

# Enhancement of essential amino acid contents in crops by genetic engineering and protein design

Marc Beauregard<sup>1,\*</sup> and Mary Alice Hefford<sup>2</sup>

<sup>1</sup>Plant Biology Group, Département de Chimie-Biologie, Université du Québec à Trois-Rivières, Trois-Rivières, QC, Canada, G9A 5H7

<sup>2</sup>Centre for Biologics Research, Biologics and Genetic Therapies Directorate, A/L:2201C, Health Canada, Ottawa, ON, Canada, K1A 0L2

Received 28 February 2006;

revised 1 May 2006;

accepted 5 May 2006.

\*Correspondence (fax (819) 376-5084;

e-mail marc.beauregard@uqtr.ca)

**Keywords:** animal feeding, nutritive proteins, plant biotechnology, protein design, protein stability, seed storage proteins.

## Summary

The importance and urgency of providing humans and animals with quality proteins are reflected in the growing scientific and industrial interest in augmenting the nutritive value of the world's protein sources. Such nutritive value is determined by the protein content in 'essential amino acids', those that cannot be synthesized *de novo* and that must be supplied from the diet. It is the object of this review to discuss recent advances in the genetic modification of crops that aim to provide enhanced quantities of essential amino acids.

## Introduction

Protein quality has always been an important consideration in both human and animal nutrition. Because all animals lack the enzymatic machinery to synthesize *de novo* some of the amino acids needed to assemble the proteins necessary for their own growth and function, these 'essential amino acids' (EAAs) must be supplied from the diet. The deleterious effects of diets that are sufficient in protein quantity but deficient in protein quality are well documented: poor growth, tissue wasting and, in severe cases, death. As the world population increases (and with it the load on our agricultural resources), the need to make good-quality protein available efficiently and economically becomes increasingly important.

On the agricultural front as well, there is ample evidence that the nutritional requirements for EAAs in today's high-producing livestock are not readily met by simple mix rations. Meat proteins are now discouraged, and producers must turn to expensive 'high-quality' proteins ('protected' proteins, soya proteins, EAAs) in order to maximize yield (Williams, 2003). The importance and urgency of providing humans and animals with quality proteins are reflected in the growing scientific and industrial interest in augmenting the nutritive value of the world's protein sources. Many look to advances in the genetic modification of crops to provide solutions (Uzogara, 2000; Dijkstra *et al.*, 2003; Williams, 2003). Until recently, modification of crops involved protection against

pests or resistance to commercial herbicides, and provided a partial answer to the issue of crop yields. A second generation of modified crops is now emerging which boast significant changes specifically aimed at enhancing their nutritional value (Flachowsky *et al.*, 2005). The object of this review is to discuss advances in the area of genetic and protein engineering that will lead to an enhancement in the nutritional value of crops.

## Engineering crops with enhanced EAA content

It is now clear that proteins within the plant can be modified to increase the complement of EAAs – to make a 'complete protein source' from an 'incomplete' source. Plants, however, are complex organisms in which proteins are expressed in a tissue-specific manner in order to meet the functional requirements of each tissue. With the growing body of knowledge in each of the component areas, protein engineering in plants is now not only possible, but can be a practical alternative to traditional feeding and nutritional practices.

Amongst the different strategies used to date, three stand out as efficient ways to achieve an increased nutritional value.

**1** The first approach is the genetic engineering of EAA metabolism. In plant seeds, in addition to amino acids present in proteins, some amino acids are 'free' in the

cytosol of seed cells and available to be digested within the seed. These 'free' amino acids also represent the pool available to the plant cell for protein synthesis and, to some extent, limit the amount and type of protein synthesized by the cell. Thus, genetic engineering to increase the level of amino acid synthesis has the potential to both remove some of the limitations to protein synthesis and enrich the 'free amino acid content' of the plant.

**2** The second major approach is to change the levels of natural 'high-quality' proteins within the plant tissue. For this, the gene copy number and transcription rate for specific genes are increased, or genes with the appropriate EAA profile from a different organism can be imported for heterologous expression.

**3** The third major route to increase the nutritional quality of plants is to improve the nutritional quality of the proteins they synthesize through protein engineering and/or design. The amino acid contents of proteins already expressed in a particular plant can be tailored, or entirely new, modified proteins with a more desirable complement of amino acids can be designed and expressed.

### Metabolic engineering

The free amino acid pool represents up to 5% of the total amino acids available in plant seeds for digestion. In theory, therefore, a substantial modification of the EAA content of the amino acid pool of crops should result in a significant improvement in their nutritional value (Ghislain *et al.*, 1995). In cereals, both high-threonine- and high-lysine-producing mutants have been identified in model systems (Nefrutin *et al.*, 1984; Frankard *et al.*, 1992). In each case, the over-production was shown to result from a modification to an enzyme in the amino acid biosynthetic pathway. Even in the maize mutant opaque-2, in which several concomitant changes result in both decreased production of lysine-poor zein proteins and increased production of lysine-rich proteins, changes in the enzymes involved in lysine synthesis and degradation have been shown to be responsible, at least in part, for the elevated lysine content and concomitant changes in protein composition (Brochetto-Braga *et al.*, 1992; Gibbon and Larkins, 2005).

These classical mutants indicate that the modification of enzymes involved in both the anabolism and catabolism of EAAs by genetic engineering will result in an enhanced EAA content. In plants, lysine is derived from aspartate, and its synthesis is controlled by two main enzymes: aspartate kinase (AK) and dihydrodipicolinate synthetase (DHDPS). Both enzymes are regulated by end-product feedback inhibition by

lysine and/or threonine; relief of this inhibition results in the over-production of lysine. A lysine-insensitive form of DHDPS has been isolated from corn and characterized, and enzymes involved in the biosynthetic pathways for lysine have been cloned and/or identified from a variety of sources (Zhu and Galili, 2004; Huang *et al.*, 2005, and references cited therein.).

Scientists from a major seed company have achieved a several 100-fold increase in free lysine in canola and soybean using insensitive AK and DHDPS, a modification that led to a two- to fivefold increase in total lysine content (Falco *et al.*, 1998). Similarly, insertion of bacterial cysteine  $\gamma$ -synthase in potato has been shown to enhance cysteine levels two- to threefold over the wild-type level (Hesse *et al.*, 2003). Fusion of the *Escherichia coli* gene encoding serine acetyltransferase (SAT) to the 5'-signal sequence of the small subunit of RuBisCo (Ribulose-1,5-bisphosphate carboxylase/oxygenase) was performed in order to target the protein into the chloroplast (Denk and Böck, 1987). Leaf extracts of successfully transformed potato plants showed significantly higher SAT activity (up to 20-fold) and exhibited two- to threefold increases in the levels of cysteine and glutathione (GSH) compared with wild-type plants. A very similar strategy has been used to improve the supply of sulphur-containing amino acids in developing soybean seeds to facilitate the production of sulphur-rich proteins to meet the nutritional needs of both livestock and poultry (Krishana, 2005).

Recently, a copy of the *Arabidopsis* cystathionine  $\gamma$ -synthase gene, driven by the RuBisCo small subunit promoter, was introduced into alfalfa in order to obtain expression of this protein (which controls the first step unique to the synthesis of methionine) in alfalfa leaves. Not only were the levels of soluble methionine and S-methylmethionine significantly increased in the ensuing transgenics, but the levels of free cysteine and GSH were also increased two- to fivefold, as was protein-bound cysteine (Avraham *et al.*, 2005). Similar success was not obtained, however, when the native potato analogue of this same gene was over-expressed in potatoes (Kreft *et al.*, 2003). As this latter example illustrates, metabolic engineering to increase the pool of free amino acids is not always straightforward. Often, feedback loops work to maintain homeostasis, and over-production of one or another amino acid can result in the activation of degradation pathways (Karchi *et al.*, 1994).

Metabolic engineering can also result in unexpected and unwanted effects. Enhanced lysine biosynthesis in mutant plants and in transgenic plants expressing bacterial genes involved in lysine biosynthetic pathways appears to be associated with phenotypic abnormalities in the plant (Frankard *et al.*, 1992). These undesirable effects on leaves and stems

have been partially eliminated using seed-specific promoters for genes in the lysine biosynthetic pathway, but further refinement is required (Falco *et al.*, 1998; Mazur *et al.*, 1999). Thus, metabolic engineering as an approach to enhance both the free amino acid content and the amino acid pool for protein synthesis, although promising, may require further refinement. Such refinement may be achievable through the use of developmentally regulated promoters (Ghislain *et al.*, 1995) or, as recently proposed (Zhu and Galili, 2004), through the coordinated over-expression and suppression of appropriate genes in transgenic plants using RNA interference (RNAi) to 'knock-down' undesirable gene expression. This RNAi method has been used successfully to control lysine catabolism in seeds genetically modified to produce increased amounts of free lysine. Control of lysine catabolism during seed development improved germination (Segal *et al.*, 2003).

