Running head: Proteome rebalancing in soybean seeds
Silencing of soybean seed storage proteins results in a rebalanced protein composition preserving seed protein content without major collateral changes in the metabolome and transcriptome

Monica A. Schmidt, Donald Danforth Plant Science Center, St. Louis, MO 63132
Brad Barbazuk, Department of Biology and the Genetics Institute, University of Florida, Gainesville, Fl 32611, Michael Sandford, Department of Biology, University of Florida, Gainesville, Fl 32611, Greg May, National Center for Genome Resources, Santa Fe, NM 87505, Zhihong Song, Department of Biochemistry, Biophysics and Molecular Biology
Wenxu Zhou, Department of Biochemistry, Biophysics and Molecular Biology Basil J. Nikolau, Department of Biochemistry, Biophysics and Molecular Biology, and WM Keck Metabolomics Research Laboratory, Iowa State University, Ames, IA50011 Eliot M Herman* Donald Danforth Plant Science Center, St. Louis. MO 63132

*Corresponding author; email eherman@danforthcenter.org. The author responsible for distribution of materials integral to the findings in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Eliot M. Herman (eherman@danforthcenter.org).
ABSTRACT

The ontogeny of seed structure and the accumulation of seed storage substances is the result of a determinant genetic program. Using RNAi, the synthesis of soybean glycinin and conglycinin storage proteins has been suppressed. The storage protein knockdown (SP-) seeds are overtly identical to the wild-type maturing to similar size, weight, and in developmental ontogeny. The SP-seeds rebalance the proteome, maintaining wild-type levels of protein and storage triglycerides. The SP-soybeans were evaluated with systems biology techniques of proteomics, metabolomics and transcriptomics using both microarray and next-generation sequencing transcript sequencing (RNA-Seq). Proteomic analysis shows that rebalancing of protein content largely results from the selective increase in the accumulation of only a few proteins. The rebalancing of protein composition occurs with small alteration to the seed’s transcriptome and metabolome. The selectivity of the rebalancing was further tested by introgressing into the SP-line a GFP glycinin allele mimic, and quantifying the resulting accumulation of GFP. The GFP accumulation was similar to the parental GFP-expressing line showing that the GFP glycinin gene mimic does not participate in proteome rebalancing. The results show that soybeans make large adjustments to the proteome during seed filling and compensate for the shortage of major proteins with the increased selective accumulation of other proteins that maintains a normal protein content.
INTRODUCTION

Seed crops are propagated for the stored protein, oil, and carbohydrates that specifically accumulate in seeds. Seeds accumulate protein as a source of carbon, nitrogen and sulfur, and triglycerides and carbohydrate reserves, which are used as a source of carbon and ultimately energy (Herman and Larkins 1999 for review). Seeds can be characterized as storing primarily either protein and triglyceride, or protein and carbohydrate. Soybean seed storage proteins are encoded by few conserved gene families (Harada et al 1989, Nielsen et al 1989, Schuler et al 1982ab). Most seed proteins are members of the cupin superfamily (e.g., legumins and vicilins), but in some dicotyledoneous seeds lectins are abundant, whereas in cereal grains the prolamin and to a lesser extent the legumins are abundant (Herman and Larkins 1999).

During seed filling, the accumulation of seed storage proteins (see Miernyk et al 2009) is regulated by an integrated genetic and physiological network (Brocard-Gifford et al 2003, Elke et al 2005, Fait et al 2006, Golombek et al 2001 for examples). However, the tight genetic control of seed development is modifiable within a certain range by maternal nutritional and environmental effects (Weber et al 2005). The composition of seeds can be broadly defined as a developmental genetic program that is modified by nutrient source availability and the demands of the forming storage substance sink. Viewed from a systems biology perspective the regulation and cross-talk between underlying developmental program modified by extrinsic conditions and nutrient source flux and nutrient sink formation results in determining the final content of the mature seed. Broadly the mature seed can be considered the product of a combinatorial output of the interaction of program, source and sink that yields a specific seed composition constrained within genetically defined limits.

transcription factors in seed development is the primary control of protein and oil and possibly other classes of storage substance accumulation (Kroi et al 2003, Gutierrez et al 2007, Santos-Mendoza 2008 for reviews) and the developmental processes that support the accumulation of storage substances. Forward and reverse genetics experiments have defined regulators of seed development that consists of four master regulator transcription factors, LEC1, LEC2, ABI3, and FUS3, and at least four other proteins (see Gutierrez et al 2007, Kroi et al 2003 for review). The four master regulators have a hierarchal relationship and interactive relationship that further regulate other transcription factors forming a control network. For instance LEC1 and LEC2 positively regulate themselves and the others (To et al 2006). The master regulators form additional regulatory networks with other transcriptional regulators, for example AP2/wrinkled family that is response to sucrose source flux and whose expression alters seed size (Ohto et al 2005) and triglyceride content (Baud et al 2007, Cernac and Benning 2004, Maeo et al 2009, Li et al 2010).

The nutrient status of the maternal plant provides a further regulatory control that modulates the output of the seed genetic program in forming the seed sink (Gayler and Sykes 1985). In this program the plant couples the nutrient source (the maternal tissue) to the nutrient sink (the seed propaqule) on the assumption there is a proportional linkage of the strength of the seed sink to input of the nutrient source perhaps regulated at the transport level. Viewed from this perspective the maturing seed would maximize utilization of the source nutrients (Hernandez et al 2005, Gu et al 2010). Sulfur availability is one example of nutrient source regulating seed protein composition (Beach et al 1985). Legumes such as soybean possess storage proteins with low sulfur amino acid content. The beta subunit of conglycinin exhibits plasticity in response to available sulfur (Hagen et al 2003, Hirai et al 1995, Holowach et al 1986, Tabe et al 2002). There is also feedback regulation of protein filling in the seed sink in response to nitrogen availability (Biermann et al 1998, Ohtake et al 2002). Another of source of regulation is the over-expression of seed-specific amino acid permeases that increase the nutrient flux into the seed resulting in an increase in seed sink protein content (Saalbach et al 2005). The plasticity induced in seed protein composition by altered nutrient source availability modulates the genetic developmental program changing for instance the expression of transcripts as the result of sulfur availability (Rolletschek et al 2005).
Because the stored seed sink will become a source in the following life-cycle, plants maintain a proportional and species-specific defined inventory of protein, triglyceride, and carbohydrate for the use by the germinating seedling. Breeders have long known that in soybean the two major reserve substances, protein and triglyceride, are metabolically linked and their level can be selected as a trait. A shortfall of accumulation of a major reserve substance limits the availability of critical nutrients for the post-germination seedling. Suppression of seed protein sink production results in compensating protein accumulation shown by mutation-induced suppression or genetic modification of storage protein synthesis in Maize, for example in opaque 2 (Geetha et al 1991; Hunter et al 2002), and soybean (Kinney et al 2001, Takahashi et al 2003) all result in rebalancing protein content by increased accumulation of other seed proteins. Moreover, seed protein and other constituents most notably triglyceride have an inverse relationship, where selection for increased protein or oil content results in a compensating decrease in the other reserve substance. The variability in protein and triglyceride content is maintained within relatively narrow limits of about +/-3-5% w/w, with seeds that would function to maintain a relatively defined inventory of storage substances to provide nutritional reserves for the plant’s next generation.

One way to view the genetic, source, sink regulation of seed protein fill is as a hierarchal series of controls, regulation, cross-talk and feedbacks from the genetic to physiological level. In such a systems-oriented model there is a determinant genetic framework that dictates overall development of the seed, including its morphology, the developmental timing of gene expression and reserve substance accumulation. But the genetic program through regulatory controls and feedback modulate the composition and balance of constituents resulting in some plasticity that serves to maintain a relatively defined ratio of storage substances and composition in the mature seed.

