Importance of Storage Protein in Soybeans

A white paper by

Stephen O. Opiyo[1], Daniel J. Davidson[2*], Eliot M. Herman[3] and Gladys A. Opiyo[4]

[1]Molecular and Cellular Imaging Center-Columbus, The Ohio State University, Columbus, OH 43210.

[2]Illinois Soybean Association, Bloomington, IL 61704.

[3] School of Plant Sciences, University of Arizona, Tucson, AZ 85721.

[4]Timbuk Farms, Granville, OH 43023.

*Corresponding author: Daniel J. Davidson

Phone: 309-533-8085.

Fax: 309-663-6981.

Importance of soybeans

Soybeans are grown because of the oil that can be extracted and marketed as vegetable oil or blended to make biodiesel and the high protein soybean meal by-product can be fed to livestock. Soybeans, at 13 percent moisture contain about 18 to 20 percent oil and 34 to 36 percent crude protein, though those ranges can vary considerably across varieties and climatic zones. A bushel of soybeans weighing about 60 pounds contains 44 pounds of meal, 11 pounds of oil, 3.5 pounds of hull and about 1.5 pounds of water (Figure 1).

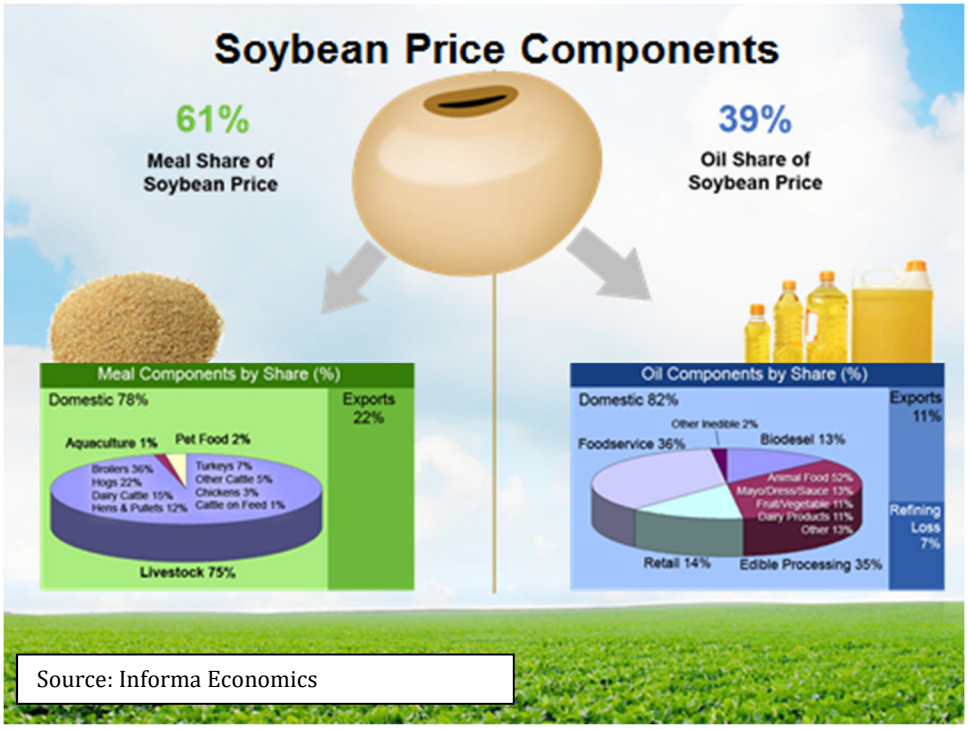


Figure 1. Marketed components in a bushel of soybeans

Soybean meal (SBM) is considered the gold standard when it comes to high protein (hi-pro) meal with a good balance of amino acid, not a perfect amino acid balance, but better than most other meal sources that come from plants. The feed industry wants a hi-pro meal, which contains 47.5 to 48 percent crude protein. Nutritionists use this value as an index of the amount of individual amino acids available in meal, which feeds directly in their formulation software.

Estimating crude protein is an easy and inexpensive measurement using NIR (Near Infrared) technology widely used for establishing price levels for meal. However it is not accurate in predicting the amino acid levels of soybean meal.

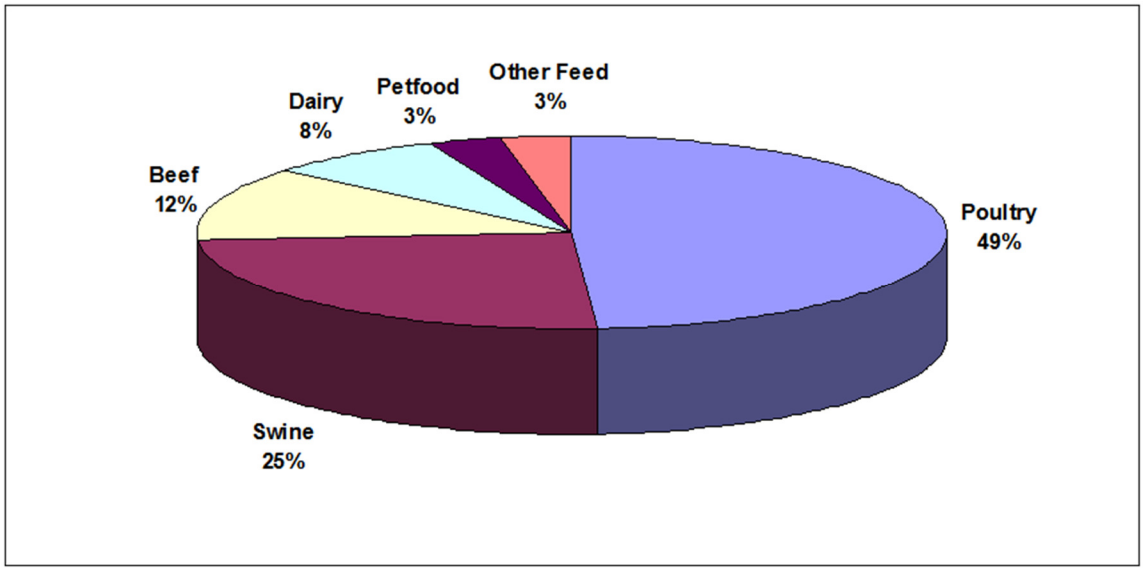
Soybean meal is traded on the basis of weight, moisture, protein, and fat. Animal growth however, is more directly related to available energy and amino acid content of soybean meal. Feed mill purchasers however often simply calculate the price per percent protein to determine the value of a soybean meal compared to other protein sources and purchase the cheapest. Meal is used to compliment the nutrient characteristics of other cheaper feed ingredients to achieve targeted nutrient levels and serves primarily as a supplemental source of essential amino acids.

The popularity of soybean meal in swine and poultry feeds is largely due to its high concentration of protein (44 to 48 percent) and its excellent profile of highly digestible amino acids. Although not a perfect blend of amino acids (it tends to be low in the sulfur containing amino acids methionine and cysteine), soybean meal is a rich source of lysine, tryptophan, threonine, isoleucine, and valine - the amino acids that are deficient in corn, grain sorghum, and other cereal grains that are commonly fed to pigs and poultry.

There is approximately 6.5% lysine in the protein of soybean meal. Other oilseed meals have less lysine in their protein. For example, the lysine percentage in canola meal protein is 5.8%, in cottonseed meal, 4.2%, in peanut meal, 3.4%, and in sunflower meal, 2.8%. However fish meal contains about 7.6% lysine.

While soybean meal is naturally high in lysine compared to other vegetable protein sources, it is not sufficient enough to meet the needs of livestock. A perfect meal would have adequate amount of all six of these limiting amino acids. However the amino acid compliment would

differ between poultry and swine. Pigs and poultry are the primary consumers of soybean meal while cattle, dairy and aquaculture consumes much smaller amounts (Figure 2).



