DE NOVO DESIGNED SYNTHETIC PLANT STORAGE PROTEINS: ENHANCING PROTEIN QUALITY OF PLANTS FOR IMPROVED HUMAN AND ANIMAL NUTRITION

JESSE MICHAEL JAYNES, Ph.D.

Vice President for Development
Demeter Biotechnologies, Ltd., Research Triangle Park, North Carolina USA

Background

The composition of storage proteins, a major food reservoir for the developing seed, determines the nutritional value of plants and grains when they are used as foods for man and domestic animals. The amount of protein varies with genotype or cultivar, but in general, cereals contain 10% of the dry weight of the seed as protein, while in legumes, the protein content varies between 20% and 30% of the dry weight. In many seeds, the storage proteins account for 50% or more of the total protein and thus determine the protein quality of seeds. Each year the total world cereal harvest amounts to some 1,700 million tons of grain (Keris et al., 1985). This yields about 85 million tons of cereal storage proteins harvested each year and contributes a majority of the total protein intake of humans and animals.

With respect to human and animal nutrition, most seeds do not provide a balanced source of protein because of deficiencies in one or more of the essential amino acids in the storage proteins. For example, humans require from foods eight amino acids: isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan and valine, to maintain a balanced diet. Consumption of proteins of unbalanced amino acid composition can lead to a malnourished state which is most often found in children in developing countries where plants are the major source of protein intake. Therefore, the development of nutritionally-balanced proteins for introduction into plants is of extreme importance.

Recently, many laboratories have attempted to improve the nutritional quality of plant storage proteins by transferring heterologous storage protein genes from other plants (Pederson et al., 1986). The development of recombinant DNA technology and the Agrobacterium-based vector system has made this approach possible. However, genes encoding storage proteins containing a more favorable amino acid balance do not exist in the genomes of major crop plants. Furthermore, modification of native storage proteins has met with difficulty because of their instability, low level of expression, and limited host range. One possible alternative would be the de novo design of a more nutritionally-balanced protein

Jaynes JM. (1994) De Novo Designed Synthetic Plant Storage Proteins: Enhancing Protein Quality of Plants for Improved Human and Animal Nutrition. In and the Feed Industry, 129-154. *Proceedings of Alltech's Tenth Annual Symposium.*

which retains certain characteristics of the natural storage proteins of plants.

Our initial work described the use of small fragments of DNA which encoded spans of protein high in essential amino acids (Jaynes et al., 1985; Yang et al., 1989). Subsequently, the genes encoding these protein domains were cloned into an existing protein and the expression level of this modified protein determined in transgenic potato plants. However, because of some of the problems mentioned above, the results were somewhat less than desirable (Yang et al., 1989).

There are at least two fundamental difficulties in achieving efficient expression of designed proteins. First, it is not yet known what stabilizes a protein against proteolytic breakdown; and second, the mechanisms for folding of an amino acid sequence into a biologically-stable tertiary structure have not yet been fully delineated. For the construction of DNP 1 (Designed Nutritional Protein), we focused on the design of a physiologically-stable as well as a highly nutritious, storage-protein-like, artificial protein. Based on what we have learned in the design and expression of lytic peptides for enhanced disease resistance in plants (Jaynes et al., 1992), it now seemed possible for us to design entirely artificial, stable nutritionally-significant proteins to improve the nutritive quality of plants.

Amino acid requirements

The biosynthesis of amino acids from simpler precursors is a process vital to all forms of life as these amino acids are the building blocks of proteins. Organisms differ markedly with respect to their ability to synthesize amino acids. In fact, virtually all members of the animal kingdom are incapable of manufacturing some amino acids. There are twenty common amino acids which are utilized in the fabrication of proteins; and essential amino acids are those protein building blocks which cannot be synthesized by the animal. It is generally agreed that humans require eight of the twenty common amino acids in their diet. Protein deficiencies can usually be ascribed to a diet which is deficient in one or more of the essential amino acids. A nutritionally adequate diet must include a minimum daily consumption of these amino acids (Figure 1).

When diets are high in carbohydrates and low in protein over a protracted period, essential amino acid deficiencies result. The name given to this undernourished condition in humans is "Kwashiorcor" which is an African word meaning "deposed child" (deposed from the mother's breast by a newborn sibling). This debilitating and malnourished state, characterized by a bloated stomach and reddish-orange discolored hair, is more often found in children than adults because of their great need for essential amino acids during growth and development. In order for normal physical and mental maturation to occur, the above mentioned daily source of essential amino acids is a requisite. Essential amino acid content, or protein quality, is as important a feature of the diet as total protein quantity or total calorie intake.

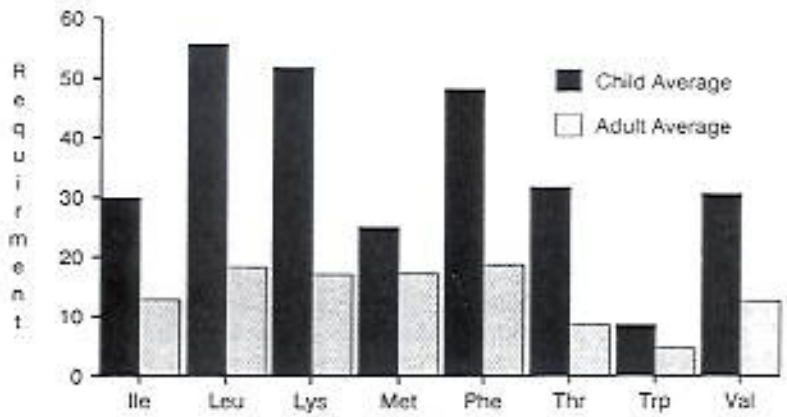

Figure 1 The average essential amino acid requirement for both children and adults in mg. per kg. body weight. Note that children, on a per weight basis, require more of the essential amino acids than do adults. this indicates the importance of diet to normal physical and mental maturation.

Some foods, such as milk, eggs, and meat, have very high nutritional values because they contain a disproportionately high level of essential amino acids. On the other hand, most foodstuffs obtained from plants possess a poor nutritional value because of their relatively low content of some or, in a few cases, all of the essential amino acids. Generally, the essential amino acids which are found to be most limiting in plants are isoleucine, lysine, methionine, threonine, and tryptophan (MLEAA) (Figure 2).

It has been difficult to produce significant increases in the essential amino acid content of crop plants utilizing classical plant breeding approaches. This is primarily due to the fact that the genetics of plant breeding is complex and that an increase in essential amino acid content may be offset by a loss in other agronomically important

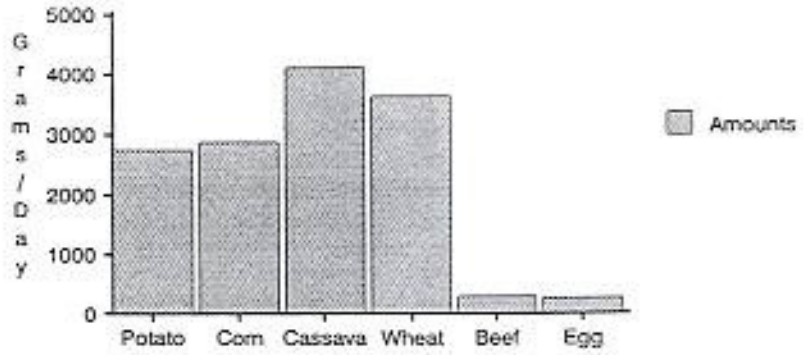

Figure 2 The amount of foodstuffs which must be consumed in grams per day in order to meet the minimum daily requirement of all essential amino acids.

characteristics. Also, it is probable that the storage proteins are very conserved in their structure and their essential amino acid composition would be little modified by these conventional techniques.