In this paper we demonstrate that response to posttranscriptional silencing of soybean storage proteins results in a control that maintains the size of seed protein sink by remodeling the proteome by greatly increased accumulation of a few proteins. Proteome rebalancing to maintain seed protein content occurs with minor collateral changes in the ontogeny of the soybean seed development including the transcriptome, metabolome, gross and ultrastructural morphology, and viability.
RESULTS

Creation of storage protein knockdown seeds

An RNAi construct designed to suppress glycinin accumulation was transferred to soybean using biolistic transformation protocols (Parrott and Clemente 2004). A FAD2 RNAi was also included in this construct to provide a marker for additional screening for high-oleic acid phenotype, and to maintain consistency with a prior conglycinin knockdown that also included the FAD2 knockdown (Kinney et al 2001). Comparisons of the RNAi sequence show a component with homology to the glycinin family of storage proteins and little homology to conglycinin family of storage proteins (Supplemental Figure SF1A). The FAD component of the RNAi has homology with the FAD family (Supplemental Figure SF1B). The regenerated somatic embryos and T0 seeds were screened for total protein distribution by 1D-SDS/PAGE and with immunoblot assays using anti-glycinin and anti-conglycinin antibodies. The recovered transgenic lines not only exhibited suppressed glycinin content, but also an essentially complete knockdown of α/α’-and β-subunits of conglycinin. These lines that exhibited knockdown of both glycinin and conglycinin, termed SP- (storage protein minus) were regenerated into plants. The growth morphology of the resulting plants and their seed set, and the mature seeds were overtly identical to wild-type controls, cv Jack. A total of five lines from two transformation experiments were created that all exhibited the same storage protein suppression.

Small chips were removed from the T0 seeds, and used to assay for seed protein phenotype, and SP-seeds were regrown and reselected for two self-propagated generations to produce a population homozygous for the RNAi transgene. The SP-phenotype was stable through each of these generations, with α/α’ and β-conglycinin subunits being undetectable and glycinin levels being greatly reduced. The SP- phenotype has proven stable with the plants and seeds appearing overtly normal in growth and development through subsequent greenhouse expansion generations and a field test. The oleic acid level in the SP-seeds was > 94% of the fatty acids indicating that the FAD2 marker knockdown was also expressed. The size and dry weight of the greenhouse grown SP-dormant seeds averaged 146 mg, which is similar to dormant seeds of the wild type, control, 163 mg. The protein and oil content of the SP-(40.2%, 19.1%) and the wild type are also (37.5%, 20.5%) similar. Together these data
show that the knockdown of the storage proteins (glycinin and conglycinin) that comprise the majority of the seed’s protein, results in a soybean that rebalances its protein composition to a nearly identical protein and oil content, and maintains its normal seed size. The SP-seeds germinate with about 100% frequency and the initial stages of seedling growth appear overtly identical to the conventional Jack seeds. The SP-trait has been stable through more than eight generations under greenhouse growth conditions. Whether the SP-trait was maintained was further assessed in a field grow-out in 2010 season using a Wooster, Ohio field. Both the proteome rebalancing of suppression of the glycinin and conglycinin storage proteins and their replacement by other intrinsic seed proteins was apparently identical in both greenhouse and field-grown seeds (Supplemental Figure SF2). The FAD2 suppression co-trait was also maintained in the field-grown seeds. Although this grow-out was not a true field-test the overall productivity of the SP-plants were comparable to that of conventional soybean cultivars.

Proteomic analysis of the SP-shows that other seed proteins compensate for the absence of storage protein polypeptides

Two-dimensional (2D) IEF/SDS-PAGE fractionation of protein extracts from SP-seeds in comparison with the Jack seed shows a large change in the spot distribution of the proteins that results from the knockdown of glycinin and conglycinin (Figure 1 A & B) (see Kinney et al 2001 and Hajduch et al 2005 for soybean seed proteome maps and Miernyk and Hajduch 2011 for seed proteomics review). The protein gel stained with Coommassie shows that there is a large-scale change in the protein distribution in SP-seeds, with the absent storage proteins being replaced by other abundant proteins. The knockdown of the glycinin and conglycinin storage proteins was confirmed by immunoblot probing a replicate gel with antibodies specific for these storage proteins. The identification of the induced proteins in the SP-seeds was achieved by mass spectrometric analysis of excised protein gel spots. Triplicate 2D gels were evaluated by visual examination with the assistance of gel spot size scanning software. By comparing the protein spot sizes between SP-and the wild type seeds significantly altered spots were excised, subjected to trypsin-digestion, and analyzed by tandem mass spectroscopy. The map of the protein spots selected for mass spectroscopy is shown in Figure 1C.
The compiled proteomic data (supplemental on-line table ST1) shows that most of the protein content rebalancing is due to greatly increased accumulation of only a few proteins and global increase of a myriad of other proteins that produce the seed proteome. The major proteins exhibiting an increase in abundance with the post-transcriptional silencing of SPs are Kunitz trypsin inhibitor (KTI) (Fig. 1C; spots # 2,7), soybean lectin (Le) (Fig. 1C; spots #17, 1, 24), and the immunodominant soybean allergen P34 or Gly m Bd 30k (Fig. 1C spt # 16). Less prominent increases in abundance were also observed in glucose binding protein (Fig. 1C spots# 1,2) and seed maturation associated protein (Fig. 1C spots # 4, 5). This is similar to the same set of proteins whose accumulation is greatly increased in a cross between two soybean lines that carry conglycinin and glycinin null alleles, respectively (Takahashi et al 2003). The relative contribution of each major protein to the total proteome is shown in a pie chart (Fig. 2) comparing the conventional Jack cv with the SP-. In Jack the integrated spot volume of the 7s and 11s storage proteins from 2D broad pH IEF/SDS-PAGE is 54% of the total protein compared with 11% in SP- that is primarily glycinin A4 with the SP- trait suppressing the more abundant group 1 glycinins. The silencing of storage proteins storage proteins in SP-resulted in increased accumulation of sucrose binding protein from 1% in the conventional to 4% total protein, P34 from 1% to 9%, lectin from 1% to 4%, and Kunitz trypsin inhibitor from 5% to 11%. These four proteins together constitute 28% of the total seed proteome replacing over half of the silenced SP sink. In addition there is a global increase in other all proteins from 35% in the conventional to 58% that together constitute 23% of the rebalanced SP- proteome. Not all proteins increase in abundance, the Bowman Birk protease inhibitor is the major sulfur amino acid sink in soybean constituting 3% of the conventional seed proteome and remains 3% in the rebalanced proteome of SP-. Similarly the major oil body protein, oleosin, constitutes about 2% of the total proteome and it is also unchanged in SP- consistent with the observation that total seed triglyceride content is unchanged. The impact of the change in protein composition on the total amino acid composition of the rebalanced soybean proteome was examined by digestion of SP- and Jack mature seed samples followed by fractionation and quantifying the resulting amino acid distribution. Figure 3 shows a bar graph comparing the amino acids and demonstrates that although the SP- phenotype and resulting protein composition rebalancing changes a majority of the protein content of the seed there is little impact on total amino acid content.
that is largely conserved in the rebalancing process.

**SP- seeds form protein storage vacuoles in a developmentally correct morphology and pattern**

Protein storage vacuoles (PSVs) of dicotyledonous seeds, such as soybean, are formed by the subdivision of the central vacuole, which occurs coordinately with synthesis and deposition of the storage proteins (Herman and Larkins 1999). This results in protein-filled PSVs that fill much of the cytoplasm of seeds accumulating seed proteins. In order to examine the cellular structure of the *SP*-seeds, maturing cotyledons prepared by high-pressure cryofixation and the resulting samples were freeze-substituted with acetone/OsO₄, and then embedded in Epson plastic. In comparison to the wild type cotyledons, the PSVs of the *SP*-lines are overtly similar in size and appearance (Figure 4), and they possess a protein-filled amorphous matrix typical of soybeans. (Note, soybean seed PSVs do not form protein-specific domains, or crystalloids in the matrix). In some *SP*-seed PSVs there appears to be an excess of autophagic inclusions (see Melroy and Herman 1991 for examples in soybean) in some of the PSVs of the maturing seeds as compared to the wild type, although this was not quantified or examined further.

