of each of the fusion proteins, confirming the function of the SBP propeptides, and to compare the yield of the foreign protein in the four different plant subcellular compartments. The cloning strategy is described here.

Plasmids for targeting SBP-GFP to different subcellular locations were based upon pCAMBIA1300. The *nos* terminator sequence was amplified from the plasmid pCAMBIA1302 by PCR using primers JM-nos For and JM-nos Rev (see Table 2.1 for primer sequences and

description), digested with *Hin*dIII and ligated into pCAMBIA1300, also digested with *Hin*dIII, to generate the vector p1300-nos. The enhanced cryptic tobacco promoter, *entCUP4*, was amplified from the transgene *entCUP4-GUS-nos* in pTZ19R plasmid (Malik et al. 2002), using primers JM-entCUP4 F and JM-entCUP4 R. The resulting PCR product was digested with *Eco*RI and *Sac*I and ligated to the p1300-nos vector digested with *Eco*RI and *Sac*I to generate the vector p1300-BV.

The portion of the *Ep* gene encoding mature SBP (nucleotides 79 to 996), plus, for convenience of cloning, the last amino acid from the N-terminal propeptide, was amplified from 5'PrxSoy03 plasmid using primers JM-SBP For and JM-SBP Rev. Additionally, the entire *Ep* gene except for the portion encoding the C-terminal propeptide was amplified using primers JM-ER TS For and JM-SBP Rev. The 5'PrxSoy03 plasmid contains the entire *Ep* cDNA cloned into the pCMV-BK vector. Both PCR products were digested with *Nco*I and *Bam*HI and ligated separately to the p1300-BV vector, which was also digested with *Nco*I and *Bam*HI, generating pBV-mSBP and pBV-TS-SBP vectors, respectively.

The *mGFP5* gene encoding a modified GFP (Siemering et al. 1996) was amplified from pCAMBIA1302 using primers JM-mGFP5 For and JM-mGFP5 Rev1. This PCR product was digested with *Bam*HI and *Xba*I as were the pBV-mSBP and pBV-TS-SBP vectors. The PCR product was ligated separately to each vector, generating constructs cSBP-GFP (Figure 3.2A) and aSBP-GFP (Figure 3.2B), respectively. Between the *Ep* and *mGFP5* genes is a linker peptide (Gly-Ser-Ala). The first two amino acids are derived from the *Bam*HI cloning site and the third amino acid was introduced with the JM-mGFP5 For primer. The same linker peptide can be found in eSBP-GFP and vSBP-GFP constructs, described below.

For the eSBP-GFP construct, the primers JM-mGFP5 For and JM-mGFP5 Rev2 were

Table 2.1 List of primers used in cloning, and PCR and RT-PCR analysis of transgenic plants.

Primer Name	Primer Sequence	Description
JM-nos For	CCCAAGCTTGGTGACCAGCTCGAATTTCC	Underline - introduced *Hin*dIII site
JM-nos Rev	CCCAAGCTTCCCGATCTAGTAACATAGATGAC	Underline -introduced *Hin*dIII site
JM-entCUP4 F	CCGGAATTCCAGGGGGATCTTCTGCAAGC	Underline - introduced *EcoR*I site
JM-entCUP4 R	GACGAGCTCCCATGGTGGAACCTCAAGACCG	Underline - introduced *Sac*I and *Nco*I sites; boldface - nucleotides changed to eliminate *Xba*I and *Bam*HI sites
JM-SBP For	CATG**CCATGGCT**CAGCTTACTCCTACGTTC	Underline - introduced *Nco*I site; boldface - start codon and extra codon to allow for *Nco*I site
JM-SBP Rev	CGCGGATCCCGAGTCTCCATTCACAAAATTAC	Underline - introduced *Bam*HI site
JM-ER TS For	CATG**CCATGG**GTTCCATGCGTCTATTAG	Underline - introduced *Nco*I site; boldface - start codon
JM-mGFP5 For	CGC**GGATCCGCA**AGTAAAGGAGAAGAACTTTTCACTG	Underline - introduced *Bam*HI site; boldface - sequence encoding linker
JM-mGFP5 Rev1	TGCTCTAGA**TCA**TTTGTATAGTTCATCCATGCCATG	Underline - introduced *Xba*I site; boldface - stop codon
JM-mGFP5 Rev2	TGCTCTAGA**TCA***CAGCTCGTCATGAGCTCCAGC*TTTGTATAG TTCATCCATGCCATG	Underline - introduced *Xba*I site; boldface - stop codon; italics - sequence encoding ER-retention signal and linker
JM-VS Rev	TGCTCTAGA**TTA**TTTAGATTGAGCAACAAGCTTTTG	Underline - introduced *Xba*I site; boldface - stop codon.
tCUP-for	GAACCCACCATGTTGGGCAGGGGGATCTTCTGCAAGC	Underline - *Bst*XI site
tCUP-rev	CAGGCTTTTTCATGGTGGATCCTCTAGACCG	Underline - region of overlap with Hygr-for primer
Hygr-for	GAGGATCCACCATGAAAAAGCCTGAACTCACCGCG	Underline - region of overlap with tCUP-rev primer
Hygr-rev	GGAGAAACTCGAGCTTGTCGATC	Underline - *Xho*I site.
Ep-1.5F	CGCAAGCTTTAGATAAAAAAATGGGATATAATTTTTC	Underline – introduced *Hin*dIII site
Ep-0.6F	CGCAAGCTTCAATGTTCCAAAACCTAATGC	Underline - introduced *Hin*dIII site

(Continued)

Table 2.1 (Continued)

Primer Name	Primer Sequence	Description
Ep-Rev	CGCCCATGGTTTGAGTTAATATGATTTTACCTCTC	Underline - introduced *Nco*I site
NOS-Rev	CCCGCCAATATATCCTGTCA	
EpcDNA-for	CTACCATGGGTTCCATGCGTCTATTAGTAG	Underline - introduced *Nco*I site
EpcDNA-Rstop	AGTGGTCACCTTATTTAGATTGAGCAACAAGCTTTTGTT	Underline - introduced *Bst*EII site
EpcDNA-Rwhole	ATAGGTCACCTGATTAATTAAGTACCTTAAAGAGATTCAA	Underline - introduced *Bst*EII site
Prx-F	CGCGAGCTCCCGGGTACCTGCAGTAGATAAAAAAATGGG	Underline - introduced *Sst*I, *Sma*I, *Kpn*I, and *Pst*I sites
Prx-R	CGCGGATCCCTACCATGGTTTGAGTTAATATG	Underline - introduced *Bam*HI and *Nco*I sites
JM-nos-F2	CTAGTCTAGAGGTGACCAGCTCGAATTTCC	Underline - introduced *Xba*I and *Bst*EII sites
JM-nos-R2	CCCAAGCTTCCCGGGAGCTCGTCGACCCGATCTAGT	Underline - introduced *Hin*dIII, *Sma*I, *Sst*I, *Sal*I sites
GmTub2-For	GTGACTTGAACCATCTGATCTAGC	
GmTub2-Rev	GTTGAAGCCATCCTCAAGCCAG	
HptII (99) For	GATTTGTGTACGCCCGACAGT	
HptII (614) Rev	CCGCAAGGAATCGGTCAAT	
tCUP (20) For	GGGATCTTCTGCAAGCATCTC	
tCUP (432) Rev	TGGCGGTTGGTGAGAAAT	
nos (13) For	AATCCTGTTGCCGGTCTT	
nos (192) Rev	TTATCCTAGTTTGCGCGCT	
Ep1.5F32	TAGATAAAAAAATGGGATATAATTTT	
Ep-ori-Rev	CCCATGGTTTGAGTTAATATG	
Ep-0.3F	CGCAAGCTTCCAACCACATTTAAGAGATTATAG	

(Continued)

Table 2.1 (Continued)

Primer Name	Primer Sequence	Description
JM-sphbA1-F1	TATGAGTGCAGCTAAAGCTAAGG	
JM-sphbA-R	TTTTCTCTCCACAGCAAGTGCC	
Ep-phbA-F	AGTGGCGAGCTGCAGTAG	
phbInt-R	CTGGTGCTGGAATCTTTGC	
JM-sphbB-F1	AAGAATTGCTTATGTTACTGGAGG	
JM-sphbB-R	CCTCCCATATGAAGTCCTCCG	
Ep-phbB-F	GATCGGGTCGACGAGCTC	
phbBInt-R	ACTCTGAATCCATCCTTAGC	
JM-sphbC-F1	AAAGGAGCAGCTGCATCTACTC	
JM-sphbC-F2	CTTCAGAACGAATTGAAGGTTCC	
JM-sphbC-R	GCTTTAACATATCTTCCTGGAGC	
Ep-phbC-F	CGGGTCGACCTGCAGTAG	
phbCInt-R	ACTGTCTTGACCATTCAAGC	
JM-Ep-For	CATGCGTCTATTAGTAGTGG	
JM-Ep-Rev	CATCACAACCTTGAACAAAGC	
BetaAct(RT)-F	TCCAAGGGGACCTAACGGAGA	
BetaAct(RT)-R	TGGGTCAAGAGCTGGATGGTG	
AMPSigF	ATGGGTTCTATGAGACTTCTTG	
AMPSigR	AACAAGCTTTTGCTTAGCATCC	
Ep-MsrA2vac-F	CGAGAAGCTTCTGCAGTAG	
MsrA2Int-R	CCTTTCCAGCATGAAGAGC	

used to introduce the ER-retention HDEL signal while amplifying *mGFP5*, also inserting a three amino acid linker peptide (Ala-Gly-Ala) between GFP and the HDEL signal. The PCR product was digested with *Bam*HI and *Xba*I and ligated to pBV-TS-SBP, which was also digested with *Bam*HI and *Xba*I, generating the eSBP-GFP construct (Figure 3.2C).

For the vSBP-GFP vector, *mGFP5* was amplified along with the sequence encoding the last four amino acids of mature SBP as well as the SBP C-terminal propeptide (nucleotides 985 to 1056), separated by a sequence encoding a linker peptide (Ala-Ser), from the PRX_CAMBIA1302 vector using the primers JM-mGFP5 For and JM-VS Rev. The PRX_CAMBIA1302 vector is pCAMBIA1302 modified such that the *mGFP5* gene is flanked by the sequence encoding the SBP N-terminal propeptide and the first four amino acids of the mature protein and the sequence encoding the last four amino acids of mature SBP followed by the C-terminal propeptide. The PCR products were digested with *Bam*HI and *Xba* and ligated with pBV-TS-SBP, which was also digested with *Bam*HI and *Xba*I, generating the vSBP-GFP construct (Figure 3.2D). The final constructs were sequenced to confirm that all elements were properly amplified and cloned.

2.2.3 Construction of EpC and EpW plasmids

Two plasmids were designed for the expression of the *Ep* gene in soybean plants. In both plasmids, the *Ep* gene is driven by the *Ep* 1.5 promoter. Both genes are also followed by the *nos* transcriptional terminator. One construct contains the *Ep* cDNA (herein referred to as *EpC*) while the other contains the *Ep* cDNA plus 3' untranslasted region (UTR) (herein referred to as *EpW*). For some genes, the 3'UTR is essential for high levels of gene expression. A comparison of expression levels in plants containing these two constructs will indicate if the 3'UTR is necessary for the high levels of *Ep* expression.

Initially plasmids were created such that the promoter driving the selectable marker and the *Ep* 1.5 promoter driving expression of the transgene were adjacent in a head-to-head configuration. However, to minimize interaction between the promoter driving expression of the selectable marker and the *Ep* 1.5 promoter, the vectors were redesigned so that the two promoters were in a head-to-tail configuration, placing the *Ep* 1.5 promoter next the right T-DNA border. In addition, the CaMV 35S promoter driving expression of the selectable marker was replaced with the *entCUP4* promoter. The steps taken to achieve this are described in the following section.

The EpC and EpW vectors for soybean transformation were created by Dr. L. Gudynaitę-Savitch. The *entCUP4* promoter and *hptII* gene were simultaneously amplified using an overlap extension technique. The tCUP-for and tCUP-rev primers were used to amplify the *entCUP4* promoter from the pUC19-tCUP plasmid. The Hygr-for and Hygr-rev primers were used to amplify *hptII* from pCAMBIA1301. The tCUP-rev primer contains a 5' extension that overlaps with the *hptII* gene and the Hygr-for primer contains a 5' extension that overlaps with the *entCUP4* promoter sequence. As a result, the PCR products annealed with each other and subsequent reactions amplified the joined sequences. The resulting *entCUP4* promoter-*hptII* gene PCR product was digested with *Bst*XI and *Xho*I and ligated with pCAMBIA1300, which was also digested with *Bst*XI and *Xho*I to generate the vector pCAMBIA1300-tCUP#1.

Both the *Ep* 1.5 promoter and the *Ep* 0.6 promoter (region 0.6 kb upstream of the *Ep* gene) were amplified with the Ep-1.5F and Ep-0.6F primers, respectively, and the Ep-Rev primer using the pPRX-1 plasmid as template. The pPRX-1 plasmid was generated by cloning an *Xba*I genomic fragment containing the *Ep* locus into pBLUESCRIPT KS(+) and was kindly provided by Dr. M. Gijzen. The resulting PCR products were digested with *Hin*dIII and *Nco*I and ligated

with pCAMBIA1301, which was also digested with *Hin*dIII and *Nco*I, to generate pCAMBIA1301-1.5Prx#2 and pCAMBIA1301-0.6Prx#2. In this plasmid, the *Ep* 1.5 and 0.6 promoters are driving expression of the *gusA* gene, which is followed by the *nos* transcriptional terminator.

The *Ep* 0.6 promoter, the *gusA* gene, and the *nos* transcriptional terminator were simultaneously amplified with the Ep-0.6F and NOS-Rev primers from the pCAMBIA1301-0.6Prx#2 plasmid. This PCR product was digested with *Hin*dIII and ligated with pCAMBIA1300-tCUP#1, which was digested with *Hin*dIII and *Sma*I, to generate pCAMBIA1300-tCUP0.6Prx#2. This step inverted the *Ep* 0.6 promoter-*gusA-nos* transcriptional terminator unit so that the *Ep* 0.6 promoter is placed farthest from the *entCUP4* promoter driving the *hptII* gene and was done to minimize the impact of the *entCUP4* promoter on the *Ep* 0.6 promoter.

The *Ep* 0.6 promoter was replaced in the pCAMBIA1300-tCUP0.6Prx#2 plasmid with the *Ep* 1.5 promoter from the pCAMBIA1301-1.5Prx#2 plasmid by digesting both plasmids with *Hin*dIII and *Nco*I, isolating the pCAMBIA1300-tCUP-0.6Prx#2 plasmid with the *Ep* 0.6 promoter removed and the *Ep* 1.5 promoter and ligating the two restriction products together to generate pCAMBIA1300-tCUP1.5Prx#6.

The *EpC* and *EpW* genes were amplified from the 5'PrxSoy03 plasmid. *EpC* was amplified using the EpcDNA-for and EpcDNA-Rstop primers and *EpW* was amplified using the EpcDNA-for and EpcDNA-Rwhole primers. Both PCR products were digested with *Nco*I and *Bst*EII and ligated with pCAMBIA1301-1.5Prx#2, which was also digested with *Nco*I and *Bst*EII, releasing the *gusA* gene, to generate pCAMBIA13Prx-EpCod#2 and pCAMBIA13Prx-EpWh#2, which contain *EpC* and *EpW*, respectively. These plasmids are orientated such that the

CaMV 35S promoter driving expression of *hptII* and the *Ep* 1.5 promoter are directly adjacent. In the subsequent steps, the *EpC* and *EpW* genes will be removed and placed into the pCAMBIA1300-tCUP1.5Prx#6 plasmid, which has the *Ep* 1.5 promoter placed farthest from the *entCUP4* promoter.

EpC and *EpW* were re-isolated by digesting the pCAMBIA13Prx-EpCod#2 and pCAMBIA13Prx-EpWh#2 plasmids with *Nco*I and *Bst*EII. They were ligated with the pCAMBIA1300-tCUP-1.5Prx#6 plasmid, which was also digested with *Nco*I and *Bst*EII, releasing the *gusA* gene, to generate pCAM1300tCUP-Prx-EpCod#3 (Figure 4.3A) and pCAM1300-tCUP-Prx-EpWh#4 (Figure 4.3B), containing *EpC* and *EpW* transgenes, respectively.

Both plasmids were sequenced to confirm that the various elements were properly amplified and cloned. Sequencing indicated that a single nucleotide error (A→G) was introduced at position -1349 in the *Ep* 1.5 promoter during amplification. However, promoter analysis work done by Drs. L. Gudynaitę-Savitch and M. Shukla indicated that this error did not affect promoter function.

2.2.4. Construction of the PHB plasmid

In order to metabolically engineer soybean seed coats to produce the biodegradable plastic PHB, three novel enzymes must be expressed in this tissue. These are the *phbA*, *phbB*, and *phbC* genes from *Ralstonia eutropha* (Peoples and Sinskey 1989a; 1989b). All three genes were included in a single cassette, each driven independently by a copy of the *Ep* 1.5 promoter and followed by a *nos* transcriptional terminator. The three *phb* genes were codon optimized prior to being synthesized in order to maximize expression of genes of bacterial origin in soybean. The *phb* genes were also modified to include the sequence encoding the SBP

propeptides separated by a small linker peptide in order to direct the enzymes to the seed coat vacuole.

The *entCUP4* promoter driving expression of the selectable marker was separated from the first *Ep* 1.5 promoter by an approximately 2.5 kb fragment of the LacZ gene in order to minimize promoter interactions between adjacent promoters. All three *phb* expression cassettes were in the same orientation with the promoters being closest to the selectable marker and the *phb* genes being closest to the right border of the T-DNA. This orientation minimizes the influence of adjacent genomic DNA on the *Ep* 1.5 promoters.

The *Ep* 1.5 promoter was amplified from the Prx_promoter (in pUC18) synthesized by the Plant Biotechnology Institute using primers Prx-F and Prx-R. The resulting PCR product was digested with *Sst*I and *Bam*HI and ligated with pUC19 plasmid (Yanisch-Perron et al. 1985), which was also digested with *Sst*I and *Bam*HI to generate the vector pUC19-Prx.

The *nos* terminator sequence was amplified from the plasmid pCAMBIA1302 by PCR using primers JM-nos-F2 and JM-nos-R2. The resulting PCR product was digested with *Xba*I and *Hin*dIII and ligated with pUC19-Prx, which was also digested with *Xba*I and *Bam*HI to generate the vector pUC19-Prx-nos.

The PRX_CAMBIA1302 plasmid (described in Section 2.2.2) was digested with *Nco*I and *Bst*EII and the gene encoding GFP with the N- and C-terminal propeptides from SBP was isolated. This was ligated to the pUC19-Prx-nos plasmid, which was also digested with *Nco*I and *Bst*EII, generating the plasmid pUC19-PGn. The 1302-tCZ-Prx plasmid was digested with *Pst*I and the 2.5 kb LacZ fragment was isolated. The LacZ fragment was ligated to the pUC19-PGn plasmid, generating pUC19-LPGn.

The pUC19-Prx-nos vector was digested with *Sal*I, and the *Ep* 1.5 promoter-*nos* transcriptional terminator insert was isolated, blunt-ended using the Klenow fragment, and digested with *Pst*I. The plasmid pCAMBIA1300-tCUP#1 was digested with *Hin*dIII, blunt-ended with the Klenow fragment, and digested with *Pst*I. pCAMBIA1300-tCUP#1 is derived from pCAMBIA1300 in which the CaMV35S promoter driving expression of the *hptII* gene has been replaced by the enhanced cryptic tobacco promoter *entCUP4* (see Section 2.2.3). The *Ep* 1.5 promoter-*nos* transcriptional terminator insert was then ligated to pCAMBIA1300-tCUP plasmid to generate p1300t-Prx-nos.

The *phbA* and *phbC* genes were codon-optimized for soybean and synthesized by GenScript Corp. The *phbB* gene was codon-optimized for soybean and synthesized by the Plant Biotechnology Institute. Each synthetic genes consisted of the first 30 codons of the *Ep* gene (encoding the N-terminal propeptide and the first four amino acids of the mature protein) then a sequence encoding the linker Gly-Ala-Gly, the *phb* gene, a sequence encoding the linker Gly-Ala-Gly and finally the last 24 codons of the *Ep* gene (encoding the last four amino acids of the mature protein and the C-terminal propeptide). *Nco*I sites were included at the start codons and *Bst*EII sites were included 3' to the stop codons.

The *phbA*, *phbB*, and *phbC* genes were digested with *Nco*I and *Bst*EII and ligated with the pUC19-Lac-Prx-nos, pUC19-Prx-nos, and p1300t-Prx-nos plasmids, respectively, which were also digested with *Nco*I and *Bst*EII. This generated pUC19-LPAn, pUC19-PBn, and p1300t-PCn, respectively.

The pUC19-PBn plasmid was digested with *Sal*I and *Sma*I, releasing a fragment containing the *Ep* 1.5 promoter, the *phbB* gene, and the *nos* transcriptional terminator. This fragment was isolated and ligated to p1300t-PCn, which was also digested with *Sal*I and *Sma*I,

generating p1300t-PBn-PCn. The pUC19-LPAn plasmid was digested with *Sma*I, releasing a fragment containing the LacZ fragment, the *Ep* 1.5 promoter, the *phbA* gene, and the *nos* transcriptional terminator. This fragment was isolated and ligated to p1300t-PBn-PCn, which was also digested with *Sma*I, generating the J13 plasmid (Figure 5.1). This plasmid was checked by sequencing to confirm that all the elements were properly amplified and cloned.

2.3 Plant transformation

Arabidopsis thaliana transformation was modified from the methods of Clough and Bent (1998). *A. thaliana* (L.) Heynh. (ecotype Columbia) seeds were grown in soil at 22°C with 16 h light and 8 h dark until flowers on primary inflorescences were beginning to open. In some cases, bolts were cut back and allowed to grow back to this stage. Any fully open flowers as well as all siliques were removed immediately prior to transformation. *Agrobacterium tumefaciens* GV3101 was transformed with plasmid DNA by electroporation. Transformed *Agrobacterium* were suspended in infiltration medium (2.2 g/L MS salts, 50 g/L sucrose, 500 µl/L Silwet L-77). *Arabidopsis* shoots were then submerged in the infiltration medium for approximately 30 seconds. Plants were grown to maturity and seeds were collected representing the T_1 generation.

Soybean (*Glycine max* (L.) Merrill cv. Jack) plants were transformed by particle bombardment of proliferating somatic embryos generated from immature cotyledons according to the method of Trick et al. (1997). All transgenic soybean plants were generated in the laboratory of D. Brown, AAFC, London, ON.

2.4 Growth of plants

Arabidopsis seeds were sterilized by washing in 70% ethanol for 10 minutes followed by 50% bleach for 10 minutes and then rinsed four times in sterile water. Seeds were spread on

selection plants (1/2 MS salts, pH ~5.7, 1% sucrose, 0.8% agar, 30 mg/L hygromycin, 250 mg/L timentin) and grown at 22°C with a 12 h light/12 h dark or a 16 h light/8 h dark photoperiod. After two weeks, plants were transferred to soil and grown to maturity under the same conditions. Plants were watered once a week with modified Hoagland's solution (2 mM $MgSO_4$, 4 mM KH_2PO_4, 10 mM KNO_3, 10 mM $Ca(NO_3)$, 19.9 μM EDTA disodium salt, 18.0 μM $FeSO_4 \cdot 7H_2O$, 92.5 μM H_3BO_3, 28.9 μM $MnSO_4 \cdot H_2O$, 1.6 μM $ZnCl_2$, 0.6 μM $CuSO_4$, 0.2 μM $Na_2MoO_4 \cdot 2H_2O$). Rosette leaf tissue was collected from *Arabidopsis* plants at maturity for analysis. Tissue was frozen in liquid nitrogen and stored at -80° until analyzed.

Soybean seeds were sterilized by washing two times in 25% bleach for 5 min followed by two washes in sterile water for 5 min. Soybean seeds were germinated in moistened vermiculite or on moistened filter paper. In some cases, if seed coats did not soften, they were punctured with the tip of a scalpel to allow water uptake. After four to six days, seed coats were collected and seeds were transferred to soil. Seed coats were frozen in liquid nitrogen and stored at -80°C until analyzed. Soybean plants were grown at 25°C with a 16 h light/8 h dark photoperiod with the exception of the plants used for the developmental analysis of SBP accumulation, which were grown at 26°C during the day and 22°C at night with a 12 h photoperiod. In some cases, the photoperiod was reduced to 14 h light/10 h dark to induce flowering. Plants were watered once a week with 20-20-20 fertilizer at a concentration of approximately 3 g/L. Young leaves (approximately 2 weeks) were collected for molecular analysis as were 21 DPA pods, seed coats, and embryos. All tissues were frozen in liquid nitrogen and stored at -80°C until analyzed.

Due to the complex genotype of soybean seeds, it is important to note that the genotype of the embryo is different from the parent plant since it represents the next generation and will have undergone segregation (Figure 2.1). Therefore, a T_n plant produces leaves, pods, and seed

coats[3] with an identical genotype (T_n generation) but the embryo is part of the T_{n+1} generation and therefore likely has a different genotype. Also, the term embryo is used to refer to the seed without its seed coat and includes both the embryo, consisting of the radicle and cotyledons, and the endosperm, which differs genetically from the embryo since it is triploid.

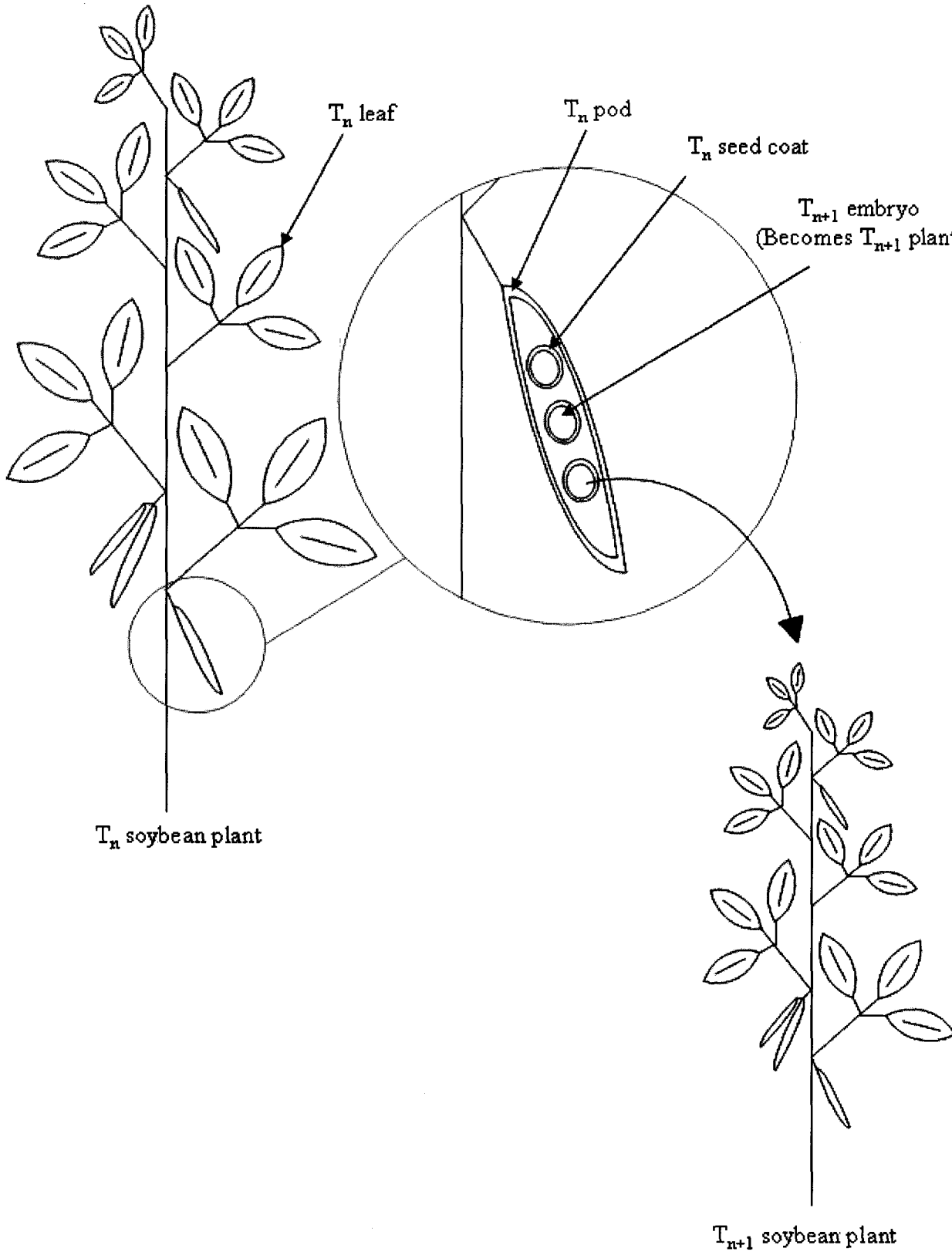
2.5 PCR analysis of transgenic plants

Genomic DNA was extracted and analyzed by PCR to confirm the presence of the T-DNA. For *Arabidopsis* plants, a leaf disc was ground in 400 µl of extraction buffer (200 mM Tris, pH 7.5, 250 mM NaCl, 25 mM EDTA, 0.5% SDS) in an Eppendorf tube with a pestle, then centrifuged at 16,100 g for 1 min. DNA in the supernatant was precipitated with an equal volume of 2-propanol and collected by centrifugation at 16,100 g for 10 minutes. The pellet was washed with 500 µl of 70% ethanol and then pelleted again by centrifugation at 16,100 g for 5 min. The ethanol was removed and the DNA pellet was dried at 37°C for 30 minutes. The DNA was then solubilised in 100 µl of water.

Since *Arabidopsis* plants were screened for the presence of the transgene by their ability to grow on hygromycin, the *hptII* gene is presumed to be present. Therefore, the *hptII* gene was amplified to test the integrity of the genomic DNA extract, using the Hygr-for/Hygr-rev primers. To confirm the presence of the remainder of the T-DNA, the *mGFP5* gene was also amplified using the JM-mGFP5 For/JM-mGFP5 Rev primers. PCR was performed in 20 µl containing 1xPCR reaction buffer, 2 mM $MgSO_4$, 0.2 mM dNTPs, 0.5 µM of each primer, 1 U of High Purity Taq DNA Polymerase, and 2 µl of genomic DNA extract. PCR cycling was as follows: initial DNA denaturation at 94°C for 2 min, followed by 35 cycles of denaturation for 15 s at

[3] With the exception of the innermost layer which is composed of aleurone cells that originate from the outermost layer of the endosperm and are therefore not maternal in origin (Miller et al. 1999).

Figure 2.1 Schematic diagram illustrating the relationship between the genotype of different tissues of a soybean plant and between plants from different generations (Drawn by J. Schnell). T_n refers to the generation of the transgenic plant line, where T_0 represents the plant that was transformed, T_1 represents the progeny of the T_0 plant, and so on.



94°C, 15 s at the annealing temperature of the primer pair, and then elongation for 1 min at 72°C, followed by a final elongation at 72°C for 10 min.

Genomic DNA was isolated from soybean leaf disks using the REDExtract-N-Amp Plant PCR Kit or the DNeasy Plant Mini Kit. For DNA isolated using the REDExtract-N-AMP Plant PCR Kit, PCR reactions were performed in 10 µl containing 5 µl of Ready Mix, 0.2 to 0.5 µM of each primer, and 2 µl of genomic DNA extract. For DNA isolated using the DNeasy Plant Mini Kit, PCR reactions were performed in 20 µl containing 1xPCR reaction buffer, 2 mM $MgSO_4$, 0.2 mM dNTPs, 0.2 to 0.5 µM of each primer, 0.05 U of Taq DNA polymerase, and 0.5 to 1 µl of genomic DNA extract. The endogenous *β-tubulin-2* gene was amplified to check the quality of the genomic DNA extract using the GmTub2For and GmTub2Rev primers.

For all transgenic soybean plants, the *hptII* gene was amplified using the HptII (99) For and HptII (614) Rev primers. The *entCUP4* promoter driving expression of the *hptII* gene was amplified using the tCUP (20) For and tCUP (432) Rev primers. The *nos* terminator was amplified using the nos (13) For and nos (192) Rev primers.

For soybean plants transformed with the EpC and EpW constructs, the *EpC* and *EpW* genes were amplified with the EpcDNA-F and EpcDNA-Rstop primers. Alternatively, to specifically amplify the *EpW* gene, the EpcDNA-F and EpcDNA-Rwhole primers were used. The *Ep* 1.5 promoter driving expression of the *EpC* and *EpW* genes was amplified with the Ep1.5F3.2 and Ep-ori-Rev primers. The EpcDNA-F and the Ep-ori-Rev primers are located in the region of the *Ep* gene that is deleted in cv. Jack and specifically detect the transgene and not the *ep* allele found in cv. Jack genomic DNA.

For soybean plants transformed with the PHB construct, the *phbA* gene was amplified using the sphbA F1 and sphbA R primers and the *Ep* 1.5 promoter driving expression of the

phbA gene was amplified using the Ep-phbA-F and phbAInt-R primers. The *phbB* gene was amplified using the sphbB F1 and sphbB R primers and the *Ep* 1.5 promoter driving expression of the *phbB* gene was amplified using the Ep-phbB-F and phbBInt-R primers. The *phbC* gene was amplified using the sphbC F1 and sphbC R primers or the sphbC F2 and sphbC R primers and the *Ep* 1.5 promoter driving expression of the *phbC* gene was amplified using the Ep-phbC-F and phbCInt-R primers. The primers that amplify the *Ep* 1.5 promoters for each of the *phb* genes were designed with a forward primer that overlaps both plasmid sequence and the *Ep* 1.5 promoter and a reverse primer located in the *phb* gene in order that amplification of the *Ep* 1.5 promoter for each of the three *phb* gene cassettes was specific. However, the amplification of *nos* transcriptional terminator was not specific for the three separate genes.

In some cases, the *Ep* 1.5 promoters for the *phb* construct were amplified with the Ep-0.3F and Ep-ori-Rev primers, which do not distinguish between the three *Ep* 1.5 promoters driving expression of the three *phb* genes but are specific for the transgene since the Ep-ori-Rev primer is located in the region of the *Ep* gene that is deleted in cv. Jack.

PCR cycling was as follows: initial DNA denaturation at 94°C for 2 min, followed by 30 cycles of 30 s at 94°C for denaturation, 30 s at the annealing temperature of the primer pair, and then 30 s to 1.5 min of elongation at 72°C, depending on the length of the product, followed by a final elongation at 72°C for 10 min.

A list of primer names and sequences can be found in Table 2.1 and an overview of the primers used to amplify the various transgene elements as well as the PCR conditions used for each primer pair can be found in Table 2.2.

2.6 Confocal scanning laser microscopy

A Zeiss LSM 510 META laser scanning microscope with a 63x C-Apochromat objective (Jena, Germany) was used for GFP imaging of *Arabidopsis* leaves. Approximately 1 cm^2 leaf sections were cut from fully expanded leaves and mounted in water. GFP was excited with a 488 nm argon laser and detected using a 505/530-nm bandpass filter.

2.7 Plant subcellular fractionation

All subcellular fractionation work was done on rosette leaf tissue from 5 week old *Arabidopsis* plants grown with a 12 h day/12 h night photoperiod.

2.7.1 Isolation of leaf apoplast fraction

Isolation of leaf apoplast fraction was performed as in Firek et al. (1993) with minor modifications. Between 1 and 3 g of *Arabidopsis* leaves were removed and submerged in infiltration buffer (0.2 M $CaCl_2$, 0.3 M mannitol, pH 6.8). Vacuum (approximately 300 mbar) was applied for 5 min. Leaves were blotted dry and then wrapped in parafilm. The parafilm-wrapped leaves were then placed into the barrel of a 10 ml syringe, which was placed inside of a 50 ml centrifuge tube. This was centrifuged for 15 min at 500 g (1800 rpm) at 20°C in a Sorvall RC5C centrifuge with an SA-600 rotor. The flow-through was collected and mixed with 3 µl of plant protease inhibitor cocktail and represents the leaf apoplast fraction. The remaining tissue was homogenized in 0.067 M phosphate buffer (8.5 mM Na_2HPO_4, 58.8 mM KH_2PO_4), pH 6.0, with plant protease inhibitor cocktail at 3 µl per 100 mg of fresh plant tissue. This was centrifuged at approximately 15,000 g (10,200 rpm) for 15 min. The supernatant was collected and represents the intracellular fraction.

Table 2.2 PCR and RT-PCR conditions used for the amplification of T-DNA elements and mRNAs and expected amplicon sizes[1].

Application	Target Sequence	Primers	Size (bp)[2]	T_M (°C)	Extension (mm:ss)	Cycles
PCR	*hptII*	Hygr-for/Hygr-rev	1072	57	01:00	35
PCR	*mGFP5*	JM-mGFP5 For/JM-mGFP5 Rev	720	61	01:00	35
PCR	*β-tubulin-2*	GmTub2 F/GmTub2 R	430	56	00:30	30
PCR	*entCUP4*	tCUP (20) For/ tCUP (432) Rev	413	65/57.6[3]	00:30	30
PCR	*hptII*	HptII (99) For/ HptII (614) Rev	516	63.2	00:30	30
PCR	*nos*	nos (13) For/nos (192) Rev	180	60.5	00:30	30
PCR	*EpC*	EpcDNA-F/EpcDNA-Rstop	1069	65	01:00	30
PCR	*EpW*	EpcDNA-F/EpcDNA-Rwhole	1216	64	01:00	30
PCR	*Ep* 1.5 promoter	Ep1.5F3.2/Ep-ori-Rev	1543	60	01:30	30
PCR	*Ep* 1.5 promoter	Ep-phbA-F/phbAInt-R	1727	59	01:30	30
PCR	*Ep* 1.5 promoter	Ep-phbB-F/phbBInt-R	1751	62	01:30	30
PCR	*Ep* 1.5 promoter	Ep-phbC-F/phbCInt-R	1757	62	01:30	30
PCR	*Ep* 1.5 promoter	Ep-0.3F/Ep-ori-Rev	324	59	00:30	30
PCR	*phbA*	sphbA F1/sphbA R	1175	60	01:00	30
PCR	*phbB*	sphbB F1/sphbB R	732	61.5	00:30	30
PCR	*phbC*	sphbC F1/sphbC R	1748	62	1:30	30
PCR	*phbC*	sphbC F2/sphbC R	425	61.5	00:30	30
RT-PCR	*β-Actin*	BetaAct(RT)-F/ BetaAct(RT)-R	179	60	01:00	25-40
RT-PCR	*Ep* gene	JM-Ep-For/ JM-Ep-Rev	221	60	01:00	25-40
PCR	*MsrA2vac*	AMPSigF/AMPSigR	222	61	00:30	30
PCR	*Ep* 1.5 promoter	Ep-MsrA2vac-F/MsrA2Int-R	1689	62	01:30	30
PCR	*Ep* 1.5 promoter	Ep-0.3F/Ep-ori-Rev	324	59	00:30	30

[1]The table represents the combined efforts Dr. Loreta Gudynaitę-Savitch, Shuyou Han and Jaimie Schnell.

[2]For each primer pair the T_M was estimated by gradient PCR.

[3]For these primers, the lower annealing temperature was required when amplification was done with the REDExtract-N-AMP Plant PCR Kit.

2.7.2 Isolation of vacuoles

Isolation of vacuoles was performed as in Matsui et al. (2006) with minor modifications. Approximately 2 mg of leaf tissue was collected and cut into 2 mm strips. Leaves were submerged in 30 ml of protoplast enzyme solution (10 mg/ml cellulose Onozuka R-10, 10 mg/ml Maceroenzyme R-10, 4 mg/ml $CaCl_2 \cdot 2H_2O$, 0.4 M mannitol, and 10 mM MES, pH 5.6) and vacuum was applied for 10 minutes. Leaves and protoplast enzyme solution were then shaken together at 70 rpm for 4 h at room temperature in the dark. After 4 h, the protoplasts were strained through a 150 mesh sieve and pelleted at approximately 80 g (800 rpm) in a Sorvall RC5C centrifuge with an SA-600 rotor. Protoplasts were washed twice with wash buffer (0.4 M mannitol, 10 mM MES, pH 5.6) and then resuspended in wash buffer and layered onto 12% Ficoll. This was centrifuged at approximately 71,000 g (24,000 rpm) for 1 h at 4°C in a Beckman Optima XL-100K ultracentrifuge with a SW28 rotor. The upper layer was collected and represents the vacuole fraction.

2.7.3 Isolation of microsomal and cytoplasmic fractions

Isolation of microsomal and cytoplasmic fractions was performed as in Klahre et al. (1998) with minor modifications. Leaf tissue was homogenized in extraction buffer (10 mM HEPES-KOH, pH 7.1, 12.5% (w/v) sucrose, 10 mM potassium acetate, 3 mM $MgCl_2$) by grinding in a mortar and pestle. The homogenate was centrifuged at approximately 1000 g (2600 rpm) for 10 min at 4°C in a Sorvall RC5C centrifuge with an SA-600 rotor. The supernatant was collected and centrifuged at approximately 100,000 g (33,000 rpm) for 1 h at 4°C in a Beckman Optima XL-100K ultracentrifuge with a SW55 Ti rotor. The supernatant was collected and represents the cytoplasmic fraction. The pellet represents the microsomal fraction.