Source: Soy Stats® 2011

Figure 2. U.S. Soybean Meal Use By Livestock 2010

The trend in the future will focus more on amino acids and less on crude protein levels. Poultry and swine have specific dietary requirements for amino acids, not crude protein. And while nutritionists use soybean meal to provide limiting amino acids in poultry and swine feeds, the relationship between crude protein level and amino acid levels quality is not accurate and estimating amino acid levels based on crude protein is considered to be inadequate by many practicing nutritionists. Lysine is the first limiting amino acid for pigs while methionine is the first limiting for poultry. It is correct that soybean protein is high in lysine and tryptophan, but compared with some other protein sources such as sunflower meal and canola meal, it is low in methionine and methionine plus cysteine. Producers don't see deficiencies in methionine if they use canola or sunflower meal in the diets but they will be short on lysine if they use those other

meal feedstock. Soybean meal has a nutrient profile that is specific to the beans that were processed (Table 1). Soybean meal contains 44 to 48% crude protein, which is an indirect measure of amino acids available in animal feed ration.

Table 1. Nutrient profile of a soybean meal sample.

Ingredient/Nutrient Code	Description	Unit	Price for Plant USBGV
100	48% Soybean Meal		16
1	Weight	Lbs	1
3	Moisture	%	10
4	Dry Matter	%	90
6	Met Energy Poultry	kcal/lb	1,109
9	Digest Energy Swine	kcal/lb	1,673
10	ME Swine	kcal/lb	1,500
11	Net Energy Swine	kcal/lb	967
12	ME Swine MCal/lb	Mcal/lb	1.5
20	Crude Protein	%	47.5
30	Lysine	%	3
31	Threonine	%	1.88
32	Tryptophan	%	0.69
33	Methionine	%	0.71
34	Meth + Cys	%	1.41
35	Arginine	%	3.67
36	Histidine	%	1.2

37	Leucine	%	3.63
38	Isoleucine	%	2.13
39	Phenylalanine	%	2.36
40	Phenylala + Tyrosine	%	4.07
41	Valine	%	2.47
50	Crude Fat	%	0.9
57	18:2 Linoleic Acid	%	0.54
65	Crude Fiber	%	3.4
84	Calcium	%	0.26
85	Phosphorous-Total	%	0.64
86	Phosphorous-Available	%	0.16
88	Sodium	%	0.01
89	Chloride	%	0.04
90	Potassium	%	2.13
91	Magnesium	%	0.3
92	Sulfur	%	0.44
95	Copper	mg/lb	9.23
97	Iron	mg/lb	59.5
98	Manganese	mg/lb	16.9
99	Selenium	mg/lb	0.0454
100	Zinc	mg/lb	25.9
103	Vitamin E	IU/lb	1
106	Choline	mg/lb	1251

Source: Nick Bajjalieh, Integrative Nutrition, Inc.

Seed proteins

Seed proteins can be broadly classified into two categories, housekeeping proteins and storage proteins. The housekeeping proteins are responsible for maintaining normal cell metabolism. The storage types provide proteins (nitrogen and sulfur source) required during germination and establishment of a soybean seedling, and triglycerides and carbohydrate reserves, which are used as a source of carbon and ultimately energy (Herman and Larkins 1999 for review). In addition, seed proteins are also classified based on their solubility as follows (Osborne 1924).

- Albumins: Water extractable (1.6S-2S);
- Globulins: Extractable in dilute salt solutions (7S -13S);
- Prolamins: Extractable in aqueous alcohol;
- Glutelins: Most difficult to solubilize; Extractable by weakly acidic or alkaline or dilute SDS solution

Albumins and globulins comprise the storage proteins of dicots (e.g. pulses including soybeans), whereas prolamins and glutelins are major proteins in monocots (e.g. cereals)(Derbyshire and others 1976). The seed storage proteins can be distinguished from other housekeeping proteins by some of their characteristics, i.e.; they accumulate in high amounts in the seed during mid-maturation stage of seed development and are used up during germination; they are synthesized only in the seed (in cotyledon or in endosperm) and not in other tissues; they lack any other functional activity besides storage; and they are deposited mostly in special storage organelles called protein bodies. Storage proteins are generally rich in asparagine,

glutamine, arginine or proline. The legume proteins (mostly 7-11S globulins) tend to be deficient in the sulfur containing amino acid cysteine and methionine and also tryptophan (Higgins 1984).

Soybean seed storage protein

Soybean seed storage proteins are encoded by conserved gene families broadly distributed in plant species (Harada and others 1989, Nielsen and others 1989, Schuler and others 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 prolamins and to a lesser extent the legumins are abundant (Herman and Larkins 1999). Soybean is an important source of high quality proteins and oils for both human consumption and animal feed. The storage protein concentration of commercial soybean cultivars calculated on a dry weight basis ranges from 37 - 42% depending on the genotype, location, and growth conditions (Krishnan and others 2007).

Soybean seed storage proteins contain approximately 88 proteins, and each falls into four broad categories, albumins, globulins, prolamins, and glutelins (Agrawal and others 2008). It is important to note that this characterization of proteins into broad categories is not an identification of the included proteins but merely a definition of comparative solubility of groups of proteins in an acid / base or salt solutions. Within the classification if proteins, for instance as an albumin there are many different types of proteins. Of the 88 proteins, two major storage proteins are glycinin and β-conglycinin, which fall into the globulin category. The other seed storage proteins (86 in number) only account for a minor portion of the total seed protein content since globulins can be up to 80% of the total seed protein (Higgins 1984).

Some of these less abundant proteins have significant impact on the composition and/or use of soybean protein. For instance about half of the seed's inventory of sulfur-containing amino acids are in the form of protease inhibitors that are considered anti-nutritional factors in animal feed applications. Protease inhibitors inhibit the activity of proteolytic enzymes, can cause a decrease in digestive efficiency and a reduction in dietary sulfur amino acid. In poultry and swine, trypsin inhibitors (one protease inhibitor) significantly reduce the digestibility and utilization of amino acids

Other significant storage proteins are classified as allergenic because of its allergen content and are included in the eight food allergens regulated under the US Food and Allergen Labeling Act (see Herman EM, Burks AW (2011). The glycinin, also called 11S protein, consists of acidic and a basic polypeptides which are linked by a disulfide bridge. β-conglycinin, also called 7S protein consists of three subunits (α'-,α-,and a- β subunits)(Tsubokura and others 2012). The amino acid compositions of glycinin and β-conglycinin vary sufficiently to affect their nutritional value in animal diets. Glycinin contains more of the limiting amino acids cysteine and methionine, than β-conglycinin (Tsubokura, Hajika, Kanamori, Xia, Watanabe, Kaga, Katayose, Ishimoto and Harada 2012). Table 2 shows the percentage of amino acids found in glycinin and β-conglycinin (Prakash and Rao 1986). Based on the amount of amino acids in a kilogram of crude protein, the primary limiting amino acid in soybeans is methionine, and its average concentration ranges from 10.7 to 12.6 g/kg of crude protein (Serretti and others 1994).

The poultry and pig industry is supplementing synthetic crystalline methionine to soybean meal in animal rations to overcome the problem. However, this causes increase in the cost of meal and also during the processing of the meal leaching of methionine takes place and bacterial degradation with the formation of undesirable, volatile sulfides (Clarke and Wiseman 2000). The

development of a soybean variety with increased methionine content will overcome the problem of low methionine.

Table 2. Amino acid of 11S (glycinin) and 7S (β-conglycinin) in soybean seed protein. Residues per 100,000 g of protein[1].

Amino acid	11S (glycinin)	7S (β-conglycinin)
Aspartic acid	106	110
Threonine	44	17
Serine	74	47
Glutamic acid	169	144
Proline	50	42
Glycine	64	33
Alanine	47	33
Valine	43	39
Methionine	9	2
Isoleucine	45	40
Leucine	56	63
Tyrosine	24	23
Lysine	33	45
Arginine	45	53
Tryptophan	7	-
Phenylalanine	34	41
Histidine	17	15

Half cystine	7	-

[1]Prakash and Rao 1986

Amino acids are the principal building blocks of enzymes and other proteins. Amino acids have been divided into two groups, essential and non essential. Non-essential amino acids are readily available or can be synthesized by animals, and essential amino acids cannot be synthesized in animals. Essential amino acids are Lysine (Lys), Histidine (His), Leucine (Leu), Isoleucine (Ile), Valine (Val), Methionine (Met), Threonine (Thr), Tryptophan (Trp) and Phenylalanine (Phe) (D'Mello 2003). In addition, the amino acids arginine, cysteine, glycine, glutamine, histidine, proline, serine and tyrosine are conditionally essential, meaning they are not normally required in the diet, but must be supplemented to specific populations that do not synthesize it in adequate amounts.