The knockdown of conglycinin, whether by directed genetic engineering (Kinney et al 2001), or as a consequence of a spontaneously arising null allele (Mori et al 2004), results in the retention of a large fraction of normally soluble transport competent glycinin in the precursor proglycinin form, which is accreted in ER-derived protein bodies. In contrast, in response to the knockdown of both glycinin and conglycinin, the storage parenchyma cells contain only PSVs, indicating that the compensating PSV-associated proteins remain vacuolar, without redirection of the proteins into ER-derived protein bodies (see Herman 2008 for discussion). The structure and distribution of all other subcellular organelles and structures appear to be identical in the *SP*-and *Jack*.

**Maturing SP-soybean seeds exhibit limited changes in the global transcript profile.**

The global transcript profile of mid-late maturation *SP*-seeds were compared to the wild type using the Affymetrix DNA Genechip platform, incorporating both biological and technical replicates in the design of the experiment. The resulting transcriptome data indicate that few transcripts show any significant change in accumulation, using a relatively stringent three-
fold up/down cutoff with a t-test p value of <0.05. Only those transcripts that had a positive correlation in both technical and biological replicates were scored as valid. The data is accessible as number GSE12314. The DNA array results were annotated using Brandon et al (2007). The transcriptome data showed RNAi suppression of all of the glycinin and conglycinin-related transcripts. Seventy sequences exhibit decreased abundances more than 3.0 fold and the 45 are increased in abundance sequences that is summarized as a scatterplot, figure 5, with the expression data shown in online supplement table ST2. Some of the transcripts represent the suppressed cupin-superfamily storage proteins. The other transcript changes are a diverse set, among the down-regulated transcripts is the AP2 gene domain related transcription factors that has the largest N-fold decrease of the transcripts assayed by DNA arrays has been associated with the deposition of storage products, particularly seed oil, (Cernac and Benning 2004; Kwong et al 2003). Other transcripts of note include decreased abundance of pre-mRNA processing related proteins, proteins related to a variety of oxygen or oxygenase responses and ferritin.

To further analyze the possible changes in transcriptome resulting from the SP- phenotype replicate RNA samples converted to cDNA were subjected to Illumina sequencing. One half plate of cDNA sequence was obtained on the Illumina Genome Analyzer for both SP- and Jack, which resulted in >12M and >10M 36mer sequences, respectively. Each sequence set was aligned to the soybean gene index (Quackenbush et al 2001) GMGI version 14 (http://compbio.dfci.harvard.edu/cgi-bin/tgi/gimain.pl?gudb=soybean), counts were normalized to RPKM values (Mortizavi et al 2008), and only alignments where both SP- and Jack were represented with a minimum RPKM value of 10.0 were examined. This yielded 12,568 distinct sequence elements of which 997 were down-regulated and 151 were up-regulated more than three-fold in SP- compared to Jack. Overall, the results of the Illumina sequencing confirmed those observed with the Affymetrix array but provided greater breadth and depth of analysis as well as actual representation of individual transcripts in the transcriptome. Since the soybean gene index that the Illumina reads were aligned against a more comprehensive collection of soybean transcript sequences than represented on the Affymetrix array provides additional depth of information that extends as well as conforms the more preliminary array dataset. Table 1 displays a summary of selected transcript abundance data comparing the Affymetrix
and Illumina data for the transcripts of the silenced proteins and the proteins that compensate as well as a few other seed proteins such as the oil body localized oleosin. The Illumina data enabled some transcript signals identified in the Affymetrix analysis to be resolved to specific members of gene families. The Illumina data set is available as on on-line supplemental table ST3, with the data accessible as downloadable in excel format and the raw dataset as accession # GSE21116 at NCBI. The transcripts down-regulated include those encoding the proteins with suppressed accumulation in SP- phenotype shown in the proteomic assay (Fig 1, 2, table 1) and in the Affymetrix transcript assay (supplemental table ST2). The Illumina dataset as a quantitative assessment of transcript abundance shows a much greater breadth of down-regulation of 40 to 175-fold compared to the Affymetrix results confirming and extending the array results (supplemental table ST2). KTI, lectin, P34 transcripts that are the proteins that rebalance the protein content in SP- do not exhibit changes in transcript abundance more than three-fold. This confirms the Affymetrix array results that also showed that the increase abundance of KTI, lectin, and P34 in SP- occurs without a large increase in transcription (Table 1). The Illumina results also showed that other major seed proteins, for instance oil body oleosin, protein storage vacuole aquaporin α-TIP and Bowman-Birk protease inhibitor (BBI) do not show significant changes in transcript abundance (Table 1). Among the metabolic enzymes with increased steady state expression in SP- is asparaginase (transcript #12528, 12529, 6.3 fold up-regulation, plus two other transcripts #7131 and 11879 with less than a 2 fold down and up regulation), an enzyme that has been associated with post-germination storage protein mobilization when the amino acid flux generated by proteolysis becomes the nutrient source for the growing embryo. The free amino acid and metabolic profile of SP- shows a 5.98 fold excess free asparagine compared to the Jack control (Figures 6 and 7 and supplemental table ST4A) suggesting that resulting increase in asparaginase expression is a feedback regulation.

The posttranscriptional knockdown of storage proteins resulted in up and down changes in the abundance of changes in selected transcriptional factors. The Affymetrix array showed that the transcription factor AP2 was one the transcripts exhibiting the largest change in abundance (supplemental table ST2). The Illumina results confirmed and extended this observation showing several different transcripts encoding AP2/wrinkled/Ethylene binding factor (EBF) are greatly decreased in abundance shown in supplemental table ST3;
transcript #s 19, 11.8 fold down; 343 and 344, 3.94 fold down-regulated; 600 to 602, 3.43 down-regulated and transcript 937, 3 fold down-regulated). Each of these individual genes presumably contributed to the over-all observation of the AP2 domain transcription down-regulation results from the DNA array. Transcript #33, 8.25 fold down-regulated is a member of the bzip family of transcription factors. Members of this family have been shown to be involved in a regulon related sucrose allocation to storage compounds. Breeders have long asserted there is a negative correlation between oil and protein content in soybean where either one decreases in abundance in response to increases likely reflecting carbon flux allocation. In SP- phenotype where if not for the observed protein rebalancing based on breeding experience oil accumulation would be expected to increase but it does not. What, if any, role the decrease in the bzip transcription factor has in maintaining the balance of protein and oil accumulation will require further investigation. Among the other >1,000 transcripts up- and down-regulated comparing SP- and Jack are a wide variety of proteins representing many presumably un-related pathways representing primary, secondary, and tertiary effects showing that suppressing SP caused broad adjustments to the expression of low abundance message in the seed transcriptome.

Metabolite profiling of SP- seeds indicate that the metabolic differences with the wild type seeds at mid-maturing stage narrow by the onset of seed maturation

The alteration of protein accumulation in SP- seeds would be expected to impose an altered dynamic on the metabolism of the influx carbon and nitrogen nutrients as they are converted to the protein and oil seed reserves. Total free amino acid analysis showed that the SP- trait resulted in the accumulation of excess free asparagine largely at the expense of other amino acids (Figure 6). Asparagine is one of the major nitrogen transport vehicles of plants including soybean. Additional insights into these metabolic changes was obtained by conducting a combination of non-targeted metabolomics analyses and metabolite profiling that was targeted for amine-containing metabolites and fatty acids. In combination, these analyses measured the relative abundance of about 320 metabolite analytes (Figure 7A-D, Supplementary table 4A&B). These analyses were conducted on extracts made from mid-maturation seeds and mature seeds (approximately 200 mg tissue per extract). Seed samples were quickly frozen in liquid N₂ to rapidly quench metabolism and three parallel analytical
procedures were used to assess the metabolite pool sizes in the samples. All three procedures used GC-MS as the analytical platform for gathering this dataset, adapting procedures that have been developed for assessing metabolite pools in Arabidopsis (Bais et al. 2010; www.plantmetabolics.org). Two of these analytical procedures targeted the analysis of amine-containing metabolites (enriched in amino acids) and fatty acids, respectively. The rationale for targeting the analysis of these metabolites is that they are intermediates and products of the major seed reserves (i.e., proteins and oils, respectively). The third analytical procedure assessed the abundance of metabolites irrespective of their chemical class (i.e., non-targeted metabolite profiling).

