Testing the Cruciferin Deficient Mutant, ssp-1, of Arabidopsis thaliana, as a Vehicle for Overexpression of Foreign Proteins

by

Yimei Lin

A thesis submitted in conformity with the requirements for the degree of Master of Science
Cell and System Biology
University of Toronto

© Copyright by Yimei Lin 2011
Testing the Cruciferin Deficient Mutant, *ssp-1*, of *Arabidopsis thaliana*, as a Vehicle for Overexpression of Foreign Proteins

Yimei Lin

Master of Science

Cell and System Biology
University of Toronto

2011

Abstract

*ssp-1* is a seed storage protein mutant which is deficient in one of the major seed storage proteins in *Arabidopsis thaliana*, the 12S cruciferins. To determine if this mutant can drive a higher level expression of a transgene than that found in wild type, the mutant was transformed with the *phytohemagglutinin (PHA)* gene and single copy PHA homozygotes were identified. These PHA transformants were crossed to wild type so that each PHA gene would be in the same copy number and chromosomal context in a wild type background. Immunoblotting was employed to compare the PHA levels of the single copy transformants in both genetic backgrounds. PHA levels ranged from 4.52% to 7.7% of the total protein in transformants. Two of the transformants showed 30.33% and 44.18% more PHA than that of their backcross. Therefore, a mutant such as *ssp-1* may provide a means for overexpression of foreign proteins.
Acknowledgments

I would like to show my gratitude to my supervisor, Dr. Daniel Riggs for his encouragement, guidance and support from the initial to the final level in my project. He was always so patient and willing to help his students with their research. As a result, my research life became smooth and rewarding for me. I also would like to thank my committee members, Dr. Clare Hasenhampf and Dr. Rongmin Zhao for their helpful direction and suggestions. In addition, I need to thank Dr. Sonia Gazzarini’s lab and Dr. Greg C. Vanlerberghe’s lab for their technical support. Recognition should be given to Dr. Peter McCourt and my supervisor for their generation of ssp-1, to Jane Sung Yun Lee for her initial studies. To all of those who supported me in any respect during the completion of my project, especially Jia (Steven) Wang, Allen Tsai, Esther Lesmana, and Patti Stronghill, I am grateful for your friendship and help. I would also like to acknowledge my parents and my brother for their unflagging love and support throughout my life. Without all of these help I could never have accomplished this dissertation.
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>iii</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>iv</td>
</tr>
<tr>
<td>List of Tables</td>
<td>vii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>viii</td>
</tr>
<tr>
<td><strong>1 Introduction</strong></td>
<td></td>
</tr>
<tr>
<td>1.1 Seed development</td>
<td>1</td>
</tr>
<tr>
<td>1.2 The importance of seeds</td>
<td>3</td>
</tr>
<tr>
<td>1.3 Production of crops with improved seed storage proteins</td>
<td>5</td>
</tr>
<tr>
<td>1.3.1 Seed storage protein</td>
<td>6</td>
</tr>
<tr>
<td>1.3.2 Enhancement of SSPs in plants</td>
<td>7</td>
</tr>
<tr>
<td>1.3.2.1 Enhancing protein-bound essential amino acids in plants</td>
<td>7</td>
</tr>
<tr>
<td>1.3.2.2 Enhancing free essential amino acids in plants</td>
<td>11</td>
</tr>
<tr>
<td>1.4 Production of ‘value added’ crops</td>
<td>13</td>
</tr>
<tr>
<td>1.4.1 Production of biomedical reagents in plants</td>
<td>14</td>
</tr>
<tr>
<td>1.4.2 Production of biodegradable plastics in plants</td>
<td>18</td>
</tr>
<tr>
<td>1.4.3 Production of nutritional fatty acids-added plants</td>
<td>18</td>
</tr>
<tr>
<td>1.5 ‘Empty container’ hypothesis</td>
<td>19</td>
</tr>
<tr>
<td>1.6 ssp-1</td>
<td>22</td>
</tr>
<tr>
<td>1.7 The objective of my thesis</td>
<td>23</td>
</tr>
<tr>
<td><strong>2 Materials and Methods</strong></td>
<td>24</td>
</tr>
<tr>
<td>2.1 Seed lots</td>
<td>24</td>
</tr>
<tr>
<td>2.2 Single pot plantings of <em>Arabidopsis thaliana</em></td>
<td>24</td>
</tr>
</tbody>
</table>
Table of Contents (continued)

2.3 Plasmids used for transformation 24
2.4 Production of DNA constructs 24
2.5 Transformation of *E. coli* 25
2.6 Transformation of *Agrobacterium tumefaciens* 26
2.7 *Arabidopsis* transformation 26
2.8 Selection of transformed plants 27
2.9 Isolation of plant DNA 27
2.10 PCR of *PHA* transgenics 29
2.11 Isolation of seed proteins 29
2.12 Quantification of seed proteins 29
2.13 SDS-PAGE 30
2.14 Immunoblotting 30
2.15 Southern blot analysis of *PHA* transgenics 31
2.16 Genetic cross of *PHA* homozygote in *ssp-1* and *Ler* 32
2.17 Sequencing preparation of DNA 33
2.18 Evaluation of *PHA* levels in transgenic plants 33

3 Results 34

3.1 Preface 34
3.2 The ‘empty container’ hypothesis 34
3.3 Strategy for testing ‘empty container’ hypothesis 36
3.4 Plasmids used for transformation 36
3.5 Generation of *ssp-1* transformants 39
3.6 Selection of single copy *PHA* homozygotes in the *ssp-1* genetic background 39
   3.6.1 Selection of single copy of *PHA* transformants in *ssp-1* 39
   3.6.2 Confirmation of *PHA* expression in single locus transformants 42
Table of Contents (continued)  

3.6.3 Screening for PHA homozygotes in ssp-1  

3.7 Crossing of PHA homozygotes into the Ler genetic background  

3.8 Comparison of PHA levels in ssp-1 verses Ler genetic backgrounds  

4 Discussion  

5 Summary  

References
List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Selected examples of expression of genes encoding seed storage protein</td>
<td>8</td>
</tr>
<tr>
<td>2. Summary of southern blot analysis of 4 single copy PHA transformants</td>
<td>46</td>
</tr>
<tr>
<td>3. Chi-square test of T3 seeds from heterozygous single copy PHA transgenic T2 plants on hygromycin selection medium.</td>
<td>50</td>
</tr>
<tr>
<td>4. Summary of PHA homozygotes in the ssp-1 background identified from segregation analysis of T3 seeds</td>
<td>50</td>
</tr>
<tr>
<td>5. Summary of PHA homozygote in Ler background</td>
<td>54</td>
</tr>
<tr>
<td>6. Summary of PHA homozygotes in homozygote Ler genetic background</td>
<td>58</td>
</tr>
<tr>
<td>7. PHA expression level in both ssp-1 and Ler genetic background background</td>
<td>60</td>
</tr>
</tbody>
</table>
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Embryogenesis of the dicot <em>Arabidopsis thaliana</em></td>
<td>2</td>
</tr>
<tr>
<td>2. The structure of a mature dicot common bean seed</td>
<td>2</td>
</tr>
<tr>
<td>3. Dicot seed (Garden bean) germination and seedling growth</td>
<td>4</td>
</tr>
<tr>
<td>4. Polypeptide pattern of seed storage proteins from <em>Ler</em> and <em>ssp-1</em></td>
<td>35</td>
</tr>
<tr>
<td>5. <em>Arabidopsis</em> seeds do not contain endogenous PHA</td>
<td>35</td>
</tr>
<tr>
<td>6. Schematic representation of two DNA constructs used for the <em>Agrobacterium</em>-mediated transformation of <em>ssp-1</em></td>
<td>37</td>
</tr>
<tr>
<td>7. Cloning of LC686</td>
<td>38</td>
</tr>
<tr>
<td>8. PCR analysis for PHA transgene in selected putative transgenic plants T₁</td>
<td>40</td>
</tr>
<tr>
<td>9. Schematic representation of the 2.8 kb PHA EcoR I/Hind III fragment.</td>
<td>40</td>
</tr>
<tr>
<td>10. Selected southern blot analysis of T₂ PHA transformants with either basta resistance or hygromycin resistance</td>
<td>43</td>
</tr>
<tr>
<td>11. Southern blot analysis of T₂ PHA transformants</td>
<td>45</td>
</tr>
<tr>
<td>12. Confirmation of PHA expression in PHA transformants</td>
<td>46</td>
</tr>
<tr>
<td>13. Strategy for screening PHA homozygote in both <em>ssp-1</em> and Ler genetic background</td>
<td>48</td>
</tr>
<tr>
<td>14. PCR analysis of PHA gene insertion in PHA backcross F₁ lines</td>
<td>52</td>
</tr>
<tr>
<td>15. SDS-PAGE of seed storage proteins from PHA homozygote backcrosses</td>
<td>53</td>
</tr>
<tr>
<td>16. Schematic representation of PCR of CRU3 fragment from 4318 bp to 5154 bp and partial amino acid sequence in the chromosome 4 of wild type CRU3 gene</td>
<td>57</td>
</tr>
<tr>
<td>17. PCR amplification of CRU3 fragments containing the mutation site</td>
<td>57</td>
</tr>
<tr>
<td>18. Sequencing chromatograms of 9 CRU3 homozygotes and one CRU3 heterozygote</td>
<td>58</td>
</tr>
<tr>
<td>19. Comparison of PHA expression in single copy of PHA transformants and their relative backcross</td>
<td>59</td>
</tr>
<tr>
<td>20. Comparison of PHA expression level in both <em>ssp-1</em> and Ler genetic background</td>
<td>60</td>
</tr>
</tbody>
</table>
1 Introduction

During the life cycle of plants, seed production is an essential process for plant reproduction. Before fertilization meiosis occurs within the anthers and ovaries to produce male and female gametes. In angiosperms, during pollination the fusion of male and female gametes in a double fertilization event produces a single-cell diploid zygote along with the triploid endosperm. The endosperm provides nutrients for embryogenesis, and is maintained in the seed as a storage tissue or is absorbed by cotyledons during seed development. During embryogenesis the zygote undergoes a series of morphological and cellular changes which result in the mature embryo, consisting of an embryonic axis with shoot and root poles required for seedling growth and development, and cotyledons, which contain high levels of storage nutrients such as proteins, carbohydrates and lipids (West and Harada, 1993; Möller and Weijers, 2009). The development of embryo is complete by the end of seed maturation.

1.1 Seed development

Seed development consists of three stages: the zygote development stage, the mid-maturation stage and the desiccation stage, where seeds are preparing for dormancy (Meinke, 1995). The first phase is involved in morphogenesis. After fertilization, the single-celled zygote goes through a series of coordinated divisions with cytologically recognizable substages, to form the mature differentiated embryo (Figure 1) (Möller and Weijers, 2009; Jürgens, 1995). In the globular stage, the embryo differentiates the embryonic axis and the cotyledons (West and Harada, 1993). At the torpedo stage, along the apical-basal axis, the upper tier of the proembryo gives rise to the shoot apical meristem and most of the cotyledons, while the lower tier of the proembryo produces the abaxial part of the cotyledons, the hypocotyl, and root apical meristem (Möller and Weijers, 2009). A mature embryo contains hypocotyl, epicotyl, cotyledons and radicle (Figure 2) (Brooker et al., 2010). The mature embryo along with the seed coat arising from the ovule gives rise to the mature seed. Therefore seed embryogenesis starts with a single-cell zygote and ends with a mature embryo.
**Figure 1. Embryogenesis of the dicot *Arabidopsis thaliana.*** Colors indicate corresponding regions in embryo. Following cell division and differentiation, the single-celled zygote undergoes 5 different stages of embryogenesis: 2-cell stage, 8-cell stage, globular stage, transition (heart shaped stage) and the torpedo stage. At the torpedo stage, the embryo gives rise to cotyledon (cot, yellow); shoot apical meristem (SAM, red); hypocotyl (hypo, blue), root apical meristem (RAM, green) (Modified from Möller and Weijers, 2009).

**Figure 2. The structure of a mature dicot common bean seed.** Within the seed, the mature embryo often contains epicotyl, hypocotyl, radicle, cotyledon, which are all covered by a seed coat (Brooker et al., 2010).
During the second stage, cell expansion of the globular shaped embryo takes place, which results in the enlargement of the embryo (West and Harada, 1993). Moreover, there is rapid synthesis and accumulation of seed storage reserve macromolecules, which are principally carbohydrates, proteins and lipids (Gallardo et al., 2008). In monocots, seed reserves are mainly accumulated in the endosperm; by contrast, in dicots, during this expansion stage, the cotyledons absorb storage nutrients from endosperm and become the major storage organ (West and Harada, 1993; Gallardo et al., 2008). These storage reserves are the nutrition source for the seedling growth and development.

During the last stage of seed development the embryo undergoes developmental arrest and the desiccation stage, which prepares the seed for dormancy and germination. When seeds undergo dormancy, they lose more than 90% of their water content, and most metabolic processes terminate (Mandal and Mandal, 2000). For example, RNA and protein synthesis drops dramatically during this stage (Mandal and Mandal, 2000). Under favorable conditions, seed germination starts after dormancy and the majority of seed reserves are mobilized and provide nutrients for seed germination and seedling establishment (Figure 3) (Brooker et al., 2010). In dicots, like grain legumes, as the seedlings grow the nutrients in the cotyledons are used as the energy source and cotyledons become desiccated (Figure 3) (Brooker et al., 2010).

1.2 The importance of seeds

In plants, seed storage reserves are of key importance to determine the success of germination and early plant growth. Depending on the species, the seed contains various amounts of carbohydrates, proteins and lipids. For example, beans are a good source of protein. Canola seeds contain high level of lipids. Rice has high levels of carbohydrate. Seeds are the main source of food for humans and animals, and used as the raw materials for industries. According to FAO (The Food and Agriculture Organization of the United Nations) estimate, 70% of the food in the world for human consumption is from cereals and legumes and the remaining 30% comes from animals, which are fed by seed meals (Mandal and Mandal, 2000). It is well known that seed storage proteins play a crucial role in supplying essential amino acids (EAAs) and are the primary source of nutrients for humans and animals (Shewry and Halford, 2002). Seed oil can be used as both fuel and chemical feedstocks (Jaworski and Cahoon, 2003), or might contain nutritional or beneficial fatty acids for humans (Singh et al., 2005). For example, the hydroxylat-
Figure 3. Dicot seed (Garden bean) germination and seedling growth. The seedling pulls the embryonic shoot (epicotyl and hypocotyl) and cotyledons above the ground, leaving the seed coat in the soil, and has absorbed most of the nutrients from cotyledons which become withered. The foliage leaves are formed after seed germinated (Brooker et al., 2010).
-ed fatty acid ricinoleic acid can serve as a raw material for producing many useful products such as lubricants, emulsifiers, and inks to biodiesel formulation and nylon precursors (Dyer and Mullen, 2008). A range of cereals are used in the production of grain whisky and ethanol (Agu et al., 2008).

With the development of genetic engineering and gene manipulation many attempts have been made to improve crop or seed products. Desirable traits have been introduced into plants to increase the yield and quality of commercial products. Due to these alterations, crops can be resistant to insects, herbicides, and environmental stresses. Moreover, by using these techniques, the composition of plant or seed products can be altered, which is of great interest to agriculture and many industries as it provides a means to increase the value of commercial crops. People have tried to modify components of fatty acids in transgenic plants to produce more favorable seed oils that can be used for many specific industrial uses, such as cosmetics, plasticizers, lubricants, and as an alternative renewable source of energy to crude oil (Dyer and Mullen, 2008; Napier, 2007).

As discussed above, due to the importance of seeds, many transgenes have been introduced into crops in order to improve the nutritional value and quality of crops. Seed storage protein (SSP) quality is one of the factors that affect crop quality. Plants have been used widely for production of recombinant proteins. To improve nutritional quality of crops, seeds could be used as a vehicle for overproduction of economic foreign proteins. Aside from being a source of dietary protein, plant-based systems also have been used to produce pharmaceutically or industrially important products including vaccines, antigens, biodegradable plastics, and nutritional oils, etc. Likewise, seeds could be a robust industry to produce these foreign components. For all these potential reasons, high level expression of transgenes is required.

1.3 Production of crops with improved seed storage proteins

One goal of agricultural biotechnology is to improve the nutritional quality of crops. The composition of seed storage proteins is an important agronomic trait. In general seed protein is deficient in some EAAs, and therefore it has poor nutritional quality. Cereal proteins are in general low in lysine (Lys) (1.5-4.5% vs. 5.5% of the World Health Organization (WHO) recommendation) and tryptophan (Trp) (0.8-2.0% vs. 1.0%), and threonine (Thr) (2.7-3.9% vs. 4%) while legumes contain only 1.0-2.0% the sulfur-containing amino acids (methionine (Met)
and cysteine (Cys), compared with the 3.5% of the WHO reference protein) (Mandal and Mandal, 2000; Shewry and Halford, 2002; Sun and Liu, 2004). The nutritional quality of cereals is not only considered in human diets but it is also important in animal feed. Cereal grains contain relatively less protein than legume seeds. As the cereal grains are fed to livestock such as pigs and poultry, it’s usual to combine cereals with other protein sources such as legume seeds or synthetic amino acids.