### Expression of natural EAA-rich proteins

#### *Protein expression in seeds*

Much of the early effort in expressing naturally occurring EAA-rich proteins has focused on the production of proteins in the seeds because, for many crops, seeds are most often used as a source of food and feed. Seed storage proteins were originally classified on the basis of the solubility of the extractable protein fractions as follows: albumins or water-soluble proteins; globulins (water insoluble, alkaline soluble); prolamines (alcohol soluble); and glutelins (insoluble in water, saline and alcohol, but soluble in alkaline solutions). Seeds from different species differ in their distribution of storage protein types, and different varieties of seed from within a species can exhibit markedly different distributions of storage proteins.

In seeds of Gramineae, there are large quantities of prolamines and glutelins; in seeds of other species, globulins predominate. Prolamines are relatively rich in proline and glutamine, but low in lysine. In cereal seeds, approximately 50% of the extractable protein is prolamine and, although the EAA content of each prolamine varies with the species and, to some extent, with the individual cultivar, lysine levels are low in all prolamines. For example, zein (the prolamine fraction of maize) often has as little as 0.1 g lysine/100 g protein (Nelson, 1969). The glutelin fraction of cereals has a higher lysine content (averaging 3.2 g lysine/100 g glutelin protein), but the glutelin content is small. Lysine is, in fact, a limiting amino acid in all cereal-based diets. In all cereals except maize (where lysine and tryptophan are co-limiting), threonine is the second limiting EAA.

The major seed proteins of legumes are globulins (e.g. glycinin and vicilin) which contain moderate to adequate supplies of lysine (Ma and Bliss, 1978). Methionine is limiting in the seeds of almost all legumes. There are, however, several naturally occurring cultivars in which the levels of methionine in the seeds are higher than those found in their close relatives. 'Anila' subtypes of the vicilin (phaseolin) storage proteins in the common bean have been found that contain additional methionine residues and enhanced levels of methionine-rich phaseolin deposition (Kami and Gepts, 1994; and references cited therein). Selection has been successful in increasing the percentage of seed protein in soybean (Brime and Buxton, 1979; Carter *et al.*, 1982) and the total protein content in wheat (Loeffler *et al.*, 1983). Thus, although seeds of many crops represent a natural source of concentrated protein, they are not ideal; the protein provided is most often limited in one EAA or another and, as a result, seeds are considered as an 'incomplete' protein source.

Several modifications to seeds and their storage proteins have been made to enhance their nutritive value, and some of these have met with at least partial success. The most famous early examples are perhaps the maize opaque-2 and floury-2 mutants. These mutants, obtained by conventional breeding programmes, contain a high lysine content as a result of a decrease in low-quality zein protein and an increase in glutelin content (Mertz *et al.*, 1964). Mutants of opaque-2 with near-normal seed phenotype have been designated 'High-Quality Maize' (see Villegas *et al.*, 1980 for a review) and are used throughout the world (Gibbon and Larkins, 2005).

One other way to complement a crop EAA profile is to express natural proteins from different species which contain sufficient quantities of the desired EAA(s) (heterologous expression). Simple expression of a methionine-rich maize protein in a methionine-deficient legume or of a lysine-rich legume protein in lysine-deficient corn, if possible, would generate a seed that could function as a complete protein source. A number of proteins have been identified as methionine sources: the maize 10-kDa zein with 30% methionine (Kirihaara *et al.*, 2001; and references cited therein); the maize 15-kDa zein with 15% methionine (Pedersen *et al.*, 1986); 2S albumin from *Bertholletia exalsa* (Brazil nut) harbouring 24% methionine and a 10-kDa seed prolamine with 25% methionine by weight (Masumura *et al.*, 1989); and an 18-kDa zein (high-sulphur zein) with 37% methionine (Chui *et al.*, 2003). Such proteins are well suited for complementation of the EAA profile in legumes.

Of the methionine-rich proteins available for EAA complementation, the 2S albumin of the Brazil nut has been

particularly well studied. This 2S albumin protein was first expressed in transgenic model plants (tobacco) by Altenbach *et al.* (1992), and resulted in a 30% enhancement of the seed methionine content. It has since been expressed in other plants with varying degrees of success. Studies by three different groups on 2S methionine-rich protein in transgenic *Brassica napus* (canola) seem to indicate that many components of the transgene construct must be considered in order to facilitate adequate expression and deposition of the desired protein. These include the choice of seed-specific promoter, the fusion of the signal peptide-encoding sequences, and selection for processed or unprocessed 2S albumin protein (Guerche *et al.*, 1990; Altenbach *et al.*, 1992; and references cited therein). Altenbach *et al.* achieved accumulation of 2S albumin protein up to levels of 4% of the total seed protein (an enrichment in total seed methionine of 33%) by expressing the unprocessed (prepro) 2S albumin protein from a phaseolin promoter. This research indicates that, in addition to a consideration of promoter strength and temporal expression, both transport and processing play key roles in the levels of heterologous protein accumulation in the seeds of transgenic plants.

Advancements in the gene transfer methods for legumes have expanded the range of plant targets for Brazil nut 2S albumin expression to include the common bean (Aragao *et al.*, 1992) and three different grain legumes (Saalbach *et al.*, 1994). Sunflower seed also expresses a 2S albumin protein that is extremely methionine rich (Kortt *et al.*, 1991). Chimeric genes for this 2S protein have been investigated in efforts to address the methionine requirement of pigs and chickens fed substantial amounts of peas or lupin (Tabe *et al.*, 1993). The successful expression of sunflower albumin in lupin seeds has been reported, resulting in a twofold improvement in methionine content, whilst lowering the sulphate detected in the plants (Molvig *et al.*, 1997). Sulphur supply, however, appears to be the limiting factor in the methionine and cysteine content in plants (Tabe and Droux, 2002). Although biotechnology has the potential to take care of excess free EAAs and relieve feedback inhibition of the EAA metabolic pathways by the expression of appropriate sink proteins, it cannot address the increased demand for chemical building blocks required for amino acid synthesis.

Work with transgenic plants has revealed another, rather unexpected, challenge that must be dealt with before this approach sees widespread use as a means to produce nutritive crops. Transgenic soybeans expressing the methionine-rich 2S albumin protein were one of the earliest nutritionally modified crops examined for commercialization by an international plant breeding and seed supply company. However,

a study (financially supported by that same company) indicated that the 2S albumin protein was probably a major Brazil nut allergen, and that transgenic soybean expressing this protein could elicit an allergic reaction in sensitive individuals (Nordlee *et al.*, 1996). Although the allergenicity of the sunflower 2S albumin protein has not been tested thoroughly, the homologous 2S albumin (Sina 1) in yellow mustard has been reported to be a major allergen in mustard seeds (Menendez-Arias *et al.*, 1988). This protein is closely related to 2S albumin found in canola, castor bean and Brazil nut. Considering that food allergies affect about 2% of adults and 6% of children, the company developing the transgenic soybean chose to cease development of the product. Because of the potentially fatal reactions in sensitive individuals, health regulatory authorities in most countries require that novel foods (such as transgenics) containing a potential allergen be labelled to alert the consumer.

Albumins, however, are not the only possible candidate for heterologous expression in transgenic plants. Some progress has also been made in the expression of soybean glycinin in transgenic potatoes. Glycinins, like most legume proteins, are sufficient in most amino acids except the sulphur-containing amino acids. Their expression in potato would complement patatin, the major storage protein in potato tubers, which is deficient in leucine, lysine and threonine. Neither patatin nor a glycinin on its own is deemed to be sufficient to meet the nutritional requirements of young children (FAO/WHO/UNU, 1985). Together, glycinin and patatin would constitute a source of complete protein. Soybean proglycinins have been expressed in model systems [*E. coli* (Kim *et al.*, 1990) and tobacco (Utsumi *et al.*, 1997, and references cited therein)]. Soybean  $\beta$ -conglycin has been expressed in petunia and transgenic tobacco (Beachy *et al.*, 1986).

Other methionine-enriched proteins, such as 10- and 15-kDa zeins, have been expressed and accumulated as protein bodies in alfalfa (Bagga *et al.*, 1999), and the EAA content in potato has been enhanced via the expression of an albumin (*AmA1* gene from *Amaranthus hypocondriacus*) by Chakraborty *et al.* (2002).

Expression of lysine-rich protein for cereal complementation has been the object of a few studies. Transgenic rice in which the lysine-rich  $\beta$ -phaseolin protein from bean is expressed to complement the lysine-deficient rice prolamines has been reported (Zheng *et al.*, 1995). The phaseolin protein could be expressed at relatively high levels (4% of total seed protein). This study was interesting, not only because of the importance of rice as a major dietary component for a significant proportion of the world population, but also because the heterologous protein was targeted to and

expressed in only one of the two types of protein body that coexist in the rice endosperm. A lysine-rich protein gene isolated from winged bean *Psophocarpus tetragonolobus* expressed in rapeseed has been patented (Sun *et al.*, 2001).

