7
Engineering Nutritious Proteins

MARY ALICE HEFFORD[1]

Centre for Food and Animal Research, Agriculture and Agri-food Canada, Ottawa, Canada K1A 0C6

Introduction

The word 'protein' will elicit a very different response when spoken to a nutritionist or a molecular biologist. For the former, it may bring to mind words like quality, quantity, digestibility, malnutrition or health; for the latter, words like sequence, structure, function or activity. These two seemingly disparate viewpoints, however, converge in recent efforts to use molecular biology techniques to augment the nutritive value of proteins in food and feedstuffs. The amino acid sequence of a protein (the order and identity of its component amino acids) dictates how the protein folds and functions, its structure and its activity. That same amino acid sequence determines the complement of amino acids available on protein digestion, the protein's nutritive value.

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' must be supplied from other sources. For humans and non-ruminant animals, dietary protein is the usual source of essential amino acids. In ruminants, some of the essential amino acid requirement can be met by bacterial fermentation of other feedstuffs within the rumen. 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 for making good quality protein available efficiently and economically becomes increasingly important.

Foodstuffs differ widely in their composition and protein quality. Animal proteins like meat and milk contain both the quantity and distribution of amino acids required to meet dietary requirements, and are usually considered 'complete' protein sources. Plant proteins, on the other hand, are usually limited in at least one essential amino acid and are considered 'partial' or 'incomplete' protein sources. Dietary requirements can

[1]Current address: Life Sciences Division, Bureau of Drug Research, Health Canada, Banting Building, Tunney's Pasture, Ottawa, Canada K1A OL2 (ph: 2201C), e-mail address: mhefford@emerald.hwc.ca.

Biotechnology and Genetic Engineering Reviews – Vol. 14, April, 1997
0264–8725/97/14/191–210 $20.00 + $0.00 © Intercept Ltd, P.O. Box 716, Andover, Hampshire SP10 1YG, UK

be met by complementing one 'incomplete' protein with a second incomplete protein rich in the missing amino acid(s). This is, of course, the basis for traditional meals from vegetarian cultures that often serve grains (in which lysine and tryptophan are limiting) with legumes (which, while rich in lysine and tryptophan, lack methionine and cysteine). It is also the reason that a variety of animals are routinely fed mixed rations.

The well-balanced diets of most individuals in developed countries result in adequate protein nutrition. In developing countries, the situation is much different. Production costs, the climatic restrictions on cropping possibilities, some traditional eating habits, and simple lack of quantity are all factors that militate against diets supplying adequate amounts of essential amino acids. The effects of inadequate protein diets are most striking in children, where the need for essential amino acid intake is increased by the demands for growth and development. The 1996 World Health Organization (WHO) Fact Sheet still reports that 40% of children in the least developed countries are underweight, and severe protein malnutrition or kwashiorkor syndrome still impairs the physical and mental development of many children worldwide.

On the agricultural front as well, there is ample evidence that the nutritional requirements for essential amino acids are not easily met in domestic animals. For both swine and poultry, for example, mixed diets of cereals and legumes do not completely meet protein requirements. Amino acid supplements are common. Even in ruminants, where microbial fermentation in the rumen augments the supply of essential amino acids, protein requirements for today's high-producing livestock are not readily met by simple mix rations. 'High quality' proteins, and 'protected' proteins and/or amino acids often supplement the feed. While these 'modern' feeding methods serve to maintain good nutritional balance for the animal, they are expensive, both in terms of the costs to the producer and the environmental costs of producing the nutrients. All of these factors are reflected in the growing scientific and industrial interest in augmenting the nutritive value of the world's protein sources.

Historical perspective

The idea of producing crops with enhanced nutritional value is not new. There is some natural variation in levels of limiting amino acids in many seed crops. Strains and varieties have been selectively bred on the basis of their essential amino acid content and, while many of these varieties are successfully marketed, the 'enhancement' of the limiting amino acid is usually insufficient to markedly affect feeding practices. Mutant varieties of maize with increased lysine content, for example, (the limiting amino acid in the zein corn proteins) have been identified (Mertz *et al.*, 1964; Nelson, 1969). Strains and varieties of other seed crops with enhanced levels of essential amino acids have also been produced by traditional breeding methods. Unfortunately, however, many of the varieties that show enhanced essential amino acid content also show undesirable agronomic or processing traits: they germinate poorly, dry-down slowly, have reduced yield or increased microbial susceptibility. (Glover and Mertz, 1987).

In the last 30 years, there has also been some effort invested in the augmentation of nutritional values in proteins during their processing. Proteins have been chemically

modified to add important amino acids like tryptophan (Puigserver *et al.*, 1986), cysteine (Friedman and Gumbmann, 1986), and methionine (Puigserver *et al.*, 1986). These methods, while successful to the extent they add essential amino acids, require a source of amino acid (often bacterial) and add considerably to the production cost. They also affect the taste of the final protein product, often negatively.

The advent of genetic engineering offered a real alternative solution to the traditional nutritional problem – proteins could potentially be designed or engineered to meet the animals' nutritional requirements.

Genetic modification of plants

Because plants provide the most universal food source, much of the research effort has been focussed on modifying the proteins within the plant to increase the complement of essential amino acids — 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 to meet the functional requirements of each tissue. Within individual cells of specific tissues, proteins are compartmentalized, again on the basis of their function. Genetic engineering of plants to enhance their nutritional value, therefore, requires not only vectors and transformation systems for the successful introduction of recombinant genes, but understanding and control of mechanisms of protein targeting and tissue-specific expression. With the growing body of knowledge in each of these component areas, protein engineering in plants is now not only possible but can be a practical alternative to traditional feeding and nutrition practices.

In the engineering of plants to meet nutritional needs, then, several factors must be considered. Of primary importance, of course, is the plant tissue or tissues in which new or modified proteins will be expressed as well as consideration of which proteins will be modified. Plant seeds and, to some extent, leaves are the major sources of dietary proteins in both human and animal nutrition. In addition to providing structural scaffolding, proteins in leaf are often associated with enzymatic actions important to tissue growth and function. Because the primary structure of a protein dictates not only its nutritional value (i.e., its amino acid complement) but also its structure-function relationships, modifications to the amino acids of specific proteins or to the amounts of those proteins in cells can have a significant, often deleterious, effect on the proper functioning of the cells where those proteins are expressed. Many of the proteins in seeds, on the other hand, have the sole function of providing nitrogen to the germinating embryo. Thus, seed storage proteins provide perhaps greater scope for modifications and additions without serious impairment of seed function. Modification of seed storage proteins, however, does require some knowledge and skill as amino acid composition can affect protein deposition (see below) as well as agronomic and processing traits.

MODIFICATION OF SEED STORAGE PROTEINS

The possible targets of genetic engineering to increase seed nutritional quality are many. In addition to amino acids comprising seed proteins (denoted 'protein-bound amino acids'), some of the amino acids in plant seeds available to be digested are 'free' in the cytosol of cells. These 'free' amino acids 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 provides a very real avenue to increasing nutritive value both in removing some of the limitations to protein synthesis and in enriching the 'free amino acid' content of plant seeds.