1.3.1 Seed storage protein

Seed storage proteins (SSPs) are mainly used to provide the nitrogen and sulphur required during seed germination and seedling establishment (Li et al., 2007). According to their solubility, seed proteins are classified as albumins (water extractable, 1.6S-2S), globulins (extractable in dilute salt solutions, 7S-13S), prolamins (extractable in aqueous alcohol) and glutelins (most difficult to solubilize; extractable by weakly acidic or alkaline or dilute SDS solution) (Fukushima, 1991; Osborne, 1924). In dicots (eg. legumes), albumins and globulins are the main storage proteins whereas prolamins and glutelins comprise the major proteins of monocots (e.g. cereals) (Matta et al., 2009). SSPs are only synthesized in the seed either in cotyledon (dicots) or in endosperm (monocots) and lack any other functional activity besides storage (Mandal and Mandal, 2000).

During the second stage of seed development, SSPs are biosynthesized on membrane-bound polysomes as precursors containing a leader pre-peptide sequence and pro-sequence which have the signal for transport, processing and targeting from the site of synthesis to the storage organelles (Müntz, 1998). An N-terminal signal peptide of the nascent chain targets the polypeptide from the rough side of the endoplasmic reticulum (ER) into the lumen of the ER (Herman and Larkins, 1999). After post-translational cleavage of the signal peptide, the polypeptides are glycosylated and folded with the help of chaperones. In the ER, disulfide bridges are formed to stabilize the structure and several polypeptides are joined together to form an oligomer which has the proper conformation to be deposited in specialized membrane-bound storage organelles, called protein bodies (Mandal and Mandal, 2000). Deposition of SSPs in protein bodies protects proteins against uncontrolled premature degradation. When the seedling starts to emerge, hydrolytic enzymes, such as proteinases and peptidases, are produced and targeted to the protein bodies. As a result, SSPs will be mobilized and metabolized by degradation and provide nutrients for seedling growth during germination.
1.3.2 Enhancement of SSPs in plants

Intense research has attempted to improve seed protein nutritive value as human food and animal feed. Although conventional breeding programs have been used extensively, plant breeding to increase the nutritional value of crops has not been so successful for any given species due to lack of genes encoding seed storage protein with high levels of essential amino acids. By using biotechnology, modified or novel genes of interest can be transferred into plants by various methods such as electroporation or Agrobacterium-mediated transformation. As a result, the composition of seed storage proteins can be potentially modified. In addition, genetic engineering of crop plants is the only way to transfer specific, modified genes between widely related species (Mandal and Mandal, 2000). As illustrated in Table 1, many seed storage proteins and their expression levels have been studied in transgenic plants. In most cases, the protein is expressed, correctly processed, and accumulated in plants. However, in some cases, the level of introduced proteins is low, usually not more than a few percent (Table 1). This could be due to many factors, such as degradation of the foreign protein and promoters that function less efficiently in heterologous plants. Alternatively, the foreign protein is not accumulated correctly in the plants and degraded (Montoya, et al., 2010). There are two broad strategies to enhance the essential amino acids quality of crops including enhancement of protein-bound and free EAAs.

1.3.2.1 Enhancing protein-bound essential amino acids in plants

The EAAs can be improved in the protein form. One direct approach is modification of the coding region of a homologous seed protein gene, followed by reintroduction of the altered gene into the plant itself. For this approach, however, it’s critical to select suitable regions in the protein that can be modified without affecting its overall structure, stability, and function (Sun and Liu, 2004). Seed storage proteins assigned specific conformations and modifications may affect the structure and proper deposition of these proteins in the seed. Like other cereals, maize (Zea mays) is deficient in Lys in its major storage protein, zein. By nucleotide substitution and oligonucleotide insertion, Lys and Trp codons were created in different positions in a 19-kDa α-zein cDNA and this Lys-rich α-zein were produced in transgenic tobacco (Nicotiana tabacum) seeds like normal zein (Ohtani et al., 1991). However, the newly synthesized modified and unmodified zeins were both degraded in tobacco seeds due to protein instability.
<table>
<thead>
<tr>
<th>Protein/Gene</th>
<th>Source Plant</th>
<th>Host Plant</th>
<th>Promoter</th>
<th>Level&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sunflower albumin 8</td>
<td><em>Helianthus annuus</em></td>
<td><em>Festuca arundinacea</em> Schreb.</td>
<td>CaMV 35S</td>
<td>0.2%</td>
<td>Wang et al. (2001)</td>
</tr>
<tr>
<td>Brazil nut 2S albumin</td>
<td><em>Bertholletia excelsa</em></td>
<td><em>Vicia narbonensis</em></td>
<td>Legumin B</td>
<td>5%</td>
<td>Müntz et al. (1998)</td>
</tr>
<tr>
<td>Sunflower albumin 8</td>
<td><em>Helianthus annuus</em></td>
<td><em>Oryza sativa</em></td>
<td>Promoter of wheat high molecular weight glutenin gene</td>
<td>7%</td>
<td>Hagan et al. (2003)</td>
</tr>
<tr>
<td>11kDa γ-zein</td>
<td><em>Zea mays</em></td>
<td><em>Glycine max</em></td>
<td>β-conglycinin α’-promoter</td>
<td>0.5%</td>
<td>Kim and Krishnan (2004)</td>
</tr>
<tr>
<td>18-kDa Lys-rich protein</td>
<td><em>Psophocarpus tetragonolobus</em></td>
<td><em>Oryza sativa</em></td>
<td>Rice glutelin Gt1</td>
<td>1%</td>
<td>Sun et al. (2001)</td>
</tr>
<tr>
<td>18-kDa Lys-rich protein</td>
<td><em>Psophocarpus tetragonolobus</em></td>
<td><em>Oryza sativa</em></td>
<td>Modified rice glutelin Gt1</td>
<td>30%</td>
<td>Wenefrida et al. (2009)</td>
</tr>
<tr>
<td>Lys-rich protein sb401</td>
<td><em>Solanum tuberosum</em></td>
<td><em>Zea mays</em></td>
<td>Maize P19z</td>
<td>11.6% to 39%</td>
<td>Yu et al. (2004)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Protein product, as a percentage of total extractable protein in transgenic seeds.
Another zein protein, γ-zein, was modified to increase the Lys content in Arabidopsis thaliana plants. The modified Lys-rich γ-zeins were accumulated to high levels in protein bodies and co-localized with the endogenous zeins in maize endosperms (Torrent et al., 1997). But these mutated proteins were modified in the post translation process, leading to mis-sorting and secretion to the leaf cell wall of the transgenic Arabidopsis (Alvarez et al., 1998).

Attempts have been made to improve the protein quality of the common bean (Phaseolus vulgaris). Phaseolin is the main storage protein in common bean seeds despite the fact that it’s deficient in Met, Cys and Trp. A 45-bp sequence containing six Met codons was inserted into the β-phaseolin genomic clone, which was introduced into tobacco to increase the Met content of phaseolin (Sun and Liu, 2004). Although the modified gene can be transcribed, the modified phaseolin was degraded when it is accumulated in the protein bodies or in the Golgi vesicles which resulted in accumulation of only 0.2% of normal phaseolin in tobacco. There are also examples of increasing the amount of EAAs in other plants. For example, the sunflower albumin 8 protein-coding region was modified by addition of nucleotides to increase the sulphur-rich protein and the gene was then transformed into tall fescue (Festuca arundinacea Schreb.) (Wang et al., 2001). This study demonstrated that transgenic tall fescue can lead to the accumulation of Met-rich sunflower albumin at levels of up to 0.2% of the total soluble protein in individual plants. Soybean protein quality is limited by the low content of sulfur-containing amino acid. Glycinin is one of the predominant seed storage proteins in soybean, which contain sulfur amino acids. To improve nutritional value of soybean (Glycine max), a modified proglycinin gene with synthetic DNA encoding Met was introduced into soybean. The results showed higher accumulation of glycinin in transgenic soybean compared with non-transgenic plants (El-Shemy et al., 2007).

Another approach for improving the protein-bound EAAs in plants is to introduce new heterologous genes encoding a protein rich in desirable EAAs into the plant of interest. Many attempts have been made to increase the sulfur amino acids in crops. For example, Zein is rich in sulfur amino acids, especially Met. In an early study, using β-conglycinin α’-promoter, Kim and Krishnan (2004) were able to express the 11kDa γ-zein in soybean. Although the expression of 11kDa γ-zein elevated the Met content of the alcohol-soluble protein fraction 1.5-1.7 fold more than the non-transgenic line, the overall Met content of seed flour was not increased as the 11kDa γ-zein accounted for less than 0.5% of the total protein. Moreover, it also resulted in the
formation of two ER-derived protein bodies designated as either spherical or complex. Brazil nut 2S albumin is also highly rich in the sulfur amino acids (18% Met and 8% Cys) (Ampe et al., 1986). This gene has been widely expressed in Arabidopsis thaliana, tobacco, potato (Solanum tuberosum) and canola (Brassica napus), with certain enhancement of the methionine content in the target plants (Clercq, et al., 1990; Altenbach, et al., 1992; Tu, et al., 1998). Münzt et al. (1998) demonstrated that the Met-rich Brazil nut 2S albumin gene, under the regulation of the seed specific legumin B promoter from field bean (Vicia faba L.), could be highly expressed and the protein correctly processed in narbon bean (Vicia narbonensis). The 2S protein accumulated up to 5% of total seed protein in transgenic plants, resulting in two-fold increase in the Met content of seed proteins over the untransformed counterpart. Also, in transgenic common bean lines expressing this 2S albumin gene, the Met content increased from 10% to 23% as compared to the non-transgenic plants (Montoya, et al., 2010). However, during seed maturation, the albumin was either not stored correctly in the cotyledon tissue and degraded prematurely, or the 2S mRNA is less stable in beans than in Brazil nut in some transgenic bean lines. Moreover, the 2S albumin exhibits high resistance to proteolytic hydrolysis both in raw and heat-treated forms. Therefore, using the 2S albumin to increase Met content may not be useful in common bean. As the 2S Brazil nut albumin has been identified as potential allergen (Nordlee et al., 1996), its use may be limited. Sunflower seed albumin (SSA) is also rich in the sulfur amino acids. For instance, the seed albumin 8 contains 23% Met and Cys (Kortt et al., 1991). Due to its resistance to degradation by stomach fluid in vitro (McNabb et al., 1994), the seed albumin 8 has become an ideal protein to improve forage for ruminant crop production. In an early study, SSA was expressed in transgenic lupin (Lupinus angustifolius L.) to increase the nutritive value (Molvig et al. 1997). This study revealed that the dramatic increase in Met resulted in a significant decrease in Cys in lupin seeds. Similarly, using an endosperm-specific promoter from a wheat high molecular weight glutenin gene, Hagan et al. (2003) demonstrated the expression and stable accumulation of the sunflower seed 2S albumin in transgenic rice (Oryza sativa), resulting in a 7% increase in SSA of total seed protein. However, the grain sulfur amino acid content showed little change as compared to the parental genotype. In addition, the limited sulfur proteins were re-allocated from endogenous proteins to the new sulfur sink in the grain. More detailed studies showed that the production of the heterologous protein came at the expense of other endogenous sulfur-rich compounds or endogenous Met-rich proteins (Amir, 2008). Unfortunately, the
sunflower seed 2S albumin was also found to be allergenic to some people, reducing the application of these proteins in crops (Kelly and Hefle, 2000; Moreno and Clemente, 2008).

Many efforts have also been made to identify Lys-rich proteins for improving crops, especially the cereals. Legume proteins usually contain more Lys residues than cereal proteins; therefore expression of an 18-kDa Lys-rich protein from winged bean (*Psophocarpus tetragonolobus*) under the control of the rice glutelin Gt1 promoter in rice was tested. Although 18-kDa Lys-rich protein was stably accumulated in the mature seeds of transgenic rice, its accumulation level in seeds reached only about 1% of the total seed storage protein which was not sufficient to increase the content of lysine (Sun et al., 2001). An improved approach was used to alleviate this problem. The rice Gt1 gene, encoding a glutelin storage protein, was fused to a heterologous cDNA encoding a Lys-rich protein from winged bean to facilitate Lys improvement in rice (Wenefrida et al., 2009; Sun and Liu, 2008). This fusion protein was successfully expressed in rice endosperm and accumulated to about 30% of the total seed protein. Additionally, accumulation of this fusion protein resulted in a significant increase in the Lys content, up to 45% more than the non-transgenic line.

Overall, several candidates genes have been targeted for improving protein content in crops. However, introduction of these genes in transgenic plants often resulted in an increase in one of the essential amino acids but a decrease in others, leading to an imbalance of the AAs in transgenic crops. To avoid these effects, a seed protein, AmA1 (Amaranth Albumin 1) was expressed in transgenic potato under the control of a tuber-specific, granule-bound starch synthase promoter (Chakraborty et al., 2010). Transgenic tubers exhibited an increase of up to 60% in total protein content and a significant increase in AAs, notably Lys, Tyr, and sulfur AAs, which is considered to rebalance the AAs in potato tubers. They demonstrated that AmA1 potato tubers are nontoxic, nonalergenic, and safe for consumption (vs. crops with Brazil nut 2S Albumin). The nutritional value of wheat is limited in Lys and Thr in the prolamin fraction (Matta et al., 2009). Using a wheat endosperm-specific promoter, AmA1 has been expressed in wheat, resulting in a higher content of EAAs (Tamas et al., 2009).

### 1.3.2.2 Enhancing free essential amino acids in plants

Another strategy for enhancing the essential amino acid content in plants is to increase the level of desirable EAAs in the free form (vs. protein-bound) by manipulating amino acid biosynthesis
and catabolism. Aspartokinase (AK) and dihydrodipicolinate synthase (DHDPS) are two key enzymes in the biosynthetic pathway of lysine and are feedback inhibited by Lys (Thu et al., 2007). Lys synthesis can be increased by expressing the mutant AK and/or DHPS enzymes(s) that are insensitive to Lys feedback inhibition. Several previous studies have shown the feasibility of this approach (Galili and Hoefgen, 2002; Galili et al., 2002; Wang et al., 2007). For example, the dhdps-r1 gene encoding a DHPS enzyme insensitive to feedback inhibition was expressed in the transgenic pigeonpea [ Cajanus cajan (L.) Millsp] seeds by using a phaseolin or Arabidopsis 2S2 promoter. It was found that overexpression of these Lys-insensitive forms of mutant DHPS enzyme resulted in over-producing free Lys by 1.6 to 8.5-fold over the wild-type plants (Thu et al., 2007). However, the total seed Lys content was not increased. In plants, Lys catabolism is controlled by a bifunctional enzyme, lysine ketoglutarate reductase (LKR)/ saccharopine dehydrogenase (SDH). In more recent years, many transgenic plants with suppressed LKR/SDH activity have been generated to increase Lys content in seeds (Azevedo and Arruda, 2010). For example, transgenic plants containing both the bacterial DHDPS and endosperm-specific RNA interference (RNAi) suppression of the LKR/SDH were produced, which resulted in a significant increased in free Lys (Frizzi et al., 2008; Reyes et al., 2009). Again, no significant major changes in total Lys content in the seeds were obtained.

Tryptophan is another limiting EAA in cereal crops. Anthranilate synthase (AS) is also feedback inhibited in the Trp biosynthetic pathway. Overexpression of feedback-insensitive mutant AS in transgenic rice lead to increased Trp accumulation up to 180 and 35 fold in callus and leaves, respectively (Tozawa et al., 2001). Likewise, cystathionine γ-synthase (CGS) is the first enzyme of the Met biosynthesis pathway which is critical in determining the Met levels in plants (Amir, 2008). Overexpression of CGS gene from Arabidopsis led to a significant increased in Met content transgenic tobacco (Amir, 2008; Hacham et al., 2006). Although an increase in the free EAAs in several important crops can be achieved by metabolic engineering, a disadvantage of enhancing free EAAs (vs. protein-bound) is that free amino acids could be leached from the plant tissues and lost during boiling and other processing (Sun and Liu, 2004).

Although expression of an EAA-rich protein and metabolic engineering of the EAA biosynthetic pathways have enhanced the levels of EAA in crops, there are some drawbacks as mentioned above. To overcome these undesirable effects, an integrated approach to enhance a specific free EAA pool (the source) though modification of the metabolic pathway and to trap the over-
produced free EAA (the sink) by over-production of a protein rich in the same EAA in the same transgenic event were applied. The strategy enhanced the expression of the EAA-rich protein because of the adequate free EAA, and reduced abnormal phenotypes by removing and trapping the excessive free EAA in the EAA-rich protein. Guenoune et al. (2003) demonstrated that coexpression of the soybean vegetative storage protein β subunit with the bacterial-insensitive DHPS gene resulted in a significant increase in protein-bound Lys in transgenic tobacco plants. A similar result was found in transgenic narbon bean (Demidov et al., 2003). Expression of Brazil nut 2S albumin (BN2S) resulted in a significant elevation in Met in transgenic crops, but the increased Met content was only 40% of the FAO standard (Galili et al., 2002). However, overexpression of Met-rich Brazil nut 2S albumin together with a feed-back insensitive bacterial AK resulted in 2.0 to 2.4 fold increase of protein-bound Met in transgenic seeds as compared to wild type plants which was close to the FAO standard (Demidov et al., 2003). These results suggested that enhancement of free EAA pools may help the accumulation of proteins rich in the respective EAAs.