Instead of importing high-quality proteins from other species, some researchers have attempted to enhance the level of selected endogenous proteins in a given crop. This strategy has been used to avoid the down-regulation of the highly expressed  $\delta$ -zein genes. Transgenic corn harbouring a zein gene with modified 3' untranslated regions showed enhanced levels of the target nutritive protein, and feeding trials revealed that the protein supplied sufficient amounts of methionine for sustained animal growth (Messing and Lai, 2005). Enhancement of high-EAA proteins may also be achieved by repression of some genes coding for lower quality proteins. Using the RNAi method, Huang *et al.* (2005) repressed expression of the 19-kDa zein, resulting in an enhanced expression of other EAA-rich storage proteins in maize.

#### *Protein expression in leaf tissue*

Seed storage proteins also have the potential to improve forage crops if they can be accumulated in leaves. Grazing animals consume large amounts of forage. In ruminants, however, the nutritional quality of the feed ingested is only part of the nutritional picture. Microorganisms in the rumen digest most of what is ingested by the animal and use the nutrients for their own growth (Hungate, 1966; Clark *et al.*, 1992). It is the products of microbial digestion and the rumen microbes themselves that largely provide the nutrient source for the animal. In high-producing ruminants, such as modern cattle and sheep, however, rumen microbial synthesis of protein is insufficient to meet the animal's needs for amino acids such as lysine, methionine, threonine and leucine (King *et al.*, 1990; Clark *et al.*, 1992). In order to overcome these deficits, high-producing ruminants are often fed additional protein supplements (proteins or amino acids that have been artificially protected against microbial digestion in the rumen) (Ashes *et al.*, 1995).

Tabe and coworkers have attempted to identify proteins rich in sulphur-containing amino acids (particularly important in wool production) that are somewhat resistant to breakdown in sheep rumen. These studies have selected hen-egg ovalbumin and 2S methionine-rich albumin of sunflower seeds for transfer into forage clovers and lucerne (Tabe *et al.*, 1993; Khan *et al.*, 1996).

Crops for non-ruminants have also been engineered for the expression of high-EAA proteins. Both pea lectin expressed in potato plants (Edwards *et al.*, 1991) and barley

lectin expressed in tobacco plants (Wilkins *et al.*, 1990) were found in significant amounts in the vacuoles of leaf cells of the respective transgenic plants. Attempts to express the vacuolar targeted  $\beta$ -phaseolin gene behind the cauliflower mosaic virus promoter in transgenic tobacco plants resulted in new protein accumulation in the seed, but the  $\beta$ -phaseolin was presumably degraded in the protease-containing leaf vacuoles and no new protein was detected in that tissue (Bagga *et al.*, 1995). In contrast, the 15-kDa zein protein (which exists as an endoplasmic reticulum-derived protein body in maize seeds) stably accumulated in both the leaves and seeds of transgenic tobacco expressing that gene (Bagga *et al.*, 1995). Whether the protein accumulation in leaves results from endoplasmic reticulum retention or the inherent ability of the protein to aggregate and resist digestion is not evident from the results reported to date. Clearly, although seed storage protein expression in heterologous hosts is a promising avenue towards the enrichment of both seed and leaf proteins, much remains to be understood at a more fundamental level.

#### **Expression of modified EAA proteins**

Strategies involving both protein engineering and protein *de novo* design are being used to increase the nutritive value of plants. Protein engineering offers the possibility of modifying a plant's existing proteins to incorporate additional residues of limiting EAAs. This strategy capitalizes on the fact that the proteins being re-engineered are already produced and properly targeted to the appropriate plant tissue. The challenge is to incorporate more of a desirable complement of amino acids without altering the folding or stability of the protein or adversely affecting its abundance. Protein design strategies, in which entirely new genes coding for *de novo* designed proteins with more desirable amino acid composition have been introduced, allow much more flexibility in the composition of the protein, but must contend with both design (folding and stability) issues as well as appropriate targeting to the tissue of interest (normally the leaf or seed). With each approach, some success has been achieved and some hurdles remain.

Early work in protein engineering of proteins focused on seed storage proteins. As plants are able to synthesize all of the amino acids required for protein synthesis and growth, seed storage proteins are merely the source of nitrogen during the development of the seedling, not of specific amino acids. Functionally, therefore, they should accommodate variation in amino acid composition with ease (Shotwell and Larkins, 1991). However, because seed storage proteins have highly

ordered structures, structural modifications have the potential to adversely affect their deposition (Habben and Larkins, 1995). Initial attempts to improve the nutritional quality of legumes (which are deficient in methionine but rich in lysine) focused on the common bean, *Phaseolus vulgaris*. A naturally occurring 15-amino-acid sequence from the 15-kDa zein storage protein of maize was introduced into the bean  $\beta$ -phaseolin. Because phaseolins represent about 50% of the total bean protein, and because six of the 15 amino acids in the insertion sequence were methionine, this modification appeared to be one that would substantially increase the methionine level in the seeds. Although the gene exhibited a normal developmental expression, the modified protein did not accumulate. On the basis of their studies and other information from unrelated systems, the researchers concluded that both tertiary and quaternary structures were important to intracellular transport (Hoffman, 1996; and references cited therein). Work on the expression of phaseolins in *Xenopus* oocytes (Cerioli *et al.*, 1991) confirmed that interruptions to the C-terminal  $\alpha$ -helix domain of phaseolin monomers prevented their assembly into trimers and ensuing transport to the Golgi from the endoplasmic reticulum. The more recent work of Nuttall *et al.* (2003) confirms these findings and emphasizes the need to consider protein folding in augmenting the complement of EAAs in proteins. Recently, Rapp and collaborators reported the successful modification of soybean storage proteins to enhance their amino acid profile, whilst still maintaining the proteins' tertiary structures and thus allowing their accumulation in seed. A patent based on this work was issued recently (Rapp *et al.*, 2005).

Dyer *et al.* (1993) used the available three-dimensional structure of phaseolins to design methionine-rich looping sequences for insertion into loops and turns. They also suggested methionine replacement of hydrophobic amino acids within the hydrophobic core of the  $\beta$ -barrel domain of the phaseolin monomers, a domain that is not involved in trimerization. Others (Sparvoli *et al.*, 1996) have compared the sequences of the N-terminal regions of phaseolins from different lima bean sources with those of the common bean phaseolins, and have suggested potential 'acceptors' of rather drastic amino acid modifications. To date, experimental results showing the practicality of these suggested modifications have not been published. A similar approach, however, has been successfully used to engineer bean seed phytohaemagglutinin (PHA) to contain an additional three to four methionine residues (Kjemtrup *et al.*, 1994). In this case, the introduction of an engineered gene for *P. vulgaris* PHA protein enriched in methionine in transgenic tobacco plants

resulted in correct processing and accumulation of the protein in the seed.

The use of homologous protein sequences to select sites for methionine insertion was also employed to modify an *Arabidopsis* 2S albumin gene. 2S albumin proteins consist of two polypeptide chains of varied length connected by disulphide bridges between conserved cysteine residues, and represent a major dicotyledon seed storage protein (Youle and Huang, 1979). In most dicots, 2S albumins are relatively lacking in methionine content. The Brazil nut 2S albumin, with its 18% methionine content, is a notable exception (Ampe *et al.*, 1986). Replacement of a variable region of sequence in one of the two 2S albumin encoding genes with a methionine-rich sequence based on a homologous sequence from Brazil nut resulted in an increase of approximately 5% in the total methionine content in seeds expressing the transgene (DeClercq *et al.*, 1990). More recent research has attempted to correct both methionine and tryptophan deficiencies by incorporating tryptophan codons into transgenes of Brazil nut 2S albumins. Genes containing up to five consecutive tryptophan residues have been successfully expressed in model plants (Marcellino *et al.*, 1996).

Similarly, the consideration of structural constraints led to the selection of an insertion site for four contiguous methionine residues in a variable region of glycinin from soybean. Such engineered soybean glycinins can be expressed in potato tubers using the potato patatin promoter and, although the glycinins were not processed in the tubers, the proglycinin proteins trimerized and were deposited at levels of up to 1% of the total soluble protein. Interestingly, these researchers attempted to alter not only the EAA content of the proteins, but also their food processing qualities, whilst recognizing the paramount importance of three-dimensional structure, solubility and stability in this endeavour (Utsumi *et al.*, 1997; and references cited therein).

Zein proteins have also been the target of engineering to correct amino acid deficiencies in lysine and tryptophan. Zein genes with two lysine and two tryptophan codons inserted were successfully expressed in oocytes (Wallace *et al.*, 1988), indicating some tolerance for lysine in the zein structure. Subsequent efforts to express lysine-containing  $\alpha$ -zein genes in tobacco, however, proved to be unsuccessful (Ohtani *et al.*, 1991). Modified zeins as a source of methionine have been reported, and a truncated version of 'high-sulphur-zein' corn storage protein called 'High Methionine Domain' has been engineered, leading to 48% methionine content (Chui *et al.*, 1999).