The second major approach possible is to change the levels of specific 'high quality' proteins within the plant tissue. In addition to increasing the availability of component amino acids, gene copy number and transcription rate for specific genes can be increased, the mRNA specifying the protein can be made more stable or the rate of its translation into protein increased, and the stability of the protein itself can be improved to encourage accumulation within the cell.

The third major route to increasing the nutritional quality of plant seeds is to increase the nutritional quality of the proteins synthesized there. New amino acids can be introduced into proteins already expressed in a particular plant and new proteins with more desirable complements of amino acids can be expressed. Each of these avenues has been explored and will be discussed.

Increasing amino acid synthesis

Because the free amino acid pool represents up to 5% of the total amino acids available for digestion in plant seeds, a substantial modification of the lysine and/or threonine content in the amino acid pool of cereal seeds (which are usually limited in these amino acids) could result in a significant improvement in their nutritional value (Ghislain *et al.*, 1995). Both high threonine- and high lysine-producing mutants have been identified in model systems (Negrutin *et al.*, 1984; Frankard *et al.*, 1991). In each case, the overproduction was shown to result from a modification to an enzyme in the amino acid biosynthetic pathway. In the maize mutant, opaque-2, the elevated lysine content and concomitant changes in protein composition have been shown to result, at least in part, from changes in the enzymes involved in lysine synthesis and degradation (Brochetto-Braga*et al.*, 1992). Barley mutants with altered lysine metabolic pathways have also been studied (Bright *et al.*, 1979).

These studies open the door to modifying enzymes involved in both the anabolism and catabolism of the desired amino acid. Lysine, in plants, is derived from aspartate, and its synthesis controlled by two main enzymes: aspartate kinase (AK) and dihydrodipicolinate synthetase (DHDPS). Both of these enzymes are regulated by end-product feedback inhibition by lysine and/or threonine and relief of this feedback inhibition results in overproduction of lysine. A lysine-insensitive form of DHDPS has been isolated from corn and characterized (Bittle *et al.*, 1992), and enzymes involved in the biosynthetic pathways for lysine have been cloned and/or identified from a variety of sources (Kaneko *et al.*, 1990; Silk and Matthews, 1992; Shur *et al.*, 1992; Perl*et al.*, 1993; Vauterin and Jacobs, 1994; Ghislain *et al.*, 1995). The enhanced lysine biosynthesis in mutant plants and in transgenic plants expressing bacterial genes involved in lysine biosynthetic pathways, however, appears to be associated with phenotypic abnormalities in the plant (Frankard *et al.*, 1992), and it would appear that this promising approach to enhancing both the free amino acid content and the amino acid pool for protein synthesis requires further refinement, possibly through the use of developmentally regulated promoters (Ghislain *et al.*, 1995).

Increasing levels of naturally occurring proteins

Seed storage proteins were originally classified on the basis of the solubilities of the extractable protein fractions. This resulted in four major classifications: albumins or water soluble proteins; globulins (water insoluble but saline soluble); prolamines (alcohol soluble) and glutelins (insoluble in water, saline or alcohol, but soluble in alkaline solutions). The distribution of proteins among these four major classifications differs with the seed type.

Not only do seeds from different species differ in their distribution of storage protein types, different varieties of seeds 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 (hence the name coined by Osborne (1924)), but low in lysine. In cereal seeds, approximately 50% of the extractable protein is prolamine and, while the essential amino acid content of each prolamine varies with the species and, to some extent, 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)). Wheat, rye, sorgum, millet and barley prolamines contain slightly more lysine than corn, but amounts are still insufficient to meet nutritional needs. Lysine content is also problematic in rice prolamine. The glutelin fraction of cereals has a higher lysine content (averaging 3.2 g lysine/100 g glutelin protein), but the glutelin content of these seeds is smaller than the prolamine content. Lysine is thus a limiting amino acid in all cereal-based diets. In each of these cereals except maize (where lysine and tryptophan are co-limiting), threonine is the second limiting amino acid.

The major seed proteins of legumes are globulins (e.g., glycinin and vicilin). Lysine content in globulin proteins varies with species, but all contain moderate to adequate supplies (Ma and Bliss, 1978). Methionine is limiting in seeds of almost all legumes. There are, however, several naturally occurring cultivars of legumes in which the levels of methionine in the seeds are higher than those found in their close relatives. 'Sanilac' subtypes of the vicilin (phaseolin) storage proteins in the common bean have been found that contain additional methionine residues and enhanced levels of phaseolin deposition. (Gepts and Bliss, 1984; Kami and Gepts, 1994). In this bean, therefore, there appears to be a case for increasing the proportion of methionine richer phaseolins in the storage proteins (Delaney and Bliss, 1991). Selection has been successful in increasing the percentage of seed protein in soybean (Brim and Buxton, 1979; Carter *et al.*, 1982) and the total protein content in wheat (Loeffler *et al.*, 1983).

In maize, two of the early mutants with altered amino acid composition, opaque-2 and floury-2, were found in the course of a search for mutants that had deficiencies in the seeds' ability to produce zein (Frey *et al.*, 1949; Mertz and Bressani, 1957). The proponents of this approach argued that, if zein does not form, the other seed storage proteins which contain a larger percentage of lysine will be formed in increased amounts (Nelson, 1969). An increased ratio of glutelin to zein has indeed been found in the opaque-2 mutant (Mertz *et al.*, 1964). Despite the increased nutritional value of opaque-2 varieties, the decrease in the zein fraction was not fully compensated by increases in other fractions and softer, smaller kernels resulted (Larkins, 1983). Recombinant methods, including antisense RNA technologies, are now replacing

traditional breeding methods in altering the levels of protein synthesis. Studies with expression of the various zeins in model systems, however, indicate that the synthesis and deposition of seed storage proteins is a complex system involving interaction between the various protein components (Wallace *et al.*, 1990) and simple adjustment of the levels of the various protein components of the seed may not be 'simple' to execute.

Expression of modified or heterologous seed storage proteins

One of the most rapidly expanding areas of plant genetic engineering is the expression of modified or heterologous seed storage proteins with enhanced nutritional value. Because of the potential commercial value of such transgenic seeds, much of the work in this field is still proprietary in nature. None-the-less, enough is now published in the scientific press or in the patent literature to give a good indication of the directions and the progress of this type of work. Generally speaking, transgenic plants being developed are of three major types: those in which the genes of normally occurring proteins have been modified to incorporate additional residues of limiting essential amino acids; those in which a gene for a naturally occurring protein from another species has been incorporated into the genome to provide the complementary essential amino acids; and those in which entirely new genes for *de novo* designed proteins with more desireable amino acid composition have been introduced. With each approach, some success has been achieved and some hurdles remain.