1.4 Production of ‘value added’ crops

New developments in biotechnology have led to the generation of plant-based systems to produce a wide rage of ‘valuable materials’. Taking advantage of photosynthesis, plants can be the robust biosynthetic machinery which produces a diverse array of compounds that support human activities, such as vaccines, diagnostic proteins, bioactive peptides, nutritional supplements, enzymes, antibodies, and biodegradable plastics (Sharma and Sharma, 2009). Plants also offer several advantages over other systems including low cost of input and raw material, free of animal pathogens and diseases, rapid scale-up production, and in some cases, direct delivery without purification (negating the requirement for needles or syringes such as are used with many vaccines) (Streatfield, 2002; Streatfield and Howard, 2003a, b). Additionally, plant-based systems also avoid undesired modifications of protein caused by microorganisms such as Escherichia coli and yeast species (Horn et al., 2003). Seeds could be genetically engineered to work as economical bio-factories for large scale production of those valuable materials for industrial and pharmaceutical uses. Since seed production is the last stage of plant development and unlike leaves and roots, accumulation of new products will not result in detrimental effects to plant viability. Proteins expressed in seeds may remain stable for longer periods at room temperature, and are easy for storage and transportation (Howard and Hood,
2005). The seed contents can be extracted more easily than green leaf tissue by established and efficient processing methods.

### 1.4.1 Production of biomedical reagents in plants

Since Mason et al. (1992) reported the first transgenic plant expressing a recombinant viral antigen, transgenic plants have been investigated widely for production of human and animal biopharmaceutical recombinant proteins. With population increases, the need of vaccination for disease can not be met in many areas in the world, especially some developing countries due to the high cost and the difficulties in mass administration of the vaccines to humans (Koprowski and Yusibov, 2001). The use of plants for production of vaccines and other biomedically important proteins could be a way to resolve this current problem. For example, it has been shown that up to 400 million doses of clean, safe and fully functional anthrax vaccine antigen could be produced from one-acre of transgenic plants (Koya et al., 2005; Watson et al., 2004). Several plant types and systems have been used to express a wide range of vaccine antigens including tobacco, *Arabidopsis*, potatoes, strawberries, seeds of maize, rice, beans and tobacco, suspension cell cultures of tobacco and maize, hairy root cultures and transformed chloroplasts of various plants (Rybicki, 2010). Plant-produced vaccine can be purified from plant material before delivery, or delivered orally in edible plant material that has been processed to a homogenous and stable product.

Maize has been used extensively to produce a number of antigens and antibodies due to its high potential yield and very well-established milling/processing technology (Rybicki, 2010). For example, the *E. coli* heat labile enterotoxin B subunit protein (LT-B) is an oral vaccine against specific pathogens that produce the diarrhea inducing toxins LT-B and cholera toxin B (CT-B) in the gut (Chikwamba, et al., 2002). Chikwamba et al. (2002) demonstrated that *E. coli* LT-B under the seed endosperm specific promoter (27kD gamma zein) was successfully expressed and accumulated stably in transgenic maize seed. The expression level of LT-B in maize seed ranged from less than 0.01% to 0.07% LT-B of total soluble proteins. When transgenic maize-fed mice were challenged with LT or CT toxins, they induced a higher anti-LT-B and anti-CT-B compared to the mice fed with the equivalent amount of bacterial LT-B, indicating maize-synthesized LT-B has biological and immunological functions comparable to the native LT-B protein that can protect immunized mice from LT or CT toxins. A similar study was carried out
to express LT-B and swine transmissible gastroenteritis (TGE) in maize seeds (Lamphear, et al., 2002). The result showed that both of the maize-produced LT-B and TGE vaccines were strongly immunogenic and protective in chicken and swine, respectively, and this study is an excellent proof of efficacy and stability of recombinant vaccines delivered in maize seed. Transgenic maize also has been used to produce the simian immunodeficiency virus (SIV) major surface glycoprotein gp130 (analogous to human immunodeficiency virus, HIV gp120) (Horn et al., 2003). SIV was glycosylated and accumulated properly in maize seeds. Expression levels ranged from 0.007 to 0.08% of the total soluble protein. The expression level observed of 0.08% of total soluble protein corresponding to approximately 10 μg of antigen per g of seed would allow antigen does of up to about 100 μg to be delivered to individual mice in several feeding studies.

Likewise, rice also has been chosen as a production and delivery vehicle. As16 is an antigen protective against the roundworm Ascaris sum, which causes gastrointestinal disease in both humans and animals. As16 was fused with cholera toxin B subunit (CTB) and expressed in transgenic rice under the control of the endosperm-specific glutelin-B1 promoter (Matsumoto et al., 2009b). The expression level of recombinant As16 in the endosperm reached 50 μg/g seed. When mice were fed with transgenic rice seeds, a weak immunogenic response was observed. However, after adding cholera toxin to the vaccine formulation, the mice produced As16 specific serum antibody against the infection from the roundworm. They suggested that the rice-delivered antigen works as a prophylactic edible vaccine for controlling parasitic infection in animals. Wu et al. (2007) expressed infectious bursal disease virus (IBDV) host-protective immunogen VP2 protein in rice seeds under the control of the rice glutelin Gt1 promoter. The expression level of VP2 protein ranged from 0.678% to 4.521% of the total soluble seed protein. The pathogen-free chickens fed with transgenic rice seeds developed antibodies against IBDV, and were protected when challenged with a highly virulent IBDV. They demonstrated that transgenic rice seeds containing IBDV VP2 can be an effective, safe and inexpensive vaccine against IBDV.

A wide rage of vaccines has also been produced in Arabidopsis and tobacco (Wu et al., 2009a; Kohl et al., 2007; Lai et al., 2010; Bendandi et al., 2010). For example, the Sl gene encoding σ-C protein of an avian reovirus (ARV) was expressed in transgenic Arabidopsis under the control of a superpromoter from a patented plant-expression vector pE1857 derived from pGPTV plasmid. This resulted in levels of σ-C protein ranging from 0.5 to 4.9 % of the total soluble protein from leaf materials (Wu et al., 2009a). Chickens fed with plant-derived σ-C protein induced a variety
of immunoglobulin G (IgG) antibody responses, and were protected from challenges with ARV. They suggested that the plant-derived σ-C protein can be potentially used in the large-scale production of a vaccine against ARV in commercial poultry.

In humans, papillomaviruses (HPVs) are the major cause of cervical cancer. HPVs-11 L1 gene lacking the region encoding the C-terminal nuclear localization signal (NLS-) was expressed in transgenic Arabidopsis and tobacco, which resulted in 12 μg/g and 2 μg/g of HPV-11 L1 NLS-proteins in transgenic plants respectively (Kohl et al., 2007). HPV-11 L1 NLS- was stably accumulated in transgenic plants. Inoculation of rabbits with plant-derived HPV-11 L1 NLS-induced a weak immune response and was not able to neutralize HPV-11 pseudovirion infectivity. However, Paz de la Rosa et al. (2009) reported expression in tomato plants of a chimeric protein containing HPV 16 L1 and a string of T-cell epitopes from HPV 16 E6 and E7 fused to the C-terminus, which resulted in the chimeric protein accumulating to 0.05 to 0.1% of the total soluble protein. Intraperitoneal inoculation of HPV 16 L1 proteins in mice was able to induce a significant antibody and cytotoxic T-lymphocytes response. Experiments in vivo of whether the chimeric proteins are able to induce regression of disease and resolution of viral infection in mice are still in process. They assumed that chimeric proteins in this study may be the potential basis for developing therapeutic vaccines.

Another example of tobacco expressing a human vaccine centers on lymphoma (Bendandi et al., 2010). One quarter of Non-Hodgkin’s lymphoma (a hematologic malignancy) cases are caused by B-cell lymphoma. Each clone of malignant B cells expresses a unique cell surface immunoglobulin (Ig). New active immunotherapy has been focused on vaccination with the patient’s own idiotype (ID). ID-containing IgG molecules of 20 lymphoma patients and 2 mouse lymphoma models conjugated with keyhole limpet hemocyanin (KLH) were expressed in tobacco at levels between 0.5 and 4.8 g/kg of leaf biomass. Several patients Igs produced in plants revealed specific cross-reactivity with sera from the same patients immunized with hybridoma-produced Id vaccine. Moreover, the plant-produced ID-KLH induced comparable protection levels in tumor challenged mice as hybridoma-produced Igs.

In addition to vaccines, plant-based systems have also been used to produce a number of peptides and proteins for pharmaceutical uses, including interferon (Wu et al., 2009b; Fukuzawa et al., 2010; Sawahel, 2002), interleukin-12 (Liu et al., 2008), human insulin-like growth factor-1
For example, interleukin-12 (IL-12) is an important cytokine involved in directing cell-mediated immunity, and is a potential vaccine adjuvant and anticancer therapeutic (Liu et al., 2008). However, there is not an effective bioproduction system for this complex heterodimeric glycoprotein which requires a eukaryotic production system to yield fully functional products. Liu et al. (2008) revealed that mouse interleukin-12 (MuIL-12) was expressed and accumulated correctly in transgenic tobacco under the control of the constitutive plant promoter (double-enhanced 35S promoter) with levels of 3-5 $\mu$g/mg of total soluble proteins. Feeding mice with plant-derived MuIL-12 induced secretion of interferon-$\gamma$ from mouse splenocytes and stimulated splenocyte proliferation, which is comparable to mice fed with commercially available animal cell-derived MuIL-12. Interferons are cytokines that can induce an antiviral state to moderate some cancers and diseases caused by viruses in vertebrate cells. Their studies suggested that plants produce fully functional MuIL-12 more effectively than commercial production in animal cells. Recently the Atlantic salmon (Salmo salar) interferon gene (SasaIFN-$\alpha$1) was transformed into potato and rice plants with low expression levels (Fukuzawa et al., 2010). In vitro antiviral activity of the SasaIFN-$\alpha$1 derived from plants showed that the survival rates of pancreatic necrosis virus infected cultured fish cells pre-treated with transgenic samples was up to 95% as compared to 30-47% of cells pre-treated with non-transgenic samples. They revealed an antiviral effect of the SasaIFN-$\alpha$1 produced from plants.

Human insulin-like growth factor-1 (hIGF-1) is essential for normal fetal growth and development. It is known to promote proliferation and survival of many cell types. In addition, it has significant therapeutic potential due to its ability to resemble insulin in many aspects of physiology (Panahi et al., 2004). Panahi et al. (2004) reported that hIGF-1 was correctly expressed around 113 ng/mg total proteins in transgenic rice and 27 ng/mg total proteins in transgenic tobacco plants under the control of maize ubiquitin 1 promoter. An in vitro study showed plant produced hIGF-1 was effective in inducing the growth and proliferation of human SH-SY5Y neuroblastoma cells, which disclosed that the plant-derived hIGF-1 was stable and biologically active. The SH-SY5Y cell line can proliferate and differentiate only in the presence of hIGF-1 or insulin. They suggested that hIGF-1 may become a substitute therapeutic agent of insulin in subjects with defects in insulin receptor signaling as hIGF-1 and insulin have some overlapping roles.
1.4.2 Production of biodegradable plastics in plants

With the price of oil rising, alternative sources of fuel and oil-based products such as plastics have been explored around the world. Petroleum-derived plastics cause lots of environment pollution, so renewable and biodegradable plastic from natural carbohydrate polymers are of interest to people. Biodegradable plastics can be completely degraded in landfills or composters or sewage treatment systems by the natural micro-organisms (Mooney, 2009). Plants have been engineered to produce biodegradable polymers for raw materials of plastics (Matsumoto et al., 2009a). Novel polymers are also introduced into plants to produce plant-based biodegradable polymers. For example, hydroxyalkanoate is stored as inclusion bodies in most bacteria under nutrient-limiting and carbon-excess conditions (Mooney, 2009). Hydroxyalkanoate is completely biodegradable, which has been used to produce films, compost bags and disposable flatware, and bottles (Mooney, 2009). Thus, hydroxyalkanoate has been selected as a source of non-polluting renewable plastics and elastics. For example, transgenic Arabidopsis expressing modified hydroxyalkanoate synthase resulted in hydroxyalkanoate content up to 1.8 mg/g cell dry weight (Matsumoto et al., 2009a). Polyhydroxybutyrate (PHB) is another source of biodegradable and renewable plastic in hydroxyalkanoate family. PHB was introduced into biomass crop switchgrass (Panicum virgatum) with expression levels up to 3.72% dry weight in leaf tissues and 1.23% dry weight in whole tillers (Somleva et al., 2008). Although these transgenic polymer levels were lower than the 7.5% dry weight estimated to be necessary for the commercialization of PHB-producing switchgrass, this study showed a promising future to achieving this goal.

1.4.3 Production of nutritional fatty acids-added plants

Plants also have the potential for producing non-native nutritional fatty acids. For example, some very long chain polyunsaturated fatty acids (VLCPUFA) play an important role in human health and nutrition. Omega-3 VLPUSA is cardiovascular-protective component of the human diet, for example, docosahexaenoic acid (DHA), eicosapentaenoic acid (EPA) are important for optimal neonatal health and brain development (Napier and Sayanova, 2005), because of which, DHA has been used in infant formula (Napier, 2007; Molvig et al., 2007). They are usually isolated from oily fish, which decline dramatically due to overfishing (Heinz, 2006). Qi et al. (2004) reported that Arabidopsis expressing genes encoding components of the alternative desaturation pathways for C20 polyunsaturated fatty acids (PUFA) resulted in accumulation of 43% AA and EPA of the total C20 PUFA products in leaf tissues. The production of these fatty
acids in transgenic plants is a sustainable and cheap way to produce alternative sources of fish oils.

1.5 ‘Empty container’ hypothesis

Despite extensive research to improve EAA content in crops as discussed above, higher levels of EAA were found only in some plant species and not in others. Although the transgenes were expressed and EAAs were increased, the gene products only accounted for a small portion of the total protein content. It has been theorized that limited expression of foreign proteins in seeds may be because of high expression levels of endogenous protein targeted to the seed and a bias of biosynthesis of natural reserves (Molvig et al., 1997). If the seed could be partially emptied of some of these proteins then a void would be made which could be filled by foreign or genetically engineered proteins or these valuable materials. This has become known as the ‘empty container’ hypothesis.

It has been discussed that alternations in the level of one component can affect the other components (Weselake et al., 2009). Often increasing or decreasing one storage component results in a compensatory change of the other components. For example, reduced starch levels found in wrinkled pea seed mutants lead to increased protein content on a per gram basis, which showed a change in partitioning between starch and protein (Golombek et al., 2001). In peas (Pisum sativum), amino acid permeases such as VfAAP1 may control storage protein synthesis. Overexpression of VfAAP1 increases the levels of N and protein in the seeds; however, it resulted in decreased in sucrose/starch and individual seed weight (Weigelt, et al., 2008). There is a mutant in Arabidopsis, shrunken seed 1 (sse1), in which oil and protein body contents are dramatically reduced, but instead starch accumulates (Lin et al., 2004). It is obvious that there must be some form of regulation that controls the level and the composition of the storage components in seeds. Perhaps this regulation occurs at multiple levels and involves some sensors for seed filling. For instance, if the total level of proteins and lipids are programmed into the sensors, so when the protein and lipid levels are adjusted to fill in the predetermined amount of space in seeds, the starch accumulation will be repressed. If protein and lipid levels can not be modified to make up any changes in the seed, the starch production will take over to compensate the differences with a maximum production capacity. This might explain why there is an accumulation of starch in sse1 (Lin et al., 2004).
However, decreasing the level of one component of seed reserves may not affect other components. For example, rape (*Brassica napus*) was transformed with an antisense gene for cruciferin to increase EAAs content in seeds (Khno-Murase et al., 1995). Increases in three EAAs, Lys, Met and Cys were revealed but there was no significant change in the total protein and lipid levels. A similar result was found with rape expressing an antisense gene for napin (Kohno-Murase et al., 1994). There was an increase in cruciferins but no change in total protein or lipid levels even though there was a change in the ratio of lipids to enhance the 18:2 levels at the expense of 18:1 lipids. Therefore, the plant can control both the levels of seed macromolecules, as well as the composition of the macromolecules. So the proteins and lipids may be regulated in their own biosynthetic pathways that determine their composition. The composition of proteins and lipids could be regulated individually. When one of the seed components changes in composition, the composition of the other could be modified due to an interaction between certain intermediates of the two biosynthetic pathways. If the total seed storage proteins are decreased, the total lipids may be increase or vice versa. But, if only one component of seed storage protein decreases, another seed storage protein can make up and lipid levels may remain the same, but the fatty acid composition of the lipids may change. So if a mutation somehow affected or bypassed this sensor, it would be possible to generate a seed with lower amounts of one component of seed storage protein without changing the total lipid and starch content. This would result in the production of a partially empty container.