2.8 Protein extraction

Plant tissues were ground to a fine powder in liquid nitrogen and then fully homogenized in extraction buffer (0.067 M phosphate buffer [8.5 mM Na_2HPO_4, 58.8 mM KH_2PO_4], pH 6.0), plus 3 µl/100 mg fresh tissue plant protease inhibitor cocktail. Samples were centrifuged at 16,100 g for 15 minutes and the supernatant was collected and represents crude protein extract. Protein concentration was estimated by the Bradford method (Bradford 1976) using the Coomassie (Bradford) Protein Assay Kit with bovine serum albumin (BSA) as the standard.

2.9 Measurement of peroxidase specific activity

The peroxidase specific activity of plant extracts was determined by measuring the rate of oxidation of ABTS based on the rate of change of absorption at A_{405} over 3 to 5 minutes (modified from Pütter and Becker 1983). Absorption measurements were done on a Thermo Spectronic BioMate 3 spectrophotometer with a 10 mm path length. The assay contained a final concentration of 1.7 mM ABTS and 0.8 mM H_2O_2 in 0.067 M phosphate buffer (8.5 mM Na_2HPO_4, 58.8 mM KH_2PO_4), pH 6.0, and was carried out at 25°C. The following equation was used to convert the change in absorbance into specific activity:

$$\text{Specific Activity (U mg}^{-1}\text{ protein)} = \frac{\Delta A_{405} \bullet V_A \bullet df}{\varepsilon \bullet d \bullet \Delta t \bullet v_E \bullet \rho_{protein} \bullet 2}$$

where ΔA_{405} is the change in absorbance measured at 405 nm, V_A is the final volume of the assay in litres, df is the dilution factor of the enzyme sample, ε is the linear millimolar absorption coefficient, which for ABTS under the above conditions, is 537.6 L µmol^{-1} mm^{-1} (Pütter and Becker 1983), d is the path length in milliliters, Δt is the length of time over which the A_{405} was recorded in minutes, and $\rho_{protein}$ is the mass concentration of protein in the extracts in mg L^{-1},

which represents the mass of protein divided by the volume of the extract. The denominator is multiplied by a factor of two because one mole of H_2O_2 yields two moles of oxidized ABTS.

2.10 Antibodies

The plasmid pETSBP, which contains the region of the *Ep* gene encoding the mature protein cloned into the *Nde*I and *Bam*HI sites of the pET-3a vector (Henrickson et al. 2001), was kindly provided by Drs. K. Welinder and K. Lehmann Nielsen. This plasmid was transformed into *E. coli* and mature SBP was recovered as an inclusion body, purified from SDS-PAGE gels, and injected into rabbits to generate a rabbit polyclonal anti-SBP antibody (kindly provided by M.C. Romero, AAFC, London, ON). It was used at a dilution of 1:500. Rabbit polyclonal anti-GFP (FL) antibody and rabbit polyclonal anti-BiP (at-95) antibodies were purchased from Santa Cruz Biotechnology, Inc. Anti-GFP was used at a dilution of 1:500 and anti-BiP at a dilution of 1:200. Rabbit polyclonal anti-cFBPase (cytosolic fructose-1,6 bisphosphatase) antibody and rabbit polyclonal anti-V-ATPase (H^+-ATPase) antibody were purchased from Agrisera and used at dilutions of 1:3000 and 1:1000 respectively. Affinity purified goat anti-rabbit IgG (H+L) alkaline phosphatase conjugate was purchased from Bio-Rad and used at a dilution of 1:3000.

2.11 Immunoblot analysis

SDS-PAGE was performed according to the method of Laemmli (1970). Samples were mixed with SDS-PAGE gel loading buffer (62.5 mM Tris-HCl, pH 6.8, 25% glycerol, 2% SDS, 0.01% bromophenol blue, 5% β-mercaptoethanol) and boiled for 5 min prior to loading on the gel. Typically, between 10 and 50 μg of protein were loaded per well onto 8 to 15% SDS-PAGE

gels. Recombinant GFP purified from transformed *E. coli* (Roche) and SBP purified from soybean seed coats (Sigma) were used as positive controls.

Some samples with low protein concentration were first precipitated by the addition of 100% TCA at a ratio of 4 volumes of sample to 1 volume of TCA. Protein was pelleted by spinning at maximum speed in a microcentrifuge for 5 min. The pellet was washed twice with 200 µl of cold acetone. Residual acetone was evaporated by incubating the sample at 95°C for approximately 5 min. The pellet was then resuspended in half-strength loading buffer, followed by boiling for 5 min prior to loading the sample on the gel.

Samples were separated by electrophoresis at 200 V for approximately 40 minutes using the Mini-Protean 3 system (Bio-Rad). Protein was transferred to PVDF at 100 V for 1 h in western transfer buffer (25 mM Tris, 192 mM glycine, pH 8.3), using the Mini-Transblot system (Bio-Rad). Efficiency of transfer and equal well loading were confirmed by staining the blot with Ponceau S, followed by complete destaining in water. Blots were blocked overnight in 5% non-fat milk at 4°C. Blots were probed with primary antibody for 2 h followed by secondary antibody for 1 h, both at room temperature. Blots were developed by overlaying them with NBT and BCIP diluted in alkaline phosphatase reaction buffer (100 mM Tris-HCl, 100 mM NaCl, 5 mM $MgCl_2$, pH 9.5) to final concentrations of 330 and 165 µg/ml, respectively.

2.12 RNA extraction and RT-PCR

Between 40 and 100 mg of plant tissue was ground to a fine powder under liquid nitrogen. RNA was extracted using the RNeasy Plant Mini Kit. RNA was quantified by measuring the absorbance at 260 nm. Contaminating genomic DNA was removed by treatment with RNase-free DNaseI. A typical reaction was performed in 10 µl and contained up to 1 µg of

RNA, 1x reaction buffer (10 mM Tris-HCl, pH 7.5, 2.5 mM $MgCl_2$, 0.1 mM $CaCl_2$) and 1 U of DNase I. The reaction was incubated at room temperature for 15 minutes. To inactivate the DNase I, 1 µl of 25 mM EDTA (final concentration of 2.27 mM) was added and the sample was heated at 65°C for 10 min.

RT-PCR was performed using the OneStep RT-PCR kit. Both primers specific for the gene of interest and internal control primers were included in the reaction. Reactions were performed in 25 µl containing 1x Qiagen OneStep RT-PCR buffer, 0.4 mM dNTPs, 0.6 µM of each of the gene-specific primers, 0.3 µM of each of the internal control primers, 2 µl of Qiagen OneStep RT-PCR enzyme mix, and 0.1 to 0.2 µg of template RNA.

Cycling was as follows: reverse transcription for 30 min at 50°C followed by an initial PCR activation step at 95°C for 15 min. Twenty-five to 40 cycles of denaturation at 94°C for 30 s, annealing at the annealing temperature of the primer pairs for 30 s, and extension at 72°C for 1 min were performed followed by a final extension at 72°C for 10 min.

For every sample, a duplicate reaction was prepared as a control for genomic DNA contamination. This reaction was kept on ice until the beginning of the PCR activation step. Since no reverse transcription is performed for these samples, only contaminating genomic DNA should be amplified, if present.

β-actin (BU082893) transcript was amplified as an internal control for all RT-PCR reactions using the BetaAct(RT)-F and BetaAct(RT)-R primers. *Ep* transcript was amplified using the JM-Ep-For and JM-Ep-Rev primers. The JM-Ep-For primer is located in the region of the *Ep* gene that is deleted in Jack so these primers should specifically amplify the *Ep* transcript originating from the transgene. The sequences of these primers can be found in Table 2.1 and the conditions used to amplify each transcript can be found in Table 2.2.

2.13 Extraction of PHB and analysis by gas chromatography

Single seed coats (approximately 10 mg) were dried at 50°C for 1 to 2 h and then ground to a fine powder under liquid nitrogen. To extract PHB, 500 µl of chloroform was added. In addition, benzoic acid, dissolved in ethanol, was added to a final concentration of 0.18 mg/ml to serve as an internal standard. The sample was incubated at 55°C for 1 h, and then 1.7 ml of ethanol and 0.2 ml of HCl were added and the sample was incubated for 4 h in a boiling water bath, resulting in the formation of the ethyl ester of 3-hydroxybutyric acid (3-HB). Following this, 4 ml of 0.9 M NaCl was added and the solution was mixed. After the aqueous and organic phases were separated, the organic phase was collected. Free hydroxyl groups were derivatized by mixing 50 µl of the organic phase with 150 µl of N-methyl N-trimethylsilyltrifluoroacetamide and then incubating the sample at room temperature for at least 1 h.

Samples were analyzed with an Agilent 7890A gas chromatograph equipped with a flame ionization detector, the Injector Agilent 7683B Series, and an HP-5 capillary column (Agilent 19091J-413, 30 m x 320 µm x 0.25 µm). The sample (1 µl) was injected with an injection split ratio of 50:1. The oven temperature was set at 100°C for 3 min and then increased at a rate of 20°C/min until a temperature of 120°C was reached after which it was increased at a rate of 30°C/min until 300°C was reached. The carrier gas used was hydrogen at a flow rate of 40 ml/min. Commercial PHB was used as a positive control. Samples were quantified by relating the height of the peak representing the response of the flame ionization detector (in picoamperes) to the concentration of PHB using a standard curve prepared with commercial PHB at initial concentrations of 0.46 and 0.09 mg/ml, according to the internal standard method of Chang and Karr (1959). In all cases, samples were normalized by using the ratio of the height of the peak

corresponding to 3-HB to the height of the peak corresponding to benzoic acid. Final amounts were expressed as percentage of seed coat DW.

2.14 Statistical analyses

Data were statistically analyzed using S-PLUS version 8.0. To test the difference between two means, a t-test was performed. For multiple means, one-way ANOVA was performed followed by a post-hoc analysis using the 95% simultaneous confidence intervals for specified linear combinations by the Bonferroni method (Bland and Altman 1995). In cases where data were not normally distributed, they were first log transformed.

CHAPTER 3 RELATIONSHIP BETWEEN PROTEIN TARGETING AND PROTEIN ACCUMULATION

3.1 Protein targeting in *Arabidopsis thaliana*: testing the putative SBP targeting signals

Several studies have shown that targeting to different subcellular locations can increase the recovery of foreign proteins in plant systems (Hood et al. 2003; Ramírez et al. 2002; Yang et al. 2005). In order to increase the yield of bioproducts in soybean seed coats, we attempted to target foreign proteins to the vacuole using the N- and C-terminal propeptides from SBP that are predicted to direct vacuolar localization. To confirm the function of the SBP propeptides and to investigate the influence of these propeptides on protein yield, the *Ep* gene was expressed in *Arabidopsis thaliana* in frame with *mGFP5*, encoding GFP, with different combinations of the sequences encoding the SBP N- and C-terminal propeptides and an ER-retention signal. The resulting fusion proteins contain both SBP, whose activity can be measured as an indication of protein yield, and GFP, which can be imaged by confocal microscopy to confirm subcellular localization.

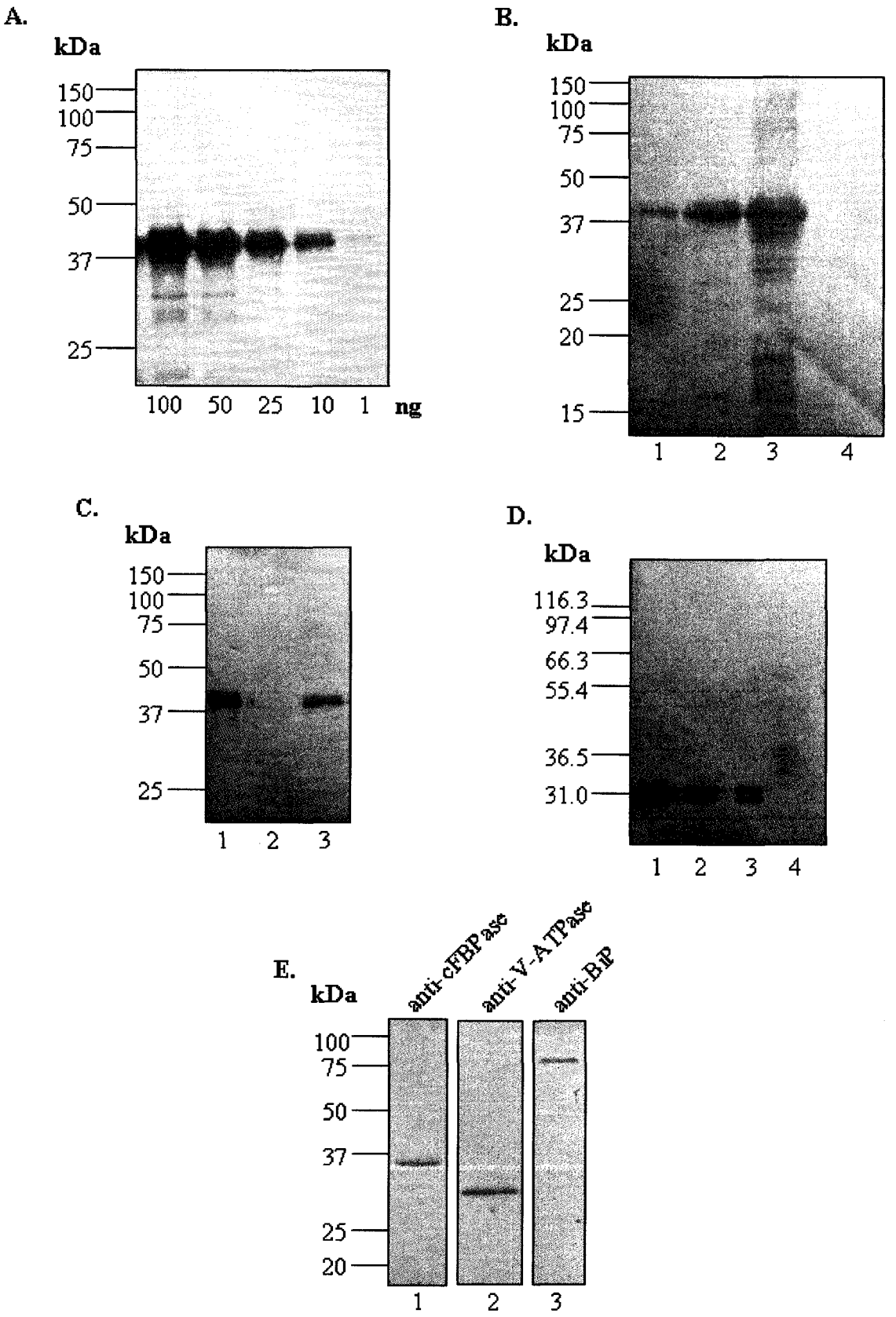
3.2 Characterization of anti-SBP, anti-GFP, anti-cFBPase, anti-V-ATPase, and anti-BiP antibodies used for protein immunoblots

A polyclonal antibody prepared from rabbits injected with SBP (prepared in *E. coli*) was obtained from M.C. Romero (AAFC, London, ON). We initially tested the sensitivity and specificity of the antiserum. The anti-SBP antibody was able to detect as low as 1 ng of purified SBP (Figure 3.1A) when the colorimetric assay was used. In seed coat extracts from the *EpEp* soybean cv. Harosoy 63, which contains abundant SBP, the anti-SBP antibody specifically recognizes a protein (Figure 3.1B lanes 2 and 3) of 43 kDa, the same size as the purified SBP

(Gijzen et al. 1993). In Figures 1A and 1B additional bands of lower molecular weight can be seen when more protein is loaded into the well. These may represent breakdown products, since they are absent in the *epep* cv. Jack, which does not contain SBP. There are also bands of unknown identity with a higher molecular weight than SBP. These are absent in cv. Jack extracts, suggesting they are specific to cv. Harosoy 63. In seed coat extracts from cv. Jack, the antibody does not detect any proteins (Figure 3.1B lane 4), nor does it detect any proteins in 50 μg of protein extract from 21 DPA cv. Jack pod or embryo tissues, nor from 2-week-old leaf tissue (data not shown), indicating that it does not cross-react with other peroxidases that can be found in these tissues (Chen and Vierling 2000; Caldwell et al. 1998; Diehn et al. 1993). Similarly, in leaf extracts from wild-type *Arabidopsis* plants, the antibody does not detect any protein while in a leaf extract from a transgenic *Arabidopsis* plant carrying the *vSBP-GFP* gene, which encodes a fusion between the *Ep* gene encoding soybean peroxidase and the *mGFP5* gene encoding GFP, a single band the same size as the purified SBP is detected (Figure 3.1C). As will be discussed below, this is significantly smaller than the expected size for the fusion protein. These results indicate that the anti-SBP antibody is able to detect as little as 1 ng of SBP and it is specific for SBP in soybean tissues and *Arabidopsis* leaves, which suggests that this antibody will be useful in the analysis of transgenic plants expressing SBP.

The commercial anti-GFP was raised against amino acids 1 to 238 of GFP, representing the full-length protein of *Aequorea victoria* origin. The anti-GFP antibody is able to detect 50 ng of recombinant GFP with the expected size of 27 kDa (Figure 3.1D, lane 2). GFP occasionally migrated as a double band, which has been documented previously and may depend upon sample preparation (Gustafson et al. 1998; Batoko et al. 2000). Anti-GFP does not detect any bands in

Figure 3.1 Immunoblot analysis demonstrating the specificity and sensitivity of anti-SBP, anti-GFP, anti-cFBPase, anti-V-ATPase, and anti-BiP, as described in Section 2.11. A colorimetric assay was used to develop the immunoblots using the reagents NBT and BCIP. **A.** Reaction of anti-SBP antibody (diluted at 1:500) to different amounts of purified soybean peroxidase (10% acrylamide gel). **B.** Reaction of anti-SBP antibody (diluted at 1:500) to soybean seed coat extracts (12% acrylamide gel). Lane 1 contains 42 ng of purified SBP. Lanes 2 and 3 contain 10 μg of protein extracted from 45 DPA and mature Harosoy 63 seed coats, respectively. Lane 4 contains 10 μg of protein from a mature Jack seed coat. **C.** Reaction of anti-SBP antibody (diluted at 1:500) to *Arabidopsis* leaf extracts (10% acrylamide gel). Lane 1 contains 42 ng of purified SBP. Lanes 2 and 3 contain 30 μg of protein from *Arabidopsis* leaf extracts from a wild-type and a transgenic plant containing the *vSBP-GFP* gene, respectively. **D.** Reaction of anti-GFP antibody (diluted to 1:500) to purified recombinant GFP and an *Arabidopsis* leaf extract (10% acrylamide gel). Lanes 1 to 3 contain 300, 100, and 50 ng of recombinant GFP, respectively. Lane 4 contains 30 μg of protein extracts from wild-type *Arabidopsis* leaves. **E.** Reaction of subcellular marker antibodies to wild-type *Arabidopsis* leaf extracts. Lane 1 is the reaction of the cytoplasmic marker anti-cFBPase antibody (diluted to 1:3000) to 20 μg of a cytoplasmic fraction. Lane 2 is the reaction of the vacuolar marker anti-V-ATPase antibody (diluted 1:1000) to 20 μg of protein from a leaf extract. Lane 3 is the reaction of the endoplasmic reticulum marker anti-BiP antibody (diluted 1:200) to 30 μg of protein of a microsomal fraction. (Gel acrylamide percentages, from left to right, 15%, 12%, 10%).



wild-type *Arabidopsis* leaf extracts (Figure 3.1D, lane 4). The anti-GFP antibody will also be useful in the analysis of *Arabidopsis* transgenics, since it is able to detect as little as 50 ng of GFP and it does not cross-react with other proteins in *Arabidopsis* leaf extracts.

For subcellular fractionation studies, several antibodies were purchased to serve as subcellular markers. Most of these antibodies specifically detected proteins of the expected size in wild-type *Arabidopsis* leaf fractions. The cytoplasmic marker α-cFBPase was raised against *Arabidopsis* cFBPase overexpressed in *E. coli* and is reactive to cFBPases from both *Arabidopsis* and *Brassica napus* (Agrisera). In a leaf cytoplasmic fraction, it detected a band of approximately 37 kDa (Figure 3.1E, lane 1), which is the predicted size for cFBPase. The tonoplast marker α-V-ATPase was raised against a synthetic peptide conserved in the E subunit of plant V-ATPases and is reactive to V-ATPases from *Arabidopsis*, tomato, petunia, and tobacco (Agrisera). It detected a band of the expected size of approximately 29 kDa in a total leaf extract (Figure 3.1E, lane 2). The ER marker α-BiP was raised against the amino acids 541-635 from the C-terminus of *Arabidopsis* BiP and is reactive to BiPs from *Arabidopsis*, tobacco, tomato, and maize (Santa Cruz Biotechnology). It detected a band of approximately 78 kDa in a microsomal fraction (Figure 3.1E, lane 3), which is the expected size for *Arabidopsis* BiP. Since these antibodies detect a single band of the correct size, they appear to be specific for the intended subcellular marker protein, suggesting that they will function effectively as subcellular markers.

Anti-ascorbate oxidase raised against ascorbate oxidase from *Cucurbita* sp. (Agrisera) and anti-Xet raised against recombinant *P. tremula x tremuloides* xyloglucan endotransglucosylases (Agrisera) were tested as apoplast markers. However, neither antibody detected an appropriately sized band in up to 60 μg of total leaf extracts or 40 μg of an apoplast

fraction, collected from *Arabidopsis* leaves using the vacuum infiltration method (Firek et al. 1993). Therefore, these antibodies were not effective as markers of the apoplast and were not used in subsequent studies.

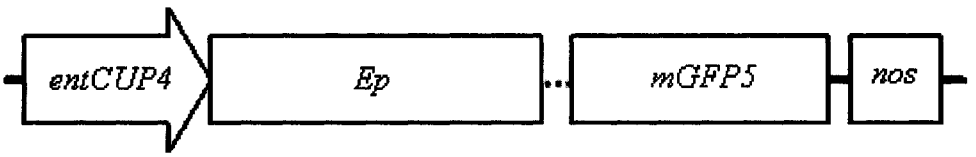
3.3 Cloning and plant transformation

Four genes encoding fusions between mature SBP and GFP were created, each with different combinations of the SBP N- and C-terminal propeptides and an ER-retention signal (Figure 3.2). *cSBP-GFP* contains no targeting signals and the resulting fusion protein is predicted to be found in the cytoplasm. To create *aSBP-GFP*, a sequence encoding the N-terminal SBP propeptide was added, and the resulting fusion protein is predicted to be secreted to the apoplast. Sequences encoding the N-terminal propeptide and a C-terminal ER-retention signal were added to create *eSBP-GFP*, which is predicted to encode an ER-localized fusion protein. Finally, sequences encoding the N- and C-terminal SBP propeptides were added to form *vSBP-GFP*, which is predicted to express a vacuolar-targeted fusion protein. Along with the SBP C-terminal propeptide, the last four amino acids of mature SBP were added in order to conserve the site of cleavage of the propeptide.

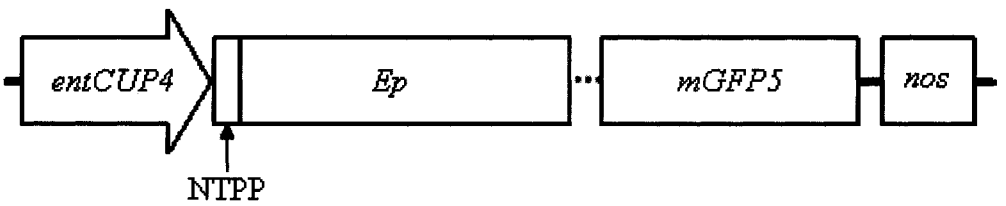
Linker peptides were inserted between the different protein domains to facilitate cloning and to increase their separation so as to aid the formation of tertiary structure of the two domains and thereby conserve protein function (F. Brandizzi, pers. comm.). Between SBP and GFP the three amino acid peptide Gly-Ser-Ala was inserted. Between GFP and the ER-retention signal HDEL the three amino acid peptide Ala-Gly-Ala was inserted. Finally, between GFP and the SBP C-terminal propeptide the two amino acid peptide Ala-Ser was inserted. Alanine, glycine, and serine were chosen to form the linker because they are small, flexible amino acids that are

Figure 3.2 Schematic diagrams illustrating the gene cassettes containing a fusion between the *Ep* gene encoding soybean peroxidase and the *mGFP5* gene encoding green fluorescent protein with either **A.** no signals, or the addition of sequences encoding **B.** the SBP N-terminal propeptide (NTPP), **C.** the NTPP plus the ER-retention signal HDEL, or **D.** the NTPP plus the SBP C-terminal propeptide (CTPP). Each contains the *entCUP4* promoter (Malik et al. 2002) and the *nopaline synthase* transcriptional terminator (*nos*). Linker peptides between separate protein domains are indicated by a dashed line (not drawn to scale).

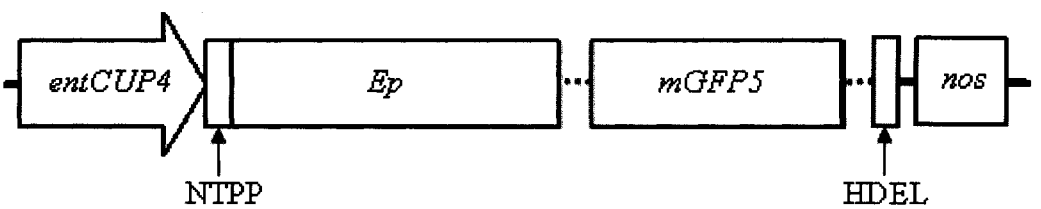
A. cSBP-GFP



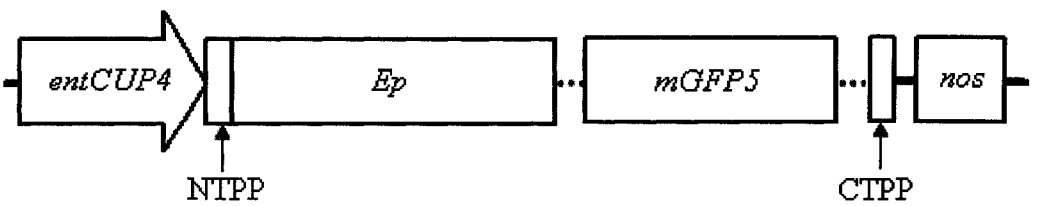
B. aSBP-GFP



C. eSBP-GFP



D. vSBP-GFP



commonly found in linker peptides that join protein domains (Argos 1990).

These genes were each introduced into a pCAMBIA1300 vector, under the control of the *entCUP4* promoter (Malik et al. 2002), which drives high, constitutive expression in plants, and a *nos* transcriptional terminator. The final constructs were sequenced to confirm that all elements were properly amplified and cloned. The four constructs were introduced into *A. thaliana* by *Agrobacterium*-mediated floral transformation, yielding T_1 seeds. These seeds were germinated on selective plates containing hygromycin to select for transformants and then transferred to soil.

3.4 Confirmation of transgene integration

Arabidopsis transformants were screened by PCR to confirm the presence of the T-DNA insertion. Plants were grown on media containing hygromycin to select for those with a T-DNA insertion, so the *hptII* gene was first amplified to test the integrity of the genomic DNA extracts. To confirm the presence of the remainder of the T-DNA, the *mGFP5* gene was also separately amplified (data not shown). The vectors used for plant transformation were included as positive controls and DNA from untransformed plants was included as a negative control. For each construct, twenty plants positive for both elements were selected for further analysis. All analyses, with the exception of GFP imaging, were done on T_1 *Arabidopsis* plants with each plant presumably representing an independent transformation event (Desfeux et al. 2000), although the site of insertion was not characterized to confirm this.

3.5 Confocal imaging of green fluorescent protein

To confirm the location of each of the four SBP-GFP proteins, *Arabidopsis* leaves were imaged by laser-scanning confocal microscopy. Autofluorescence can be detected in wild-type

plants (Figure 3.3A). In plants containing *cSBP-GFP* (Figure 3.3B) and *aSBP-GFP* (Figure 3.3C), no GFP signal could be detected except for autofluorescence similar to what is seen in the wild-type plants. This was observed in multiple plants from at least three different transformation events for each construct. Those plants containing *eSBP-GFP* gave a signal in a reticulate pattern that is consistent with ER localization (Figure 3.3D). This pattern was observed in at least three T_2 plants from each of three independent transformation events. In most plants containing *vSBP-GFP*, GFP could not be detected. However, in a single plant from each of two separate transformation events, a GFP signal could be detected throughout the cell, indicative of a large, central vacuole (Figure 3.3E).

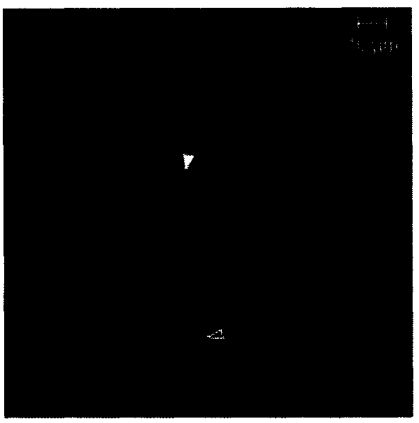
Based on these results, eSBP-GFP is localized in the ER. This suggests that the SBP N-terminal propeptide is functioning as an N-terminal signal peptide directing import into the ER and the ER-retention signal is functioning to retain the fusion protein in the ER. These results also suggest that vSBP-GFP is localized in the vacuole, providing additional support that the N-terminal propeptide is a functional signal peptide and suggesting that the C-terminal propeptide is functioning as a vacuolar sorting signal, although further analysis is required to confirm this result.

3.6 Subcellular localization of SBP-GFP fusion proteins

In order to determine the location of aSBP-GFP and confirm the location of vSBP-GFP, apoplast and vacuole fractions were isolated and analyzed for the presence of the recombinant fusion protein. To confirm the identity and integrity of these fractions, they were also analyzed for the presence of known subcellular markers (Quail 1979). Based on immunoblotting and activity assays, discussed below, the yield of cSBP-GFP was generally too low to be detected, so

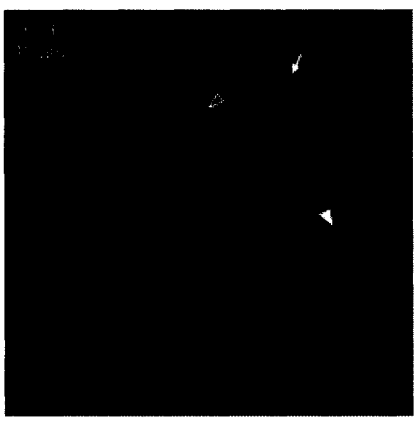
Figure 3.3 Confocal laser microscopy of *Arabidopsis* leaves from **A.** a wild-type plant or transgenic plants containing the **B.** *cSBP-GFP*, **C.** *aSBP-GFP*, **D.** *eSBP-GFP*, or **E.** *vSBP-GFP* gene. The images were taken with a Zeiss LSM 510 META laser scanning confocal microscope. In **E.** panels show, from left to right, GFP fluorescence, transmissible light, and a merging of the first two panels. Closed and open arrowheads and arrows show chloroplast, cell wall, and guard cell autofluorescence, respectively.

A.

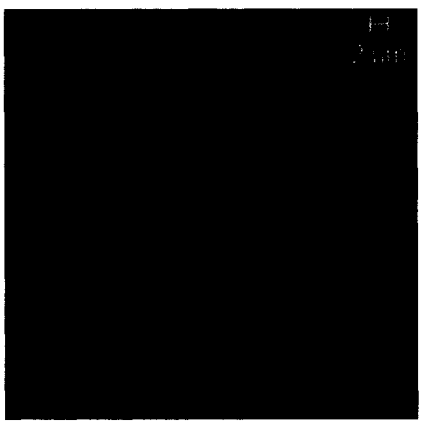


B.

C.



D.

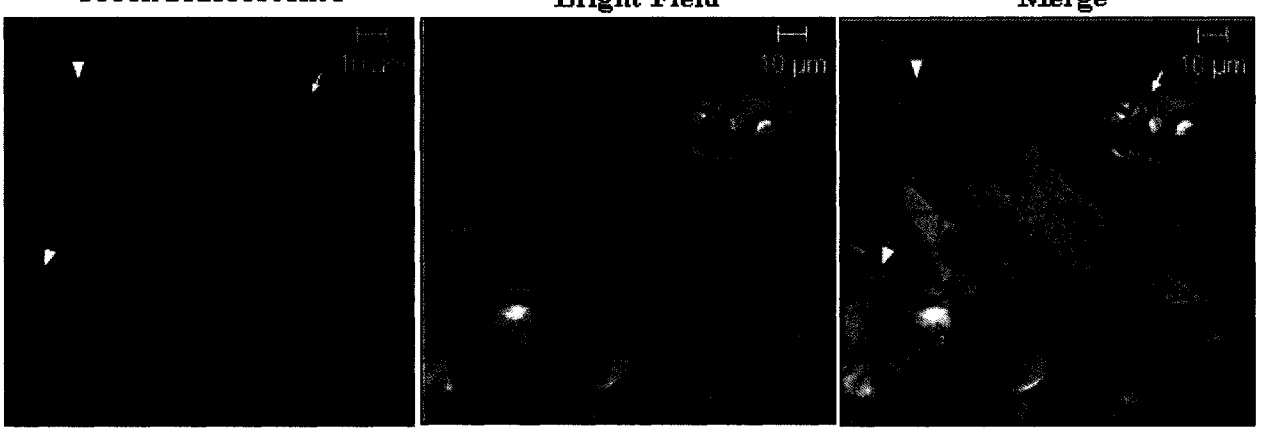


E.

Green Fluorescence

Bright Field

Merge

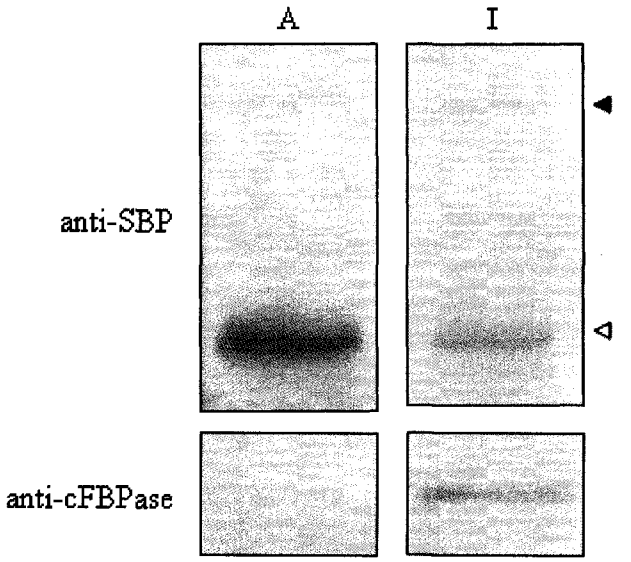


I was unable to confirm the cytoplasmic location of this fusion protein. Isolation of microsomal fractions from *eSBP-GFP* plants was attempted to confirm the ER-localization of eSBP-GFP, but it was difficult to produce fractions without contamination from other compartments, such as the vacuole. Since the ER-localization of eSBP-GFP was clear and reproducible based on the GFP imaging, subcellular fractionation of this line was not pursued further.

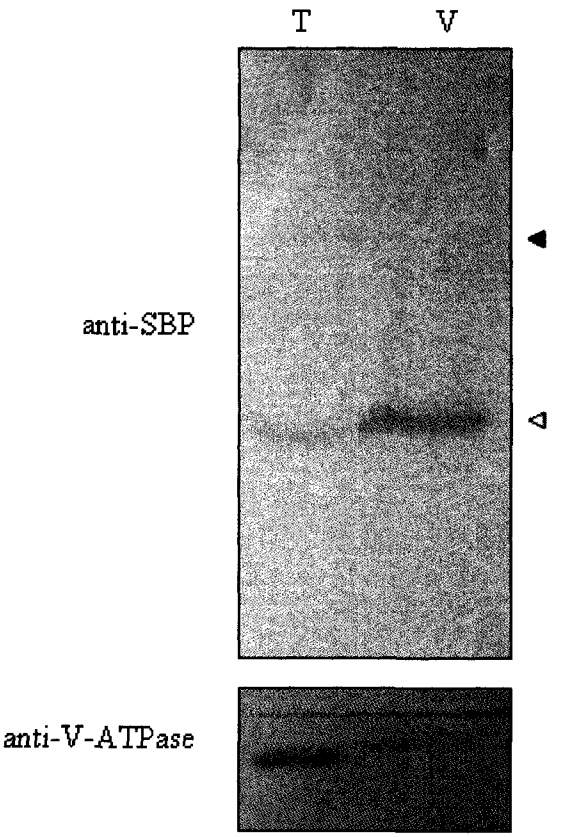
To confirm the location of the aSBP-GFP protein, apoplastic fluid was collected from *Arabidopsis* leaves using the vacuum infiltration method followed by western blotting (Firek et al. 1993). No apoplast marker was available to confirm the identity of the apoplast fraction but antibody against the cytoplasmic marker cFBPase was used to detect cytoplasmic contamination. A similar approach was used by Boudart et al. (2005) with the cytoplasmic marker malate dehydrogenase. Based on blotting with an anti-SBP antibody, aSBP-GFP was more concentrated in the apoplast fraction, although a signal could typically still be detected in the intracellular fraction (Figure 3.4A). The aSBP-GFP associated with the intracellular fraction could represent protein traveling through the secretory system, either passing through the ER or in transit to the plasma membrane for secretion (Denecke et al. 1990; Di Sansebastiano et al. 1998; Matsui et al. 2006). Alternatively, it may be that aSBP-GFP was not fully removed from the apoplast fraction. Proteins in the apoplast may bind to the cell wall with varying strength, making it difficult to fully remove them by vacuum infiltration (Lee et al. 2004). HRP C expressed in tobacco was partially removed by vacuum infiltration by including $CaCl_2$ in the infiltration buffer, but a fraction of the peroxidase activity did remain with the cell wall fraction and could only be released by its digestion, suggesting a strong association with the cell wall (Kis et al. 2004). It is possible that aSBP-GFP is similarly strongly associated with cell wall components, preventing its complete removal from the apoplast.

Figure 3.4 Immunoblot analysis, as described in Section 2.11, of subcellular fractions from *A. thaliana* leaves from plants containing the *aSBP-GFP* and *vSBP-GFP* genes. **A.** Immunoblot analysis of apoplast (A) and intracellular (I) fractions (20 μg) of *A. thaliana* leaves. The upper panels (10% acrylamide gel) show the reaction of anti-SBP antibody (diluted at 1:500) and the lower panels (12% acrylamide gel) show the reaction of the cytoplasmic marker anti-cFBPase (diluted at 1:3000). **B.** Immunoblot analysis (10% acrylamide gel) of total (T) and vacuolar (V) fractions of *A. thaliana* leaves (20 μg). The upper panel shows the reaction of anti-SBP antibody (diluted at 1:500) and the lower panel shows the reaction of the vacuolar marker anti-V-ATPase (diluted at 1:1000). A colorimetric assay was used to develop the immunoblots using the reagents NBT and BCIP. The closed arrowhead shows the band with the expected size of the SBP-GFP fusion protein and the open arrowhead shows the smaller band with the expected size of SBP alone.

A.



B.



To confirm the location of the vSBP-GFP protein, a vacuolar fraction was isolated from *Arabidopsis* leaves followed by western blotting (Matsui et al. 2006). The identity of the fraction was confirmed with an antibody against the tonoplast protein V-ATPase (Shimaoka et al. 2004). Contamination of the fraction was tested with an antibody against the ER resident protein, BiP (Shimaoka et al. 2004). Anti-BiP did not detect any bands in the vacuolar fractions, even when up to 50 μg are loaded on the gel, suggesting that they were not contaminated by other cellular compartments (data not shown). Based on blotting with anti-SBP, vSBP-GFP was enriched in the vacuolar fraction, confirming vacuolar targeting (Figure 3.4B)[4]. This suggests that the SBP C-terminal propeptide is functioning as a vacuolar sorting signal, as predicted.

Based on these results and the results of GFP imaging, aSBP-GFP is primarily secreted to the apoplast, eSBP-GFP is retained in the ER, and vSBP-GFP is transported to vacuoles. This suggests that the SBP N-terminal propeptide functions as a signal peptide directing import into the ER and the C-terminal propeptide functions as a vacuolar sorting signal, as predicted. These signals are functional in a heterologous plant system, suggesting functional conservation which is a common observation for plant targeting signals (Bednarek and Raikhel 1991; Brown et al. 2003; Frigerio et al. 1998).

3.7 Peroxidase activity assays in *Arabidopsis* leaves

As a measure of accumulation of the SBP-GFP protein in leaves from transgenic *Arabidopsis*, peroxidase activity assays were performed by measuring the rate of oxidation of

[4] In an attempt to concentrate the vacuoles from one vacuolar preparation, vacuoles were pelleted by centrifugation at 21,000 g for 15 min. Unexpectedly, when analyzed by immunoblotting with anti-SBP antibody, no vSBP-GFP was detected in the resuspended pellet. However, a strong band was detected in the supernatant. The vacuolar marker could be detected in both the pellet and supernatant. It is possible that this centrifugation step lysed the vacuoles, releasing their content into the supernatant while pelleting only the tonoplast. The vacuolar marker used in these experiments is more specifically a marker of the tonoplast, which would explain why it was detected in the vacuolar pellet while SBP was not, if this explanation were correct.

