Update on Seed Quality Traits

Can We Improve the Nutritional Quality of Legume Seeds?

Trevor L. Wang*, Claire Domoney, Cliff L. Hedley, Rod Casey, and Michael A. Grusak

John Innes Centre, Norwich Research Park, Colney, Norwich NR4 7UH, United Kingdom (T.L.W., C.D., C.L.H., R.C.); and United States Department of Agriculture-Agricultural Research Service Children’s Nutrition Research Center, 1100 Bates Street, Houston, Texas 77030–2600 (M.A.G.)

The Food and Agriculture Organization statistics for 2001 (http://apps.fao.org/page/collections?subset=agriculture) show that 274 million metric tons of grain legumes were produced across the world, of which 177 million were soybeans (Glycine max; one-half of which were produced in the U.S.) compared with 2 trillion metric tons of cereals. Legume seeds are put to a myriad of uses, both nutritional and industrial, and in some parts of the developing world they are the principle source of protein for humans. They form a very important part of our diet and that of other animals. However, compared with meat, our main source of protein, legumes are deficient in sulfur-containing amino acids. Legume seeds are also an important source of dietary minerals, with the potential to provide all 15 of the essential minerals required by man (Grusak, 2002). The concentrations of certain minerals (especially Fe, Zn, and Ca), however, are low relative to animal food products.

Some legumes contain compounds detrimental to our diet, so for this reason and those above, it is desirable to improve their quality. Here, we concentrate on the protein, carbohydrate, and mineral content of legume seeds, although a major component of some legume seeds is their oil. This oil is mainly consumed indirectly through processed foods such as peanut (Arachis hypogaea) butter or margarine, or via cooking oils. A major goal in this area, on health grounds, is the reduction of the “saturated” forms of lipid in soybean oil, mainly by reduction of palmitate, for which there are both conventional breeding and genetic engineering approaches (Kinney, 1998). Modification of storage compounds, however, can have unforeseen consequences if there are pleiotropic effects or if the plant requires the material for vital processes and not just for storage. Such challenges to plant modification are also outlined here.

**LEGUME PROTEINS ARE OF TWO DISTINCT CLASSES**

The majority of protein in legume seeds consists of salt-soluble globulins, or storage proteins, that are synthesized during seed development, stored in protein bodies, and hydrolyzed during germination to provide nitrogen and carbon skeletons for the developing seedling. The remainder are albumins that include many “housekeeping” proteins, lectins, and lipoxygenases. The globulins comprise two classes, termed 7/8S and 11/12S on the basis of their sedimentation coefficients. The 11/12S are hexameric and are generally known as legumins (or glycinin in soybean), whereas the 7/8S are trimers and are variously known as vicilin, convicilin, β-conglycinin, phaseolin, canavalin, and other trivial names, depending on their species of origin. Both are nutritionally deficient in Cys and Met—the 7/8S more so than the 11/12S—and both have important physicochemical or “functional” properties that are significant to their use in foods.

**SEED STORAGE PROTEINS ARE CONSERVED AMONG LEGUMES**

7/8S and 11/12S globulins are present in all major grain legumes, in the model legumes Medicago truncatula and Lotus japonicus, and also in a wide range of nonlegume species (Casey, 1999). Within the 7/8S and 11/12S classes, there are several sequence subclases; sequence relationships suggest early divergence of the subclasses during legume speciation (Casey et al., 1986).

Sequence comparisons indicate little homology between 7/8S and 11/12S globulins, but determination of crystal structures makes it clear that the two classes share common ancestry (Maruyama et al., 2001, and refs. therein) and support the existence of relatively polar, “hypervariable” (Nielsen et al., 1989) regions at the molecular surface that may be potential sites for directed protein engineering. Attempts at this have proved profitable in soybean (Utsumi et al., 2002) but are limited by the necessity to produce engineered glycinin in Escherichia coli as an unprocessed “proglycinin” trimer, rather than the mature hexamer. Expression of 11/12S globulins in the ap-
Antinutritional proteins are classified as undesirable or antinutritional. Many of these belong to the albumin fraction, components that are classified as undesirable or antinutritional. Antinutritional proteins are especially poor for ruminant nutrition, however, because the protein is degraded in the rumen. Processing using heat treatments may be helpful in improving nitrogen values for these animals (Aranda et al., 2001).