In pigs and poultry cysteine is an important amino acid. However, cysteine can be synthesized by the animals using methionine as a substrate, so it is considered a semi-essential. If a diet has excess methionine, animals can synthesize cysteine. However if the diet is deficient in methionine, they synthetic cysteine must be added to the formulation. For monogastric animals like pigs and poultry, there will be a requirement for methionine or methionine plus cysteine. So while the animals have a requirement for cysteine, nutritionists only calculate the total for methionine and cysteine and then assume the animals can convert methionine to cysteine if needed.

Development of seed protein in soybean seed

Storage proteins are synthesized in the rough endoplasmic reticulum located in the cytoplasm

and then transported to the storage organelles. Once a seed sets the stage for accumulation of reserves there follows stages with high metabolic activity yielding an increase in dry weight, proliferation of protein and lipid bodies, increase in size, and in the final stage seeds lose moisture and the amount of starch declines (Wilson 1987). During development and maturation of seeds, reserve protein is deposited as protein storage vacuoles in cotyledon cells (Tombs 1967). Protein storage vacuoles in soybean seeds sequester most of the protein of the mature seed. Protein storage vacuoles are formed by the deposition of reserve protein in the central vacuole, that then subdivides into small protein storage vacuoles differentiating the central vacuole into individual vacuoles. (Adams and others 1985).

Adams and coworkers reported that formation of small vacuoles from the central vacuole of cotyledon cells is an early event in soybean seed development. Filling small vacuoles with protein, which then forms protein bodies. According to Adams and coworkers, early subdivision of central vacuoles seventeen days after flowering appears to be the source of protein storage vacuoles that are subsequently filled with storage proteins. In this regard soybean protein storage vacuole appear to have a different pattern of origin compared to the common pea, another well described model where the protein-filled subdivisions of the vacuole appear to be completely filled (Adams, Norby and Rinne 1985) (there is a lot of additional literature- see for review Herman EM and Larkins BA (1999).

Developing improved high protein concentration in soybean seeds

Several processes are involved in the control of seed protein accumulation. These include biosynthesis of amino acids and organic acids in source leaves, control in phloem loading of transported amino acids, uptake of amino acids by the embryo mediated by amino acid

transporters, and enhanced provision of organic acids for amino acids and interconversion of amino acids (Rainbird and others 1984).

During seed fill, the accumulation of seed storage proteins is regulated by an integrated genetic and physiological network (Brocard-Gifford and others 2003, Elke and others 2005, Fait and others 2006, Golombek and others 2001 for examples). However, the tight genetic control of seed development is modifiable within a certain range by the plant nutritional and environmental effects (Weber and others 2005). The composition of seeds can be loosely defined as a developmental genetic program that is modified by nutrient source availability; environmental stresses impact availability and the demands of the forming storage substance sink.

Viewed from a systems biology perspective, the regulation and cross-talk between an 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.

The seed's genetic program specifies a series of stage-specific gene expression patterns and regulons that provide a determinant framework for the formation and development of the seed (Goldberg and others 1981ab, 1983, Harada and others 1989, Hill and Breidenbach 1974, Mienke and others 1981, Naito and others 1988 Nielsen and Nam 1999, Perez-Grau and Goldberg 1989, Walling and others 1986, To and others 2006). Seed-specific DNA binding proteins have been discovered using sequence-specific probes and promoter-deletion analysis with reporter gene constructs (Allen and others 1989, Baumlein and others 1992, Chen and others 1986, Kwong and others 2003, Lessard and others 1991, Jofuku and others 1987, Wang

and others 2007). A primary role of transcription factors in seed development is the primary control of protein and oil and possibly other classes of storage substance accumulation (Kroij and others 2003, Gutierrez and others 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 and others 2007, Kroij and others 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 and others 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 and others 2005) and triglyceride content (Baud and others 2007, Cernac and Benning 2004, Maeo and others 2009, Li and others 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 and others 2005, Gu and others 2010).

Sulfur availability is one example of nutrient source regulating seed protein composition (Beach and others 1985). Legumes such as soybean possess storage proteins with low sulfur

amino acid content (cysteine and methionine). The beta subunit of conglycinin exhibits plasticity in response to available sulfur (Hagen and others 2003, Hirai and others 1995, Holowach and others 1986, Tabe and others 2002). There is also feedback regulation of protein filling in the seed sink in response to nitrogen availability (Biermann and others 1998, Ohtake and others 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. 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 and others 2005).

Because the stored seed sink will become a nutrient source in the following life-cycle with the germinating seed, 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 and others 1991; Hunter and others 2002), and soybean (Kinney and others 2001, Takahashi and others 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. The variability in protein and

triglyceride content in the seed, within a specific genetic makeup is maintained within relatively narrow limits of about +/-3 to 5% w/w. Seeds have 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 plasticity that serves to maintain a relatively defined ratio of storage substances and composition in the mature seed

In improving soybean protein, emphasis has primarily been focused on increasing protein content, increasing essential amino acids in soybean proteins, and altering 11S/7S globin ratio. Increasing 11S ratio improves the levels of sulfur amino acids in soybean protein. Soybean storage proteins are rich in asparagine, glutamine and arginine or proline. On the other hand, methionine is the most limiting amino acids in soybean storage proteins (Tsubokura, Hajika, Kanamori, Xia, Watanabe, Kaga, Katayose, Ishimoto and Harada 2012). Therefore, one of the focuses on the developing high protein concentration in soybean seeds should be on improving methionine content of soybean seeds.

Elevating the methionine content of endogenous storage protein can increase methionine content. For example, when Brazil nut 2S albumin with 18% methionine was introduced in soybeans, it significantly increased the quality of seed protein (Müntz and others 1998). A Brazil nut, methionine-rich protein successfully increased the percentage of methionine in soybean by 26% (Nordlee and others 1996). However, Brazil nut has an allergen protein. In another

approach, expressing methionine-zein from maize increased methionine in soybean seed (Dinkins and others 2001).

There have been other attempts to use less controversial proteins to achieve the same goals such as prolamine zeins from maize by inserting maize gene into soybean seed. While it has been shown that zeins are produced in soybean and other dicot seeds and correctly assembled in ER-derived protein bodies the level of the heterologous protein accumulation even with strong seed promoters has been insufficient to greatly alter the overall sulfur content of soybeans. There are two distinct problems with this approach that have emerged, first the allocation of carbon and nitrogen in a seed appears to be largely proscribed and as a consequence when a new gene is introduced it will have to compete the for nutrient flux. As a result the foreign gene appears not be favored and the accumulation of its product is limited. One technical solution appears to be to have the foreign gene product mimic an endogenous gene thereby fooling the seed to accumulate more of that protein under the guise that it is a protein that would normally be produced in large amounts. Schmidt and Herman (2008), Herman EM, Schmidt MA (2011).

Proteome rebalancing in soybean seeds can be exploited to enhance foreign protein accumulation. Schmidt and Herman (2008) showed that this strategy can be used to increase foreign protein accumulation many fold although the researchers tested the ability of this technique to alter protein abundance and have not yet shown that it will also alter amino acid balance. Furthermore, Scientist from Pioneer Hi-Bred International generated soybean plants that have higher amounts of essential amino acid (Pioneer Hi-Bred International). The 11S and 7S globulins tend to be deficient in the sulfur-containing amino acid methionine and cysteine, and also tryptophan. The 11S and 7S proteins differ in their nutritional properties; increasing 11S and reducing 7S protein in soybean can improve the nutritional quality of soybean proteins

(Kitamura 1995). Krishnan and Nelson reported that high protein concentration is associated with the accumulation of specific subunits of the 11S globulin glycinin (Krishnan and Nelson 2011). Mutants that lack 7S proteins in soybeans contain higher protein bound sulfur amino acids (Kitamura 1995). In general, 11S contains 3-4 times more methionine per unit protein than 7S (Kitamura 1995).