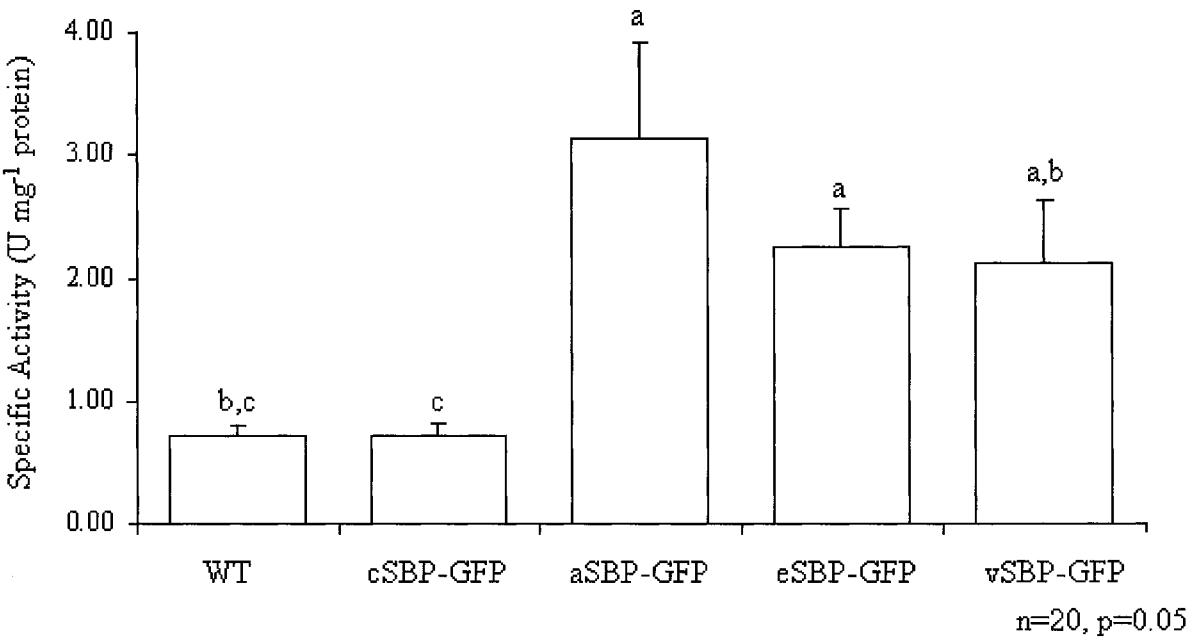
ABTS. Twenty independent transgenics were analyzed for each construct and for each plant protein was extracted from three equally sized leaves collected from mature plants grown with a 12 h day/12 h night photoperiod. The mean peroxidase activity for each construct as well as for untransformed plants is shown in Figure 3.5A.

Peroxidase specific activities were significantly different, $p<0.0005$ [$F(4,95)=8.90$], as determined by one-way analysis of variance (ANOVA). Post-hoc analyses using the 95% simultaneous confidence intervals for specified linear combinations by the Bonferroni method (as described in Section 2.14) indicated that mean peroxidase activities in plants in which SBP-GFP is targeted to any of the three locations, apoplast, ER, or vacuole, were significantly different than the activity in plants which contained non-targeted, cytoplasmic SBP-GFP, although only plants in which SBP-GFP is targeted to the apoplast or ER but not the vacuole were significantly different than non-transformed (wild-type) plants. These results indicate that targeting to any of the three compartments of the secretory system in *Arabidopsis* leaves can equally increase the accumulation of foreign proteins.

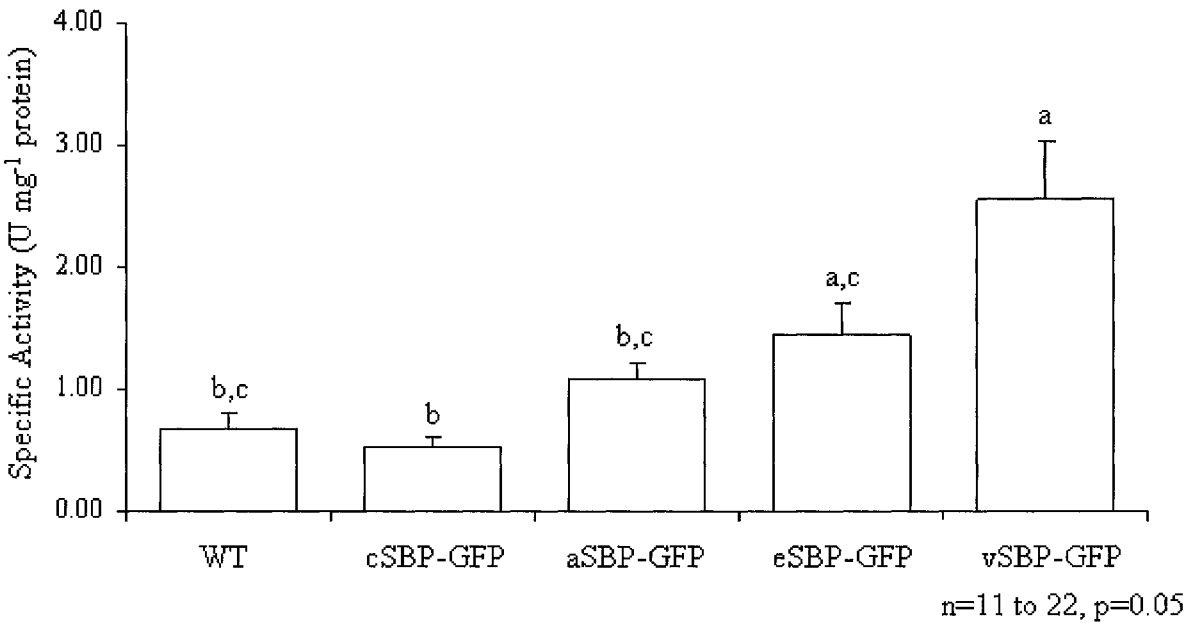
Different results were obtained when this experiment was repeated under long day conditions (16 h day/8 h night) (Figure 3.5B). The mean peroxidase specific activities of the leaf extracts were again significantly different, $p<0.0005$ [$F(4,91)=8.92$ and the peroxidase activity in plants in which SBP-GFP is targeted to the vacuole is significantly different than activity in all other locations and wild-type plants, except for plants in which SBP-GFP is targeted to the ER. The peroxidase activity in *eSBP-GFP* plants was only significantly different from the activity in *cSBP-GFP* plants. Activity levels in both *eSBP-GFP* and *aSBP-GFP* plants were not significantly different from activity levels in wild-type plants and the activity of *aSBP-GFP* plants was also not significantly different than the non-targeted *cSBP-GFP* plants. These results

Figure 3.5 Mean peroxidase specific activity in leaf extracts from untransformed *A. thaliana* plants (WT) or plants transformed with the *cSBP-GFP*, *aSBP-GFP*, *eSBP-GFP*, and *vSBP-GFP* genes. **A.** Results from plants grown with a 12 h photoperiod. **B.** Results from plants grown under long day conditions (16 h day/8 h night). Means sharing letter(s) are not significantly different based on post-hoc analyses using the 95 % simultaneous confidence intervals for specified linear combinations by the Bonferroni method, as described in Section 2.14.

A.



B.



are different from the experiment described above in which the plants were grown with a 12 h photoperiod, which showed that the activity of plants containing the three targeted SBP-GFP fusion proteins was greater than activity in plants containing the non-targeted fusion protein. This indicates that growth conditions can impact the accumulation of foreign proteins in different compartments. The impact of these results on experiments that study protein targeting and accumulation will be further examined in the Discussion.

3.8 Immunoblot analysis of *Arabidopsis thaliana* leaf extracts

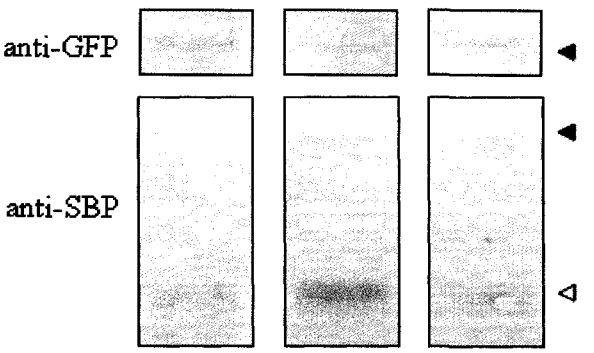
Arabidopsis leaf extracts were also analyzed by immunoblotting to determine if the activity of the extracts is comparable to the yield of protein and to confirm the size of the protein.

SBP-GFP protein was detected by both anti-SBP and anti-GFP antibodies (Figure 3.6). As a positive control, either SBP purified from soybean seed coats or recombinant GFP was used. Extracts from each of the twenty plants for each construct were analyzed with at least one immunoblot probed with anti-SBP and a second immunoblot probed with anti-GFP.

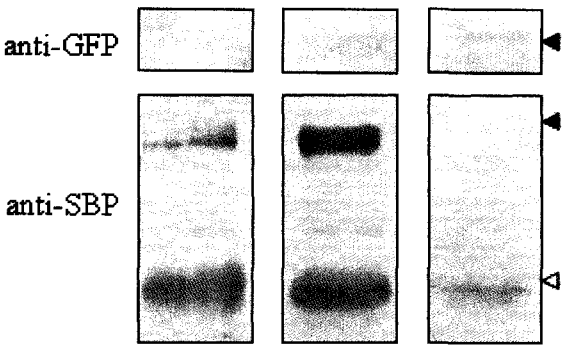
SBP-GFP protein was rarely detected in extracts from the plants containing *cSBP-GFP* and no bands were seen in extracts from non-transformed plants. The leaf extracts from the other three transgenic *Arabidopsis* plant lines showed two bands when blotted with anti-SBP antibody. The larger band was approximately 75 kDa, which is close to the expected size of the fusion protein. A second smaller band that is the same size as SBP purified from soybean seed coats, approximately 43 kDa, could also be detected. When the leaf extracts were blotted with anti-GFP antibody, only the larger band could be detected. Since the smaller protein was unreactive to the anti-GFP antibody and is indistinguishable in size from commercial SBP, it is likely that the GFP portion of the fusion protein is being degraded, leaving only the SBP portion. Based on the

Figure 3.6 Immunoblot analysis (10% acrylamide gels) of leaf extracts from *Arabidopsis* plants containing the **A.** *aSBP-GFP*, **B.** *eSBP-GFP*, and **C.** *vSBP-GFP* genes, as described in Section 2.11. A colorimetric assay was used to develop the immunoblots using the reagents NBT and BCIP. The upper immunoblots were tested with anti-GFP antibody (diluted at 1:500) and the lower immunblots were tested with anti-SBP (diluted at 1:500). The closed arrowheads show the bands with the expected size of the SBP-GFP fusion protein and the open arrowheads show the smaller bands with the expected size of SBP alone.

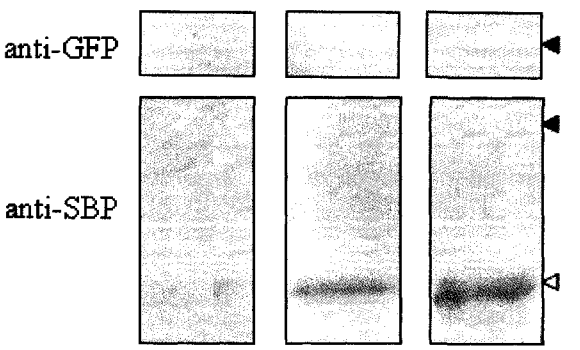
A.



B.



C.



density of the chromogenic precipitate, a larger proportion of the recombinant protein was present as the higher molecular weight form than as the lower molecular weight form in the extracts from the plants containing *eSBP-GFP* compared to any of the other transgenic lines, suggesting the degree of degradation of GFP is less severe in the ER than in the vacuole or apoplast.

The total density of both of the bands was roughly correlated with the specific activity of the extract, such that extracts with higher specific activity typically had more dense bands. While this suggests that most of both forms of the recombinant fusion protein are functional, quantification of the density of both species is needed to confirm this.

As a result of the partial degradation of the fusion protein, the plants contain two distinct populations of marker protein that could have different levels of activity, which could affect the interpretation of the targeting results. It is possible that the full-length fusion protein containing both SBP and GFP has a different activity level than the degraded fusion protein consisting primarily of SBP. In particular, eSBP-GFP plants, which typically contained a larger proportion of the full-length fusion protein could have their activity either under- or over-represented. However, based on immunoblots, the total amount of protein detected by anti-SBP roughly correlated with the activity levels as detected by activity assay, which suggests that differences in activity between the full-length and degraded fusion protein did not greatly impact the results.

Degradation of the fusion proteins could also impact their targeting, depending on where in the secretory system the degradation is occurring. For aSBP-GFP, there is no C-terminal targeting information, so degradation is not likely to impact its targeting. For eSBP-GFP, the ER-retention signal is C-terminal and degradation could result in its loss. If this occurred, it is possible that the degraded fusion protein would be secreted. An apoplast fraction prepared from

eSBP-GFP Arabidopsis leaves did not contain any detectable eSBP-GFP, suggesting that there is not a great amount of eSBP-GFP being secreted (data not shown). Similarly, the vacuolar sorting signal in vSBP-GFP is C-terminal. If degradation occurs prior to sorting to the vacuole, then vSBP-GFP could potentially be secreted. In an apoplast fraction, a small amount of vSBP-GFP could be detected, but the majority of protein was intracellular (data not shown). This suggests that despite degradation, the majority of the fusion protein is being properly sorted in *eSBP-GFP* and *vSBP-GFP* plants.

Due to degradation of the fusion protein, the results of this study should be interpreted with caution. However, degradation does not appear to have greatly impacted the targeting of the fusion protein and the level of activity of the SBP portion of the protein, suggesting that the results are reliable.

CHAPTER 4 EXPRESSION OF RECOMBINANT SOYBEAN PEROXIDASE IN SOYBEAN SEED COATS

4.1 Expression of the *Ep* transgene in *epep* soybean plants

For the soybean seed coat system, we used the *Ep* 1.5 promoter to drive expression of foreign genes. However, since the *Ep* promoter has not yet been fully characterized, it is possible that elements necessary for the high expression levels and seed coat specificity of the *Ep* gene were not included in the *Ep* 1.5 promoter. Since regulatory elements may be found at locations distant to their site of action (Clark et al. 2006; Jack et al. 1991; Krumm et al. 1998; Lettice et al. 2003), it is possible that some of the elements that regulate expression of the *Ep* gene were further upstream than the 1.5 kb region we selected. In addition, other genetic elements may also contribute to the regulation of *Ep* expression, such as introns and UTRs (Gallie and Young 1994; Vasil et al. 1989; Tanaka et al. 1990; Bailey-Serres and Dawe 1996; Gallie 1993; Bate et al. 1996). In order to be able to directly compare the level of expression from the *Ep* 1.5 promoter to the endogenous *Ep* promoter, we chose to express the *Ep* gene in an *epep* cultivar in the transgenic soybean seed coat system. Two cDNA constructs were made, one containing the *Ep* cDNA, which will be referred to as *EpC*, and the other containing the *Ep* cDNA plus 3'UTR, which will be referred to as *EpW*. A comparison of the accumulation of SBP in these two transgenic lines will test the contribution of the 3'UTR to the expression of the *Ep* gene. We also designed a construct containing the entire *Ep* genomic locus in order to determine if there are additional elements that contribute to *Ep* expression, such as intron regions, but no transformants were recovered.

To facilitate the comparison between transgenic and wild-type *Ep* soybean plants, the developmental accumulation of SBP in the *EpEp* cv. Harosoy 63 was analyzed by measuring the

increase in peroxidase activity and the accumulation of SBP by immunoblotting, complementing what is already known about the development of seed coats and the accumulation of *Ep* transcript (Miller et al. 1999; Gijzen et al. 1999).

4.2 Accumulation of soybean peroxidase in *EpEp* plants

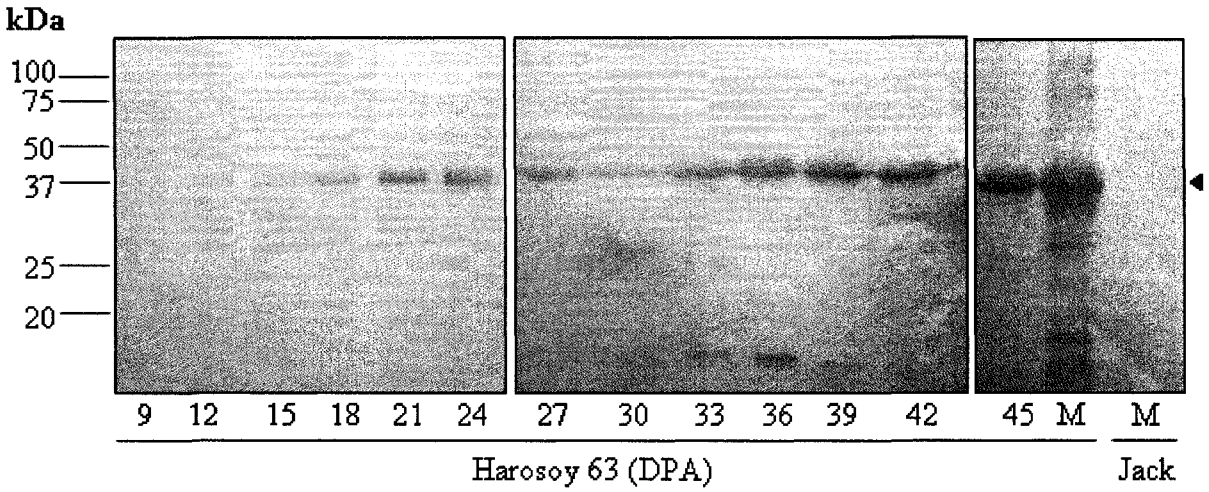
Based on *in situ* hybridization analysis, *Ep* mRNA can be detected as early as 6 DPA (Gijzen et al. 1999). Immunoblots and specific activity assays were performed on protein extracted from soybean seed coats at different developmental stages to determine how the accumulation of SBP correlates with the expression of the *Ep* gene. Every three days between 9 and 45 DPA, seed coats were collected from the *EpEp* cv. Harosoy 63. At each stage, three protein extracts were analyzed.

Accumulation of SBP was quite low at 9 and 12 DPA followed by a steady increase from 15 DPA to approximately 33 DPA after which levels remain relatively constant, showing no significant difference in mean peroxidase activity between 33 and 45 DPA ($p<0.05$) (Figure 4.1). For 9 and 12 DPA, as much as 60 μg of protein was required on an immunoblot before a band could be detected (Figure 4.1B). Therefore, accumulation of SBP begins early in seed coat development and steadily increases over approximately 30 days of development.

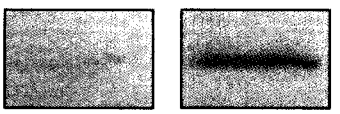
For this study, the *EpEp* cv. Harosoy 63 was selected for comparison to transgenic *Ep* plants. However, to determine if variability exists in the level of peroxidase in seed coats from different *EpEp* cultivars, the mean peroxidase activity of four additional *EpEp* cultivars was measured. For each cultivar, protein was extracted from three mature seed coats and the peroxidase activity was measured. The peroxidase activity of the six different cultivars was significantly different, $p<0.0005$ [$F(4, 10)=17.25$]. All five *EpEp* cultivars had peroxidase

Figure 4.1 Endogenous production of soybean peroxidase in soybean seed coats. **A.** Immunoblot analysis (12% acrylamide gel) of seed coat extracts (10 μg) from the *EpEp* cv. Harosoy 63 at different developmental stages, and mature cv. Jack seed coats with anti-SBP antibody (diluted at 1:500), as described in Section 2.11. Closed arrowhead indicates the expected size of SBP. A colorimetric assay was used to develop the immunoblots using the reagents NBT and BCIP. **B.** Immunoblot analysis, as in **A.**, with 60 μg of protein from 9 and 12 DPA Harosoy 63 seed coat extracts. **C.** Mean peroxidase specific activity (U mg^{-1} protein) of seed coat extracts from the *EpEp* cv. Harosoy 63 at different developmental stages, and mature Jack seed coats (n=3).

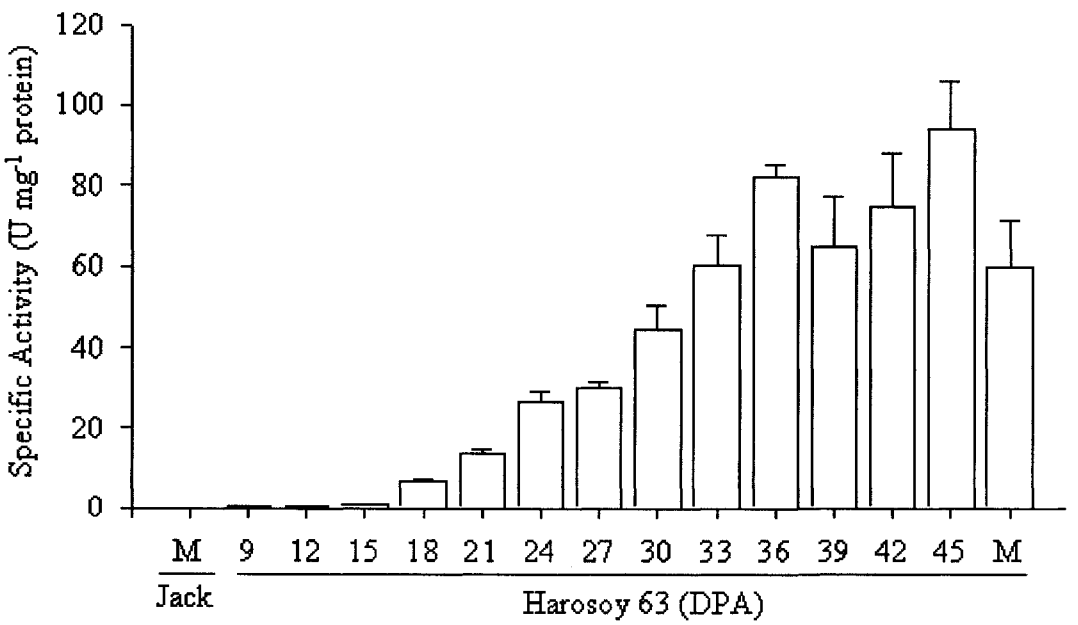
A.



B.



C.



specific activity that was significantly higher than the *epep* cultivar Jack. Although the specific activity amongst the five *EpEp* cultivars was not significantly different, activity ranged from an average of 20 U mg^{-1} protein for cv. Fayette to almost 140 U mg^{-1} protein for cv. Kunitz (Figure 4.2). The variability in the peroxidase activity levels in the different *Ep* cultivars may also reflect different growth conditions as well as different genetic backgrounds since the seeds from all five cultivars were obtained from AAFC (London, ON) and were likely grown at different times, under different conditions. Furthermore, since the seeds were not genotyped, it is possible that some of them were actually heterozygous at the *Ep* locus (*Epep* instead of *EpEp*), which could also affect the accumulation of SBP through a gene dosage effect. These results demonstrate that there is a wide range of peroxidase activity levels produced in *EpEp* soybean cultivars.

4.3 Cloning and plant transformation

The *EpC* and *EpW* genes were cloned by Dr. L. Gudynaitę-Savitch into a modified pCAMBIA1301 plasmid between the *Ep* 1.5 promoter and the *nos* transcriptional terminator (Figure 4.3). When designing transgenes, it is important to try to limit the influence of nearby promoter elements on the transgene promoter. These influences can come from within the transgene, from the promoter driving expression of the selectable marker, as well from outside the transgene, from genomic DNA that borders the site of insertion (Wilson et al. 1990; Jagannath et al. 2001; Yoo et al. 2005; Breyne et al. 1992a; Hampf and Gossen 2006; Eszterhas et al. 2002; Rubinchik et al. 2001). This is particularly important for tissue-specific promoters that can be influenced by constitutive promoter elements, such as the CaMV 35S promoter typically used to drive expression of selectable markers (Jagannath et al. 2001; Yoo et al. 2005; Rubinchik et al. 2001).

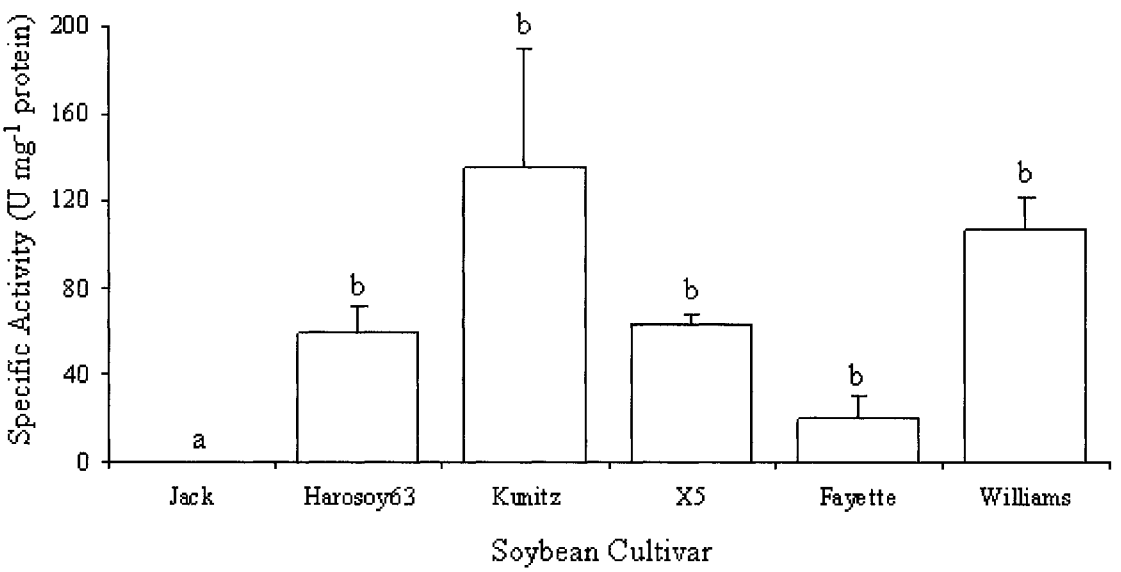
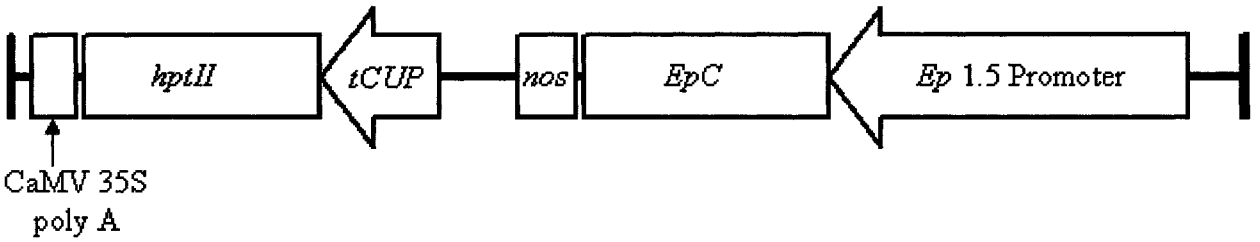


Figure 4.2 Mean peroxidase specific activity (U mg^{-1} protein) of different *EpEp* soybean cultivars as well as the *epep* cv. Jack (n=3). Means sharing a letter are not significantly different based on post-hoc analysis using the 95% simultaneous confidence intervals for specified linear combinations by the Bonferonni method, as described in Section 2.14.

A.



B.

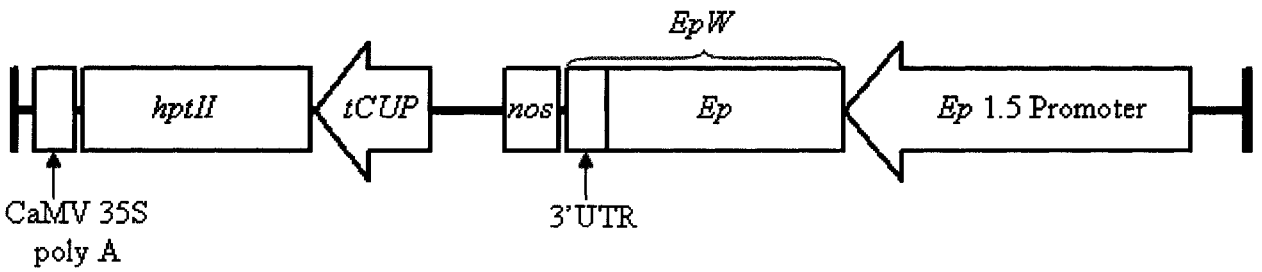


Figure 4.3 Schematic diagrams illustrating the T-DNA regions of the **A.** pCAM1300tCUP-Prx-EpCod#3 construct containing the *Ep* cDNA (*EpC*) and **B.** the pCAM1300-tCUP-Prx-EpWh#4 construct containing the *Ep* cDNA plus 3'UTR (*EpW*). Both constructs contain the *entCUP4* promoter (*tCUP*) driving the *hptII* selectable marker gene. The *Ep* 1.5 promoter was used to drive expression of the *Ep* genes. During PCR amplification of the promoter, a single nucleotide error (A→G) was introduced at position -1349. Following both *Ep* genes is the *nopaline synthase* transcriptional terminator (*nos*). Black bars on left and ride sides of constructs represent left and right T-DNA borders, respectively.

Based upon work done by Dr. L. Gudynaitę-Savitch to optimize transgene design (Gudynaitę-Savitch et al. submitted), the CaMV 35S promoter in pCAMBIA1301 used to drive expression of *hptII* was replaced with the *entCUP4* promoter from tobacco (Malik et al. 2002) and the orientation of the *Ep* transgenes were inverted, placing the *Ep* 1.5 promoter adjacent to the right T-DNA border and the *nos* transcriptional terminator adjacent to the *entCUP4* promoter to produce the plasmids pCAM1300tCUP-Prx-EpCod#3 (Figure 4.3A) and pCAM1300-tCUP-Prx-EpWh#4 (Figure 4.3B). These plasmids were used to transform soybean. Furthermore, genomic DNA adjacent to the site of transgene insertion can also influence promoter expression (Campisi et al. 1999; Springer 2000), suggesting that steps should be taken to limit the influence of elements both within and outside of the transgene. These results were taken into account for subsequent constructs, as will be seen in Chapter 5 for the *phb* construct.

The final constructs were sequenced to confirm that all elements were properly amplified and cloned. Sequencing revealed a single nucleotide error (A→G) at position -1349 in the *Ep* 1.5 promoter of both transgenes. However, promoter analysis work indicated that this error does not affect the activity of the *Ep* 1.5 promoter (Drs. L. Gudynaitę-Savitch and M. Shukla, unpublished results). The two *Ep* constructs were introduced into soybean (cv. Jack) by particle bombardment, and T_0 plants were regenerated and grown to maturity to produce T_1 seeds. T_1 seeds were germinated on moistened filter paper and after 4 to 6 days seed coats were collected for analysis and seedlings were transplanted into soil. Seed coats are maternal in origin, therefore the seed coats from T_1 seeds are from the T_0 generation and seed coats from the T_2 seeds are from the T_1 generation (Figure 2.1).

Transformation with the construct containing the *EpC* gene yielded several plants, 8 of which survived to maturity and yielded seeds. All 8 plants were from a single transformation

event, i.e. regenerated from the same somatic embryo, and therefore presumably represent a single clonal line (Trick et al. 1997), although the insertion site was not characterized. Of these 8 plants, only 3 had seeds that germinated, providing T_0 seed coats for analysis and T_1 plants. Between 9 and 33 T_0 seed coats were analyzed per plant. Three T_1 plants grown to maturity and yielding seeds were also analyzed. Of these 3 plants, all had seeds that germinated, providing T_1 seed coats for analysis and T_2 plants. Between 1 and 10 T_1 seed coats were analyzed per plant.

Transformation with the construct containing the *EpW* gene yielded several plants. These plants were from three different transformation events and therefore presumably represent three independent clonal lines (Trick et al. 1997), although the sites of insertion were not characterized. One transformation event yielded 26 plants that survived to maturity and produced seeds. Of these 26 plants, 9 were analyzed but only 7 had seeds that germinated, providing between 2 and 9 T_0 seed coats for analysis as well as T_1 plants. Two T_1 plants grown to maturity and yielding seeds were also analyzed. Of these 2 plants, both had seeds that germinated, providing T_1 seed coats for analysis as well as T_2 plants. For one plant 9 seed coats were analyzed, and for the second plant 10 seed coats were analyzed. The second transformation event yielded 18 plants that survived to maturity and produced seeds. Of these 18 plants, 13 had seeds that germinated, providing T_0 seed coats for analysis as well as T_1 plants. Between 1 and 27 seed coats were analyzed per plant. The third transformation event yielded 3 plants that survived to maturity and produced seeds. All 3 had seeds that germinated, providing T_0 seed coats for analysis as well as T_1 plants. Between 1 and 16 seed coats were analyzed per plant.

4.4 Confirmation of transgene insertion

T_1 transgenic soybean plants were screened by PCR to confirm the presence of the

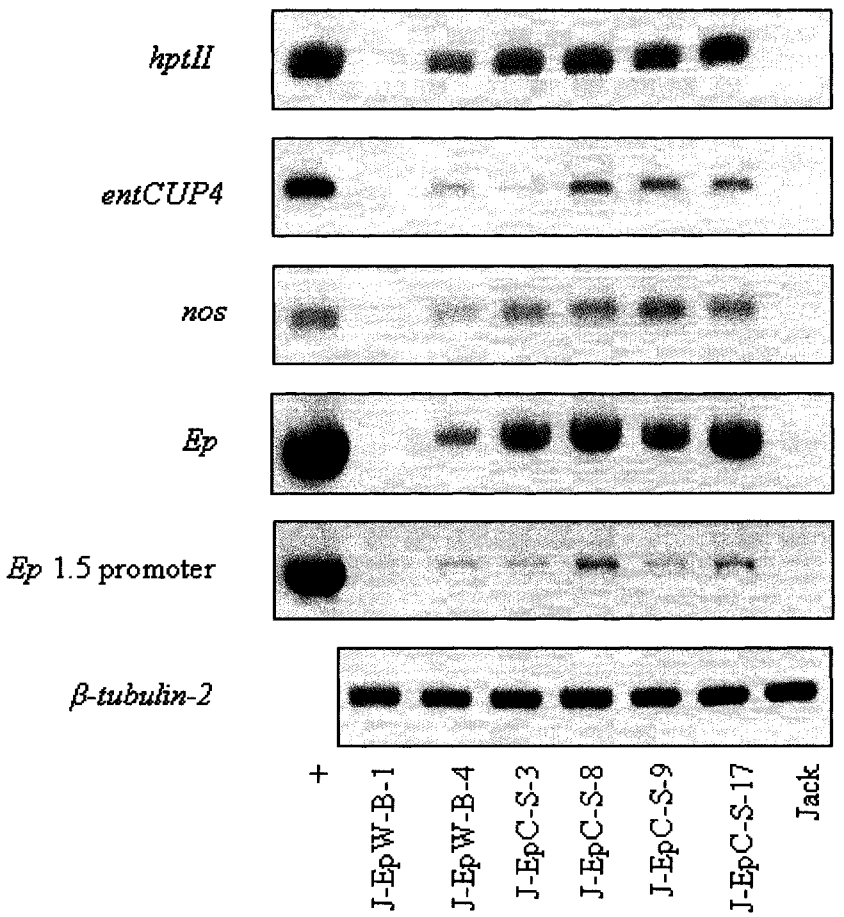


Figure 4.4 Amplification of transgene elements from soybean genomic DNA extracts. The elements encoding the *hptII* and *Ep* genes, the *entCUP4* and *Ep* 1.5 promoters, and the *nopaline synthase* transcriptional terminator (*nos*) were amplified separately by PCR. As a control for genomic DNA quality, the endogenous *β-tubulin-2* gene was amplified. The pCAM1300tCUP-Prx-EpCod#3 and pCAM1300-tCUP-Prx-EpWh#4 plasmid were used as positive controls (+) and DNA from untransformed soybean plants (cv. Jack) was used as a negative control. The plant J-EpW#4-B.1 has lost the transgene due to segregation while the remaining plants are positive for the transgene. The primers used are listed in Table 2.1 and conditions for amplification of each element in Table 2.2.

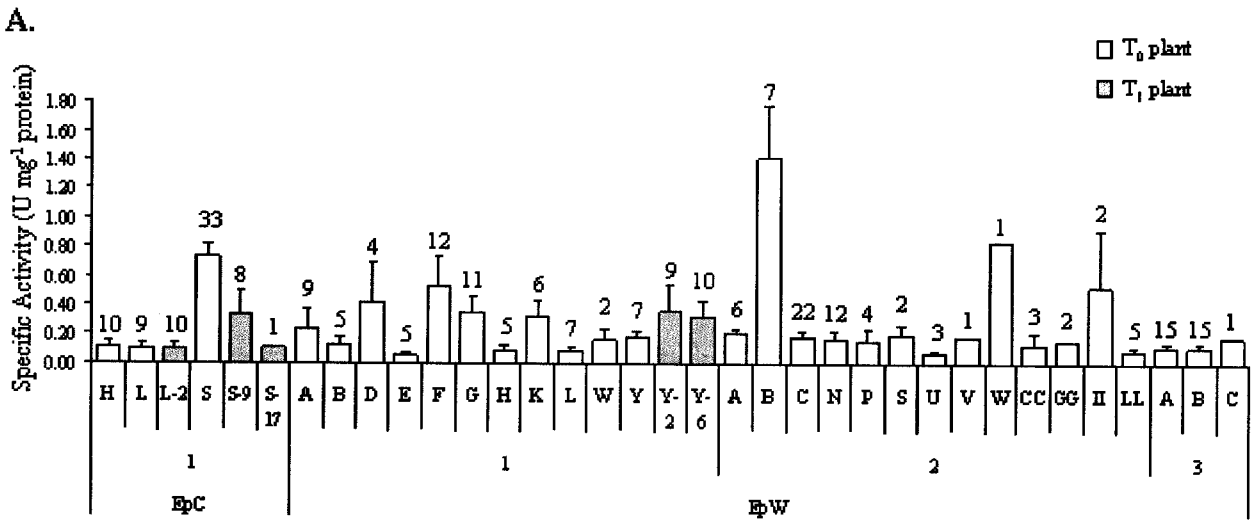
various T-DNA elements and identify any plants that may have lost some or all of the T-DNA through segregation or instability. PCR was done on genomic DNA template with primers that specifically amplify the *hptII* gene encoding hygromycin resistance, the *entCUP4* promoter, the *Ep* gene, the *Ep* 1.5 promoter, and the *nos* transcriptional terminator (Figure 4.4). The *β-tubulin-2* gene was also amplified to confirm the quality of the genomic DNA extracts. The pCAM1300tCUP-Prx-EpCod#3 and pCAM1300-tCUP-Prx-EpWh#4 plasmids were used as a positive control and DNA from untransformed soybean plants (cv. Jack) was used as a negative control. Of the 13 T_1 plants screened by PCR, two had lost the transgene. A second genomic DNA extract was prepared from those two plants to confirm the absence of transgene.

4.5 Peroxidase activity in soybean seed coats

To measure the accumulation of SBP in the transgenic soybean seed coats, peroxidase activity assays were performed by measuring the rate of oxidation of ABTS. For each plant, between 1 and 33 protein extracts were prepared, each from a half or a whole seed coat from a single seed. The mean peroxidase activity for each plant is shown in Figure 4.5.

Mean seed coat peroxidase activity ranged from 3 to 70 times higher than the mean activity measured in seed coats from the *epep* cv. Jack (n=3), which is deficient in peroxidase activity in the seed coat. However, the mean activity was 42 to 1200 times lower than the mean activity in seed coats from the *EpEp* cv. Harosoy 63 (n=3), which naturally produces high levels of SBP in its seed coat. Therefore, even though soybean peroxidase was produced in soybean seed coats through recombinant technologies, the system presently does not achieve levels comparable to endogenous expression. Mean peroxidase activity of plants from a single transformation event varied as much as 22-fold. In contrast, there was little overall difference

Figure 4.5 Mean peroxidase specific activity in soybean plants containing the *EpC* and *EpW* transgenes. **A.** Mean peroxidase specific activity (U mg^{-1} protein) + SE of plants from a single transformation event for the *EpC* transgene and from three transformation events for the *EpW* transgene. Sample size for each plant is indicated above each bar. **B.** Table showing the mean ± SE, maximum, and minimum peroxidase activity (U mg^{-1} protein) for each plant for a single transformation event from plants transformed with the *EpC* transgene and for three transformation events from plants transformed with the *EpW* transgene. Also indicated is sample size (n).



B.

Construct	Event	Plant	Generation	Mean	n	Maximum	Minimum
EpC	1	H	T_0	0.11 ± 0.04	10	0.45	0.02
		L	T_0	0.10 ± 0.04	9	0.24	0.02
		L-2	T_0	0.10 ± 0.04	10	0.44	0.00
		S	T_0	0.73 ± 0.09	33	2.12	0.05
		S-9	T_1	0.34 ± 0.16	8	1.25	0.02
		S-17	T_1	0.110	1	0.11	0.11
EpW	1	A	T_0	0.23 ± 0.14	9	1.34	0.03
		B	T_0	0.13 ± 0.06	5	0.34	0.04
		D	T_0	0.41 ± 0.29	4	1.26	0.07
		E	T_0	0.05 ± 0.02	5	0.14	0.02
		F	T_0	0.17 ± 0.20	5	0.53	0.03
		G	T_0	0.35 ± 0.11	11	1.11	0.03
		H	T_0	0.08 ± 0.04	5	0.24	0.02
		K	T_0	0.32 ± 0.11	6	0.78	0.02
		L	T_0	0.09 ± 0.02	7	0.17	0.04
		W	T_0	0.16 ± 0.08	2	0.24	0.07
		Y	T_0	0.18 ± 0.04	7	0.36	0.06
		Y-2	T_1	0.35 ± 0.19	9	1.75	0.03
		Y-6	T_1	0.32 ± 0.12	10	1.24	0.04
	2	A	T_0	0.20 ± 0.04	6	0.33	0.09
		B	T_0	1.41 ± 0.36	7	3.21	0.54
		C	T_0	0.19 ± 0.03	22	0.63	0.02
		N	T_0	0.17 ± 0.05	12	0.69	0.03
		P	T_0	0.15 ± 0.08	4	0.35	0.00
		S	T_0	0.20 ± 0.07	2	0.26	0.13
		U	T_0	0.06 ± 0.02	3	0.09	0.03
		V	T_0	0.190	1	0.19	0.19
		W	T_0	0.830	1	0.83	0.83
		CC	T_0	0.13 ± 0.08	3	0.28	0.03
		GG	T_0	0.15 ± 0.00	2	0.15	0.15
		II	T_0	0.52 ± 0.40	2	0.92	0.12
		LL	T_0	0.08 ± 0.02	5	0.16	0.03
	3	A	T_0	0.12 ± 0.01	15	0.24	0.05
		B	T_0	0.11 ± 0.02	15	0.40	0.02
		C	T_0	0.18	1	0.18	0.18

observed between plants from the different transformation events, with the mean peroxidase activity of these groups of plants varying up to 2.5-fold. This includes both those containing the *EpC* gene and those containing the *EpW* gene.