Since 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. Since amino acid sequence is also the major determinant of protein structure, however, substitutions and additions of amino acids can have considerable impact on a protein's ability to fold. The initial assumption that conservation of seed storage protein primary structure was unimportant because they lack catalytic function (Shotwell and Larkins, 1991) has given way to the idea that, 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) focussed on the common bean, *Phaseolus vulgaris*. The researchers inserted a naturally occurring 15 amino acid sequence from the 15 kD zein storage protein of maize (maize proteins are rich in methionine but deficient in lysine) into the bean β-phaseolin. Because phaseolins represent about 50% of the total bean protein, and because 6 of the 15 amino acids in the insertion sequence are methionine, this modification appeared to be one that would substantially increase the methionine level of seeds from transgenic bean plants and, hence, nutritional quality. Unfortunately, the modified protein was degraded despite the fact that the gene followed a normal developmental expression (Hoffman *et al.*, 1988). Based on their studies and other information from unrelated systems, the authors concluded that both tertiary and quaternary structures were important to intracellular transport and that the insertion site chosen caused changes in structure that interfered with the proteins's Golgi-mediated transport before deposition in protein bodies. Work on the expression of

phaseolins in *Xenopus* oocytes (Ceriolli *et al.*, 1991) confirmed that interruptions to the C-terminal α-helix domain of phaseolin monomers prevented their assembly into trimers and ensuing transport to the Golgi from the endoplasmic reticulum.

In light of this work and the available three-dimensional structure of phaseolins (Lawrence *et al.*, 1990), Dyer and coworkers (1993) have designed methionine-rich looping sequences for insertion into loops and turns in phaseolin three-dimensional structure and suggested methionine replacement of hydrophobic amino acids within the hydrophobic core of the β-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. Based on the diversity of sequences in these regions compared with relative conservation of sequence in other regions of the protein, the N-terminal portions of phaseolins are suggested as 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 phytohemagglutinin (PHA) to contain an additional three to four methionine residues (Kjemtrup *et al.*, 1994). In this case, the sequences of *P. vulgaris* PHA (a lectin which can account for 5%–10% of the total protein in bean seeds) was compared to homologous proteins from other species to determine where methionine codons could be inserted. Introduction of the resultant engineered gene in transgenic tobacco plant 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 used 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, 1981). In most dicots, 2S albumins are relatively lacking in methionine content. The Brazil nut (*Bertholletia exalsa*) 2S albumin, with its 18% methionine content, is a notable exception (Anpe *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 an homologous sequence from Brazil nut resulted in an increase of approximately 5% in the total methionine content in seeds expressing the transgene (De Clercq *et al.*, 1990).

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 (Kim *et al.*, 1990). This modified glycinin has also been successfully expressed in transgenic tobacco and the protein was shown to accumulate in the seed (Utsumi *et al.*, 1993a, b).

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 zein structure. Subsequent efforts to express lysine-containing α-zein genes in tobacco, however, proved unsuccessful. While the genes were expressed and mRNA levels appeared high enough to direct significant protein synthesis, the α-zein protein did not accumulate in seed. (Ohtani *et al.*, 1991). Because unmodified α-zein proteins did not accumulate either, the researchers postulate that the problem resides in the inability of α-zeins to form

protein bodies in the absence of β- and γ-zeins, and is not inherent to the lysine modification.

One obvious genetic solution to the need for dietary and feeding regimes where incomplete grain seed proteins are complemented with legume seed proteins is to engineer a seed which is internally complemented. 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. One popular target protein for expression in heterologous plant systems has been the Brazil nut 2S albumin seed storage protein, mainly because of its high methionine content. The properties of this protein and its processing and deposition during seed development in the Brazil nut have been well studied (Sun *et al.*, 1987) and the cognate gene cloned (Altenbach *et al.*, 1987; Gander *et al.*, 1991). In addition, nutritional studies in quail indicate that the protein is sufficient to promote growth and development in animals (Tao *et al.*, 1987).

This 2S albumin protein was first expressed in transgenic model plants (tobacco) in 1989 by Altenbach *et al.*, 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. Using the seed specific promoter from the soybean lectin gene and a coding sequence consisting of a fusion of the signal peptide-encoding sequence of the lectin and the codons for the processed 2S albumin protein, Guerche and co-workers (1990) achieved levels of heterologous protein ranging from 0.02% to 0.06% of total seed protein. De Clercq and co-workers (1990) used the promoter and signal sequence from one (AT2S1) of the two 2S albumin encoding genes in *Arabidopsis* to drive and target expression of the processed methionine-rich protein and achieved only marginally better expression in canola (the transgene product represented 0.1% of total seed protein). While 2S albumins represent some 40% of total protein in *Arabidopsis* seed, the AT2S1 gene, however, accounts for only about one fifth of the mRNA for these proteins. The much stronger AT2S2 gene promoter produces most of the mRNA for 2S albumin protein production. More recently, the third group (Altenbach *et al.*, 1992) has 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 using a chimeric gene in which a phaseolin promoter is used to drive the expression of the coding sequence for the unprocessed (pre-pro) 2S albumin protein. This research indicates that, in addition to 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. Recent advancement in gene transfer methods for legumes have expanded the range of plant targets for Brazil nut 2S albumin expression to include the common bean (Aragoa *et al.*, 1992) and three different grain legumes (Saalbach *et al.*, 1994).

Like Brazil nut, the sunflower seed expresses a 2S albumin protein that is extremely methionine-rich (Kortt *et al.*, 1991). In efforts to address the methionine requirement of pigs and chickens fed substantial amounts of the peas or lupin (as part of a grain/legume mixed ration regime) Australian researchers have begun to investigate the transformation of legume with chimeric genes for the sunflower 2S protein (Tabe *et*

al., 1993). 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 5 consecutive tryptophan residues have been successfully expressed in model plants. (Marcellino *et al.*, 1996).

Transgenic soybeans expressing the methionine-rich 2S albumin protein were recently examined for commercialization by an international plant breeding and seed supply company. A recent study (financially supported by that same company) indicates that the 2S albumin protein is 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). While the allergenicity of the sunflower 2S albumin protein has not been thoroughly tested, the homologous 2S albumin (Sina I) in yellow mustard has been reported to be a major allergen in mustard seeds (Menéndez *et al.*, 1988). This protein is closely related to 2S albumin found in canola, castor bean and Brazil nut. While food allergies are relatively rare, affecting 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.

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. Patatin, the major storage protein in potato tubers, is deficient in leucine, lysine and threonine. Neither patatin nor a glycinin on its own is deemed 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.*, 1993a)). Soybean β-conglycin has been expressed in petunia (Lawton *et al.*, 1987) and transgenic tobacco. Both native soybean glycinins and soybean glycinins which have been genetically engineered to contain four additional methionine residues can be expressed in potato tubers using the potato patatin promoter. 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 (Utsumi *et al.*, 1994).

Transgenic rice in which the lysine-rich β-phaseolin protein from bean is expressed to complement the lysine-deficient rice prolamines has recently 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 portion of the world population, but also because the heterologous protein was targeted to and expressed in only one of the two types of protein bodies that co-exist in the rice endosperm.