*Arabidopsis* is considered to be a model plant system in plant biology due to its small genome size, low abundance of repetitive sequences, well defined RFLP maps, rapid generation time, and the extensive knowledge of its genetics. In addition, the ease of mutagenesis and transformation make it a useful system for testing the ‘empty container’ hypothesis. In *Arabidopsis*, the main seed storage proteins are the 2S albumins (napin) and 12S globulins (legumin) (Heath et al., 1986; Krebbers et al., 1988; Kroj et al., 2003; Li et al., 2007; Pang et al., 1988; Van der Klei et al., 1993). The 12S globulins, also referred to as cruciferins, usually are hexameric complexes composed of six subunit pairs- each acidic (α) and basic (β) subunits linked by disulphide bonds (Li et al., 2007; Pang et al., 1988). Each pair of α and β subunits are proteolytically cleaved from the same precursor protein after disulphide bond formation (Higgins, 1984). There are four genes encoding 12 cruciferins, including *CRA1* (At5g44120), *CRB* (At1g03880), *CRC* (At4g28520) and *CRU2* (At1g03890) are located on chromosome 5, 1, 4 and 1, respectively in the *Arabidopsis*
To determine the validity of the ‘empty container’ hypothesis, mutants that are defective in the accumulation of seed storage proteins must first be identified. In *Arabidopsis*, two mutants, *abi3* (abscisic acid insensitive) and *fus3* (anthocyanin biosynthesis) have also been indentified that are defective in the accumulation of the two major seed storage proteins, the 12S cruciferins and the 2S napins at both the protein and mRNA levels (Nambara et al., 1995; Bâumlein et al., 1994; Finkelstein and Somerville, 1990). However, both mutations caused pleiotropic effects that show they may regulate the expression of many different genes associated with seed maturation. For example, *abi3* mutant has reduced in seed dormancy and effect on the quality or composition of storage protein or lipid (Finkelstein and Somerville, 1990). Hence the use of these types of mutants in this study will be difficult as more than one gene is affected. Hou et al. (2005) found a naturally occurring *Arabidopsis* mutant, Cape Verde Islands which missed one of the 12S cruciferin bands in the seed protein profile with the present of a new band of lower molecular mass. The mutation is due to deletion of a few amino acids in a binding region which is important for the hexamer formation although the mutant accumulates normally relative to other seed storage protein subunits. However, they thought they still need to assess if there are some specific effects from the mutations in the mutant. In addition, they did not evaluate the amount of 12S cruciferin defective in this mutant. Moreover, they have not done any investigation on testing the utility of this mutant such as introducing the foreign gene into the mutants and determine if the foreign gene can be overexpressed.

It would be more useful to find a seed storage protein mutant that is defective in only one class of seed storage proteins that may decrease the complexity of the experiments and analysis of the results. Therefore, *Arabidopsis* seeds were mutagenized by ethylmethane sulfonate (EMS) method and screened specifically for a seed storage protein mutant. A mutant was identified by Dr. Peter McCourt and Dr. Dan Riggs through the analysis of seed extracts by SDS-PAGE where the gel profile of storage proteins from mutant seeds showed two missing bands at 35 kDa and 25 kDa. This mutant has been named *ssp-1* (seed storage protein mutant 1).
1.6 *ssp-1*

*ssp-1* is a recessive, non-lethal mutant deficient in one of the 12S major seed storage proteins in *Arabidopsis*, the 12S cruciferins (legumin-type globulin), with no other observable phenotypic effects. Moreover, according to the study by a previous graduate student Lee (2000), *ssp-1* is defective in cruciferins that did not affect other components such as other proteins, lipids or starch levels. The mutation of *ssp-1* is due to a G-A transition in the *CRC* gene residing on the distal arm of chromosome 4 near cer2 (74.5 cM), which results in turning a tryptophan codon (TGG) into a premature stop codon (TGA) (Lee, 2000). Therefore, there is no precursor protein CRC produced. As a result, there is no production of CRC, which is also called CRU3. The significant characteristic of *ssp-1* is that *ssp-1* seeds contain 20-27% less protein than wild type seeds. Therefore *ssp-1* has been termed the ‘empty container’, and that this might provide a tool to drive higher-level expression of foreign proteins than might be accomplished in a wild type genetic background. Moreover, *ssp-1* might be developed as protein-expression systems for production of hard-to-make proteins, such as monoclonal antibodies, peptides, and other human recombinant therapeutic proteins.

To validate the ‘empty container’ hypothesis, Lee (2000) transformed *ssp-1* plants with a DNA construct using a reporter gene encoding phytohemagglutinin (PHA) and identified PHA homozygotes in *ssp-1*. Eight PHA homozygotes with the known copy number of the PHA gene were crossed to wild type *Arabidopsis* and screened for PHA homozygotes in a wild type background. Lee (2000) evaluated the PHA expression level of 6 PHA homozygotes with more than 2 copies of PHA in both genetic backgrounds using immunoblots combined with densitometry. The PHA levels ranged from 0.5% to 1.4% of the total protein in transformants similar to the levels in the backcross of 0.17% to 1.5%. What she found interesting was that the transgenics with more copy number of PHA had higher levels of PHA. As a result, for 5 sets of plants, the levels of PHA detected did not show significant differences. Only one PHA transformant showed a significant enhancement of PHA expression (1% of the total seed protein) compared to its backcross (0.2% of the total seed protein). This showed some evidence to support the validity of the ‘empty container’ hypothesis. It may be a matter to find transgenic lines with high expression levels of PHA in the seed. Also, expression of the PHA may be contributed by the copy number of the transgenes, and also might be affected by the integration site of the PHA gene in the genome that can not transcribe at high rates. Since Lee’s results were
inconclusive, it is difficult to determine if the hypothesis is validated or not. A continued study must be done to find the concrete evidence to support this hypothesis. For example, it is necessary to identify transformants with a single copy of the \textit{PHA} gene, which simplify genetic analysis and quantification of expression levels because the effects of multiple copy number are unknown. In addition, it will be useful to screen for a homozygote transgenic line that expresses PHA to high levels.

### 1.7 The objective of my thesis

My project is to test the feasibility of this ‘empty container’ hypothesis by introducing a reporter gene construct into the \textit{ssp-1} mutant and measuring reporter gene expression levels as compared to wild type. To accomplish my project I have 3 objectives to be achieved: (1) to identify transgenics with a single copy of the transgene in the \textit{ssp-1} genetic background and identify transgenic homozygotes in \textit{ssp-1}; (2) to cross single copy transgenics with the wild type to generate transgenic homozygotes in the wild type genetic background; (3) to compare transgene expression levels in both an \textit{ssp-1} and a wild type genetic background. Theoretically, if the ‘empty container’ hypothesis is valid, the expression level of the reporter gene will be higher in the \textit{ssp-1} mutant than wild type.
2 Material and Methods

2.1 Seed lots

The seed slots used were Landsberg erecta (Ler) and ssp-1 Arabidopsis thaliana plants which were obtained from Dr. Riggs (University of Toronto).

2.2 Single pot plantings of Arabidopsis thaliana

The required numbers of 25-cm² pots were filled with freshly mixed Premier Pro-mix PGX®. Seeds of ssp-1 and Ler were stored at -20°C overnight and the next day they were surface sterilized through a series of wash solutions by rocking on a tipper: 20 min in tap water, 5 min in 70% ethanol, and 5 min in 10% bleach, 0.1% SDS and 5 washes of tap water. The sterilized seeds were suspended in tap water and spread evenly on the surface of the moist potting mix by Pasteur pipettes. Every 10 single pots were placed in a plastic tray which was covered with transparent plastic cover. They were grown in a 16-h light, 8-h darkness growth chamber at 23°C. The seedlings of each pot were thinned and left six plants per pot.

2.3 Plasmids used for transformation

Plasmid pGPTV-HPT carries the hygromycin resistance gene as a selectable marker and was used as a vector for the reporter gene encoding phytohemaglutinin (PHA). This construct was mobilized into Agrobacterium tumefaciens strain GV3101 and termed LC286. Another DNA construct which is plasmid pEGAD-link carried the bar gene as a selectable marker under the control of the cauliflower mosaic virus 35S (CaMV35S) promoter along with the PHA gene was made. Both of the DNA constructs were used for ssp-1 transformation.

2.4 Production of DNA constructs

The PHA gene (Riggs et al., 1989) was cloned previously in other labs. The plasmid, pDR214, containing the PHA gene and the vector, pEGAD-link carried the bar gene were both cloned in Escherichia.coli, strain DH5α. The two clones (from Dr. Riggs’ lab) were designated as Lab Clone (LC) 175 and LC 505, respectively. LC175 and LC 505 were grown separately in 250 mL Erlenmeyer flasks containing 50 mL of LB (2% LB Broth Lennox, BioShop®) with antibiotic Kanamycin (Kan) (50 μg/mL) for overnight at 37 °C in a Lab-line Orbit Environ shaker at 300 rpm. The cells were harvested by centrifuging at 4000× g for 10 min in 2 sterile 50 mL Corex
tubes in Sorvall® RC 5C Plus centrifuge (Rotor: SS-34). The plasmid DNAs of these harvested cells was purified by using a QIAGEN® Plasmid Midi Kit (Cat. No. 12143). The plasmid, pDR214, and the vector, pEGAD-link were first digested separately with Hind III (New England Biolabs® Inc.) in a total volume of 95 μL (20 μL plamid DNA, 62 μL sterile double distilled water (DDW), 10 μL NEB buffer and 3 μL Hind III) at 37 °C water bath for 2 h, and then transferred to 65 °C water bath for 20 min to heat inactivate Hind III. An aliquot of 2 μL of 2.5 M Sodium Chloride (for adjusting the salt concentration in the buffer for EcoR I) and 3 μL EcoR I (Invitrogen™ life technologies) were added to the 95 μL digestion reaction solution, then the reaction was incubated at 37 °C for another 2h. The digested DNA products were resolved on a 0.9% agarose gel. The 2.8 kb PHA fragment and the 12.6 kb pEGAD-link fragment were cut out with razor blade and gel purified using the QIAquik® Gel Extraction Kit (Cat. No. 28704). The ligation of the PHA gene EcoR I/Hind III to the pEGAD-link EcoR I/Hind III was carried out in a total volume of 20 μL containing 1 unit T4 ligase (Fermentas) and 5X T4 ligase buffer. There were four ligation reactions with various amounts of digested PHA and vector as follows: Ligation reaction 1 (L1): 1.5 μL PHA and 2 μL vector; L2: 2.5 μL PHA and 3 μL vector; L3: 0 μL PHA and 2 μL vector; L4: 2.5 μL PHA and 2 μL vector. Each reaction mixture was adjusted volume to 15 μL with DDW and heated at 50 °C water bath for 2 min, and then chilled on ice, and 1 unit T4 ligase and 4 μL 5X T4 ligase buffer were added to the final volume of 20 μL. These ligations were incubated at room temperature (RT) overnight.

2.5 Transformation of E.coli

All of the four ligation products were used in E.coli transformation. 3 μL of each ligation solution was added to four 13 mL Reagent and Centrifuge tubes on ice. Competent E.coli, strain DH5α, thawed 10 min on ice and using cold pipette tips four 75 μL aliquots were transferred to the tubes on ice. The cells in each tube were mixed gently and kept on ice. The cells were transferred to 42 °C water bath for 30 sec, and put back on ice for 2 min. After adding 1 mL of LB to each tube, cells were grown 1h at 37 °C in the shaker at 300 rpm. Aliquots of 500 μL of each tube were plated on LB + kanamycin (kan, 50 μL/mL) respectively and incubated overnight at 37 °C. Randomly picked colonies were inoculated into 50 μL minicultures of LB + Kan (50 μg/mL) on multiwell plate and grown for 3 h at 37 °C in a Precision® Gravity Convection incubator with occasionally mixing. PCR of the PHA fragment was carried out to confirm the
presence of the PHA gene in the cells. After PCR two independent colonies were identified and used to inoculate 5 mL of LB + Kan (50 μg/mL) and grown overnight (< 16 h) at 37 °C in the shaker at 300 rpm. Plasmid DNA was isolated from the two cell cultures using QIAprep® Spin Miniprep Kit (Cat. No. 27104). The plasmid was digested with EcoR I and Hind III and the presence of a 2.8 kb fragment was the criteria used to confirm the selection of the correct PHA construct.

2.6 Transformation of *Agrobacterium tumefaciens*

*Agrobacterium tumefaciens* strain GV3101 with gentamycin resistance was obtained from Dr. Riggs. 1.5 μL of plasmid DNA was added to 100 μL aliquots of frozen competent GV3101 in a 1.5 ml Eppendorf tube. The tube was incubated at 37 °C water bath for 5 min and the cells grown for 2 to 4 h at RT in a Lab-line Orbit shaker at 200 rpm. After transformation the cell culture was plated on LB + Kan (50 μg/mL) + Gentamycin (Gent, 10 μg/mL) and incubated at RT for 2 d. PCR of the PHA fragment was carried out to validate the correct constructs. The construct was designated LC686. One colony of LC686 was inoculated into 2 mL LB + Kan (50 μg/mL) + Gent (10 μg/mL) and shaken at RT for overnight at 200 rpm. The next day 1 mL of overnight culture was archived by adjusting the solution to 15% glycerol and was stored at -70 °C.

2.7 *Arabidopsis* transformation

18 pots of *ssp-1* and 2 pots of *Ler* seeds were surface sterilized and planted on soil for each round of transformation. When the primary inflorescences emerged from the plants, plants were clipped to favor the growth of multiple secondary bolts (Martinez-Trujillo et al., 2004). *Agrobacterium* LC286 harboring pGPTV-HPT-PHA and LC686 containing pEGAD-link-PHA were grown separately overnight at RT in 50 mL of LB with Gent (10 μg/mL) and Kan (50 μg/mL) in 250 mL Erlenmeyer flasks. The overnight culture was added to 200 mL of fresh LB with the same antibiotics and grown to the stationary phase (OD600>2.0). Cells were chilled on ice and harvested by centrifuging at 6000 rpm (SS-34 rotor) for 10 min at 4°C in 2 sterile 50 mL Corex tubes. The pellet was resuspended in infiltration medium [0.5X Murashige and Skoog (MS) (SIGMA M519), 0.5 g/L MES, pH 5.75, 5% sucrose, 0.05% Silwet L-77] to obtain the desired density (OD600 >1.5) (Clough and Bent, 1998; Martinez-Trujillo et al., 2004). Plants were inoculated by performing direct drop-by-drop inoculation to every flower by using a Pasteur pipette. To increase transformation efficiency, these plants were re-inoculated with the
same *Agrobacterium* strain one week after the first round of transformation. These *Agrobacterium* infected plants were allowed to grow and mature. Seeds T₁ were collected until all siliques were dry and stored at RT in labeled Eppendorf tubes.

### 2.8 Selection of transformed plants

T₁ seeds produced by each of the inoculated plants were surface sterilized in deionized water for 20 min, 70% ethanol for 5 min, 10% bleach, 0.1% SDS for 5 min and washed 5 times in sterile double distilled water by rocking in a tipper. The seeds were resuspended in 5mL of liquid pre-warmed antibiotics selection medium [0.5X MS, 5mM MES, pH 5.75, 1% planta agar (Phytotechnology Laboratories A111), 10 μg/mL Hygromycin b or 10 μg/mL BASTA] and spread evenly in Petri dishes containing hygromycin selection medium (0.5X MS, 5mM MES, pH 5.75, 0.8% planta agar, 10 μg/mL hygromycin b or 10 μg/mL BASTA). Seeds were stratified at 4°C for 4 d prior to transferring to RT and incubated under constant white light. Seedlings susceptible to hygromycin or basta would germinate but become bleached and die. Resistant seedlings with dark green true leaves and well-developed branched roots were selected and transferred into soil to grow to maturity in the growth chamber. T₂ seeds of these resistant plants were harvested.

### 2.9 Isolation of plant DNA

For PCR, genomic DNA was isolated from leaves of transformants and control plants before formation of siliques by the DNA mini-prep method from Dr. Riggs’ lab. A part of cauline leaf (<0.5 cm²) were taken and transferred to a clean 1.5 mL Eppendorf tube. After the plant tissue was ground with a mini pestle, 500 μL of Extraction buffer (100 mM Tris-HCl pH 8.0, 250 mM NaCl, 10 mM EDTA and 0.5% SDS) was added and the tissue was reground and vortexed thoroughly. The samples were placed on ice temporarily until all samples had been homogenized. The samples were centrifuged in a Sorvall® MC 12V centrifuge at 13,000 rpm for 2 min to remove cell debris. Only 300 μL of supernatant containing the DNA was transferred to a 1.5 mL Eppendorf tube containing 300 μL isopropanol and mixed thoroughly by inverting the tubes, and incubated at RT for 5 min. To pellet the DNA the samples were centrifuged at 13,000 rpm for 5 min. After the supernatant was removed the DNA pellet was washed with 750 μL of ice cold 70% Ethanol and centrifuged at 13,000 rpm for 5 min to pellet the DNA. The supernatant was removed and the DNA pellet was vacuum dried in a Savant SVC 100 speed vac.
The pellet was resuspended in 100 μL TE (10 mM Tris-HCl, 1 mM EDTA, pH 8.0) and mixed by gently flipping and then vortexing, and centrifuged at 13,000 rpm for 2 min. The upper 30-50 μL of the supernatant containing was carefully transferred to a clean labeled tube. For PCR amplification, 1 μL of the extracted genomic DNA was used.