Silencing and position effects may have played a role in reducing transgenic *Ep* expression in these plants. Since only four different transformation events were examined, it is possible that all four contain transgenes inserted in an unfavorable location. However, low levels of expression have been similarly observed in transgenic plants transformed with other constructs (S. Han, unpublished results), suggesting that it may be more likely that the *Ep* 1.5 promoter does not contain all of the regulatory elements responsible for the high levels of endogenous expression. Regulatory elements may be located greater than 1.5 kb from the *Ep* gene or they may be located in the intron sequences that were not included in these constructs. The potential contributions of these factors to the lower level of expression of the transgenic *Ep* gene will be further examined in the Discussion.

4.6 Expression of *Ep* is reduced in transgenic soybean plants

Based on the reduced peroxidase levels detected in the transgenic *Ep* plants compared to the wild-type *EpEp* cv. Harosoy 63, we suspect that regulatory elements that promote the high level of expression of the *Ep* gene were omitted in the *Ep* 1.5 promoter. In order to determine if the reduced level of peroxidase activity is correlated with reduced levels of *Ep* expression, RT-PCR was performed on 21 DPA seed coat tissues. We chose 21 DPA because at this point in wild-type *EpEp* cultivars the *Ep* gene is expressed and expression is restricted to the hourglass cell layer (Gijzen et al. 1999). Also, most of the changes that will occur to the hourglass cell layer are complete at this stage (Miller et al. 1999). At 21 DPA, SBP is clearly detectable both by

immunoblotting and peroxidase activity assay in Harosoy 63 seed coat extracts (Figure 4.1). In order to be able to determine the relative quantity of transgenic *Ep* transcript compared to *Ep* transcript from wild-type plants, transcript from the internal control *β-actin* (BU082893) was simultaneously amplified in all samples. In order to relate transcript levels to protein levels, immunoblots and activity assays were similarly performed for seed coats at 21 DPA.

In order to specifically amplify *Ep* transcript, the forward primer was designed to hybridize to the region of the *Ep* gene that is deleted in *epep* cultivars. This will prevent amplification of the low levels of transcript that are produced by the *ep* gene (Gijzen 1997). The reverse primer for *Ep* transcript amplification was designed to cross one of the introns in the genomic *Ep* gene. This will prevent amplification of genomic DNA from the *Ep* locus, allowing *Ep* cDNA to be specifically amplified. Both the primers for the *Ep* transcript and the *β-actin* transcript gave rise to bands of the expected size.

RNA was isolated from 21 DPA seed coats from the wild-type *EpEp* cv. Harosoy 63 and the T_1 soybean plant J-EpW-B-4. This was also done for the wild-type *epep* cv. Jack, as a negative control. This transgenic plant was chosen because it was one of the progeny of the T_0 plant with the highest average specific activity. PCR was performed, as described in Section 4.4, to confirm that all T-DNA elements were present in this plant. For each plant, three separate extracts, made from 2 to 3 seed coats, were prepared and analyzed. *Ep* transcript and the internal control *β-actin* transcript were reverse transcribed using gene-specific primers and then amplified by PCR with the same primers. As a control, PCR amplification was always performed simultaneously with a sample that did not undergo reverse transcription, in order to detect the presence of genomic DNA contamination. After 35 cycles, *Ep* transcript was only amplified from the Harosoy 63 seed

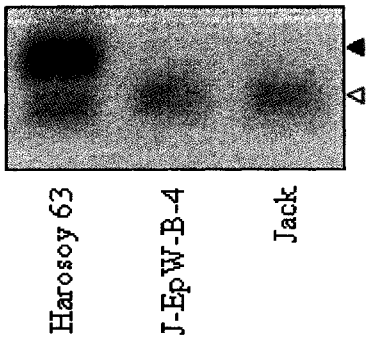
coat mRNA samples (Figure 4.6A). Therefore, expression of *Ep* mRNA in the transgenic plant J-EpW-B-4 is greatly reduced compared to the *Ep* cv. Harosoy 63.

Protein extracts were also prepared from 21 DPA seed coats from cv. Harosoy 63 and the J-EpW-B-4 plant and SBP accumulation was determined by immunoblotting with anti-SBP antibody and by peroxidase activity assay. Again, this was also done for wild-type cv. Jack plants, as a negative control. For each plant, three separate extracts, each from 2 to 3 seed coats, were analyzed. Extract from cv. Harosoy 63 gave a strong band while extract from cv. Jack did not give any band nor did the transgenic plant J-EpW-B-4 (Figure 4.6B). Based on activity assays, peroxidase activity in the three plants is significantly different, $p<0.0005$ [$F(2,6)=41.44$]. The activity in in Harosoy 63 extracts was significantly higher than in either Jack or J-EpW-B-4 extracts, which have equally low activity (Figure 4.6C). Therefore, accumulation of SBP is also much reduced in J-EpW-B-4 seed coats compared to the *EpEp* cv. Harosoy 63 at 21 DPA.

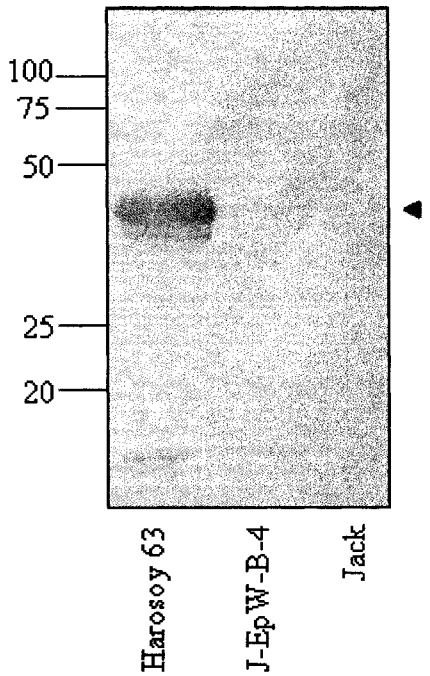
While these results support the hypothesis that the lower level of peroxidase activity in transgenic *Ep* plants compared to wild-type *EpEp* plants is due to reduced expression of the *Ep* gene, more analysis will be necessary to confirm this conclusion. These results were obtained from a single plant that could have reduced *Ep* expression due to silencing. It will be necessary to repeat this analysis on several other T_1 plants from this line as well as from other high-expressing lines, such as J-EpW-W, J-EpW-II, and J-EpC-S. RT-PCR measures transcript abundance and further analysis will be necessary to determine if the reduction in *Ep* transcript is correlated with reduced transcription due to the omission of regulatory elements from the *Ep* 1.5 promoter. It is possible that the reduction in *Ep* transcript could be due to other mechanisms, such as reduced transcript stability.

Figure 4.6 *Ep* expression and SBP accumulation in 21 DPA wild-type cv. Harosoy 63, J-EpW-B-4, and wild-type cv. Jack soybean seed coats. **A.** RT-PCR amplification (35 cycles) of *Ep* transcripts. Closed arrowhead indicates expected size of *Ep* amplicon and open arrowhead indicates expected size of *β-actin* amplicon. **B.** Immunoblot analysis (12% acrylamide gel) of SBP, as described in Section 2.11, with anti-SBP antibody (diluted at 1:500). For each sample, 50 μg of protein was loaded. A colorimetric assay was used to develop the immunoblots using the reagents NBT and BCIP. Closed arrowhead indicates expected size of SBP. **C.** Mean peroxidase specific activity + SE (n=3) of soybean seed coat extracts. Means sharing a letter are not significantly different based on post-hoc analysis using the 95% simultaneous confidence intervals for specified linear combinations by the Bonferroni method, as described in Section 2.14.

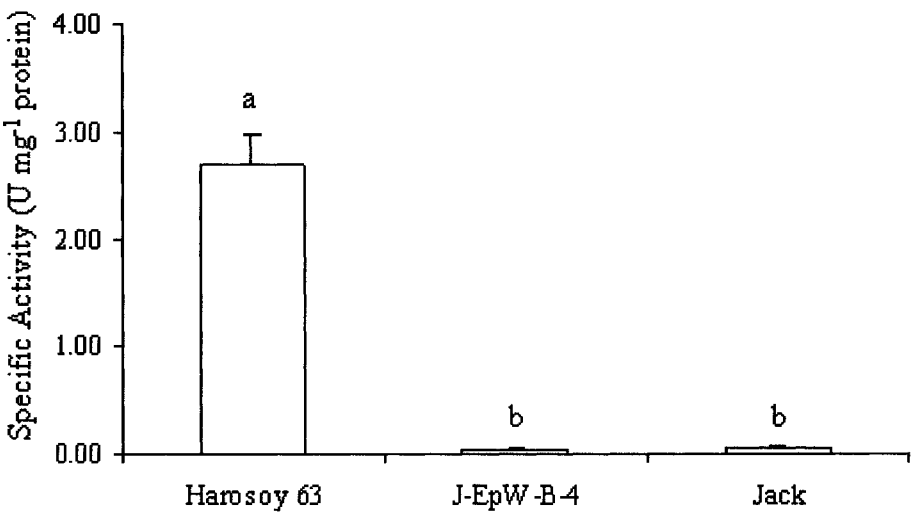
A.



B.



C.



CHAPTER 5 ENGINEERING SOYBEAN SEED COATS TO PRODUCE POLYHYDROXYBUTYRIC ACID

5.1 Production of polyhydroxybutyric acid in soybean seed coats

In addition to using soybean seed coats to produce proteins of economic value, they can also be used to produce other non-protein biochemicals. This is achieved through metabolic engineering either by altering existing metabolic pathways to enhance production of a naturally occurring metabolite or by diverting metabolism towards the production of novel metabolites. In order to test the applicability of metabolically engineering soybean seed coats, we sought to introduce a novel metabolic pathway. We designed a construct that contained the genes encoding the three enzymes that form the PHB biosynthetic pathway. This pathway was selected because it is relatively simple, requiring the introduction of only three new genes, and it has previously been successfully introduced into several plant systems (Table 1.1 and references therein).

5.2 Cloning and plant transformation

A single construct was created containing all three of the genes encoding the enzymes necessary for the biosynthesis of PHB. Previous work in *Arabidopsis* has indicated that PHB production could be maximized by including the three genes in a single construct as opposed to individually transforming plants with each gene and then crossing them to create a line with all three genes (Bohmert et al. 2000; Nawrath et al. 1994). The three *phb* genes are bacterial in origin and, therefore, likely contain a codon bias that is different from that of soybean genes, which could potentially reduce the rate at which the mRNA can be translated (Gustafsson et al. 2004). The process of codon optimization changed the codon bias of the *phb* genes to more closely resemble that of soybean genes, a process which has been shown to enhance expression

in plant systems (Horvath et al. 2000; Batard et al. 2000; Gustafsson et al. 2004). Each *phb* gene was flanked by sequences encoding the SBP N-terminal propeptide followed by the first four amino acids of mature SBP and the SBP C-terminal propeptide preceded by the last four amino acids of mature SBP. The additional amino acids from mature SBP were included in order to conserve the site of cleavage of the propeptides. Each of the SBP propeptide signals was separated from the PHB enzyme by a small linker peptide so as to facilitate formation of tertiary structure of the enzyme and thereby maintain its function (F. Brandizzi, personal communication).

Each of the three genes was flanked by the *Ep* 1.5 promoter and the *nos* transcriptional terminator (Figure 5.1). Work done by Dr. L Gudynaitę-Savitch to optimize transgene design suggested that the optimal arrangement of the T-DNA is with the selectable marker promoter and *Ep* 1.5 promoter in a head-to-head orientation thus placing both promoters as far as possible from the T-DNA borders. In addition to this manipulation, a 2.5 kb DNA fragment from the LacZ gene was inserted between the two promoters and the CaMV 35S promoter driving expression of the selectable marker was replaced with the *entCUP4* promoter (Malik et al. 2002). The three *phb* gene cassettes in plasmid J13 were all positioned in a head-to-tail orientation in order to separate the three *Ep* 1.5 promoters, provide the maximum distance between the right T-DNA border and each of the promoters, and to minimize silencing that can occur when duplicate sequences are introduced as an inverted repeat (Stam et al. 1997).

The J13 plasmid was introduced into soybean cv. Jack by particle bombardment, and T_0 plants were regenerated and grown to maturity to produce T_1 seeds. T1 seeds were germinated on moistened filter paper or in moistened vermiculite and after 4 to 6 days seed coats were collected for analysis and seedlings were transplanted into soil. Seed coats are maternal in origin, therefore

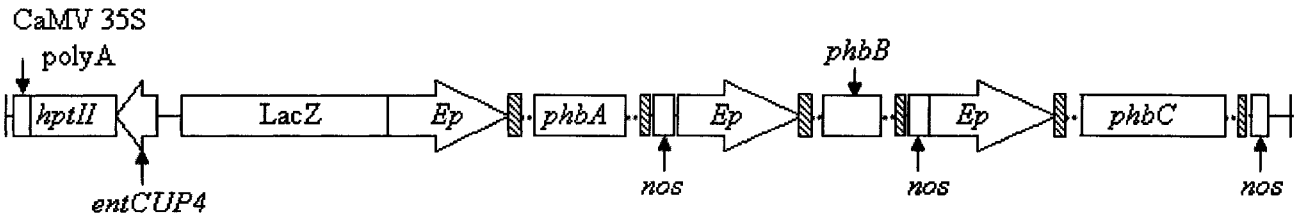


Figure 5.1 Schematic diagram illustrating the T-DNA region of the J13 construct containing the *phb* genes for soybean transformation. This construct contains the *entCUP4* promoter driving expression of the *hptII* selectable marker gene. The *Ep* 1.5 promoter (*Ep*) was used to drive expression of each of the *phb* genes. Each *phb* gene was modified to include the sequences encoding the SBP N- and C-terminal propeptides (hatched boxes) 5' and 3' to each gene, respectively. Sequences encoding a linker peptide, represented by dotted lines, were included between each of the propeptide sequences and the *phb* genes (not drawn to scale). Following all three *phb* genes is the *nopaline synthase* transcriptional terminator (*nos*). Black bars on left and ride sides of constructs represent left and right T-DNA borders, respectively.

the seed coats from T_1 seeds are from the T_0 generation and seed coats from the T_2 seeds are from the T_1 generation (Figure 2.1).

Plants from five independent transformation events were regenerated; however, one transformation event, represented by a single plant, did not survive to produce seeds leaving a total of four independent transformation events. One transformation event yielded two plants, both of which produced seeds that were analyzed. A second transformation event yielded 9 plants that survived to produce seeds, and 3 of these plants were analyzed. The third transformation event yielded 5 plants that survived to produce seeds, and four of these were analyzed. The final transformation event yielded a single plant that survived to produce seeds, which was also analyzed. In addition, seed coats from 1 T_1 plant from the first transformation event, 1 T_1 plant from the second transformation event, 3 T_1 plants from the third transformation event, and 2 T_1 plants from the fourth transformation event were also analyzed.

5.3 Confirmation of transgene insertion

Transgenic soybean plants were screened by PCR to confirm the presence of all of the elements of the T-DNA. PCR was done on genomic DNA template with primers that specifically amplify the *hptII* gene encoding hygromycin resistance, the *entCUP4* promoter, the three *phb* genes, and the three *Ep* 1.5 promoters. In addition, the *nos* transcriptional terminator was also amplified; however, the primers used were not able to distinguish between the *nos* regions from the three different *phb* gene cassettes. The *β-tubulin-2* gene was also amplified to confirm the quality of the genomic DNA extracts. J13 plasmid was used as a positive control and DNA from untransformed soybean plants cv. Jack was used as a negative control. For the T_0 generation, all of the T-DNA elements were present in each of the plants analyzed, as determined by PCR, with

the exception of plant 2F for which the analysis was incomplete. However, since all T-DNA elements were present in one of the progeny from this plant, the T_0 plant likely contained all of the T-DNA elements. In the T_1 generation, 10 plants were screened by PCR, and 3 were found to be negative for all T-DNA elements, likely as a result of segregation or possibly transgene instability (Figure 5.2). A second genomic DNA extract was prepared from those three plants to confirm the absence of insert. Plants that lost the insert were not further analyzed. Representative PCRs from T_1 plants are shown in Figure 5.3.

5.4 Analysis of PHB in soybean seed coats

PHB was extracted with chloroform, transesterified to form the ethyl ester of 3-HB, and derivatized according to the method of Bohmert et al. (2000). However, the procedure was simplified by omitting the initial washes with ethanol and methanol. Despite omitting these washes, the region around the peak corresponding to the ethyl ester of 3-HB was not contaminated by other compounds, allowing accurate analysis of this peak.

The seed coat chloroform extracts were analyzed by gas chromatography. Benzoic acid was added to all extracts at a final concentration of 0.18 mg/ml to serve as an internal standard. A comparison of control samples with and without the internal standard revealed a unique peak at a retention time of approximately 2.85 minutes in the sample containing benzoic acid. For each plant, three extracts were made, each from an individual seed coat. The one exception to this was the plant 2F, for which only two seed coats were available for analysis. A peak was detected at a retention time of 1.95 minutes in several of the transgenic seed coats (Figure 5.3A), but was not present in wild-type cv. Jack seed coats (Figure 5.3B, red line). A unique peak with the same retention time was identified in control samples prepared with commercial PHB (Figure

Figure 5.2 Amplification of transgene elements from T_1 soybean genomic DNA extracts. The elements encoding the *hptII* and *phb* genes, the *entCUP4* and *Ep* 1.5 promoters (*Ep*), and the *nopaline synthase* transcriptional terminator (*nos*) were amplified separately by polymerase chain reaction. As a control for genomic DNA quality the endogenous *β-tubulin-2* gene was amplified. Positive control (+) is J13 plasmid DNA. As a negative control, genomic DNA from wild-type cv. Jack was used as template. The plants 1C-12.2, 1C-23.1, and 1E-35.1 have lost the transgene. The primers used are listed in Table 2.1 and conditions for amplification of each element in Table 2.2.

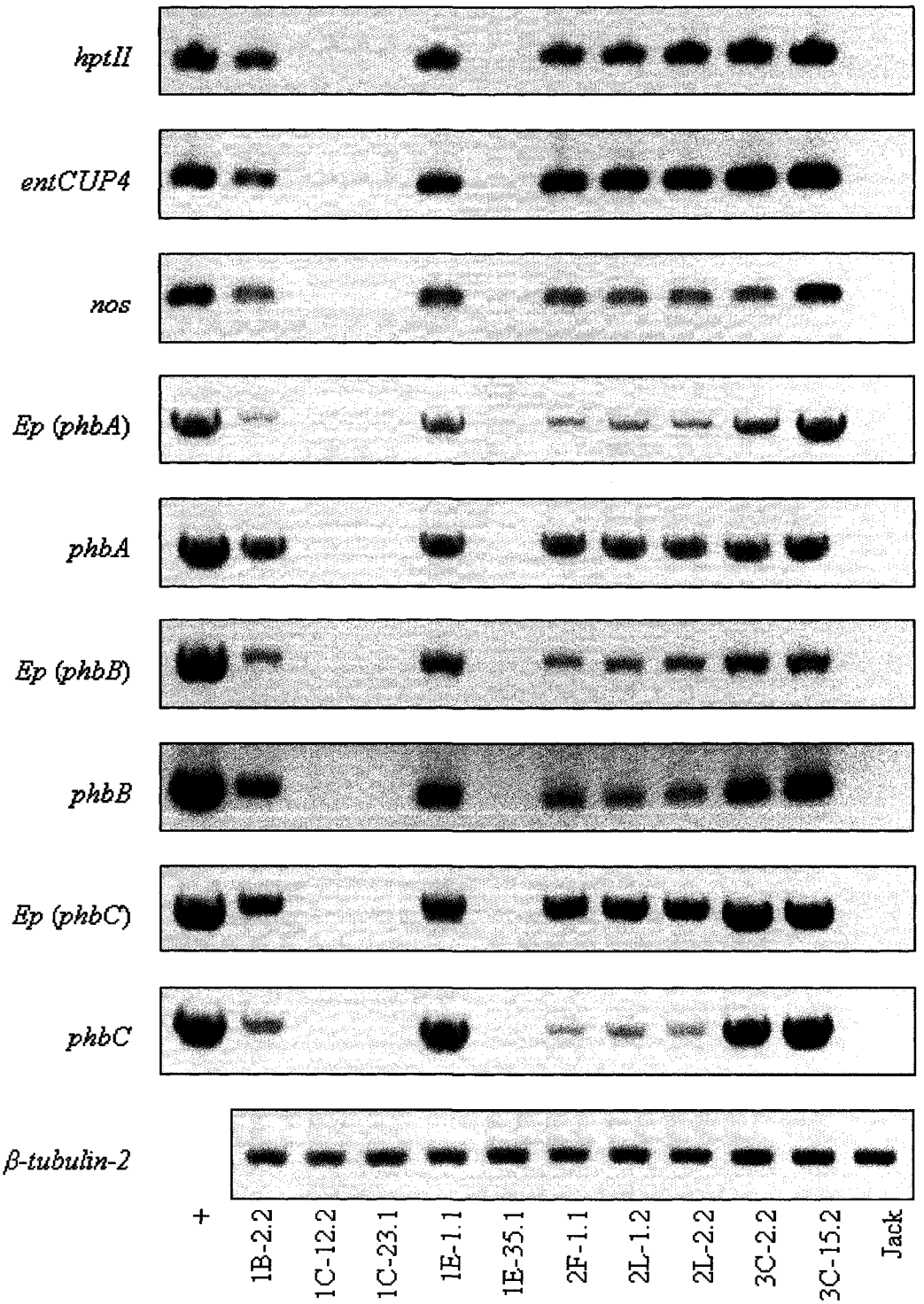
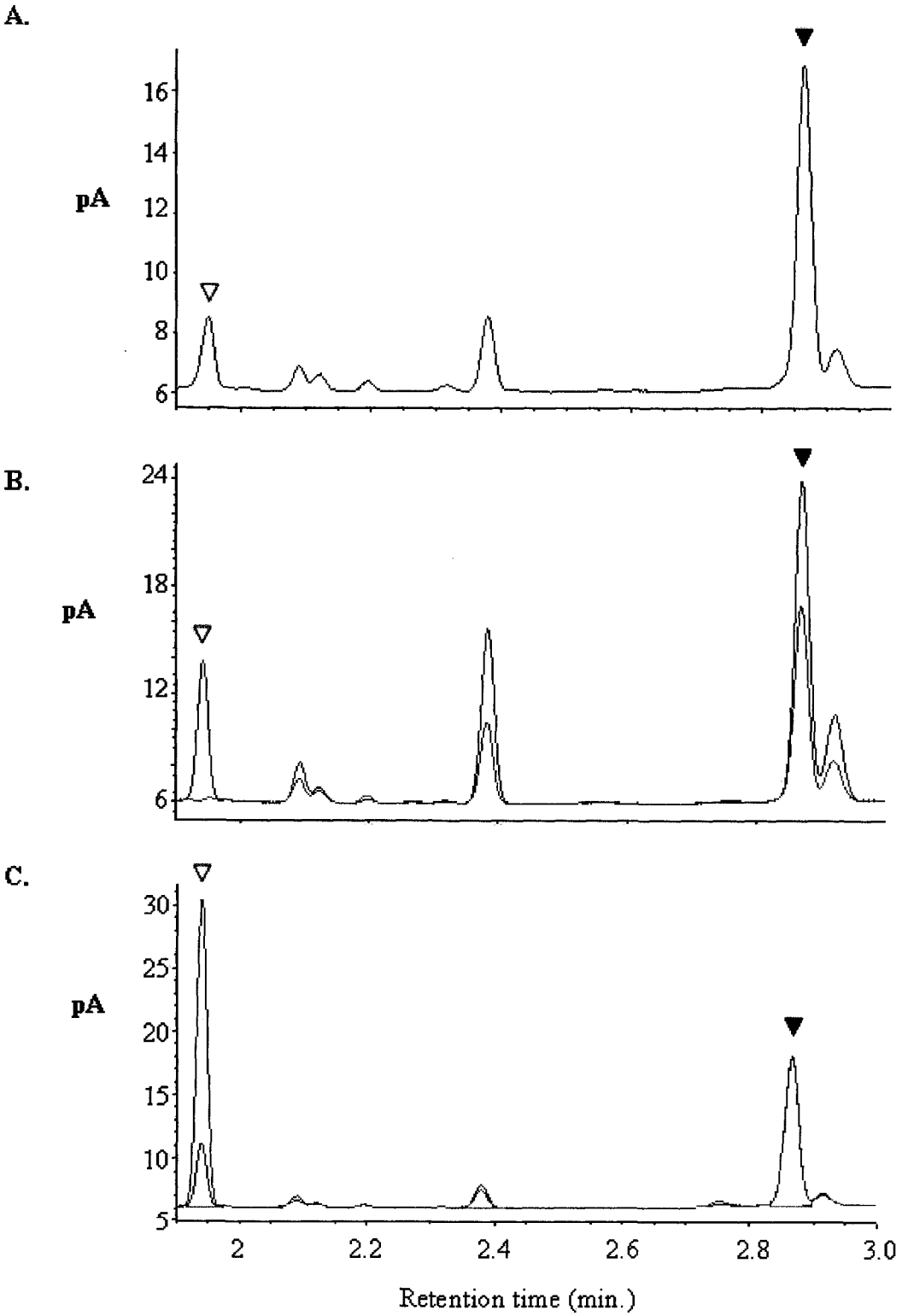


Figure 5.3 Gas chromatograms of chloroform extracts prepared from **A.** a seed coat from the transgenic plant 1A, **B.** a wild-type cv. Jack seed coat (red line) and a wild-type cv. Jack seed coat spiked with commercial PHB at a concentration of 0.1 mg/ml (blue line), overlayed, and **C.** commercial PHB at a concentration of 0.5 mg/ml (red line) and 0.1 mg/ml (blue line), overlayed. Chromatograms show flame ionization detector response in picoamperes (pA) versus retention time in minutes (min). Open arrowheads indicate the peaks corresponding to the ethyl ester of 3-HB and closed arrowheads indicate the peaks corresponding to the internal standard benzoic acid (0.18 mg/ml).



5.3C) as well as in a chloroform extract from a wild-type seed coat spiked with PHB (Figure 5.3B, blue line) and this peak was therefore assumed to represent the ethyl ester of 3-HB. The amount of 3-HB ethyl ester in each final extract, which we assume to be reflective of the original amount of PHB extracted, was quantified by the internal standard method (Chang and Karr 1959). The mean PHB content expressed as percentage of seed coat DW for all plants analyzed is shown in Figure 5.4.

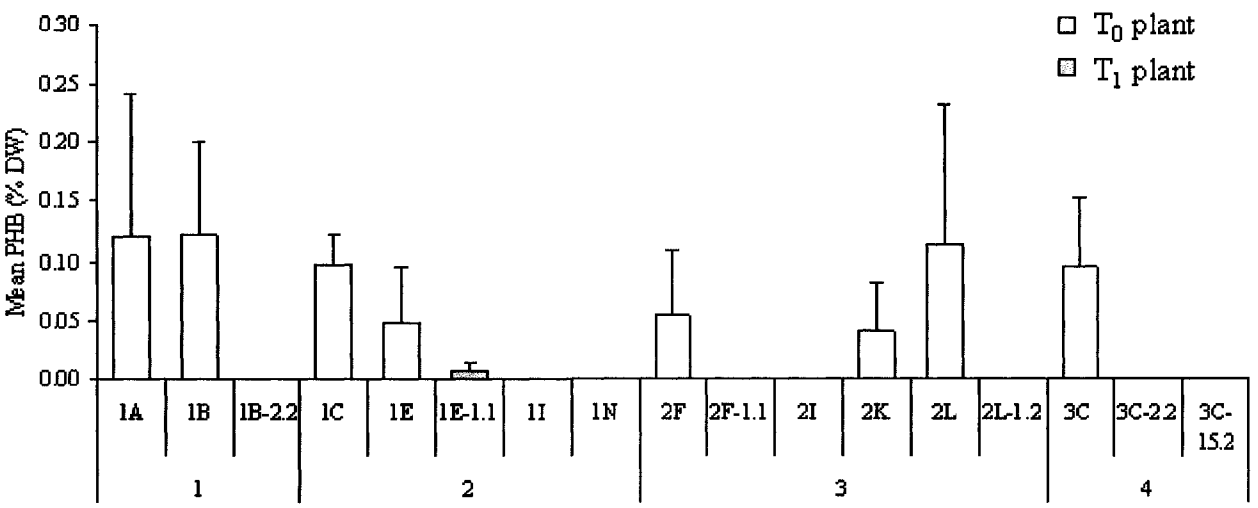
The amount of PHB detected in seed coat extracts was low, with the maximum amount detected in a single seed coat being 0.36% DW. Many seed coats did not have any detectable PHB. These results demonstrate that it is possible to metabolically engineer soybean seed coats to produce a novel biochemical. However, there do appear to be limitations to the amount of biochemical produced. Variability in the level of PHB was observed both between seeds from a single plant and between plants from the same transformation event, although there was no appreciable difference between plants from different transformation events.

One possible explanation for the low levels of PHB produced is that the PHB had a negative impact on seed health, which limited the level to which it could be produced in seed coats. Overall, plants were similar in appearance to wild-type cv. Jack plants grown at the same time. Pod production, depending on whether plants were primarily grown with 16 h or 14 h day length, was typically in the range of 20 to 200 pods per plant.

Some plants produced seeds that showed a poor rate of germination (data not shown), which could be indicative of PHB inhibiting germination. For example, seeds from the plant 1I, which was otherwise normal in appearance, had a rate of germination of 7.5% (n=40). Production of PHB in tobacco also had an impact on seed germination (Arai et al. 2004). If higher levels of PHB prevented germination, the results would be skewed for seed coats

Figure 5.4 Mean PHB content in soybean seed coats from plants transformed with the PHB biosynthetic pathway, expressed as percentage of seed coat dry weight (% DW). The amount of PHB was measured by gas chromatography using the internal standard method, as described in section 2.13. **A.** Mean PHB + SE of plants from four independent transformation events. Sample size is 3 for all plants except for 2F, which had a sample size of 2. **B.** Table showing the mean ± SE, maximum, and minimum seed coat PHB content for each of the plants analyzed, expressed % DW. Also indicated is sample size (n).

A.



B.

Event	Plant	Generation	Mean	n	Maximum	Minimum
1	1A	T_0	0.12 ± 0.12	3	0.36	0.00
	1B	T_0	0.120 ± 0.08	3	0.27	0.00
	1B-2.2	T_1	0.00	3	0.00	0.00
2	1C	T_0	0.10 ± 0.02	3	0.14	0.06
	1E	T_0	0.05 ± 0.05	3	0.14	0.00
	1E-1.1	T_1	0.01 ± 0.01	3	0.02	0.00
	1I	T_0	0.00	3	0.00	0.00
	1N	T_0	0.00	3	0.00	0.00
3	2F	T_0	0.05 ± 0.05	2	0.10	0.00
	2F-1.1	T_1	0.00	3	0.00	0.00
	2I	T_0	0.00	3	0.00	0.00
	2K	T_0	0.04 ± 0.04	3	0.12	0.00
	2L	T_0	0.12 ± 0.12	3	0.35	0.00
	2L-1.2	T_1	0.00	3	0.00	0.00
4	3C	T_0	0.09 ± 0.06	3	0.20	0.00
	3C-2.2	T_1	0.00	3	0.00	0.00
	3C-15.2	T_1	0.00	3	0.00	0.00

containing lower PHB amounts, since the seed coats analyzed primarily came from seeds that germinated. Two of the seed coats with higher PHB levels (0.36 and 0.27% DW) came from seeds that did not germinate, although one seed containing PHB at 0.35% DW did germinate. In addition, a few plants appeared to have a higher than normal rate of seed abortion (data not shown), such as the T_1 plant 2L-1.2, which had 42 pods containing at least one aborted seed out of a total of 104 pods. This plant was otherwise normal in appearance. Seed abortion was observed in other plants, including wild-type plants, but it was typically not as severe.

Alternatively, the low levels of PHB could result from low level expression of the biosynthetic enzymes, which, as discussed in Chapter 4, appears to be a limitation of the *Ep* 1.5 promoter. Another limitation to the production of PHB in seed coats could be the amount of acetyl-CoA within the vacuole, since this is the starting metabolite required for the production of PHB. These possibilities will all be further examined in the Discussion.

CHAPTER 6 DISCUSSION

6.1 Development of a soybean seed coat expression system

Our goal was to develop a plant system for the high level synthesis of bioproducts with the aim of adding value to soybean, an important Canadian crop. Soybean seed coats are a low-value byproduct of the soybean industry and are removed during processing. This makes the soybean seed coat an ideal tissue for adding value to soybean crops.

The system we have developed consists of several components.

1. Large quantities of the enzyme SBP accumulate primarily in the seed coat with only a small amount being produced in root tissues (Gijzen et al. 1993). The *Ep* promoter that regulates expression of SBP can therefore be used to drive high levels of seed coat-specific expression of foreign genes. Although the promoter is not fully characterized, we chose to use the 1.5 kb region upstream from the *Ep* gene as our promoter, designated the *Ep* 1.5 promoter, assuming that fragment was likely to contain the elements necessary both for high level and seed coat-specific expression. It does regulate seed coat-specific expression of GUS in transgenic plants (D. Simmonds; L Gudynaitę-Savitch, unpublished).

2. Several naturally-occurring soybean cultivars, designated *epep*, have a deletion at the *Ep* locus that results in greatly reduced expression of the *Ep* gene (Gijzen 1997). Since high level expression of foreign proteins can impact plant health, we chose to transform an *epep* cultivar, reasoning that with the absence of SBP these seed coats may have a greater capacity to host large amounts of protein without affecting plant health.

3. SBP contains N- and C-terminal propeptides that were predicted to direct vacuolar targeting (Gijzen 1997). Subcellular targeting of proteins can influence their level of expression, and therefore the addition of N- and C-terminal propeptides to our transgenes could contribute to

their high level expression as well targeting their product to the vacuole. However, it is possible that targeting proteins to other subcellular compartments would provide a better environment for high level and stable accumulation of foreign proteins. To investigate this possibility, I compared the yield of a foreign protein targeted to different compartment of the plant secretory system in the model plant *A. thaliana*. The results of this study will be discussed in Section 6.2.

In order to determine if introducing novel genes under control of the *Ep* 1.5 promoter has the potential to produce expression levels similar to what is naturally achieved in soybean seed coats, the first protein I chose to produce in soybean seed coats was SBP. By reintroducing the *Ep* cDNA into an *epep* background, the levels of expression of the *Ep* transgene can be directly compared to levels naturally produced in *EpEp* cultivars. Since SBP is an enzyme with a number of industrial applications, this approach would also demonstrate the ability of soybean seed coats to produce such bioproducts. By further enhancing levels of expression as well as modifying SBP to enhance its activity, the value of soybean seed coats could be further enhanced through the production of SBP. The results from this study will be discussed in Section 6.3.

To investigate the feasibility of altering the metabolism of soybean seed coats to produce other non-protein biochemicals, I chose to introduce the biosynthetic pathway for the biodegradable plastic PHB. This relatively simple, three enzyme pathway converts acetyl-CoA into PHB and has been successfully introduced into a number of plant species (Table 1.1 and references therein). The results of this study will be discussed in Section 6.4.

6.2 Targeting foreign proteins in the plant secretory system to increase accumulation

Several studies have demonstrated that the yield of foreign proteins expressed in plants systems can be increased through protein targeting to subcellular compartments (Hood et al.

2003; Ramírez et al. 2002; Yang et al. 2005). It was suggested that SBP possesses two targeting signals that direct it to the vacuole in the seed coat (Gijzen 1997) that could be used to target heterologous proteins to the vacuole and improve their yield and stability. Targeting of foreign proteins in plant systems was investigated in the model system *A. thaliana* with two main objectives. First, to determine if the SBP propeptides are functional targeting signals, as predicted. And second, to compare the yield of a foreign protein following targeting to different compartments of the plant secretory system.

To investigate targeting, a gene encoding a fusion protein consisting of mature SBP followed by GFP was designed and sequences encoding the N- and C-terminal SBP propeptides as well as an ER-retention signal were added in different combinations to target the protein to the apoplast, ER, and vacuole. GFP was intended to function as a marker of cellular location and SBP activity was used as a marker of protein yield.

6.2.1 Characterization of antibodies for immunoblot analysis

A polyclonal antibody against SBP was raised in rabbit. This antibody can detect commercial SBP at levels as low as 1 ng. In soybean seed coat extracts from *EpEp* cultivars, it detects a band of the expected size of 43 kD. When large amounts of SBP are present, minor bands are also detected above and below the 43 kDa. The lower molecular weight bands likely represent breakdown products of SBP as they are detected both in purified SBP from a commercial source (Figure 3.1A) and in a crude extract from mature Harosoy 63 seed coats (Figure 3.1B, lane 3). The origin of the larger bands is unclear. While SBP is heterogeneously glycosylated, this would likely only account for minor fluctuations in the size (Gray and Montgomery 2006). One possible explanation is that they represent aggregation products that resisted separation even under the denaturing conditions used during SDS-PAGE (Carroll et al.

1997; Howell et al. 2001; Marolda et al. 2004). Alternatively, these bands may be non-specific reactions to other proteins found in soybean seed coats. However, these bands are not detected in Jack seed coat extracts (Figure 3.1B, lane 4), and therefore would need to be specific to the Harosoy 63 seed coats.

The anti-SBP antibody was produced using the mature SBP protein. In plants, there are typically multiple class III peroxidases that have the potential to cross-react with the anti-SBP antibody. In *Arabidopsis*, 73 genes and over 100 ESTs have been identified encoding class III plant peroxidases, and these are expressed in a wide range of tissues, including roots, stems, flowers, and leaves (Hiraga et al. 2001; Tognolli et al. 2002; Valério et al. 2004; Welinder et al. 2002). In addition to SBP, other peroxidases can be identified in soybean seed coats based on 4-chloro-1-napthol staining of isoelectric focused polyacrylamide gels, although these are minor compared to SBP since it required loading 50 times the amount of protein to detect them by staining (Gillikan and Graham 1991). In addition, peroxidases have been identified in soybean roots, leaves, and cotyledons, although some of these are only expressed in response to wounding or infection (Gijzen et al. 1993; Graham and Graham 1991; Cakmak and Horst 1991; Felton et al. 1994). Furthermore, other classes of peroxidases are also expressed in soybean plants, such as the cytosolic ascorbate peroxidases identified in leaves (Caldwell et al. 1998).

These peroxidases could potentially cross-react with the anti-SBP antibody, limiting its effectiveness at detecting SBP. Peroxidases share a low degree of identity within an individual plant species but may show a much higher degree of similarity across multiple species (Hiraga et al. 2001). SBP is most closely related to four peroxidases from alfalfa with percent identity of their amino acid sequences in the range of 65-67% (Gijzen 1997). However, antibodies typically recognize short stretches of amino acids and therefore cross-reactivity can arise so long as there

are small stretches of identical amino acids. The anti-SBP antibody showed some cross-reactivity to HRP C with anti-SBP antibody able to detect commercial HRP C at levels of 50 ng or higher. An alignment of the mature sequences of SBP and HRP C demonstrated an overall sequence identity of 58%; however, several smaller regions in the two proteins that share approximately 80% identity and that typically contain stretches of 8 to 10 identical amino acids (data not shown) can be found. These regions are most likely to be responsible for the cross-reactivity of anti-SBP antibody to HRP C. Despite the potential for anti-SBP antibody to cross-react with endogenous plant peroxidases in *Arabidopsis* and soybean, no significant reactions were detected in wild-type tissues (Figure 3.1B, lane 4, and Figure 3.1C, lane 2).

In order to detect the second part of the SBP-GFP fusion protein, anti-GFP antibody was purchased. This antibody was able to detect recombinant GFP at the expected size (Figure 3.1D, lanes 1 to 3) and did not detect any non-specific bands in wild-type *Arabidopsis* leaf extracts (Figure 3.1 D, lane 4). The anti-GFP antibody detected the recombinant GFP as a double band. This has been documented previously and may depend upon sample preparation (Gustafson et al. 1998; Batoko et al. 2000). Three antibodies were also obtained to function as subcellular markers. These are anti-cFBPase, a cytoplasmic marker, anti-BiP, an ER marker, and anti-V-ATPase, a marker of the tonoplast which will function as a vacuolar marker. Each of these antibodies specifically detected appropriately sized proteins (Figure 3.1E), indicating that they will function appropriately as subcellular markers.

6.2.2 Subcellular localization of the SBP-GFP fusion proteins

In order to determine the subcellular location of each of the fusion proteins, GFP fluorescence in *Arabidopsis* leaves was imaged by laser-scanning confocal microscopy. Confocal imaging of GFP fluorescence from the eSBP-GFP protein revealed a reticulate pattern that is

characteristic of the ER (Figure 3.4A), suggesting that eSBP-GFP is accumulating in the ER. eSBP-GFP contains the SBP N-terminal propeptide which was predicted to direct entry of the fusion protein into the ER. It additionally contains an ER-retention signal which functions by retrieving proteins exported from the ER. Therefore, eSBP-GFP was predicted to be localized in the ER, and this was confirmed by GFP imaging.