Although no plant expression systems have been reported, a series of improved proteins developed by Rao and colleagues

is worth mentioning. Using rational approaches, barley seed protein  $\alpha$ -hordothionin (Rao *et al.*, 1994) and barley chymotrypsin inhibitor-2 (Roesler and Rao, 1999, 2000) were mutated to substantially enhance their EAA content. These modified proteins were characterized after production by microbes. This elegant work was based on rational design and site-directed mutagenesis and, as a consequence, the number of new and potentially improved proteins remained limited to a handful in each study. More recently, departing from rational engineering, this group took nutritive protein design into the era of high-throughput engineering using accelerated evolution. In this approach, high-throughput, random mutagenesis is coupled with parameter-specific screening to allow for the isolation of improved protein variants. In principle, this technology allows for rapid improvement of any protein, well characterized or not, in any field of application, in as much as a screening method which is sensitive to the parameter(s) to be improved is available. Rao authored some of the very first publications of accelerated evolution, work that culminated in a patent in 2005 (Rao and Sleister, 2005). The goal sought was an increase in the methionine level of *Glycine max* vegetative storage protein (VSP). To begin, a rational approach based on structural predictions and comparison with homologous proteins was used. Then, a second cycle of evolution involving high-throughput, error-prone polymerase chain reaction (PCR) was carried out, which led to the generation of a large library of further modified proteins. The screening method chosen by these scientists was based on antibody recognition of properly folded VSP (Sleister and Rao, 2002). Proteins that carried an improved content of EAAs and that were folded properly were retained by antibodies, whereas other derivatives were lost. It will be interesting to follow the fate of these promising proteins when transgenic crops engineered for their expression become available.

### Combined approaches

Strategies to enhance the expression of enzymes through the synthesis of EAAs or of proteins with desirable amino acid profiles often meet with unforeseen difficulty resulting from two main sources: (i) gene silencing of both the introduced gene and any naturally occurring homologue by a phenomenon known as co-suppression; and (ii) feedback responses designed to maintain cell homeostasis, including inhibition of synthetic pathways and activation of catabolic pathways (Zhu and Galili, 2004). Recent efforts to enhance the nutritional quality of plant protein are beginning to address these problems with unique approaches that utilize

our increasing understanding of regulatory pathways in plants.

In order to exploit a possible synergy between the metabolic engineering of free EAA levels on the one hand, and the removal of free EAAs from the cell via their inclusion into proteins on the other, two seed companies have proposed a combined approach. Falco *et al.* (1999) reported the transformation of maize with several genes coding for feedback-insensitive enzymes. Together with these enzyme genes, another coding for a methionine-rich 10-kDa zein from maize (Mertz *et al.*, 1964) was integrated to act as a protein 'sink' for the excess methionine produced. This combined strategy, coined 'binary system', was also claimed by Beach and Tarczynski (2000) using a Brazil nut protein as a sink protein. In another variation of the binary approach, the metabolic engineering using feedback-insensitive enzymes was replaced by an RNAi strategy. The decrease in low-methionine 19- and 22-kDa zein expression, combined with the introduction of the methionine-rich 10-kDa zein encoding gene, resulted in a several-fold increase in the total methionine content of seeds (Kirihaara *et al.*, 2001).

Workers from a major seed supplier have patented an interesting combined approach to improve the free amino acid content and to provide a protein sink that they claim solves the problem of a limited sulphur pool (Rudolf *et al.*, 2005). The strategy involves the down-regulation of the genes for five low-quality zein proteins, as well as the down-regulation of the lysine ketoglutarate reductase gene, using inverted repeats of a fusion of the five genes. The down-regulation of these genes is coupled with the over-expression of  $\alpha$ -globulin, 15-kDa  $\beta$ -zein and 18-kDa  $\delta$ -zein with various promoters. This combined approach is based on the premise that, without the suppression of genes for proteins with a poor EAA profile, the EAAs would be depleted to make undesirable proteins. Meals prepared from plants with transgenic constructs had a slightly lower seed protein content, but their lysine and tryptophan contents were increased by 70% and 60%, respectively.

A slightly different strategy, yet one that relies on a combination of two approaches, was taken by Takaiwa and Tada (2004). Using the full-length 5' untranslated region of the glutelin *GluB-1* gene, the levels of recombinant glycinin protein and mRNA were increased. Use of a crop mutant defective in a storage protein ensures that the maximal capacity for transcription and translation (i.e. the EAA pool) is not exhausted. A two- to sixfold increase in glycinin in rice was observed, but no EAA profile was provided. This 'maximal capacity' story is in line with the sulphur shortage that others (Brunold and Perez, 2004) have tried to address. In this case,

a plant was transformed with a cassette for the expression of a sulphate assimilation pathway. High levels of adenosine 5'-phosphosulphate reductase were obtained in *Lemna minor*, a result which the researchers claim will address the issue of the availability of EAAs and other building blocks for nutritive protein production. Other binary systems using *de novo* designed (or 'designer') proteins have been developed and are discussed below.

## Engineering novel proteins as sources of EAAs

### Simple repetitive polypeptides

The end-product inhibition of the metabolic enzymes involved in EAA synthesis has been the root of problems encountered in most strategies for the enhancement of crop EAA content reported to date. Although it is possible to engineer or import enzymes with reduced sensitivity to such inhibition, it is simpler to remove the end-product causing the inhibition by using it in protein synthesis. Although natural storage proteins appear to be the normal choice for sink proteins in most strategies, the potential of synthetic sink proteins or polypeptides has also captured the imagination of a large number of scientists. Bacterial cells were first explored as sources of such synthetic sink proteins. These EAA-rich proteins could be used to supplement diets that were deficient in only certain amino acids. Such feeding practices are common with livestock. As the simplest example, one could elect to express a chain of a selected EAA to remove it from the free AA pool. For this, a cloned synthetic gene containing several copies of a single codon would, in theory, result in the synthesis of a homopolymer 'sink' for the encoded amino acid. Kangas *et al.* (1982) were successful in using this concept to produce a proline-rich protein in the bacterium *E. coli*. A synthetic DNA for poly(Asp-Phe) was also cloned and shown to express (Doel *et al.*, 1980), as was a poly(Glu-Glu-Pro-Glu-Tyr-Gly-Glu) encoding sequence (Niehrs *et al.*, 1992). An attempt to express an artificial gene encoding poly(Phe-Trp-Pro-Lys), however, was unsuccessful (Gupta *et al.*, 1983). Biernat and Koster (1987) constructed a new plasmid to express synthetic genes for proteins of enhanced nutritional quality. In addition, Jaynes *et al.* (1985) reported the bacterial expression of repetitive peptides, coded in genes prepared from the random ligation of small oligonucleotides. In this latter case, radiolabelling was necessary for the detection of the EAA-rich peptides. The same polypeptides were mentioned later in a patent (Jaynes and Derrick, 1998) as potential sink proteins for crops, but no precise account of the expression yields

from various plants transformed with such constructs was reported.

In each of the successful cases, the levels of protein expression were not particularly high despite the fact that the copy number of the encoding gene was quite sufficient to support high protein production. As in the plant seed storage protein experience, the problems appeared to be with the inherent instability of the protein product rather than with any inefficiency in translation of the mRNA into protein. Clearly, a rational approach to the enhancement of protein stability *in vivo* needed to be integrated into this concept of simple synthetic sink polypeptides.

In 1986, the 'N-terminal rule' (which stresses the importance of the N-terminal amino acid residue in conferring proteolytic stability to a protein) was proposed as the key to stable protein expression (Bachmair *et al.*, 1987). Biernat and Koster (1987) suggested that it was the N-terminal amino acid sequence, rather than the identity of only the first amino acid in that sequence, that was significant for the correct folding of the polypeptide chain. In our laboratory, we attempted to express synthetic genes encoding repeating pentapeptides of nutritional importance fused to the N-terminal sequence of  $\beta$ -galactosidase (referred to as  $\alpha$ -peptide). The  $\alpha$ -peptide is the same amino acid sequence that provides the N-terminal folding signals for the stable accumulation of  $\beta$ -galactosidase in the native enzyme, and thus should fulfil the requirements of both variations of the 'N-terminal rule'. However, we were unable to detect any of the recombinant proteins without resorting to a sensitive, antibody-based detection system (Beauregard *et al.*, 1994). Because the levels of mRNA transcript for each of our fusion proteins were extremely high, it appeared that protein degradation rather than synthesis was limiting the level of accumulation of the gene product. Indeed, when the same proteins were expressed in protease-deficient bacterial mutants, the *de novo* designed proteins accumulated. Predictive methods indicated that all of our *de novo* designed proteins were expected to be unstructured in solution and thus likely targets for proteolytic degradation. Thus, in bacteria, as well as in plants, protein structure is an important factor to consider in designing new or altered proteins of nutritional value.

Recently, a patent has been filed describing the design of chimeric proteins consisting of an N-terminal carrier protein and various poly amino acid extensions at the C-terminal (Wahlroos *et al.*, 2004). Up to six copies of these simple oligonucleotide repeats (coding for about 10–20 EAAs) were polymerized, in a manner somewhat reminiscent of early (and largely unsuccessful) efforts with modifications to naturally occurring phaseolins cited above. Given the success

in expressing high levels of at least one such protein (with a polyhistidine extension), it appears that the carrier protein (tobacco mosaic virus 30-kDa movement protein) may fulfil the 'N-terminal rule' and, unlike the phaseolins, its folding may not be obstructed by the addition of amino acids at its C-terminus.