For southern blotting, a large scale preparation of genomic DNA of transgenic plants was carried out. The protocol for maize DNA miniprep by Dellaporta et al. (1985) was modified as follows. About 40 to 200 mg of leaf tissue of transformants was weighed and frozen in liquid nitrogen in a mortar, a pestle was used to grind the plants into a fine powder, and the powder was then poured evenly into two 1.5 mL Eppendorf tubes. To each tube 750 μL of Extraction buffer (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 25 mM EDTA; 10 mM β-mercaptoethanol) at 65°C was added. The powdered tissues were dispersed briefly by vortexing, and then a 50 μL of 20% SDS were added to each tube. The sample solution was mixed well and incubated at 65°C for 10 to 15 min with occasional mixing. The tubes were transferred to ice and 250 μL of cold 5M potassium acetate was added. The tubes were shaken vigorously to mix the contents and placed on ice for 15 min. After incubation, the tubes were shaken again and centrifuged at 13,000 rpm for 5 min. For each tube, a 1 ml blue pipette tip stuffed with Calbiochem® Miracloth was prepared. After the spin, the supernatant was filtered through the miracloth to remove particulates. An aliquot of 500 μL of isopropanol was added to the tubes, the contents were mixed well and placed at -20 °C for 20 min, and then centrifuged at RT at 13,000 rpm for 10 min. After discarding the supernatant, the pelleted DNA was resuspended in 150 μL of TE buffer (50 mM Tris-HCl pH 8.0 and 2 mM EDTA) and left at 4 °C for overnight. The following day the samples were warmed to RT and mixed to solubilize, and then centrifuged at 13,000 rpm for 5 min to remove any insoluble debris. The supernatant was transferred to a 1.5 mL Eppendorf tube. The DNA was precipitated by addition of 0.1 volume 3M Sodium acetate (15 μL) and 0.6 volumes of isopropanol (100 μL), and then mixed well and pelleted immediately in a microcentrifuge at 13,000 rpm for 2 min. After the supernatant was removed the pelleted DNA was washed twice with 70% ethanol and vacuum dried for 5 min. The pellet was resuspended in 50 μL of TE buffer (10 mM Tris-HCl pH 8.0 and 0.2 mM EDTA). One microliter of 10mg/ml RNAsel was added to degrade RNA at RT for 30 min and the two samples were pooled into one tube to give about 100 μL of DNA.
2.10 PCR of PHA transgenics

PCR amplification of a 776 bp DNA fragment of the PHA gene was carried out in a Bio-Rad MJ Mini™ Personal Thermal Cycler using specific primers 5′-TCTTCACTGTCCCTCTTCCCTTGTG-C-3′ (PHA Forward) and 5′-GGCGAGATTCAAACCTTCAGATG-3′ (Reverse) annealing at position 17 bp and 792 bp on the coding strand, respectively. The PCR reaction mixture contained 1 μL (about 50ng) extracted leaf genomic DNA of PHA transformants in a final volume of 50 μL containing 5 μL of 10X PCR buffer [100 mM (NH₄)₂SO₄, 100 mM KCl, 20 mM MgCl₂, 200 mM Tris-HCl pH 8.4, 0.1% gelatin and 0.1% Triton X-100], 1 μL dNTPs (10 mM), 1 μL of each primer (10 μM), 40 μL sterile double distilled water and 1 μL Taq DNA Polymerase (0.1 units). Amplification was carried out by denaturation at 94 °C for 45 seconds, annealing at 60 °C for 35 seconds, and extension at 72 °C for 1.5 min for 35 cycles. A pre-denaturation step of 2 min and a final elongation step of 5 min were included. The PCR products were electrophoresed on a 0.9% agarose gel and stained with 0.5 μg/mL ethidium bromide in 0.5X TBE (45 mM Tris, 45 mM Boric acid and 1 mM EDTA) using a Bio-Rad Wide Mini-Sub™ Cell gel electrophoresis apparatus for 1h at 100 V. Following electrophoresis the DNA samples in the gel were visualized with UV light and photographed through a Syngene Genomic and Proteomic Gel Documentation (Gel Doc) System.

2.11 Isolation of seed proteins

Crude seed protein was extracted from the seeds of mature plants. About 100 seeds were ground by a mini-pestle in a 1.5 mL Eppendorf tube. A 300 μL aliquot of 1X Denaturing buffer (1% SDS, 15 mM Tris-HCl pH 8, 25 mM β-mercaptoethanol) at 100 °C were added to the tube. The tube was placed in 100 °C for 5 min, and then centrifuged at 13,000 rpm for 2 min. The supernatant containing clarified seed protein extracts was transferred into a clean tube and stored at -20 °C.

2.12 Quantification of seed proteins

Extracted seed proteins of PHA transformants were quantified by using the Bradford Reagent as described by the manufacturer (Sigma® life science, B6916, Batch #: 080M4359). After quantification, all samples were diluted to the same protein concentration with 1X Denaturing buffer. The same amount of seed proteins of each sample was subjected to SDS-PAGE and
Coomassie staining to verify the concentration. These samples were then subjected to immunoblotting to examine and compare the PHA level in each transformant.

### 2.13 SDS-PAGE

The seed storage protein samples were separated on a 15% polyacrylamide gel by SDS-PAGE at 100 V for 3h in SDS-PAGE running buffer (0.02 M Tris-HCl pH 7.5, 0.2 M Glycine, and 0.1% SDS) using a BioRad Mini-2D gel apparatus. The gel was fixed in destaining solution (9% Acetic acid and 25% Methanol) for 5 min, stained in 0.25% Coomassie brilliant blue in 5% Acetic acid and 50% Methanol for 1 h, rinsed twice with destaining solution, and then destained overnight. Next day, the gel was incubated in distilled water for 45 min by rocking, and dried in a Novex™ DryEase™ Gel Drying System using Novex™ Gel-Dry™ Drying Solution or 5% Glycerol with incubation for 1 h.

### 2.14 Immunoblotting

Seed protein extracts were separated on a 15% polyacrylamide gel by SDS-PAGE at 100 V until the bromophenol blue tracking dye was run off the gel. Using a BioRad Trans-Blot® Cell apparatus, the proteins were transferred to nitrocellulose membrane (BioRad Trans-Blot® Transfer Medium) at 100 V for 1 h in cold transfer buffer (25 mM Tris, 192 mM glycine and 15% methanol). Upon completion of transfer, the blot was immersed in 50 mL of 3% gelatin in 1X TTBS (20 mM Tris pH 7.5, 0.5 M NaCl and 5% Tween 20) in a suitable size of plastic container. The blot was incubated for 30 min at RT by gently rocking and placed in the refrigerator overnight at 4 °C.

The next day, the gelatin was liquefied by placing the container in a 37 °C water bath for 5 min. After discarding the solution, 50 mL of 1% gelatin, 1X TTBS and 17 μL of anti-PHA antibody [880(4), a gift of Dr. Arnd Sturm, UCSD] was added to give a final dilution of 1:3000. The primary antibody reaction was performed for 1.5 h at RT by rocking, and then the blot was washed 5X with 1X TTBS. Another 50 mL of 1% gelatin and 1X TTBS was added with 17 μL of the secondary antibody which was Bio-Rad Blotting Grade Affinity Purified Goat anti-Rabbit IgG (H+L) Horseradish Peroxidase Conjugate (1:3000). The secondary antibody was incubated with the blot for 1h at RT by rocking, and then washed 5X with 1X TTBS, 2X with 1X TBS (20 mM Tris pH 7.5 and 0.5 M NaCl).
Chemiluminescent western blotting substrate was used to detect PHA on the blot. Pierce® ECL Western Blotting Substrate (Thermo Scientific, product #: 32106) was prepared by mixing equal volumes of the two provided solutions, and added on the surface of the blot with gently swirling after decanting the TBS. The blot with the substrate was placed at RT for 10 min, drained, wrapped in SaranWrap and exposed to Bioflex® Scientific Imaging Films for Maximum Sensitivity multiple times to obtain the best image. The film was developed using an x-ray developer [model SRX-101A (Konica Minolta Ltd.)].

2.15 Southern blot analysis of PHA transgenics

Leaf genomic DNA (~5-6 μg) of PHA transgenics and control plants was digested with the restriction endonuclease EcoR I or Hind III and separated on a 1% agarose gel in 1X TBE buffer. DNA was blotted onto an Amersham Hybond™-N+ nylon membrane according to the manufacturer’s instructions (GE Healthcare UK Limited). The gel was trimmed, then agitated gently in Depurination solution (0.125 N HCl) for 10 min, then submerged in Denaturation buffer (1.5 M NaCl, 0.5 N NaOH) for 30 min with agitating, and then placed in Neutralization buffer (1.5 M NaCl, 0.001M EDTA, 0.5 M Tris-HCl pH 7.5) for 30 min with agitating. Between each step the gel was rinsed in distilled water. The capillary blot device was set up as follows. A glass dish of suitable size was half filled with the transfer buffer 20X SSC (0.3 M Sodium citrate, 3 M NaCl pH 7). A piece of glass used as the platform was placed on the glass tray and covered with a wick made from 2 sheets of Whatman 3MM filter paper saturated in transfer buffer. The treated gel was placed on the wick platform. To avoid trapping air bubbles beneath the gel a glass pipette was used to roll on the gel. The gel was surrounded with cling film to prevent the transfer buffer being absorbed directly into the paper towels. A sheet of Hybond N+ membrane of the same size of the gel was placed on top of the gel avoiding with air bubbles. Six pieces of Whatman 3MM paper cutting to the same size as the gel were saturated in transfer buffer and placed on top of the membrane avoiding air bubbles, and then a stack of absorbent paper towels at least 5 cm high with the same size as the gel were placed on the top of the Whatman 3MM paper. A glass plate and a suitable weight (< 750 g for a 20 X 20 cm gel) were placed on top of the paper stack. The transfer was allowed to proceed overnight. After blotting, the membrane was marked to allow identification of the tracks with pencil and air dried. The DNA was fixed to the membrane by baking at 80°C for 2 hours. The blot may be rinsed in 2X SSC before storage or hybridization and stored wrapped in Saran Wrap desiccated at RT.
A 776 bp fragment of the *PHA* gene from PCR was used to make the probe. The *PHA* fragment was separated on an agarose gel, purified, and radio-labeled with ($\alpha$-$^{32}$P) dCTP using the random primer DNA labeling kit (Amersham Rediprime II DNA Labeling System, GE Healthcare UK Limited). To remove the unincorporated nucleotides, this radio-labeled sample was run through Micro Bio-Spin™ 30 Chromatography column (BIO-RAD) according to the instruction. Prehybridization and hybridization were carried out in Modified Church and Gilbert buffer containing 7% SDS, 10 mM EDTA and 0.5 M Phosphate buffer pH 7.2 according to the manufacturer’s instructions (GE Healthcare UK Limited). After the blot was pre-wet first in sterile double distilled water, then 0.5 M Phosphate buffer pH 7.2 in a hybridization tube, it was pre-hybridized for 1 h at 65°C in pre-hybridization buffer in the HybriLinker HL-2000 Hybridizer (UVP Laboratory Products). After pre-hybridization the purified denatured radioactive probe was added to the pre-hybridization buffer. Hybridization was then performed overnight (16 h) at 65°C.

Following hybridization, the blot was rinsed briefly in 2X SSC, 0.1% SDS at RT, washed twice in 2X SSC, 0.1% SDS at RT for 5min, washed twice in 1X SSC, 0.1% SDS at 65 °C for 10 min and washed four times in 0.1X SSC, 0.1% SDS at 65 °C for 5 min. The blot was removed from the last stringency wash, drained and wrapped in Saran-Wrap, and then carried out autoradiography. Hybridized DNA was visualized on Bioflex® Scientific Imaging Films for Maximum Sensitivity after exposure of the blot for 3-4 days (time varies depending on the radioactivity). The film was developed in an x-ray developer [model SRX-101A (Konica Minolta Ltd.)].

### 2.16 Genetic cross of *PHA* homozygote in *ssp-1* and *Ler*

Single pot plantings were made for both of the *PHA* homozygotes and *Ler*. A bud of *Ler* that has not yet pushed their stigma into view through the sepals of was chosen as the female recipient and the interfering parts of the plant such as other buds, leaves and siliques around the bud were removed from that stem. The bud of *Ler* was carefully dissected under a dissecting microscope using watchmaker’s (FST by Dumont, Electronic Switzerland) forceps, leaving only an undamaged carpel including stigma, style and ovary. The anthers of a slightly open flower from the *PHA* homozygote were dissected and used to dust the pollen onto the stigma. The crossed bud was tied loosely with a piece of sewing thread at the base of the carpel as the label. After one
week if the cross was successful a silique would form. When the silique was dried the seeds were collected and designated as F1.

2.17 Sequencing preparation of DNA

The miniprep method was used to prepare the genomic DNA for PCR. PCR was carried out to amplify a 837 bp DNA fragment of the CRU3 gene using specific primers 5′-ATGAGGTCC-CACGAGAACATTG-3′ (CRU3 Forward) and 5′-CGTTAAAGGTAAAGGGACGTGA-3′ (CRU3-1 Backward) annealing at position 4318 bp and 5154 bp on the coding strand, respectively. The mutation of ssp-I is at 4873 bp which is included in this amplified fragment. Amplification was carried out as above except for annealing at 57 °C and extension for 1.5 min. The CRU3 fragment was separated on an agarose gel, and purified using the EZ-10 Spin Column DNA Gel Extraction Kit (Biobasic Inc.). Samples for sequencing were prepared containing 7 μL (50 ng) purified DNA and 1 μL (4 pmol) of primer (CRU3-1 Backward). DNA sequencing was performed by The Centre for Applied Genomics at the Hospital for Sick Children, Toronto.

2.18 Evaluation of PHA levels in transgenic plants

Crude seed proteins were extracted from 100-200 seeds of homozygote PHA transformants and the homozygote backcross. Protein concentrations were determined and equal amounts of total seed proteins were subjected to SDS-PAGE and immunoblotting. In addition, aliquots of purified PHA (14 ng, 22 ng, and 29 ng) were also used to gauge the relative expression levels of the samples. Following immunoblotting, the autoradiograms were scanned by a Canon scanner (CanoScan LiDE100) and Image-J (NIH) was used to evaluate the intensity of the immunoreactive signals. Most lanes of samples had several bands appearing at lower molecular weights than the band for purified PHA. These were processing products of PHA and were still indicative of PHA expression. Because the relative amounts of the processing products were similar to the uncleaved PHA, I decided to use only the uncleaved PHA to compare the samples. The levels of PHA present in each lane were estimated by comparison to the standard. PHA as a percentage of total cellular protein was estimated based on this value, divided by the total amount of protein loaded on the gel.
3 Results

3.1 Preface

ssp-1 is a mutant defective in one of the 12S cruciferins, CRU3. The polypeptide pattern of total seed protein extracts is shown in Figure 4. The cruciferins are synthesized as preproteins and undergo specific cleavage reactions to give rise to two polypeptides, \( \alpha \) and \( \beta \) (Hou et al., 2005). In ssp-1, two polypeptides of 25kD (\( \beta \)) and 35kD (\( \alpha \)) are missing, whereas these bands are readily observed to be abundant polypeptides in the parental Ler samples. Earlier molecular work determined that the nature of the mutation is a G (Guanine) to A (Adenine) transition, which alters a tryptophan codon to a premature stop codon in exon 4 (Lee, 2000). As such, no CRU3 polypeptide is generated in this background and RNA gel blot experiments showed that no CRU3 RNA is present (Riggs, unpublished results).

3.2 The ‘empty container’ hypothesis

The objective of this project is to determine if foreign proteins can be overexpressed in a seed storage protein mutant defective in seed filling (ssp-1). The previous study showed that ssp-1 seeds contain 20-27% less protein than wild type seeds and therefore the ssp-1 seed may work as the ‘empty container’ (Lee, 2000). To test the feasibility of ssp-1 genetic background as the ‘empty container’, my aim was to compare the expression of PHA in plants containing a single copy of the reporter gene in the ssp-1 background verses a wild type background. I transformed ssp-1 plants with a DNA construct using a reporter gene encoding phytohemagglutinin (PHA). PHA is a normally occurring seed storage protein gene of the common bean, Phaseolus vulgaris. The construct contains the promoter and structural gene and as PHA is a storage protein, it has the appropriate signals/domains that would target it to protein bodies in the seed. A western blot with seed storage proteins from bean, ssp-1, Ler, and purified PHA protein indicated that Arabidopsis seeds do not contain any PHA (Figure 5). Thus, PHA can be used as a reporter to test the ‘empty container’ hypothesis.
Figure 4. Polypeptide pattern of seed storage proteins from Ler and ssp-1. The ssp-1 mutation is indicated in the gel profile as two missing bands at 35 kDa and 25 kDa as compared to Ler.

Figure 5. Arabidopsis seeds do not contain endogenous PHA. Immunoblotting showing that neither ssp-1 nor Ler contain PHA. The lane labeled bean contains total seed protein of Phaseolus vulgaris; PHA contains 200ng of purified commercially available PHA.
3.3 Strategy for testing ‘empty container’ hypothesis

When plants are transformed using *Agrobacterium*-mediated transformation method, the transgenes are inserted into the plant genome randomly at single or multiple sites. Thus, it is necessary to produce transformants with a single copy of the *PHA* gene in order to be able to compare the amount of PHA that is being produced on a per gene basis. It is also necessary to cross transformants with a single copy of *PHA* with wild type *Arabidopsis*, which would eliminate the *ssp-1* mutation and this would generate a copy of the *PHA* gene in a wild type genetic background in the same chromosome context and copy number. Therefore it is critical to compare levels of PHA in the homozygote *ssp-1* with the homozygote cross so as to determine if PHA is overexpressed to a greater extent in the ‘empty container’, *ssp-1*.