Confocal imaging of GFP fluorescence from the aSBP-GFP protein did not yield any detectable fluorescence in the apoplast. Immunoblotting subsequently indicated that this was likely a result of degradation of the GFP portion of the fusion protein (Figure 3.7). As an alternative approach, subcellular fractionation was performed to isolate the apoplast and intracellular fractions. This was done by infiltrating *Arabidopsis* leaves with buffer and then gently centrifuging them to collect the buffer along with the apoplast fraction, which is a common approach for isolating this fraction (Firek et al. 1993; Boudart et al. 2005). Remaining tissue represents the intracellular fraction. The cytoplasmic marker cFBPase was associated with the intracellular fraction, and its presence in the apoplast fraction indicated intracellular contamination. This approach demonstrated that aSBP-GFP was predominantly associated with the apoplast in fractions free from the cytoplasmic contaminant cFBPase (Figure 3.5A). aSBP-GFP was also typically detected in the intracellular fraction, although at a lower concentration. This could represent aSBP-GFP that is in transit through the secretory system, either in the ER or being transported to the plasma membrane for secretion (Denecke et al. 1990; Di Sansebastiano et al. 1998; Matsui et al. 2006).

An alternative explanation is that the infiltration and centrifugation of *Arabidopsis* leaves did not result in the complete isolation of all of the aSBP-GFP found in the apoplast. Secreted proteins may be bound to cell wall components through a number of interactions of varying

strength, limiting the extent to which they can be extracted (Lee et al. 2004). The composition of the infiltration buffer, particularly the presence of different salts and chelating agents, can influence the amount and type of proteins isolated from the apoplast (Boudart et al. 2005). We included 0.2 M $CaCl_2$ in our infiltration buffer, which has been shown to be particularly effective at extracting apoplast proteins (Boudart et al. 2005). In tobacco plants expressing HRP C, the recovery of peroxidase activity in the apoplast fraction was improved by infiltrating leaves with $CaCl_2$ but a portion of peroxidase activity remained associated with the cell wall and could only be released by its digestion, suggesting a very strong association with cell wall components (Kis et al. 2004). Therefore it is possible that some aSBP-GFP may have remained with the intracellular fraction as a result of incomplete disassociation from the cell wall.

These results suggest that aSBP-GFP is primarily secreted in *Arabidopsis* leaves. The only targeting information in aSBP-GFP is the SBP N-terminal propeptide. This propeptide is expected to function by directing co-translational import into the ER (Vitale and Hinz 2005). In the absence of additional targeting information, the fusion protein would be secreted outside of the cell along the default pathway (Denecke et al. 1990). Therefore, as predicted, aSBP-GFP is primarily secreted, which confirms that the N-terminal propeptide is a functional targeting signal.

Most *Arabidopsis* leaves from plants expressing *vSBP-GFP* did not have a detectable GFP signal. In two plants, fluorescence was localized throughout the cell (Figure 3.4B), suggestive of a large central vacuole, but this pattern was not apparent in any other plants. Imaging of GFP in plant vacuoles is a common problem (Di Sansebastiano et al. 2004; Tamura et al. 2003). In *Arabidopsis*, light-dependent degradation of GFP at acidic pH has been observed (Tamura et al. 2003). Typically, this can be prevented by placing plants in the dark for 2 days prior to GFP imaging. Incubation in the dark prior to GFP imaging did not improve our results

nor did incubation of plant tissues with the cysteine proteinase inhibitor E64-d (results not shown). Immunoblots revealed that the GFP portion of vSBP-GFP was being degraded (Figure 3.7), as will be discussed in Section 6.2.5.

To further investigate the subcellular location of the vSBP-GFP fusion protein, vacuoles were isolated from *Arabidopsis* leaves (Matsui et al. 2006). The identity of this fraction was confirmed by immunoblotting with the tonoplast marker anti-V-ATPase. To confirm that this fraction was not contaminated by other membrane compartments, it was also immunoblotted with the ER marker anti-BiP, which could not be detected. SBP-GFP was concentrated in the vacuole fraction (Figure 3.5B), suggesting that vSBP-GFP is targeted to the vacuole. vSBP-GFP contains the SBP N-terminal propeptide, which is functioning as a signal directing entry into the plant secretory system. In addition, it contains the SBP C-terminal propeptide, which was predicted to function as a vacuolar sorting signal (Gijzen 1997). The localization of vSBP-GFP in the vacuole, as predicted, confirms that in *Arabidopsis* leaves, the C-terminal propeptide is functioning as a vacuolar sorting signal, as well as providing further evidence that the N-terminal propeptide is a functional targeting signal.

The confocal imaging and subcellular fractionation work indicates that each of the three targeted fusion proteins are being directed to the expected subcellular location. Due to the low level of expression of cSBP-GFP, we were unable to confirm its location in the cytoplasm. This work also supports the hypothesis that the SBP N-terminal propeptide can function as an N-terminal signal peptide and that the SBP C-terminal propeptide can function as a vacuolar sorting signal. The targeting signals are functional in a heterologous system, indicating functional conservation which is often observed for plant targeting signals (Bednarek and Raikhel 1991;

Brown et al. 2003; Frigerio et al. 1998). These results suggest that these signals will be able to direct vacuolar localization of foreign proteins in other plant systems in addition to soybean.

During the development of the vacuole isolation protocol, I attempted to concentrate vacuoles by centrifugation followed by resuspension in a smaller volume. Immunoblotting revealed the presence of only the vacuolar marker V-ATPase in the pellet while vSBP-GFP and V-ATPase were both present in the supernatant. One possible explanation for this observation is that centrifugation resulted in vacuole lysis and the pellet that was collected consisted only of tonoplast while the lumenal proteins were released in the supernatant. Since V-ATPase is specifically a marker of the tonoplast, it would associate with the pellet. However, vSBP-GFP, which is predicted to be soluble and therefore in the vacuole lumen, would only be detected in the supernatant. V-ATPase was also detected in the supernatant and could originate from unlysed vacuoles or tonoplast remnants that remained unpelleted or from fragments of V-ATPase cleaved from the tonoplast.

An alternative hypothesis is that the vacuole fraction represented a heterogeneous population of vacuoles, and only the larger and/or denser vacuoles were pelleted by centrifugation. SBP would therefore be associated with the smaller and/or less dense population of vacuoles that were not pelleted at the selected speed. Different types of vacuoles have been found to coexist within a single cell (Park et al. 2004; Di Sansebastiano et al. 1998; Paris et al. 1996). The lack of a consensus sequence in the SBP C-terminal propeptide and its position at the C-terminus suggests that it functions as a ctVSS, which are typically found in seed storage proteins such as β-phaseolin, the α'-subunit of β-conglycinin, and the 2S albumins of *Bertholletia excelsa*, *Ricinus communis*, and *Curcubita maxima* that are targeted to PSVs (Vitale and Hinz 2005). In a homogenate of pumpkin seeds, PSVs, which are typically smaller than LVs,

were pelleted at a speed of 41,000 g (Jiang et al. 2000). In contrast, I centrifuged the vacuolar preparation at 21,000 g to pellet the vacuoles. Thus, this speed may not have been great enough to completely pellet smaller vacuoles such as PSVs.

If this explanation is correct, it would suggest that the vSBP-GFP protein is targeted to a smaller vacuole that is more consistent with a PSV. However, this would contradict what was seen with GFP imaging of plants containing the *vSBP-GFP* gene, although it is possible that vSBP-GFP could be associated with both types of vacuoles. This has been observed in tobacco protoplasts, where GFP accumulated in different types of vacuoles depending on the type of protoplast cell (Di Sansebastiano et al. 1998). While the results of subcellular fractionation confirm that the SBP C-terminal propeptide does function as a vacuolar sorting signal, they are insufficient for determining if the signal is specific for a single vacuole type.

6.2.3 Impact of subcellular location on foreign protein yield

Peroxidase activity in leaves of plants containing *aSBP-GFP*, *eSBP-GFP*, and *vSBP-GFP* was significantly higher than activity in plants containing the non-targeted *cSBP-GFP* (Figure 3.6A). There was no significant difference in the peroxidase activity levels found between the three different targeted fusion proteins. This suggests that targeting to the secretory system, regardless of the final destination, is important for increasing the yield of foreign proteins.

Variability was observed in peroxidase activity between plants containing the same construct (Figure 3.6). Variability in transgene expression is a common observation (Hobbs et al. 1990; Schubert et al. 2004; Peach and Velten 1991; Gendloff et al. 1990) and can be attributed to a number of different factors. In some cases, a higher number of transgene inserts has been found to be correlated with higher levels of expression (Finnegan and McElroy 1994; Gendloff et al. 1990) and this could account for some of the observed variability. Variability is also often

attributed to position effects. The site of transgene integration influences the level of transgene expression due to the influence of regulatory elements near the site of integration or from insertion in a region of a chromosome that does not favor high levels of gene expression (Matzke and Matzke 1998; Wilson et al. 1990). There are a number of examples of one promoter or regulatory region influencing the expression of a second nearby promoter region (Eszterhas et al. 2002; Yoo et al. 2005; Hampf and Gossen 2006), and this phenomenon has been adapted into the technique known as promoter or enhancer trapping (Campisi et al. 1999; Springer 2000). Integration in telomeres and centromeres as well as intercalary heterochromatin regions appears to be associated with variable levels of gene expression, although transgenes with more stable expression levels have also been found near telomeres, possibly because gene-rich regions tend to be found near the distal ends of chromosome arms (Matzke and Matzke 1998; Iglesias et al. 1997; Jakowitsch et al. 1999). In addition, location of a transgene adjacent to an AT-rich region that may function as a nuclear matrix attachment region is also associated with stable levels transgene expression (Iglesias et al. 1997).

Alternatively, host gene-silencing of transgene expression could contribute to the high variability that is observed in transgene expression (Stam et al. 1997). Transgene silencing is a homology-based effect that can influence expression at the transcriptional or the post-transcriptional level of gene expression (Matzke et al. 2000). With *Agrobacterium* transformation systems, in many cases, the T-DNA insertion is accompanied by sequence from the binary vector outside of the T-DNA left and right borders and this sequence has been associated with transgene silencing (Iglesias et al. 1997; Jakowitsch et al. 1999). Also associated with transgene silencing is the pattern of transgene insertion with multiple, incomplete copies leading to silencing and single, complete insertions being less affected (Iglesias et al. 1997;

Jakowitsch et al. 1999). Some studies have indicated transgene silencing to be the predominant source of transgene expression variability (Schubert et al. 2004; Hobbs et al. 1990). While increasing gene copy number may correlate with increased transgene expression in a dosage effect, it may also correlate with decreased transgene expression as a result of gene silencing (Gendloff et al. 1990; Elomaa et al. 1995; Finnegan and McElroy 1994; Peach and Velten 1991). This was observed in tobacco plants transformed with the *uidA* gene and was associated with transgene methylation, a marker of host silencing (Hobbs et al. 1990). In a large-scale study conducted in *Arabidopsis* involving several different transgenes, a gene-specific transcript-dosage effect was observed in which increased copy number was associated with increased expression up until a threshold transcript level was reached, after which expression was silenced (Schubert et al. 2004).

The variability observed in the level of peroxidase activity between different plants transformed with the same construct is most likely attributable to some of the above effects. However, by using a large number of plants in this analysis, we can be confidant that the average specific activity for each construct is reflective of the influence of the targeting signals on protein accumulation. Alternatively, an examination of the highest expressing plant produced for each construct can provide an idea of the maximum attainable level of expression for each subcellular location. The maximum level of activity observed in a single plant (mean of three extracts) containing the non-targeted construct was 1.51 U mg^{-1} protein while the maximum activity observed for a single plant containing the targeted constructs was 10.64, 6.80, and 8.49 U mg^{-1} protein for aSBP-GFP, eSBP-GFP, and vSBP-GFP, respectively. These values provide additional support for the conclusion that targeting results in an increase in protein accumulation.

These results suggest that targeting foreign proteins in the soybean seed coat system will be an effective means of increasing their yield. In addition to targeting foreign proteins to the vacuole, targeting them to the ER and the apoplast can additionally increase their yield.

6.2.4 Effect of growth conditions on foreign gene expression

Interestingly, a different pattern of accumulation of the SBP-GFP fusion proteins was observed when plants were grown under long day conditions (16 h day/8 h night) rather than the conditions used above (12 h day/12 h night). Under these conditions, peroxidase activity in plants containing *vSBP-GFP* was significantly higher than all other plant lines, with the exception of plants containing *eSBP-GFP*. This analysis was based on a single extract from each plant, since long day plants typically yield much less leaf material.

Photoperiod is an important cue that regulates both the juvenile to adult phase change and floral transition (Bäurle and Dean 2006). In *Arabidopsis*, longer photoperiods result in a more rapid transition through the three main phases. Photoperiod may also serve as a signal for a number of other processes in plants. In poplar, short days result in the accumulation of the mRNA encoding bark storage protein, while long days correspond to a decline in this mRNA (Coleman et al. 1993). The rate of starch degradation in *Arabidopsis* was found to be responsive to photoperiod, as were the levels of transcripts of starch-degrading enzymes (Lu et al. 2005). In cucumber, the rate of root exudation increased with increasing day length (Pramanik et al. 2000). These studies demonstrate that photoperiod can impact a number of physiological processes in plants. In our study, photoperiod could have induced some physiological processes that differentially affected protein accumulation in one or more compartments. For example, it could have altered the rate of transport along one of the branches of the plant secretory system or the

nature of each of the subcellular compartments, resulting in changes in pH or the amount of degradative enzymes, leading to differences in protein accumulation.

This result does highlight the importance of plant growth conditions on experimental results, particularly for plants that are grown under conditions determined by the environment. Understanding the impact of photoperiod on the accumulation of foreign proteins in the different compartments of a plant cell would provide valuable information for the creation of effective plant expression systems.

6.2.5 Degradation of GFP

When leaf extracts were analyzed by immunoblotting with anti-SBP antibody, two distinct bands could be detected. The first was approximately 75 kDa, the size predicted for the fusion protein. This band could also be detected with anti-GFP antibody, further supporting its identity as the SBP-GFP fusion protein. The second band was approximately 43 kDa, the size of commercial SBP, and could not be detected by anti-GFP. This suggests that the GFP portion is being selectively degraded from the fusion protein. The 43 kDa protein could also represent a portion of the SBP-GFP fusion containing both SBP and GFP amino acid sequence. However, it was not reactive with anti-GFP antibody, which was raised against full-length GFP, suggesting that if GFP sequence remained, it would likely account for only a small proportion of the protein. Peroxidase activity is detected in protein extracts containing primarily the 43 kDa protein, which suggests that the 43 kDa protein does possess peroxidase activity and therefore must contain most of the SBP protein.

Although GFP has been characterized as having high stability, even across a wide range of pH, from 5.5 to 12 (Bokman and Ward 1981), several studies have reported its degradation in different plant subcellular compartments. Batoko et al. (2000) noted that plants expressing

secreted GFP showed much lower fluorescence than those expressing ER-retained GFP, which they suggested may result from the lesser stability of GFP in the apoplast. They also observed two bands in immunoblots with anti-GFP, one of which is approximately 2 kDa smaller than full-size GFP, and is likely to result from partial proteolysis, which is similar to what we observed during immunoblotting of recombinant GFP with anti-GFP antibody (Figure 3.1). As discussed in Section 6.2.2, GFP fluorescence is sometimes much reduced in plant vacuoles and GFP is also occasionally degraded in this compartment (Tamura et al. 2003). This degradation is dependent on the presence of light and an acidic pH and was shown to be caused by vacuolar papain-type cysteine proteases.

In addition to the impact of the cellular environment, it is also possible that the design of our fusion protein affected GFP stability. It has been observed that the length of the linker peptide separating GFP from its fusion partner can influence the stability of GFP fusion proteins, with longer linkers being generally better at achieving proper expression (Prescott et al. 1999). Therefore, it is possible that the linker that we inserted between SBP and GFP was not long enough to allow the proper folding of GFP, leading it to be destabilized and ultimately degraded.

Degradation of GFP limited the information we could obtain through confocal microscopy on the subcellular location of the fusion protein. However, through subcellular fractionation analysis we were able to confirm the subcellular location of the three targeted SBP-GFP fusion proteins. The question does arise, however, as to whether or not the degradation of GFP influences the results obtained in the remainder of the study. As a result of GFP degradation, there are two populations of foreign proteins with potential peroxidase activity – the remaining full-length SBP-GFP fusion protein and the degradation product, which we suggest represents SBP. The SBP in these two populations could conceivably have different activity

levels. In particular, the presence of GFP in the SBP-GFP fusion protein could impact the activity of SBP by destabilizing SBP or by preventing it from folding properly. Since the SBP-GFP targeted to the ER was less degraded than those found in either the apoplast or vacuole, in this case, the activity levels of the ER would be underrepresented. Immunoblotting was able to demonstrate that the peroxidase activity level was roughly correlated with the total amount of both forms of SBP-GFP detected by anti-SBP antibody, which suggests that if differences in activity do exist between the two proteins, they are minor. However, a more accurate quantification of the two populations of fusion protein is necessary to determine if both provide equal levels of peroxidase activity.

Degradation of GFP could also present another challenge in the interpretation of the targeting results since it could also act to remove essential targeting information, resulting in the missorting of the degraded fusion protein. This would not affect both cSBP-GFP and aSBP-GFP, which do not carry C-terminal targeting information. However, eSBP-GFP has a C-terminal ER-retention signal and vSBP-GFP has a C-terminal vacuolar sorting signal. In the case of eSBP-GFP, degradation of this signal should result in its secretion as opposed to retention in the ER. An apoplast fraction was isolated from *Arabidopsis* leaves from plants containing *eSBP-GFP* and no eSBP-GFP could be detected following immunoblotting with anti-SBP (data not shown). This does suggest that the population of eSBP-GFP that is degraded is not being missorted to the apoplast. There is evidence that ER-resident proteins can form complexes that may be inefficiently exported (Hadlington and Denecke 2000) and a similar mechanism could prevent the export of partially degraded eSBP-GFP. Alternatively, assembly defective proteins are retained within the ER and the partially degraded eSBP-GFP could be recognized as such, thereby preventing its secretion (Hadlington and Denecke 2000).

For vSBP-GFP, loss of the C-terminal propeptide will only affect its localization if it occurs prior to transport to the vacuole. Degradation of the eSBP-GFP suggests that some degradation can occur in either the ER or Golgi apparatus. vSBP-GFP degraded in either compartment risks secretion as opposed to vacuolar targeting. Apoplast fractions were prepared from the leaves of plants transformed with *vSBP-GFP* and a small amount of fusion protein was detected by immunoblotting with anti-SBP, although the majority was intracellular (data not shown). These results suggest that a small proportion of vSBP-GFP could be secreted but that the majority is likely correctly sorted to the vacuole.

Degradation of GFP did not appear to be equal in all compartments. Based on the density of the chromogenic precipitate, a larger proportion of the recombinant protein was present as the higher molecular weight form than as the lower molecular weight form in the extracts from the plants containing *eSBP-GFP* compared to any of the other transgenic lines (Figure 3.7). This indicates that of the three compartments, the ER may provide a more stable environment for foreign proteins, particularly those proteins that are naturally less stable. The ER is the site of folding for most proteins and contains many proteins that help achieve this function, including the chaperonins and enzymes such as protein disulfide isomerase (Vitale and Denecke 1999). It is for this reason that the ER likely presents a more stable environment for foreign proteins. In addition, many proteases are secreted or vacuolar, suggesting that proteins targeted to these locations may be more likely to be degraded (Callis 1995; Müntz 2007). Protein degradation represents a significant hurdle to the successful production of foreign proteins in plant systems, and protein targeting is one of the approaches that can be used to minimize its impact (Benchabane et al. 2008). Another approach for limiting protein degradation is tissue-specific expression in tissues that are more stable, such as the seed, which typically has lower protease

activity, suggesting that proteins could be targeted to the vacuole or apoplast in this tissue without being degraded (Benchabane et al. 2008).

6.3 Comparison of transgenic *Ep* expression versus endogenous *Ep* expression

In order to develop a commercially successful plant expression system, it is important to achieve high levels of expression. For the soybean seed coat system, we attempted to adapt a gene expression system naturally functioning in the soybean seed coat for the expression of bioproducts. The *Ep* gene is expressed at high levels in a primarily seed coat specific pattern producing large amounts of the enzyme SBP (Gijzen 1997; Gijzen et al. 1993). Our goal was to replicate the elements responsible for the high levels of expression from the *Ep* locus in our transgenes in order to achieve similarly high levels of expression of the desired bioproducts. This is done in *epep* cultivars that are deficient in SBP in order to minimize toxic effects that result from protein overexpression as well as to simplify the identification of recombinant SBP.

6.3.1 Accumulation of soybean peroxidase during seed coat development

An analysis of the accumulation of SBP during seed coat development of an *EpEp* cultivar was performed using an antibody specific to SBP. This will complement the work that has already been done on the development of the soybean seed coat and the expression of the *Ep* gene during development (Miller et al. 1999; Gijzen et al. 1999) and will provide an important reference for the comparison of transgenic *Ep* plants to *EpEp* cultivars.

Based on immunoblotting and peroxidase assays, SBP in the *EpEp* cv. Harosoy 63 begins to accumulate as early as 9 DPA, although at very low levels (Figure 4.1A and B). These levels remain low until 15 DPA, at which point they increase steadily until approximately 33 DPA, after which they remain relatively steady. This correlates well with the pattern of *Ep* expression

based on *in situ* hybridization, which shows expression of the *Ep* gene beginning at 6 DPA (Gijzen et al. 1999). Both *Ep* transcript and SBP begin to accumulate prior to the development of the hourglass cell layer, which begins to differentiate at 12 DPA in the hilum portion of the seed and is complete throughout the seed by 15 DPA (Miller et al. 1999). At 6 DPA, *Ep* transcript can be found in the thin-walled parenchyma of the outer integument as well as flanking the hilum region. At 9 DPA, transcript can be found between the two parenchyma layers, but by 12 DPA, *Ep* transcript can be found in the newly developing hourglass cells (Gijzen et al. 1999). We would predict that SBP would be found in similar locations prior to 12 DPA. However, the majority of SBP accumulation occurs after 15 DPA, following the differentiation of the hourglass cells, and the exclusive expression of the *Ep* gene in this layer would account for the predominant accumulation of peroxidase activity in this cell layer (Gijzen et al. 1993).

6.3.2 Variability in soybean peroxidase levels between *EpEp* cultivars

The amount of peroxidase activity measured in mature seed coats from five soybean cultivars (Figure 4.1C) varied over 6-fold demonstrating that factors other than the presence of the *Ep* locus can influence the yield of SBP. It also provides a more thorough framework for the comparison of SBP accumulation in transgenic plants and wild-type *EpEp* cultivars. The differences can be attributed to a number of different factors. First, each of the five cultivars has a different genetic background, which could impact *Ep* expression and, correspondingly, SBP accumulation. Since the seeds were not genotyped, it is also possible that some were heterozygous at the *Ep* locus (*Epep*), which could lower the accumulation of SBP through a gene dosage effect. In addition, since we obtained the seeds for this analysis from an external source, the variation in peroxidase activity could also be accounted for by differences in growth conditions. Both of these factors are important as they illustrate the importance of cultivar and

growth conditions on the level of accumulation of a gene product. This also reinforces what was demonstrated by the study in *Arabidopsis*, which similarly demonstrated the influence of growth conditions on protein accumulation.

6.3.3 Expression of *Ep* in transgenic soybean plants

Two constructs were designed to express recombinant *Ep* in an *epep* background. Both contained a gene encoding the *Ep* cDNA with expression driven by the *Ep* 1.5 promoter. One of the two constructs additionally contained the *Ep* 3'UTR.

Transformation of both constructs in soybean plants resulted in production of SBP in the seed coat, as indicated by peroxidase activity assays (Figure 4.4). Although the level of peroxidase activity was higher than is observed in *epep* soybean plants, it was much lower than is observed in any of the *EpEp* cultivars. The plant with the highest expression level had an average seed coat specific activity of 1.41 U mg^{-1} protein with levels in individual seed coats ranging from 0.54 to 3.21 U mg^{-1} protein, while the average specific activity in wild-type *EpEp* cultivars ranged from 20.11 to 135.69 U mg^{-1} protein. Peroxidase activity levels in plants from the same transformation event varied more than did plants between different transformation events, including both those containing the *Ep* gene alone and those additionally containing the *Ep* 3'UTR.

We used RT-PCR to test the hypothesis that SBP activity was related to transcription by comparing the *Ep* transcript abundance in a transgenic plant to the *EpEp* cv. Harosoy 63. We selected the T_1 plant J-EpW-B-4 for this analysis since it is one of the progeny of the T_0 plant J-EpW-B, which had the highest mean peroxidase activity at 1.41 U mg^{-1} protein. At 21 DPA, *Ep* transcript was absent in the seed coats of J-EpW-B-4, in constrast to the Harosoy 63 seed coats which clearly contained *Ep* transcript (Figure 4.6A). Similarly, SBP protein accumulation in the

transgenic plant was undetectable compared to the abundance of SBP protein in Harosoy 63 seed coats (Figure 4.6B and C). These results suggest that there is reduced expression of the transgenic *Ep* gene compared to the endogenous *Ep* gene. However, these results were obtained from a single T_1 plant carrying the transgene and low expression could be the result of position effects or silencing in this plant. More plants will need to be analyzed from this line and other high-expressing lines to demonstrate that the low accumulation of SBP in the transgenic lines is correlated with a decrease in the amount of *Ep* transcript compared to wild-type *EpEp* cultivars.

In order to create a successful system for the high level production of bioproducts, we need both high level and seed coat-specific expression levels. Levels of SBP produced were much lower than expected, so the issue of whether the *Ep* 1.5 promoter maintained seed coat-specific expression was not pursued (see Appendix II). Once high levels of expression are achieved in the seed coat, it will be necessary to screen plants to identify those that also have seed coat-specific expression. Although we were successful in producing SBP in *epep* cultivars through transformation with the *Ep* gene, we were unable to achieve levels of expression similar to endogenous *Ep*. The low levels of expression observed in the *EpC* and *EpW* transgenic plants may result from some of the inherent limitations of transgenic technologies, such as position effects or host defense silencing. Alternatively, it may be that our transgene is missing elements that are necessary for the high levels of expression observable in *EpEp* soybean cultivars.

6.3.4 Influence of the host genome on *Ep* expression in transgenic plants

As discussed in Section 6.2.4, variability in transgene expression has often been attributed to position effects and transgene silencing. Mean peroxidase activity of plants from different transformation events did not vary considerably, which likely indicates that position effects did not contribute to variability in transgene expression, although it is possible that all

four sites of insertion experienced a similar position effect. The mean peroxidase activity of individual plants from the same transformation event varied as much as 22-fold, which could indicate that silencing is occurring to varying degrees in these plants. In particular, some forms of transgene silencing are homology-dependent (Matzke and Matzke 1998) and, even in the case of a single transgene insertion there will still remain sequence identity between the *Ep* transgene and the *ep* locus which are identifcal with the exception of the 87 bp deletion at the *ep* locus that encompasses the initial 78 bp of the open reading frame and 9 bp upstream (Gijzen 1997) as well as the intron sequences that are present at the *ep* locus but not in the *Ep* transgene. This sequence identity could contribute to transgene silencing and might be one reason why lower levels of SBP accumulated in the *EpC* and *EpW* transgenic plants than can typically accumulate in wild-type *EpEp* cultivars.

At least two methods have been proposed to limit the variability in transgene expression that can result from position effects and host genome silencing and these could be applied to improve the soybean seed coat system. The inclusion of scaffold/matrix attachment regions in transgene design can reduce transgene expression variability (Grosveld et al. 1987; Stief et al. 1989; Bonifer et al. 1990; Breyne et al. 1992b; Petersen et al. 2002) and reduce transgene silencing (Brouwer et al. 2002; Ulker et al. 1999; Abranches et al. 2005). In addition, these scaffold/matrix attachment regions have been shown to increase transgene expression (Allen et al. 1993; Vain et al. 1999; Liu and Tabe 1998; Petersen et al. 2002; Ulker et al. 1999; Cheng et al. 2001). Matrix attachment regions have been shown to attach to the nuclear matrix *in vitro* and are suspected to do the same *in vivo*, creating an independently regulated transcriptional domain that is insulated from surrounding chromatin (Allen et al. 2000; Spiker and Thompson 1996).

The second possibility is the targeted insertion of the transgene into well-characterized genome locations that are known to provide stable transgene expression. This approach can be achieved through site-specific recombination, for example by using the Cre/*lox* recombination system (Gilbertson 2003; Srivastava and Ow 2004). Additionally, this approach can be used to reliably generate single copy insertions. While this approach can help to minimize position effects, transgenes may still be subject to host silencing (Day et al. 2000).

6.3.5 Limitations of the *Ep* 1.5 promoter

Given the low levels of SBP accumulation in the *EpC* and *EpW* transgenic plants, and the fact that similarly low levels of expression have been observed in transgenic plants transformed with other constructs (S. Han, unpublished results), it is likely that in addition to expression variability resulting from host genomic influences, elements that contribute to the high expression levels of the endogenous *Ep* gene were not included in our transgene design. We selected a region 1.5 kb upstream from the *Ep* gene as our promoter. In this region, immediately upstream of the *Ep* gene is the 5'UTR which is relatively small, consisting of either 13 nucleotides, based on 5'RACE analysis, or 16 nucleotides, based on *in silico* predictions (M. Gijzen, unpublished data). A TATA box was also identified from position -37 to -47 (M. Gijzen, unpublished data). Additionally, several regulatory sequences were identified in this segment of DNA using the Motif Finder and Pattern Search programs (M. Gijzen, unpublished data). This suggests that this region encompasses the core promoter as well as containing additional regulatory regions, but does not guarantee that all of the regulatory elements necessary for expression of the *Ep* gene are present.

It is possible that the *Ep* 1.5 promoter omitted other important regulatory elements perhaps located further upstream. Enhancer elements have been found thousands of base pairs

from their site of action (Potenza et al. 2004; Clark et al. 2006; Jack et al. 1991; Krumm et al. 1998; Lettice et al. 2003). Furthermore, enhancer elements can also be found downstream of the coding region as well as within the coding region (Potenza et al. 2004).

Regulatory elements have also been found within introns (Sternberg et al. 1988; Bruhat et al. 1990; Deyholos and Sieburth 2000; Rippe et al. 1989) and the inclusion of introns in transgenes has been shown to boost expression levels (Gallie and Young 1994; Buchman and Berg 1988; Huang and Gorman 1990; Brinster et al. 1988; Callis et al. 1987; Vasil et al. 1989; Tanaka et al. 1990; Rose and Last 1997). Introns can function to increase gene expression in many ways, including promotion of transcription initiation, pre-mRNA editing and polyadenylation, mRNA nuclear export, translation, and decay (Le Hir et al. 2003).

The *Ep* gene contains three introns (Gijzen 1997), which could contribute to the high expression of the *Ep* gene. These introns could potentially contain positive regulatory elements or they could enhance accumulation of *Ep* transcript post-transcriptionally. A construct was designed by Dr. L. Gudynaitę-Savitch containing the entire *Ep* genomic DNA sequence, including all three introns. However, no plants survived to maturity to permit analysis. A comparison between transgenic plants containing the *Ep* gene with and without different combinations of the three introns is necessary to determine the influence of the introns on *Ep* expression.

In addition to introns, studies have demonstrated that 5' and 3' UTRs can influence gene expression (Bailey-Serres and Dawe 1996; Bate et al. 1996; Gallie et al. 1987; De Loose et al. 1995; Boado and Pardridge 1998; Ingelbrecht et al. 1989). Both 5' and 3'UTRs can influence the rate of translation and sites have been identified within these regions that bind regulatory proteins (Bate et al. 1996; McCarthy and Kollmus 1995; Kaufman 1994; Klausner et al. 1993;

Curtis et al. 1995; Mignone et al. 2002; Chabouté et al. 2002). The 3'UTR also plays a role in regulating mRNA stability (Decker and Parker 1994; Mignone et al. 2002; Green 1993; Mazumder et al. 2003; Powell et al. 2000; Boado and Pardridge 1998).

The *Ep* 5'UTR was included in both the *EpC* and *EpW* constructs, so its impact on gene expression could not be analyzed. However, the 5'UTR is quite small, either 13 or 16 nucleotides, as determined by 5'RACE and *in silico* analysis, respectively, and is therefore unlikely to contain many regulatory sequences.

The *Ep* 3'UTR was only present in the *EpW* construct. There was no major difference in mean specific activity of seed coats from plants containing the *EpC* gene versus those containing the *EpW* gene. This would suggest that the *Ep* 3'UTR does not contribute substantially to *Ep* expression levels. However, only a small number of independent insertions were analyzed, and more should be studied to confirm this conclusion. The 3'UTR region of the *Agrobacterium tumefaciens nos* gene, also known as a transcriptional terminator or poly (A) signal, was included in both the *EpC* and *EpW* constructs. This sequence, and other related sequences such as the CaMV 35S poly (A) signal, are typically included in plant expression vectors, and are believed to promote stable transgene expression. It is possible that the contribution of the *Ep* 3'UTR was masked by the presence of the *nos* transcriptional terminator. To accurately determine the influence of the *Ep* 3'UTR on transgene expression levels, soybean plants transformed with a construct that contains only the *Ep* 3'UTR and poly (A) signal following the *Ep* gene should be compared to plants transformed with the *EpC* transgene.

6.3.6 Enhancing expression of transgenes in the soybean seed coat system

In order to create a plant expression system that can be successful commercially, it is critical that high, stable expression levels can be reliably generated. There are two avenues that

must be explored to achieve this aim. The first is to complete the characterization of the *Ep* locus in order to identify the elements missing in our constructs that are responsible for the high levels of endogenous *Ep* expression. To do this, regions further upstream of the *Ep* 1.5 promoter, as well as the *Ep* introns must be analyzed for the presence of positive regulatory elements. A complete characterization of the *Ep* promoter is also necessary to achieve seed coat-specific expression. It is important to identify any elements responsible for the seed coat-specific expression so that they can be retained in future promoter designs while elements responsible for expression in other tissues, such as the low level expression detectable in root tissues (Gijzen 1997), are removed.

The second avenue that can be explored to improve transgene expression in the seed coat system is the introduction of novel elements. Promoter elements from different genes can be combined to create strong synthetic promoters (Lee et al. 2007; Ni et al. 1995; Comai et al. 1990; Rancé et al. 2002). Strong positive regulatory elements could be combined with the seed coat-specific elements of the *Ep* promoter to enhance gene expression. The *Ep* 5'UTR could be replaced with a 5'UTR that is known to promote high levels of gene expression. Commonly used 5'UTRs for this purpose include those from tobacco mosaic virus (De Jaeger et al. 2002; Gallie and Walbot 1992; Gallie et al. 1987), the alfalfa mosaic virus (Charest et al. 1993; DeMattos et al. 1998; Sutton et al. 1992), and the potato virus X (Pooggin and Skryabin 1992; Hefferon et al. 1997). As discussed in Section 6.3.4, global regulatory sequences such as scaffold/matrix attachment regions can be used to increase transgene expression. If the *Ep* introns do not contribute to gene expression, then a synthetic intron could be included in the transgene that is known to have such an effect (Fiume et al. 2004; Tanaka et al. 1990; Norris et al. 1993). By using a combinatorial approach to enhance the expression level of the *Ep* promoter, it will

become possible to equal or exceed expression levels that are naturally produced by the wild-type *Ep* promoter, which will be necessary to create a successful plant expression system.

Promoter optimization can be a lengthy process, which is made more difficult by lengthy soybean transformation technologies, although as soybean transformation technology improves this process will become easier. Still it may be necessary to complete some of the optimization process in an alternative model plant, such as *Lotus japonicus* or *Medicago sativa*, for which there are more efficient transformation methodologies (Somers et al. 2003). An alternative approach may be to select a different promoter that can drive high levels of seed coat-specific expression. A proteomics approach has been used by M.C. Romero to identify proteins that are produced at high levels specifically in soybean seed coats, which could yield a number of candidate genes containing promoters with the desired properties (unpublished results). It may be simpler to identify a promoter with all of the necessary regulatory elements clustered in a more localized region than to try to locate the missing regulatory elements of the *Ep* promoter.

6.4 Metabolic engineering of soybean seed coats

In addition to the production of proteins, the soybean seed coat system could also be used to produce biochemicals, through the introduction or modification of metabolic pathways. This approach could be used to alter the nutritional content of soybean seed coats, providing a more desirable feed component. It could also be used to produce a whole range of products, including fatty acids, secondary metabolites such as terpenoids, flavonoids, carotenoids, and complex polymers such as lignin (Capell and Christou 2004). This approach can also be used to introduce novel metabolites, of which the classic example is PHB. To investigate metabolic engineering of soybean seed coats, we introduced the PHB biosynthetic pathway. This pathway was selected

because it is relatively simple, consisting of only three novel enzymes, and it has been successfully introduced into a number of plant species (Table 1.1 and references therein).

6.4.1 Synthesis of polyhydroxybutyric acid in soybean seed coats

A single construct containing gene cassettes encoding each of the three PHB biosynthetic enzymes driven by the *Ep* 1.5 promoter as well as the selectable marker was introduced into soybean plants. Soybean seed coat extracts were analyzed by gas chromatography. PHB was detected up to a maximum of 0.36% of seed coat DW in a single seed coat, although most seed coats had little or no detectable PHB. This result demonstrates that it is possible to metabolically engineer soybean seed coats, although the level of PHB produced was relatively low. As with the *Ep* gene, the variability in PHB production was greater between plants from a single transformation event than between the different transformation events.

Initial work in *Arabidopsis* yielded low levels of PHB, in the range of 0.002 to 0.01% of leaf FW, following crossing of separate transgenic lines containing the *phbB* and *phbC* genes without targeting signals (Poirier et al. 1992a; 1992b). Similarly low levels were achieved in a number of other plant species, including Black Mexican Sweet maize suspension cultures (Hahn et al. 1999), flax (Wróbel et al. 2004), cotton (John and Keller 1996), alfalfa (Saruul et al. 2002), tobacco (Bohmert et al. 2002), and potato (Bohmert et al. 2002), using a variety of gene combinations, targeting signals, and promoters, as outlined in Table 1.1. Further work in *Arabidopsis* has increased the yield of PHB to 1% of FW (14% DW) by crossing of separate transgenic lines containing the three *phb* genes altered to include chloroplast targeting signals (Nawrath et al. 1994). These levels were then further improved to 4% FW (40% DW) by introducing all three genes with chloroplast targeting signals simultaneously in a single construct (Bohmert et al. 2000). This demonstrates that it is possible to enhance the production of PHB

through a variety of techniques, including transformation technologies and targeting. Similar strategies could be used to further enhance the accumulation of PHB in soybean seed coats.

6.4.2 Relationship between PHB production and expression of *phb* genes

It is possible that the low level of PHB accumulation in soybean seed coats was the result of the low expression of the genes encoding the three PHB biosynthetic enzymes. As discussed in Section 6.3.3, expression from the *Ep* 1.5 promoter appears to be much lower than the level of expression normally produced by the endogenous *Ep* promoter. This suggests that we omitted regulatory elements that contribute to the high levels of expression of the *Ep* gene, perhaps located further upstream from the *Ep* gene or in one of the three introns found within the *Ep* coding region. Based on northern blot analysis of transgenic *Arabidopsis* plants expressing the three *phb* genes, there is a positive correlation between the expression of the *phb* genes and the level of PHB accumulation (Bohmert et al. 2000). At the protein level, a positive correlation has also been observed between the level of 3-ketothiolase and acetoacetyl-CoA reductase activity and the level of accumulation of PHB (Nawrath et al. 1994). This suggests that the low level of PHB produced in soybean seed coats could be due to low levels of expression of the *phb* genes.

In addition to limitations of the *Ep* 1.5 promoter, low levels of expression of the *phb* genes could also result from the influence of the host genome through silencing and/or position effects, as discussed in Sections 6.2.4 and 6.3.4. With the J13 construct, there is an increased risk that silencing could impact gene expression, since there are three repeats of the *Ep* 1.5 promoter, the *nos* transcriptional terminator, as well as the sequences encoding the SBP N- and C-terminal propeptides. Multiple copies of similar sequences in transgenes are more likely to result in silencing than are unique sequences (Stam et al. 1997). This is further compounded by the fact that a large portion of the *Ep* 1.5 promoter is also present in the host genome as is the sequence

encoding the SBP C-terminal propeptide. The N-terminal propeptide and a portion of the *Ep* 1.5 promoter are not found in the host genome since these are deleted in *epep* cultivars. We attempted to minimize silencing by introducing these sequences in a head-to-tail orientation, which would prevent the formation of inverted repeats that are known to be more likely to cause silencing (Stam et al. 1997). However, the potential for silencing still exists with this construct and may be a limitation that needs to be overcome. A fusion of the 3-ketothiolase and acetoacetyl-CoA reductase enzymes has been generated and expressed in *Arabidopsis* (Kourtz et al. 2005). This is one approach that could reduce the number of independent transgenes that must be introduced, and thereby reduce the number of copies of the promoter and regulatory regions, although two repeats of each would still be required. Alternatively, unique promoters and regulatory regions could be used for each of the genes, although this will require the identification of new high level, seed coat-specific promoters.