The identity of the metabolites was determined by a combination of chromatographic behavior (i.e., retention indices relative to a series of hydrocarbon standards), and matches with mass-spectra using either authentic commercially available standards or chemical databases (i.e., NCBI, www.plantmetabolomics.org and PubChem). In combination the three analytical protocols identified about 320 metabolite entities, and of these, 180 are metabolites whose chemical identities were determinable via the above two criteria, and the remaining 143 metabolite entities that were not chemically defined, but were labeled by a convention that has been defined by the plant metabolomics research community (Bino et al. 2004). These latter analytes were mainly detected by the non-targeted metabolite profiling platform.

Because the response of the MS-detector used to assess the abundance of the metabolites is dependent on the chemical nature of each metabolite, it is not possible to generate absolute abundance data for those metabolite entities whose chemical nature was not determined. Moreover, even for those metabolites whose chemical identity was determinable, the response of the MS-detector is variable between each metabolite. Therefore, in order to integrate all the metabolite data irrespective of whether the chemical identity of the metabolite is known, we assessed the effect of the SP- transgene on the relative abundances of all metabolite entities.

Figures 7A-D show the relative abundances of the 320-metabolite entities in the two genotypes assessed at two different stages of seed development. Figures 7A and 7B plot the log-ratio of the 180 metabolites whose chemical identity is known, and the order of these metabolites on the y-axis is the same in both plots. Similarly, Figures 7C and 7D plot the
log-ratio of the 140 analytes whose chemical identity is unknown, and the order of these analytes on the y-axis is the same in both of these plots. In these log-ratio plots, those metabolites that hyper- or hypo-accumulate as a consequence of the genetic manipulation plot to the extreme left or right of the zero-ordinate of the x-axis, respectively, whereas those metabolites, whose abundance was minimally affected by the genetic manipulation plot near the zero-ordinate. Cursory examination of these plots indicates that at the mid-maturation stage of seed development the relative abundance of more metabolites is affected by the SP-transgene than in seeds at near maturity. Specifically, of the 320 metabolites that were assessed, 95 either hyper- or hypo-accumulate more than 4-fold in the SP-lines relative to wild-type, whereas the equivalent number is 38 metabolites at seed maturation. Therefore, these data indicate that although the metabolic state of the seeds is altered by the transgenic event at the mid-maturation stage, as the seeds develops to maturity compensatory mechanisms appear to be expressed to that bring the metabolic status of the seeds to near wild-type state.

Supplemental Table 4A&B, list the metabolites that show altered accumulation at mid-maturation stage and at seed maturity, and these are listed in the order of magnitude-difference in abundance, from those that hyper-accumulate (most red-shaded cell) to those that hypo-accumulate (most-blue shaded cell) in response to the SP-transgenic event, and the chemical nature of these metabolites is indicated in Figures 7A and 7B. These data indicate that in response to the SP-transgenic event, at mid-maturation stage of development, amino acid metabolites hyper-accumulate (as indicated by the enrichment of the red-colored symbols at the top of Figure 7A), and carbohydrate metabolites hypo-accumulate (as indicated by the enrichment of the green-colored symbols at the bottom part of Figure 7B). However although source free amino acids accumulate the process of proteome rebalancing maintains a fixed protein and total amino acid content (Fig 3), albeit with a very different proteome composition. The accumulation of carbohydrate likely results from the interplay between carbon allocation to amino acids and other polymer precursors with the rebalancing process tipping the ratio of allocation to favor carbohydrates.

A foreign protein introgressed into the SP-background does not contribute to seed protein rebalancing

The impact of expressing a foreign protein gene was tested in the context of the protein
rebalancing process occurring in the SP-lines. The expression of the extrinsic gene product consisted of the G2 glycinin promoter sequence, glycinin terminator, and a modified GFP that included a chitinase signal sequence and C-terminus ER-retention sequence. This mimics a G2 glycinin allele with an exchanged open reading frame (GFP). This transgene is designed to accrete in the ER, forming ER-derived protein bodies (PB), which are inert, de novo created organelles (Schmidt and Herman 2008a) that are not normally found in soybean. Because the GFP-tagged protein bodies stably accumulate during seed maturation (Schmidt and Herman 2008a), the fluorescence associated with this protein/organelle can be quantified to measure its accumulation throughout seed development.

The GFP construct was introduced into soybean by biolistic transformation followed by selection and regeneration of the plants. GFP positive seeds were re-grown for T1 and T2 generations producing a homozygous line of seed-specific GFP-HDEL expressing seeds. Accumulation of the GFP-HDEL protein was assayed using a fluorometer assay with a standard curve control using commercially obtained GFP (Schmidt and Herman 2008a). These assays showed that the GFP-HDEL protein in the parental homozygous seeds accumulates to 1.6-2% GFP of seed protein (Fig. 8). GFP-HDEL homozygous plants were then used as pollen donors in a cross to homozygous SP- seeds. Successful crosses were obtained, subjected to recurrent selection producing homozygous GFP-HDEL/SP- seeds that were assayed GFP fluorescence. GFP fluorescence in the SP- background was found to be approximately equal to that of the GFP-HDEL parental line indicating the GFP-HDEL glycinin allele mimic does not appear to participate in the proteome rebalancing in the SP-.
DISCUSSION

Protein rebalancing remodels the seed into an alternate variant of its conventional developmental ontogeny.
These results show that there is a remarkable adaptive response to shortage of SP accumulation in soybean. It is a fascinating question as what was the selection process that led to the development of physiological process that will precisely compensate for the lesion of a shortage of most of the conserved SPs maintaining protein content, the ratio of protein/oil content, and the seed’s amino acid content. That seeds possess compensatory mechanism(s) by which an alternate maturation program results in an overtly normal conventional soybean albeit with a greatly altered proteome may require the utilization of unrecognized feedback and control mechanisms. That the soybean SP- and Jack soybean content diverges in metabolome during maturation but that all of these differences narrow by seed maturation shows that even with the large perturbation in composition the seed strives to rebalance to a variation of normal composition. The overt development of the SP-soybean seeds is indistinguishable from cv Jack. The present results suggest that soybean seeds and likely other seeds possess far more plastic developmental programs than is inferred by plant breeding experience. The capacity to absorb large perturbations and to rebalance composition and developmental program to reach an specific composition endpoint indicates that the interplay of all of the various metabolic and synthetic pathways are capable of adjusting to a new program leading to an overtly normal seed. The 11S and 7S SP families are conserved in seeds from their first evolution. That the soybean seed proteome rebalancing process functions equally well in field-grown as well as greenhouse seeds exchanging SPs for non-SPs raises many interesting questions about what are the selective factors have maintained SPs throughout plant evolution. The results shown here show that soybean seeds have the capacity to possess an extensively remodeled protein composition without overt change in the seed’s developmental pattern of maturation and content in plants that appear to be equally productive to conventional soybeans.