GENETIC ENGINEERING OF PLANT LEAF TISSUE

Most heterologous protein expression in plants has been targeted to the seeds. Seed storage proteins, because of their high nutritive value, also have potential use in improving forage crops if they could accumulate 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 to synthesize cell components (polysaccharides, protein, lipid and nucleic acids) needed for their own survival and growth (Hungate, 1964; Clark *et al.*, 1992; Merchen and Titgemeyer, 1992). While some foodstuff escapes digestion in the rumen, it is the products of microbial digestion (mainly volatile fatty acids) and the rumen microbes themselves that largely provide the nutrient source for the animal itself. The rumen fermentation process has both advantages and disadvantages. Rumen bacteria are able, for instance, to digest cellulose and plant fibre; they are as a group also able to synthesize all amino acids *de novo*, an ability that allows the production of 'high quality' complete proteins like milk and meat from 'low quality' incomplete proteins found in forage. In high producing ruminants like modern cattle and sheep, however, rumen microbial synthesis of protein is insufficient to meet the demands of production. In dairy cattle at peak lactation, for example, rumen bacterial protein production is not sufficient to meet the animal's need for the amino acids lysine, methionine, threonine and leucine (King *et al.*, 1990; Munneke *et al.*, 1991; Clark *et al.*, 1992). Protection or post rumen infusions of these limiting amino acids have been shown to increase the protein content of milk significantly (Donkin *et al.*, 1989; Le Henaff *et al.*, 1990; Rulquin *et al.*, 1993). In order to overcome these deficits, high producing ruminants are often fed additional protein supplements. Only a small portion of this additional protein in a ruminant's diet escapes microbial digestion to reach the ruminant (Beever and Siddons, 1986; Beever and Cotrill, 1994). While the amount of protein escaping ruminal digestion can increase with protein consumption (Slatter, 1986; Minson, 1990) feeding of digestible protein supplements is an inefficient way to increase amino acid availability. It is now common practice to supplement ruminant diets with proteins or amino acids that have been artificially protected against microbial digestion in the rumen (Ashes *et al.*, 1995).

Because individual proteins appear to have different susceptibility to rumen degradation (see, for example, Spencer *et al.*, 1988), Tabe and co-workers (1993) have attempted to identify proteins rich in sulfur-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. Ovalbumin in leaves of transgenic plants represented less than 0.01% of the total extractable protein (Wandelt *et al.*, 1991). The sunflower seed protein, on the other hand, was expressed at much higher levels in leaves, particularly when genetic engineering was used to introduce amino acid sequence changes that anchored the protein in the endoplasmic reticulum (Tabe *et al.*, 1993).

Seed storage proteins are generally targeted either to the endoplasmic reticulum (ER) or to vacuolar compartments via the ER. In seeds, vacuoles are largely storage compartments where protein bodies form and accumulate. In non-seed tissue, like leaf, vacuoles are protease-containing vesicles in which protein is generally degraded and does not accumulate (Van der Valk and van Loon, 1988). Seed storage proteins modified to contain the ER-anchoring peptide KDEL have been shown to be retained in the endoplasmic reticulum (Wandelt *et al.*, 1992; Tabe *et al.*, 1993). Some seed proteins without such targeting signals, however, have been shown to accumulate stably in leaves of transgenic plants. Both pea lectin expressed in potato plants

(Edwards *et al.*, 1991) and barley lectin in tobacco plants (Wilkins *et al.*, 1990) were found in significant amounts in vacuoles of the leaf cells of the respective transgenic plants. Attempts to express the vacuolar targeted β-phaseolin gene behind the cauliflower mosaic virus promoter in transgenic tobacco plants resulted in some new protein accumulation in the seed, but the β-phaseolin was presumably degraded in the leaf vacuoles and no new protein was detected in that tissue (Bagga *et al.*, 1995). The 15-kilodalton zein protein (which exists as an ER-derived protein body in maize seeds) did, on the other hand, stably accumulate in both leaves and seeds of transgenic tobacco expressing the gene (Bagga *et al.*, 1995). Whether the protein accumulation in leaves results from ER retention or the inherent ability of the protein to aggregate and resist digestion is not evident from the results reported to date. Clearly, while seed storage protein expression in heterologous hosts is a promising avenue toward the enrichment of both seed and leaf proteins, much remains to be understood at a more fundamental level.

Genetic engineering for nutritional quality in bacteria

Bacterial cell lines are a popular target for genetic engineering to produce new proteins, particularly those that do not require post-translational modifications. Their frequent use is partly because of the relative simplicity of the system and partly because, in the case of the more common 'workhorse' bacterial species, expression vectors and transformation systems are quite well developed. In these well-studied species, effective promoters, enhancers, terminator signals, etc are readily available and the temporally controlled, targeted expression of proteins is rather routine. Much of the early work using genetic engineering to increase the nutritional value of foodstuffs focussed on the creation of nutritional products synthesized in, and purified from, bacterial cells. These products would then be added back to a food or feed preparation. Single cell protein production using recombinant micro-organisms to convert, for example, methane into protein have been explored by a variety of researchers as a means of converting inexpensive or waste sources of carbon and nitrogen into useful proteins. (For a review of early work in this area, see Spinks, 1982). Work to make single cell, 'industrial food' proteins for human use was well underway in the early 1980's (Rehm, 1982). In most of these early examples, bacterial cells produced the usual complement of proteins and the engineering was at the level of modifications which allowed the bacteria to use hitherto 'unusable' sources of component molecules. The resultant products were envisioned as being additives rather than whole foods themselves.

Expression of new, de novo design proteins

Bacterial cells have also been explored as sources of specific amino acids. These amino acids could be used to supplement diets that are defficient in only certain amino acids. Such feeding practices are common with livestock. Until recently, the amino acid were usually produced by protein hydrolysis and purification rather than by recombinant bacterial fermentation. With advances in bacterial genetic engineering, the situation is quickly changing. The key to successful fermentation for amino acid production is the ability to overcome the end-product inhibition of the enzymes in the

biosynthetic pathway for amino acid production. While it is possible to engineer enzymes to be insensitive to such inhibition, it is simpler, conceptually at least, to remove the end-product causing the inhibition by using it in protein systhesis. A cloned synthetic gene containing several copies of a single codon will, in theory, result in the synthesis of a homopolymer 'sink' for the encoded amino acid. Kangas and co-workers (1982) were successsful in using this concept to produce a proline-rich protein in *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 artifical 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 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 appear to be with inherent instability of the protein product rather than with any inefficiencies in translation of the mRNA into protein.

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.*, 1986). Biernat and Koster (1987) suggested that it is the N-terminal amino acid sequence rather than the identity of only the first amino acid in that sequence that has significance for the proper folding of the polypeptide chain. In our laboratory, we attempted to express synthetic genes encoding repeating pentapepetides of nutritonal importance fused to the N-terminal sequence (or α-peptide) of β-galactosidase. The α-peptide is the same amino acid sequence that provides the N-terminal folding signals for the stable accumulation of β-galactosidase in the native enzyme and thus should fulfill the requirements of both variations of the 'N-terminal Rule'. We were not, however, able 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 acculatin of the gene product. 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.