Structure and classification of natural storage proteins

Seed storage proteins can be characterized by several main features (Pernollet and Mosse, 1983): 1) their main function is to provide amino acids or nitrogen to the young seedling; 2) the general absence of any other known function; c) their peculiar amino acid composition in cereal and legume seeds; and d) their localization within storage organelles called protein bodies, at least during seed development. Several classes of storage proteins are generally recognized based on their solubilities in different solvents. Proteins soluble in water are called "albumins"; proteins soluble in 5% saline, "globulins"; and proteins soluble in 70% ethanol, "prolamins". The proteins that remain following these extractions are treated further with dilute acid or alkali, and are named "glutelins". Most cereals contain primarily prolamin-type proteins and can be classified into different groups on the basis of the relative proportions of prolamins, glutelins, and globulins, and the subcellular location of these proteins in the mature seed. The first group corresponds to the Panicoidae sub-family, the second group the Triticeae tribe, and the last one to oat and rice storage proteins.

The principal members of the Panicoideae sub-family are maize, sorghum, and millet. Their major storage proteins are prolamins (50 to 60% of seed protein) and glutelins (35 to 40% of seed protein; Pernollet and Mosse, 1983). Prolamins are stored within protein bodies, but glutelins are located both inside and outside these organelles. The Triticeae tribe which includes wheat, barley, and rye, differ from the Panicoideae mainly in storage protein localization and structure. In the starchy endosperm of the seeds belonging to this tribe, no protein bodies are left at maturity . Clusters of proteins are then deposited between starch granules, but are no longer surrounded by a membrane.

In legumes and most other dicots, the major storage proteins are salt-soluble globulins (80%) and prolamins (10–15%). Globulins can be divided into vicillins and legumins (Agros, 1985), based on their sedimentation coefficient (7S/11S), oligomeric organization (trimeric/hexameric), and polypeptide chain structure (single chain/disulphide-linked pair of chains). In the legume seed cotyledon, protein bodies are embedded between starch granules (Pernollet and Mosse, 1983). They are membrane-bound organelles, a few microns in diameter, mainly filled with storage proteins and phytates. Besides storage proteins, protein bodies also contain other proteins, such as enzymes or lectins, although in lesser amounts.

The structure of soluble globulins were studied more than the insoluble prolamins and glutelins. Vicillin appears as a homo- or heterotrimer, sometimes able to associate into hexameric form. Soybean β–conglycin and french bean phaseolin (Bollini and Chrispeels, 1978) are the structurally best known vicillins. Recently, the three-dimensional structure of phaseolin

was determined by X-ray crystallographic analysis (Lawrence et al., 1990). However, unlike other vicilins, the phaseolin trimer can associate into a dodecamer (tetramer of trimer) below pH 4.5. Each polypeptide of the trimeric form comprises two structurally similiar units each made up of a β–barrel and an α–helical domain.

Glycinin, the soybean legumin, has a quaternary structure that was suggested by Badley et al. (1975) to be twelve subunits packed in two identical hexagons. In general, the legumin molecule is a polymer formed by the association of six monomers. Each monomer consists of two subunits, acidic and basic. Sometimes these subunits are associated by disulfide linkages. On the other hand, arachin, the peanut legumin, was found to consist of different kinds of subunits. The arachin hexamer association does not need different kinds of subunits, which suggests that the subunits have a very similar structure.

The most studied storage proteins, in terms of structure, are the corn prolamines called zeins. These proteins perform no known enzymatic function. Three types of zeins (α, β and γ) (Esen, 1986) are synthesized on rough endoplasmic reticulum and aggregate within this membrane as protein bodies. The zein protein readily self-associates to form protein bodies and is insoluble in water even in low concentrations of salt. The presence of all types of zeins is not necessary for the formation of a protein body as a single type of zein can aggregate into a dense structure and is generally found at the surface of protein bodies (Lending et al., 1988; Wallace et al., 1988). The mechanism responsible for protein body formation is thought to involve hydrophobic and weak polar interactions between individual zein molecules (Wallace et al., 1988; Agros et al., 1982), while they require a high amount of ethanol in aqueous systems to maintain their strict molecular conformation (Agros et al., 1982).

Circular dichroic measurements, amino acid sequence analysis, and electron microscopy of a zein protein suggests that zein secondary structure to be primarily helical with nine adjacent, topologically anti-parallel helices clustered within a distorted cylinder (Agros et al., 1982; Larkins, 1983; Larkins et al., 1984). Polar and hydrophobic residues appropriately distributed along the helical surfaces allowing intra- and intermolecular hydrogen bonds and van der Waals interactions among neighboring helices, such that rod-shaped zein molecules could aggregate and then stack through glutamate interactions at the cylindrical caps. Because of this structure, zein is much less soluble under physiological conditions than the globulin phaseolin, and precipitation of insoluble zein in the tightly packed protein body may make them less available for proteolytic degradation (Greenwood and Chrispeels, 1985). The storage protein structures are adapted to a maximal packing within protein bodies (Pernollet and Mosse, 1983). Maximal packing is achieved in at least either of two ways. The folding of the polypeptide chain may favor the maximal packing of amino acids within the protein molecule, or the compacting of proteins is increased by the formation of closely packed quaternary structure. High degrees of polymerization can be observed in pearl millet pennisetin (Pernollet and Mosse, 1983) or zein (Lending et al., 1988; Wallace et al., 1988). Also, wheat prolamins and glutelin associate into aggregates arising in the formation of insoluble gluten.

These insoluble forms of protein deposits are osmotically inactive and stable during the long period of storage between the time of seed maturation and germination.

Regulation of storage protein genes

All storage proteins which have been investigated are encoded by multigene families (Bartels and Tompson, 1983; Crouch et al., 1983; Forde et al., 1985; Kasarda et al., 1984; Lycett et al., 1985; Rafalski et al., 1984; Slightom et al., 1983). The structure of these families varies, and in some cases, such as in wheat or barley, two major subgroups can be noted: the α– and γ–gliadins and the B- and C-hordeins, respectively (Forde et al., 1985; Kasarda et al., 1984; Rafalski et al., 1984). Within each subgroup, several subfamilies can be distinguished. Often short repeats account for at least part of the structure of the polypeptides. These repeats constitute links through which different subfamilies within the same species are related.

Storage protein genes, like most other plant genes characterized to date, are transcribed in a regulated rather than a constitutive fashion. Expression is frequently tissue-specific and(or) temporally regulated. Cis-acting DNA sequences involved in developmental and(or) tissue-specific regulation of gene expression can be defined by introducing plant storage protein gene regulatory regions coupled to bacterial reporter genes (Twell and Ooms, 1987; Wenzler et al., 1989, Marries et al., 1988; Chen et al., 1988), or by introducing entire or dissected genes (Colot et al., 1987; Chen et al., 1986) into a transgenic environment. Unfortunately, a transformation system for the nutritionally important cereal species has not yet been well established. Therefore, most regulation mechanisms have been studied with transgenic dicot plants. However, there is increasing evidence that gene expression is controlled, at least partly, by the interaction between regulatory molecules and short sequences that are present in the 5' flanking region of the gene.