### *De novo* design of novel nutritive proteins

The development of novel nutritive proteins was made possible thanks to advances in the area of '*de novo* designed' proteins. In this area, researchers attempted to achieve a specific fold using simple *ab initio* design principles. The production and characterization of such folded polypeptides allowed the expansion of our knowledge of protein stability and of protein folding determinants and provided valuable information for protein design and protein engineering projects. *De novo* proteins have been designed to form some of the simpler tertiary structure motif: the four  $\alpha$ -helix bundle (e.g. DeGrado, 1988; Hecht *et al.*, 1990), coiled-coil proteins (e.g. Monera *et al.*, 1996), an  $\alpha/\beta$ -barrel protein (e.g. Goraj *et al.*, 1990) and all- $\beta$  proteins (e.g. Quinn *et al.*, 1994). Recent work also includes the design of all- $\beta$  structures, particularly those that are able to avoid aggregation into amyloid-like sheets (Wang and Hecht, 2002).

On the basis of such advances, the ambitious goal of creating efficient sink proteins based on repetitive EAA sequences was taken closer to reality when Falco *et al.* (1998) patented the use of repetitive, EAA-rich peptides with a sequence that would fold into helical structures and adopt a dimeric tertiary organization. In our group, we moved a step further and encoded folding information in the first nutritive *de novo* protein reported: Milk Bundle-1 (MB-1) (Beauregard *et al.*, 1995). The basic assumptions for MB-1 design were as follows: (i) the range of possible residues in each recurring position in the helices would be determined by the physicochemical properties of the amino acid residues and the topological constraints resulting from their particular location in the tertiary structure; (ii) the driving force for bundle formation would be a 'hydrophobic collapse', where hydrophobic side-chains are buried in the interior of the bundle; and as a result (iii) the loops would be relatively unimportant. The design process was iterative, in which criteria such as secondary structure propensities, the distribution of charged residues and other global factors found to be determinant for folding were subsequently used to refine the initial design (Beauregard *et al.*, 1995; and references cited therein). The resulting protein was expected to fold into a four-helix bundle, whilst hosting a heavily biased EAA content (67 of the 100 amino

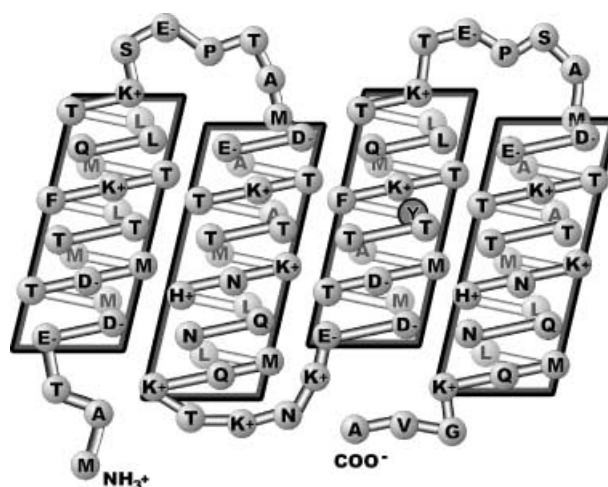


**Figure 1** Flattened view of Milk Bundle-1 (MB-1), external face. Helical segments are represented as four parallelograms. Core residues are shown shaded in the background. Tyrosine in core position 62, mutated for a tryptophan in MB-1Trp, is highlighted. Figure from Beauregard *et al.* (1995), reprinted with permission of Nature Publishing Group.

acid residues were lysine, methionine, leucine or threonine; see model in Figure 1). The protein was expressed from a synthetic gene introduced into recombinant *E. coli* at high levels (7%–12% of the total cellular protein). Its behaviour and expression levels *in vivo* were found to be far superior to those obtained in earlier attempts of high-EAA polypeptide design (MacCallum *et al.*, 1997).

The expression of MB-1 as a sink protein in transgenic crops has been attempted in three different plants. Kirihaara *et al.* (2001) designed a strategy based on antisense RNA technology (to repress the expression of low-quality proteins), coupled with the use of a sink protein (for EAAs), in maize. Transformation of maize with the *MB-1* gene was reported, but information on expression was not divulged. MB-1 has also been used in an attempt to improve the quality of soy protein. Simmonds and Donaldson (2000) successfully targeted the MB-1 protein to seed protein storage bodies using seed-specific promoters and leader sequences. The levels of protein expressed, although adequate, were not outstanding, and work is ongoing to improve the amount of synthetic protein deposited in the seed using MB-1 derivatives that fold more stably and compactly (D. Simmonds, Agriculture and Agri-food Canada, personal communication). More recently, Khoudi and Beauregard (2005) attempted to express a derivative of MB-1 in alfalfa. In spite of a high efficiency of transformation and the detection of large amounts of mRNA, no *de novo* designed protein could be detected using normal immunological methods.

Detailed analysis of these experiments indicates that the design for expression and stability in one system (bacteria)

may not always lead to facile achievement of expression and stability in another (transgenic plants). Although the basic determinants of protein stability (compact fold, lack of exposed protease-susceptible sites, etc.) are common in all organisms, the specificities of each system (specific proteases, transit and deposition mechanisms) differ.

Further improvement of the folding and stability of the MB-1 protein has been reported that should result in increased accumulation of the protein in the production host. Efforts to improve MB-1 stability have considered several potential contributing factors: core packing (Gagnon *et al.*, 2000), helix 'capping' (Parker and Hefford, 1998), loop optimization (Simons *et al.*, 2001), addition of salt bridges (Sasseville *et al.*, 2004) and addition of disulphide bridges (Doucet *et al.*, 2002). Some of these projects have resulted in substantial improvement in the properties of MB-1 as a potential nutritive protein. Further, they have revealed a strong correlation between resistance to bacterial proteolysis and conformational stability. Transgenic crop expression experiments of some improved MB-1 variants have not resulted in enhanced protein levels, however. The reasons for these disappointing results in transgenic crops remain to be elucidated.

As designers of potential nutritive proteins, however, scientists are faced with an apparent dichotomy: those factors considered to be important for the stabilization of most of the naturally occurring proteins characterized to date (stability, solubility, specificity of fold, ability to remain non-aggregated) may not be conducive to the expression and deposition in crops, particularly in seeds (in which natural nutritive proteins are not particularly soluble). It must be considered that a designer nutritive protein may be exposed to various cellular backgrounds during its development: in *E. coli* at the laboratory scale, in production microbes (inoculant, rumen microbe) and/or in transgenic crops for large-scale expression.

Although rational approaches, such as *de novo* design and site-directed mutagenesis, have proven to be useful, they cannot compete with powerful approaches of directed evolution techniques. In this context, Claveau *et al.* (2004) have identified mutagenic PCR conditions that provide a high bias on both sides of the mutagenesis process (i.e. nucleotide

mutation, nucleotide insertion), leading to enrichment in adenine and thymine. The method generates novel variants with different properties, and favours enrichment in EAAs when compared with standard error-prone PCR (Cadwell and Joyce, 1992), as described in Table 1. The use of such mutagenic methods with the correct screening approach may provide an answer to the challenges described above when adapting a novel nutritive protein to expression in crops.

## General conclusion and future directions

The potential of biotechnology to help meet the ever-increasing nutritional needs of both livestock and humans has long been recognized. The advances in this field, as evidenced by both the quantity and quality of the research efforts discussed in this article, are impressive. Indeed, a number of major companies have filed patents for genetically engineered plants with improved nutritive value. For now, however, no crops with tailored EAA content have made it to the market, and a number of hurdles remain. Some of the modified plants produced to date have had difficulty with allergenicity. In other cases, the difference in nutritive value of the genetically altered plants and their naturally occurring counterparts is too small, and protein and/or EAA production and deposition need to be enhanced to make such crops viable alternatives to current feeds and foods. Clearly, a better understanding of protein transit, association and deposition in the many cellular compartments of plants will help achieve the yields that will make such crops cost efficient. Even as our knowledge in these areas continues to increase, concerns regarding proteolytic stability, solubility and allergenicity will need to be addressed. Here, both rational, iterative design and strategies using high-throughput screening of multiple random mutations promise to be useful. Advances in metabolic engineering, in particular combined methods involving the simultaneous enhancement and inhibition of anabolic and catabolic pathways for EAAs, coupled with the production of protein sinks, indicate that this strategy will become an increasingly viable option. As our knowledge of the structure–function relationships of specific storage proteins develops, protein engineering will continue to play an important role.

| Method                                 | Total mutations | New EAA codons | New non-EAA codons | Other mutations | Net change in EAA codons (%) |
|----------------------------------------|-----------------|----------------|--------------------|-----------------|------------------------------|
| Standard method                        | 138             | 38             | 76                 | 24              | –28                          |
| Method of Claveau <i>et al.</i> (2004) | 73              | 25             | 20                 | 28              | +7                           |

EAA, essential amino acid; MB-1, Milk Bundle-1.