Given the knowledge that structural stability was the key to ensuring protein stability, we then set out to design a new protein with high levels of essential amino acids that would fold into a compact, protease-resistant structure. *De novo* design of proteins with specific tertiary structures or folds has been undertaken by several other research groups in order to explore in more detail the rules which govern protein folding in general. Proteins have been designed to form some of the simpler tertiary structure motifs – the four-α-helix-bundle (e.g., DeGrado, 1988; Hecht *et al.*, 1990), coiled-coil proteins (Monera *et al.*, 1996), an αβ- barrel protein (Goraj *et al.*, 1990; Beauregard *et al.*, 1991; Houbrechts *et al.*, 1995) and all β proteins (Yan and Erickson, 1994; Quinn *et al.*, 1994). In each of these designs, however, the goal was to

choose from the twenty common amino acids those that would best promote the desired fold. The composition of the pool of amino acids that will result on hydrolysis or digestion of the protein was of no concern. In the design of proteins for nutritional purposes, the situation is different. The overall amino acid composition of the protein is of paramount importance; the actual linear sequence of amino acids is inconsequential as long as the protein can adopt a stable, compact, folded structure. Thus, in the design of our first nutritional protein, the milk bundle protein or MB-1, we used a novel design approach. Instead of choosing amino acids which best promote a given fold, we searched the known, simple tertiary folds in naturally occurring proteins for one that could support the incorporation of the largest percentages of the desired amino acids using a 'consensus residue' approach (Beauregard *et al.*, 1995).

Our particular interest is in meeting the nutritional requirements of high-production, lactating dairy cows. Several studies have shown that mixed ration diets routinely fed to dairy cows in lactation are severely limited in the essential amino acids methionine and lysine and secondarily limited in leucine and threonine (King *et al.*, 1990; Clark *et al.*, 1992; Rulquin *et al.*, 1993). There is also ample evidence to show that when these amino acids are supplied post-ruminally, levels of protein in the milk increase significantly (Donkin *et al.*, 1989; Rulquin *et al.*, 1993). The question for us, then, was what known tertiary fold will best accommodate large amounts of lysine, methionine, leucine and threonine. The results of our consensus residue analysis indicated that the four-α-helix-bundle motif was a good candidate fold. We then designed MB-1, a bundle protein in which 67 of the 100 amino acid residues are lysine, methionine, leucine or threonine. The protein could be expressed from a synthetic gene introduced into recombinant *E. coli* at high levels (7% to 12% of the total cellular protein) and was shown to be largely helical as designed (Beauregard *et al.*, 1995; Hefford *et al.*, in preparation.)

Because our target was a nutritional protein for ruminants, a bacterial expression system seemed obvious. As noted, most of the protein available to a ruminant comes from the microbial protein produced by rumen fermentation rather than directly from the feed. Protein expressed in plants or as single cell protein products that are fed to ruminants need to be protected from rumen degradation. MB-1 was designed to be expressed in the cytosol of a recombinant rumen bacterium where it would be protected from the digestive enzymes present in the rumen fluid. The intracellular location of MB-1 necessitated refinement of the protein design to eliminate any known sites for bacterial cytosolic proteolytic enzymes (Gottesman, 1989). In addition, the gene for MB-1 must be configured to be expressed at high levels in strains of microorganisms that are abundant in the rumen. Genetic engineering methods for rumen bacteria, while much less developed than those for such bacteria as *E. coli* or *B. subtilis*, are now emerging (see Teather *et al.*, 1996, for a recent review) and we are currently refining the expression vector in order to optimize MB-1 production in recombinant strains of the rumen bacterium *B. fibrisolvens*.

The successful *de novo* design of proteins with specific tertiary structures indicates that incorporation of at least some of the essential amino acids into novel proteins is quite feasible. Leucine, for example, is the amino acid most commonly used by nature in the hydrophobic 'a' and 'd' positions in the heptad repeats that form the helices in both four-α-helix-bundle proteins and in coiled-coil proteins (Paliakasis and Kokkinidis, 1992; West and Hecht, 1995). Lysine (in conjunction with glutamic acid

to neutralize charge effects) is often found in the hydrophobic positions of both α-helices and β-sheets in proteins of all shapes and sizes (Regan and DeGrado, 1988; Hecht *et al.*, 1990; Kamtekar *et al.*, 1993; Houbrechts *et al.*, 1995). The consensus residue approach used in MB-1 design, however, was most useful in that it identified methionine as a potiential hydrophobe in bundle proteins. Methionine, while compatible with the physical and chemical properties required for the 'a' and 'd' positions of the heptad repeats in helices, is not often found in naturally-occurring bundles (Paliakasis and Kokkinidis, 1992) nor has it often been incorporated in designed proteins. Others have found that methionine can, in fact, destabilize α-helices when inserted in some structural environments (Dyer *et al.*, 1993; Gassner *et al.*, 1996). The consensus residue approach to helix design permits one to choose 'allowed' structural environments for methionine insertion. We have recently expanded our consensus residue approach to look at the loops between the helices in four-α-helix-bundle proteins as well as the helices themselves (Parker and Hefford, 1997). The results of this study indicate that the essential amino acid, tryptophan, among others, can be accommodated in some positions within the loop regions of these proteins.

The principles used in the design of MB-1 are not limited to expression of nutritional proteins in bacteria and are currently being employed for the design of new proteins, enriched in essential amino acid, in plant seeds and leaves. As in the production of other nutritionally important proteins in plant seed, *de novo* protein expression will depend not only on the proper folding of the new protein itself but also on the ease with which it can be incorporated into protein bodies. This is, at present, an area of very active research. As a result of the economic import of the work, however, most of this research is proprietary.

General conclusions

The nutritional quality of protein sources in both animal and human diets has enormous social and economic impacts. Nutrition influences both the health and productivity of our livestock and human vitality and mortality. The ability to supply adequate amounts of essential amino acids in diets while reducing the economic and environmental costs of doing so is becoming increasingly important as the world population expands. Genetic engineering to produce recombinant bacteria and plants with enhanced levels of either free or protein-bound essential amino acids provides a very feasible alternative to traditional methods of enhancing the essential amino acid supply in diets. While there are still several technical challenges to be met, the research available at this date indicates that novel, nutritionally enhanced food products from genetic engineering will become increasingly common in the not too distant future.

Acknowledgements

The helpful suggestions of Dr Shea Miller and Dr Ron Teather (Agriculture and Agri-food Canada) on the preparation of this manuscript and the many discussions with Dr Marc Beauregard (Department of Chemisty, University of Prince Edward Island) on protein design are gratefully acknowledged.

References

ALTENBACH, S.B., PEARSON, K., LEUNG, F.W. AND SUN, S.S.M. (1987). Cloning and sequence analysis of a cDNA encoding a Brazil nut protein exceptionally rich in methionine. *Plant Molecular Biology* **8**, 239–250.

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 Molecular Biology* **18**, 235–245.

ALTENBACH, S.B., PEARSON, K.W., MEEKER, G., STARCI, L.C. AND SUN, S.S.M. (1989). Enhancement of the methionine content of seed proteins by the expression of a chimeric gene encoding a methionine-rich protein in transgenic plants. *Plant Molecular Biology* **13**, 513–522.