3.4 Plasmids used for transformation

Two DNA constructs containing the *PHA* gene were used to transform *ssp-1* plants in this project. One DNA construct utilized the vector pGPTV-HPT that carries the hygromycin resistance gene as a selectable marker under the control of the nopaline synthase promoter (Becker, et al. 1992). Lab clone 286 consists of this vector into which a reporter gene encoding phytohemagglutinin (PHA) was cloned (Figure 6A). Another DNA construct used was plasmid pEGAD-link that carries the *bar* gene as a selectable marker under the control of the cauliflower mosaic virus 35S (CaMV35S) promoter, along with the *PHA* gene was made (Figure 6B). The plasmid, pDR214 (Riggs et al., 1989), containing the *PHA* gene and the vector, pEGAD-link carried the *bar* gene were both digested with Hind III and EcoR I (Figure 7A). The 2.8 kb *PHA* gene and the 12.6 kb pEGAD-link fragments were purified from the gel (Figure 7B), ligated and transformed into *E.coli*, strain DH5α. This construct was called LC685. A PCR of the *PHA* gene was carried out to confirm the successful *E.coli* transformation (Figure 7C). LC685 plasmid DNA was then transformed into *Agrobacterium* strain GV3101, to generate lab clone 686. PCR of the *PHA* gene was performed to test for the presence of the *PHA* gene (Figure 7D).
A. pGPTV-HPT-PHA

![Diagram A](image)

B. pEGAD-link-PHA

![Diagram B](image)

**Figure 6.** Schematic representation of two DNA constructs used for the *Agrobacterium*-mediated transformation of *ssp-1*. A. Plasmid pGPTV-HPT-PHA (LC286) carried the hygromycin resistance gene as a selectable marker under the control of the NOS promoter. B. Plasmid pEGAD-link-PHA (LC685) carried the *bar* gene as a selectable marker under the control of the cauliflower mosaic virus 35S (CaMV35S) promoter along with the PHA gene. Restriction enzymes Hind III and EcoR I cut on the both sides of PHA gene respectively. LB: left border, RB: right border, 35S: CaMV35S promoter, *bar*: phosphinothricin acetyl transferase coding sequence, *Hgy*: hygromycin resistant gene. pNOS: nopaline synthase promoter.
Figure 7. Cloning of LC686. Panel A: Restriction enzyme digestion of plasmids pEGAD-link and pDR214. The reaction produced the vector about 12.6 kb in the lane of pEGAD-link, and about 2.8 kb PHA gene in the lane of pDR214. Panel B: Gel extracted vector pEGAD-link and PHA gene from the restriction enzyme digestion as shown in A. Panel C and D: PCR of PHA gene in LC685 and LC686, both of which showed a 776bp PHA fragment indicating the success of transformation.
3.5 Generation of ssp-1 transformants

ssp-1 plants were transformed with two DNA constructs, pGPTV-HPT-PHA and pEGAD-link-PHA, respectively. A total of 53 pots of ssp-1 plants (T0 generation) transformed with pGPTV-HPT-PHA were produced, 32 pots of which were subjected to first round selection on Hygromycin-containing medium. 74 resistant seedlings (T1 generation) with healthy green true leaves and well developed branched roots were selected from hygromycin plates. In addition, 15 pots of ssp-1 were transformed with pEGAD-link-PHA, screened on Basta selection medium, and 23 T1 seedlings were identified as resistant transformants. These selected plants were transferred into soil, allowed to self and produce T2 seeds.

3.6 Selection of single copy PHA homozygotes in the ssp-1 genetic background

Genomic DNA was isolated from leaf samples of all T1 transformants and negative control plants by the DNA mini-prep method. PCR analysis was carried out to detect the presence of the PHA transgenes in these lines. The expected 776 bp amplification band corresponding to the PHA gene was detected in most of the selected resistant plants (Figure 8). As expected, no amplification was observed when genomic DNA from untransformed ssp-1 plants was used as template. Of the 74 hygromycin resistant lines T1, all of them tested positive for the PHA gene. Similarly, all 23 T1 basta resistant lines were PCR positive.

3.6.1 Selection of single copy of PHA transformants in ssp-1

In order to get the transformants with single copy of PHA, southern blot analysis of T2 plants was performed. Because the southern blot analysis was not working at the beginning of my project, it was too late to evaluate the T1 plants. Genomic DNAs were extracted from each T2 transformant and analyzed by southern blotting. A 776 bp fragment of the PHA gene from PCR was used to make the probe (Figure 9). DNA extracted from the leaves of T2 plants was digested either with EcoRI, which cuts once at 1.5 kb from the 5’ end of the gene sequence in the PHA promoter sequence, or with HindIII, which cuts the construct once at 0.5 kb from the 3’ end of the gene sequence (Figure 9). Digested genomic DNA was resolved, blotted, and probed with the radiolabelled PHA DNA fragment, revealing the integration pattern and number of copies of the PHA gene inserted in the host genome.
Figure 8. **PCR analysis for PHA transgene in selected putative T1 transgenic plants.** PCR amplification of *PHA* gene showed the 776 bp *PHA* fragment in T1 plants (lanes 1, 2, 4, 5, 6, 7, 8, 9, 11, 12). Lanes: *P* Positive control (bean genomic DNA); *M* molecular size makers; *N* negative control *ssp-1* (T0) DNA.

Figure 9. **Schematic representation of the 2.8 kb PHA EcoR I/Hind III fragment.** The DNA probe for southern blot analysis is 776 bp as indicated. The restriction enzyme EcoR I cuts once in the *PHA* promoter about 1.5 kb from the 5’ end of the gene. Hind III cuts once in the *PHA* gene about 0.5 kb from the 3’ end of the gene. Xba I cuts once really close to the probe.
By using restriction enzymes that cut the sequence only once, each hybridizing band would represent a unique insertion site, as the other end of the hybridizing fragment would be delimited by an EcoR I site or a Hind III site present in the *ssp-1* genome. Thus, the number of hybridizing bands is indicative of the number of insertions into the genome and is a measure of copy number. Of the 16 lines transformed with pEGAD-link-*PHA* analyzed, none of them showed a single copy of *PHA* insertion. Figure 10 shows some selected examples of *PHA* transgenics with basta resistance or hygromycin resistance had multiple copies of *PHA*. PHA 8 and PHA 19 could be single copy lines since they were only digested with EcoR I; additional restriction digests with Hind III must be performed to confirm the result (Figure 10A). The hybridization band shown for PHA 12A and 12B were too faint to determine if PHA 12 has a single copy of *PHA* (Figure 10A). The same situation applies to PHA 17 and also when PHA 17 was digested with Hind III it seemed that there was another band at about 6 kb (Figure 10A). Figure 10B showed all transgenic lines had multiple copy of *PHA* except for PHA 7, which will be discussed below. Figure 10B had some background signals, because the southern blot did not working perfectly at the beginning of my project. In addition, the imaging film of Figure 10B was over exposed. Of the 37 lines transformed with pGPTV-HPT-*PHA* analyzed, DNA from 4 lines (PHA 1, 3, 5, 7) showed a single hybridization band indicating a single copy of *PHA* insertion (Figure 11). From the number and intensity of bands in the autoradiograph of Figure 11, it can be concluded that PHA 1, 3, 5, 7 lines have single gene insertions and PHA 2 and 4 have multiple gene insertions. In order to confirm the results, genomic DNAs of the same *PHA* transgenic lines plus PHA 6 were then digested with Hind III. PHA 1, 3, 5, 7 again only generated single copy hybridization banding patterns (Figure 11C). In PHA 3 (Figure 11C), there was a faint band above the intense band, which may be due to some undigested genomic DNAs. Thus, PHA3 was digested again with Hind III. From Figure 11B it can be seen that PHA 3 had only a single band of *PHA*. The size of the hybridizing band of PHA 6 was less than 3 kb, so it may be that the *PHA* gene was not fully transferred into the plant genome or that some rearrangement has taken place. Other than that there was also a faint band around 3 kb in PHA 6 which was not clear, but when the film was exposed for more days it could be seen. We did not consider PHA 6 as a single copy *PHA* transgenic. When PHA 2 was digested with Hind III it showed an obvious single band of about 5 kb (Figure 11C), but it also showed a smaller band under the band around 4 kb (Figure 11C), and a band of more than 7 kb which probably was the undigested genomic DNA (Figure 11A). Therefore, PHA 2 carries at least two copies of PHA at two distinct integration sites. *ssp-1*
was the negative control which did not show any bands (Figure 11C). A summary of PHA gene integration pattern in the 4 transgenic lines is presented in Table 2.

3.6.2 Confirmation of PHA expression in single locus transformants

Seed protein extracts were made from ssp-1 seeds and T3 seeds of the seven transgenic lines. Although the crude protein was not quantified, as based on the Coomassie blue stained gel approximately the same amount of protein of each sample was used. The samples were run on a SDS-PAGE gel, along with some purified PHA. This gel was used in western blotting which was performed to confirm the expression of PHA (Figure 12). The results showed that PHA was expressed in all 4 lines (PHA 1, 3, 5, and 7) giving rise to strong signals at around 30 kDa. ssp-1 was the negative control which did not contain any PHA, and as expected, no signal was observed. There was some variability in expression. PHA 4, containing multiple copies of the PHA transgene, gave the most intense signal. PHA 1, 5, 7 gave the approximately same amount of signal, which was stronger than that of PHA 3. PHA 6 was expressed poorly, which likely indicates that the truncated band observed in the southern blotting experiments was a rearranged gene (Figure 11C). I observed some anti-PHA reactive proteins of low molecular weight which are likely derived from specific processing events that normally occur in dicot seeds. It is unlikely that these are due to non-specific proteolysis, as distinct bands of the same size are observed for all samples, and the denaturing extraction buffer used would limit proteolytic activity.
A.

```
<table>
<thead>
<tr>
<th>bar</th>
<th>Hygromycin</th>
</tr>
</thead>
<tbody>
<tr>
<td>EcoR I</td>
<td>Hind III</td>
</tr>
</tbody>
</table>
```

```
PHA 8  PHA 9  PHA 10  PHA 11  PHA 12A  PHA 13A  PHA 14A  PHA 15A  PHA 16A  PHA 17A  PHA 18  PHA 19  PHA 20  PHA 12B  PHA 13B  PHA 14B  PHA 15B  PHA 16B  PHA 17B
```

B.

```
<table>
<thead>
<tr>
<th>Hygromycin</th>
<th>bar</th>
<th>Hygromycin</th>
</tr>
</thead>
<tbody>
<tr>
<td>EcoR I</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