Other targets for legume seed protein improvement include removal of antinutritional factors and activities that generate undesirable flavors, removal of potential allergens, improved digestibility, and improved functional behavior for processing. The molecular basis of many of these has not been sufficiently well defined to enable directed improvement, either by breeding with genetic variation or by genetic manipulation. There have been several attempts to alter the amino acid composition of the globulins through the use of natural variation, and through genetic manipulation either to directly modify globulin amino acid sequence or to express exogenous sulfur-rich proteins (see Krishnan, 2000). The great potential of these approaches is still largely unrealized but should yield seed protein with enhanced quality in the future. Studies of model legumes can play an important role in this, through improved understanding of the regulation of the amounts of the individual seed proteins and of the effects of environment and genetic background on protein quality (Casey et al., 1993).

Legumes also contain antinutritional proteins

Legume seed protein contains a variety of components that are classified as undesirable or antinutritional. Many of these belong to the albumin fraction, often considered to be of more favorable amino acid composition than the globulins. It is unclear to what extent these proteins are dispensable either to the seed or to the germinated plant because in only a few cases have appropriate mutants been identified. Where mutants have been studied, for example, for lipoxygenases in pea (Pisum sativum) and soybean (Forster et al., 1999), a concomitant loss of seed or seedling vigor has not been reported.

Enzyme inhibitors have received much attention as legume albumin proteins with negative attributes. It may be inferred from studies of near-isogenic pea lines, differing in their quantity of trypsin inhibitor (TI) proteins, that benefits in terms of animal performance could be derived from the use of null mutants (Hedemann et al., 1999). Mutants and/or transgenic "knockouts" could also provide answers to the question of whether or not TI proteins are essential for seed/plant viability.

In soybean, analysis of TI proteins is confounded by the existence of two unrelated gene families (Kunitz and Bowman-Birk). Mutants have been described for both of these, though the mutation in the latter family is in a wild Glycine species and may not easily be combined with the mutation described for Kunitz (Vollmann et al., 2002). Such mutants may prove advantageous to the processing of soybean, in that the high temperatures required for inactivation of inhibitors for animal feed purposes may be avoided. However, this possible advantage is more relevant to other legumes that are primarily processed for animal feed where cost savings may be achieved; in the case of soybean, where protein is a by-product of oil extraction, alterations to processing conditions may be less tolerated.

The definition in pea of multigene families encoding closely related TI proteins with discrete expression patterns (Domoney et al., 2002) has facilitated the development for use in breeding of "ideal" DNA markers (based directly on the genes of interest). The markers based on the Tri loci distinguish genes from pea lines that are genetic variants for quantitative expression of the seed trypsin-chymotrypsin inhibitors. Such markers promise to be reasonably robust and to circumvent the requirement for cumbersome assessments of seeds in breeding programs and should pave the way for large-scale screening for seed quality parameters (Page et al., 2002). Furthermore, these studies provide some explanation for variation in quantitative expression of TI proteins.

The major pea seed albumin, PA2, has a number of characteristics that are undesirable for various end uses. This protein is cytosolic and is not hydrolyzed during germination in line with seed storage proteins. PA2 has been implicated in partial insolubility of pea protein isolates, due to a free sulfydryl group, and, furthermore, has been shown to resist digestion in the digestive tract of chickens (Gallus domesticus; Crevieu et al., 1997, and refs. therein). A PA2-homologous protein, isolated from chickpea (Cicer arietinum), has been shown to have lectin-like properties and has been implicated in allergic responses.

Can seed protein quality be manipulated?