Soybean protein is deficient in sulfur amino acids, therefore, altering 11S/7S protein ratios should improve the essential amino acid composition of soybean proteins. Genetic accumulation through breeding can affect changes in 11S and 7S composition. For example, Kwanyuen and coworkers reported that 11S protein from cultivar *Prolina* contained approximately 9.0 kg more methionine, cysteine and lysine than 11S protein from cultivar *Dare*, which represented 1.6 fold increase (Kwanyuen and others 1997). Furthermore, cysteine plus methionine account for 30-45 g/kg of amino acid residues in 11S (Nielsen and others 1989), and sulfur amino acids account for less than 10 g/kg of amino acid in 7S (Harada and others 1989). Similarly changes in protein composition can be easily achieved in elite soybean lines by directly altering gene expression patterns using biotechnology approaches. For instance 7S deficient soybeans have been produced (Kinney and others 2001) that result in creating a proteome dominated by glycinin that seed over-produces to compensate of the 20% of the total protein lost when the conglycinin is suppressed. Similarly if both storage proteins are suppressed the seed again compensates to maintain protein content, albeit with a quite different proteome (Schmidt and others 2011).

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 (Schmidt and others 2011). One of the significant aspects of silencing one or more storage proteins is that in rebalancing the seed's protein content the amino acid composition is

barely altered even when as much as two thirds of the total seed proteins are altered. This indicates that the seed's biology and regulatory mechanisms are highly determinant with respect to amino acid content and that the seed strives to maintain a normalized amino acid balance even if a large fraction of the total seed proteins are altered.

Increasing the levels of free amino acids could be an alternative approach in changing the content of seed storage protein. The pool of any free amino acid is less than 0.05% of the protein-bound quantity and the size of free amino acids is tightly regulated (Bright and Shewry 1983). This level has been increased somewhat by stacking mutant lesions of storage protein nulls raising the asparagine content in particular, but even so, free amino acids are still only a tiny fraction of those sequestered in proteins. The fundamental problem with this approach is that free amino acids need to be stored in compartment where the elevated content of charged molecules will not perturb other solute considerations such as pH and organic and inorganic ions. Even uncharged molecules such as sucrose are specifically sequestered in intracellular compartments.

In most instances the vacuole with its large volume and as physiological distal compartment is the location of choice, examples being both sucrose accumulation in many plants and other substances such as betaine, *N,N,N*-trimethylglycine, accumulated by plants such as sugar beets. Whether it is feasible to induce the protein storage vacuoles to accumulate free amino acids at high levels remains to be determined. There may be some reason to believe this may be more feasible than might be possible with vegetative cells.

The post-germination mobilization of storage proteins by in situ proteolysis results in at least the transient accumulation of massive amounts of free amino acids that are subsequently transferred from the vacuole, processed into transport amino acids such asparagine and then

moved to the growing seedling for use. This suggests that protein storage vacuoles could tolerate and be optimized to store free amino acids if a means were to be developed to over-produce free amino acids and move the molecules into the vacuole. Possible approaches to increase biosynthesis of free amino acids include altering the regulatory of the enzymes or introduction of unregulated enzyme with the same catalytic specificity. Most of the work has been done on the aspartate biosynthesis pathway, which produces lysine, threonine, and methionine, and on the aromatic pathway, which produces tryptophan, tyrosine and phenylalanine (Bright and Shewry 1983).

The average concentration of methionine, cysteine and threonine in soybean seed range from 10.7 to 12.6, 12.4 to 13.7 and 40.0 to 42.0, respectively (Serretti, Schapaugh and Leffel 1994). As mentioned before, methionine is the primary limiting amino acid in soybeans. The most promising approach to improving methionine content is to introduce genes from foreign plants that code for high methionine levels. Genes encoding high levels of methionine includes 2S albumins from Brazil nut (Altenbach and others 1987), and *Helianthus annuus* (Kortt and Caldwell 1990). There have been a number of studies on other dicot seeds where the flux of amino acids is altered either by perturbing the amino acid biosynthesis pathway and/or altering transporters. The effect of these alterations have been shown to alter protein composition in other seeds and that these experiments need to be duplicated and extended in soybean.

Breeding

Increasing seed protein concentrations in soybean has been the objective for many soybean-breeding programs. However the struggle has been that increasing protein leads to significant reductions in oil and or yield. As breeding stock lines have been identified that are null for

storage proteins, anti-nutritional proteins, and allergens that together constitute a rich toolbox from which breeders could conceivably stack traits to produce higher protein and higher quality beans optimizing the best of soybean's proteins while limiting the presence of the least desirable proteins. Because each cross brings with it the entire genome of the parental lines the problem with breeding some of these rare lesions is to isolate those traits and then incorporate this into more elite lines suitable for US agriculture methods and expected productivity. Various populations structures, including biparental, 3-way, backcross, and inter-mated populations have been evaluated for their effectiveness in producing progenies with increased seed protein content.

Selections have been based on inbred line means or evaluated for single plants. Traditionally, an individual with the highest level of seed protein concentrations is included in the base populations. Recurrent selection has been effective in increasing seed protein content in soybean (Miller and Fehr 1979). Brim and Burton (Brim and Burton 1979) realized average gain in seed protein content in two selection program using a base population derived from a cross between a high protein, low oil, low maturity line and a low protein, high oil, early maturity line. Thorne and Fehr (Thorne and Fehr 1970) crossed three exotic plant introductions with average seed protein content of 466 g Kg^{-1} with two high yield parents with average seed protein content of 423 g Kg^{-1}. Similarly, Xu and Wilcox (Hu and Wilcox 1992) selected single S0 from recurrent selection population formed by a combination of high protein line and two F2 populations segregation for male sterility. Average gain for seed protein content was 8.0 g Kg-1 per cycle after five cycles of selection.

All cited studies above for high protein soybean protein included either adapted or unadapted high-protein lines in the base population. In the soybean breeding program at the University of

Nebraska-Lincoln, a population known as UP3 protein was created using 100% adapted germplasm (Opiyo 2003). The parents were either cultivars or high yielding lines, none of which were high protein sources. Mean seed protein content increased 9.6% from 400 g Kg^{-1} for parents of Cycle 0 to 438.3 g Kg^{-1} for parents in Cycle 3 of recurrent selections (Opiyo 2003).

Marker-assisted selection increases the heritability of selectable traits, and increases the rate of genetic gain, by decreasing the variance of the selected trait. Marker-assisted selection is one of the tools used by breeders to increase protein levels in soybean (Brim and Burton 1979). Genetic control of seed protein is inherited in a quantitative manner. Many quantitative trait loci (QTLs) associated with seed protein have been identified in soybean (Fasoula and others 2004);(Panthee and others 2005). In addition, Brummer and coworkers identified QTL affecting both protein and oil contents in soybean seeds (Brummer and others 1997).

Genomics

The recent sequencing of soybean genome (Schmutz and others 2010), together with the availability of tools for new genomic, transcriptomic, proteomics, and metabolomics, offers new possibilities for increasing seed protein content in soybean. Genomics is the study of the structure and function of the genomes of a living organism (Baer and others 1984). The genomics study can help in the development of DNA markers (molecular markers) for molecular characterization and relationship with soybean seed protein content using genetic linkage maps and QTL (quantitative trait loci) analysis. The populations typically assayed in the linkage studies include F1, F2 or BC1, with high diversity and low levels of linkage disequilibrium. DNA markers technology that has been used in soybean seed protein content includes RFLPs (restriction fragment length polymorphism; (Lee and others 1996), RAPDs (random amplified

polymorphic DNAs; (Diers and others 1992); (Chung and others 2003), SSR (simple sequence repeat; (Panthee, Pantalone, West, Saxton and Sams 2005), (Jun and others 2007)), ESTs (expressed sequence tags; (Mooney and others 2004)), and SNPs (single nucleotide polymorphisms; (Phansak 2010)). Thus, the field of genomics will contribute to the increase of seed protein content in seed.