```
```
Figure 10. Selected southern blot analysis of T$\textsubscript{2}$ *PHA* transformants with either basta resistance or hygromycin resistance. Genomic DNA was restricted with EcoR I or Hind III and hybridized with a radiolabelled 776 bp *PHA* fragment. Transformed plants PHA 13, 14, 15, 16, 17 have multiple hybridization bands upon digestion with either EcoR I or Hind III; PHA 9, 10, 11, 18, 20, 21, 22, 23, 24 have multiple hybridization bands upon restriction digested with EcoR I; PHA 8, 12 and 19 could have single copy of *PHA*. PHA 12A and PHA 12B are different plants from the same parent transgenic line, so as 7A/B, 13A/B, 14A/B, 15A/B, 16A/B, 17A/B, 21A/B, 22A/B, 23A/B and 24A/B.
Figure 11. Southern blot analysis of T$_2$ PHA transformants. Genomic DNA was restricted with EcoR I or Hind III and hybridized with a radiolabelled 776 bp PHA fragment. Transformed plants PHA 1, 3, 5 and 7 showed a single hybridization band upon digestion with either EcoR I or Hind III; PHA 2 and 6 showed multiple hybridization bands; ssp-1 was the negative control; PHA 1A, PHA 1B and PHA 1C are different plants from the same parent transgenic line. Thus, lines 1, 3, 5, and 7 are single locus insertion lines.
Table 2. Summary of southern blot analysis of 4 single copy PHA transformants

<table>
<thead>
<tr>
<th>T1</th>
<th>T2 individual</th>
<th>RI bands (kb)</th>
<th>HIII bands (kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHA1 &lt;br&gt;A</td>
<td>B</td>
<td>4.3</td>
<td>3.4</td>
</tr>
<tr>
<td>PHA1 &lt;br&gt;C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHA3 &lt;br&gt;A</td>
<td>B</td>
<td>6</td>
<td>4.7</td>
</tr>
<tr>
<td>PHA3 &lt;br&gt;C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHA5 &lt;br&gt;A</td>
<td>B</td>
<td>5.5</td>
<td>4.5</td>
</tr>
<tr>
<td>PHA5 &lt;br&gt;A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHA7 &lt;br&gt;A</td>
<td>B</td>
<td>7.5</td>
<td>6.5</td>
</tr>
</tbody>
</table>

RI bands (kb) and HIII bands (kb) represent the sizes of the fragments of EcoR I and Hind III digested DNA hybridizing with the PHA probe, respectively.

**Figure 12. Confirmation of PHA expression in PHA transformants.** Immunoblot of PHA transformants and ssp-1 along with 100ng, 150ng and 200ng of purified PHA. PHA was expressed well in most of the transformants at around 30 kDa except for PHA 6. ssp-1 was the negative control.
3.6.3 Screening for PHA homozygotes in ssp-1

Southern blotting experiments indicated that transformants 1, 3, 5, and 7 were single locus transformants which express the PHA transgene at relatively high levels. Thus, these plants were selected for further analyses as the starting lines for testing the ‘empty container’ hypothesis. The strategy for generating the appropriate lines is shown in Figure 13.

The first step in this process is to generate homozygous lines. T2 seeds of single copy PHA transformants were first screened on hygromycin plates. More than six resistant seedlings of each transformant line were selected and transferred into soil to allow them to self-cross to increase the probability of getting homozygotes instead of the original strategy showing in Figure 13. Moreover, I need a lot of T2 plants for southern blot analysis. These resistant seedlings formed dark green true leaves and well-developed branched roots. The T3 seeds were collected and about 100-160 seeds were screened on selection plates. For heterozygous T2 plants, the T3 seeds showed both of the resistant and non-resistant seedlings on selection plates, so the number of hygromycin resistant seedlings and hygromycin susceptible seedlings of these T3 seedlings were analyzed in Chi-square test (Table 3). Chi-square test showed that $X^2$ values of most samples are smaller than the critical $X^2$ value, which indicated that the T3 seeds showed the inheritance of the PHA gene in a 3:1 ratio of Mendelian manner (Table 3). This also confirmed the southern blot analysis that PHA 1, 3, 5, 7 have single copy of PHA (Table 3) since if a transgenic have more than one copy of PHA, the segregation ratio will be bigger than 3:1 ratio. For homozygous T2 plants, the T3 seeds produced 100% resistant seedlings. PHA homozygotes were identified from all the four lines of single copy PHA transgenics (Table 4). Homozygote T3 seedlings were selected and transferred to soil to grow, and crossed with wild type plants. Homozygote T4 seeds were harvested for subsequent analysis (Table 4). These results along with PCR and western blot analysis confirmed that the PHA gene was efficiently inherited and transcribed in the progeny of the transgenic plants.
**T₀ ssp.1 + LC286 or LC686**

Selected on medium with antibiotics

**T₁ plants (PHA-1, PHA-2…PHA-27) are analyzed in PCR to verify PHA gene insertion.**

**cru3cru3 PHApha**

**T₂ seeds were harvested and screen on selection medium.**
- 3 (resistant): 1 (non-resistant)
- 6 T₂ seedlings of each PHA line were transferred into soil.

**T₂ leaf genomic DNAs were used in southern blot analysis to identify single copy of PHA transformants.**

**T₃ seeds were harvested from single copy of PHA transformants T₂ and screened on selection medium to get PHA homozygote. A homozygote T₃ will give 100% hygromycin resistant seedlings.**

**PHA homozygote T₃**

**Cross**

**Pollen cru3cru3 PHApha**

**Wild Type**

**F₁ seeds from the cross were screened on selection medium to get F₁ (CRU3cru3 PHApha) seedlings, which were also analyzed in PCR to verify the presence of PHA gene and the success of the cross.**
F₂ seeds were harvested and screened on selection medium.  
12 (resistant): 4 (non-resistant)  
32 seedlings F₂ of each backcross line were transferred into soil to get F₃ seeds.

F₃ seeds were used in segregation analysis to get PHA homozygote which will give 100% resistant seedlings F₃ (PHAPHA __).  

SDS-PAGE of seed proteins from F₃ was used to identify PHA homozygote in Ler background (PHAPHA CRU3__).  

Sequencing CRU3 fragment containing ssp-1 mutation site of F₃ plants to get PHA homozygote in homozygous CRU3 background (PHAPHA CRU3CRU3).

Figure 13. Strategy for screening for PHA homozygotes in both ssp-1 and Ler genetic backgrounds.
<table>
<thead>
<tr>
<th>T3 seeds</th>
<th>Gene copy No.</th>
<th>HygR seedlings No.</th>
<th>HygS seedlings No.</th>
<th>Expected ratio</th>
<th>X² value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHA 1-11</td>
<td>1</td>
<td>118</td>
<td>23</td>
<td>3:1</td>
<td>5.676</td>
</tr>
<tr>
<td>PHA 1-10</td>
<td>1</td>
<td>92</td>
<td>14</td>
<td>3:1</td>
<td>7.862</td>
</tr>
<tr>
<td>PHA 3-4</td>
<td>1</td>
<td>107</td>
<td>30</td>
<td>3:1</td>
<td>0.703</td>
</tr>
<tr>
<td>PHA 3-9</td>
<td>1</td>
<td>105</td>
<td>28</td>
<td>3:1</td>
<td>1.105</td>
</tr>
<tr>
<td>PHA 5-6</td>
<td>1</td>
<td>84</td>
<td>36</td>
<td>3:1</td>
<td>1.600</td>
</tr>
<tr>
<td>PHA 5-9</td>
<td>1</td>
<td>129</td>
<td>26</td>
<td>3:1</td>
<td>5.594</td>
</tr>
<tr>
<td>PHA 7-12</td>
<td>1</td>
<td>126</td>
<td>32</td>
<td>3:1</td>
<td>1.899</td>
</tr>
<tr>
<td>PHA 7-19</td>
<td>1</td>
<td>120</td>
<td>37</td>
<td>3:1</td>
<td>0.172</td>
</tr>
</tbody>
</table>

Table 3. Chi-square test of T3 seeds from heterozygous single copy PHA transgenic T2 plants on hygromycin selection medium. The heterozygotes plants were supposed to show a Mendelian segregation ratio (3:1). P-value = 0.05. The critical X² value = 3.841. HygR = hygromycin resistance, HygS = susceptible to hygromycin.

Table 4. Summary of PHA homozygotes in the ssp-1 background identified from segregation analysis of T3 seeds

<table>
<thead>
<tr>
<th>PHA homozygote in ssp-1 (T3)</th>
<th>PHA homozygote in ssp-1 (T4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHA 1</td>
<td>PHA 1-3, PHA 1-8, PHA 1-12</td>
</tr>
<tr>
<td>PHA 3</td>
<td>PHA 3-8</td>
</tr>
<tr>
<td>PHA 5</td>
<td>PHA 5-1, PHA 5-3, PHA 5-5,</td>
</tr>
<tr>
<td></td>
<td>PHA 5-15</td>
</tr>
<tr>
<td>PHA 7</td>
<td>PHA 7-1, PHA 7-9, PHA 7-10</td>
</tr>
</tbody>
</table>
3.7 Crossing of PHA homozygotes into the Ler genetic background

The four single locus PHA homozygous lines (PHA 1, 3, 5, 7) were crossed to wild type Arabidopsis, Ler, to get PHA homozygotes into the Ler genetic background (Figure 13). When the F₁ seeds were collected, they were screened on hygromycin selection medium. The resistant seedlings were transferred into soil and allowed to grow to maturity to get F₂ seeds. PCR of extracted leaf DNA from F₁ plants was carried out to verify the presence of the PHA gene and the success of the backcrosses (Figure 14). F₂ seeds were plated on the selection medium and 32 of the healthiest seedlings of each cross line were transferred into soil and allowed to self cross to get F₃ seeds. The F₃ seeds were placed on selection media and the PHA homozygote plants F₃ giving 100% resistant seedlings were selected.

Although the homozygous state of the F₃ PHA homozygotes fixed one of the two genes being tracked (PHA), the F₃ PHA homozygote could be in either in the Ler or ssp-1 genetic background (either CRU3/CRU3, CRU3/cru3, or cru3/cru3). My aim was to bring the PHA locus into the Ler background, so the following method was devised to select for homozygous wild type CRU3 in these lines. Seed proteins extracts were made from F₃ seeds and seeds of ssp-1. These seed extracts were subjected SDS-PAGE and Coomassie staining to score for the missing CRU3 bands; such a pattern is indicative of the ssp-1 (cru3/cru3 null) genotype (Figure 15). PHA homozygotes with the ssp-1 mutation showed two missing bands around 35 kDa and 25kDa as shown in Figure 15. Table 5 shows the number of PHA homozygotes in Ler background of each backcross line.
Figure 14. PCR analysis of *PHA* gene insertion in *PHA* backcross F$_1$ lines. All four PHA backcrosses 1, 3, 5, 7 showed the *PHA* fragment around 776 bp similar to the PHA transgenic PHA 1. M: molecular weight marker.
Figure 15. SDS-PAGE of seed storage proteins from *PHA* homozygote backcrosses. *PHA* homozygotes in the *ssp-l* background do not contain the two CRU3 bands at 35 kDa and 25 kDa.

Table 5. Summary of *PHA* homozygote in Ler background

<table>
<thead>
<tr>
<th></th>
<th><em>PHA</em> homozygote backcross in Ler (F₃)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHA 1 backcross</td>
<td>1, 9, 11, 14, 16, 18, 20, 26, 27</td>
</tr>
<tr>
<td>PHA 3 backcross</td>
<td>3, 15, 20, 24, 25</td>
</tr>
<tr>
<td>PHA 5 backcross</td>
<td>2, 9, 12, 13, 14, 18, 19, 20, 22, 24, 25, 27, 30</td>
</tr>
<tr>
<td>PHA 7 backcross</td>
<td>3, 8, 13, 17, 20, 24, 25</td>
</tr>
</tbody>
</table>
Using this strategy I was able to discard some of the segregants. The segregants that give rise to the wild type CRU3 polypeptide pattern could be either homozygous or heterozygous for CRU3. The only difference between the Ler and ssp-1 in the CRU3 gene was a single point mutation where an adenine replaces a guanine (G to A) in ssp-1 (Figure 16). Sequencing of CRU3 gene fragment containing the mutation site of all the backcross lines was used to distinguish between homozygotes and heterozygotes for CRU3. PCR amplification of CRU3 fragments was performed to prepare templates for sequencing (Figure 16 and 17). As shown in Figure 17, PHA 1 F2 segregants 1 and 9; PHA 3 segregants 15 and 25; PHA 5 segregants 14 and 20; and PHA 7 segregants 3, 17, and 23 showed 837 bp CRU3 fragment. These CRU3 fragments were purified from the gel for sequencing. For homozygous CRU3 segregants, the wild type sequence should be obtained, whereas for CRU3/cru3 heterozygotes, there should be an equal signal for both the wild type and mutant nucleotide at the same position. Because the primer used for sequencing was complimentary to the mRNA stand, the position in question would generate a ‘C’ instead of a ‘G’ (sense strand), and heterozygotes would have both a ‘C’ and a ‘T’ at this position. In total I sequenced 28 templates for the four transformants. I identified nine CRU3 homozygotes including PHA 1 backcross 1, 9, PHA 3 backcross 15, 25, PHA 5 backcross 14, 20, and PHA 7 backcross 3, 17, 23, which showed one ‘C’ peak at the 4873bp position (Figure 18). CRU3 heterozygotes gave two equally intense peaks (C and T) at the mutation site, for example PHA 1 backcross 20 (Figure 18A). PHA homozygotes in the homozygote Ler (CRU3/CRU3) background are summarized in Table 6.

3.8 Comparison of PHA levels in ssp-1 versus Ler genetic backgrounds

Immunoblot analysis was carried out to compare the amount of PHA protein expressed in seeds of PHA homozygotes in both the ssp-1 and Ler genetic backgrounds. Total soluble proteins were extracted from 4 lines of single copy PHA homozygotes (PHA 1-12-1, PHA 3-8-1, PHA 5-1-1 and PHA 7-9-1) and 4 lines of these PHA homozygotes backcrossed to Ler (PHA 1 backcross 1, PHA 3 backcross 25, PHA 5 backcross 14 and PHA 7 backcross 23). Protein concentrations were determined and equal amounts of total seed proteins were subjected to SDS-PAGE and immunoblotting (Figure 19). In addition, aliquots of purified PHA (14 ng, 22 ng and 29 ng) were also used to gauge the relative expression levels of the samples. Each PHA transgenic line was run with 500 ng of total soluble protein in duplicates. Following immunoblotting, the
autoradiograms were scanned and Image-J (NIH) was used to evaluate the intensity of the immunoreactive signals. Most lanes of samples had several bands appearing at lower molecular weights than the band for purified PHA. These were processing products of PHA and were still indicative of PHA expression. Because the relative amounts of the processing products were similar to the uncleaved PHA, I decided to use only the uncleaved PHA to compare the samples. The levels of all PHA around 30kDa in each lane were measured through comparison to the signal from 22 ng of PHA standard. The level of PHA in each sample was from the average of the PHA levels of the duplicate samples (Table 7).

All the transgenic plants expressed variable levels of the PHA protein in seeds. The PHA levels ranged from 4.52% to 7.7% of total soluble protein in the four transgenic ssp-1 lines, whereas the levels in the backcross ranged from 3.13% to 7.26% of total soluble protein (Table 7). PHA 1, 3, 5 and 7 revealed higher expression of PHA than their corresponding backcrosses (Figure 20). For PHA 5, PHA levels were increased 5.87% compared to its backcross. PHA 7 only showed a 0.31% increase in PHA relative to its backcross. However, PHA 1 and 3 showed a significant increase of PHA expression, 30.33% and 44.18%, respectively, in comparison to their backcrosses (Table 7 and Figure 20). Therefore, there is some evidence to support the validity of the ‘empty container’ hypothesis. Figure 20 also shows that the immunoreactive PHA band in all four backcross lines have two bands really close together which can be easily seen in the original film, but still can be seen from PHA 3 backcross 25. The upper PHA band could be some modified PHA protein containing only a few more amino acids, perhaps as a result of aberrant processing events.
Figure 16. Schematic representation of PCR of *CRU3* fragment from 4318 bp to 5154 bp and partial DNA and amino acid sequence of the wild type *CRU3* gene. The black arrow indicates the mutation of *ssp-1* at position 4873 bp where an adenine replaces a guanine (G to A) in *ssp-1*. This resulted in altering a tryptophan codon to a premature stop codon. The blue arrows indicate the two primers used in amplification of *CRU3* fragments which were used as the templates for sequencing and the *CRU3*-1 Backward primer was the primer used for sequencing.

Figure 17. PCR amplification of *CRU3* fragments containing the mutation site.
Figure 18. Sequencing chromatograms of 9 CRU3 homozygotes and one CRU3 heterozygote. CRU3 heterozygotes gave two equally intense peaks (C and T) on the mutation site such as PHA 1 backcross 20. CRU3 homozygotes showed one peak (C). The black arrow indicates the mutation site.

Table 6. Summary of PHA homozygotes in homozygote Ler genetic background

<table>
<thead>
<tr>
<th>Backcross</th>
<th>Backcross homozygote lines</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHA 1 backcross</td>
<td>PHA 1 backcross 1, 9</td>
</tr>
<tr>
<td>PHA 3 backcross</td>
<td>PHA 3 backcross 15, 25</td>
</tr>
<tr>
<td>PHA 5 backcross</td>
<td>PHA 5 backcross 14, 20</td>
</tr>
<tr>
<td>PHA 7 backcross</td>
<td>PHA 7 backcross 3, 17, 23</td>
</tr>
</tbody>
</table>
Figure 19. Comparison of PHA expression in single copy of PHA transformants and their relative backcross. Immunoblot of PHA transformants and their relative backcrosses along with 14ng, 22ng and 29ng of purified PHA. Each PHA transgenic line was run with 500 ng of total soluble protein in duplicates. PHA was expressed well in all of the transformants at around 30 kDa.
<table>
<thead>
<tr>
<th>PHA 1</th>
<th>PHA 3</th>
<th>PHA 5</th>
<th>PHA 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>ssp-1</td>
<td>Ler</td>
<td>ssp-1</td>
<td>Ler</td>
</tr>
<tr>
<td>1.564± 0.001</td>
<td>1.2± 0.030</td>
<td>1.028± 0.046</td>
<td>0.713± 0.046</td>
</tr>
</tbody>
</table>

The relative intensity value of PHA as compared to 22ng PHA standard.

PHA of total soluble protein (%)

<table>
<thead>
<tr>
<th></th>
<th>ssp-1</th>
<th>Ler</th>
<th>ssp-1</th>
<th>Ler</th>
<th>ssp-1</th>
<th>Ler</th>
<th>ssp-1</th>
<th>Ler</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.88%</td>
<td>5.28%</td>
<td>4.52%</td>
<td>3.13%</td>
<td>7.70%</td>
<td>7.26%</td>
<td>7.23%</td>
<td>7.21%</td>
<td></td>
</tr>
</tbody>
</table>

Increased PHA level (%) in ssp-1 as compared to Ler

Table 7. PHA expression level in both ssp-1 and Ler genetic backgrounds. The relative intensity values of PHA represent mean ± standard deviation.

PHA Expression Level Comparison in ssp-1 and Ler

Figure 20. Comparison of PHA expression level in both ssp-1 and Ler genetic backgrounds. The amount of PHA was calculated from immunoblot signals analyzed by Image-J (NIH). The values represent mean ± standard deviation. The mean value of PHA 1 is 1.564 ± 0.001, PHA 3 is 1.028 ± 0.046, PHA 5 is 1.750 ± 0.031, and PHA 7 is 1.643 ± 0.023. Whereas, the mean value of PHA 1 backcross is 1.200 ± 0.030, PHA 3 backcross is 0.713 ± 0.046, PHA 5 backcross is 1.653 ± 0.03, PHA 7 backcross is 1.638 ± 0.011. P-values for T-test pair wise comparisons: for PHA1 p<0.001, for PHA 3 p<0.001, for PHA 5 p<0.05, for PHA 7 p>0.05.
4 Discussion

Transgenic plants as bioreactors have been studied extensively. By modification of endogenous genes or introducing novel genes of interest, not only improvement of crop quality and nutritional value can be achieved such as enhancing essential amino acid content, increasing lipid and starch level, but also plants can be used as a robust bioreactor to produce various valuable materials such as biopharmaceuticals (e.g. vaccines and mammalian antibodies), biodegradable plastics, and nutritional supplements for human health. Likewise, seeds have the potential to provide virtually unlimited amounts and types of recombinant proteins for use in human health and bioscience. The most important reasons to use a plant system are its low cost input compared to animal and microorganism systems, rapid large-scale production, and lack of animal disease. In addition, seeds are easy for storage, delivery, and can be purified efficiently using established processing methods. Overexpression of foreign or recombinant proteins in plants is limited by many factors such as the promoter, transgene insertion site and copy number, and foreign protein modification and/or degradation. Moreover, it can be also due to the high levels of endogenous proteins in the seed and a bias toward accumulation of natural reserves. If the seed can be emptied of some endogenous proteins, the void could be filled by foreign or genetically engineered proteins. This idea has become known as the ‘empty container’ hypothesis.

To explore this hypothesis, a mutant must be found which does not fill its seeds. Such a mutant was identified and called *ssp-1*, which is deficient in one of the 12S cruciferin proteins, CRU3. *ssp-1* contains 20-27% less seed storage proteins than wild type and is missing two bands in the gel profile that correspond to the α and β subunits of 12S crucifers. Thus, *ssp-1* has been considered as the ‘empty container’. The *ssp-1* mutation is caused by G-A transition in the *CRU3* gene located on the distal arm of chromosome 4 near cer2 (74.5cM) (Riggs, unpublished results). Based on a previous study there was no significant difference in the levels of starch within the *ssp-1* seed as compared to wild type (Lee, 2000). In addition, the lipid composition was relatively unchanged. As discussed before, alteration of one storage component would result in a compensatory change in other seed storage components. It also has been studied that if there is a decrease in one type of seed storage protein, other protein components will be increased to compensate resulting in unchanged or increased total protein levels. Recently, the overall AA composition of genetically related lines of common beans defective in seed storage proteins, phaseolin, phytohemagglutinin, and arcelin individually were evaluated (Taylor et al., 2008).
They found that several changes in the free AA content in these bean lines including a reduction of S-methyl-cysteine and γ-glutamyl-S-methyl-cysteine (non-protein AAs that can not substitute for Met or Cys in the diet). However, the absence of major seed storage proteins (phaseolin, phytohemagglutinin, and arcelin) was compensated by increases in sulphur-containing AA (Met and Cys) contents up to 40% as compared to beans with only high phytohemagglutinin, and arcelin contents (Taylor et al., 2008). Also, overexpression of the milk protein α-lactalbumin gene resulted in significantly increased in content of Lys levels in transgenic maize (Bicar, et. al, 2008). They found total protein content in endosperm from transgenic seeds was not significantly different from total protein content in endosperm from non-transgenic seeds. However, the content of some other AAs was reduced. These results provide evidence that there must be a regulation of seed storage protein, which might be in the form of a sensor which can detect the levels of macromolecules in the seed. If abnormal levels of seed storage components were detected, the sensor would regulate expression levels of other seed components to compensate. The regulation mechanism of a sensor may lead to that foreign transgenes are not expressed at high levels since the amount of each seed components are somehow set and modulated by the sensor if any components are changed. In addition, the transgene mRNA would be in competition with endogenous mRNAs and translate a small percentage of protein to the total protein which is measure by the sensor. Also, because they are foreign mRNAs, they may be prone to degradation. That’s probably why the transgenic seed fill normally and low levels of foreign protein, which could be considered as normal.