Vegetable, notably soybean, proteins have been used in the food industry for many years and their beneficial effects are attributed to their so-called "functional" properties, particularly those that relate to gelation, emulsifying, and foaming behavior. Few studies have been carried out on engineering the properties of legume globulins to alter functional behavior, but those of Utsumi’s group (Maruyama et al., 2002; Utsumi et al., 2002) have given valuable insights into the relationship between structure and functionality for soybean proteins. The lack of Cys and Met in legume seed globulins can readily be overcome in feed production by mixing with cereal protein (which has a complementary composition; Shewry and Tatham, 1999), but is still regarded as a target for improvement (Krishnan, 2000). Legume proteins are especially poor for ruminant nutrition, however, because the protein is degraded in the rumen. Processing using heat treatments may be helpful in improving nitrogen values for these animals (Aranda et al., 2001).

Other targets for legume seed protein improvement include removal of antinutritional factors and activities that generate undesirable flavors, removal of potential allergens, improved digestibility, and improved functional behavior for processing. The molecular basis of many of these has not been sufficiently well defined to enable directed improvement, either by breeding with genetic variation or by genetic manipulation. There have been several attempts to alter the amino acid composition of the globulins through the use of natural variation, and through genetic manipulation either to directly modify globulin amino acid sequence or to express exogenous sulfur-rich proteins (see Krishnan, 2000). The great potential of these approaches is still largely unrealized but should yield seed protein with enhanced quality in the future. Studies of model legumes can play an important role in this, through improved understanding of the regulation of the amounts of the individual seed proteins and of the effects of environment and genetic background on protein quality (Casey et al., 1993).

Legumes also contain antinutritional proteins

Legume seed protein contains a variety of components that are classified as undesirable or antinutritional. Many of these belong to the albumin fraction, often considered to be of more favorable amino acid composition than the globulins. It is unclear to what extent these proteins are dispensable either to the
in chickpea-sensitive individuals (Vioque et al., 1998). In combination, these features indicate that a reduction or removal of PA2 could lead to significant improvements in seed quality for food and feed end uses. Introduction of variant PA2 alleles (Vioque et al., 1998, and refs. therein), if proven desirable, into elite genetic backgrounds could be simplified using molecular marker techniques.

ARE LEGUME CARBOHYDRATES GOOD FOR US?

Starch, the dominant carbohydrate in our diets, is used as the main carbon reserve in many grain legumes (e.g. pea), but grain legumes are also high in soluble carbohydrates, especially the raffinose family of oligosaccharides (RFO). There is much genetic variation for both the total seed content and the composition of starch (Wang et al., 1998) and of RFOs (Jones et al., 1999). In pea, over 30 novel starch mutants at five loci (r, rb, rug3, rug4, and rug5) have been characterized. These mutations, which alter the shape of the seed from round to wrinkled (hence the term *rugosus* loci) lead to changes in starch content, polymer (amylose and amylopectin) composition, and the physical structure of the starch granule. The original *r* (*rugosus*) mutant was used by Mendel as one character in his studies of inheritance. There is a sixth locus, *lam* (*low amylose*), which does not affect the shape of the seed, changes the starch content very little, but decreases the amylose content considerably. Also in pea, there are lines with very high and very low verbascose content where the latter is due to impairment of the enzyme verbascose synthase (Peterbauer et al., 2001).

In humans, starch is normally consumed as part of cooked or processed food. After this process, a proportion of the starch is recrystallized on cooling to become highly resistant to pancreatic amylase (retrograded) and cannot be digested—so-called “resistant starch” (RS). RS contributes to the total unavailable carbohydrates believed to be important in combating certain forms of cancer (Aranda et al., 2001). Grain legumes are characterized by a relatively low glycaemic index (the blood Glc-raising potential) that is about one-half that of white bread. Foods with a low glycaemic index are considered to be beneficial in reducing postprandial blood Glc and insulin responses; therefore, it is especially useful to include legumes in the diet of people with insulin-dependent diabetes (type 2). Vegetarian diets that are high in grain legumes reduce the incidence of digestive tract cancers by reducing the consumption of saturated fats and increasing the content of unavailable carbohydrates in the diet (Aranda et al., 2001).