\*This gene contains a 49/51 content of (A or T) vs. (G or C) prior to mutagenesis.

**Table 1** Comparison of mutations using two error-prone polymerase chain reaction (PCR) methods performed on *MB-1Trp* gene\*

The *de novo* design of proteins, although still a relatively young field, offers much scope for diversity to meet specific nutritional challenges. All of these approaches will only be augmented as knowledge in the basic sciences advances. In addition, customized evolution techniques, such as those put forward by Claveau *et al.* (2004) and Rao and Sleister (2005), offer the potential to produce novel nutritive proteins even in the absence of a complete knowledge of their underlying biochemistry.

What then is the future likely to hold for proponents of novel nutritive crops? Based on the progress achieved during the last decade, one can forecast the coming of a new generation of crops with tailored EAA profiles in the very near future. These new crops will undoubtedly lead to further developments in the agro-food industry, perhaps even towards nutrigenomics. Particularly in the developed world, there is an ever-increasing awareness of the value of nutrition, and specific nutritional regimes for meeting both health and fitness challenges in humans and production issues in livestock and other domestic animals. The successful application of nutrigenomics demands that the food industry develop a new generation of functional foods tailored to meet the needs of individuals with specific genetic traits. This new approach can also be applied to whole populations based on their nutritional habits. Here, too, biotechnology holds much promise for a future in which a wide variety of efficient crops with tailored EAA profiles are readily available to meet both existing and impending challenges in the agro-food industry.

## References

- Altenbach, S.B., Kuo, C.-C., Staraci, L.C., Pearson, K.W., Wainwright, C., Georgescu, A. and Townsend, J. (1992) Accumulation of Brazil nut albumin in seeds of transgenic canola results in enhanced levels of seed protein methionine. *Plant Mol. Biol.* **18**, 235–245.
- Ampe, C., Van Damme, J., deCastro, L., Sampaio, M.J., Van Montagu, M. and Vanderkerckhove, J. (1986) The amino acid sequence of the 2S sulfur-rich proteins from seeds of Brazil nut (*Bertholletia excelsa* HBK). *Eur. J. Biochem.* **159**, 597–604.
- Aragao, F.J., Grossi de Sa M.F., Almeida, E.R., Gander, E.S. and Rech, E.L. (1992) Particle bombardment-mediated transient expression of Brazil nut methionine-rich albumin in bean (*Phaseolus vulgaris* L.). *Plant Mol. Biol.* **20**, 357–359.
- Ashes, J.R., Gulati, S.K. and Scott, T.W. (1995) The role of rumen-protected proteins and energy sources in the diet of ruminants. In *Animal Science Research and Development: Moving Toward a New Century* (Ivan, M., ed.), pp. 177–188. Ottawa: Agriculture and Agri-Food Canada.
- Avraham, T., Badani, H., Galili, S. and Amir, R. (2005) Enhanced levels of methionine and cysteine in transgenic alfalfa (*Medicago sativa* L.) plants over-expressing the *Arabidopsis* cystathionine  $\gamma$ -synthase gene. *Plant Biotechnol. J.* **3**, 71–79.
- Bachmair, J., Hasselmann, H., Hofer, B., Kennedy, N. and Koste, H. (1987) In vivo half-life of a protein is a function of its amino-terminal residue. *Science*, **234**, 179–186.
- Bagga, S., Adams, H., Kemp, J.D. and Sendgupta-Gopalan, C. (1995) Accumulation of 15-kilodalton zein in novel protein bodies in transgenic tobacco. *Plant Physiol.* **107**, 13–23.
- Bagga, S., Sengupta-Gopalan, C. and Kemp, J.D. (1999) Co-expression of proteins. US Patent 5,990,384.
- Beach, L. and Tarczynski, M.C. (2000) Methods of increasing accumulation of essential amino acids in seeds. US Patent 6,127,600.
- Beachy, R.N., Chen, Z.-L., Bray, E., Lawton, M., Koemda, Y., Nakamura, I., Naito, S., Anderson, E., Dube, P., Rogers, S.G. and Fraley, R.T. (1986) Expression of subunits of beta-conglycinin in transgenic tobacco and petunia plants. In *Molecular Biology of Seed Storage Proteins and Lectins* (Shannon, L. and Chrispeels, M., eds), pp. 53–63. Boston: Jones and Bartlett.
- Beauregard, M., Hefford, M.A. and Teather, R.M. (1994) Use of a monoclonal antibody to beta-galactosidase for the detection of alpha peptide fusions. *Biotechniques*, **16**, 832–839.
- Beauregard, M., Dupont, C., Teather, R.M. and Hefford, M.A. (1995) Design, expression and initial characterization of MB-1, a *de novo* protein enriched in essential amino acids. *Biotechnology (N.Y.)*, **13**, 974–981.
- Biernat, J. and Koster, H. (1987) Expression of synthetic genes coding for completely new, nutritionally rich, artificial proteins. *Protein Eng.* **1**, 353–358.
- Brime, C.A. and Buxton, J.W. (1979) Recurrent selection in soybean. II. Selection for increased protein in seeds. *Crop Sci.* **19**, 494–498.
- Brochetto-Braga, M.R., Leite, A. and Arruda, P. (1992) Partial purification and characterization of lysine-ketoglutarate reductase in normal and opaque-2 maize endosperm. *Plant Physiol.* **98**, 1139–1147.
- Brunold, C. and Perez, P. (2004) Method for obtaining plants enriched in cysteine and glutathione content. US Patent Application 20040123341.
- Cadwell, R.C. and Joyce, G.F. (1992) Randomization of genes by *in vitro* mutagenesis. *PCR Meth. Appl.* **2**, 28–33.
- Carter, T.E., Burton, J.W. and Brime, C.A. (1982) Recurrent selection for percent protein in soybean seed – indirect effects on plant N accumulation and distribution. *Crop Sci.* **22**, 513–519.
- Cerioli, A., Pedrazzini, E., Fabbrini, M.S., Zoppe, M., Bollini, R. and Vitale, A. (1991) Expression of the wild-type and mutated vacuolar storage protein phaseolin in *Xenopus* oocytes reveals relationships between assembly and intracellular transport. *Eur. J. Biochem.* **202**, 959–968.
- Chakraborty, S., Chakraborty, N. and Datta, A. (2002) Increased nutritive value of transgenic potato by expressing a nonallergenic seed albumin gene from *Amaranthus hypochondriacus*. *PNAS-USA*, **97**, 3724–3729.
- Chui, C.-F.C., Falco, S.C., Rice, J.A. and Knowlton, S. (1999) High sulfur seed protein gene and method for increasing the sulfur amino acid content of plants. US Patent 5,939,599.
- Chui, C.-F.C., Falco, S.C., Rice, J.A. and Knowlton, S. (2003) High-sulphur seed protein gene and method for increasing the sulphur amino acid content of plants. Canadian Patent CA 2104022.
- Clark, J.H., Kusmeyer, T.J. and Cameron, M.R. (1992) Microbial protein synthesis and flows of nitrogen fractions to the duodenum of dairy cows. *J. Dairy Sci.* **75**, 2304–2323.