ANPE, C., VAN DAMME, J., DECASTRO, L., SAMPAIO, M.J., VAN MONTAGU, M. AND VANDERKEREKHOVE, J. (1986). The amino acid sequence of the 2S sulfur-rich proteins from seeds of Brazil nut (*Bertholletia exalsa* H.BK) *European Journal of Biochemistry* **159**, 597–604.

ARAGOA, F.J.L., GROSSI DE SA, M.F. AND ALMEIDA, E.R. (1992). Particle bombardment-mediated transient expression of Brazil nut methionine-rich albumin in bean (*Phaseolus vulgaris* L.) *Plant Molecular Biology* **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*, (M. Ivan, ed.) Agriculture and Agri-Food Canada, Ottawa, pp. 177–188.

BACHMAIR, J., HASSELMANN, H., HOFER, B., KENNEDY, N. AND KOSTER, 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 SENGUPTA-GOPALAN, C. (1995). Accumulation of 15-kilodalton zein in novel protein bodies in transgenic tobacco. *Plant Physiology* **107**, 13–23.

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. *Bio/Technology* **13**, 974–981.

BEAUREGARD, M., GORAJ, K., GOFFIN, V., HEREMANS, K., GOORMAGHTIGH E., RUYSSCHAERT J. M. AND MARTIAL, J. A. (1991). Spectroscopic investigation of structure in octarellin (a *de novo* protein designed to adopt the alpha/beta-barrel packing). *Protein Engineering* **4**, 745–749.

BEAUREGARD, M., HEFFORD, M. A. AND TEATHER, R. M. (1994). Use of a monoclonal antibody to beta-galactosidae for the detection of alpha peptide fusions. *BioTechniques* **16**, 832–839.

BEEVER, D.C. AND COTRILL, B.R. (1994). Symposium: Protein Systems for Feeding Ruminant Livestock: a European assessment. *Journal of Dairy Science* **77**, 2031–2043.

BEEVER, D. AND SIDDONS, R.C. (1986). Digestion and metabolism in the grazing ruminant. In: *Control of Digestion and Metabolism in Ruminants*. (L.P. Milligan, W.L. Grovum and A. Dobson, Eds). Prentice Hall, New Jersey, pp. 479–496.

BITTLE, D.C., KEITH, R., SHAVER, J., SELLNER, D.A. AND GENGENBACH, B.G. (1992). Characterization of a lysine-insensitive form of DHD PS from maize. In: *Biosynthesis and Molecular Regulation of Amino Acids in Plants*. (B.K. Singh, H.E. Flores and J.C. Shanon, eds) American Society of Plant Physiologists, Rockville, MD, pp. 322–323.

BRIGHT, S.W.J., FEATHERSTONE, L.C. AND MIFFLIN, B.J. (1979). Lysine metabolism in a barley mutant resistant to s-(z-aminoethyl)-L-cysteine. *Planta* **146**, 629–633.

BRIM, C.A. AND BUXTON, J.W. (1979) Recurrent selection in soybean. II. Selection for increased protein in seeds. *Crop Science* **19**, 494–498.

BROCHETTO-BRAGA, M.R., LEITE, A. AND ARRUDA, P. (1992). Partial purification and characterization of lysine-ketoglutamate reductase in normal and opaque-2 maize endosperm. *Plant Physiology* **98**, 1139–1147.

CARTER, T.E., BURTON, J.W. AND BRIM, C.A. (1982). Recurrent selection for percent protein

in soybean seed – indirect effects on plant N accumulation and distribution. *Crop Science* **22,** 513–519.

CERIOLLI 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. *European Journal of Biochemistry* **202,** 959–968.

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. *Journal of Dairy Science* **75,** 2304–2323.

DE CLERCQ, A., VANDEWIELE, M., VAN DAMME, J., GUERCHE, P., VAN MONTAGU, M., VANDERKERCHOVE, J. AND KEBBERS, E. (1990). Stable accumulation of modified 2S albumin seed storage proteins with higher methionine contents in transgenic plants. *Plant Physiology* **94,** 970–979.

DEGRADO, W. F. (1988). Design of peptides and proteins. *Advances in Protein Chemistry* **39,** 51–124.

DELANEY, D.E. AND BLISS, F.A. (1991). Selection for increased percentage phaseolin in common bean. *Theoretical and Applied Genetics* **81,** 301–305.

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(l-asparyl-L-phenylalanine). *Nucleic Acids Research* **8,** 4575–4592.

DONKIN, S. S., VARGA, G. A., SWEENEY, T.F. AND MULLER, L. D. (1989). Rumen-protected methionine and lysine: effects on animal performance, milk protein yield, and physiological measures. *Journal of Dairy Science* **70,** 977–982.

DYER, J. M., NELSON, J. W. AND MURAI, N. (1993). Strategies for selecting mutation sites for methionine enhancement in the bean seed stroage protein phaseolin. *Journal of Protein Chemistry* **12,** 545–560.

EDWARDS, G.A., HEPHER, A., CLERK, S.P. AND BOULTER, D. (1991). Pea lectin is correctly processed, stable and active in the leaves of transgenic potato plants. *Plant Molecular Biology* **17,** 89–100

FAO/WHO/UNU (1985). Energy and protein requirements: Report on a joint FAO/WHO/UNU Expert Consultation. WHO Tech Rep. Ser. No. 724. WHO Geneva.

FRANKARD, V., GHISLAIN, M., NEGRUTIN, I. AND JACOBS, M. (1991). High threonine producer mutant in *Nicotiana sylvestris* (Spegg and Comes). *Theoretical and Applied Genetics* **82,** 273–282.

FRANKARD, V., GHISLAIN, M. AND JACOBS, M. (1992). Two feedback-insensitive enzymes of the aspartate pathway in *Nicotiana sylvestris*. *Plant Physiology* **99,** 1285–1293.

FREY, K.J., BRIMHALL, B. AND SPRAGUE, G. F. (1949). Effects of selection upon protein quality in the corn kernel. *Agronomy Journal* **41,** 399–403.

FRIEDMAN, M. AND GUMBMANN, M. R. (1986). Nutritional improvement of legume proteins through disulfide interchange. *Advances in Experimental and Medical Biology* **199,** 357–389.

GANDER, E. S., HOLMSTROEM, K. -O. DE PAIVA, G. R., DE CASTRO, L. A. B., CARNEIRO, M. AND GROSSI DE SA, M.-F. (1991). Isolation, characterization and expression of a gene coding for a 2S albumin from *Berholletia excelso* (Brazil nut). *Plant Molecular Biology* **16,** 437–448.

GASSNER, N. C., BAASE, W. A. AND MATTHEWS, B. W. (1996). Methionine cores of T4 lysozyme which preserve protein activity, folding cooperativity and shape. Abstract in 'Protein Folding and Design: an International Conference', NIH, Bethesda, Md, USA.

GEPTS, R. AND BLISS, F.A. (1984). Enhanced availability methionine concentration associated with higher phaseolin levels in common bean seeds. *Theoretical and Applied Genetics* **69,** 47–53.

GHISLAIN, M., FRANKARD, V. AND JACOBS, M. (1995). A dinucleotide mutation in DHDPS of *Nicotiana sylvestris* leads to lysine overproduction. *Plant Journal* **8,** 733–743.