The regulatory sequences of potato storage protein were investigated using transgenic potato plants. A 2.5 kb 5' flanking DNA fragment containing the promoter and the patatin gene was used to construct a transcriptional fusion gene with chloramphenicol acetyl transferase (CAT) or the β–glucuronidase (GUS) gene (Twell and Ooms, 1987; Wenzler et al., 1989). When reintroduced into potato, these chimeric genes were expressed in tubers, but not in leaves, stems or roots.

The expression pattern of storage protein genes of cereals is retained in tobacco, not only with respect to tissue, but also to temporal expression. The 5' upstream regions of wheat glutenin genes possess regulatory sequences that determine endosperm-specific expression in transgenic tobacco (Colot et al., 1987). Deletion analysis of the low molecular weight (LMW) glutenin sequence indicated that sequences present between 326bp and 160bp upstream of the transcriptional start point are necessary to confer endosperm-specific expression. Furthermore, cis-acting elements determining the regulation of each gene in the cluster are recognized by the tobacco trans-acting factor but also that cis-acting elements directing

expression of one gene do not affect expression of neighboring genes. This was demonstrated by the transfer of a 17.1 kb soybean DNA containing a seed lectin gene with at least four nonseed protein genes to transgenic tobacco plants (Okamuro, 1986). The genes in this cluster were expressed in a manner similar to that in soybean; i.e., the lectin gene products accumulated in seeds, and the other genes were expressed in tobacco leaves, stems, and roots.

The expression of several DNA deletion mutants with a 257 bp 5' flanking sequence of the α'-conglycin gene indicates that this region contained enhancer like elements (Chen et al., 1986). Only a low level of expression of the α' gene occurred in developing seeds of transgenic plants that contain the α' gene flanked by 159 nucleotides 5' of the transcriptional start site. However, a 20 fold increase in expression occurred when an additional 98 nucleotides of upstream sequence were included. The DNA sequence between 143 and 257 contained five repeats of the sequence AA(G)CCCA, and played a role in conferring tissue-specific and developmental regulation. The 35S promoter containing this sequence in different positions and different orientations could enhance the expression of the CAT gene by 25 to 40 fold (Chen et al., 1988).

Trans-acting factors directly involved in storage protein gene regulation have not yet been reported. However, in some cases, the level of amino acids can control the expression of storage protein. Vegetative storage protein (VSP) gene expression in leaves, stems and seed pods is closely related to whether these organs are currently a sink for nitrogen or a source for mobilized nitrogen for other organs (Staswick, 1989). The leaves have a sensitive mechanism for detecting changes in sink demand of mobilizing reserves, and VSP gene expression can be rapidly adjusted accordingly. Sequestering excess amino acids in this way may prevent their accumulation to toxic levels.

Genetic engineering using *Agrobacterium Tumefaciens*

One of the most significant recent advances in the area of plant molecular biology has been the development of the *Agrobacterium tumefaciens* Ti plasmid as a vector system for the transformation of plants. In nature, *A. tumefaciens* infects most dicotyledonous and some monocotyledonous plants by entry through wound sites. The bacteria bind to cells in the wound and are stimulated by phenolic compounds released from these cells to transfer a portion of their endogenous, 200 kb Ti plasmid into the plant cell (Weiler and Schroder, 1987). The transferred portion of the Ti plasmid, T-DNA, becomes covalently integrated into the plant genome where it directs the biosynthesis of phytohormones using enzymes which it encodes. The vir gene in the bacterial genome is known to be responsible for this process. In addition to vir gene products, directly repeating sequences of 25 bases called "border" sequences are essential, but only the right terminus has been shown to be used for T-DNA transfer and integration.

Expression of the T-DNA gene inside the plant results in the uncontrolled growth of these and surrounding cells, leading to formation of a gall (Weiler and Schroder, 1987). Ti plasmids, from which these disease-producing genes have been removed or replaced, are referred to as "disarmed" and can be used for the introduction of foreign genes into plants. The great size of the disarmed Ti plasmid and lack of unique restriction endonuclease sites prohibit direct cloning into the T-DNA. Instead, intermediate vectors such as pMON237 or pBI121 can be used to introduce genes into the Ti plasmid. Currently, two kinds of vector systems are available as intermediate vectors: cointegrating vectors and binary vectors. A cointegrating transformation vector must include a region of homology between the vector plasmid and the Ti plasmid. Once recombination occurs, the cointegrated plasmid is replicated by the Ti plasmid origin of replication. The cointegrate system, while more difficult to use, does offer advantages. Once the cointegrate has been formed, the plasmid is stable in Agrobacterium.

A binary vector contains an origin of replication from a broad host-range plasmid instead of a region of homology with the Ti plasmid. Since the plasmid does not need to form a cointegrate, these plasmids are considerably easier to introduce into Agrobacterium. The other advantage to binary vectors is that this vector can be introduced into any Agrobacterium host containing any Ti or Ri plasmid, as long as the vir helper function is provided. Using these systems, the gene regulation mechanism of storage proteins has been elucidated.

Improvement of nutritional qualities of plants

The amino acid composition of the cereal endosperm protein is characterized by a high content of proline and glutamine while the amount of essential amino acids, lysine and tryptophan in particular, is a limiting factor (Pernollet and Mosse 1983). In legumes, sulfur containing amino acids such as methionine and cysteine are the major limiting essential amino acids for the efficient utilization of plant protein as animal or human food while roots and tubers are deficient in almost all of the essential amino acids.

There has been a great deal of effort to overcome these amino acid limitations by breeding and selecting for more nutritionally balanced varieties. Plants have been mutated in hopes of recovering individuals with more nutritious storage proteins. Neither of these approaches has been very successful, although some naturally occurring and artificially produced mutants of cereals were shown to contain a more nutritionally balanced amino acid composition. These mutations cause a significant reduction in the amount of storage protein synthesized and thereby result in a higher percentage of lysine in the seed; however, the softer kernels and low yield of such strains have limited their usefulness (Pernollet and Mosse, 1983). The reduction in storage protein also causes the seeds to become more brittle; as a result, these seeds shatter more easily during storage. The lower levels of prolamin also result in flours with unfavorable functional properties which cause brittleness in the

Expression of the T-DNA gene inside the plant results in the uncontrolled growth of these and surrounding cells, leading to formation of a gall (Weiler and Schroder, 1987). Ti plasmids, from which these disease-producing genes have been removed or replaced, are referred to as "disarmed" and can be used for the introduction of foreign genes into plants. The great size of the disarmed Ti plasmid and lack of unique restriction endonuclease sites prohibit direct cloning into the T-DNA. Instead, intermediate vectors such as pMON237 or pBI121 can be used to introduce genes into the Ti plasmid. Currently, two kinds of vector systems are available as intermediate vectors: cointegrating vectors and binary vectors. A cointegrating transformation vector must include a region of homology between the vector plasmid and the Ti plasmid. Once recombination occurs, the cointegrated plasmid is replicated by the Ti plasmid origin of replication. The cointegrate system, while more difficult to use, does offer advantages. Once the cointegrate has been formed, the plasmid is stable in Agrobacterium.