Suppression of seed storage protein accumulation induces proteome rebalancing to maintain protein content and amino acid balance within a fixed box
The results presented here provide a new perspective on the developmental process of protein filling of soybean seeds and, perhaps by implication, seeds in general. The results of this study indicate that there is an intrinsic process(s) that evaluates the progress of protein filling during seed development and can alter the mix of proteins synthesized to rebalance the system to produce a mature seed with the correct protein and oil content with a greatly altered protein composition. The results in this paper extend a prior investigation in which $\alpha/\alpha'$ conglycinin is suppressed (Kinney et al 2001) resulting in transgenic soybeans that rebalance protein content by increasing the level of glycinin to compensate. This results in soybean seeds that possess an identical total protein content but a different proteome dominated by glycinin. A similar rebalancing of glycinin for conglycinin shortage occurs in a conglycinin null obtained from screening a germplasm collection (Mori et al 2004). A cross of the conglycinin mutant with a glycinin mutant also results in rebalancing with several PSV proteins along with an increase in free amino acids replacing the SP shortage (Takahashi et al 2003). The post-transcriptional RNAi knockdown of all three SP proteins, $\alpha/\alpha'$ conglycinin, $\beta$-subunit of conglycinin and glycinin results in a rebalancing of protein distribution to maintain total protein content with a complex alteration of the proteome resulting in maintaining the seed’s protein content. Taken together these results presented here and in the prior publications (Kinney et al 2001, Mori et al 2004, Takahashi et al 2003) show that soybean seeds have the capacity to compensate for a protein accumulation shortage and to rebalance that shortage by accumulating additional quantities of other intrinsic seed proteins to attain a predetermined protein content of approx. 40%. The rebalancing process appears to function whether the SP shortage is generated by mutation (Mori et al 2004) or by directed knockdown in transgenic soybeans (Kinney et al 2001). This indicates that seed size and composition is regulated so that the seed rebalances the proteome to maintain overall protein content.

The introduced $SP$- lesion is post-transcriptional RNAi mediated suppression implying the mechanism(s) that mediates proteome rebalancing has components that must recognize the protein shortage and induce sufficient alternate protein synthesis to compensate for the protein shortfall and maintain the amino acid balance. Such mechanisms could involve primarily translational control, or alternately the shortage of storage protein may be recognized by a signal transduction system that induces changes in gene expression.
The RNAi used in this study was designed to target glycinin; the accumulation of both glycinin and conglycinin was suppressed is surprising. This may support a role for a signaling mechanism that functions to repress transcription of the mRNA of other related proteins when the RNAi targets one of the gene families perhaps by an epigenetic process. Proteomic analysis shows that in SP- no other cupin-superfamily SP protein compensates in the rebalancing of seed protein composition. Instead other PSV proteins normally produced are accumulated at higher proportional levels resulting in seeds with a normal protein content but a different protein composition. The use of PSV proteins KTI, SBA and P34 as major contributors to the rebalancing process is similar to the rebalancing previously shown to occur in the cross of glycinin and conglycinin null mutants (Takahashi et al 2003).

Protein content rebalancing occurs without significant collateral changes in transcriptome

The Affymetrix arrays show comparatively few changes in transcript abundance in SP- compared to the conventional Jack seeds. In contrast the Illumina cDNA sequencing of the same samples showed 997 transcripts down-regulated and 151 transcripts up-regulated comparing SP- and Jack. There are several reasons for the apparent increase in sensitivity including only partial representation of the soybean transcriptome on the Affymetrix array, that DNA array hybridization is at best a compromise average of binding conditions for the sequences, and that the DNA array poorly assess abundance of low abundance transcripts. The Illumina transcriptome results point to more complex and subtle changes in transcription resulting from silencing SPs. The Illumina results confirmed and extended the Affymetrix array results by quantifying the transcript populations that the SP silencing and resulting proteome rebalancing does not result from a large increase in transcription of the genes encoding the compensating proteins. But the apparent changes in a many unrelated pathways indicate there is the potential for complex feedback control, regulation, and cross-talk indicating that overt proteome rebalancing has components that are primary, secondary and tertiary events resulting from the SP suppression. How these unrelated pathways intersect and have regulatory connections remains to be determined given the complications in both data gathering and interpretation necessary. The complexities of regulatory cross-talk in seed development is beginning to be established. Gu et al (2010) recently described changes in
amino acid catabolism of seeds results in cross-talk in what they described as a complex network of regulation of accumulation of amino acids generated by un-related pathways.

The large-scale changes in SP- proteome that occur with minimal changes in transcription shown with the Affymetrix array was confirmed by Illumina cDNA sequencing. KTI, lectin, and P34 transcript abundance of SP- were within two-fold of that observed in conventional cv Jack and represent examples of the primary proteins that compensate for the SP shortfall. Bowman-Birk Inhibitor (BBI) abundance provides another contrasting observation, where BBI protein accumulation is not altered in SP- seeds and its transcript level is essentially the same as in the control. That BBI does not exhibit altered transcript levels and/or protein accumulation likely indicates another separate proteome regulation within the conserved protein content box. BBI contains a large fraction of the total sulfur inventory of the seed and its regulation is tied to sulfur availability (Hirai et al 1997) and therefore its regulation might be expected to be separated from other seed proteins that do contribute to the protein rebalancing in SP-. This indicates that there is a source availability component to proteome rebalancing where the amino acid balance of the seed is conserved along with its protein content.

For the SP silencing the transcript changes represent the repercussion of silencing large fraction of the total seed protein and therefore a change in the majority of the seed protein sink. Some of the transcript changes are likely significant to the physiological changes imposed by the SP- lesion, for instance the change in transcript level of the wrinkled/AP2/ethylene response protein transcription factors that have been shown to regulate processes resulting from changes in sugar source and the formation of the TAG sink (Baud et al 2007, Cernac and Benning 2004, Maeo et al 2009, Li et al 2010) and seed size (Ohto et al 2005; Okamuro et al 2007 for review). Similarly the bzip transcription factor (Weise et al 2004) that is highly down-regulated in SP- is related to sucrose regulation of transcription. Although the SP- maintains the same protein content as the wild-type cv Jack the regulatory processes and cross-talk between pathways that occur presumably impacts the processes of control of the other storage substance accumulation even if in the final seed the proportions of the storage substances in the mature seed remains fixed. One speculative possible role for the decrease of AP2/wrinkled/EBF related transcripts in SP- seeds is if the
SP-trait would induce a reduction of seed size that is then mitigated by decrease in AP2 and this through interaction and cross-talk maintains a normal seed protein content and seed size in SP-seeds. This should be testable by combining SP- and an over-expression of AP2/wrinkled/EBF transcripts, experiments currently in progress. Similarly the excess of free Asn in SP- may be an example of feed-back gene regulation in SP- where the excess of Asn is sensed and induces the increase of asparaginase expression.

Protein rebalancing does not significantly impact the metabolome

The metabolomic analysis does show some minor changes occurring as a consequence of the SP-trait and proteome rebalancing. However the remodeling of protein composition within a fixed protein content box would not be expected to require major metabolic changes exchanging like sink. The free amino acids do show differences comparing immature green seeds during the process of protein fill. Some of these may be due to differential composition of the now dominant proteins that present a different amino acid sink for their synthesis. It is likely the seed is capable of making minor adjustments in amino acid availability to compensate for the altered amino acid sink requirements. The amino acid differences in maturing seeds between SP- and Jack narrow but are not entirely eliminated in the mature seed. Similarly many of the other metabolites including sugars and oligosaccharides exhibit some variance between the Jack and SP-in immature seeds during protein filling but the differences narrow and often are eliminated in the mature seed. This again shows rebalancing occurs with the mature SP-seed possessing a composition that is similar to that of the Jack. This indicates that although there may be some perturbations of the metabolome by altering the SP protein sink, the rebalancing process adjusts the metabolome by the end point of seed maturation to produce seeds close to the wild-type metabolome configuration. The most significant difference between SP- and Jack in the metabolomics assays is the greater abundance of mannose and various oligosaccharides that may reflect a minor redistribution of reserve carbon in favor of carbohydrate. This again indicates that protein rebalancing is largely a physiological process contained within the context of protein and in particular vacuolar protein accumulation that has little collateral impact on the remaining cellular processes. Although the metabolomics analysis did not analyze TAG precursor synthesis, the SP-seeds possess an essentially wild-type content of TAG so the inference even if not demonstrated is that rebalancing protein has little impact on the carbon flow to
TAG. However although there is no apparent change in TAG accumulation the changes in transcription factors and metabolic alterations may be adjustments to maintain the differential flux of source carbon into the various reserve substances. Reversing the changes in expression in transcription factors in the SP- is one approach under way to examine how the rebalancing process leads to a fixed content of protein seed with a very different composition.