6.4.3 Relationship between availability of acetyl-CoA and level of PHB production

It is also possible that production of PHB was limited by the availability of acetyl-CoA. The maximum amount of PHB produced in *Arabidopsis* increased approximately 100-fold when the chloroplast-targeting signals were added to the PHB biosynthetic enzymes (Nawrath et al. 1994). This was attributed to the larger pool of acetyl-CoA located in the chloroplast for fatty acid biosynthesis than can be found in the cytoplasm, where acetyl-CoA is primarily used as a precursor for the mevalonate pathway (Nawrath et al. 1994).

We chose to target the PHB biosynthetic pathway to the vacuole in soybean seed coats, using the SBP N- and C-terminal propeptides. It is predicted that SBP is located within the vacuole in the seed coat, and this is supported by the demonstration that the SBP C-terminal propeptide can function as a vacuolar sorting signal in *Arabidopsis*. The seed coat vacuoles can

therefore host a large amount of protein, suggesting that they will have the capacity to host the PHB biosynthetic enzymes and the resulting PHB, particularly since we are using a cultivar in which SBP is not produced. In addition, vacuoles serve as storage compartments and often contain high concentrations of a variety of metabolites (Marty 1999). As an example, in spinach leaves vacuolar malate concentrations are 5-fold and 8-fold greater than in the chloroplast stroma and cytosol, respectively (Winter et al. 1994). Therefore, we hypothesized that the vacuoles might contain a pool of stored acetyl-CoA that could be used for the production of PHB, although there is no data indicating the potential levels in this organelle in hourglass cells. However, if this hypothesis was incorrect and the pool of acetyl-CoA in the vacuole of seed coats was not large, this could also have contributed to the low levels of PHB that were observed in soybean seed coats.

Targeting the PHB biosynthetic enzymes to a different subcellular compartment in soybean seed coats with a larger pool of acetyl-CoA could help to improve the production of PHB. In *Arabidopsis*, PHB production was significantly improved by targeting the PHB biosynthetic enzymes to the chloroplast because of the large pool of acetyl-CoA in this organelle (Nawrath et al. 1994; Bohmert et al. 2000). High levels of PHB have also been obtained in *Brassica napus* seeds by targeting the PHB biosynthetic enzymes to leucoplasts (Valentin et al. 1999; Houmiel et al. 1999). There are numerous starch granules observable throughout development of the hourglass cell layer (Miller et al. 1999). This suggests that there is likely to be high metabolic flux through amyloplasts, some of which could be redirected towards PHB biosynthesis. The biosynthetic enzymes could, therefore, be targeted to plastids in the soybean seed coat in order to enhance the production of PHB. Alternatively, PHB has also been produced up to 0.2% FW by targeting the biosynthetic enzymes to the peroxisome in Black Mexican Sweet

maize suspension cultures (Hahn et al. 1999). Targeting the PHB biosynthetic pathway to a different subcellular compartment may be another approach that could be taken to increase the yield of PHB.

A powerful approach for identifying changes in metabolic flux following the introduction of novel metabolic pathways is metabolic profiling using a combined gas chromatography-mass spectrometry approach. This approach was used to analyze *Arabidopsis* plants producing PHB and revealed that while levels of fatty acids remained stable, other important metabolites had altered concentrations, including fumarate and isocitrate (Bohmert et al. 2000). This approach could be employed to analyze changes in metabolic flux in soybean seed coats and could help to identify if there are limitations at the level of metabolism in the production of PHB.

6.4.4 Influence of glycosylation on the PHB biosynthetic enzymes

Another factor that could have contributed to the low production of PHB in soybean seed coats is glycosylation of the PHB biosynthetic enzymes. Since the enzymes were targeted to seed coat vacuoles with the SBP N- and C-terminal propeptides, they are expected to pass through the secretory system where they could potentially be glycosylated. The phbC enzyme does not have any potential N-glycosylation sites but the phbA enzyme has two and the phbB enzyme has four. These enzymes are bacterial in origin and most bacterial proteins are not glycosylated (Schäffer et al. 2001). Therefore, glycosylation could potentially have a negative impact on the function of the enzymes. This is one situation where targeting proteins to the plant secretory system could be disadvantageous. Although these enzymes have been expressed in plant systems, this is the first time that they have been targeted to the secretory system where they could potentially be glycosylated and therefore the impact of glycosylation on their function is unknown.

Antibodies against each of the PHB biosynthetic enzymes have been made (Nawrath et al. 1994) and could be used to determine if the enzymes have been glycosylated by comparing their speed of migration before and after treatment with endoglycosidase H to remove glycosyl groups. If the enzymes are glycosylated, the impact of glycosylation on enzyme activity can be assessed by enzyme assay (Nawrath et al. 1994; Poirier et al. 1992b) of glycosylated and unglycosylated forms. Alternatively, to avoid this problem, the glycosylation consensus sites can be mutated to prevent their glycosylation or the biosynthetic enzymes can be targeted to an alternative subcellular compartment where they will not be glycosylated.

6.4.5 Impact of PHB production on plant health

Another possible explanation for the low levels of PHB observed in soybean seed coats is related to the potential impact of PHB on plant health. There was some indication that production of PHB had an impact on fertility at the level of seed abortion and germination. One plant showed a germination rate of only 7.5%. Another plant had at least one aborted seed in almost half of all pods produced. These trends will need to be carefully documented in the high-expressing soybean lines to characterize the impact of PHB production on seed viability.

Seed coats were primarily obtained from seeds that were able to germinate, unless the line only produced a few seeds and most failed to germinate, such as was seen for the plants from event 1 (Figure 5.4). If high levels of PHB prevented seed germination, this could skew the results since only seed coats with lower levels of PHB would have been analyzed. Two of the seed coats with the highest level of PHB came from seeds that failed to germinate; however, another seed coat with a similar level of PHB did come from a seed that could germinate. Therefore the impact of PHB production on seed health remains unclear and will need to be

further examined to determine if soybean seed coats can undergo metabolic engineering without compromising seed health.

Production of PHB in plants has often been observed to have a negative impact on plant health, even when produced at levels of less than 1% (Suriyamongkol et al. 2007 and references therein). Negative effects have included reduced growth and fertility, reduced seed germination, leaf chlorosis, and male sterility. Nawrath et al. (1994) were able to eliminate the impact of PHB on plant health by transferring the PHB biosynthetic pathway from the cytoplasm to the chloroplast, but when PHB levels were further increased, the negative effects on plant health returned (Bohmert et al. 2000). Sequestering PHB biosynthesis in the vacuoles of the hourglass cells could similarly help to minimize the impact of PHB on seed health. However, there is some evidence that PHB can cross the vacuolar membrane (Poirier et al. 1992a; 1992b; Mittendorf et al. 1998), which could limit the ability of the vacuole to sequester PHB and minimize its impact on plant health.

PHB could directly affect plant health or the negative effects could be related to depletion of acetyl-CoA pools. Reductions in the levels of fumarate and isocitrate in *Arabidopsis* plants producing PHB indicate that depletion of acetyl-CoA in the chloroplast can impact the tricarboxylic acid cycle (Bohmert et al. 2000). Bohmert et al. (2002) demonstrated that the *phbA* gene, encoding 3-ketothiolase, had a negative effect on plant transformation efficiency in *Arabidopsis*, tobacco, and potato. This could again be accounted for simply by a depletion of the acetyl-CoA pool. Alternatively, it could result from toxic effects of metabolic intermediates, produced by 3-ketothiolase, such as acetoacetyl-CoA, or the 3-ketothiolase enzyme could itself be toxic to plants (Bohmert et al. 2002).

It is important that metabolic engineering of soybean seed coats not impact seed health. Careful analysis will need to be done to observe changes in metabolic flux and relate these changes to seed health. This will most effectively be accomplished through metabolic profiling techniques. The choice of subcellular compartment could also play an important role in limiting the impact of metabolic engineering on the seed.

6.4.6 Complexity of metabolic engineering

The production of PHB in soybean seed coats demonstrates that it is feasible to produce a novel biochemical in this tissue. However, it also demonstrates the difficulties typically associated with metabolic engineering. It is increasingly apparent that effective metabolic engineering will likely require the manipulation of several metabolic pathways in order to direct metabolic flux towards the desired biochemical (Capell and Christou 2004). In the case of PHB, it may be necessary to alter other metabolic pathways in order to create a large acetyl-CoA pool. This could help to minimize the impact of PHB production on plant health as well as increasing the total yield of PHB.

6.5 Conclusions

It was our goal to develop a transgenic plant system for the heterologous expression of bioproducts in soybean seed coats. We successfully demonstrated the production of the industrial enzyme SBP and the biochemical PHB in this tissue, although the overall yield of these products was low. In order to create a system that can be commercially successful, high yields of the desired product are fundamental. By expressing SBP, we were also able to directly compare the levels of accumulation of SBP in our transgenic system to what is naturally produced in *Ep*

cultivars. This analysis suggested that the *Ep* 1.5 promoter likely did not encompass all of the necessary regulatory regions for the high level of expression of the *Ep* gene.

Introduction of the PHB biosynthetic pathway additionally demonstrated the complexity of metabolic engineering of plant systems. While novel metabolites can be produced in plant systems, maximizing expression while minimizing impact on plant health is a difficult balance to achieve that will likely require multipoint engineering.

We added the SBP N- and C-terminal propeptides to all foreign proteins expressed in the soybean seed coat system. These signals are putative targeting signals and we hypothesized that they may be necessary to achieve the high levels of accumulation observed for SBP in wild-type *EpEp* cultivars. To further investigate the influence of targeting on the yield of foreign proteins, a model system was created in *A. thaliana*. This system involved a fusion between SBP as a marker enzyme and GFP as a marker of subcellular location, targeted to different compartments of the plant secretory system with combinations of the SBP N- and C-terminal propeptides as well as an ER retention signal. In this system, the yield of the fusion protein was greater when targeting signals were present, although the choice of final location did not have an influence on yield. It did have an influence on protein stability, as protein degradation was greater in the apoplast and vacuole than in the ER. This system additionally allowed us to demonstrate that the SBP N-terminal propeptide can function as an N-terminal signal peptides and the C-terminal propeptide can function as a vacuolar sorting signal in a heterologous plant system.

These results are an important first step in the development of the soybean seed coat system, although more work is necessary to enhance the expression of the bioproducts in order to create a viable system.

6.6 Future directions

The research described in this thesis outlines a first attempt to engineer soybean plants to produce important bioproducts in the seed coat. Based on the results, there are a number of future directions that should be pursued to further improve the system. The most promising avenues are outlined here.

1. Characterization of the *Ep* promoter. The results presented here suggest that the *Ep* 1.5 promoter is missing elements that are necessary for the high expression levels observed for the *Ep* gene in wild-type *EpEp* cultivars. A complete characterization of the *Ep* 1.5 promoter as well as other regions of the *Ep* locus is necessary to identify the regulatory elements that promote both high levels and seed coat-specific expression. This can be done both through sequence analysis and through reporter assay, using promoter deletions and mutations to identify the important regulatory regions. Due to limitations in soybean transformation, it may be necessary to conduct this study in an alternative model system, such as *Lotus japonicus* or *Medicago sativa*, although advances in soybean transformation are improving the efficiency of this method (Somers et al. 2003). A complete characterization of the *Ep* promoter is necessary in order to design a promoter that both promotes high level expression and maintains a seed coat-specific pattern of expression as this is an essential feature of the soybean seed coat system.

2. Enhancement of the *Ep* promoter. Once the *Ep* promoter elements have been identified, they can be combined with other regulatory elements to further enhance the level of expression that can be generated by the *Ep* promoter. As discussed in Section 6.3.6, this can include regulatory elements from other promoters, stronger 5' and 3'UTRs, and the addition of an intron, all of which can enhance transgene expression. More promoter analysis work will need to be done using a reporter assay to identify the combination of elements that produces the

highest level of transgene expression. Again, this work may be limited by the available transformation technology and may require work done in an alternative model system. The largest barrier in the development of plant expression systems is achieving a high level of expression and this work is therefore essential in order to maximize the potential of the soybean seed coat system.

3. Metabolic profiling of soybean seed coats. Expression of the PHB biosynthetic pathway in soybean seed coats demonstrated that this tissue can successfully divert metabolism towards the production of a novel metabolite. Comparing the metabolic profile of wild-type and transgenic soybean seed coats will indicate which pathways were altered in order to produce PHB. Metabolic profiling by gas chromatography-mass spectrometry is an effective means of analyzing how the introduction of the PHB biosynthetic pathway can alter metabolism (Bohmert et al. 2000). This information can be used to determine the types of biochemicals that can be successfully produced in soybean seed coats. It could also provide information that can allow us to understand how alterations in metabolism can affect host health, allowing us to find a balance between metabolic engineering and seed health.

4. In addition, it may be possible to boost expression in seed coats by introducing transgenes with elements specific for the epidermal layer in order to maximize expression of bioproducts throughout the seed coat. The epidermal layer is involved in the synthesis of pigments as well as the waxy cuticle of the seed (Moïse et al. 2005), and may therefore have a high metabolic flux that can be diverted towards the production of novel bioproducts.

References

Abranches, R, RW Shultz, WF Thompson, and GC Allen. 2005. Matrix attachment regions and regulated transcription increase and stabilize transgene expression. Plant Biotechnol J 3: 535-543.

Ahmed, SU, M Bar-Peled, and NV Raikhel. 1997. Cloning and subcellular location of an Arabidopsis receptor-like protein that shares common features with protein-sorting receptors of eukaryotic cells. Plant Physiol 114: 325-336.

Ahmed, SU, E Rojo, V Kovaleva, S Venkataraman, JE Dombrowski, K Matsuoka, and NV Raikhel. 2000. The plant vacuolar sorting receptor AtELP is involved in transport of NH2-terminal propeptide-containing vacuolar proteins in Arabidopsis thaliana. J Cell Biol 149: 1335-1344.

Allefs, SJHM, ER De Jong, DEA Florack, C Hoogendoorn, and WJ Stiekema. 1993. *Erwinia* soft rot resistance of potato cultivars expressing antimicrobial tachyplesin I. Mol Breed 2: 97-105.

Allen, GC, GE Hall, Jr., LC Childs, AK Weissinger, S Spiker, and WF Thompson. 1993. Scaffold attachment regions increase reporter gene expression in stably transformed plant cells. Plant Cell 5: 603-613.

Allen, GC, S Spiker, and WF Thompson. 2000. Use of matrix attachment regions (MARs) to minimize transgene silencing. Plant Mol Biol 43: 361-376.

Amiche, M, F Ducancel, A Mor, JC Boulain, A Menez, and P Nicolas. 1994. Precursors of vertebrate peptide antibiotics dermaseptin b and adenoregulin have extensive sequence identities with precursors of opiod peptides dermorphin, dermenkephalin, and deltorphins. J Biol Chem 269: 17847-17852.

Anderson, AJ and EA Dawes. 1990. Occurrence, metabolism, metabolic role, and industrial uses of bacterial polyhydroxyalkanoates. Microbiol Rev 54: 450-472.

Andrès, E and JL Dimarcq. 2005. Clinical development of antimicrobial peptides. Int J Antimicrob Agents 25: 448-452.

Arai, Y, T Shikanai, Y Doi, S Yoshida, I Yamaguchi, and H Nakashita. 2004. Production of polyhydroxybutyrate by polycistronic expression of bacterial genes in tobacco plastid. Plant Cell Physiol 45: 1176-1184.

Argos, P. 1990. An investigation of oligopeptides linking domains in protein tertiary structures and possible candidates for general gene fusion. J Mol Biol 211: 943-958.

Bailey-Serres, J and RK Dawe. 1996. Both 5' and 3' sequence of maize *adh1* mRNA are required for enhanced translation under low-oxygen conditions. Plant Physiol 112: 685-695.

Barton, MD. 2000. Antibiotic use in animal feed and its impact on human health. Nutr Res Rev 13: 279-299.

Batard, Y, A Hehn, S Nedelkina, M Schalk, K Pallett, H Schaller, and D Werck-Reichhart. 2000. Increasing expression of P450 and P450-reductase proteins from monocots in heterologous systems. Arch Biochm Biophys 379: 161-169.

Bate, N, C Spurr, GD Foster, and D Twell. 1996. Maturation-specific translational enhancement mediated by the 5'-UTR of a late pollen transcript. Plant J 10: 613-623.

Batoko, H, H-Q Zheng, C Hawes, and I Moore. 2000. A Rab1 GTPase is required for transport between the endoplasmic reticulum and Golgi apparatus and for normal Golgi movement in plants. Plant Cell 12: 2201-2217.

Bäurle, I and C Dean. 2006. The timing of developmental transitions in plants. Cell 125: 655-664.

Bednarek, SY and NV Raikhel. 1991. The barley lectin carboxyl-terminal propeptide is a vacuolar protein sorting determinant in plants. Plant Cell 3: 1195-1206.

Bednarek, SY, TA Wilkins, JE Dombrowski, and NV Raikhel. 1990. A carboyl-terminal propeptide is necessary for proper sorting of barley lectin to vacuoles of tobacco. Plant Cell 2: 1145-1155.

Belaid, A, M Aouni, R Khelifa, A Trabelsi, M Jemmali, and K Hani. 2002. *In vitro* antiviral activity of dermaseptins against Herpes Simplex Virus Type 1. J Medical Virology 66: 229-234.

Benchabane, M, C Goulet, D Rivard, L Faye, V Gomord, and D Michaud. 2008. Preventing unintended proteolysis in plant protein biofactories. Plant Biotechnol J 6: 633-648.

Bethke, PC, SJ Swanson, S Hillmer, and RL Jones. 1998. From storage compartment to lytic organelle: The metamorphosis of the aleurone protein storage vacuole. Ann Bot 82: 399-412.

Bland, JM and DG Altman. 1995. Multiple significance tests: the Bonferroni method. BMJ 310: 169-170.

Boado, RJ and WM Pardridge. 1998. Ten nucleotide *cis* element in the 3'-untranslated region of the GLUT1 glucose transporter mRNA increases gene expression via mRNA stabilization. Mol Brain Res 59: 109-113.

Bohmert, K, I Balbo, J Kopka, V Mittendorf, C Nawrath, Y Poirier, G Tischendorf, RN Tretheway, and L Willmitzer. 2000. Transgenic *Arabidopsis* plants can accumulate polyhydroxybutyrate to up to 4% of their fresh weight. Planta 211: 841-845.

Bohmert, K, I Balbo, A Steinbüchel, G Tischendorf, and L Willmitzer. 2002. Constitutive expression of the β-ketothiolase gene in transgenic plants. A major obstacle for obtaining polyhydroxybutyrate-producing plants. Plant Physiol 128: 1281-1290.

Bokman, SH and WW Ward. 1981. Renaturation of *Aequorea* green-fluorescent protein. Biochem Biophys Res Commun 101: 1372-1380.

Boller, T and H Kende. 1979. Hydrolytic enzymes in the central vacuole of plant cells. Plant Physiol 63: 1123-1132.

Bonifer, C, M Vidal, F Grosveld, and AE Sippel. 1990. Tissue specific and position independent expression of the complete gene domain for chicken lysozyme in transgenic mice. EMBO J 9: 2843-2848.

Boudart, G, E Jamet, M Rossignol, C Lafitte, G Borderies, A Jauneau, M-T Esquerré-Tugayé, and R Pont-Lezica. 2005. Cell wall proteins in apoplastic fluids of *Arabidopsis thaliana* rosettes: Identification by mass spectrometry and bioinformatics. Proteomics 5: 212-221.

Bradford, MM. 1976. A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72: 248-254.

Bregitzer, P and D Tonks. 2003. Inheritance and expression of transgenes in barley. Crop Sci 43: 4-12.

Breyne, P, G Gheysen, A Jacobs, M Van Montagu, and A Depicker. 1992a. Effect of T-DNA configuration on transgene expression. Mol Gen Genet 235: 389-396.

Breyne, P, MV Montagu, A Depicker, and G Gheysen. 1992b. Characterization of a plant scaffold attachment region in a DNA fragment that normalizes transgene expression in tobacco. Plant Cell 4: 463-471.

Brinster, RL, JM Allen, RR Behringer, RE Gelinas, and RD Palmiter. 1988. Introns increase transcriptional efficiency in transgenic mice. Proc Natl Acad Sci USA 85: 836-840.

Brogden, KA. 2005. Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? Nat Rev Microbiol 3: 238-250.

Brouwer, C, W Bruce, S Maddock, Z Avramova, and B Bowen. 2002. Suppression of transgene silencing by matrix attachment regions in maize: A dual role for the maize 5' *ADH1* matrix attachment region. Plant Cell 14: 2251-2264.

Brown, KL and REW Hancock. 2006. Cationic host defense (antimicrobial) peptides. Curr Opin Immunology 18: 24-30.

Brown, JC, NA Joliffe, L Frigerio, and LM Roberts. 2003. Sequence-specific, Golgi-dependent vacuolar targeting of castor bean 2S albumin. Plant J 36: 711-719.

Bruhat, A, S Tourmente, S Chapel, ML Sobrier, JL Couderc, and B Dastugue. 1990. Regulatory elements in the first intron contribute to transcriptional regulation of the β_3 tubulin gene by 20-hydroxyecdysone in *Drosophila* Kc cells. Nucleic Acid Res 18: 2861-2867.

Buchman, AR and P Berg. 1988. Comparison of intron-dependent and intron-independent gene expression. Mol Cell Biol 8: 4395-4405.

Cakmak, I and WJ Horst. 1991. Effect of aluminum on lipid peroxidation, superoxide dismutase, catalase, and peroxidase activities in root tips of soybean (*Glycine max*). Physiol Plant 83: 463-468.

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

Callis, J. 1995. Regulation of protein degradation. Plant Cell 7: 845-857.

Callis, J, M Fromm, and V Walbot. 1987. Introns increase gene expression in cultured maize cells. Genes Devel 1: 1183-1200.

Campisi, L, Y Yang, Y Yi, E Heilig, B Herman, AJ Cassista, DW Allen, H Xiang, and T Jack. 1999. Generation of enhancer trap lines in *Arabidopsis* and characterization of expression patterns in the inflorescence. Plant J 17: 699-707.

Cao, X, SW Rogers, J Butler, L Beevers, and JC Rogers. 2000. Structural requirements for ligand binding by a probable plant vacuolar sorting receptor. Plant Cell 12: 493-506.

Capell, T and P Christou. 2004. Progress in plant metabolic engineering. Curr Opin Biotechnol 15: 148-154.

Carroll, SL, ML Miller, PW Frohnert, SS Kim, and JA Corbett. 1997. Expression of neuroregulins and their putative receptors, ErbB2 and ErbB3, is induced during wallerian degeneration. J Neurosci 17: 1642-1659.

Carter, C, S Pan, J Zouhar, EL Avila, T Girke, and NV Raikhel. 2004. The vegetative vacuole proteome of *Arabidopsis thaliana* reveals predicted and unexpected proteins. Plant Cell 16: 3285–3303.

Castelli, S and A Vitale. 2005. The phaseolin vacuolar sorting signal promotes transient, strong membrane association and aggregation of the bean storage protein in transgenic tobacco. J Exp Bot 56: 1379-1387.

Caza, N, JK Bewtra, N Biswas, and KE Taylor. 1999. Removal of phenolic compounds from synthetic wastewater using soybean peroxidase. Wat Res 33: 3012-3018.

Chabouté, M-E, B Clément, and G Philipps. 2002. S phase and meristem-specific expression of the tobacco *RNR1b* gene is mediated by an E2F element located in the 5' leader sequence. J Biol Chem 277: 17845-17851.

Chakrabarti, A, TR Ganapathi, PK Mukherjee, and VA Bapat. 2003. MSI-99, a magainin analogue, imparts enhanced disease resistance in transgenic tobacco and banana. Planta 216: 587-596.

Chang, T-CL and C Karr, Jr. 1959. Gas-liquid chromatographic analysis of aromatic hydrocarbons boiling up to 218° in a low-temperature coal tar. Anal Chim Acta 21: 474-490.

Charest, PJ, N Caléro, D Lachance, RSS Datla, LC Duchêsne, and EWT Tsang. 1993. Microprojectile-DNA delivery in conifer species: factors affecting assessment of transient gene expression using the β-glucuronidase reporter gene. Plant Cell Rep 12: 189-193.

Chen, H and RA Vierling. 2000. Molecular cloning and characterization of soybean peroxidase gene families. Plant Sci 150: 129-137.

Cheng, Z, J Targolli, and R Wu. 2001. Tobacco matrix attachment region sequence increased transgene expression levels in rice plants. Mol Breeding 7: 317-327.

Cho, M-J, HW Choi, BB Buchanan, and PG Lemaux. 1999. Inheritance of tissue-specific expression of barley hordein promoter-*uidA* fusions in transgenic barley plants. Theor Appl Genet 98: 1253-1262.

Choffnes, DS, R Philip, and LO Vodkin. 2001. A transgenic locus in soybean exhibits a high level of recombination. In Vitro Cell Dev Biol – Plant 37: 756-762.

Chowdhury, B and ME John. 1998. Thermal evaluation of transgenic cotton containing polyhydroxybutyrate. Thermochim Acta 313: 43-53.

Clark, RM, TN Wagler, P Quijada, and J Doebley. 2006. A distant upstream enhancer at the maize domestication gene *tb1* has pleiotropic effects on plant and inflorescent architecture. Nat Genet 38: 594-597.

Clough, SJ and AF Bent. 1998. Floral dip: A simplified emthod for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. Plant J 16: 735-743.

Coleman, GD, JM Englert, THH Chen, and LH Fuchigami. 1993. Physiological and environmental requirements for poplar (*Populus deltoides*) bark storage protein degradation. Plant Physiol 102: 53-59.

Comai, L, P Moran, and D Maslyar. 1990. Novel and useful properties of a chimeric plant promoter combining CaMV 35S and MAS elements. Plant Mol Biol 15: 373-381.

Conly, J. 2002. Antimicrobial resistance in Canada. CMAJ 167: 885-891.

Conrad, U. 2005. Polymers from plants to develop biodegradable plastics. Trends Plant Sci 10: 511-512.

Craddock, CP, PR Hunter, E Szakacs, G Hinz, DG Robinson, and L Frigerio. 2008. Lack of vacuolar sorting receptor leads to a non-specific missorting of soluble vacuolar proteins in Arabidopsis seeds. Traffic 9: 408-416.

Crofts, AJ, N Leborgne-Castel, S Hillmer, DG Robinson, B Phillipson, LE Carlsson, DA Ashford, and J Denecke . 1999. Saturation of the endoplasmic reticulum retention machinery reveals anterograde bulk flow. Plant Cell 11: 2233-2247.

Curtis, D, R Lehmann, and PD Zamore. 1995. Translational regulation in development. Cell 81: 171-178.

Daly, JW, J Caceres, RW Moni, F Gusovsky, M Moos, Jr., KB Seamon, K Milton, and CW Myers. 1992. Frog secretions and hunting magic in the upper Amazon: Identification of a peptide that interacts with an adenosine receptor. Proc Natl Acad Sci USA 89: 10960-10963.

Day, CD, E Lee, J Kobayashi, LD Holappa, H Albert, and DW Ow. 2000. Transgene integration into the same chromosome location can produce alleles that express at a predictable level, or alleles that are differentially silenced. Genes Develop 14: 2869-2880.

De Jaeger, G, S Scheffer, A Jacobs, M Zambre, O Zobell, A Goossens, A Depicker, and G Angenon. 2002. Boosting heterologous protein production in transgenic dicotyledonous seeds using *Phaseolus vulgaris* regulatory sequences. Nature Biotechnol 20: 1265-1268.

De Loose, M, X Danthinne, E Van Bockstaele, M Van Montagu, and A Depicker. 1995. Different 5' leader sequences modulate β-glucuronidase accumulation levels in transgenic *Nicotiana tabacum* plants. Euphytica 85: 209-216.

Decker, CJ and R Parker. 1994. Mechanisms of mRNA degradation in eukaryotes. Trends Biochem Sci 19: 336-340.

DeGray, G, K Rajasekaran, F Smith, J Sanford, and H Daniell. 2001. Expression of an antimicrobial peptide via the chloroplast genome to control phytopathogenic bacteria and fungi. Plant Physiol 127: 852-862.

DeMattos, RB, LK Curtiss, and DL Williams. 1998. A minimally lipidated form of cell-derived apolipoprotein E exhibits isoform-specific stimulation of neurite outgrowth in the absence of exogenous lipids or lipoproteins. J Biol Chem 273: 4206-4212.

Denecke, J, J Botterman, and R Deblaere. 1990. Protein secretion in plant cells can occur via a default pathway. Plant Cell 2: 51-59.

Desfeux, C, SJ Clough, and AF Bent. 2000. Female reproductive tissues are the primary target of *Agrobacterium*-mediated transformation by the Arabidopsis floral-dip method. Plant Physiol 123: 895-904.

Dewey, CE, BD Cox, BE Straw, EJ Bush, and S Hurd. 1999. Use of antimicrobials in swine feeds in the United States. Swine Health Prod 7: 19-25.

Deyholos, MK and LE Sieburth. 2000. Separable whorl-specific expression and negative regulation by enhancer elements within the *AGAMOUS* second intron. Plant Cell 12: 1799-1810.

Di Sansebastiano, G, N Paris, S Marc-Martin, and J Neuhaus. 1998. Specific accumulation of GFP in a non-acidic vacuolar compartment via a C-terminal propeptide-mediated sorting pathway. Plant J 15: 449-457.

Di Sansebastiano, G-P, L Renna, G Piro, and G Dalessandro. 2004. Stubborn GFPs in *Nicotiana tabacum* vacuoles. Plant Biosyst 138: 37-42.

Dickinson, CD, LA Floener, GG Lilley, and NC Nielsen. 1987. Self-assembly of proglycinin and hybrid proglycinin synthesized *in vitro* from cDNA. Proc Natl Acad Sci USA 84: 5525-5529.

Diehn, SH, W Burkhart, and JS Graham. 1993. Purification and partial amino acid sequence of a wound-inducible, developmentally regulated anionic peroxidase from soybean leaves. Biochem Biophys Res Comm 195: 928-934.

Dombrowski, JE, MR Schroeder, SY Bednarek, and NV Raikhel. 1993. Determination of the functional elements within the vacuolar targeting signal of barley lectin. Plant Cell 5: 587-596.

Doyle, E. 2001. Alternatives to antibiotic use for growth promotion in animal husbandry. Food Research Institute Briefings. Food Research Institute, Univ. Wisc. Madison WI.

Elomaa, P, Y Helariutta, RJ Griesbach, M Kotilainen, P Seppänen, and TH Teeri. 1995. Transgene inactivation in *Petunia hybrida* is influenced by the properties of the foreign gene. Mol Gen Genet 248: 649-656.

Emanuelsson, O and G von Heijne. 2001. Prediction of organellar targeting signals. Biochim Biophys Acta 1541: 114-119.
Eszterhas, SK, EE Bouhassira, DIK Martin, and S Fiering. 2002. Transcriptional interference by independently regulated genes occurs in any relative arrangement of the genes and is influenced by chromosomal integration position. Mol Cell Biol 22: 469-479.
Faye, L, A Boulaflous, M Benchabane, V Gomord, and D Michaud. 2005. Protein modification in the plant secretory pathway: current status and practical implications in molecular pharming. Vaccine 23: 1770-1778.
Felton, GW, CB Summers, and AJ Mueller. 1994. Oxidative responses in soybean foliage to herbivory by bean leaf beetle and three-cornered alfalfa hopper. J Chem Ecol 20: 639-650.
Finnegan, J and D McElroy. 1994. Transgene inactivation: Plants fight back! Nature Biotechnol 12: 883-888.
Firek, S, J Draper, MRL Owen, A Gandecha, B Cockburn, and GC Whitelam. 1993. Secretion of a functional single-chain Fv protein in transgenic tobacco plants and cell suspension cultures. Plant Mol Biol 23: 861-870.
Fiume, E, P Christou, S Gianì, and D Breviario. 2004. Introns are key regulatory elements of rice tubulin expression. Planta 218: 693-703.
Florack, DEA, WJ Stiekema, and D Bosch. 1996. Toxicity of peptides to bacteria present in the vase water of cut roses. Postharvest Biol Technol 8: 285-291.
Frigerio, L, M de Virgilio, A Prada, F Faoro, and A Vitale. 1998. Sorting of phaseolin to the vacuole is saturable and requires a short C-terminal peptide. Plant Cell 10: 1031-1042.
Frigerio, L, NA Joliffe, A Di Cola, DH Felipe, N Paris, J-M Neuhaus, JM Lord, A Ceriotti, and LM Roberts. 2001. The internal propeptide of the ricin precursor carries a sequence-specific determinant for vacuolar sorting. Plant Physiol 126: 167-175.
Gallie, DR. 1993. Posttranscriptional regulation of gene expression in plants. Annu Rev Plant Physiol Plant Mol Biol 44: 77-105.
Gallie, DR, DE Sleat, JW Watts, PC Turner, and TMA Wilson. 1987. The 5'-leader sequence of tobacco mosaic virus RNA enhances the expression of foreign gene transcripts *in vitro* and *in vivo*. Nucleic Acid Res 15: 3257-3273.
Gallie, DR and V Walbot. 1992. Identification of the motifs within the tobacco mosaic virus 5'-leader responsible for enhancing translation. Nucleic Acid Res 20: 4631-4638.
Gallie, DR and TE Young. 1994. The regulation of gene expression in transformed maize aleurone and endoperm protoplast. Plant Physiol 106: 929-939.
Gao, A-G, SM Hakimi, CA Mittanck, Y Wu, BM Woerner, DM Stark, DM Shah, J Liang, and CMT Rommens. 2000. Fungal pathogen protection in potato by expression of a plant defensin peptide. Nat Biotechnol 18: 1307-1310.
Gao, D, MR Knight, AJ Trewavas, B Sattelmacher, and C Plieth. 2004. Self-reporting Arabidopsis expressing pH and [Ca^{2+}] indicators unveil ion dynamics in the cytoplasm and in the apoplast under abiotic stress. Plant Physiol 134: 898-908.
Gendloff, EH, B Bowen, and WG Buchholz. 1990. Quantitation of chloramphenicol acetyl transferase in transgenic tobacco plants by ELISA and correlation with gene copy number. Plant Mol Biol 14: 575-583.
Geng, Z, AS Bassi, and M Gijzen. 2004. Enzymatic treatment of soils contaminated with phenol and chlorophenols using soybean seed hulls. Water Air Soil Pollut 154: 151-166.
Geng, Z, KJ Rao, AS Bassi, M Gijzen, and N Krishnamoorthy. 2001. Investigation of biocatalytic properties of soybean seed hull peroxidase. Catal Today 64: 233-238.

Ghosh, JK, D Shaoo, P Guillaud, L Cicéron, D Mazier, I Kustanovich, Y Shai, and A Mor. 1997. Selective cytotoxicity of dermaseptin S3 toward intraerythrocytic *Plasmodium falciparum* and the underlying molecular basis . J Biol Chem 272: 31609-31616.

Gijzen, M. 1997. A deletion mutation at the *ep* locus causes low seed coat peroxidase activity in soybean. Plant J 12: 991-998.

Gijzen, M, SS Miller, L-A Bowman, AK Batchelor, K Boutilier, and BLA Miki. 1999. Localization of peroxidase mRNAs in soybean seeds by *in situ* hybridization. Plant Mol Biol 41: 57-63.

Gijzen, M, R van Huystee, and RI Buzzell. 1993. Soybean seed coat peroxidase: A comparison of high-activity and low-activity genotypes. Plant Physiol 103: 1061-1066.

Gilbertson, L. 2003. Cre-*lox* recombination: Cre-ative tools for plant biotechnology. Trends Biotechnol 21: 550-555.

Gillespie, J, SW Rogers, M Deery, P Dupree, and JC Rogers. 2005. A unique family of proteins associated with internalized membranes in protein storage vacuoles of the *Brassicaceae*. Plant J 41: 429-441.

Gillikan, JW and JS Graham. 1991. Purification and developmental analysis of the major anionic peroxidase from the seed coat of *Glycine max*. Plant Physiol 96: 214-220.

Graham, MY and TL Graham. 1991. Rapid accumulation of anionic peroxidases and phenolic polymers in soybean cotyledon tissues following treatment with *Phytophthora megasperma* f. sp. *Glycinea* wall glucan. Plant Physiol 97: 1445-1455.

Gray, JSS and R Montgomery. 2006. Asymmetric glycosylation of soybean seed coat peroxidase. Carb Res 341: 198-209.

Green, PJ. 1993. Control of mRNA stability in higher plants. Plant Physiol 102: 1065-1070.

Grignon, C and H Sentenac. 1991. pH and ionic conditions in the apoplast. Annu Rev Plant Physiol Plant Mol Biol 42: 103-128.

Grosveld, F, GB van Assendelft, DR Greaves, and G Kollias. 1987. Position-independent, high-level expression of the human β-globin gene in transgenic mice. Cell 51: 975-985.

Gustafson, CE, S Levine, T Katsura, M McLaughlin, MD Aleixo, BK Tamarappoo, AS Verkman, and D Brown. 1998. Vasopressin regulated trafficking of a green fluorescent protein-aquaporin 2 chimera in LLC-PK1 cells. Histochem Cell Biol 110: 377-386.

Gustafsson, C, S Govindarajan, and J Minshull. 2004. Codon bias and heterologous protein expression. Trends Biotechnol 22: 346-353.

Hadlington, JL and J Denecke. 2000. Sorting of soluble proteins in the secretory pathway of plants. Curr Opin Plant Biol 3: 461-468.

Hahn, JJ, AC Eschenlauer, UB Sleytr, DA Somers, and F Srienc. 1999. Peroxisomes as sites for synthesis of polyhydroxyalkanoates in transgenic plants. Biotechnol Prog 15: 1053-1057.

Hampf, M and M Gossen. 2007. Promoter crosstalk effects on gene expression. J Mol Biol 365: 911-920.

Hancock, REW, and R Lehrer. 1998. Cationic peptides: a new source of antibiotics. Trends Biotechnol 16: 82-88.

Happel, N, S Höning, J-M Neuhaus, N Paris, DG Robinson, and SHE Holstein. 2004. *Arabidopsis* μA-adaptin interacts with the tyrosine motif of the vacuolar sorting receptor VSR-PS1. Plant J 37: 678-693.

Hara-Nishimura, I, T Shimada, K Hatano, Y Takeuchi, and M Nishimura. 1998. Transport of storage proteins to protein storage vacuoles is mediated by large precursor accumulating vesicles. Plant Cell 10: 835-836.

Harding, KG, JS Dennis, H von Blottzitz, and STL Harrison. 2007. Environmental analysis of plastic production processes: Comparing petroleum-based polypropylene and polyethylene with biologically-based poly-β-hydroxybutyric acid using life cycle analysis. J Biotechnol 130: 57-66.

Håversen, LA, I Engberg, L Baltzer, G Dolphin, LÅ Hanson, and I Mattsby-Baltzer. 2000. Human lactoferrin and peptides derived from a surface-exposed helical region reduce experimental *Escherichia coli* urinary tract infection in mice. Infect Immun 68: 5816-5823.

He, F, F Hunag, KA Wilson, and A Tan-Wilson. 2007. Protein storage vacuole acidification as a control of storage protein mobilization in soybeans. J Exp Bot 58: 1059-1070.

Hefferon, KL, H Khalilian, MG AbouHaider. 1997. Expression of the PVY^0 coat protein (CP) under the control of the PVX CP gene leader sequence: protein under greenhouse and field conditions against PVY^0 and PVY^N infection in three potato cultivars. Theor Appl Genet 94: 287-292.

Henriksen, A, O Mirza, C Indiani, K Teilum, G Smulevich, KG Welinder, and M Gajhede. 2001. Structure of soybean seed coat peroxidase: A plant peroxidase with unusual stability and haem-apoprotein interactions. Protein Sci 10: 108-115.

Hillmer, S, A Movafeghi, DG Robinson, and G Hinz. 2001. Vacuolar storage proteins are sorted in the *cis*-cisternae of the pea cotyledon Golgi apparatus. J Cell Biol 152: 41-50.

Hinz, G, S Hillmer, M Bäumer, and I Hohl. 1999. Vacuolar storage proteins and the putative vacuolar sorting receptor BP-80 exit the Golgi apparatus of developing pea cotyledons in different transport vesicles. Plant Cell 11: 1509-1524.

Hinz, G, A Menze, I Hohl, and D Vaux. 1997. Isolation of prolegumin from developing pea seeds: its binding to endomembranes and assembly into prolegumin hexamers in the protein storage vacuole. J Exp Bot 48: 139-149.

Hiraga, S, K Sasaki, H Ito, Y Ohashi, and H Matsui. 2001. A large family of class III plant peroxidases. Plant Cell Physiol 42: 462-468.

Hobbs, SLA, P Kpodar, and CMO DeLong. 1990. The effect of T-DNA copy number, position and methylation on reporter gene expression in tobacco transformants. Plant Mol Biol 15: 851-864.

Hohl, I, DG Robinson, MJ Chrispels, and G Hinz. 1996. Transport of storage proteins to the vacuole is mediated by vesicles without a clathrin coat. J Cell Sci 109: 2539-2550.

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

Holwerda, BC, HS Padgett, and JC Rogers. 1992. Proaleurain vacuolar targeting is mediated by short contiguous peptide interactions. Plant Cell. 4: 307-318.

Hood, EE, MR Bailey, K Beifuss, M Magallanes-Lundback, ME Horn, E Callaway, C Drees, DE Delaney, R Clough, and JA Howard. 2003. Criteria for high-level expression of a fungal laccase gene in transgenic maize. Plant Biotech J 1: 129-140.