Starch is the primary energy source in many animal diets, but legume starch generally provides less available energy, especially in monogastric animals, than do cereals because of its high amylose content (almost double) and the properties of the granules (Aranda et al., 2001). This could be one explanation for the difference between the digestibility of starch from wild-type (RR) and high-amylose (rr) peas. Processes involving heat, such as pelleting or extrusion, increase the digestibility of legume starches in meal when fed to chickens. The high content of dietary fiber in legumes, however, can have a negative effect on digestibility in animals.

The presence of the RFO in seeds is one of the major reasons why legumes do not play a more major role in animal and human nutrition. The degradation of these compounds occurs in the lower bowel, where they are fermented by bacteria with the release of hydrogen and methane causing discomfort in humans and diarrhea in animals. This occurs because higher animals, including humans, lack the enzyme (*α*-galactosidase) necessary to break the α(1→6) linkage that characterizes this group of compounds. It has been shown that intestinal digestion of the *α*-galactosides can be increased if animal diets are supplemented with exogenous *α*-galactosidase. The presence of *α*-galactosidases in the colon, however, may have a beneficial effect by increasing the bifidobacteria population. These bacteria produce short-chain fatty acids that reduce the incidence of colon cancer in humans (Aranda et al., 2001).

CAN CARBOHYDRATE QUALITY BE ALTERED?

Processing is a standard way of manipulating carbohydrate utilization. Soaking and sprouting legume seed can enhance starch digestibility and reduce the level of RFOs by up to 100% through release of *α*-galactosidase (Aranda et al., 2001). Fermentation is used in more exotic grain legumes, the RFO being hydrolyzed by bacterial *α*-galactosidase.

Genetically manipulating the level of RFO—by inhibiting galactinol synthase activity—has been patented (Kerr et al., 1998). This is the first committed reaction in the pathway and involves the synthesis of galactinol from UDP-Gal and myo-inositol. The individual members of the RFO are then synthesized by distinct galactosyltransferases (e.g. raffinose synthase and stachyose synthase). The physiological importance of the RFO during seed development and storage (see below) suggests that a better strategy would be based on the activation of *α*-galactosidase to degrade the RFO after harvesting or based on the transfer of *α*-galactosidase from a thermophilic bacterium (*Thermotoga neapolitana*) into grain legumes (Griga et al., 2001). This has a temperature optimum close to 100°C and could be activated by, for example, canning. Frias et al. (1999) have suggested an alternative: reducing the level of the RFO while promoting the synthesis of related compounds such as the galactosyl cyclitols. This would maintain the protective nature of these compounds but decrease their flatus potential, because there is evidence that ciceritol is more slowly hydrolyzed by *α*-galactosidase.
than the RFO. Ciceritol is present in chickpea and lentil (*Lens culinaris*) but has not been detected in pea. The key to introducing galactinol cyclitols into pea with an accompanied reduction in the RFO content appears to lie with stachyose synthase, which has a central role in the synthesis of the galactinol cyclitols and in the synthesis of stachyose (Peterbauer and Richter, 2001). It represents a link, therefore, between the RFO and galactinol cyclitol pathways.

Manipulating starch content, composition, and granule structure genetically has been performed primarily in pea through the generation of the *rugosus* and *lam* mutants (see above). The information gained from this “model,” however, has been applied successfully to narbon beans (*Vicia narbonensis*) genetically modified by a targeted reduction in ADP-Glc phosphorylase (Rolletschek et al., 2002). The effects of changes in starch granular structure on the nutritional quality of the seed have been examined in the *r* and *rb* pea mutants, which differ greatly for structural characteristics. The glycaemic index of products from the *r* mutant seed was predicted to be lower than for that from the *rb* seed. However, there were several other pleiotropic effects of the mutations that could have contributed to this difference. For example, there was a large difference in amylose content and a 3-fold difference in the proportion of RS (Skrabanja et al., 1999).

**SEEDS CONTAIN IMPORTANT MINERALS. CAN LEVELS BE MODIFIED?**

As noted earlier, grain legumes can contain all 15 of the essential minerals required by man, although concentrations will vary in response to both genetic and environmental factors. In regions where legumes are a significant component of the human diet, mineral deficiencies (especially Fe and Zn) can be quite prevalent. Therefore, efforts to understand how minerals move from the soil, through the plant, and into developing seeds have gained much interest in recent years (Grusak, 2002), with the hope that this knowledge will facilitate strategies to increase seed mineral density.