GLOVER, D. V. AND MERTZ, E. T. (1987). Corn. In: *Nutritional quality of cereal grains: genetic and agronomic improvement*. (R.A. Olsen and K.J. Frey, eds.) ASA – CSSA – SSA, Madison, pp. 183–336.

GORAJ, K., RENARD, A. AND MARTIAL, J. A. (1990). Synthesis, purification and initial structural characterization of octarelline, a *de novo* polypeptide modelled on the alpha/beta-barrel proteins. *Protein Engineering* **3**, 259–266.

GOTTESMAN, S. (1989). Genetics of proteolysis in *E. coli*. *Annual Review of Genetics* **23**, 163–198.

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*. *Molecular and General Genetics* **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 repilicons in *E. coli*. *Bio/Technology* **1**, 602–609.

HABBEN, J.E. AND LARKINS, B.A. (1995). Genetic modification of seed proteins. *Current Opinions in Biotechnology* **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 protein of native-like structure. *Science* **249**, 884–891.

HOFFMAN, L. M., DOLALDSON, D. D. AND HERMAN, E. M. (1988). A modified storage protein is synthesized, processed and degraded in the seeds of transgenic plants. *Plant Molecular Biology* **11**, 717–729.

HOUBRECHTS, A., MOREAU, B., ABAGYAN, R., MAINFROID, B., PREAUX, G., LAMPROYE, A., PONCIN, A., GOORMAGHTIGH, E., RUYSSCHAERT, J.-M., MARTUAKM J. A. AND GORAJ, K. (1995). Second-generation octarellins: two new *de novo* $(\beta/\alpha)_8$ polypetides designed for investigating the influence of β-residue packing on the α/β-barrel structure stability. *Protein Engineering* **8**, 249–259.

HUNGATE, R. E. (1966). *The rumen and its microbes*. Academic Press Inc., New York.

KAMI, J.A. AND GEPTS, P. (1994). Phaseolin nucleotide sequence diversity in Phaseolins. I. Intraspecific diversity in *Phaseolus vulgaris*. *Genome* **37**, 751–757.

KAMTEKAR, S., SCHIFFER, J. M., XIONG, H., BABIK J. M. AND HECHT, M. H. (1993). Protein design by binary patterning of polar and nonpolar amino acids. *Science* **262**, 1680–1685.

KANEKO, T., HASHINOTO, T., KUMPAISAL, R. AND YAMADA, Y. (1990). Molecular cloning of wheat DHDPS. *Journal of Biological Chemistry* **265**, 17451–17455.

KANGAS, T. T., COONEY, C. L. AND GOMEZ, R. F. (1982). Expression of a proline-enriched protein in *E. coli*. *Applied Environmental Microbiology* **43**, 629–635.

KIM, C-S., KAMIYA, S., KANAMORI, J., UTSUMI, S. AND KITO, M. (1990). High level expression, purification and functional properties of soybean proglycinin from *E. coli*. *Agricultural Biological Chemistry* **54**: 1543–1550.

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. *Journal of Dairy Science* **73**: 3208–3216.

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. *European Journal of Biochemistry* **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 annus* L.) *FEBS Letters* **195**, 329–334.

LARKINS, B.A. (1983). Genetic engineering of seed storage proteins. In: *Genetic engineering of plants: An agricultural perspective*. (T. Kosuge and C.P. Meredith, eds). Plenum Press, New York, pp. 93–120.

LAWRENCE, M.C., SUZUKI, E., VARGHESE, J.N., DAVIS, P.C., VAN DONTELAAR, A., TULLOCH, P.A. AND COLMAN, P.N. (1990). The three-dimensional structure of the seed storage protein phaseolin at 3 A° resolution. *EMBO Journal* **9**, 9–15.

LAWTON, M.A., TIERNEY, M.A., NAKAMURA, I., ANDERSON, E., KOMEDA, Y., DUBÉ, P., HOFFMAN, N., FRALEY, R.T. AND BEACHY, R.N. (1987). Expression of a soybean β-conglycinin gene under the control of the cauliflower mosaic virus 35S and 19S promoters in transformed petunia tissues. *Plant Molecular Biology* **9**, 315–324.

LE HENAFF, L., RULQUIN, H. AND VERITE, R. (1990). Lactational responses to postruminal

infusions of amino acids in dairy cows fed with maize silage, hay or grass silage. *Reproduction and Nutrition Development* (Suppl. 2): 237S.

LOEFFLER, C.M., BUSCH, R.H. AND WIERSMA, J.V. (1983). Recurrent selection for grain protein percentage in hard red spring wheat. *Crop Science* **23,** 1097–1101.

MA, Y. AND BLISS, F.A. (1978). Seed proteins of common bean. *Crop Science* **17,** 431–437.

MARCELLINO, L.H., NESBICH, G., GROSSIDESA, M.F., KREBBES, E., AND LANDER, E.S. (1996). Modified 2S albumins with improved tryptophan content are correctly expressed in transgenic tobacco plants. *FEBS Letters* **385,** 154–158

MENÉNDEZ, L., MONEA, I., DOMINGURZ, J. AND RODRIGUEZ, R. (1988). Primary structure of the major allergen of yellow mustard (*Sinapis alba L.*) seed, Sin a I. *European Journal of Biochemistry* **177,** 159–166.

MERCHEN, N. R. AND TITGEMEYER, E. C. (1992). Manipulation of amino acid supply to the growing ruminant. *Journal of Animal Science* **70,** 3238–3247.

MERTZ, E.T. AND BRESSANI, R. (1957). Studies on corn proteins: I. A new method of extraction. *Cereal Chemistry* **34,** 63–69.

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.

MINSON, D.J. (1990). Forage. In: *Ruminant nutrition.* Academic Press. San Diego, CA, USA.

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. *Journal of Biological Chemistry* **271,** 3995–4001.

MUNNEKE, R. L., SCHINGOETHE, D. J. AND CASPER, D. P. (1991). Lactational evaluation of ruminally protected methionine in diets containing extruded soybeans and urea. *Journal of Dairy Science* **74,** 227–233.

NEGRUTIN, 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). *Theoretical and Applied Genetics* **68,** 11–20.

NELSON, O.E. (1969). Genetic Modification of Protein Quality. *Advances in Argronomy* **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. *Journal of Biological Chemistry* **267,** 15938–15942.

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. *New England Journal of Medicine* **394,** 688–692.

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 Molecular Biology* **16,** 117–128.

OSBORNE, T. B. (1924). *The vegetable proteins*. Longmans Green, New York, NY, USA.

PALIAKASIS, C. D. AND KOKKINIDIS, M. (1992). Relationships between sequence and structure for the four-alpha-helix bundle tertiary motif in proteins. *Protein Engineering* **5,** 739–748.

PARKER, M. AND HEFFORD, M.A. (1997). A consensus residue analysis of loop and helix capping residues in four-alpha-helical bundle proteins. *Protein Engineering, in the press.*

PERL, A., GALILI, S., SHAUL, O. AND GALILI, G. (1993). Bacterial DHDPS and desensitized aspartate kinase: two novel selectable markers for plant transformation. *Bio/Technology* **11,** 715–718.