Horvath, H, J Huang, O Wong, E Kohl, T Okita, G Kannangara, and D von Wettstein. 2000. The production of recombinant proteins in transgenic barley grains. PNAS 97: 1914-1919

Houmiel, KL, S Slater, D Broyles, L Casagrande, S Colburn, K Gonzalez, TA Mitsky, SE Reiser, D Shah, NB Taylor, M Tran, HE Valentin, and KJ Gruys. 1999. Poly(β-hydroxybutyrate) production in oilseed leucoplasts of *Brassica napus*. Planta 209: 547-550.

Howell, JA, AD Matthews, KC Swanson, DL Harmon, and JC Matthews. 2001. Molecular identification of high-affinity glutamate transporters in sheep and cattle forestomach, intestine, liver, kidney, and pancreas. J Anim Sci 79: 1329-1336.

Huang, MTF and CM Gorman. 1990. Intervening sequences increase efficiency of RNA 3' processing and accumulation of cytoplasmic RNA. Nucleic Acid Res 18: 937-947.

Huang, Y, RO Nordeen, M Di, LD Owens, and JH McBeath. 1997. Expression of an engineered cecropin gene cassette in transgenic tobacco plants confers disease resistance to *Pseudomonas syringae* pv. *Tabaci*. Phytopathology 87: 494-499.

Hunter, PR, CP Craddock, S Di Benedetto, LM Roberts, and L Frigerio. 2007. Fluorescent reporter proteins for the tonoplast and the vacuolar lumen identify a single vacuolar compartment in Arabidopsis cells. Plant Physiol 145: 1371-1382.

Iglesias, VA, EA Moscone, I Papp, F Neuhuber, S Michalowski, T Phelan, S Spiker, M Matzke, and AJM Matzke. 1997. Molecular and cytogenetic analyses of stably and unstably expressed transgene loci in tobacco. Plant Cell 9: 1251-1264.

Ingelbrecht, ILW, LMF Herman, RA Dekeyser, MC Van Montagu, and AG Depicker. 1989. Different 3' end regions strongly influence the level of gene expression in plant cells. Plant Cell 1: 671-680.

Jack, J, D Dorsett, Y Delotto, and S Liu. 1991. Expression of the *cut* locus in the *Drosophila* wing margin is required for cell type specification and is regulated by a distant enhancer. Development 113: 737-747.

Jagannath, A, P Bandyopadhyay, N Arumugam, V Gupta, PK Burma, and D Pental. 2001. The use of a Spacer DNA fragment insulates the tissue-specific expression of a cytotoxic gene (*barnase*) and allows high-frequency generation of transgenic male sterile lines in *Brassica juncea* L. Mol Breeding 8: 11-23.

Jakowitsch, J, I Papp, EA Moscone, J van der Winden, M Matzke, and AJM Matzke. 1999. Molecular and cytogenetic characterization of a transgene locus that induces silencing and methylation of homologous promoters in *trans*. Plant J 17: 131-140.

James, Clive. 2007. Global Status of Commercialized Biotech/GM Crops: 2007. *ISAAA Brief* No. 37. ISAAA: Ithaca, NY.

Jauh, G-Y, AM Fischer, HD Grimes, CA Ryan, Jr., and JC Rogers. 1998. δ-Tonoplast intrinsic protein defines unique plant vacuole functions. Proc Natl Acad Sci USA 95: 12995-12999.

Jauh, G-Y, TE Phillips, and JC Rogers. 1999. Tonoplast intrinsic protein isoforms as markers for vacuolar function. Plant Cell 11: 1867-1882.

Jaynes, JM, P Nagpala, L Destéfano-Beltrán, JH Huang, J Kim, T Denny, and S Cetiner. 1993. Expression of a cecropin B lytic peptide analog in transgenic tobacco confers enhanced resistance to bacterial wilt caused by *Pseudomonas solanacearum*. Plant Sci 89: 43-53.

Jendrossek, D, A Schirmer, and HG Schlegel. 1996. Biodegradation of polyhydroxyalkanoic acids. Appl Microbiol Biotechnol 46: 451-463.

Jiang, L, TE Phillips, CA Hamm, YM Drozdowicz, PA Rea, M Maeshima, SW Rogers, and JC Rogers. 2001. The protein storage vacuole: A unique compound organelle. J Cell Biol 155: 991-1002.

Jiang, L, TE Phillips, SW Rogers, and JC Rogers. 2000. Biogenesis of the protein storage vacuole crystalloid. J Cell Biol 150: 755-769.

Jiang, L and SSM Sun. 2002. Membrane anchors for vacuolar targeting: application in plant bioreactors. Trend Biotechnol 20: 99-102.

John, ME and G Keller. 1996. Metabolic pathway engineering in cotton: Biosynthesis of polyhydroxybutyrate in fiber cells. Proc Natl Acad Sci USA 93: 12768-12773.
Jolliffe, NA, CP Craddock, and L Frigerio. 2005. Pathways for protein transport to seed storage vacuoles. Biochem Soc Trans 33: 1016-1018.
Jürgens, G. 2004. Membrane trafficking in plants. Annu Rev Cell Dev Biol 20: 481-504.
Kalfa, VC, D Palmquist, MR Ackermann, and KA Brogden. 2001. Suppression of *Mannheimia (Pasteurella) haemolytica* serovar 1 infection in lambs by intrapulmonary administration of ovine antimicrobial anionic peptide. Int J Antimicrob Agents 17: 505-510.
Kamal, JKA and DV Behere. 2002. Thermal and conformational stability of seed coat soybean peroxidase. Biochemistry 41: 9034-9042.
Kamal, JKA and DV Behere. 2003. Activity, stability and conformational flexibility of seed coat sobyean peroxidase. J Inorg Biochem 94: 236-242.
Kamal, JKA and DV Behere. 2008. Kinetic stabilities of soybean and horseradish peroxidases. Biochem Eng J 38: 110-114.
Kaufman, RJ. 1994. Control of gene expression at the level of translation initiation. Curr Opin Biotechnol 5: 550-557.
Kirsch, T, N Paris, JM Butler, L Beevers, and JC Rogers. 1994. Purification and initial characterization of a potential plant vacuolar targeting receptor. Proc Natl Acad Sci USA 91: 3403-3407.
Kis, M, E Burbridge, IW Brock, L Heggie, PJ Dix, and TA Kavanagh. 2004. An N-terminal peptide extension results in efficient expression, but not secretion, of a synthetic horseradish peroxidase gene in transgenic tobacco. Ann Bot 93: 303-310.
Klahre, U, T Noguchi, S Fujioka, S Takatsuto, T Yokoto, T Nomura, S Yoshida, and N-H Chua. 1998. The Arabidopsis *DIMINUTO/DWARF1* gene encodes a protein involved in steroid synthesis. Plant Cell 10: 1677-1690.
Klausner, RD, TA Rouault, and JB Harford. 1993. Regulating the fate of mRNA: the control of cellular iron metabolism. Cell 72: 19-28.
Knutson, K, S Kirzan, and A Ragauskas. 2005. Enzymatic biobleaching of two recalcitrant paper dyes with horseradish and soybean peroxidase. Biotechnol Lett 27: 753-758.
Koide, Y, H Hirano, K Matsuoka, and K Nakamura. 1997. The N-terminal propeptide of the precursor to sporamin acts as a vacuole-targeting signal even at the C terminus of the mature part in tobacco cells. Plant Physiol 114: 863-870.
Kornfeld, S and Mellman I. 1989. The biogenesis of lysosomes. Annu Rev Cell Biol 5: 483-525.
Kourtz, L, K Dillon, S Daughtry, LL Madison, O Peoples, and KD Snell. 2005. A novel thiolase-reductase gene fusion promotes the production of polyhydroxybutyrate in *Arabidopsis*. Plant Biotechnol J 3: 435-447.
Kourtz, L, K Dillon, S Daughtry, OP People, and KD Snell. 2007. Chemically inducible expression of the PHB biosynthetic pathway in *Arabidopsis*. Transgenic Res 16: 759-769.
Krumm, A, L Madisen, X-J Yang, R Goodman, Y Nakatani, and M Groudine. Long-distance transcriptional enhancement by the histone acetyltransferase PCAF. Proc Natl, Acad Sci USA 95: 13501-13506.
Kurkdjian, A and J Guern. 1989. Intracellular pH: Measurement and importance in cell activity. Annu Rev Plant Physiol Plant Mol Biol 40: 271-303.
Laemmli, UK. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227: 680-685.

Lazarev, VN, L Stipkovits, J Biro, D Miklodi, MM Shkarupeta, GA Titova, TA Akopian, and VM Govorun. 2004. Induced expression of the antimicrobial peptide melittin inhibits experimental infections by *Mycoplasma gallisepticum* in chickens. Microbes Infect 6: 536-541.

Le Hir, H, A Nott, and MJ Moore. 2003. How introns influence and enhance eukaryotic gene expression. Trends Biochem Sci 28: 215-220.

Lee, L-Y, ME Kononov, B Bassuner, BR Frame, K Wang, and SB Gelvin. 2007. Novel plant transformation vectors containing the superpromoter. Plant Physiol 145: 1294-1300.

Lee, S-J, RS Saravanan, CMB Damasceno, H Yamane, B-D Kim, and JKC Rose. 2004. Digging deeper into the plant cell wall proteome. Plant Physiol Biochem 42: 979-988.

Lettice, LA, SJH Heaney, LA Purdie, L Li, P de Beer, BA Oostra, D Goode, G Elgar, RE Hill, and E de Graaff. 2003. A long-range *Shh* enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Hum Mol Genet 12: 1725-1735.

Li, Y-B, SW Rogers, YC Tse, SW Lo, SSM Sun, G-Y Jauh, and L Jiang. 2002. BP-80 homologs are concentrated on post-Golgi, probable lytic prevacuolar compartments. Plant Cell Physiol 43: 726-742.

Liu, Q, J Ingersoll, L Owens, S Salih, R Meng, and F Hammerschlag. 2001. Response of transgenic Royal Gala apple (*Malus* x *domestica* Borkh.) shoots carrying a modified cecropin *MB39* gene, to *Erwinia amylovora*. Plant Cell Rep 20: 306-312.

Liu, J-W and LM Tabe. 1998. The influences of two plant nuclear matrix attachment regions (MARs) on gene expression in transgenic plants. Plant Cell Physiol 39: 115-123.

Lorin, C, H Saidi, A Belaid, A Zairi, F Baleux, H Hocini, L Be'lec, K Hani, and F Tangy. 2005. The antimicrobial peptide Dermaseptin S4 inhibits HIV-1 infectivity *in vitro*. Virology 334: 264-275.

Lössl, A, K Bohmert, H Harloff, C Eibl, S Mühlbauer, and H-U Koop. 2005. Inducible trans-activation of plastid transgenes: expression of the *R. eutropha phb* operon in transplastomic tobacco. Plant Cell Physiol 46: 1462-1471.

Lössl, A, C Eibl, H-J Harloff, C Jung, and H-U Koop. 2003. Polyester synthesis in transplastomic tobacco (*Nicotiana tabacum* L.): significant contents of polyhydroxybutyrate are associated with growth reduction. Plant Cell Rep 21: 891-899.

Lu, Y, JP Gehan, and TD Sharkey. 2005. Daylength and circadian effects on starch degradation and maltose metabolism. Plant Physiol 138: 2280-2291.

Malik, K, K Wu, X-Q Li, T Martin-Heller, M Hu, E Foster, L Tian, C Wang, K Ward, M Jordan, D Brown, S Gleddie, D Simmonds, S Zheng, J Simmonds, and B Miki. 2002. A constitutive gene expression system derived from the *tCUP* cryptic promoter elements. Theor Appl Genet 105: 505-514.

Marolda, CL, J Vicarioli, and MA Valvano. 2004. Wzx proteins involved in biosynthesis of O antigen function in association with the first sugar of the O-specific lipopolysaccharide subunit. Microbiology 150: 4095-4105.

Marty, F. 1999. Plant vacuoles. Plant Cell 11: 587-599.

Maruyama, N, LC Mun, M Tatsuhara, M Sawada, M Ishimoto, and S Utsumi. 2006. Multiple vacuolar sorting determinants exist in soybean 11S globulin . Plant Cell 18: 1253-1273.

Matanic, VCA and V Castilla. 2004. Antiviral activity of antimicrobial cationic peptides against Junin virus and herpes simplex virus. Int J Antimicrobiol Agents 23: 382-389.

Matsui, T, M Hori, N Shizawa, H Nakayama, A Shinmyo, and K Yoshida. 2006. High-efficiency secretory production of peroxidase C1a using vesicular transport engineering in transgenic tobacco. J Biosci Bioeng 102: 102-109.

Matsui, T, H Nakayama, K Yoshida, and A Shinmyo. 2003. Vesicular transport route of horseradish C1a peroxidase is regulated by N- and C-terminal propeptides in tobacco cells. Appl Microbiol Biotechnol 62: 517-522.

Matsuoka, K, DC Bassham, NV Raikhel, and K Nakamura. 1995. Different sensitivity to wortmannin of two vacuolar sorting signals indicates the presence of distinct sorting machineries in tobacco cells. J Cell Biol 130: 1307-1318.

Matsuoka, K and K Nakamura. 1991. Propeptide of a precursor to a plant vacuolar protein required for vacuolar targeting. Proc Natl Acad Sci USA 88: 834-838.

Matsuoka, K and J-M Neuhaus. 1999. *Cis*-elements of protein transport to the plant vacuoles. J Exp Bot 50: 165-174.

Matzke, AJM and MA Matzke. 1998. Position effects and epigenetic silencing of plant transgenes. Curr Opin Plant Biol 1: 142-148.

Matzke, MA, MF Mette, and AJM Matzke. 2000. Transgene silencing by the host genome defense: implications for the evolution of epigenetic control mechanisms in plants and vertebrates. Plant Mol Biol 43: 401-415.

Mazumder, B, V Seshadri, and PL Fox. 2003. Translational control by the 3'-UTR: the ends specify the means. Trends Biochem Sci 28: 91-98.

McCarthy, JEG and H Kollmus. 1995. Cytoplasmic mRNA-protein interactions in eukaryotic gene expression. Trends Biochem Sci 20: 191-197.

Mignone, F, C Gissi, S Liuni, and G Pesole. 2002. Untranslated regions of mRNAs. Genome Biol 3: reviews0004.1-0004.10.

Miller, SS, LA Bowman, M Gijzen, BLA Miki. 1999. Early development of the seed coat of soybean (*Glycine max*). Ann Bot 84: 297–304.

Mittendorf, V, V Bongcam, L Allenbach, G Coullerez, N Martini, and Y Poirier. 1999. Polyhydroxyalkanoate synthesis in transgenic plants as a new tool to study carbon flow through β-oxidation. Plant J 20: 45-55.

Mittendorf, V, EJ Robertson, RM Leech, N Krüger, A Steinbüchel, and Y Poirier. 1998. Synthesis of medium-chain-length polyhydroxyalkanoates in *Arabidopsis thaliana* using intermediates of peroxisomal fatty acid *β*-oxidation. Proc Natl Acad Sci USA 95: 13397-13402.

Moïse, JA, S Han, L Gudynaitę-Savitch, DA Johnson, and BLA Miki. 2005. Seed coats: structure, development, composition, and biotechnology. In Vitro Cell Dev Biol-Plant 41: 620-644.

Mor, A, VH Nguyen, A Delfour, D Migliore-Smour, and P Nicolas. 1991. Isolation, amino acid sequence, and synthesis of dermaseptin, a novel antimicrobial peptide of amphibian skin. Biochemistry 30: 8824-8830.

Mor, A and P Nicolas. 1994. The NH2-terminal α-helical domain 1-18 of dermaseptin is responsible for antimicrobial activity. J Biol Chem 269: 1934-1939.

Mor, A and P Nicolas. 1994. Isolation and structure of novel defensive peptides from frog skin. Eur J Biochem 219: 145-154.

Mori, T, N Maruyama, K Nishizawa, T Higasa, K Yagasaki, M Ishimoto, and S Utsumi. 2004. The composition of newly synthesized proteins in the endoplasmic reticulum determines the transport pathways of soybean seed storage proteins. Plant J 40: 238-249.

Müntz, K. 2007. Protein dynamics and proteolysis in plant vacuoles. J Exp Bot 58: 2391-2407.
Murray, C, PW Sutherland, MM Phung, MT Lester, RK Marshall, and JT Christeller. 2002. Expression of biotin-binding proteins, avidin and streptavidin, in plant tissues using plant vacuolar targeting sequences. Transgenic Res 11: 199-212.
Nakashita, H, Y Arai, T Shikanai, Y Doi, and I Yamaguchi. 2001. Introduction of bacterial metabolism into higher plants by polycistronic transgene expression. Biosci Biotechnol Biochem 65: 1688-1691.
Nakashita, H, Y Arai, K Yoshioka, T Fukui, Y Doi, R Usami, K Horikoshi, and I Yamaguchi. 1999. Production of biodegradable polyester by a transgenic tobacco. Biosci Biotechnol Biochem 63: 870-874.
Nawrath, C, Y Poirier, and C Somerville. 1994. Targeting of the polyhydroxybutyrate biosynthetic pathway to the plastids of *Arabidopsis thaliana* results in high levels of polymer accumulation. Proc Natl Acad Sci USA 91: 12760-12764.
Neuhaus, J-M, M Pietrzak, and T Boller. 1994. Mutation analysis of the C-terminal vacuolar targeting peptide of tobacco chitinase: low specificty of the sorting system, and gradual transition between intracellular retention and secretion into the extracellular space. Plant J 5: 45-54.
Ni, M, D Cui, J Einstein, S Narasimhulu, CE Vergara, and SB Gelvin. 1995. Strength and tissue specificity of chimeric promoters derived from the octopine and mannopine synthase genes. Plant J 7: 661-676.
Nishizawa, K, N Maruyama, R Satoh, Y Fuchikami, T Higasa, and S Utsumi. 2003. A C-terminal sequence of soybean β-conglycinin α' subunit acts as a vacuolar sorting determinant in seed cells. Plant J 34: 647-659.
Nishizawa, K, N Maruyama, and S Utsumi. 2006. The C-terminal region of α' subunit of soybean β-conglycinin contains two types of vacuolar sorting determinants. Plant Mol Biol 62: 111-125.
Norris, SR, SE Meyer, and J Callis. 1993. The intron of *Arabidopsis thaliana* polyubiquitin genes is conserved in location and is a quantitative determinant of chimeric gene expression. Plant Mol Biol 21: 895-906.
Okamoto, T, T Tanida, B Wei, E Ueta, T Yamamoto, and T Osaki. 2004. Regulation of fungal infection by a combination of amphotericin B and peptide 2, a lactoferrin peptide that activates neutrophils. Clin Diagn Lab Immun 11: 1111-1119.
Okamoto, T, A Yuki, N Mitsuhashi, and T Mimamikawa. 1999. Asparaginyl endopeptidase (VmPE-1) and autocatalytic processing synergistically activate the vacuole cysteine proteinase (SH-EP). Eur J Biochem 264: 223-232.
Osusky, M, L Osuska, RE Hancock, WW Kay, and S Misra. 2004. Transgenic potatoes expressing a novel cationic peptide are resistant to late blight and pink rot. Transgenic Res 13: 181-190.
Osusky, M, L Osuska, W Kay, and S Misra. 2005. Genetic modification of potato against microbial diseases: *in vitro* and *in planta* activity of a dermaseptin B1 derivative, MsrA2. Theor Appl Genet 111: 711-722.
Osusky, M, G Zhou, L Osuska, RE Hancock, WW Kay, and S Misra. 2000. Transgenic plants expressingcationic peptide chimeras exhibit broad-spectrum resistance to phytopathogens. Nat Biotechnol 18: 1162-1166.

Papo, N, Z Oren, U Pag, H-G Sahl, and Y Shai. 2002. The consequence of sequence alteration of an amphipathic a-helical antimicrobial peptide and its diastereomers. J Biol Chem 277: 33913-33921.

Paris, N, SW Rogers, L Jiang, T Kirsch, L Beevers, TE Phillips, and JC Rogers. 1997. Molecular cloning and further characterization of a probably plant vacuolar sorting receptor. Plant Physiol 115: 29-39.

Paris, N, CM Stanley, RL Jones, and JC Rogers. 1996. Plant cells contain two functionally distinct vacuolar compartments. Cell 85: 563-572.

Park, M, SJ Kim, A Vitale, and I Hwang. 2004. Identification of the protein storage vacuole and protein targeting to the vacuole in leaf cells of three plant species. Plant Phys 134: 625-639.

Park, M, D Lee, G-J Lee, and I Hwang. 2005. AtRMR1 functions as a cargo receptor for protein trafficking to the protein storage vacuole. J Cell Biol 170: 757-767.

Park, JH, M Oufattole, and JC Rogers. 2007. Golgi-mediated vacuolar sorting in plant cells: RMR proteins are sorting receptors for the protein aggregation/membrane internalization pathway. Plant Sci 172: 728-745.

Peach, C and J Velten. 1991. Transgene expression variability (position effect) of CAT and GUS reporter genes driven by linked divergent T-DNA promoters. Plant Mol Biol 17: 49-60.

Peoples, OP and AJ Sinskey. 1989a. Poly-β-hydroxybutyrate biosynthesis in *Alcaligenes eutrophus* H16: Characterization of the genes encoding β-ketothiolase and acetoacetyl-CoA reductase. J Biol Chem 264: 15293-15297.

Peoples, OP and AJ Sinskey. 1989b. Poly-β-hydroxybutyrate (PHB) biosynthesis in *Alcaligenes eutrophus* H16: Identification and characterization of the PHB polymerase gene (*phbC*). J Biol Chem 264: 15298-15303.

Petersen, K, R Leah, S Knudsen, and V Cameron-Mills. 2002. Matrix attachment regions (MARs) enhance transformation frequencies and reduce variance of transgene expression in barley. Plant Mol Biol 49: 45-58.

Petrasovits, LA, MP Purnell, LK Nielsen, and SM Brumbley. 2007. Production of polyhydroxybutyrate in sugarcane. Plant Biotechnol J 5: 162-172.

Phillipson, BA, P Pimpl, LL Pinto-daSilva, AJ Crofts, JP Taylor, A Movafeghi, DG Robinson, and J Denecke. 2001. Secretory bulk flow of soluble proteins is efficient and COPII dependent. Plant Cell 13: 2005-2020.

Pimpl, P, SL Hanton, JP Taylor, LL Pinto-daSilva, and J Denecke. 2003. The GTPase ARF1p controls the sequence-specific vacuolar sorting route to the lytic vacuole. Plant Cell 15: 1242-1256.

Poirier, Y. 1999. Production of new polymeric compounds in plants. Curr Opin Biotechnol 10: 181-185.

Poirier, Y, D Dennis, K Klomparens, C Nawrath, and C Somerville. 1992a. Perspectives on the production of polyhydroxyalkanoates in plants. FEMS Microbiol Rev 103: 237-246.

Poirier, Y, DE Dennis, K Klomparens, and C Sommerville. 1992b. Polyhydroxybutyrate, a biodegradable thermoplastic, produced in transgenic plants. Science 256: 520-523.

Ponti, D, ML Mangoni, G Mignogna, M Simmaco, and D Barra. 2003. An amphibian antimicrobial peptide variant expressed in *Nicotiana tabacum* confers resistance to phytopathogens. Biochem J 378: 121-127.

Pooggin, MM and KG Skryabin. 1992. The 5'-untranslated leader sequence of potato virus X RNA enhances the expression of a heterologous gene *in vivo*. Mol Gen Genet 234: 329-331.

Potenza, C, L Aleman, and C Sengupta-Gopalan. 2004. Targeting transgene expression in research, agricultural, and environmental applications: promoters used in plant transformation. In Vitro Cell Dev Biol-Plant 40: 1-22.

Powell, MJ, SAJ Thompson, Y Tone, H Waldmann, and M Tone. 2000. Posttranscriptional regulation of IL-10 gene expression through sequences in the 3'-untranslated region. J Immunology 165: 292-296.

Pramanik, MHR, M Nagai, T Asao, and Y Matsui. 2000. Effects of temperature and photoperiod on phytotoxic root exudates of cucumber (*Cucumis sativus*) in hydroponic culture. J Chem Ecol 26: 1953-1967.

Prescott, M, S Nowakowski, P Nagley, and RJ Devenish. 1999. The length of polypeptide linker affects the stability of green fluorescent protein fusion proteins. Anal Biochem 273: 305-307.

Purnell, MP, LA Petrasovits, LK Nielson, and SM Brumbley. 2007. Spatio-temporal characterization of polyhydroxybutyrate accumulation in sugarcane. Plant Biotechnology J 5: 173-184.

Pütter, J and Becker, R. 1983. In *Methods of Enzymatic Analysis* (Bergmeyer, HU, ed.) 3rd ed., Vol III, pp. 286-293, Verlug Chemie, Deerfield Beach, FL.

Quail, PH. 1979. Plant cell fractionation. Ann Rev Plant Physiol 30: 425-484.

Ramírez, N, M Ayala, D Lorenzo, D Palenzuela, L Herrera, V Doreste, M Pérez, JV Gavilondo, and P Oramas. 2002. Expression of a single-chain Fv antibody fragment specific for the Hepatitis B surface antigen in transgenic tobacco plants. Transgenic Res 11: 61-64.

Rancé, I, F Norre, V Gruber, and M Theisen. 2002. Combination of viral promoter sequences to generate highly active promoters for heterologous therapeutic protein over-expression in plants. Plant Sci 162: 833-842.

Rayon, C, P Lerouge, and L Faye. 1998. The protein N-glycosylation in plants. J Exp Bot 49: 1463-1472.

Reddy, KVR, RD Yedery, and C Aranha. 2004. Antimicrobial peptides: premises and promises. Int J Antimicrob Agents 24 : 536-547.

Regalado, C, BE García-Almendárez, and MA Duarte-Vázquez. 2004. Biotechnological applications of peroxidases. Phytochem Rev 3: 243-256.

Register III, JC, DJ Peterson, PJ Bell, WP Bullock, IJ Evans, B Frame, AJ Greenland, NS Higgs, I Jepson, S Jiao, CJ Lewnau, JM Sillick, and HM Wilson. 1994. Structure and function of selectable and non-selectable transgenes in maize after introduction by particle bombardment. Plant Mol Biol 25: 951-961.

Rippe, RA, S-I Lorenzen, DA Brenner, and M Breindl. 1989. Regulatory elements in the 5'-flanking region and the first intron contribute to transcriptional control of the mouse alpha 1 type I collagen gene. Mol Cell Biol 9: 2224-2227.

Ritzenthaler, C, A Nebenführ, A Movafeghi, C Stussi-Garaud, L Behnia, P Pimpl, LA Staehelin, and DG Robinson. 2002. Reevaluation of the effects of brefeldin A on plant cells using tobacco Bright Yellow 2 cells expressing Golgi-targeted green fluorescent protein and COPI antisera. Plant Cell 14: 237-261.

Robinson, DG and G Hinz. 1999. Golgi-mediated transport of seed storage proteins. Seed Sci Res 9: 267-283.

Robinson, DG, P Oliviusson, and G Hinz. 2005. Protein sorting to the storage vacuoles of plants: A critical appraisal. Traffic 6: 615-625.

Rose, AB and RL Last. 1997. Introns act post-transcriptionally to increase expression of the *Arabidopsis thaliana* tryptophan pathway gene PAT1. Plant J 11: 455-464.

Rubinchik, S, S Lower, Z Jia, J Norris, and J Dong. 2001. Creation of a new transgene cloning site near the right ITR of AD5 results in reduced enhancer interference with tissue-specific and regulatable promoters. Gene Therapy 8: 247-253.

Sakharov, IY, IS Alpeeva, and EE Efremov. 2006. Use of soybean peroxidase in chemiluminescent enzyme-linked immunosorbent assay. J Agric Food Chem 54: 1584-1587.

Sakurai, N. 1998. Dynamic function and regulation of apoplast in the plant body. J Plant Res 111: 133-148.

Sambrook. F, EF Fritsch, and T Maniatis. 1989. Molecular Cloning: A Laboratory Manual. 2nd ed. Col Spring Harbor Laboratory Press. Cold Spring Harbor, N.Y.

Saruul, P, F Srienc, DA Somers, and DA Samac. 2002. Production of a biodegradable plastic polymer, poly-β-hydroxybutyrate in transgenic alfalfa. Crop Sci 42: 919-927.

Schäffer, C, M Graninger, and P Messner. 2001. Prokaryotic glycosylation. Proteomics 1: 248-261.

Schägger, H. 2006. Tricine-SDS-PAGE. Nature Protocols 1: 16-22.

Schmitz, N, M Gijzen, and R van Huystee. 1997. Characterization of anionic soybean (Glycine max) seed coat peroxidase. Can J Bot 75: 1336-1341.

Schubert, D, B Lechtenberg, A Forsback, M Gils, S Bahadur, and R Schmidt. 2004. Silencing in Arabidopsis T-DNA transformants: The predominant role of a gene-specific RNA sensing mechanism versus position effects. Plant Cell 16: 2561-2572.

Schubert, P, A Steinbüchel, and GG Schlegel. 1988. Cloning of the *Alcaligenes eutrophus* genes for synthesis of poly-β-hydroxybutyric acid (PHB) and synthesis of PHB in Escherichia coli. J Bacteriol 170: 5837-5847.

Sharma, A, R Sharma, M Imamura, M Yamakawa, and H Machii. 2000. Transgenic expression of cecropin B, an antibacterial peptide from *Bombyx mori*, confers enhanced resistance to bacterial leaf blight in rice. FEBS Lett 484: 7-11.

Shimada, T, E Watanabe, K Tamura, Y Hayashi, M Nishimura, and I Hara-Nishimura. 2002. A vacuolar sorting receptor PV72 on the membrane of vesicles that accumulate precursors of seed storage proteins (PAC vesicles). Plant Cell Physiol 43: 1086-1095.

Shimada, T, K Fuji, K Tamura, M Kondo, M Nishimura, and I Hara-Nishimura. 2003. Vacuolar sorting receptor for seed storage proteins in *Arabidopsis thaliana*. Proc Natl Acad Sci USA 100: 16095-16100.

Shimaoka, T, M Ohnishi, T Sazuka, N Mitsuhashi, I Hara-Nishimura, K-I Shimazaki, M Maeshima, A Yokota, K-I Tomizawa, and T Mimura. 2004. Isolation of intact vacuoles and proteomic analysis of tonoplast from suspension-cultured cells of Arabidopsis thaliana. Plant Cell Physiol 45: 672-683.

Shimoni, Y and G Galili. 1996. Intramolecular disulfide bonds between conserved cysteines in wheat gliadins control their deposition into protein bodies. J Biol Chem 271: 18869-18874.

Siemering, KR, R Golbik, R Sever, and J Haseloff. 1996. Mutations that suppress the thermosensitivity of green fluorescent protein. Curr Biol 6: 1653-1663.

Slater, SC, WH Voige, and DE Dennis. 1988. Cloning and expression in *Escherichia coli* of the *Alcaligenes eutrophus* H16 poly-β-hydroxybutyrate biosynthetic pathway. J Bacteriol 170: 4431-4436.

Slater, S, TA Mitsky, KL Houmiel, M Hao, SE Reiser, NB Taylor, M Tran, HE Valentin, DJ Rodriguez, DA Stone, SR Padgette, G Kishore, and KJ Gruys. 1999. Metabolic engineering of *Arabidopsis* and *Brassica* for poly(3-hydroxybutyrate-*co*-3-hydroxyvalerate) copolymer production. Nat Biotechnol 17: 1011-1016.

Somers, DA, DA Samac, and PM Olhoft. 2003. Recent advances in legume transformation. Plant Physiol 131: 892-899.

Somleva, MV, KD Snell, JJ Beaulieu, OP Peoples, BR Garrison, and NA Patterson. 2008. Production of polyhydroxybutyrate in switchgrass, a value-added co-product in an important lignocellulosic biomass crop. Plant Biotechnol J 6: 663-678.

Spiker, S and WF Thompson. 1996. Nuclear matrix attachment regions and transgene expression in plants. Plant Physiol 110: 15-21.

Springer, PS. 2000. Gene traps: tools for plant development and genomics. Plant Cell 12: 1007-1020.

Srivastava, V and DW Ow. 2004. Marker-free site-specific gene integration in plants. Trends Biotechnol 22: 627-629.

Stam, M, JNM Mol, and JM Kooter. 1997. The silence of genes in transgenic plants. Ann Bot 79: 3-12.

Steinbüchel, A and HE Valentin. 1995. Diversity of bacterial polyhydroxyalkanoic acids. FEMS Microbiol Lett 128: 219-228.

Sternberg, EA, G Spizz, WM Perry, D Vizard, T Weil, and EN Olson. 1988. Identification of upstream and intragenic regulatory elements that confer cell-type-restricted and differentiation-specific expression on the muscle creatine kinase gene. Mol Cell Biol 8: 2896-2909.

Stief, A, DM Winter, WH Strätling, and AE Sippel. 1989. A nuclear DNA attachment element mediates elevated and position-independent gene activity. Nature 341: 343-345.

Strahilevitz, J, A Mor, P Nicolas, and Y Shai. 1994. Spectrum of antimicrobial activity and assembly of dermaseptin-b and its precursor form in phospholipid membranes. Biochem 33: 10951-10960.

Suriyamongkol, P, R Weselake, S Narine, M Moloney, and S Shah. 2007. Biotechnological approaches for the production of polyhydroxyalkanoates in microorganisms and plants - A review. Biotechnol Adv 25: 148-175.

Sutton, DW, PK Havstad, and JD Kemp. 1992. Synthetic *cryIIIA* gene from *Bacillus thuringiensis* improved for high expression in plants. Transgenic Res 1: 228-236.

Takahashi, H, Y Saito, T Kitagawa, S Morita, T Masumura, and K Tanaka. 2005. A novel vesicle derived directly from endoplasmic reticulum is involved in the transport of vacuolar storage proteins in rice endosperm. Plant Cell Physiol 46: 245-249.

Taiz, L. 1992. The plant vacuole. J Exp Biol 172: 113-122.

Tamura, K, T Shimada, E Ono, Y Tanaka, A Nagatani, S Higashi, M Watanabe, M Nishimura, and I Hara-Nishimura. 2003. Why green fluorescent fusion proteins have not been observed in the vacuoles of higher plants. Plant J 35: 545-555.

Tanaka, A, S Mita, S Ohta, J Kyozuka, K Shimamoto, and K Nakamura. 1990. Enhancement of foreign gene expression by a dicot intron in rice but not in tobacco is correlated with an

increased level of mRNA and an efficient splicing of the intron. Nucleic Acids Res 18: 6767-6770.

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

Trick, HN, RD Dinkins, ER Santarem, R Di, V Samoylov, CA Meurer, DR Walker, WA Parrott, JJ Finer, and GB Collins. 1997. Recent advances in soybean transformation. Plant Tissue Culture Biotechnol 3: 9-26.

Ulker, B, GC Allen, WF Thompson, S Spiker, and AK Weissinger. 1999. A tobacco matrix attachment region reduces the loss of transgene expression in the progeny of transgenic tobacco plants. Plant J 18: 253-263.

Vain, P, VA James, B Worland, and JW Snape. 2002. Transgene behaviour across two generations in a large random population of transgenic rice plants produced by particle bombardment. Theor Appl Genet 105: 878-889.

Vain, P, B Worland, A Kohli, JW Snape, P Christou, GC Allen, and WF Thompson. 1999. Matrix attachment regions increase transgene expression levels and stability in transgenic rice plants and their progeny. Plant J 18: 233-242.

Valentin, HE, DL Broyles, LA Casagrande, SM Colburn, WL Creely, PA DeLaquil, HM Felton, KA Gonzalez, KL Houmiel, K Lutke, DA Mahadeo, TA Mitsky, SR Padgette, SE Reiser, S Slater, DM Stark, RT Stock, DA Stone, NB Taylor, GM Thorne, M Tran, and KJ Gruys. 1999. PHA production, from bacteria to plants. Int J Biol Macromol 25: 303-306.

Valério, L, M De Meyer, C Penel, and C Dunand. 2004. Expression analysis of the Arabidopsis peroxidase multigenic family. Phytochemistry 65: 1331-1342.

VanCompernolle, SE, RJ Taylor, K Oswald-Richter, J Jiang, BE Youree, JH Bowie, MJ Tyler, M Conlon, D Wade, C Aiken, TS Dermody, VN KewalRamani, LA Rollins-Smith, and D Unutmazi. 2005. Antimicrobial peptides from amphibian skin potently inhibit human immunodeficiency virus infection and transfer of virus from dendritic cells to T cells. J Virol 79: 11598-11606.

van der Biezen, EA. 2001. Quest for antimicrobial genes to engineer disease-resistant crops. Trends Plant Sci 6: 89-91.

Vasil, V, M Clancy, RJ Ferl, IK Vasil, and LC Hannah. 1989. Increased gene expression by the first intron of maize *Shrunken-1* locus in grass species. Plant Phyiol 91: 1575-1579.

Vitale, A and J Denecke. 1999. The endoplasmic reticulum – gateway of the secretory pathway. Plant Cell 11: 615-628.

Vitale, A and G Hinz. 2005. Sorting of proteins to storage vacuoles: How many mechanisms? Trends Plant Sci 10: 316-323.

Walters, DA, CS Vetsch, DE Potts, and RC Lundquist. 1992. Transformation and inheritance of a hygromycin phosphotransferase gene in maize plants. Plant Mol Biol 18: 189-200.

Watanabe, E, T Shimada, K Tamura, R Matsushima, Y Koumoto, M Nishimura, and I Hara-Nishimura. 2004. An ER-localized form of PV72, a seed-specific vacuolar sorting receptor, interferes the transport of an NPIR-containing proteinase in Arabidopsis leaves Plant Cell Physiol 45: 9-17.

Watanabe, E, T Shimada, M Kuroyanagi, M Nishimura, and I Hara-Nishimura. 2002. Calcium-mediated association of a putative vacuolar sorting receptor PV72 with a propeptide of 2S albumin. J Biol Chem 277: 8708-8715.

Welinder, KG and YB Larsen. 2004. Covalent structure of soybean seed coat peroxidase. Biochim Biophys Acta 1698: 121-126.

Welinder, KG, AF Justesen, IVH Kjærsgård, RB Jensen, SK Rasmussen, HM Jespersen, and L Duroux. 2002. Structural diversity and transcription of class III peroxidases from *Arabidopsis thaliana*. Eur J Biochem 269: 6063-6081.

Wenzel, D, G Scharuermann, A von Lüpke, and G Hinz. 2005. The cargo in vacuolar storage protein transport vesicles is stratified. Traffic 6: 45-55.

Wilson, C, HJ Bellen, and WJ Gehring. 1990. Position effects on eukaryotic gene expression. Annu Rev Cell Biol 6: 679-714.

Winter, H, DG Robinson, and HW Heldt. 1994. Subcellular volumes and metabolite concentrations in spinach leaves. Planta 193: 530-535.

Wróbel, M, J Zebrowski, and J Szopa. 2004. Polyhydroxybutyrate synthesis in transgenic flax. J Biotechnol 107: 41-54.

Yang, J, LA Barr, SR Fahnestock, and Z-B Liu. 2005. High yield recombinant silk-like protein production in transgenic plants through protein targeting. Transgenic Res 14: 313-324.

Yanisch-Perron, C, J Vieira, and J Messing. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33: 103-119.

Yedery, RD and KVR Reddy. 2005. Antimicrobial peptides as microbicidal contraceptives: prophecies for prophylactics – a mini review. Eur J Contracept Reprod Health Care 10: 32-42.

Yevtushenko, DP, R Romero, BS Forward, RE Hancock, WW Kay, and S Misra. 2004. Pathogen-induced expression of a cecropin A-melittin antimicrobial peptide gene confers antifungal resistance in transgenic tobacco. J Exp Bot 56: 1685-1695.

Yoo, SY, K Bomblies, SK Yoo, JW Yang, MS Choi, JS Lee, D Weigel, and JH Ahn. 2005. The 35S promoter used in a selectable marker gene of a plant transformation vector affects the expression of the transgene. Planta 221: 523-530.

Yoshida, K, T Matsui, and A Shinmyo. 2004. The plant vesicular transport engineering for production of useful recombinant proteins. J Mol Catal B: Enzym 28: 167-171.

Zairi, A, A Belaïd, A Gahbiche, and K Hani. 2005. Spermicidal activity of dermaseptins. Contraception 72: 447-453.

Zasloff, M. 2002. Antimicrobial peptides of multicellular organisms. Nature 415: 389-395.

Appendix I

INTRODUCTION

The goal of this research project was to develop a transgenic system for the heterologous expression of bioproducts in soybean seed coats. Soybean seed coats were chosen because they presented a number of advantages, as described in Section 1.1. In addition to production of the enzyme SBP and the bioplastic PHB, I attempted to produce an antimicrobial peptide (AMP) in this system. This attempt was unsuccessful but the experimental system will be described as well as the limitations that resulted in a lack of success in this Appendix. The following introduction will provide relevant background information about AMPs and their expression in plant systems, as well as the AMP selected for expression in soybean seed coats, known as MsrA2.

Structure and function of antimicrobial peptides

AMPs are a group of small peptides (12 to 50 amino acids) that form α-helices and/or β-sheets in hydrophobic environments (Hancock and Lehrer 1998; Zasloff 2002). They typically have multiple positively charged lysine and arginine residues and a high hydrophobic content (Brown and Hancock 2006). Their function is dependent on their size, amino acid content, overall charge, three-dimensional structure and degree of hydrophobicity and amphipathicity (Brogden 2005).

AMPs are an important part of host defense in most organisms. They exhibit antimicrobial activity against a wide range of targets, including prokaryotes, fungi, protozoa (Zasloff 2002) and enveloped viruses, including the human immunodeficiency virus (HIV) and the Herpes Simplex Virus (HSV) (Matanic and Castilla 2004; Lorin et al. 2005; VanCompernolle et al. 2005). *In vitro* these proteins can function at concentrations as low as 0.25 to 4 μg/ml (Hancock and Lehrer 1998).