The introgression of foreign protein accumulation into the protein content rebalancing background indicates rebalancing is specific

In a prior study it was shown that suppression of conglycinin results in the increased accumulation of proglycinin/glycinin to compensate (Kinney et al 2001). Foreign proteins introduced into a conglycinin suppressed background as a mimic of the glycinin gene will participate in the proteome rebalancing of glycinin exchanging for a conglycinin shortage (Schmidt and Herman 2008). The introgression of a reporter foreign protein, GFP-HDEL, into the SP background does not result in a large change either up, or down, of the GFP-HDEL’s accumulation (Figure 8). The proteomics assays of the rebalancing shows that the selection of intrinsic proteins is selective. The lack of an increase in a foreign protein’s expression introgressed into the SP- background substantiates the selective choices used in protein rebalancing. Although the present experiment to test the potential of increased abundance of a foreign protein by the rebalancing was unsuccessful if the underlying process can be exploited this would open the potential to obtain high yields of foreign proteins in soybean.
MATERIALS AND METHODS

Construction of Storage Protein suppression RNAi Cassette

An RNAi cassette specific for the simultaneous suppression of both endogenous soybean storage proteins, β-conglycinin and glycinn, and FAD-2 was produced as described in Schmidt & Herman 2008b with the inverted arms of the construct produced by gene specific amplifications as follows. cDNA from 150 mg soybean cotyledon was amplified using Superscript II (Invitrogen) and primers specific to glycinn subunit A1bB2 (genbank accession AB030495) and FAD-2 (Genbank AB188250) open reading frames in two separate PCR reactions. Primer pairs

5’Gly 5’ (TTCTAGACTCGAGTTATATTGACGAGACCATTTGCACA 3’) and
3’Gly (5’ CAGTGGCGGATATCGAGCTCCAGCCAACCGCAAGTTTTGT 3’)

including restriction sites XbaI, XhoI (underlined and bold, respectively) on the 5’ primer located at 939 bp, and homologous FAD-2 (underlined) on the 3’ primer located at 1270 bp were used to amply a 331 bp region of the glycinn A1bB2 gene. Similarly, primer pairs

5’FAD2 (5’ GGAGCTCGATATCCGCCACTGCTTCTCTCTGACGT 3’) and
3’FAD2 (5’ TAAGCTTAGTTACGCGTTTAGAATATATGGG 3’),

including restriction sites (underlined and bold, respectively) on the 5’ primer located at 618 bp, and restriction sites HindIII and SpeI (underlined and bold, respectively) on the 3’ primer located at 746 bp were used to amply a 128 bp region of the FAD-2 gene. The two resultant amplification products, 331 bp glycinn and 128 bp FAD2, were gel purified (Qiagen gel extraction kit) and used as template in a single PCR reaction using the 5’Gly and 3’FAD2 primers as above. The single 459 bp amplicon, consisting of a 5’ glycinn region with XbaI and XhoI restriction sites and a 3’ FAD2 region flanked by HindIII and SpeI restriction sites, was cloned into TOPOvector (Invitrogen) and subjected to two separate double digestions, XbaI/HindIII and XhoI/SpeI which would in turn be placed on either side of an intron to make up the inverted repeats in the hairpin RNAi cassette as described previously (Schmidt & Herman, 2008b). The cassette with the hairpin for both storage proteins and FAD-2 in
tandem was then placed into a vector previously described (Schmidt & Herman, 2008a; Moravec et al., 2007) under the regulatory control of the glycinin promoter and terminator and also containing the hygromycin resistance gene this vector is hereby referred to as pRNAiSP-

**Transgenic Seeds and Proteomic Analysis**

Transgenic soybean plants expressing the pRNAiSP-cassette were produced as previously described (Schmidt & Herman, 2008a,b) and the resultant seeds analyzed by proteomic analysis. Initial screening to identify phenotype positive seeds was preformed by both one-dimensional sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) stained for total soluble protein by 0.1% coomassie brilliant blue stain and a replicate immunoblot probed using a mixture of polyclonal antibodies one specific to glycinin and another to β-conglycinin produced previously in this laboratory. Non-transformed soybean seed was used as a positive control. Seeds which corresponding chips were shown to have the desired phenotype were grown into the next generation. Two generations were grown and screened in this manner until homozygosity was obtained.

**Two-Dimensional protein analysis and mass spectroscopy analysis**

Total soluble protein was isolated from mature seeds as previously described (Schmidt & Herman, 2008a,b). The soluble protein extract (150 µg) from both a nontransformed soybean seed and a homozygote pRNAiSP-seed was separated on a first dimension 11 cm immobilized pH gradient (IPG) gel strips (pH 3-10 non-linear) (BioRad Hercules CA) and then a second dimension SDS-PAGE gels (8-16% linear gradient) and subsequently stained in 0.1% (w/v) Coomassie Brilliant Blue R250 in 40% (v/v) methanol, 10% (v/v) acetic acid overnight and subsequently de-stained for approximately 3 hrs in 40% methanol, 10% acetic acid. Individual spots of interest were excised, digested with trypsin, and the fragments analyzed and identified by tandem mass spectroscopy as previously described (Schmidt and Herman 2008b).

**DNA array analysis**

Total RNA was extracted from 150 mg cotyledons from both wildtype and pRNAiSP-cotyledons and microarray analysis performed using Soybean Arrays (Affymetrix) by Iowa State University Service Center. RNA used in the array experiment was from 3 biological samples: a non-transformed control, and two individual pRNAiSP-cotyledons with each
sample performed in duplicate. The array data was analyzed by ArrayAssist (Strategene) and only data with a stringent positive correlation coefficient and significant data in both technical and biological replicates were scored. The Affymetrix annotation of the Genechip was supplemented with a search of additional annotations by Brandon et al 2007.

**Illumina transcriptome sequencing**

A portion of the RNA isolated above (DNA array analysis) was converted to cDNA, hence the Illumina sequencing can be considered a technical replicate of the Affymetrix assay. Following second strand synthesis, end repair, and A-tailing, adapters complementary to sequencing primers were ligated to cDNA fragments (mRNA-Seq Sample Prep Kit, Illumina, San Diego, CA, USA. Resultant cDNA libraries were size fractionated on agarose gels, and 250 bp fragments were excised and amplified by 15 cycles of polymerase chain reaction. Resultant libraries were quality assessed using a Bioanalyzer 2100 and sequenced for 36 cycles on an Illumina GA II DNA sequencing instrument using standard procedures. All Illumina reads from the wildtype and pRNAiSP- were aligned to the soybean gene index (Quackenbush et al 2001) GMGI version 14 (http://compbio.dfci.harvard.edu/cgi-bin/tgi/gimain.pl?gudb=soybean) with the BOWTIE aligner ver 0.9.9.3 (Langmead et al 2009) with default parameters that results in only one alignment reported for each Illumina sequence. Counts on each gene index sequence were converted to RPKM values (reads per kilobase of target, per million mappable reads), which normalizes for transcript length and for the total read number in each experiment (Mortazavi et al 2008). Target length was simply determined by the length of the gene index target sequence that was aligned. Since many of the gene index sequences are likely not to be full length, many target lengths should be considered as estimates of actual lengths.

**Quantitation of Hydrolyzed Amino Acids and Free Amino Acids**

Samples of midmaturation and mature Jack and SP- cotyledons were submitted for hydrolyzed and free amino acid analysis. For hydrolyzed amino acids, samples were hydrolyzed for 24h at 116 C in 6N HCl containing 0.5% phenol. Samples are dried down and resuspended in 20 mM HCl, derivatized with the AccQ-tag reagent (Waters, Milford, MA, USA), separated by a Waters Acquity UPLC system and quantified as per the manufacturer’s method. Samples for free amino acid analysis were extracted according to
Hacham et al., 2002 and subjected to analogous derivitization, detection and quantifying methods. Triplicates were assayed for each biological sample.