A binary vector contains an origin of replication from a broad host-range plasmid instead of a region of homology with the Ti plasmid. Since the plasmid does not need to form a cointegrate, these plasmids are considerably easier to introduce into Agrobacterium. The other advantage to binary vectors is that this vector can be introduced into any Agrobacterium host containing any Ti or Ri plasmid, as long as the vir helper function is provided. Using these systems, the gene regulation mechanism of storage proteins has been elucidated.

Improvement of nutritional qualities of plants

The amino acid composition of the cereal endosperm protein is characterized by a high content of proline and glutamine while the amount of essential amino acids, lysine and tryptophan in particular, is a limiting factor (Pernollet and Mosse 1983). In legumes, sulfur containing amino acids such as methionine and cysteine are the major limiting essential amino acids for the efficient utilization of plant protein as animal or human food while roots and tubers are deficient in almost all of the essential amino acids.

There has been a great deal of effort to overcome these amino acid limitations by breeding and selecting for more nutritionally balanced varieties. Plants have been mutated in hopes of recovering individuals with more nutritious storage proteins. Neither of these approaches has been very successful, although some naturally occurring and artificially produced mutants of cereals were shown to contain a more nutritionally balanced amino acid composition. These mutations cause a significant reduction in the amount of storage protein synthesized and thereby result in a higher percentage of lysine in the seed; however, the softer kernels and low yield of such strains have limited their usefulness (Pernollet and Mosse, 1983). The reduction in storage protein also causes the seeds to become more brittle; as a result, these seeds shatter more easily during storage. The lower levels of prolamin also result in flours with unfavorable functional properties which cause brittleness in the

baked products (Pernollet et al., 1983). Thus, no satisfactory solution has yet been found for improving the amino acid composition of storage proteins.

One direct approach to this problem would be to modify the nucleotide sequence of genes encoding storage proteins so that they contain high levels of essential amino acids. To achieve this aim, several laboratories have tried to modify and express storage proteins in the host plants. Modified storage proteins have been created and expressed by changing their codon sequences. In vitro mutagenesis was used to supplement the sulfur amino acid codon content of a gene encoding β–phaseolin, a *Phaseolus vulgaris* storage protein (Hoffman et al., 1988). The nutritional quality of β–phaseolin was increased by the insertion of 15 amino acids, six of which were methionine. The inserted peptide was essentially a duplication of a naturally occurring sequence found in the maize 15kd zein storage protein (Pederson et al., 1986). However, this modified phaseolin achieved less than 1% of the expression level of normal phaseolin in transformed seeds. Recently it has been found that this insertion was made in part of a major structural element of the phaesolin trimer (Lawrence et al., 1990). Therefore, an inclusion of 15 residues at this site could distort the structure at the tertiary and(or) quaternary level.

Lysine and tryptophan-encoding oligonucleotides were introduced at several positions into a 19kD α–type zein complementary DNA by oligonucleotide-mediated mutagenesis (Wallace et al., 1988). Messenger RNA for the modified zein was synthesized in vitro and injected into *Xenopus laevis* oocytes. The modified zein aggregated into structures similar to membrane-bound protein bodies. This experiment suggested the possibility of creating high-lysine corn by genetic engineering.

There are alternative approaches that might be more practical. One of these is to transfer heterologous storage protein genes that encode storage proteins with higher levels of the desired amino acids. For this purpose, a chimeric gene encoding a Brazil nut methionine-rich protein which contains 18% methionine has been transferred to tobacco and expressed in the developing seeds (Altenbach et al., 1989). The remarkably high level of accumulation of the methionine-rich protein in the seed of tobacco results in a significant increase in methionine levels of ~30%.

The maize 15 kd zein structural gene was placed under the regulation of French bean β–phaseolin gene flanking regions and expressed in tobacco (Hoffmann et al., 1987). Zein accumulation was obtained as high as 1.6% of the total seed protein. Zein was found in roots, hypocotyls, and cotyledons of the germinating transgenic tobacco seeds. Zein was deposited and accumulated in the vacuolar protein bodies of the tobacco embryo and endosperm. The storage proteins of legume seeds such as the common bean (*Phaseolus vulgaris*) and soybean (*Glycine max*) are deficient in sulfur-containing amino acids. The nutritional quality of soybean could be improved by introducing and expressing the gene encoding methionine-rich 15 kd zein (Pederson et al., 1986).

A 292 bp synthetic gene (HEAAE I = .High Essential Amino Acid Encoding) which encoded a protein domain high in essential amino acid

was expressed as a CAT-HEAAE I fusion protein in potato (Jaynes et al., 1986; Yang et al., 1989). However, structural instability limited the high level expression of this fusion protein in the potato system. Also, the content of essential amino acids was diluted to less than 40% of the original encoded protein by constructing this fusion.

There are several precautions that should be considered in engineering storage proteins (Larkins, 1983). First, in vitro mutational change must not be in regions of the protein that perturb the normal protein structure; otherwise, the proteins might be unstable. Second, when attempting to increase nutritional quality by introducing a gene encoding a heterologous protein in crop plants, it is important that the protein encoded by an introduced gene does not produce any adverse effects in humans or livestock, the ultimate consumers of the engineered seed proteins (Altenbach et al., 1989). Finally, it is critical that the amino acids present in the introduced protein are able to be utilized by the animal for growth and development.

De Novo design of proteins

Recently, a new field in protein research, de novo design of proteins, has made remarkable progress due to a better understanding of the rules which govern protein folding and topology. Protein design has two components: the design of activity and the design of structure. This review will concentrate on the design of structurally stable storage protein-like proteins.

The usual approach for the design of helical bundle proteins consists of linking sequences with a propensity for forming an a-helix via short loop sequences to get linear polypeptide chains. This chain can fold into the predetermined 'globular type' tertiary structure in aqueous solution (Mutter 1988; DeGrado et al. 1989). α–helical secondary structures are stabilized by interatomic interactions that can be classified according to the distance between interacting atoms in the sequence of the protein (DeGrado et al., 1989).

Short range interactions account for different amino acids having different conformational preferences. Both statistical (Chou and Fasman, 1978) and experimental (Sueki et al., 1984) methods show that residues such as Glu, Ala, and Met tend to stabilize helices, whereas residues such as Gly and Pro are destabilizing. However, these intrinsic preferences are not sufficient to determine the stability of helices in globular proteins.

Analysis of the free-energy requirements for helix nucleation and propagation indicates that peptides of 10 to 20 residues should show little helix formation in water (Bierzynski et al., 1982) when the Zimm-Bragg equation (Zimm and Bragg, 1959) is used, with parameters (s and S) determined by host-guest experiments where s is the helix nucleation constant, n is the number of H-bonded residues in the helix and S is an average stability constant for one residue:

$$sSn\text{-}1/(S\text{-}1)$$

Nevertheless, the 13 amino acid C-peptide obtained from RNase A does show measurable helicity (~25%) at low temperature (Bierzynski et al., 1982; Brown and Klee, 1981). The stability of this peptide is 1000–fold greater than the value calculated from the Zimm-Bragg equation. Specific side-chain interactions, factors that are not considered in the Zimm-Bragg model, are responsible at least in part for the fact that the C-peptide is much more helical than predicted (Scherega, 1985).