Transcriptomics

The discipline of transcriptomics deals with the analysis of transcriptome. Transcriptome of plants are the complete list of all types of RNA molecules expressed in cells, tissue, or all organism (Zhang and others 1997). Unlike the genome, which is fixed for a give organism, transcriptome vary with environmental conditions. The transcriptome reflects the genes that are being expressed at any given time under specific conditions. Post translational RNA modifications might have functions beyond the structure and function of RNA (epigenetics; (Goto and Nakayama 2012)). Many of the epigenetics mechanisms are critical to protein function. Non-coding small RNAs play known regulatory roles in seed protein synthesis by affecting gene expression through gene silencing (Gingeras 2009). Transcriptomics analysis, which uses cDNA biochips (microarrays) to analyze global gene expression profile in developing soybean (Asakura and others 2012). Increasing deep-sequencing using total transcriptome assay is replacing chips. As sequencing costs continue to decrease and with the advantage of providing direct counts of individual transcripts sometimes differentiating among gene family members sequencing if cost effective have an advantage over the qualitative analysis with chips that require an averaged hybridization conditions. If sequencing is done deeply enough (more

genome coverage) these assays have a marked statistical advantage over averaging duplicate or triplicate chip assays.

RNA-silencing has been used as a tool to uncover gene function and engineer novel traits in soybean ((Schmidt and others 2011);(Arase and others 2012)). Schmidt and coworkers, used gene silencing to silenced storage protein 11S and 7S in soybean seeds. Their gene silencing resulted in a rebalanced protein composition preserving seed protein content without major collateral changes in the transcriptome (Schmidt, Barbazuk, Sandford, May, Song, Zhou, Nikolau and Herman 2011). Lastly, RNA-seq data from next generation sequencing are now being used to identify SNP based on short read sequences taken from protein coding regions. RNA-seq approach is beneficial because SNP markers identified via RNA-seq are specific to the genetic materials of interest; and RNA-seq data may be mined for transcriptional differences on genetics alterations between high and low seed protein varieties and that may identify genes that differentiate the varieties.

Proteomics

Proteomics is the study of the proteome, which is the complete list of all proteins expressed in a cell, tissue, or whole living organism and expressed by genome (Pandey and Mann 2000). Proteins are the active molecules involved in metabolic (enzyme), structural, functional, storage and transport processes. Proteins are assembled from 20 amino acids using information encoded in genes (DNA). Proteomic analysis was first used in soybeans to analyze the consequences of suppression of the immunodominant soybean alleegen P34. Proteome maps showed that suppression of the minor abundant P34 protein although immunologically dominant did not result in collateral changes in the proteome (Herman and others 2003). Schmidt and Herman

used proteomic approach to study rebalancing in soybean seeds proteome (Schmidt and Herman 2008). Similarly, Mooney and coworkers used high-throughput peptide mass fingerprinting of soybean seed proteins to create automated workflow and utility of expressed sequence tag databases for protein identification useful to the soybean community. (Mooney, Krishnan and Thelen 2004). Furthermore, Hajduch and coworkers used two-dimensional electrophoresis to determine the expression profile and identity of hundreds of proteins during seed filling in soybean (Hajduch and others 2005).

Recent methodology developments to move from labor intensive initial gel fractionation and spot picking to LC/MS where complex mixtures in solution can be digested with protease and the resulting fragments fractionated by liquid chromatography and analyzed by mass spectroscopy hold the promise of far less expensive high throughput proteome analysis that will be useful for screening seed collections whether accessions in collections of those produced by conventional and transgenic procedures. Combined with technologies such as automated chippers already in use for high throughput lipid analysis there is the potential to apply similar technology to proteins to produce protein profiles from hundreds or thousands of individual seeds. Other potential emerging technologies is NMR, commercial machines that are inexpensive enough for individual laboratories can assay protein and oil. Adapting these machines to robotics could allow these machines to perform an initial NMR screen of individual whole seeds, identify those most promising and then move to a secondary tier screening using LC/MS technology of seed chips. Taken together these technologies could be useful for both breeders and biotechnologists to select the most promising lines and for academics, to advance their project goals at a rate similar to the largest biotechnology firms. It would take only a modest investment on the part of the commodity boards to create one or more seed-phenotype centers with

appropriate analysis capability that although primarily oriented to soybean might prove equally advantageous to other commodities.

Metabolomics

Metabolomics is the study of the molecules involved in metabolism (metabolites) in plants and living organisms (Weckwerth and Fiehn 2002). Gas or liquid chromatography coupled with mass spectrometry (GS-MS or LS-MS, respectively) is the most important tools for non-targeted metabolite analysis (Shi and others 2011). Metabolomics analysis is now recognized as a crucial component of functional genomics in plants. Bally and coworkers recently use metabolomics analysis to study the accumulation of recombinant proteins in the plastid genome of transgenic tobacco (Bally and others 2011). Furthermore, the change in amino acids levels during seed ripening in peach was identified using metabolomics analysis (Lombardo and others 2011). As mentioned before, Schmidt and coworkers, used gene silencing to silence storage protein 11S and 7S in soybean seeds. However, their gene silencing resulted in a rebalanced protein composition preserving seed protein content without major collateral changes in the metabolome (Schmidt, Barbazuk, Sandford, May, Song, Zhou, Nikolau and Herman 2011). The later study shows the power of a combined systems approach in which changes in composition, whether through conventional or biotechnology means, can be analyzed downstream by proteome, transcriptome, and metabolome in the context of trait. Such comprehensive analysis then forms a database to design next and further generation changes that can step by step be structured in a strategy to produce the desired output product.

Many of the compounds that are considered diet healthy in soybean are in classes of metabolites. Current technologies using high-throughput HPLC analysis is capable of screening

these compounds if combined with automated chipping technology. Much like large-scale high throughput proteome analysis all of the required technology exists and indeed for metabolome this is already a reality for molecules such as the lipodome by industry. To make best use of metabolomics there first needs to be an expansion of studies of the metabolic consequences in seed biology of not only collections, breeding stock, and transgenics but also the metabolome is responsive to growing conditions, Understanding how specific field conditions impact the metabolome can be a fundamental database to create strategies to control the allocation of the plants source to produce specific seed sinks of protein and oil. Most significantly metabolomics plays a crucial role in understanding how to control and enhance the essential amino acid inventory in soybean.

Recommendations for Research

Create a new multidisciplinary integrated programs. To redirect some of the resources currently being expended or to create new resources into an interlinked program of basic research and targeted output goals to build toward creating new value-added soybeans. Such industry funded programs might mimic the former USDA NRI CAP programs of integrated research, extension, training all toward specific goals such as better, or more protein, enhanced and better oil content, etc.

- Create seed phenomics center(s). A number of plant phenomic centers have been constructed to create a high throughput analysis of greenhouse grown plants. A similar approach can be taken with seeds although a seed phenomics center would more closely resemble a biotechnology facility with robotic compositional analysis.

- Initiate a dialog with industry and state/national soybean boards to create one or more academic seed-phenomic centers that would have the robotic capacity for chipping and analyzing large numbers of seeds and provide down stream compositional analysis for proteome, lipodome, and metabolome at costs reasonable enough to be used by breeders as well as biotechnologists. Such a phenomic center would likely cost no more than one of the large genome projects that the US soybean industry has subsidized.
- Promote basic science research to understand how seed reserve accumulation is controlled. Conduct projects aimed at understanding how soybeans control the balance of stored reserves and to use this data to create new very high protein or very high oil beans targeted for specific industry end uses.
- A long-term goal of "super-beans". To create one or more varieties of a super-bean that is highly optimized for animal feed uses with varieties specifically targeted at large end-users such as pork, poultry, or aquaculture that have different specific requirements for stacked traits that not only include protein traits but also specific lipid profiles, micro-nutrients, and potentially also including species-specific oral vaccination. As part of this strategy for an optimized bean would be to create varieties that require less processing and energy input for example beans null for significant anti-nutritional proteins stacked with other traits.
- Use science to enhance the economic viability of soybean producers. To develop new and better Winter crops for annual rotation with soybean so that soybean farms can be productive for most of the year creating added economic viability for farmers and increasing global food/feed output.