AMPs are attracted to the net negative charge of the outer membrane of bacteria, where they insert into the lipid membrane, disrupting the membrane structure and ultimately inducing the

formation of pores that inhibit or destroy the microbes (Brogden 2005). AMPs can additionally cause microbial death through a number of other mechanisms. Some AMPs can act outside of the microbe by targeting autolysins and phospholipases while others are capable of crossing the cytoplasmic membrane of the microbes and acting intracellularly, by inhibiting vital cellular processes such as nucleic acid synthesis or enzymatic activity (Brogden 2005). Also, AMPs appear to play other important roles in innate defence, including activating other immune responses, stimulating wound healing, and suppressing inflammation (Brown and Hancock 2006).

MsrA2 - a modified dermaseptin

Dermaseptins are a group of helical AMPs that have been isolated from the skin of the frogs *Phyllomedusa sauvagii* and *P. bicolour* (Mor et al. 1991; Daly et al. 1992; Mor et al. 1994). They are active against yeast, bacteria, and fungi, and some are also active against HSV and HIV, but they have low hemolytic activity (Mor et al. 1994; Mor and Nicholas 1994; Ghosh et al. 1997; Belaid et al. 2002; Lorin et al. 2005). Dermaseptin b is 27 amino acids long with alternating hydrophobic and hydrophilic sequences, creating 85% α-helical content (Mor et al. 1994). Synthesized as a 78 amino acid precursor, its signal peptide and acidic spacer are removed at maturity as are an additional four amino acids at the N-terminus and three amino acids at the C-terminus (Amiche et al. 1994). MsrA2 is a synthetic dermaseptin b derivative composed of both the dermaseptin b peptide as well as the four amino acid N-terminal propeptide plus an additional N-terminal Met for a total of 32 amino acids (Osusky et al. 2005). This derivative has significantly improved antimicrobial activity over dermaseptin b (Strahilevitz et al. 1994; Osusky et al. 2005).

MsrA2 is active against the bacteria *Escherichia coli* and *Erwinia carotovora* and the fungi *Fusarium solani*, *F. oxysporum*, *Alternaria alternate*, *Cercospora beticola*, *Botrytis cinerea*, *Rhizoctonia solani*, *Phytophthora cactorum*, *P. erythroseptica*, *Pythium irregulare*, and *P.*

paroecandrum. When MsrA2 was expressed in potato, it showed activity against *E. carotovora*, *Phytophthora infestans*, *P. erythroseptica*, *Rhizoctonia solani*, *Pythium paroecandrum*, *P. splendens*, and *Fusarium solani* (Osusky et al. 2005).

MsrA2 is the AMP that we selected to express in soybean seed coats. It has been shown to have activity against a broad range of bacteria and fungi without affecting plant health in potato (Osusky et al. 2005) and, additionally, it has low hemolytic activity.

Expression of AMPs in plant systems

Bacterial and fungal diseases can devastate crops and are responsible for billions of dollars in crop loss every year. As a result, one goal of agricultural research is to increase plant resistance to such diseases. Genetic engineering has provided a new avenue for improving crop resistance, and AMPs have been considered as good candidates for such an approach. One advantage of AMPs is that some demonstrate broad antimicrobial activity against multiple, unrelated pathogens and such AMPs would therefore be able to produce broadly resistant plants. In addition, because they are small peptides, the cost of production for the plants would be low and few studies have indicated that they have any negative impact on plant health. Heterologous expression of AMPs has been achieved in a number of plants, including tobacco, potato, banana, rice, and apple (Table 8.1 and references therein) and the transgenic plants typically display improved resistance to many economically important pathogens. These studies demonstrate that it is possible to express functional AMPs in plant systems.

Applications of AMPs

Expression of AMPs in plants has, to date, focused on enhancing disease resistance in order to develop more sustainable agricultural practices (van der Biezen 2001). However, a number of other possible applications for AMPs have been considered. In order for these applications to be

Table AI.1 Summary of studies in which antimicrobial peptides were expressed in plants. *In planta* activity refers to experiments in which plant tissue was used for activity measurements. *In vitro* activity refers to experiments in which plant extracts were used for activity measurements.

AMP	Plant Species	*In planta* activity	*In vitro* activity	Phenotypic Effects	References
MSI-99	Tobacco; Banana (*Musa* spp. cv. Rasthali (AAB))	Tobacco-*Sclerotinia sclerotiorum; Alternario alternate; Botrytis cinerea*; Banana-*Fusarium oxysporum* f.sp. *cubense; Mycospaerella musicola*	N/A	N/A	Chakrabarti et al. 2003
MSI-99	Tobacco	*Pseudomonas syringae* cv. *tabaci; Colletotrichum destructivum*	*Pseudomonas syringae* cv. *tabaci; Aspergillus flavus; Fusarium moniliforme*; *Verticillium dahliae*	No effects	DeGray et al. 2001
alfAFP	Potato	*Verticillium dahliae*	N/A	No effects	Gao et al. 2000
MsrA3	Potato	*Erwinia carotovora*; *Phytophthora infestans*; *P. erythroseptica*	N/A	N/A	Osusky et al. 2004
MsrA2	Potato	*Erwinia carotovora*; *Phytophthora infestans*; *P. erythroseptica*; Rhizoctonia *solani*; *Pythium paroecandrum*; *P. splendens*; *Fusarium solani*	*Erwinia carotovora*	N/A	Osusky et al. 2005
MsrA1	Potato	*Phytophthora cactorum*; *Fusarium solani*; *Erwinia carotovora*	N/A	Lesion-mimic phenotype in cv.. Russet Burbank	Osusky et al. 2000
Esculentin-1	Tobacco	*Pseudomonas syringae* pv. *tabaci*; *P. aeruginosa*; *P. nicotianae*	*Pseudomonas syringae pv. tabaci*; *P. aeruginosa*	N/A	Ponti et al. 2003
Cecropin A-melittin chimera	Tobacco	*Fusarium solani*	N/A	No effects	Yevtushenko et al. 2005
Tachyplesin I	Potato	*Erwinia carotovora*	*Erwinia* spp.	No effect	Allefs et al. 1996
Cecropin B substitution analogs	Tobacco	*Pseudomonas solanacearum*	*Clavibacter michiganense, Erwinia carotovora, Pseudomonas solanacearum, P. syringae, Xanthomonas campestris*	No effect	Jaynes et al. 1993
Modified cecropin	Tobacco	*Pseudomonas syringae* pv. *tabaci*	N/A	No effects	Huang et al. 1997
Cecropin B	Rice	*Xanthomonas oryzae* pv. *oryzae*	N/A	No effects	Sharma et al. 2000
Modified cecropin *SB37* gene	Royal Gala apple (*Malus* x *domestica* Borkh.)	*Erwinia amylovora*	N/A	No effects	Liu et al. 2001

useful, an economically affordable source of these peptides must be developed. Expression of AMPs in plants has the ability to fill this need and the soybean seed coat system has the potential for producing AMPs on a large scale.

A number of AMPs are currently being considered for clinical applications, and a few have already reached clinical trials (Reddy et al. 2004; Andrès and Dimarcq 2005). However, expression of AMPs in soybean seed coats may be particularly suitable for the treatment of livestock diseases, which could help reduce the high reliance on antibiotics in this sector (Dewey et al. 1999; Doyle 2001). Seed coats containing AMPs could be incorporated into livestock feed, replacing the addition of antibiotics. Since seed coats are already a component of livestock feed, no additional processing of seed coats or feed would be required. There is increasing concern about the rise of antibiotic-resistant strains and studies are now showing the potential that such strains can arise in livestock and pass through the food chain to humans (Barton 2000; Conly 2002). Therefore, there is a growing demand for alternatives to antibiotics, and AMPs have been proposed as one such candidate. However, efficient systems for high level production of AMPs must be developed before this can be realized.

Despite this potential application of AMPs, few studies have looked at the applicability of treating animals with AMPs. Lazarev et al. (2004) demonstrated that treatment of chickens with a recombinant plasmid from which the AMP melittin could be expressed was able to inhibit their subsequent infection by *Mycoplasma gallisepticum*. Similarly, the survival of mice infected by *Candida albicans* or *Aspergillus fumigatus* was extended by injection with α-defensin 1 (Okamoto et al. 2004). Mice treated orally with lactoferrin, a multifunctional immunoregulatory protein that contains an AMP-like domain, 30 minutes after introducing *Escherichia coli* into the urinary bladder showed significantly lower numbers of bacteria 24 hours later (Håversen et al. 2000). In a lamb

model of acute pneumonia, the application of ovine antimicrobial anionic peptide after infection with *Mannheimia* (*Pasteurella*) *haemolytica* was able to significantly reduce the number of bacteria in infected lung tissue as well as reduce inflammation (Kalfa et al. 2001). These studies demonstrate the feasibility of administering AMPs to animals to treat diseases, although more studies will need to be conducted to determine if this is a safe and effective means of treating livestock diseases.

It is also important to determine that the AMPs do not harm the animals. Osusky et al (2000) demonstrated that mice that ingested transgenic potato tubers expressing the antimicrobial peptide MsrA1 showed no adverse health effects compared to control mice and, in addition, their gut microflora also appeared to be unaffected, demonstrating that ingesting MsrA1 in transgenic tubers does not affect animal health. This result suggests that the AMP MsrA1 did not have adverse, unexpected effects on animal health, but more research is required to state this conclusively and to determine if this result is widely applicable to most AMPs.

AMPs may also be useful in a number of other applications. Some AMPs have spermicidal activity and *in vivo* studies have demonstrated that intravaginal application can completely arrest sperm mobility, preventing contraception (Yedery and Reddy. 2005; Zairi et al. 2005). This ability, combined with antimicrobial and antiviral activities, could make AMPs effective as part of vaginal contraceptives that both prevent pregnancy and minimize the contraction of sexually transmitted diseases (Yedery and Reddy 2005). The AMP tachyplesin A is able to control bacterial growth in the vase water of cut roses, suggesting that it might be an alternative to the less environmentally friendly chemicals that are typically used for this purpose (Florack et al. 1996). This study demonstrates the applicability of using AMPs as a general antibacterial agent and could be expanded to any number of applications of a similar nature. Therefore, AMPs can be used as antimicrobial agents in numerous applications in addition to being used to engineer disease resistance plants.

Hypothesis and Objective

Hypothesis: Seed coat-specific expression of a gene encoding an AMP targeted to the vacuole with the SBP N- and C-terminal propeptides results in accumulation of a functional AMP in soybean seed coats.

Objective: Introduce a gene encoding MsrA2 into soybean plants, driven by the seed coat specific *Ep* 1.5 promoter and modified to include the SBP N- and C-terminal propeptides, and measure the accumulation of the AMP by immunoblotting and the function by antimicrobial assay.

MATERIALS AND METHODS

For a description of the general materials and methods used, refer to Chapter 2. Only materials and methods specific to the analysis of plants transformed with the gene encoding modified MsrA2 will be described here.

Sources of chemicals and materials

Coomassie Brilliant Blue G-250 (Schwarz/Mann Biotech, 808274 10 GM); tricine, electrophoresis grade (MP Biomedicals, LLC, 807416); MsrA2 peptide (synthesized by the Brain Research Centre, University of British Columbia, Vancouver, BC); rabbit polyclonal Anti-AMP-1 and Anti-AMP-2 antibodies (SACRI Antibody Services, Calgary, AB).

Construction of the J12 plasmid for soybean transformation

A construct was designed for expression of the antimicrobial peptide MsrA2, modified to include the SBP N- and C-terminal propeptides, in soybean seed coats. The gene encoding the modified MsrA2 was designated *MsrA2vac* and was codon optimized prior to being synthesized by GenScript Corp in order to maximize expression of the gene in soybean (Horvath et al. 2000; Batard et al. 2000). This gene encodes the antimicrobial peptide MsrA2 flanked by the sequence encoding

the N- and C-terminal propeptides from SBP. An *Nco*I site was included at the start codon and a *Bst*EII site was included 3' to the stop codon for cloning purposes.

Unlike the *phb* construct, the *MsrA2vac* gene was not designed to include linkers between the SBP N- and C-terminal propeptides and MsrA2, nor were the four additional codons from the *Ep* gene that follow the N-terminal propeptide or precede the C-terminal propeptide included. The function of AMPs is highly dependent upon their amino acid sequence and small changes can drastically affect their activity (Zasloff 2002). Therefore, it was decided to attempt to modify the mature sequence of the AMPs as minimally as possible by omitting these additional amino acids.

The *MsrA2vac* gene was digested with *Nco*I and *Bst*EII and ligated to the plasmid 1302-tCZ-Prx, which was also digested with *Nco*I and *Bst*EII, generating plasmid J12 (Figure AI.1A). This plasmid is derived from the pCAMBIA1302 plasmid, in which the CaMV 35S promoter driving expression of the *hptII* gene has been replaced by the enhanced cryptic tobacco promoter *entCUP4* (Malik et al. 2002). In addition, the CaMV 35S promoter driving expression of the gene encoding GFP has been replaced by the *Ep*1.5 promoter (synthesized by the Plant Biotechnology Institute). Inserted upstream of the *Ep* 1.5 promoter is a 2.5 kb fragment of the *LacZ* gene, which provides spatial separation between the *entCUP4* and *Ep* 1.5 promoters to minimize promoter interaction. The final construct was sequenced to confirm that all elements were properly cloned.

PCR analysis of transgenic plants

Soybean plants were screened by PCR as described in Section 2.5. The *MsrA2vac* gene was specifically amplified using the AMPSigF and AMPSigR primers and the *Ep* 1.5 promoter was specifically amplified using the Ep-MsrA2vac-F and MsrA2Int-R primers. The Ep-MsrA2vac-F primer overlaps both plasmid sequence and the *Ep* 1.5 promoter and the MsrA2Int-R primer is located in the *MsrA2vac* gene in order that amplification of the *Ep* promoter is specific for the

transgene and does not detect the endogenous promoter driving expression of the *ep* gene. A list of the primers used and their sequences can be found in Table 2.1 and the PCR conditions used for each primer pair can be found in Table 2.2.

Antibodies

A peptide corresponding to amino acids 1 to 12 of the MsrA2 peptide was synthesized, conjugated to keyhole limpet hemocyanin (KLH), and injected into two rabbits to generate anti-AMP-1 and anti-AMP-2 antibodies (SACRI Antibody Services, Calgary, AB). The more active anti-AMP-1 antibody was used for immunobloting at a dilution of 1:500.

Immunoblot Analysis

SDS-PAGE was performed for the separation of proteins in the range of 20 to 200 kDa, as described in section 2.11. The MsrA2 peptide is approximately 3.3 kDa and is not fully resolved using SDS-PAGE. For improved resolution of small peptides, tricine SDS-PAGE was used to separate soybean seed coat extracts, according to the method of Schägger (2006). Samples were diluted 1:1 with tricine sample buffer (200 mM Tris-HCl, pH 6.8, 2% SDS, 40% glycerol, 0.04% Coomassie Brilliant Blue G-250, 2% β-mercaptoethanol) and boiled for 5 min prior to loading on the gel. Typically, anywhere from 20 to 50 μg of protein was loaded in a single well on a 16.5% gel. Where necessary, samples were first precipitated with trichloroacetic acid (TCA) as described in Section 2.11. Samples were separated by electrophoresis at 200 V for approximately 1 h using the Mini-Protean 3 system (Bio-Rad).

Immunoblotting was performed as described in Section 2.11, except that protein was transferred to PVDF for only 50 min. Three positive controls were used: the MsrA2 fragment used to inject the rabbits for antibody production, the same fragment conjugated to BSA, or the entire MsrA2 peptide (Brain Research Centre, University of British Columbia, Vancouver, BC).

Antimicrobial Assays

Two separate methods were used to measure the antimicrobial activity of soybean seed coat extracts. The first method was a bacterial growth inhibition assay, based on the method of Osusky et al. (2005) with minor modifications. In this method, 100 mg of *Arabidopsis* leaf tissue was ground to a fine powder in liquid nitrogen in a microcentrifuge tube with a pestle. The leaf tissue was centrifuged at 16,100 g (13,200 rpm) for 10 min at 4°C. The supernatant was collected and 50 µl was mixed with 100 µl of 1xPBS (137 mM NaCl, 2.7 mM KCl, 10.1 mM Na_2HPO_4, 1.8 mM KH_2PO_4) containing approximately 2 x 10^5 *E. coli* DH5α. incubated at room temperature for 2 h and plated on LB agar. After growth overnight at 37°C, percent survival was calculated from the colony counts. The second method was an inhibition zone assay (Mor et al., 1991 with minor modifications). In this method, 100 µl of an *E. coli* DH5α culture diluted to an OD_{550} of 0.1 was spread on a plate containing LB media. Plant extracts were prepared as described for the bacterial growth inhibition assay and 10 µl of the supernatant was placed on a sterile filter disk, which was then allowed to dry. The disks were placed on the plate immediately after spreading the bacteria. The plate was grown overnight at 37°C and the diameter of inhibition zones surrounding the filter disks were measured.

RESULTS

Characterization of anti-MsrA2 antibody used for protein immunoblotting

A polyclonal antibody prepared from rabbits injected with the first twelve amino acids of the MsrA2 peptide conjugated to KLH was made by SACRI Antibody Services (Calgary, AB). The antibody was able to detect only as low as 0.5 µg of the fragment of the MsrA2 peptide used to generate the antibody conjugated to BSA following SDS-PAGE (Figure AI.1B). It was also able to

detect as low as 50 ng of the same fragment of the MsrA2 peptide alone following tricine SDS-PAGE (Figure AI.1C). The antibody detected a few faint bands in wild-type soybean cv. Jack seed coat extracts if blots were developed for an extended period of time (data not shown). This antibody is therefore sensitive for the MsrA2 peptide down to 50 ng with low cross-reactivity.

Cloning and Plant Transformation

The *MsrA2vac* gene was cloned into 1302-tCZ-Prx (Figure AI.1A) for transformation of soybean cv. Jack by particle bombardment. Three T_0 plants were regenerated, all from a single transformation event, although only two survived to maturity. From these two plants, eight T_1 seeds were germinated and after 4 to 6 days seed coats were collected for analysis and seedlings were transplanted to soil. Seed coats are maternal in origin, therefore the seed coats from T_1 seeds are from the T_0 generation and seed coats from the T_2 seeds are from the T_1 generation (Figure 2.1).

Confirmation of transgene insertion

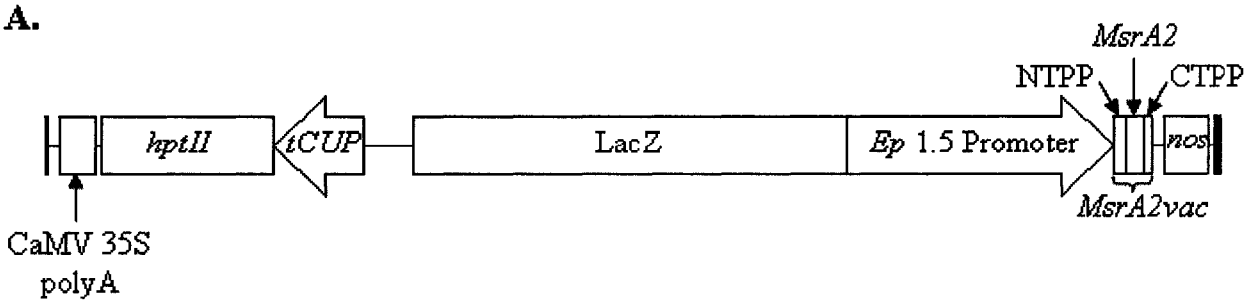
T_1 soybean plants were screened by PCR to investigate the completeness of the J12 expression cassette (Fig AI.1D), using the *β-tubulin-2* gene to test the quality of the DNA, J12 plasmid DNA as a positive control, and DNA from cv. Jack as a negative control. All transgene elements were successfully amplified from plasmid DNA; however, both the *MsrA2vac* gene and the *nos* transcriptional terminator could not be amplified from any of the eight T_1 plants, nor could the *Ep* 1.5 promoter (data not shown). In addition, the *hptII* and *entCUP4* promoter were also absent in at least two of the eight plants. A second DNA prepared from two of the T_1 plants confirmed these results. This suggests either that the transgene was unstable and therefore lost or that the transgene was lost in these plants through segregation and more plants need to be screened.

Antimicrobial Assays

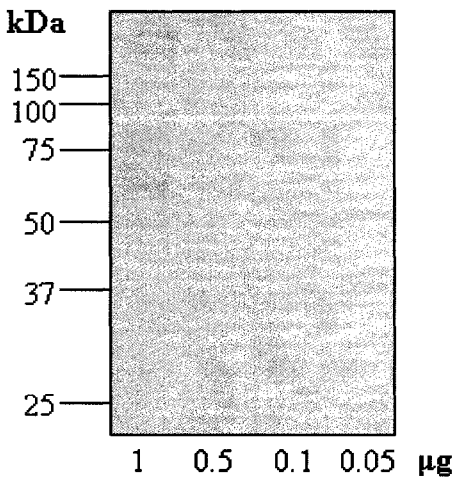
Two different methods were developed to measure antimicrobial activity - a bacterial growth

Figure AI.1 Experimental design and analysis of the soybean plants transformed with *MsrA2vac*. **A.** Schematic diagram illustrating the T-DNA region of the J12 construct containing the gene *MsrA2vac*, which encodes MsrA2 modified to include the SBP N- and C-terminal propeptides (NTPP and CTPP, respectively) for soybean transformation. This construct contains the *entCUP4* promoter (*tCUP*) driving expression of the *hptII* selectable marker gene. The *Ep* 1.5 promoter was used to drive expression of the *MsrA2vac* gene. Following the *MsrA2vac* gene is the *nopaline synthase* transcriptional terminator (*nos*). Black bars on left and right sides of the constructs represent left and right T-DNA borders, respectively. **B.** and **C.** Immunoblot analysis demonstrating the specificity and sensitivity of anti-AMP-1 antibody, as described in Sections 2.11 and this Appendix. A colorimetric assay was used to develop the immunoblots using the reagents NBT and BCIP. **B.** Reaction of anti-AMP-1 antibody (diluted at 1:500) to different amounts of MsrA2 fragment used to inject rabbits to generate the anti-AMP-1 antibody conjugated to BSA (10% acrylamide gel). **C.** Reaction of anti-AMP-1 antibody (diluted at 1:500) to different amounts of the MsrA2 fragment (16.5% tricine SDS-PAGE gel). **D.** Amplification of transgene elements from T_1 soybean genomic DNA extracts. The *MsrA2vac* gene and the *nos* were amplified separately by PCR. As a control for genomic DNA quality the endogenous *β-tubulin-2* gene was amplified. Positive control (+) is J12 plasmid DNA. As a negative control, genomic DNA from wild-type cv. Jack was used as template. Neither *MsrA2vac* nor *nos* were amplified from any of the T_1 genomic DNA samples. The primers used are listed in Table 2.1 and conditions for amplification of each element in Table 2.2.

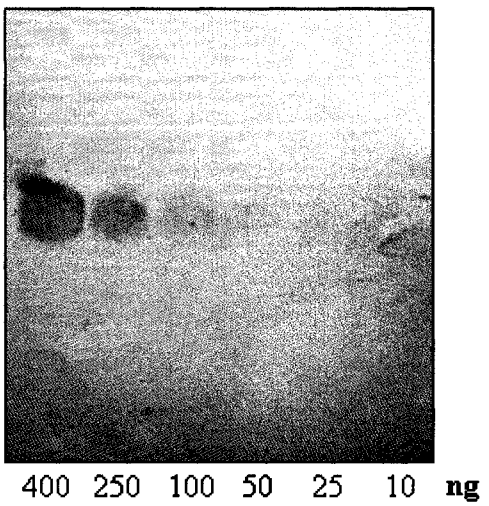
A.



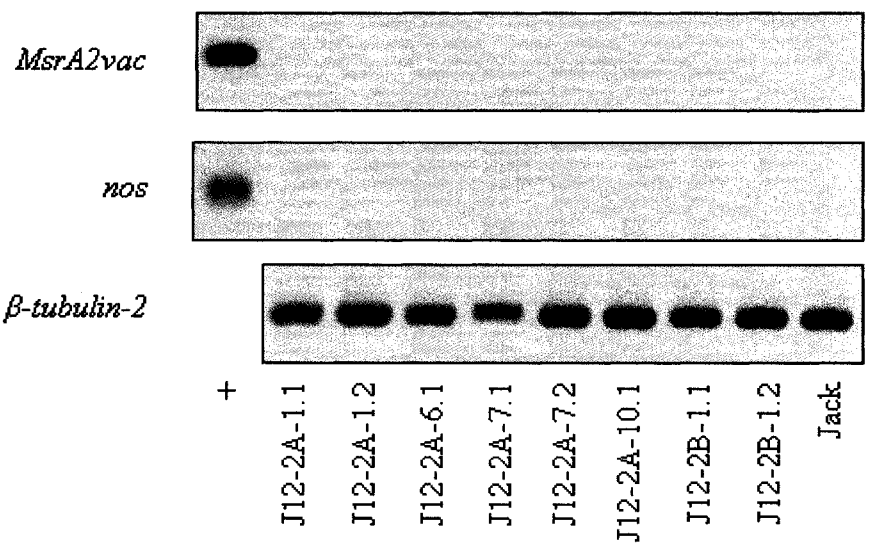
B.



C.



D.



inhibition assay and an inhibition zone assay. The synthetic MsrA2 peptide used as a positive control did not display AMP activity with either assay, at concentrations up to 1 μg/μl. The MsrA2 peptide has been demonstrated to completely inhibit growth of *E. coli* at 75 μg/ml (Osusky et al. 2005). This suggests that the MsrA2 peptide we had synthesized is not functional in our hands.

Immunoblot analysis of soybean seed coat extracts

Tricine SDS-PAGE was developed to detect MsrA2 in transgenic plants. This method was able to resolve the MsrA2 peptide used to generate the anti-AMP-1 antibody, which consists of only 12 amino acids. However, the full length synthetic MsrA2 peptide could not be detected by this method, even when up to 10 μg of the pure peptide was loaded.

DISCUSSION

Due to the absence of the J12 expression cassette in T_1 soybean plants, the analysis of transgenic soybean plants containing the *MsrA2vac* gene by immunoblotting was not further pursued.

We encountered several limitations during our attempt to express *MsrA2vac* in soybean. Legume transformation is a lengthy and difficult process that does not reliably return an abundance of unique transgenic plants (Trick et al. 1997; Somers et al. 2003). In the case of the *MsrA2vac* construct, only three plants were generated, all from a single transformation event, and one did not survive to maturity. Therefore, our analysis was restricted to a single clonal plant line. Eight T_1 plants were grown to provide more tissue for analysis but PCR screening failed to identify any elements from the J12 expression cassette in these plants. It is possible that the eight T_1 plants lost the transgene as a result of segregation. Assuming a single insertion locus, 25% of T_1 seeds would lose the expression cassette due to segregation, and this percentage would decrease with an increase

in the number of insertions. Another potential explanation is that the expression cassette was unstable, leading to its deletion in the T_1 generation. Transgene deletion has been observed in soybean plants (Choffnes et al. 2001) as well as other species (Vain et al. 2002; Walters et al. 1992; Register et al. 2004; Bregitzer and Tonks 2003; Cho et al. 1999). The best solution for transgenes that demonstrate such instability is to screen a large number of transformants in order to identify those in which the transgene are stably inherited. This remains a challenge that will need to be overcome.

The second major limitation was the lack of a reliable positive control in the development of immunoblotting and antimicrobial assays. We had the MsrA2 peptide synthesized but were unable to either detect it with the anti-AMP-1 antibody or measure its activity. The size and sequence of MsrA2 makes its synthesis difficult, although it is possible (Osusky et al. 2005). These remain as technical barriers to the analysis of AMP production in soybean.

Appendix II

INTRODUCTION

Tissue-specific expression of foreign genes is often a desirable goal in the development of value-added plant systems. Limiting expression to a specific tissue can help to minimize toxic effects that can be associated with the overexpression of transgenes. In some cases, tissue-specific promoters can produce higher expression levels than the commonly used constitutive promoters, such as the CaMV 35S promoter (Hood et al. 2003; Yang et al. 2005). In addition, this approach may be more likely to meet public approval (Potenza et al. 2004).

SBP is most abundant in seed coat tissues although a small amount is produced in root tissues as well (Gijzen et al. 1993). We chose to use the *Ep* 1.5 promoter to drive expression of foreign genes to achieve seed coat specific expression. It was our intention to restrict expression to this tissue in order to minimize any negative impact on plant health. In addition, it maintains the quality of the seed, which is the major commodity produced by soybean plants, while adding value to a tissue that is essentially a waste from the processing of seeds.

As part of the development of the soybean seed coat system, it is important to determine that the promoter elements that direct seed coat-specific elements were included in the *Ep* 1.5 promoter. Previous work has indicated that the *Ep* 1.5 promoter does regulate seed coat-specific expression of GUS in transgenic plants (D. Simmonds; L. Gudynaitę-Savitch unpublished). However, seed coat-specific expression will need to be additionally demonstrated in the soybean seed coat system.

When using tissue-specific promoters, it is important to limit the influence of nearby regulatory elements, originating either from within the T-DNA or in genomic DNA located near to the site of insertion (Jagannath et al. 2001; Yoo et al. 2005; Rubinchik et al. 2001; Campisi et al.

1999; Springer 2000). As discussed in Section 4.3, several steps were taken when designing constructs for soybean transformation in order to minimize these influences.

Despite these manipulations, the potential still exists for these influences to occur. It is therefore important to screen high expressing lines to confirm that expression is limited to the seed coat. Most of the transgenic soybean plants generated for this thesis gave relatively low levels of expression. We selected one of the T_1 plants containing the *EpW* construct to test the applicability of analyzing the pattern of expression of the *Ep* 1.5 promoter to determine that it is seed coat-specific; namely, J-EpW-B-4, the same plant for which we analyzed the level of seed coat expression of the *Ep* gene (Section 4.6). However, under the experimental conditions described, there was no *Ep* transcript detected in seed coats at 21 DPA (Figure 4.6). Since transcript was not detected in the seed coat, it is perhaps unlikely that it will be detected in other plant tissues, which suggests that this plant was not a good candidate for examining expression in other tissues. However, the results of this work will be presented here in order to serve as an example of how analysis of expression patterns can be done.

MATERIALS AND METHODS

For a description of the general materials and methods used, refer to Chapter 2.

RESULTS

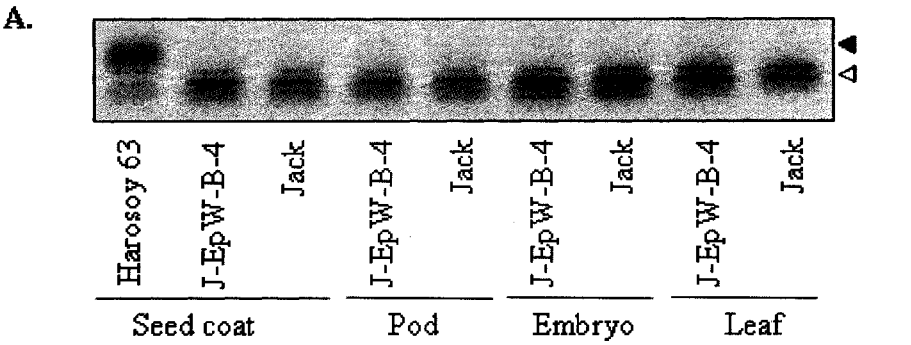
In order to examine the pattern of expression of the *Ep* gene in transgenic plants, we attempted to detect both *Ep* transcript and SBP in 2 week-old leaf, as well as 21 DPA pod and embryo tissues. This stage of pod and embryo tissue was selected for the reasons discussed in section 4.6. Two-week-old leaf tissue was chosen because at this stage leaves were still young

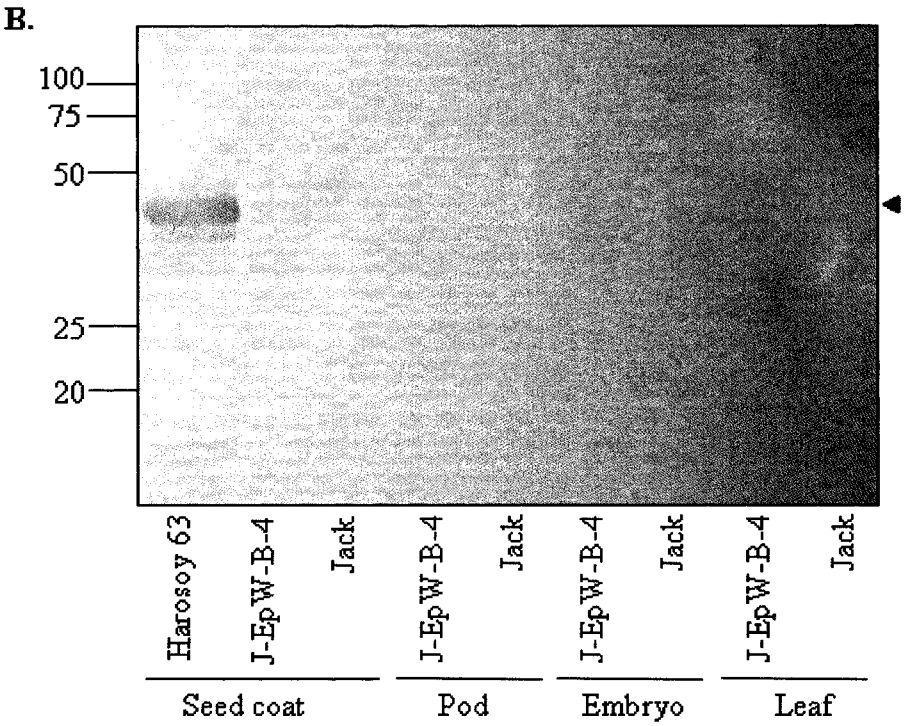
enough to obtain good RNA and protein extracts, while being old enough to have accumulated some SBP if expressed in this tissue. While pod and leaf tissues are from the T_1 generation, it is important to note that embryo tissue is primarily of the T_2 generation (see Section 2.4 and Figure 2.1). It is possible that some of the embryos have lost the transgene through segregation or instability, so the results for the embryo tissues may underestimate the accumulation of *Ep* transcript and SBP in this tissue.

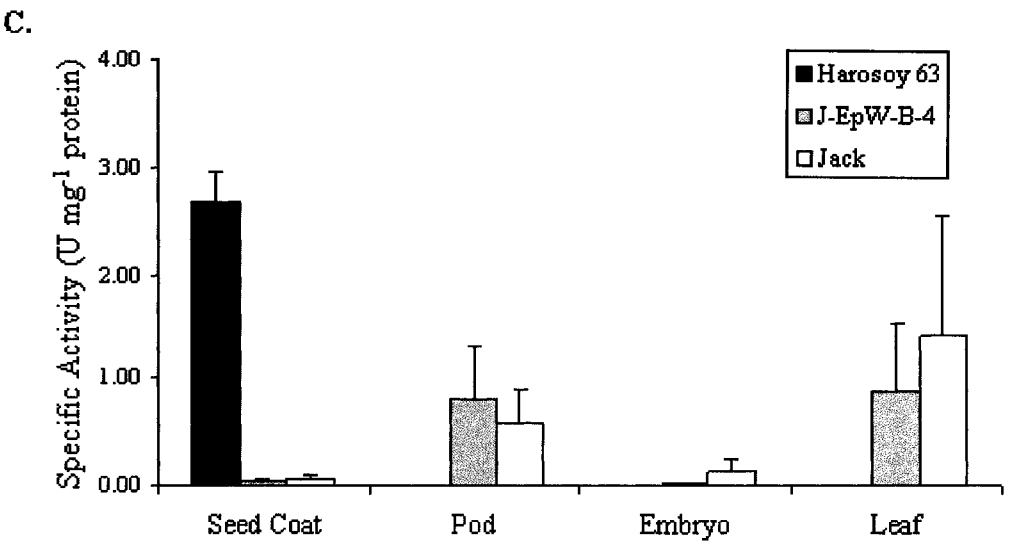
RNA was isolated from 21 DPA pod, embryo, and 2-week-old leaf tissue from the T_1 plant J-EpW-B-4 and wild-type cv. Jack as a negative control. The plant J-EpW-B-4 was selected for the reasons stated in section 4.6. For each plant, three separate extracts from each tissue were prepared and analyzed. Seed coat RNA samples (Section 4.6) were included for comparison. *Ep* transcript and the internal control *β-actin* were reverse transcribed using gene-specific primers, as described in Section 4.6, and then amplified by PCR using the same primers. As a control, PCR amplification was always performed simultaneously with a sample that did not undergo reverse transcription, in order to detect the presence of genomic DNA contamination. After 35 cycles, *Ep* transcript was only amplified in Harosoy 63 seed coat samples (Figure AII.1A). In one replicate, a band the size of the *Ep* amplicon was detected in the J-EpW-B-4 leaf extract; however, a band was similarly detected in the Jack leaf extract, suggesting that this was contamination (data not shown). No bands the size of the *Ep* transcript were detected in the other two leaf extracts for both J-EpW-B-4 and Jack. Therefore, *Ep* transcript was not detected in 21 DPA pod and embryo tissues or in 2-week-old leaf tissues in the transgenic plant J-EpW2-B-4.

Protein extracts were also prepared from 21 DPA pod and embryo and 2-week-old leaf tissues from J-EpW-B-4 as well as cv. Jack plants, as a negative control, and the accumulation of SBP was determined by immunoblotting with anti-SBP antibody and by peroxidase activity assay.

Figure AII.1 *Ep* expression and SBP accumulation in wild-type cv. Harosoy 63, J-EpW-B-4, and wild-type cv. Jack 21 DPA seed coats, pod, and embryo, and 2-week-old leaf tissues. **A.** RT-PCR amplification (35 cycles) of *Ep* transcripts. Closed arrowhead indicates expected size of *Ep* amplicon and open arrowhead indicates expected size of β-*actin* amplicon. **B.** Immunoblot analysis (12% acrylamide gel) of SBP with anti-SBP antibody (diluted at 1:500), as described in Section 2.11. For each sample, 50 μg of protein was loaded. A colorimetric assay was used to develop the immunoblots using the reagents NBT and BCIP. Closed arrowhead indicates the expected size of SBP. **C.** Mean peroxidase specific activity + SE (n=3) of soybean plant tissues.







For each plant, three separate extracts were prepared and analyzed. As for RNA analysis, extracts from seed coats described in Section 4.6 were included for comparison. Anti-SBP antibody detected a strong band in extract from cv. Harosoy 63 seed coats (Figure AII.1B) but did not detect any bands in the leaf and embryo extracts from the J-EpW-B-4 transgenic plants. In one of the three pod extracts, a very faint band of the correct size was visible, but this could not be seen in the other two extracts. Based on activity assays, there was no significant difference in mean peroxidase specific activity between J-EpW2-B-4 and Jack seed coat [t(4)=-0.3118, p=0.7707], pod [t(4)=0.3812, p=0.7224], embryo [t(4)=-1.0906, p=0.3368], and leaf [t(4)=-0.4138, p=0.7002] tissues (Figure AII.1C). The activity detected in these samples likely originates from endogenous peroxidases produced in these tissues (Gijzen et al. 1993; Graham and Graham 1991; Cakmak and Horst 1991; Felton et al. 1994; Caldwell et al. 1998). Their presence will likely render the use of activity assays to detect SBP in these tissues ineffective, unless SBP is produced at very high levels. The one exception to this is the seed coat, where endogenous peroxidase activity is low since we chose to use *epep* cultivars and the amount of SBP produced should be relatively high. Since the anti-SBP antibody does not appear to cross-react with these endogenous peroxidases, immunoblotting will likely be the most effective means of detecting SBP at the protein level in these tissues.

Based on these results, the *Ep* 1.5 promoter is not being expressed in leaf and embryo tissues in the transgenic plant J-EpW-B-4. In pod tissues, there may be a small level of expression, but as this was detectable in only a single protein extract by immunoblotting and could not be detected by either activity assay or RT-PCR, the expression is relatively minor. These results suggest that the seed coat-specific pattern of expression of the *Ep* 1.5 promoter was maintained in this plant. Alternatively, it is also possible that *Ep* expression in these tissues is undetectable due to transgene

silencing. These results were obtained from a single transgenic plant and will need to be repeated in several other plants from this line as well as from other high-expressing lines.

DISCUSSION

A T_1 plant from a high-expressing line containing the *EpW* gene was screened for *Ep* expression in leaf, pod, and embryo tissues, both at the mRNA and protein levels. Based on RT-PCR, *Ep* transcript was only detected in one of the three leaf RNA extracts but it was similarly detected in a negative control sample consisting of RNA extracted from leaves of a wild-type *epep* soybean plant (cv. Jack). This suggests the RT-PCR was detecting contamination in these samples as opposed to *Ep* transcript. Based on peroxidase activity assays, there was no significant difference in the levels of peroxidase activity in leaf, pod, and embryo tissue from the transgenic plant compared to the wild-type *epep* cultivar. Activity assays revealed some endogenous peroxidase activity, particularly in pod and leaf tissues, that could mask the presence of low levels of SBP. Based on immunoblotting, however, SBP was only ever detectable in one of the three pod protein extracts, and this was observable as a very faint band in 50 μg of total protein. This would suggest that there could be a small amount of expression in non-target tissues, but only at a minor level. This analysis should be completed on several other plants from the same line as well as from a few other independent lines to gain a more thorough analysis of the potential expression of the *Ep* 1.5 promoter in non-target tissues.

Determining the pattern of expression of foreign genes in soybean plants will be necessary in order to ensure that seed coat-specific expression is maintained. However, this type of analysis should follow identification of high-expressing lines, which are more realistic candidates for commercial production.