Medium-range interactions are responsible for the additional stabilization of secondary structures (DeGrado et al., 1989). Interaction between the side-chains are regarded as important medium range interactions (Shoemaker et al., 1987; Marqusee and Baldwin, 1987). These include electrostatic interactions, hydrogen bonding, and the perpendicular stacking of aromatic residues (Blundell et al., 1986). An α–helix possesses a dipole moment as a result of the alignment of its peptide bonds. The positive and negative ends of the amide group dipole point toward the helix NH_2-terminus and COOH-terminus, respectively, giving rise to a significant macrodipole. Appropriately charged residues near the ends of the helix can favorably interact with the helical dipole and stabilize helix formation. It was estimated that the electrostatic interaction between a pair of antiparallel α–helices is about 20Kcal/mol less than a parallel α–helices pair (Hol and Sanders, 1981). Hydrogen bonds between side chains and terminal helical N-H and C=O groups also participate in the stabilization of helical structure (Richardson and Richardson, 1988; Presta and Rose, 1988; Richardson and Richardson, 1989).

Protein structures contain several long-range stabilizing interactions which include hydrophobic and packing interactions, and hydrogen bonds. Among these, the hydrophobic effect is a prime contributor to the folding and stabilizing of protein structures. The driving force for helix formation in RNase A arises from long-range interactions between C-peptide and S-protein, a large fragment of the protein from which C-peptide was excised (Komoriya and Chaiken, 1985).

The role of hydrophobic interactions in determining secondary structures was studied for a series of peptides containing only Glu and Lys in their sequence (DeGrado and Lear, 1985). Glu and Lys residues were chosen as charged residues for the solvent-accessible exterior of the protein to help stabilize helix formation by electrostatic interaction.

Stability of designed proteins

Hydrophobic residues often repeat every three to four residues in an α–helix and form an amphiphilic structure (DeGrado et al., 1989). Amphiphilicity is important for the stabilization of the secondary structures of peptides and proteins which bind in aqueous solution to extrinsic apolar surfaces, including phospholipid membranes, air, and the hydrophobic binding sites of regulatory proteins (Degrado and Lear, 1985). This amphiphilic secondary structure can be stabilized relative to other conformations by self-association. Therefore, short peptides often form the α–helix in water only because the helix is amphiphilic and is stabilized by peptide aggregation along the hydrophobic surface. Natural globular proteins

are folded by a similar mechanism, involving hydrophobic interaction between neighboring segments of secondary structure (Presnell and Cohen, 1989). Using the concept of an amphiphilic helix, DeGrado and coworkers have successfully built peptide-hormone analogs with minimal homology to the native sequences. These peptides, like the native ones, are not helical in solution but do form helices at the hydrophobic surfaces of membranes.

Designed synthetic peptides have been used to show how hydrophobic periodicity in a protein sequence stabilizes the formation of simple secondary structures such as an amphiphilic α–helix (Ho and DeGrado, 1987). The strategies used in the design of the helices in the four-helix bundles are: 1) the helices should be composed of strong helix forming amino acids, and 2) the helices should be amphiphilic; i.e., they should have an apolar face to interact with neighboring helices and a polar face to maintain water solubility of the ensuing aggregates. The results show that hydrophobic periodicity can determine the structure of a peptide. Therefore, the peptides tend to have random conformations in very dilute solution, but form secondary structures when they self-associate (at high concentration) or bind to the air-water surface.

The free energy associated with dimerization or tetramerization of the designed peptides could be experimentally determined from the concentration dependence of the CD spectra for the peptides (DeGrado et al., 1989; Lear et al., 1988; DeGrado and Lear, 1985). At low concentrations, the peptides were found to be monomeric and have low helical contents, whereas at high concentration they could self-associate and stabilize the secondary structure. Therefore, possible hairpin loops between helices can affect the stability of the secondary structure by enhancing the self-association between the helical monomers. A strong helix breaker (Chou and Fasman, 1978; Kabsch and Sander, 1983, Sueki et al., 1984, Scheraga, 1978) was included as the first and last residue to set the stage for adding a hairpin loop between the helices. A single proline residue appeared capable of serving as a suitable link if the C and N terminal glycine residues are slightly unwound. Glycine lacks a β–carbon, which is essential for the reverse turn where positive dihedral angles are required. The pyrrolidine ring of proline constrains its f dihedral angle –60°. Thus, proline should be destabilizing at positions where significantly different backbone torsion angles are required. This amino acid, as well as glycine, has a high tendency to break helices and occurs frequently at turns (Creighton, 1987).

The direct evidence for stabilization of protein structure by adding the linking sequence was observed by comparing the guanidine denaturation curve for a monomer, dimer and tetramer (Degrado et al., 1989). The gene encoding tetrameric protein was expressed in *E. coli* and purified to homogeneity. In the series of mono-, di-, and tetramer, the stability toward guanidine denaturation increases concomitantly with the increase in covalent cross-links between helical monomer. At equivalent peptide concentrations, the midpoints of the denaturation curves occurred at 0.55, 4.5 and 6.5M guanidine for the mono-, di, and tetramer. Furthermore, as the number of covalent cross-links was increased, the curves became increasingly cooperative. Thus, the linker sequence

stabilized the formation of the four helix structures at low concentration of the peptides (<1mg/ml).

Structural stability of proteins is directly related to in vivo proteolysis (Parasell and Sauer, 1989). Proteolysis depends on the accessibility of the scissile peptide bonds to the attacking protease. The sites of proteolytic processing are generally in relatively flexible interdomain segments or on the surface of the loops, in contrast to the less accessible interdomain peptide bonds (Neurath, 1989). This suggests that the stability of the folded state of the protein is the most important determinant of its proteolytic degradation rate. The effect of a folded structure on the proteolytic degradation has been proven by several experiments. First, proteins that contain amino acid analogs or are prematurely terminated are often degraded rapidly in the cells (Goldberg and St. John, 1976). Second, there are good correlations between the thermal stabilities of specific mutant proteins and their rates of degradation in *E. coli* (Pakula and Sauer, 1986, Parasell and Sauer, 1989). Finally, second-site suppressor mutations that increase the thermodynamic stability of unstable mutant proteins have also been shown to increase resistance to intracellular proteolysis (Pakula and Sauer, 1989). The solubility of proteins could also affect their proteolytic resistance as some proteins aggregate to form inclusion bodies that escape proteolytic attack (Kane and Hartley, 1988). Metabolic stability is another factor influencing the in vivo stability of proteins. Usually, damaged and abnormal proteins are metabolically unstable in vivo (Finley and Varshavsky, 1985; Pontremoli and Melloni, 1986). In eukaryotes, covalent conjugation of ubiquitin with proteins is essential for the selective degradation of short-lived proteins (Finley and Varshavsky, 1985). It was found that the amino acid at the amino-terminus of the protein determined the rate of ubiquitination (Bachmair et al., 1986). Both prokaryotic and eukaryotic long-lived proteins have stabilizing amino acids such as methionine, serine, alanine, glycine, threonine, and valine at the amino terminus end. On the other hand, amino acids such as leucine, phenylalanine, aspartic acid, lysine, and arginine destabilize the target proteins.