Conclusions

Soybean seed storage proteins contain approximately 88 proteins. Of the 88, glycinin and β-conglycinin, contribute to overall total seed storage proteins. However, the amino acid compositions of glycinin and β-conglycinin vary sufficiently to affect their nutritional value in animal diets. Glycinin and β-conglycinin are deficient in sulfur amino acids; therefore, there is a need to understand these storage proteins in order to improve sulfur amino acids soybean proteins. The trend in the future should focus more on improving sulfur amino acids and less on storage protein levels.

There are new challenges and opportunities in applying 'omics' sciences (including genomics, transcriptomics, proteomics, and metabolomics) in the development of efficient marker assisted selection strategies in soybeans to increase seed protein content and essential amino acids. These new technologies offer new challenges and opportunities for soybean breeders in the post genomic era. For example, from Schmidt and coworkers study, where RNA gene silencing was used to silence two major soybean seed proteins; there seems to be an intrinsic process that evaluates the progress of protein filling during seed development and can alter the mix of protein synthesized to rebalance the system to produce mature seed with correct protein and amino acid in the silenced seed storage proteins.

It is often said about most things the more one knows the more one discovers one doesn't know, and so it is with how seed's produce proteins, oil and starch. So much has been learned about how this process occurs but with this it is apparent so much needs to be learned about how to alter this process toward desired end goals of composition. The rapid increases in systems biology databases from omics assays now being extended to mutants, transgenics, An initial goal should be to identify genes that are involved in the control of seed protein content that can

be probes by analyzing both transgenics with introduced lesions and by analysis of seeds created by breeders with conventional germplasm. Analyzing for commonalities between the transgenics and with breeder's lines may distill out those common factors that are the primary regulators of protein content. Thus, omics technology could help in identifying those genes in transgenics, mutants, or breeder's lines. There is no doubt that next generation sequencing could help in identifying genes involved in the increase of seed storage proteins and essential amino acids for animal nutrients.

Cost for sequencing is decreasing rapidly, so much so a complete human genome is available for 1000 dollars is almost in reach that may revolutionize individual medicine. Low cost sequencing will enable breeders to use sequencing as a selection tool while all researchers can also use sequencing as an analysis tool. The growing databases will enable others using bio-informatic tools to merge datasets from breeding, mutant, and transgenic lines to discover common features and new approaches to enhance seed value. Next generation sequencing will not solve the immediate problem of the deficiency of essential amino acids in soybean meal. However, this is a new technology, and the time it takes to bring a product into the market using this technology is at least 8-10 years. Even with identification of genes either in conventional lines or by transgenic modification the reality is that it will be a decade or more to use this information, produce and assess new lines, and to have the seeds available for widespread deployment as an agricultural product.

All the technologies to enhance soybeans require the same long term view of investment, development, and hopefully deployment. Fortunately the soybean industry and its members have been at the forefront of making their own long-term investments in technology and their investments have been further leveraged by federal and state government research support. The

past support of the soybean community to developing knowledge as well as products with the understanding that one leads to the other is an shining example for all of the commodities. Documents such as this one should be discussion points for future directions for the soybean communities support, both financial and political to define what knowledge is needed to create the products that its members will need to meet the world's needs for food and feed. Because of the lead time involved it is critical that research occur constantly building on prior work and training new investigators that bring the best evolving concepts and approaches to further develop tomorrow's crops.

REFERENCES CITED

Adams CA, Norby SW and Rinne RW. (1985). Production of Multiple Vacuoles as an Early Event in the Ontogeny of Protein Bodies in Developing Soybean Seeds. Crop Science 25:255-62.

Agrawal GK, Hajduch M, Graham K and Thelen JJ. (2008). In-depth investigation of the soybean seed-filling proteome and comparison with a parallel study of rapeseed. Plant Physiol 148:504-18.

Allen RD, Bernier F, Lessard PA and Beachy RN. (1989) Nuclear factors interact with a soybean beta-conglycinin enhancer. Plant Cell 1:623-31.

Altenbach SB, Pearson KW, Leung FW and Sun SSM. (1987). Cloning and sequence analysis of a cDNA encoding a Brazil nut protein exceptionally rich in methionine. Plant Mol Biol 8:239-50.

Asakura T, Tamura T, Terauchi K, Narikawa T, Yagasaki K, Ishimaru Y and Abe K. (2012). Globalgeneexpressionprofiles in developingsoybean seeds. Plant Physiology and Biochemistry 52:147-53.

Arase S, Kasai M and Kanazawa A. (2012). In planta assays involving epigenetically silenced genes reveal inhibition of cytosine methylation by genistein. Plant Methods 8:10.

Pandey A and Mann M. (2000). Proteomics to study genes and genomes. Nature 405:837-46.

Schmidt MA and Herman ME. (2008). Proteome rebalancing in soybean seeds can be exploited to enhance foreign protein accumulation. Plant Biotechnology Journal 6:832-42.

Baer R, Bankier AT, Biggin MD, Deininger PL, Farrell PJ, Gibson TJ, Hatfull G, Hudson GS, Satchwell SC, Seguin C and et al. (1984). DNA sequence and expression of the B95-8 Epstein-Barr virus genome. Nature 310:207-11.

Bally J, Job C, Belghazi M and Job D. (2011). Metabolic adaptation in transplastomic plants massively accumulating recombinant proteins. PLoS One 6:e25289.

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.

Baumlein H, Nagy I, Villarroel R, Inze D and Wobus U (1992) *Cis*-analysis of a seed protein gene promoter—the conservative RY repeat CATGCATG within the legumin box is essential for tissue-specific expression of a legumin gene. Plant J. 2:233–239.

Beach, LR, Spencer D, Randall PJ and Higgins TJV (1985) Transcriptional and posttranscriptional regulation of storage protein gene expression in sulfur-deficient pea seeds. Nucl. Acids Res. 13:999–1013.

Biermann BJ, Debanzie JS, Handelsman J, Thompson JF, Madison JT (1998) Methionine and sulfate increase a Bowman-Birk-type protease inhibitor and its messenger RNA in soybeans. J Agric Food Chem. 46:2858-2862.

Boerma HR. (1996). RFLP loci associated with soybean seed protein and oil content across populations and locations. Theor Appl Genet 93:649-57.

Bright SW and Shewry PR. (1983). Improvement of protein quality in cereals. Crit Rev Plant Sci 1:49-93.

Brim CA and Burton JW. (1979). Rucurrent selection in soybeans: II. Selections for increased percent protein in seeds. Crop Science 19:494-98.

Brocard-Gifford IM, Lynch TJ, Finkelstein RR (2003) Regulatory networks in seeds integrating developmental abscisic acid, sugar, and light signaling. Plant Physiology 131:78-92.

Brummer CE, Graef LG, ORF J, WIlcox JR and Shoemarker CR. (1997). Mapping QTLfor Seed Protein and Oil Content in Eight Soybean Populations. Crop Science 37:370-78.

Clarke EJ and Wiseman J. (2000). Developments in plant breeding for improved nutritional quality of soya beans I. Protein and amino acid content. Journal of Agricultural Science 134:111-24.

Cernac A, Benning C (2004) WINKLED1 encodes an AP2/EREB domain involved in the control of storage compound biosynthesis in Arabidopsis. Plant J. 40:575-85.

Chen ZL, Schuler MA, Beachy RN (1986) Functional analysis of regulatory elements in a plant embryo-specific gene. Proc. Natl. Acad. Sci. USA 83: 8560–8564.

Chung J, Babka HL, Graef GL, Staswick PE, Lee DJ, Cregan PB, Shoemaker RC and Specht JE. (2003). The seed protein, oil, and yield QTL on soybean linkage group I. Crop Science 43:1053-67.