Genetic diversity for seed mineral concentration has been studied in several legumes, although usually involving the characterization of only a limited number of genotypes (e.g. Meiners et al., 1976). In general, analysis of field-grown material has demonstrated comparable ranges of mineral concentrations among seeds of most legume species. Recently, seed mineral levels were characterized in the 500-accession *Pisum* Core Collection (part of the U.S. Department of Agriculture’s germplasm holdings), using plants grown under controlled, nutrient-replete conditions. In this study, broad genetic diversity was observed for seed micronutrient concentrations (accessions varied 3.5-fold for Fe to 6.9-fold for Mn) and for seed macronutrient concentrations (accessions varied 1.6-fold for Mg to 8.6-fold for Ca; data for individual accessions available at http://www.ars-grin.gov/cgi-bin/npgs/html/crop.pl?177). The identification of lines exhibiting high or low seed mineral levels is important, because these genotypes can now be used in comparative studies to decipher the underlying genetic and physiological mechanisms regulating mineral transport to developing seeds. They also will be useful to evaluate whether the enhancement of one mineral influences the concentration of another. Correlation analysis of seed minerals in recombinant inbred lines of bean (*Phaseolus vulgaris*; homozygous lines derived from an initial two-parent cross) has shown positive associations between most minerals (Beebe et al., 2000).

Mutants with altered seed mineral profiles only have been identified in pea. The *dg1* (degenerative leaves) mutation confers an uncontrolled hyperaccumulation of Fe into vegetative tissues and the ability to transport excess Fe to seeds (3-fold increase; Martens and Grusak, 1998). Along with another Fe hyperaccumulator that does not move excess Fe to its seeds (the *brz* [brone leaves] mutant), these mutants have been used to understand several aspects of whole-plant Fe homeostasis and the importance of phloem transport in seed Fe delivery (Grusak, 2000). Unfortunately, no other seed mineral mutants have been identified in any legume; more mutants are clearly needed.

Based on the characterization of existing germplasm, it would appear that 2-fold increases over current mean mineral concentrations might be feasible through classical breeding efforts, but that transgenic approaches will be needed to facilitate more dramatic changes. To devise useful biotech strategies, one needs to first understand the basic mechanisms by which minerals are transported to developing seeds. Minerals, of course, must be acquired initially from the soil environment and are delivered to vegetative tissues through the xylem pathway in response to the transpirational activity of various organs. Because legume seeds develop within the confines of enclosing pod walls (a region of high relative humidity), the seeds exhibit no transpiration, and, thus, they import no minerals via the xylem pathway. Instead, almost all minerals (Ca being a probable exception) must enter seeds via the phloem pathway, along with photoassimilates and other organic nutrients that are synthesized in various source regions (Grusak, 2002). Vegetative tissues, therefore, play a central role in the collection, temporary storage, and subsequent redistribution of minerals. In fact, the leaf concentration of several minerals declines as seeds develop (Hocking and Pate, 1977). What is not known is: (a) whether the remobilized minerals are phloem-loaded apoplastically or symplastically, (b) if phloem mineral loading can occur throughout all minor and major veins, (c) what role the mesophyll tissues play in...
the storage and possible release of minerals, and (d) which transporters might be required at the sieve element-companion cell periphery. Because many of these unknowns pertain to issues of spatial mineral distribution and leaf anatomy, it is worth noting that a multitude of leaf morphology mutants are available in pea (Hofer et al., 2001) that could serve as unique experimental tools to investigate the contribution of various tissue types and vascular patterns on seed mineral accretion.

DOES MODIFYING NUTRIENTALLY IMPORTANT COMPOUNDS HAVE CONSEQUENCES FOR THE PLANT?