- Claveau, S., Sasseville, M. and Beauregard, M. (2004) Alcohol mediated error-prone PCR. *DNA Cell Biol.* **23**, 789–795.
- DeClercq, A., Vandewiele, M., Van Damme, J., Guerche, P., Van Montagu, M., Vanderkerkhove, J. and Kebbers, E. (1990) Stable accumulation of modified 2S albumin seed storage proteins with higher methionine contents in transgenic plants. *Plant Physiol.* **94**, 970–979.
- DeGrado, W.F. (1988) Design of peptides and proteins. *Adv. Protein Chem.* **39**, 51–124.
- Denk, D. and Böck, A. (1987) 1-cysteine biosynthesis in *Escherichia coli*: nucleotide sequence and expression of the serine acetyltransferase (*cysE*) gene from the wild-type and a cysteine-excreting mutant. *J. Gen. Microbiol.* **133**, 515–525.
- Dijkstra, D.S., Linneman, A.R. and van Boekel, T.A. (2003) Towards sustainable production of protein-rich foods: appraisal of eight crops for Western Europe. Part II: Analysis of the technological aspects of the production chain. *Crit. Rev. Food Sci. Nutr.* **43**, 481–506.
- Doel, M.T., Eaton, M., Dook, E.A., Lewis, H., Patel, T. and Carey, N.H. (1980) The expression in *E. coli* of synthetic repeating polymeric genes coding for poly(1-aspartyl-L-phenylalanine). *Nucleic Acids Res.* **8**, 4575–4592.
- Doucet, A., Williams, M., Gagnon, M.C., Sasseville, M. and Beauregard, M. (2002) Engineering nutritious proteins: improvement of stability in the designer protein MB-1 via introduction of disulfide bridges. *J. Agric. Food Chem.* **50**, 92–98.
- Dyer, J.M., Nelson, J.W. and Muraj, N. (1993) Strategies for selecting mutation sites for methionine enhancement in the bean seed storage protein phaseolin. *J. Protein Chem.* **12**, 545–560.
- Edwards, G.A., Hephner, A., Clerk, S.P. and Boulter, D. (1991) Pea lectin is correctly processed, stable and active in the leaves of transgenic potato plants. *Plant Mol. Biol.* **17**, 89–100.
- Falco, S.C., Keeler, S.J. and Rice, J.A. (1998) Chimeric genes and methods for increasing the lysine and threonine content of the seeds of plants. US Patent 5,773,691.
- Falco, S.C., Guida, A.D. Jr and Hartnett, M.E. (1999) Nucleic acid fragments, chimeric genes and methods for increasing the methionine content of the seeds of plants. US Patent 5,912,414.
- FAO/WHO/UNU (1985) Energy and protein requirements: Report of a joint FAO/WHO/UNU Expert Consultation. In *WHO Technical Report Series No. 724*. Geneva: World Health Organisation.
- Flachowsky, G., Chesson, A. and Aulrich, K. (2005) Animal nutrition with feeds from genetically modified plants. *Arch. Anim. Nutr.* **59**, 1–40.
- Frankard, V., Ghislain, M. and Jacobs, M. (1992) Two feedback-insensitive enzymes of the aspartate pathway in *Nicotiana glauca*. *Plant Physiol.* **99**, 1285–1293.
- Gagnon, M.C., Williams, M., Doucet, A. and Beauregard, M. (2000) Replacement of Tyr-62 by Trp in the designer protein Milk Bundle-1 results in significant improvement of conformational stability. *FEBS Lett.* **484**, 144–148.
- Ghislain, M., Frankard, V. and Jacobs, M. (1995) A dinucleotide mutation in DHDPs of *Nicotiana glauca* leads to lysine overproduction. *Plant J.* **18**, 733–743.
- Gibbon, B.C. and Larkins, B.A. (2005) Molecular genetic approaches to developing quality protein maize. *Trends Genet.* **21**, 227–233.
- Goraj, K., Renard, A. and Martial, J.A. (1990) Synthesis, purification and initial structural characterization of octarellin, a de novo polypeptide modelled on the  $\alpha/\beta$ -barrel proteins. *Protein Eng.* **3**, 256–266.
- Guerche, P., De Almeida, E.R.P., Schwarstein, M.A., Gander, E., Krebbers, E. and Pelletier, G. (1990) Expression of the 2S-albumin from *Bertholletia excelsa* in *Brassica napus*. *Mol. Gen. Genet.* **221**, 306–314.
- Gupta, S.C., Weigh, H.L. and Somerville, R.L. (1983) Biological limitations on the length of highly repetitive DNA sequences that may be stably maintained within plasmid replicons in *E. coli*. *Bio/Technology*, **1**, 602–609.
- Habben, J.E. and Larkins, B.A. (1995) Genetic modification of seed proteins. *Curr. Opin. Biotechnol.* **6**, 171–174.
- Hecht, M.H., Richardson, J.S., Richardson, D.C. and Ogden, R.C. (1990) De novo design, expression, and characterization of Felix: a four-helix bundle of native-like sequence. *Science*, **249**, 884–891. [Erratum in: (1990) *Science*, **249**, 973.]
- Hesse, H., Harms, K. and Hofgen, R. (2003) Means and methods for enhancing the content of sulfur compounds in plants. US Patent 6,608,239.
- Hoffman, L.M. (1996) Modified 7S legume seed storage proteins. US Patent 5,576,203.
- Huang, S., Kruger, D.E., Frizzi, A., D'Ordine, R.L., Florida, C.A., Adams, W.R., Brown, W.E. and Luethy, M.H. (2005) Altering lysine metabolism: high-lysine corn produced by the combination of enhanced lysine biosynthesis and reduced zein accumulation. *Plant Biotechnol. J.* **3**, 555.
- Hungate, R.E. (1966) *The Rumen and its Microbes*. New York: Academic Press.
- Jaynes, J.M. and Derrick, K.S. (1998) Plants genetically enhanced for nutritional quality. US Patent 5,811,654.
- Jaynes, J.M., Langridge, P., Anderson, K., Bond, C., Sands, D., Newman, C.W. and Newman, R. (1985) Construction and expression of synthetic DNA fragments coding for polypeptides with elevated levels of essential amino acids. *Appl. Microbiol. Biotechnol.* **21**, 200–205.
- Kami, J.A. and Gepts, P. (1994) Phaseolin nucleotide sequence diversity in phaseolins. I. Intraspecific diversity in *Phaseolus vulgaris*. *Genome*, **37**, 751–757.
- Kangas, T.T., Cooney, C.L. and Gomez, R.F. (1982) Expression of a proline-enriched protein in *E. coli*. *Appl. Environ. Microbiol.* **43**, 629–635.
- Karchi, H., Shaul, O. and Galili, G. (1994) Lysine synthesis and catabolism are co-ordinately regulated during tobacco seed development. *PNAS-USA*, **91**, 2577–2581.
- Khan, M.R., Ceriotti, A., Tabe, L., Aryan, A., McNabb, W., Moore, A., Craig, S., Spencer, D. and Higgins, T.J. (1996) Accumulation of a sulphur-rich seed albumin from sunflower in the leaves of transgenic subterranean clover (*Trifolium subterraneum* L.). *Transgenic Res.* **5**, 179–185.
- Khoudi, H. and Beauregard, M. (2005) The *de novo* designed protein MB-1Trp does not accumulate in transgenic alfalfa plants. *Plant Physiol. Biochem.* **43**, 1039–1043.
- Kim, C.-S., Kamiya, S., Sato, T., Utsumi, S. and Kito, M. (1990) Improvement of nutritional value and functional properties of soybean glycinin by protein engineering. *Protein Eng.* **3**, 725–731.
- King, K.J., Huber, J.T., Sadik, M., Bergen, W.G., Grant, A.L. and King, V. (1990) Influence of dietary protein sources on the amino acid profiles available for digestion and metabolism in lactation cows. *J. Dairy Sci.* **73**, 3208–3216.
- Kirihara, J.A., Hibberd, K.A. and Janice, A. (2001) Method for altering the nutritional content of plant seed. US Patent 6,326,527.