The ssp-1 mutant is abnormal since the cruciferins do not accumulate to normal levels but there seems to be no obvious increases in the other proteins or in the lipids and starch levels. Therefore, ssp-1 is a non-compensating mutant which might have bypassed the sensor regulation. A similar mutant was characterized in soybean. A soybean mutant was found to be defective in 7S globulin subunits controlled by a single recessive gene (Hayashi et al., 1998). The mutation is not caused by a defect in the structural genes of the 7S globulin subunits, but happens at the mRNA level, either at the level of transcription or it may be due to rapid degradation of mRNA. However, transient expression of β-glucuronidase from the 7S promoter of chimeric genes in seeds and gel mobility shift assays using soybean embryo extracts both revealed that transcription factors were present. The gene was not cloned or sequenced, so the exact mechanism of the mutation is not
known, but this mutant may also provide an ‘empty container’ in the improvement of the quality of seed proteins in soybean.

The aim of this project is to test if foreign proteins can be overexpressed in ssp-1, a mutant defective in seed filling. To test the feasibility of ssp-1 as the ‘empty container’, ssp-1 plants were transformed with a DNA construct using a seed specific gene from bean called phytohemagglutinin (PHA) as the reporter gene. This project has three objectives: (1) to identify single locus PHA homozygotes in ssp-1; (2) to cross the homozygous PHA lines with wild type plants (Ler), and to identify PHA homozygotes in a homozygous Ler genetic background; (3) to compare the PHA expression level in both ssp-1 and Ler genetic backgrounds.

Many transgenic studies have demonstrated variability in the level of expression of foreign proteins in seeds. PHA-L was transformed into transgenic tobacco under the control its own promoter and the expression levels of PHA were low from 0.1% to 0.2% of total seed protein (Voelker et al., 1989). In another study, the pea lectin gene was expressed from 0.2% to 0.9% of the total seed protein content in transgenic tobacco (de Pater et al., 1996). Perhaps the low expression level of the transgene is due to the use of the gene’s own promoter or the seeds cannot tolerate high expression levels of foreign proteins. However, a high expression level of a transgene was found in some studies using a heterologous promoter. For example, sunflower seed albumin (SSA) is rich in Met and Cys, which was transferred to chickpea (Cicer srietinum L.) under a seed-specific promoter from a pea vicilin gene (Chiaiese et al., 2004). The first generation of transgenic seeds accumulated 6-12% SSA of total seed protein. The amount of SSA depended on the zygosity of the transgene loci. In a homozygous transgenic of the third generation SSA constitutes about 3% of total seed protein. However, the total seed Cys concentrations were slightly lower and the total seed sulphur amino acids were higher for the transgenic seeds than for the non-transgenic control. They found sulfur from endogenous proteins was reallocated to the heterologous sulfur-rich protein. Similarly, a milk protein, α-lactalbumin, which is rich in Lys, was expressed in transgenic maize under the control of the maize 27 kDa γ-zein promoter (Bicar, et. al, 2008). It was demonstrated that the protein was properly processed and was found in the endosperm of maize. The lysine content of transgenic seeds was significantly increased to 29%-47% compared with non-transgenic seeds, therefore increasing its nutritional value. However, the total protein content in endosperm from transgenic seeds was not significantly different from total protein content in endosperm from non-transgenic
seeds. Lastly, the human growth hormone (hGH) is a protein that induces growth and cell reproduction. The hGH gene was transformed into soybean under the control of alpha prime subunit of β-conglycinin tissue-specific promoter from soybean which gave rise to an expression levels of bioactive hGH up to 2.9% of the total soluble seed protein content (about 9 g/kg of dry seeds) (Cunha et al., 2010). Based on these studies, it is shown that the expression of foreign genes in plants is variable, and may depend on factors such as the type of host plant, the type of promoter used and the source of transgene.

A heterologous sb401 gene from potato encoding a pollen-specific protein with high lysine content was transferred and expressed in maize under a maize seed-specific expression storage protein promoter (P19z), resulting in a 16.1-54.8% accumulation of lysine content, and total protein content increased by 11.6% to 39% as compared to the non-transgenic maize control (Yu et al., 2004). Such a high level of transgene expression in the seed may suggest that an ‘empty container’ may not be needed as it is possible to achieve high levels of expression without further modification of any other seed protein fractions. However, the study also showed that after six continuous generations of self-pollination, the Lys protein content for transgenic seeds ranged from 14.0% to 17.25% whereas non-transgenic seeds averaged 10.8%. The transgenic plants started with two or more copies. After several generations the Lys protein content was decreased probably because some transgenes may have lost function due to chromosome crossing over or after chromosome recombination the transgenes are moved to some locations which are not transcriptionally active. It would be useful if they used homozygous transgenic plants with a single copy of the transgene to track the heredity of Lys expression, as protein content could be vary between seeds with different copies of the transgene. In another transgenic study, Arcelin-5 was transformed into Tepary bean (Phaseolus acutifolius A. Gray) and Arabidopsis, and resulted in an accumulation of arcelin-5 to 15% and 25% of the total protein, respectively (Gossen et al., 1999a). Arcelin-5 is an abundant seed storage protein in some wild type common bean genotypes, and is genetically closely linked with and related to PHA. Such a high level may also indicate that sensor systems do not exist and that it is not necessary to generate an ‘empty container’. However, Gossen et al. (1999a) used only 1-2 seeds per assay for quantification of bean protein which may cause some errors. It has been shown that analysis of small seed samples or individual seeds of soybean may not reflect the real protein levels for an
individual plant since the protein content has been shown to vary between seeds from different parts of plants (Escalante and Wilcox, 1993).

It also has been reported that the nutritional quality of seed proteins can be improved through alteration of the storage protein composition. This can be achieved by suppressing expression of a gene encoding a major endogenous protein and therefore increasing the levels of other endogenous or foreign seed storage protein of interest. Hansen et al. (2007) examined the expression of antisense C-hordein gene in transgenic barley. It turned out that the antisense lines had a lower C-hordein protein but were increased in the sulfur-rich B/γ- and D-hordeins significantly. They demonstrated that antisense suppression of one major seed storage protein may up-regulate other proteins in the seeds. Similar approaches were used to improve the nutritional quality of rice, since rice endosperm contains a lot of Met-rich prolamine. Maruta et al. (2001) produced transgenic rice with expression of a glutelin A antisense gene. Glutelin is the most abundant storage protein in rice. Transgenic lines showed 20% to 40% reduction in the glutelin content of seeds and an increase in the prolamine content. The total protein level was not significantly different from non-transgenic rice. However, glutinin was decreased variably, depending on the copy number of antisense gene. The two-copy integrated line was less effective in decreasing glutelin content than the four-copy integrated line. Therefore, the amount of protein that can be emptied to generate the ‘empty space’ is not consistent among seeds or different generations of seeds due to changes in copy number of transgene after chromosome crossing over. Moreover, they also discussed that there must be a mechanism by which the total protein content is maintained. This means not all seeds can be genetically modified to generate the permanent ‘empty space’ for overexpression of foreign genes since the seeds have already been full with other endogenous protein other than the suppressed protein before induction of new genes unless both of the antisense and foreign genes were introduced into plants at the same time. For example, Goosen et al. (1999b) transformed Arabidopsis with the common bean arcelin gene which also carried a 2S albumin antisense gene. This research was to increase the synthesis and accumulation of arcelin seed proteins by repressing the endogenous albumin genes. This study would also seem to test the empty container hypothesis. The result showed that the arcelin-5 was enhanced to more than 24% of the total seed protein content than the Arabidopsis seeds transformed with only arcelin-5. The expression levels between transgenic lines were verified and strongly corresponded with the transgene copy number. Therefore, the comparison
must also consider the transgene copy number since it might affect the expression level of the transgene. However, these findings also suggest that the seed has the biosynthetic capacity to direct its energy towards the production of a foreign protein or other endogenous proteins when an endogenous protein is repressed. This is supported by Hansen et al. (2007) and Maruta et al. (2001), which were previously discussed. Kohno-Murase et al. (1994) also transformed rapeseed with an antisense gene for napin driven by the 5'-flanking region of the rapeseed napin gene. They found that transgenic seeds contained reduced amounts of napin or accumulated no napin. Seeds deficient in napin accumulated 1.4 to 1.5 times more cruciferin (one of the major seed proteins) than non-transgenic seeds. As compared to Kohno-Murase et al. (1994), Goosen et al. (1999b) did not find a significant increased in cruciferin. Instead, when the resources for the production of napin were freed, the available resources were used in the production of arcelin as opposed to cruciferin. Maybe the foreign arcelin gene out competed with the endogenous cruciferin genes for transcription factors and machinery. Evidence for this is that overexpression of a novel arcelin gene in common bean resulted in a 50% decrease in endogenous phaseolin levels in the seeds, which indicated that arcelin may have a more active promoter than the endogenous phaseolin gene (Romero-Andreas et al., 1986). Although antisense technology could be a great utility in high expression of transgenes, antisense suppression of the biosynthesis of endogenous proteins may have some potential pleiotropic effects on interconnected pathways and affect the agronomic quality traits such as seed weight, seed fertility and morphological irregularities (Hansen et al., 2007). For example, a maize mutant, opaque 2 (o2), which dramatically reduces the level of the γ-zeins, was shown to contain twice the normal levels of lysine and tryptophan, and causes a soft endosperm texture that limits its commercial use (Wu et al., 2010).

So far it has been shown that the seed has the capacity to highly express a foreign protein when a major endogenous storage protein was deficient. What remains is to demonstrate whether or not a seed can be genetically modified in order to increase agricultural value by increases in traits such as yield, quality and nutritional value. If so, the seed also has the potential to contain and express genes for pharmaceutically or industrially interesting products.

The first objective of my project was to generate PHA transformants in ssp-1 (the empty container) background and then identify single copy PHA homozygotes in ssp-1. Arcelin and PHA are related and so perhaps it might be expected to produce a line with the same level of
expression seen in the study by Gossen et al. (1999a). Since the copy number of the transgene will affect its expression levels between transgenic lines (Mandal and Mandal, 2000), it will be useful to screen for a homozygous transgenic line with single copy of PHA. This will simplify genetic analysis and quantification of expression levels of the inserted gene since the effects of multiple copy number are unknown. It will be more comparable for investigating if there is difference in expression among transgenics containing the same copy number. 74 independent transformants (T₁) of hygromycin resistance and 23 independent transformants (T₁) conferring basta resistance were identified and screened to obtain the desired genotype. The transgenic lines developed normally and are fertile and the T₁ seeds retain normal germination capacity compared with wild type seeds. In order to compare the amount of PHA produced on a per gene basis, it is necessary to identify transformants with single copy of transgene. Southern blot analysis of T₂ plants was performed. Of 37 lines of PHA transformants with hygromycin resistance analyzed, 4 independent PHA transgenic lines (PHA 1, 3, 5, 7) were identified to contain only one insertion of the PHA gene (Figure 11A and B). All four transgenic lines demonstrated variable levels of PHA but which indicated PHA gene was expressed well in plants (Figure 12). However, there were some proteins of small size which could be some processing products of PHA. This was also seen in the study by Riggs et al. (1989), where tobacco was transformed with PHA and immunoblot analysis also showed the same protein profile pattern of PHA. The four single locus PHA lines were self-pollinated and T₃ seeds were used in the segregation analysis to identify homozygous lines. Chi-square test showed that some of the T₃ seeds didn’t have the inheritance of the PHA gene in a 3:1 ratio of Mendelian manner (Table 3). The genetic data did not correspond well to the copy number possibly due to that not many seeds used in segregation analysis might not reflect actual segregation ratio.

The second objective was to cross single locus PHA homozygotes with wild type plants to produce an individual with a single locus PHA gene and in the same chromosomal context but in a wild type background (full container). Thus, it is also critical to find the PHA homozygotes in homozygous wild type genetic background in order to compare PHA in the homozyote ssp-1 with the homozygote cross so as to determine if PHA is overexpressed in ssp-1. All of the four single copy of PHA homozygotes were crossed to Ler. The first generation of the cross, F₁ was selected on hygromycin medium for PHA transgenics. F₁ transgenic plants were allowed to self-pollinated and grown for F₂ seeds, which were screened on selection medium. 32 resistant
seedlings F2 of each backcross were transferred into soil to get F3 seeds, which were used in segregation analysis to identify PHA homozygotes. In order to exclude the PHA homozygotes in ssp-1 genetic background (PHAPHA cruceru) (Figure 13), the F3 seed proteins of these homozygotes were analyzed by SDS-PAGE. PHA homozygotes with the ssp-1 mutation showed two missing band around 35 kDa and 25 kDa on the gel profile (Figure 15). To get the PHA homozygotes in a homozygous CRU3 background (PHAPHA CRUCRU), the CRU3 gene fragments including ssp-1 mutation site of F3 plants were PCR amplified and purified for sequencing. PHA homozygotes in heterozygous CRU3 background showed two equally intense peaks at the site of ssp-1 mutation (C and T) on the sequencing map (Figure 18A, PHA 1 backcross 20). As a result, all the PHA homozygotes in homozygous CRU3 background were identified. The F4 seed proteins of these homozygotes were used in comparison of PHA expression level with PHA homozygotes in the ssp-1 genetic background.

The third objective was to compare the PHA expression levels in both ssp-1 and Ler genetic backgrounds. In theory, if the empty container hypothesis is correct then the transformants would exhibit higher expression levels as compared to their backcrosses. Seeds of transgenic plants and their relative backcrosses were examined for expression levels of the PHA gene by western blot methods; the amount of PHA detected varied from 4.52% to 7.70% of total soluble protein for 4 single locus PHA homozygote lines as compared to the levels in the backcross ranged from 3.13% to 7.26% of total soluble protein (Table 7). The low expression levels of PHA in PHA 3 could be because of that PHA 3 plants were grown weaker than other transgenic lines. Probably more generations of PHA 3 should be produced in order to get some healthier plants. The expression level of PHA varied in extracts prepared from seeds derived from different transformants. Such variation in expression may be caused by alternative sites of PHA integration into the plant genome. For all four of the single copy of PHA transformants, PHA 1, 3, 5 and 7 and their corresponding backcrosses, the results showed that the levels of PHA were higher in the ssp-1 genetic background for transformants with single copy of PHA (PHA 1, 3, 5 and 7) versus in the wild type genetic background. PHA 1 and 3 showed a significant enhancement of expression from 30.33% and 44.18% higher than PHA accumulation in the seeds of its backcross, although PHA 7 only showed 0.31% more PHA as compared to the backcross (Table 7 and Figure 20). These results provide evidence to support the ‘empty container’ hypothesis. PHA band in all four backcross lines have two bands really close together
(Figure 20). The upper PHA band could be some modified PHA protein containing a few more amino acids which could be due to aberrant processing events or caused by glycosylation or phosphorylation.

The result of this project does not show a consistent increase of PHA in all transgenic lines as compared to the backcross, but it does show some promise of the ‘empty container’ as we expected. It would be wise to repeat evaluation of PHA levels in seeds which are from the plants grown and harvested at the same time. Also, it would be beneficial to screen for more transgenic lines that express and accumulate PHA protein to high levels. As discussed before, an increase in one component of seed reserves may lead to a decrease in other seed components or their composition. Therefore, an evaluation of amino acids and lipids of these PHA transformants and their backcross could be done to investigate if there are changes in metabolite composition and levels. Nevertheless, further research needs to be done in order to gain a greater understanding into the control of the seed storage gene expression. This might provide some information for researchers to manipulate seeds to overproduce recombinant proteins for crop improvement and as a vector for low cost production of proteins of interest to pharmaceutical and other industries. For instance, *Brassica* species and *Arabidopsis* are closely related, and this ‘empty container’ strategy obtained from *Arabidopsis* could be translated into *Brassica* species for crop improvement. Currently, a well-established method called TILLING (Targeting Induced Local Lesions In Genomes) has been used to generate a stop-codon mutation like *ssp-1* mutation in diploid *Brassica* species, *B. rapa* (Stephenson et al., 2010). TILLING is a reverse genetic tool for identifying plants with point mutations in genes of interest, which was originally developed from *Arabidopsis*. TILLING normally uses EMS to induce point mutations in genetic material by nucleotide substitution leading to GC→AT transition changes, and then screening point mutations using established high throughout SNP discovery methods (Stephenson et al., 2010). TILLING has also been applied to a wide range of other crops such as corn, wheat, rice, soybean, tomato and lettuce (Bush and Krysan, 2010). By using approaches such as TILLING, the ‘empty container’ concept could be used to generate an appropriate genetic background in other species such that overexpression could be attempted.
5 Summary

ssp-1 is a recessive mutant deficient in one of the major seed storage proteins in Arabidopsis, the 12S cruciferins (legumin-type globulin), with no other observable phenotypic effects. The ssp-1 mutation is due to a G-A transition in the CRU3 gene, which alters a trptophan codon to a premature stop codon in exon 4 (Lee, 2000). Since ssp-1 seeds contain 20-27% less protein than that of wild type plant seeds we propose that this void could be potentially filled by foreign or genetically engineered proteins or other valuable materials. This is the ‘empty container’ hypothesis. Therefore ssp-1 has been termed the ‘empty container’, which might provide a tool to drive higher-level expression of foreign proteins than might be accomplished in a wild type genetic background.

To validate the ‘empty container’ hypothesis, I transformed ssp-1 plants with a phytohemagglutinin (PHA) gene and identified single locus PHA homozygotes in ssp-1. Four single locus PHA transgenic lines were identified and were crossed to wild type Arabidopsis and screened for PHA homozygotes in a wild type background. The rationale for this was to generate transgenic lines in both the ssp-1 and Ler backgrounds, so that the reporter gene would be present in the same copy number and chromosomal context in both the ‘empty container’ background and the wild type background. Using immunoblotting, I evaluated the PHA expression level of the eight lines I generated. The PHA levels ranged from 4.52% to 7.7% of the total protein in ssp-1 transformants whereas the levels in the backcross ranged from 3.13% to 7.26% of total soluble protein.

PHA 1 and 3 showed a significant increase in PHA expression, from 30.33% and 44.18% in comparison to their backcross. These results suggest the validity of the ‘empty container’ hypothesis. ssp-1 seed might be developed as protein-expression systems for production of hard-to-make proteins for pharmaceutical and industrial uses.
References


Kelly, J. D., Hefle, S. L. 2000. 2S methionine-rich protein (SSA) from sunflower seed is an IgE-binding protein. Allergy. 55: 556-560.