Metabolomics

a. Chemicals
Standards such as ribitol, nonadecanoic acid, n-hydrocarbon mix (C8-C40), BSTFA/TMCS (N,O-Bis(trimethylsilyl)trifluoroacetamide with 1% trimethylchlorosilane), were purchased from Sigma-Aldrich (St. Louis, MO). “EZ:fast” kit for free amino acid analysis was purchased from Phenomenex (Torrance, CA). All other chemicals were purchased from Fisher Scientific (Fairlawn, NJ).

b. Soybean sample preparation
For non-targeted metabolite analysis, soybean seeds were pulverized in liquid nitrogen with cryogenic grinder (SPEX CertiPrep, Metuchen, NJ). The samples (20 mg) were homogenized with 0.35 ml hot methanol (60°C), spiked with 25 µg ribitol and 25 µg nonadecanoic acid as internal standards. The mixture was immediately incubated at 60°C for 10 min and sonificated for 10 min. The mixture was homogenized with 0.35 ml chloroform and 0.3 ml water and centrifuged. 200 l of the upper polar fraction and lower non-polar fraction were transferred into 2 ml glass vials respectively and dried by Speed Vac Concentrator (Savant, NY). The extracts were derivatized using methoxyamine hydrochloride at 30°C for 90 min and BSTFA/TCMS at 60°C for 30 min. One sample was spiked with n-hydrocarbon standards mix for retention index.

For amino acids analysis, soybean samples (25 mg) were homogenized with 0.5 ml 10% trichloroacetic acid, spiked with 5 nmol norvaline as an internal standard.

Extracts were centrifuged and the supernatants were transferred to a 2 ml glass vial. Purification of extracts and derivatization of amino acids were performed using the “EZ:faast” kit from Phenomenex according to the manual provided by the manufacturer. For fatty acids analysis, soybean samples (50 mg) were homogenized with 0.5 ml 10% barium hydroxide and 0.55 ml 1,4-dioxane, containing 20 µg/ml nonadecanoic acid as an internal standard. The mixture was incubated at 110°C for 24 hr. After cooling to room temperature, the mixture was acidified using 6N HCl, and fatty acid analytes were recovered by extracting the aqueous phase with hexane. The recovered fatty acids were derivatized by methylation
with 1N HCl in MeOH at 80°C for 60 min, and by silyation with BSTFA/TCMS at 60°C for 30 min. The mixture was transferred to a 2 ml glass vial for GC-MS analysis.

**Gas chromatography-mass spectrometry.** GC-MS analyses were performed with an Agilent 6890N GC interfaced to a 5973 MSD detector (Agilent Technologies). A HP5ms column (30 m x 0.25 mm; 0.25 m film thickness, Agilent Technologies) was used. For non-targeted metabolites analysis and fatty acids analysis, the temperature gradient was programmed from 80 to 320 °C at a rate of 5°C/min with He flow rate at 2.2 ml/min. Operating parameters were set to 70 eV for ionization voltage and 280 °C for interface temperature. Collected GC/MS data was deconvoluted and analyzed using AMDIS program (NIST) with retention index (RI) information. Metabolites were identified based on their mass spectral by comparison with those of authentic standards in our laboratory’s standard compounds library and the NIST 05 mass spectral library. Amino acids analysis was performed on an Agilent 6890 GC equipped with a flame ionization detector (FID). A Phenomenex ZB-AAA GC column was used. The temperature gradient was programmed from 110 to 290 °C at a rate of 30°C/min with He flow rate at 1.2 ml/min. Analytes were identified based on comparison of their retention time to standard mixtures provided in

**Storage Protein knockdown cross GFP seeds**

Homozygous seeds expressing an endoplasmic reticulum (ER) targeted and retained green fluorescent protein (GFP) in a seed-specific manner were previously produced, analyzed and described in Schmidt & Herman, 2008a. Homozygous pRNAiSP-plants were crossed pollinated using homozygous GFP-HDEL plants as a pollen source. Putative seeds from resultant cross were analyzed under blue light for fluorescence. Crossed seeds were heterozygote for both traits of interest, storage protein knockdown and GFP fluorescence, and were grown to homozygosity for both traits by simultaneously analyzing resultant seeds for fluorescence under blue light and SDS-PAGE analysis for storage protein suppression phenotype. Seeds homozygous for both traits, GFP and $SP^-$, were obtained and analyzed for the amount of GFP by fluorescence on a spectrophotometer in triplicate as previously described (Schmidt & Herman, 2008a) and calculated GFP units/mg protein +/- standard error. The quantity of GFP produced in seeds from the two parental lines (GFP-HDEL and $SP^-$) and 4 homozygote cross lines were compared.
Electron Microscopy

Mid-and late maturation soybean cotyledons were excised and cryofixed in a Balzers high-pressure device. The cryofixed tissue was freeze-substituted in the presence of osmium tetroxide. The fixed dehydrated tissue was embedded in epoxy resin, thin sectioned, counter stained with 5% w/v aqueous uranyl acetate and visualized with a LEO electron microscope.

ACKNOWLEDGEMENTS: We thank Dr. Prachuab Kwanyuen, USDA/ARS, North Carolina State Univ. for measuring the protein and oil composition of the SP- seeds. We also acknowledge the technical assistance of Dr. Howard Berg and Dr. Sophie Alverez of the Danforth Center’s Integrative Microscopy Facility and Mass Spectroscopy Facility. We greatly appreciate Dr. Rouf Mien’s (USDA-ARS) grow-out and harvest of the SP-seeds.
LITERATURE CITED


Baud S, Mendoza MS, To A, Harscoet E, Lepiniec L, Dubreucq B (2007) WRINKELD1 specifies the regulatory action of LEAFY COTYLEDON2 towards fatty acid metabolism during seed maturation in Arabidopsis. Plant J 50:825-838


Brandon L, Wagmaister J, Bui, A (2007) Affymetrix soybean genome array design,
annotation and analysis. 


Herman E (2008) ER bodies; solving the insoluble. Curr Opin Plant Biol 11:672-679


32
gene expression in *Arabidopsis*. Development **130**:6065-6073


domain of APETLA2 defines a large new family of DNA binding proteins in Arabidopsis. Proc Natl Acad Sci 94:7076-7081


Perez-Grau L, Goldberg RB Soybean Seed Protein Genes Are Regulated Spatially during Embryogenesis. Plant Cell 1:1095-1109


Schmidt MA, Herman EM (2008a) Proteome rebalancing in soybean seeds can be exploited to enhance foreign protein accumulation. Plant Biotech J 6:832-842

Schmidt MA, Herman EM (2008b) Suppression of soybean oleosin produces micro-oil bodies that aggregate into oil body/ER complexes. Molec Plant 1:910- 924

Schuler MA, Ladin BF, Fryer G, Pollaco JG, Beachy RN (1982b) Structural sequences are conserved in the genes coding for the \{alpha\}, \{alpha\}' and \{beta\} subunits of the soybean seed storage protein and other seed proteins. Nucleic Acids Res 10:8245–8260


Figure Captions

**Figure 1.** The two-dimensional gel fractionation of the total proteins of mature *Jack* (A) and *SP*- (B) seeds is shown. Note abundant storage proteins in *Jack* seed are absent on the *SP*-seeds that accumulate a few alternate proteins that account for a large fraction of the total seed proteins. (C); The spot selection for excision from the *SP*-seeds is shown to identify protein spots with changes in comparison with *Jack*. The spots were excised, digested with trypsin and processed by tandem mass spectroscopy to identify the proteins. The spot numbers on the gel correspond to the identification and mass spectroscopy data outlined in supplementary table 1.

**Figure 2.** Shown is a pie chart representation of the relative abundance of proteins in the proteome of mature *SP*- and *Jack* determined by fractional spot volume. This illustrates the suppression of storage protein accumulation in *SP*- and its partial replacement by increased lectin, Kunitz trypsin inhibitor and P34 accumulation. In contrast Bowman-Birk inhibitor that is a major sulfur sink of soybean seeds remains unchanged in *SP*- compared to *Jack* indicating selectivity of the proteins that rebalance the shortage seed storage proteins.