Designed nutritional proteins

We designed the synthetic protein DNP 1 to contain a high content of those amino acids which are essential to the diet of animals. The optimized content of essential amino acids for this new protein was obtained empirically by determining the amounts of essential amino acids necessary for normal metabolism of the animal. We also determined the "deficiency values" or the ratios of deficient essential amino acids for the 10 primary crops animals consume throughout the world (Figure 3). From these data, we then found the ratio of essential amino acids needed to totally complement each particular plant foodstuff. We merely averaged these values and came up with a set of numbers we call the 'Average Ratio for All Crops Idealized to the DNP 1 Monomer' (Figure 4). This set of numbers represents the ratio of essential amino acids necessary

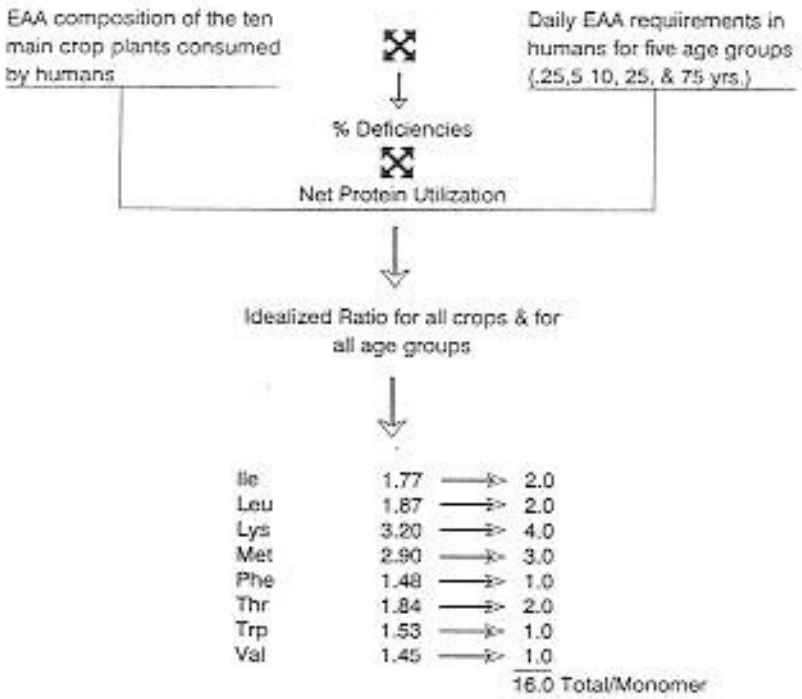

Figure 3 How the composition ratio of the HDNP 1 monomer was calculated. The ratio is an overall average which takes into consideration all daily requirements for five age groups: the deficiencies of the ten major crop plants people consume: rice, wheat, barley, sorghum, maize, potato, sweet potato, plantain, cassava, and taro; and how well these plant foodstuffs are utilized as a protein source.

to complement the deficiencies found in all 10 crops for all human age groups.

From the above set of numbers, we designed the DNP 1 protein for humans (HDNP 1). It has 1.8 times more of the essential amino acids compared to zein or phaseolin. The difference in MLEAA is much higher, containing three times more than phaseolin and 6.5 times more than zein. The helical region of HDNP 1 is amphipathic (hydrophobic residues clustered on one face of the helix while hydrophilic residues are found on the other face) and is stabilized by several GLU – LYS salt bridges (Figure 5). The helix breaker Gly-Pro-Gly-Arg has been used as a turn sequence. The design results in an antiparallel tetramer which achieves an extraordinarily stable secondary and tertiary structure even at low concentration.

The structural stability of a protein is important in determining its susceptibility to proteolysis. Most native proteins are relatively resistant to cleavage by proteolytic enzymes, whereas denatured proteins are much more sensitive (Pace and Barret, 1984). Several findings suggest that the stability of a folded protein is an important determinant of its rate of degradation. Therefore, in addition to improved nutritional quality, HDNP 1 has been designed to have a stable storage protein-like

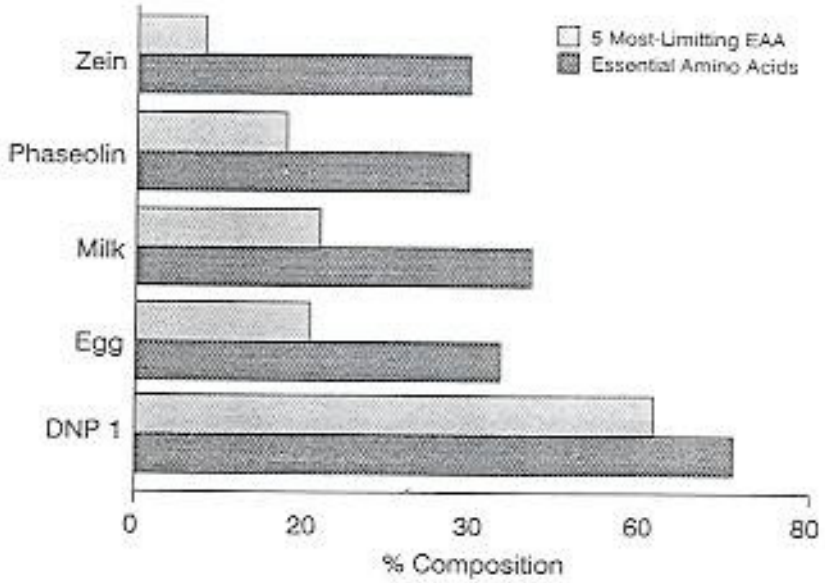

Figure 4 Percentage of essential amino acids (EAA) and percent of most limiting essential amino acids (MLEAA) in DNP 1 tetramer compared to natural proteins.

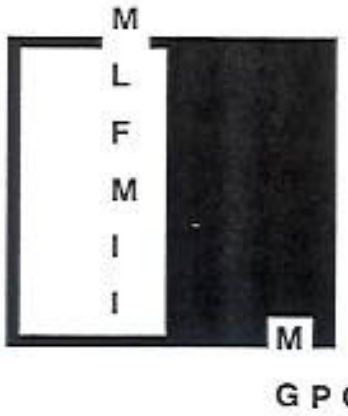

Figure 5 A depiction of the amphiphilicity of the HDNP 1 monomer where hydrophobic amino acids are in the white rectangle and hydrophilic in the shaded rectangle. Note the interaction between the Glu (E) and Lys (K) residues (dark lines depict salt-bridges).

structure in plants. Its design is based on the structurally well-studied corn storage zein proteins (Z19 and Z22), which are comprised of nine repeated helical units (Agros et al., 1982). Each helical unit of zein, 16 to 26 amino acids long, is flanked by turn regions and forms an antiparallel helical bundle. Most of the amino acids in the helices are hydrophobic residues. On the other hand, HDNP 1 is comprised of four helical repeating units, each 20 amino acids long (Figure 6). Increased

Figure 6 The amino acid sequence of the HDNP 1 tetramer. Hydrophilic amino acids are underlined and β–turns so indicated.

gene copy number by concatenation can increase the protein yields. At the same time, gene concatenation gives the increased molecular mass of the encoded protein. Such an increase in size and concatenation can significantly stabilize an otherwise unstable product (Shen, 1984).

The 284bp gene encoding this novel peptide was chemically synthesized and cloned into an *E. coli* expression vector. This gene contains plant consensus sequences at the 5' end of the translation initiation site to optimize the expression of proteins in vivo. It was placed under the control of the 35S cauliflower mosaic virus (CaMV) promoter in order to permit the constitutive expression of this gene in tobacco.

Prediction of the structure of HDNP 1

The secondary structures of the HDNP 1 monomer and tetramer were predicted by PREDICT-SECONDARY in β–SYBYL. The percentage of α–helix content predicted by information-theory showed a higher α–helix content compared to the other two prediction methods (Bayes-statistic and neural-net) in PREDICT-SECONDARY. The predicted secondary structures by information-theory gave 100% helical content for the monomer and 74% for the tetramer.