- Kjemtrup, S., Herman, E.M. and Chrispeels, M.J. (1994) Correct post-translational modification and stable vacuolar accumulation of phytohemagglutinin engineered to contain multiple methionine residues. *Eur. J. Biochem.* **226**, 385–391.
- Kortt, A.A., Caldwell, J.B., Lilley, G.G. and Higgins, T.J.V. (1991) Amino acid and cDNA sequences of a methionine-rich 2S protein from sunflower seed (*Helianthus annuus* L.). *FEBS Lett.* **195**, 329–334.
- Kreft, O., Höfgen, R. and Hesse, H. (2003) Functional analysis of cystathionine-synthase in genetically engineered potato plants. *Plant Physiol.* **131**, 1843–1854.
- Krishana, H.B. (2005) Engineering soybean for enhanced sulphur amino acid content. *Crop Sci.* **45**, 454–461.
- Loeffler, C.M., Busch, R.H. and Wiersma, J.V. (1983) Recurrent selection for grain protein percentage in hard red spring wheat. *Crop Sci.* **23**, 1097–1101.
- Ma, Y. and Bliss, F.A. (1978) Seed proteins of common bean. *Crop Sci.* **17**, 431–437.
- MacCallum, J.D., Hefford, M.A., Omar, S. and Beauregard, M. (1997) Prediction of folding stability and degradability of the *de novo* designed protein MB-1 in cow rumen. *Appl. Biochem. Biotechnol.* **66**, 83–93.
- Marcellino, L.H., Neshich, G., Grossi de Sa, M.F., Krebbers, E. and Gander, E.S. (1996) Modified 2S albumins with improved tryptophan content are correctly expressed in transgenic tobacco plants. *FEBS Lett.* **385**, 154–158.
- Masumura, T., Shibata, D., Hibino, T., Kato, T., Kawabe, K., Takeba, G., Tanaka, K. and Fujii, S. (1989) cDNA cloning of an mRNA encoding a sulfur-rich 10 kDa prolamin polypeptide in rice seeds. *Plant Mol. Biol.* **12**, 123–130.
- Mazur, B., Krebbers, E. and Tingey, S. (1999) Gene discovery and product development for grain quality traits. *Science*, **285**, 372–375.
- Menendez-Arias, L., Moneo, I., Dominguez, J. and Rodriguez, R. (1988) Primary structure of the major allergen of yellow mustard (*Sinapis alba* L.) seed, Sina 1. *Eur. J. Biochem.* **177**, 159–166.
- Mertz, E.T., Bates, L.S. and Nelson, O.D. (1964) Mutant gene that changes protein composition and increases lysine content of maize endosperm. *Science*, **145**, 279–280.
- Messing, J. and Lai, J. (2005) Method for producing high methionine corn seeds. US Patent 6,849,779.
- Molvig, L., Tabe, L.M., Eggum, B.O., Moore, A.E., Craig, S., Spencer, D. and Higgins, T.J. (1997) Enhanced methionine levels and increased nutritive value of seeds of transgenic lupins (*Lupinus angustifolius* L.) expressing a sunflower seed albumin gene. *PNAS-USA*, **94**, 8393–8398.
- Monera, O.D., Zhou, N.E., Lavigne, P., Kay, C.M. and Hodges, R.S. (1996) Formation of parallel and anti-parallel coiled-coils controlled by the relative positions of alanine residues in the hydrophobic core. *J. Biol. Chem.* **271**, 3995–4001.
- Nefrutin, I., Cattoir-Reynaerts, A., Verbruggen, I. and Jacobs, M. (1984) Lysine over-producer mutants with an altered dihydrodipicolinate synthase from protoplast culture of *Nicotiana sylvestris* (Speggazzini and Comes). *Theor. Appl. Genet.* **68**, 11–20.
- Nelson, O.E. (1969) Genetic modification of protein quality. *Adv. Agron.* **21**, 171–194.
- Niehrs, C., Huttner, W.B. and Ruther, U. (1992) In vivo expression and stoichiometric sulfation of the artificial protein sulfophilin, a polymer of tyrosine sulfation sites. *J. Biol. Chem.* **267**, 15 938–15 942.
- Nordlee, J.A., Taylor, S.L., Townsend, J.A., Thomas, L.A. and Bush, R.K. (1996) Identification of a Brazil nut allergen in transgenic soybean. *N. Engl. J. Med.* **394**, 688–692.
- Nuttall, J., Vitale, A. and Frigerio, L. (2003) C-terminal extension of phaseolin with a short methionine-rich sequence can inhibit trimerisation and result in high instability. *Plant Mol. Biol.* **51**, 885–894.
- Ohtani, T., Galili, G., Wallace, J.C., Thompson, G.A. and Larkins, B.A. (1991) Normal and lysine-containing zeins are unstable in transgenic tobacco-seeds. *Plant Mol. Biol.* **16**, 117–128.
- Parker, M. and Hefford, M.A. (1998) Introduction of potential helix-capping residues into an engineered helical protein. *Biotechnol. Appl. Biochem.* **28**, 69–76.
- Pedersen, K., Argos, P., Naravana, S.V. and Larkins, B.A. (1986) Sequence analysis and characterization of a maize gene encoding a high-sulfur zein protein of Mr 15 000. *J. Biol. Chem.* **261**, 6279–6284.
- Quinn, T.W., Tweedy, N.B., Williams, R.W., Richardson, J.S. and Richardson, D.C. (1994) Beta-doublet: De novo design, synthesis and characterization of a beta-sandwich protein. *PNAS-USA*, **91**, 8747–8751.
- Rao, A.G., Hassan, M. and Hempel, J.C. (1994) Structure–function validation of high lysine analogs of alpha-hordothionin designed by protein modeling. *Protein Eng.* **7**, 1485–1493.
- Rao, G.G. and Sleister, H.M. (2005) Compositions and methods for altering amino acid content of proteins. US Patent 6,905,877.
- Rapp, W.D., Peng, J., Nadig, G. and Venkatesh, T. (2005) Enhanced proteins and methods for their use. US Patent 6,936,696.
- Roesler, K.R. and Rao, A.G. (1999) Conformation and stability of barley chymotrypsin inhibitor-2 (CI-2) mutants containing multiple lysine substitutions. *Protein Eng.* **12**, 967–973.
- Roesler, K.R. and Rao, A.G. (2000) A single disulfide bond restores thermodynamic and proteolytic stability to an extensively mutated protein. *Protein Sci.* **9**, 1642–1650.
- Rudolf, J., Hu, W.-N., Meeley, R.B., Sewalt, V.J.H. and Nair, R. (2005) Grain quality through altered expression of seed proteins. US Patent Application 20050160488A1.
- Saalbach, I., Pickardt, T., Machemehl, F., Saalbach, G., Schieder, O. and Muntz, K. (1994) A chimeric gene encoding the methionine-rich 2S albumin gene of the Brazil nut (*Bertholletia excelsa* HBK) is stably expressed and inherited in transgenic grain legumes. *Mol. Gen. Genet.* **242**, 226–236.
- Sasseville, M., Claveau, S., Gagnon, M.C. and Beauregard, M. (2004) Single-point mutations at the surface of MB-1Trp lead to important changes in its conformational properties. *J. Peptide Res.* **64**, 1–9.
- Segal, G., Song, R. and Messing, J. (2003) A new opaque variant of maize by a single dominant RNA-interference-inducing transgene. *Genetics*, **165**, 387–397.
- Shotwell, M.A. and Larkins, B.A. (1991) Improvement in the protein quality of seeds by genetic engineering. In *Molecular Approaches to Crop Improvement* (Dennis, E.S. and Llewellyn, D.J., eds), pp. 33–61. New York: Springer-Verlag.
- Simmonds, D.H. and Donaldson, P.A. (2000) Genotype screening for proliferative embryogenesis and biolistic transformation of short-season soybean genotypes. *Plant Cell Rep.* **19**, 485–490.
- Simons, B., Scholl, D., Cyr, T. and Hefford, M.A. (2001) Effects of increased loop flexibility on the structure and stability of a *de novo* designed helical protein. *Prot. Pept. Lett.* **8**, 89–96.
- Sleister, H.M. and Rao, A.G. (2002) Subtractive immunization: a tool

- for the generation of discriminatory antibodies to proteins of similar sequence. *Immunol. Methods*, **261**, 213–220.
- Sparvoli, F., Daminati, M.G., Lioi, L. and Bollini, R. (1996) In vivo endoproteolytically cleaved phaseolin is stable and accumulates in developing *Phaseolus lunatus* L. seeds. *Biochim. Biophys. Acta*, **1292**, 15–22.
- Sun, S.S.M., Xiong, L., Jing, Y. and Liu, B. (2001) Lysine rich protein from winged bean. US Patent 6,184,437.
- Tabe, L.M. and Droux, M. (2002) Limits to sulphur accumulation in transgenic lupin seeds expressing a foreign sulphur-rich protein. *Plant Physiol.* **128**, 1137–1148.
- Tabe, L.M., Higgins, C.M., McNabb, W.C. and Hoggins, T.S.V. (1993) Genetic engineering of grain and pasture legumes for improved nutritive value. *Genetica*, **90**, 181–200.
- Takaiwa, F. and Tada, Y. (2004) Method of accumulating foreign gene product in plant seed at high level. US Patent Application 20040031075.
- Utsumi, S., Katsube, T., Ishige, T. and Takaiwa, F. (1997) Molecular design of soybean glycinins with enhanced food qualities and development of crops producing such glycinins. *Adv. Exp. Med. Biol.* **415**, 1–15.
- Uzogara, S.G. (2000) The impact of genetic modification of human foods in the 21st century: a review. *Biotechnol. Adv.* **18**, 179–206.
- Villegas, E.M., Eggum, B.O., Vasal, S.K. and Kohli, M.M. (1980) Progress in nutritional improvement of maize and triticale. In *Food and Nutrition Bulletin 2*. Tokyo, Japan: United Nations University Press.
- Wahlroos, T.M., Atabekov, J.G., Dorokhov, Y.L., Susi, P.E., Makela, M.J. and Korpela, T.K. (2004) Methods and constructs for increasing the content of selected amino acids in seeds. US Patent Application 20040177400.
- Wallace, J.C., Galili, G., Kawata, E.E., Cuellar, R.E., Shotwell, M.A. and Larkins, B.A. (1988) Aggregation of lysine-containing zeins into protein bodies in *Xenopus* oocytes. *Science*, **241**, 662–664.
- Wang, W. and Hecht, M.H. (2002) Rationally designed mutations convert de novo amyloid-like fibrils into monomeric  $\beta$ -sheet proteins. *PNAS-USA*, **99**, 2760–2765.
- Wilkins, T.A., Bednarek, S.Y. and Raikhel, N.V. (1990) Role of propeptide glycan in post-translational processing and transport of barley lectin to vacuoles in transgenic tobacco. *Plant Cell*, **2**, 301–313.
- Williams, P.E. (2003) Engineering plants for animal feed for improved nutritional value. *Proc. Nutr. Soc.* **62**, 301–305.
- Youle, R.J. and Huang, A.H. (1979) Albumin storage protein and allergens in cottonseeds. *J. Agric. Food Chem.* **27**, 500–503.
- Zheng, Z., Sumi, K., Tanaka, K. & Murai, N. (1995) The bean seed storage protein phaseolin is synthesized, processed, and accumulated in the vacuolar type 11 protein bodies of transgenic rice endosperm. *Plant Physiol.* **109**, 777–786.
- Zhu, X. and Galili, G. (2004) Lysine metabolism is concurrently regulated by synthesis and catabolism in both reproductive and vegetative tissues. *Plant Physiol.* **135**, 129–136.