PUIGSERVER, A. J., SEN, L. C., CLIFFORD, A. J., FEENEY, R. E. AND WHITAKER, J. R. (1986). A method for improving the nutritional value of food proteins: covalent attachment of amino acids. *Advances in Experimental Medicine and Biology* **105,** 587–612.

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. *Proceedings of the National Academy of Sciences USA* **91,** 8747–8751.

REGAN, L. AND DEGRADO, W. F. (1988). Characterization of a helical protein designed from first principles. *Science* **241,** 976–978.

REHM, H.-J. (1982). Biotechnology research in Europe. *Journal of Chemistry and Technological, Biotechnology*. **32**, 1.

RULQUIN, H., PISULEWSKI, P. M., VERITE, R. AND GUINARD, J. (1993). Milk production and composition as a function of postruminal lysine and methionine supply: a nutrient response approach. *Livestock Production and Science* **37**, 69–90.

SAALBACH, I., PICHARDT, T., MACHEMEHL, F., SAALBECH, G., SCHIEDER, O. AND MÜNTZ, 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. *Molecular and General Genetics* **242**, 226–236.

SHOTWELL, M.A. AND LARKINS, B.A. (1991). Improvement in the protein quality of seeds by genetic engineering. In: *Molecular Approaches to Crop Improvement*. (E.S. Dennis and D.J. Llewellyn, eds) Springer Verlag, Wein, NY. pp 33–61.

SHUR Y-Y.J., HEPBURN, A.G. AND WIDHOLM, J.M. (1992). Glyphosate selected amplification of the s-enol pyruvylshikimate-3-phosphate synthase gene in cultured carrot cells. *Molecular and General Genetics* **232**, 377–382.

SILK, G.W. AND MATTHEWS, B.F. (1992). Molecular cloning of the soybean encoding DHDPS. In: *Biosyntheses and molecular genetics of amino acids in plants*. (B.K. Singh, H.E. Floves and J.C. Shannon, eds). American Society of Plant Physiologists, Rockville, MD, pp. 341–342.

SLATTER, L.D. (1986). Protein supply from undegraded proteins. Symposium: Protein and fibre digestion passage and utilization in dairy cows. *Journal of Dairy Science* **69**, 2734–2749.

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. *Biochimica Biophysica Acta* **1292**, 15–22.

SPENCER, D., HIGGINS, N.V., FREER, M., DOVE, H. AND COOMBE, J.B. (1988). Monitoring the fate of dietary proteins in rumen fluid using gel electrophoresis. *British Journal of Nutrition* **60**, 241–247.

SPINKS, A. (1982). Targets in biotechnology. *Proceedings of the Royal Society of London* **214**, 289–303.

SUN, S.S.M., ALTENBACH, S.B. AND LEUNG, F.W. (1987). Properties, biosynthesis and processing of a sulfur-rich protein in Brazil nut (*Bertholletia excelsa* H.B.K.). *European Journal of Biochemistry* **159**, 597–604.

TABE, L.M., HIGGINS, C.M., MCNABB, W.C. AND HIGGINS, T.S.V. (1993). Genetic engineering of grain and pasture legumes for improved nutritive value. *Genetica* **90**, 181–200.

TAO, S.-H., FOX, M.R.S., FRY, B.E., JOHN, M.L., LEE, Y.H., TOMIS, J.C. AND SUN, S.M. (1987). Methionine bioavailability of a sulfur-rich protein from Brazil nut. *Federation Proceedings* **46**, 891.

TEATHER, R. M., HEFFORD, M. A. AND FORSTER, R. J. (1996). The genetics of rumen bacteria. In: *The rumen microbial ecosystem*. Second edition. (P. Hobson and C. Stewart, eds.) Andover, England.

UTSUMI S., KITAGAWA, S., KATSUBE, T., KANG, I. J., GIDAMIS, A. B., TAKAIWA, F. AND KITO, M. (1993a). Synthesis, processing and accumulation of modified glycinins of soybean in the seeds, leaves and stems of transgenic tobacco. *Plant Science* **92**, 191–202.

UTSUMI, S., KITAGAWA, S., KATSUBE, T., GIDAMIS, A.B., TAKAIWA, F. AND KITO, M. (1993b). Synthesis, processing and accumulation of modified glycinins of soybean in the seeds, leaves and stems of transgenic tobacco. *Plant Science* **92**, A1–202.

UTSUMI, S., KITAGAWA, S., KATSUBE, T., HIGASA, T., KITO, M., TAKAIWA, F. AND ISHIGE, T. (1994). Expression and accumulation of normal and modified soybean glycinins in potato tubers. *Plant Science* **102**, 181–188.

VAN DER VALK, H.C.P.M. AND VAN LOON, L.C. (1988). Subcellular localization of proteases in developing leaves of oats (*Avena Sativa* L.) *Plant Physiology* **87**, 536–541.

VAUTERIN, M. AND JACOBS, M. (1994). Isolation of a poplar and an *Arabidopsis thaliana* DHDPS cDNA clone. *Plant Molecular Biology* **25**, 545–550.

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* **240,** 662–664.

WALLACE, J.C., OHTANI, T., LENDING, C.R., LOPES, M., WILLIAMSON, J.D., SHAW, K.L., GELVIN, S.B. AND LARKINS, B.A. (1990). Factors affecting physical and structural properties of maize protein bodies. In:*Plant gene transfer.* (C.Lamb and R.N. Beachy, eds) Alan R. Liss Inc., NY, pp 205–216.

WANDELT, C., KNIBB, H.E., SCHROEDER, H.E., KHAN M.R.I., SPENCER, D., CRAIG, S. AND HIGGINS, T.J.V. (1991). The expression of an ovalbumin and a seed storage protein gene in the leaves of transgenic plants. In: *Plant molecular biology.* (R.G. Herman and B. Larkins, eds) Plenum Press, NY, pp 471–478.

WANDELT, C.E., KHAN, M.R.I., CRAIG, S., SCHROEDER, H.E., SPENCER, D. AND HIGGINS, T.J.V. (1992). Vicillin with carboxy-terminal KDEL is retained in the endoplasmic reticulum and accumulates to high levels in the leaves of transgenic plants. *Plant Journal* **2,** 181–192.

WEST, M. W. AND HECHT, M. H. (1995). Binary patterning of polar and nonpolar amino acids in the sequences and structures of native proteins. *Protein Science* **4,** 2032–2039.

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.

YAN, Y. AND ERICKSON, B. W. (1994). Engineering of betabellin 14D: Disulfide-induced folding of a Beta-sheet protein. *Protein Science* **3,** 1069–1073.

YOULE, R. AND HUANG, A. H. C.(1981). Occurrence of low molecular weight and high cysteine containing albumin storage protein in oilseeds of diverse species. *American Journal of Botany* **68,** 44–48.

ZHENG, Z., SUMI, K., TANAKA, K. AND MURAI, N. (1995). The bean seed storage protein β-phaseolin is synthesized, processed, and accumulated in the vacuolar type II protein bodies of transgenic rice endosperm. *Plant Physiology* **109,** 777–786.