If stored compounds have alternative roles in the plant, then modifying their types and amounts potentially could be catastrophic. The role of TI in plant defense against insect and nematode attack has been demonstrated clearly in many cases, including, for example, an interaction between pea TI and the pea aphid (Acyrthosiphon pisum; Quillien et al., 1998). However, the major seed-expressed TI may not fulfill a primarily protective function. Dissection of the expression patterns of three closely related TI classes in pea reveals that the mainly root-expressed TI genes encode inhibitors with two active sites that both inhibit trypsin, in contrast to the major seed-expressed TIs that are trypsin/chymotrypsin inhibitors (Domoney et al., 2002). It could be argued, therefore, that the former protein class, even though minor within the seed, is more likely to fulfill a defense role, because these proteins are predicted to have much higher affinity for trypsin-like enzymes. A related trypsin/trypsin class is expressed in developing flowers, particularly within the ovary wall, and persists in young pods. These may confer an indirect benefit to humans by protecting pods from pests. Recent research on the homologous TI proteins from soybean suggests a direct nutritional role in the prevention of certain proteolytic and related processes in tumorigenic cells (Zhang et al., 1999).

Other seed proteins may also contribute to plant defense, based on indirect evidence. A group of proteins, called PA1 or leginsulin in pea and soybean, respectively, is related to a large class of cereal enzyme inhibitors, but appears to lack any inhibitory activity and cannot be described as anti-nutritional. Limited homology to legume inhibitors is also apparent, however, particularly in the conservation of a Cys-Pro motif. The effectiveness of PA1 from pea as an insecticide against the cereal pests Sitophilus oryzae, Sitophilus granarius, and Sitophilus zeamais has led to a patent (Delobel et al., 1998). PA1 and TI proteins are likely to be ancestrally related, at least for the subunit (PA1b) that shows the insecticidal activity and the homology described above. PA1b is the smaller of two unrelated PA1 subunits that are not associated with each other in vivo. Soybean leginsulin is homologous to PA1b and can compete with insulin for binding to an insulin-binding protein (Watanabe et al., 1994). There is no evidence that leginsulin is a plant peptide hormone, but it may be part of a signal transduction pathway involving phosphorylation. The unusual structure of PA2, in particular its similarity to human vitronectin, suggests possible functions in the transport and storage of heme and/or in the control of cytolytic pathways through interaction with serine protease inhibitors (Vioque et al., 1998, and refs. therein). It is clear that the essential/nonessential plant function of many of these proteins may be more easily studied in other model legumes, where targeted mutants and/or transgenic “knockouts” may be obtained more readily.

During maturation in many grain legumes, seeds develop desiccation tolerance and accumulate RFO. It is thought that these compounds work in combination with Suc to stabilize membranes and a reduction in vigor and viability of stored seeds is accompanied by a decline in RFO content (Obendorf, 1997). RFO utilization also appears to play a very important role in germination, hence the use of germination to reduce the RFO before consumption. In contrast to RFO, removing starch does not seem to have much impact on the ability of the seed to develop or germinate because starchless pea mutants at the rug3 locus yield a viable crop (Wang et al., 1998).

With respect to minerals, elevated concentrations of macronutrients (Ca, Mg, P, and K) should cause no harm to the seed or seedling, but some care must be taken with respect to the manipulation of micronutrient metals (Fe, Zn, Mn, and Cu). Negative consequences for seed viability are a concern for the redox-active metals if they are not properly sequestered, and unintended increases in certain toxic metals (e.g. Cd and Ni) are possible, if some membrane-localized metal transporters were to be up-regulated, due to their broad selectivity.

THE FUTURE

The question in the title of this article is not new. Over the years it has been explored in many ways for each of the seed products mentioned here—for example, see Lambert and Yarwood (1992). Many of the challenges faced in the past have yet to be met fully, but our knowledge of the products and our ability to manipulate them has improved considerably. It is important that researchers continue both to utilize existing resources and to explore the new genomic tools (VandenBosch and Stacey, 2003) for modifying the nutritional components of legume seeds. Undoubtedly, any potential negative consequences to plant growth and development will need to be evaluated and monitored in newly developed lines. When indirect consequences do arise, however, they should be viewed merely as additional challenges to be overcome as we strive to develop a more nutritious food supply for a growing world.
LITERATURE CITED


Received November 13, 2002; returned for revision December 18, 2002; accepted January 6, 2003.