**Figure 3.** The distribution and abundance of total amino acids in *Jack* and *SP*- seeds is shown. Note the relative conservation of amino acid content in the *SP*- phenotype compared to *Jack* control. Of particular note is the lack of substantial change in sulfur amino acids cysteine and methionine that corresponds to the lack of change in abundance of the main sulfur amino acid containing protein, Bowman Birk inhibitor (Fig. 2).

**Figure 4.** Transmission electron microscopy of the *SP*-seed cotyledon storage parenchyma cells shows that the alternate set of proteins induced by proteome rebalancing in *SP*- are sequestered in filled protein storage vacuoles. PSV. bar = 1 μm

**Figure 5.** Scatter-plot of 3-fold up- and down-regulated transcript abundance of midmaturation *SP*- and *Jack* cotyledons assayed with Affymetrix soybean DNA array using technical and biological replicates. The data displayed represents transcript hits filtered to be present in all replicates.
Figure 6. A graphic distribution of the relative free amino content of midmaturation and mature Jack and SP- is shown. Note the divergence in free amino acid content of the midmaturation seeds narrows in upon seed maturation resulting in little difference in Jack and SP- mature seeds.

Figure 7. Log2 ratio of soybean metabolites of mid-maturation seeds and mature seeds.

7A: known metabolites in mid-maturation seeds; 7B: known metabolites in mature seeds. 1 7C: unknown metabolites in mid-maturation seeds; 7D: unknown metabolites in mature seeds. Data is sorted by log2 ratios in mid-maturation seeds on the values from the smallest to the largest. Red dot, free amino acids extracted by “EZ:faast” kit; black square, fatty acids extracted by barium hydroxide; red triangle, amino acids from total metabolites; black dot, fatty acids; black triangle, unknown fatty acids; brown dot, organic acids; blue star, phenolic acids; green dot, sugars; green dash, sugar acids; green diamond, sugar alcohols; light green star, unknown sugars; sky blue square, sterols; yellow triangle, vitamins; purple diamond, volatiles. Standard error bars were calculated on the three replicates.

Figure 8. The fluorometric analysis of GFP concentration resulting from the introgression of the GFP-hdel glycinin gene mimic (Schmidt and Herman 2008a) into the SP- background. The measurement of the GFP accumulation shows that the GFP-hdel abundance is slightly reduced when expressed in the SP-background indicating that it is not recruited to participate in the proteome rebalancing of the SP-phenotype.

Table 1. The Illumina transcriptome results of selected transcripts encoding proteins altered by the storage protein suppression in SP- along with selected additional seed proteins is shown as quantitative data. The transcript number corresponds to its number in the excel dataset (supplementary table ST3). The RPKM is adjusted transcript count that was adjusted abundance for transcript length and N-fold is Jack RPKM/SP- RPKM reflecting differential transcript abundance. The glycinin (G1-3 isoforms) and α, α’ and β subunits of conglycinin exhibit nearly complete silencing. While the transcripts of the major proteins that rebalance the storage protein shortfall, lectin, P34, and Kunitz trypsin inhibitor exhibit two-fold or less increase in transcript abundance. Transcript # corresponds to the number in the excel dataset; RPKM is adjusted transcript count, adjusted abundance for transcript length; N fold is Jack RPKM/SP- RPKM reflecting differential transcript abundance.
<table>
<thead>
<tr>
<th>Transcript #</th>
<th>Name</th>
<th>Jack RPKM</th>
<th>SP- RPKM</th>
<th>N-fold Jack/SP-</th>
</tr>
</thead>
<tbody>
<tr>
<td>1527</td>
<td>Glycinin G1</td>
<td>23402</td>
<td>133</td>
<td>-175.0</td>
</tr>
<tr>
<td>1526</td>
<td>Glycinin G2</td>
<td>20923</td>
<td>134</td>
<td>-155.0</td>
</tr>
<tr>
<td>0379</td>
<td>Glycinin G3</td>
<td>7547</td>
<td>47</td>
<td>-75.0</td>
</tr>
<tr>
<td>1525</td>
<td>Glycinin G4</td>
<td>12601</td>
<td>14459</td>
<td>1.14</td>
</tr>
<tr>
<td>644</td>
<td>Glycinin A3B4</td>
<td>6257</td>
<td>15885</td>
<td>2.55</td>
</tr>
<tr>
<td>5230</td>
<td>Conglycinin α/α'</td>
<td>8642</td>
<td>57</td>
<td>-150</td>
</tr>
<tr>
<td>1365</td>
<td>Conglycinin α</td>
<td>3277</td>
<td>81</td>
<td>-40</td>
</tr>
<tr>
<td>4119</td>
<td>Conglycinin β</td>
<td>1173</td>
<td>2971</td>
<td>2.53</td>
</tr>
<tr>
<td>5734</td>
<td>Conglycinin β</td>
<td>1244</td>
<td>6095</td>
<td>4.90</td>
</tr>
<tr>
<td>4080</td>
<td>Lectin</td>
<td>3639</td>
<td>8435</td>
<td>2.32</td>
</tr>
<tr>
<td>7124</td>
<td>KTI</td>
<td>6399</td>
<td>6478</td>
<td>1.01</td>
</tr>
<tr>
<td>382</td>
<td>KTI</td>
<td>2935</td>
<td>2031</td>
<td>-1.44</td>
</tr>
<tr>
<td>3256</td>
<td>Sucrose Binding Protein</td>
<td>2970</td>
<td>3851</td>
<td>1.29</td>
</tr>
<tr>
<td>10331</td>
<td>P34 Allergen</td>
<td>12047</td>
<td>16089</td>
<td>1.33</td>
</tr>
<tr>
<td>6408</td>
<td>BBI</td>
<td>13024</td>
<td>8228</td>
<td>-1.58</td>
</tr>
<tr>
<td>Protein Code</td>
<td>Protein Description</td>
<td>Mw</td>
<td>Mtc</td>
<td>Scale Factor</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------------------------</td>
<td>----</td>
<td>-----</td>
<td>-------------</td>
</tr>
<tr>
<td>8253</td>
<td>7S-28k Allergen</td>
<td>764</td>
<td>1418</td>
<td>1.85</td>
</tr>
<tr>
<td>9370</td>
<td>2S albumin</td>
<td>4033</td>
<td>4579</td>
<td>1.13</td>
</tr>
<tr>
<td>9393</td>
<td>2S albumin</td>
<td>3996</td>
<td>6465</td>
<td>1.61</td>
</tr>
<tr>
<td>1002</td>
<td>α-TIP</td>
<td>508</td>
<td>946</td>
<td>1.86</td>
</tr>
<tr>
<td>2622</td>
<td>P24 oleosin</td>
<td>6060</td>
<td>8251</td>
<td>1.36</td>
</tr>
<tr>
<td>2933</td>
<td>Low Mr cysteine rich inhibitor</td>
<td>2364</td>
<td>3792</td>
<td>1.66</td>
</tr>
<tr>
<td>1041</td>
<td>FAD2</td>
<td>569</td>
<td>19</td>
<td>-28</td>
</tr>
</tbody>
</table>
Comparison of the proteome of Jack and SP- seeds

**Jack**
- 1. Beta-conglycinin: 21%
- 2. Glycinin: 33%
- 3. Sucrose binding protein: 5%
- 4. Bowman Birk Inhibitor: 4%
- 5. P34: 3%
- 6. Lectin: 1%
- 7. Kunitz Trypsin Inhibitor: 1%
- 8. Other (summed): 1%

**SP-**
- 1. Beta-conglycinin: 11%
- 2. Glycinin: 4%
- 3. Sucrose binding protein: 4%
- 4. Bowman Birk Inhibitor: 9%
- 5. P34: 6%
- 6. Lectin: 11%
- 7. Kunitz Trypsin Inhibitor: 0%
- 8. Other (summed): 58%