CS, Lara MV, Fernie AR and Drincovich MF. (2011). Metabolic profiling during peach fruit development and ripening reveals the metabolic networks that underpin each developmental stage. Plant Physiol 157:1696-710.

Derbyshire E, Wright DJ and Boulter D. (1976). Legumin and Vicilin, storage proteins of legume seeds. Phytochemistry 15:3-24.

Diers BW, Keim P, Fehr WR and Shoemaker RC. (1992). RFLP analysis of soybean seed protein and oil content. Theor Appl Genet 83:608-12.

Dinkins RD, Srinivasa MS, Meurer CA, Yan B, Trick H and Thibaud-Nissen F. (2001). Increaded sulphur amino acids in soybean plants over expressing the maize 15 kDa Zein protein. In Vitro Cell Dev Biol-Plant 37:742-47.

D'Mello JPF. (2003). Amino acids as multifunctional mole-cules. In: D. M. J.P.F editor. Amino Acids in Animal Nutrition. 2nd ed. Cambridge, MA: CABI Publishing,. p 1-14.

Elke GR, Blackmore D, Offler CE, Patrick JW (2005) Increased capacity for sucrose uptake leads to earlier onset of protein accumulation in developing pea seeds. Functional Plant Biology 32: 997-1007.

Fait A, Angelovici, R, Less H, Ohad I, Urbanczyk-Wochniak E, Fernie AR, Galili G (2006) *Arabidopsis* seed development and germination is associated with temporally distinct metabolic switches. Plant Physiology 142:839-854.

Fasoula VA, Harris DK and Boerma HR. (2004). Validation and designation of quantitative trait loci for seed protein, seed oil, and seed weight from two soybean. Crop Science 44:1218-25.

Gayler KR, Sykes GE. (1985) Effects of nutritional stress on the storage proteins of soybeans. Plant Physiol. 78:582–585.

Geetha KB, Lending CR, Lopes MA, Wallace, JC, Larkins BA (1991) Opaque-2 modifiers increase [gamma]-zein synthesis and alter its spatial distribution in maize endosperm. Plant Cell 3:1207-1219.

Gingeras T. (2009). Post-transcriptional processing generates a diversity of 5'-modified long and short RNAs. Nature 457:1028-32.

Goldberg RB, Hoschek G, Ditta GS, Breidenbach RW. (1981a) Developmental regulation of cloned superabundant embryo mRNAs in soybean. Dev. Biol. 83: 218–231.

Goldberg RB, Hoschek G, Tam SH, Ditta GS, Breidenbach RW (1981b) Abundance, diversity, and regulation of mRNA sequence sets in soybean embryogenesis. Dev. Biol. 83: 201–217.

Golombek S, Rolletschek H, Wobus U, Weber H (2001) Control of storage protein accumulation during legume seed development. Journal of Plant Physiology 158: 457–464.

Goto DB and Nakayama J. (2012). RNA and epigenetic silencing: Insight from fission yeast. Develop Growth Differ 54:129-41.

Gu L. Jones AD, Last RL (2010) Broad connections in the *Arabidopsis* seed metabolic network revealed by metabolite profiling of an amino acid catabolism mutant. Plant J 61:579-590.

Gutierrez L, Van Wuytswinkel O, Castelain M, Bellini C (2007) Combined networks regulating seed maturation. Trends Plant Sci 12:294-300.

Hagan ND, Upadhyaya N, Tabe LM, Higgins TJ (2003) The redistribution of protein sulfur in transgenic rice expressing a gene for a foreign, sulfur-rich protein. Plant J. 34: 1-11.

Hajduch M, Ganapathy A, Stein JW and Thelen JJ. (2005). A systematic proteomic study of seed filling in soybean. Establishment of high-resolution two-dimensional reference maps, expression profiles, and an interactive proteome database. Plant Physiol 137:1397-419.

Harada JJ, Barker SJ and Goldbery RB. (1989). Soybean β-Con-glycinin Genes Are Clustered in Several DNA Regions and Are Regulated by Transcriptional and Posttranscriptional Processes. Ibid 1:414-25.

Herman EM, Burks AW (2011) The impact of plant biotechnology on food allergy. Curr. Opinion Biotechnology. 22: 224-230.

Herman EM, Larkins BA (1999) Protein storage bodies. Plant Cell 11:601-61.

Herman EM, Schmidt MA, M.A. Improved protein storage in plants, provisional patent filing application US # 3536553, (2011).

Hernandez-Sebastia C, Marsolais F, Saravitz C, Israel D, Dewey RE, Huber SC. (2005) Free amino acid profiles suggest a possible role for asparagine in the control of storage-product accumulation in developing seeds of low-and high-protein soybean lines. J. Exp. Bot. 56: 1951-1963.

Higgins TJV. (1984). Synthesis and Regulation of Major Proteins in Seeds. Ann Rev Plant Physiol 35:191-221.

Hill JE, Breidenbach RW (1974) Proteins of soybean seeds. II. Accumulation of the major protein components during seed development and maturation. Plant Physiol. 53:747–751.

Hirai MY, Fujiwara T, Chino M, Naito S (1995) Effects of sulfate concentrations on the expression of a soybean seed storage protein gene and its reversibility in transgenic *Arabidopsis thaliana*. Plant Cell Physiol. 36:1331–1339.

Holowach LP, Madison JT, Thompson JF (1986) Studies on the mechanism of regulation of the mRNA level for a soybean storage protein subunit by exogenous l-Methionine. Plant Physiol. 80: 561–567.

Hu H and Wilcox JR. (1992). Rucurrent selection for maturity and percent protein in Glycine max based on S0 plant evaluations. Euphytica 62:51-57.

Jofuku KD, Okamuro JK, Goldberg RB (1987) Interaction of an embryo DNA binding protein with a soybean lectin gene upstream region. Nature 328:734–737.

Kinney AJ, Jung R, Herman EM (2001) Cosuppression of the □-subunits of conglycinin in transgenic soybean seeds induces the formation of endoplasmic reticulum-derived protein

bodies. Plant Cell 13:1165-1178.

Kitamura K. (1995). Genetic improvement of nutrional and food processing quality in soybean. Japan Agricultural Research Quarterly 29:1-8.

Krishnan HB, Natarajan SS, Mahmoud AA and Nelson RL. (2007). Identification of glycinin and beta-conglycinin subunits that contribute to the increased protein content of high-protein soybean lines. J Agric Food Chem 55:1839-45.

Krishnan HB and Nelson RL. (2011). Proteomic analysis of high protein soybean (Glycine max) accessions demonstrates the contribution of novel glycinin subunits. J Agric Food Chem 2432-39.

Kroij T, Savino G, Valon C, Giraudat J, Parcy F (2003) Regulation of storage protein gene expression in *Arabidopsis*. Development 130: 6065-6073.

Kortt AA and Caldwell JB. (1990). L o w molecular weight albumins from sunflower seed: identification of a methionine-rich albumin. Phytochemistry 29:2805-10.

Kwanyuen P, Pantalone VR, Burton JW and Wilson RF. (1997). A new approach to genetic alteration of soybean protein composition and quality. Journal of the American Oil Chemists' Society 74:983-87.

Kwong RW, Bui AQ, Lee H, Kwong LW, Fischer RL, Goldberg RB, Harada JJ (2003) LEAFY COTYLEDON1-LIKE defines a class of regulators essential for embryo development. Plant Cell 15:5–18.

Lee SH, Bailey MA, Mian MAR, Carter TE, Shipe ER, Ashley DA, Parrott WA, Hussey RS and Jun TH, Van K, Kim MY, Lee SH and Walker DR. (2007). Association analysis using SSR markers to find QTL for seed protein content in soybean. Euphytica 162:179-91.

Lessard PA, Allen RD, Bernier F, Crispino JD, Fujiwara T, Beachy RN (1991) Multiple nuclear

factors interact with upstream sequences of differentially regulated beta-conglycinin genes. Plant Mol. Biol. 16:397-413.

Li J, Hua W, Zhan G, Wei, F, Wang X, Liu G, Wamg H (2010) Increasing seed mass and oil content in transgenic *Arabidopsis* by the over-expression of *wri1*-like gene from *Brassica napus.* Plant Physiol Biochem 48:9-15.