However, the accuracy of the three widely used prediction methods ranged from 49% to 56% for prediction of three states; helix, sheet, and coil (Kabsch and Sander, 1983). This inaccuracy might be due to the small size of the data base and(or) the fact that secondary structure is determined by tertiary interactions which are not included in the local sequences. For further predictions of structure, the structures predicted by information-theory were energy minimized using SYBYL MAXIMIN2.

A perfect amphiphilic α–helical conformation was predicted for the HDNP 1–monomer after minimization. The tertiary structure of the HDNP 1–tetramer after minimization showed the antiparallel confor-

mation as was designed. These minimization results suggested the high probability of stable secondary structure (α-helix and β–turn) formation of the HDNP 1–monomer and –tetramer. However, it might be too early to conclude the tertiary structure of the HDNP 1–tetramer with minimization only without considering longer range interactions which are the most important determinant of protein folding.

The predominant driving force for folding the HDNP 1–tetramer might be the longer range hydrophobic interactions between the α–helical monomers, because HDNP 1 was designed to be amphiphilic. However, no currently available force field for the minimization of tertiary structure contains these parameters and could not give the perfect predicted tertiary structure of the protein. Actually, minimization schemes alone have failed to predict chain folding accurately (Fasman, 1989). Therefore, we might be able to obtain indications of a much more stable secondary and tertiary structures if we considered this factor for the minimization of HDNP 1–tetramer.

Structural analysis of HDNP 1 protein

The structural stability of HDNP 1–monomer and tetramer could not be determined by minimization only. Therefore, the stability of the α–helical secondary structure of HDNP 1–monomer was investigated. HPLC analysis of the gel filtered synthetic HDNP 1–monomer showed that purity was more than 90% and amino acid analysis of the purified fraction gave the expected molar ratios. This fraction was also analyzed by mass spectrometry, and the molecular weight peak corresponding to the HDNP 1–monomer (2896.5) was present. Since the structural stability of HDNP 1–monomer and tetramer could not be determined by minimization only, the stability of the α–helical secondary structure of HDNP 1–monomer was investigated by circular dichroism (CD) analysis. CD spectra of HDNP 1–monomer showed the typical pattern of – helical proteins with double minima at 208 and 222 nm in aqueous solution (data not shown). The stability of the secondary structure can be induced by the inter-molecular interaction between the helical chains (DeGrado et al., 1989). Therefore, stable aggregation between monomers, presumably through hydrophobic interactions, could stabilize the helical structure. Besides, proper packing of the apolar side chains and proper electrostatic interaction might play important roles in stabilizing the secondary structure of HDNP 1. The stable interaction among the monomeric HDNP 1 molecules is an important determinant for the proper folding into the tertiary structure of the HDNP 1–tetramer. Therefore, the self-association capability of the HDNP 1–monomers was investigated by using size exclusion chromatography. The hydrodynamic behavior of this peptide showed that it was aggregated into a hexamer form with an apparent molecular weight of about 17 kD. This hexameric aggregate could be maintained in either low or high ionic strength solutions. This result provides proof of the stable globular type tertiary structure formation of tetrameric HDNP 1.

Three potential β–turn (Gly-Pro-Gly-Arg) sequences were inserted between four monomers for the HDNP 1–tetramer construction. The β–turn could play an important role for structural stability of the HDNP 1–tetramer when it is expressed in vivo. It can also help stabilize tertiary structure formation. The interactions between the helical monomers might be much faster due to the proximate effect when they are connected. This proximate effect might be critical for folding at the low concentrations of HDNP 1–tetramer that are possible when they are expressed in vivo. At the same time, the stability of the secondary structure is increased by the hydrophobic interactions between helical monomers. In addition, this β–turn sequence has a tryptic digestion site (Gly-Arg) which could increase the digestibility of this protein when it is consumed by animals.

The stability of the folded structure of a protein has a close relation to its proteolytic degradation rate (Pace and Barret, 1984; Pakula and Sauer, 1986; Parasell and Sauer, 1989; Pakula and Sauer, 1989). In this respect, we expected high stability of folded HDNP 1–tetramer against proteolytic degradation when it is expressed in vivo. Stable quaternary structure is essential for the formation of protein bodies of storage proteins in zein or phaseolin (Lawrence et al., 1990). These higher order structures can be achieved through the interaction and close packing of the stable tertiary structures. The major driving force for this quaternary structure formation is also hydrophobic interaction between the tertiary structures. At this moment, designing and predicting of the quaternary structure is not easy but our data suggest that precisely this might be occurring with HDNP 1–tetramer.

Introduction of HDNP 1 gene into tobacco

The correct insertion and orientation of the pBI derivative containing the HDNP 1 tetramer was screened by EcoRI and HindIII digestion (it was found in *E. coli* that the most stable form of the gene was the tetramer form). The EcoRI digestion gave a fragment of the expected size, 3.2 kb, which consisted of 3'NOS of HDNP 1 and the GUS gene (data not shown). Also, the HDNP 1 gene with its 35S promoter and 3'NOS sequences was detected as a 1.4 kb band by HindIII digestion. Stable transformation of the HDNP 1 gene into *A. tumefaciens* LBA4404 was confirmed by HindIII digestion of isolated plasmid DNA. It could be isolated from Agrobacterium and detected by enzyme digestion because pBI121 is a binary vector. Leaf discs, transformed with LBA4404 carrying the HDNP 1 gene, gave about five to seven shoots two to three weeks after infection. A total of 565 kanamycin-resistant shoots were regenerated from 120 leaf discs. These shoots were excised from the leaf discs and transferred to new media to grow several more weeks and then transferred to rooting media. After three weeks in rooting medium, 126 rooted shoots were analyzed for β–glucuronidase (GUS). Root tips of 56 out of 126 plants showed various levels of GUS activity. Not all the kanamycin-resistant shoots showed the GUS positive result. Although kanamycin resistance was due to the expression of neomycin

phosphotransferase (NTP II gene), regeneration of nontransgenic shoots in the presence of kanamycin has been reported. Therefore, escapes from the screening based on kanamycin sensitivity might have occurred in the nontransformed plants, making them kanamycin resistant.

Thirty six plantlets which showed high levels of β–glucuronidase activity were transplanted into jiffy pots. After establishment of the plants, a more accurate fluorogenic assay for GUS activity was done to quantify the expression level of this gene (data not shown). Some of these transformed tobacco plants showed higher levels of β–glucuronidase activity compared to other plants. The level of expression might be primarily affected by whether the gene is incorporated into an active or inactive site of chromatin. Activity of chromatin, methylation of DNA and nuclease hypersensitivity are closely related to each other. It has been found that the nuclease hypersensitive sites correlate to active transcription (Gross and Garrard, 1987). The degree of methylation of DNA is inversely related to gene expression. Furthermore, if the gene is located near the plant's endogenous promoter or enhancer sites, the level of expression of this gene will be increased by these near-by enhancing factors. Therefore, the difference in the levels of GUS activity between the transformed plants might be due to this positional effect, which was determined by the sites of incorporation of this gene into the tobacco genome.