Lombardo VA, Osorio S, Borsani J, Lauxmann MA, Bustamante CA, Budde CO, Andreo, María LV, Alisdair FR and María DF (2011). Metabolic profiling during peach fruit development and ripening reveals the metabolic networks that underpin each developmental stage. Plant Physiology 157: 1696-1710.

Maeo K, Tokuda T, Ayame A, Mitsui N, Kawai, T, Tsukagoshi H, Ishiguro S, Nakamura K (2009) An AP2-type transcription factor. WRINKLED1, of *Arabidopsis thaliana* binds to the AW-box sequence conserved among proximal upstream regions of genes involved in fatty acid synthesis. Plant J 60:476-487.

Mienke DW, Chen J, Beachy RN (1981) Expression of storage-protein genes during soybean seed development. Planta 153:130–139

Miller JE and Fehr RW. (1979). Direct and indirect selection of protein in soybean. Crop Science 4:196-98.

Mooney BP, Krishnan HB and Thelen JJ. (2004). High-throughput peptide mass fingerprinting of soybean seed proteins: automated workflow and utility of UniGene expressed sequence tag databases for protein identification. Phytochemistry 65:1733-44.

Naito S, Dube PH, Beachy RN (1988) Differential expression of conglycinin alpha-subunit and beta-subunit genes in transgenic plants. Plant Mol. Biol. 11:109–123.

Nielsen NC and Nam YW. (1999) Soybean globulins. In Shewry P.R., Casey R., eds. Seed

Proteins. Kluwer Academic Publishers, Dordrecht, The Netherlands. 285–313

Nielsen NC, Dickinson CD, Cho TJ, Thanh VH, Scallon BJ, Fischer RI, Sims TL, Drews GN and Goldbery RB. (1989). Characterization of the glycinin family in soybean. Plant Cell 1:313-28.

Nordlee JA, Taylor SL, Townsend JA, Thomas LA and Bush RK. (1996). Identification of a Brazil –nut allergin in transgenic soybeans. New England J Med 334:680-92.

Ohtake N, Kawachi T, Okuyama I, Fujikake H, Sueyoshi K, Ohyama T (2002) Effect of short-term application of nitrogen on the accumulation of beta-subunit of betaconglycinin in nitrogen-starved soybean (*Glycine max* L.) developing seeds. Soil Science and Plant Nutrition 48:31–41.

Ohto MA, Fischer RL, Goldberg RB, Nakamura K, Harada J. (2005) Control of seed mass by APETLA2. Proc. Natl. Acad. Sci. USA 22: 3123-8.

Opiyo SO. (2003). Rucurrent selection in soybean for increased seed protein content from a base population of high-yielding parents. In: Agronomy. Lincoln: Unversity of Nebraska-Lincoln. p DOI:oclc/52343051.

Osborne TB. (1924). The vegetable proteins.London: Longmans Green.

Panthee DR, Pantalone VR, West DR, Saxton AM and Sams CE. (2005). Quantitative trait loci for seed protein and oil concentration and seed size in soybean. Crop Science 45:2015-22.

Phansak P. (2010). Detection of soybean seed protein QTLS using selective genotyping. In: Agronomy and Horticulture. Lincoln: University of Nebraska-Lincoln.

Prakash V and Rao MS. (1986). Physicochemical properties of oilseed proteins. CRC Crit Rev Biochem 20:265-363.

Perez-Grau L, Goldberg RB. (1989) Soybean Seed Protein Genes Are Regulated Spatially during

Embryogenesis. Plant Cell 1:1095-1109.

Pioneer Hi-Bred International I. In.

Rainbird RM, Thorne JH and Hardy RW. (1984). Role of amides, amino acids, and ureides in the nutrition of developing soybean seeds. Plant Physiol 74:329-34.

Müntz K, Christov V, Saalbach G, Saalbach I, Waddell D, Pickardt T, Schieder O and Wuestenhagen T. (1998). Genetic engineering for high methionine grain legumes. Nahrung 42:125-27.

Rolletschek H, Hosein F, Miranda M, Heim U, Götz KP, Schlereth A, Borisjuk L, Saalbach I, Wobus U, Weber H. (2005) Ectopic expression of an amino acid transporter (VfAAP1) in seeds of Vicia narbonensis and pea increases storage proteins. Plant Physiol. 137:1236-49.

Santos-Mendoza M, Dubreucq B, Baud S, Parcy F, Carboche M, Lepiniec L (2008) Deciphering gene regulatory networks that control seed development and maturation in *Arabidopsis.* Plant J 54:608-620.

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 and Herman EM. Designing oilseeds for value-added protein co-products. In; Designing Soybeans for 21st Century Markets. Ed. R.Wilson. AOCS Press. pp 227-238, Schmidt MA, and Herman EM (2008)

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J,

Rokhsar D, Stacey G, Shoemaker RC and Jackson SA. (2010). Genome sequence of the palaeopolyploid soybean. Nature 463:178-83.

Schmidt MA, Barbazuk WB, Sandford M, May G, Song Z, Zhou W, Nikolau BJ and Herman EM. (2011). 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. Plant Physiol 156:330-45.

Schuler MA, Schmitt ES, Beachy RN. (1982a) Closely related families of genes code for the alpha and alpha' subunits of the soybean 7S storage protein complex. Nucleic Acids Res. 10: 8225–8244.

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.

Serretti C, Schapaugh WT and Leffel RC. (1994). Amino acid profile of high-seed-protein soybean. Crop Science 34:207-09.

Shi CY, Yang H, Wei CL, Yu O, Zhang ZZ, Jiang CJ, Sun J, Li YY, Chen Q, Xia T and Wan XC. (2011). Deep sequencing of the Camellia sinensis transcriptome revealed candidate genes for major metabolic pathways of tea-specific compounds. BMC Genomics 12:131.

Tabe L, Hagan N, Higgins TJ (2002) Plasticity of seed protein composition in response to nitrogen and sulfur availability. Curr. Opin. Plant Biol. 5:212-217.

Takahashi M, Uematsu Y, Kashiwaba K, Yagasaki K, Hajika M, Matsunaga R, Komatsu K, Ishimoto M (2003) Accumulation of high levels of free amino acids in soybean seeds through integration of mutations conferring seed protein deficiency. Planta. 217:577-86.

Thorne JC and Fehr RW. (1970). Incorporation of high protein exotic germplasm into soybean

populations by 2-and 3-way crosses. Crop Science 10:625-55.

To A, Valon C, Savino G, Guilleminot J, Devic M, Giraudat J, Parcy F (2006) A network of local and redundant gene regulation governs *Arabidopsis* seed maturation. Plant Cell 18: 1642-1651.

Tombs MP. (1967). Protein Bodies of the Soybean. Plant Physiol 42:797-813.

Tsubokura Y, Hajika M, Kanamori H, Xia Z, Watanabe S, Kaga A, Katayose Y, Ishimoto M and Harada K. (2012). The beta-conglycinin deficiency in wild soybean is associated with the tail-to-tail inverted repeat of the alpha-subunit genes. Plant Mol Biol 78:301-9.

Walling L, Drews GN, Goldberg RB (1986) Transcriptional and post-translational regulation of soybean seed protein mRNA levels. Proc. Natl. Acad. Sci. USA 83:2123-2127.

Weber H, Borisjuk L, Wobus U (2005) Molecular physiology of legume seed development. Annu. Rev. Plant Biol. 56:253–279.

Weckwerth W and Fiehn O. (2002). Can we discover novel pathways using metabolomic analysis? Curr Opin Biotechnol 13:156-60.

Wilson RF. (1987). Seed metabolism. In: A. Monogr editor. In Soybeans: Improvement, Production and Uses. 2nd ed. Madison, WI: ASA-CSSA-SSSA. p 643-86.

Zhang L, Zhou W, Velculescu VE, Kern SE, Hruban RH, Hamilton SR, Vogelstein B and Kinzler KW. (1997). Gene expression profiles in normal and cancer cells. Science 276:1268-72.