Expression of HDNP 1 in transformed plants

Polyclonal antibody raised against synthetic HDNP 1 monomer was used to detect the production of stable HDNP 1 protein in tobacco. High levels of the tetrameric form (11.2 kd) of the HDNP 1 protein were detected from plant #17 by western blot analysis (data not shown). Therefore, direct correlation was found between gene copy number, number of genetic NPTII loci, GUS expression, accumulation of HDNP 1 transcript and protein expression level in the case of plant #17. Some heterologous seed proteins undergo specific degradation when expressed in transgenic plants. A significant amount of the immunoreactive protein accumulated in tobacco seed expressing the phaseolin gene is smaller than the final processed protein (Sengupta et al., 1985). A similar result was found when β–conglycinin was expressed in transgenic petunia (Beachy et al., 1985). In contrast to these results, the HDNP 1 protein appears to be quite stable in transgenic tobacco plants.

Amino acid and total protein analyses were conducted on leaf tissue from several of the transgenic plants which produced detectable levels of HDNP 1. Surprisingly, we found that the overall levels of all amino acids were increased with some of the plants being remarkably high. (Figure 7). This rather disconcerting result has been repeated numerous times and the overall levels of all amino acids in the transgenic plants remain significantly elevated. Other methods of determining overall protein content have been used with similar trends observed. For example, comparison of total protein densitometric values derived from SDS-PAGE of equivalent samples (on a weight basis) yield the same results

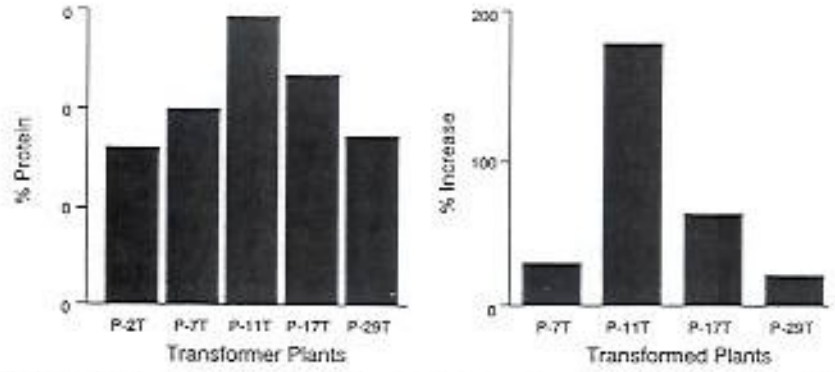

Figure 7 Overall protein content determined by amino acid analysis from control (P-2 TC) and various HDNP 1 tetramer plants. These data were derived from seedlings obtained from transformed mother plants. A minimum of four separate assays were used and the variation was no more than 30%. Percentage increase of HDNP 1 plants over the control is also shown.

(data not shown). At this time, we can offer no definitive explanation. Therefore, in addition to being a very stable protein in a plant cell, HDNP 1 must function as a general 'protein-stabilizer' and reduces overall protein turnover without apparent deleterious effects to the plants, since there is no observable difference in growth characteristics in the plants producing high amounts of HDNP 1 over control plants.

Conclusions

HDNP 1 is the first reported de novo designed protein expressed in plant systems. We have found that it is possible to construct a protein of high nutritional value mimicking the well known physical characteristics of plant storage proteins. Plans for future studies of tetrameric HDNP 1 begin with further analysis of its structural properties such as: 1) stability against proteolytic attack, and 2) solubility and aggregation pattern. Additionally, we are intensively studying the import of its apparent ability to reduce overall protein turnover. Subsequent plans are to pursue high level tissue-specific expression of this gene in the more economically important plants such as potato, soybean, and selected cereals.

It should be clear that the application of the above-described techniques for protein design could have a profound impact on how humans feed themselves and their livestock. It is possible that within the very near future (less than five years) traditional food and feed plants could be produced that contain the full-complement of essential amino acids to meet the specific needs of each species considered. Thus, foundation seed-stock of so-called genetically-engineered "human-corn", "swine-corn", "poultry-corn", etc. could be provided to farmers for high-level production, distribution, and sale to the food and feed industry.

References

Agros, P., Pederson, K., Marks, D. and Larkins, B.A. 1982. A structural model for maize zein proteins. J. Biol. Chem. 257: 9984–9990.

Agros, P., Naravana, S.V.L., and Nielsen, N.C. 1985. Structural similarity between legumin and vicillin storage proteins from legumes. The EMBO J. 4: 1111–1117.

Altenbach, S.B., Pederson, K.W., Meeker, G. Staraci, L.C., and Sun, S.S.M. 1989. Enhancement of the methionine content of seed proteins by the expression of seed proteins by the expression of a chimeric gene encoding a methionine-rich protein in transgenic plants. Plant Mol. Biol. 13: 513–522.

Bachmair, A., Finley, D., and Varshavsky, A. 1986. In vivo halflife of a protein is a function of its amino-terminal residue. Science 234: 179–186.

Badley, R.A., Atkinson, D., Hauser, H., Oldani, D., Green, J.P., and Stubbs, J.M. 1975. The structure; physical and chemical properties of the soybean protein glycinin. Biochim. Biophys. Acta. 412: 214–228.

Bartels, D., and Tompson, R.D. 1983. The characterization of cDNA clones coding for wheat storage proteins. Nucleic Acid Res. 11: 2961–2977

Beachy, R.N., Chen, Z.L., Horsch, R.B., Rogers, S.G., Hoffman, N.J., and Fraley, R.T. 1985. Accumulation and assembly of soybean '-conglycinin in seeds of transformed petunia plants. EMBO J. 4: 3047–3053.

Bierzynski, A., Kim, P.S., and Baldwin, R.L. 1982. A salt bridge stabilizes the helix formed by isolated c-peptide of RNAse A. Proc. Natl. Acad. Sci. U. S.A. 79: 2470–2474.

Blundell, T.L., Thornton, S.J., Burley, S.K., and Petsco, G.A. 1986. Atomic interactions. Science 234: 1005–1009.

Bollini, R. and Chrispeels, M.J. 1978. Characterization and subcellular localization of vicillin and phyto-hemaglutinin, the two major reserve proteins of *Phaseolus vulgaris*. Planta 142: 291–298.

Brown, J.E. and Klee, W.A. 1971. Helix-coil transition of the isolated amino terminus of Ribonuclease. Biochemistry 10: 470–476.

Chen, Z.L., Pan, N.S., and Beachy, R.N. 1988. A DNA sequence element that confers seed-specific enhancement of a constitutive promoter. The EMBO J. 7: 297–302.

Chen, Z.L., Schuler, M.A. and Beachy, R.N. 1986. Functional analysis of regulatory elements in a plant embryo-specific gene. Proc. Natl. Acad. Sci. U. S.A. 83: 8560–8564.

Chou, P.Y. and Fasman, G.D. 1978. Prediction of the secondary structure of proteins from amino acid sequence. Adv. Enzymol. 47: 45–148.

Colot, V., Robert, L.S., Kavanagh, T.A., Beavan, M.W. and Tompson, R.D. 1987. Localization of sequences in wheat endosperm protein genes which confer tissue-specific expression in tobacco. The EMBO J. 6:3559–3564.

Creighton, T.E. 1984. Proteins. New York: Freeman.

Crouch, M., Tenberge, K., Simone, N.E., and Ferl, R. 1983. Sequence

USE THIS SIDE ONLY

of the 1.7K storage protein of Brassica napus. Mol. Appl. Genet. 2:273–283.

Degrado, W.F., and Lear, J.D. 1985. Induction of peptide conformation at apolar/water interfaces. J. Am. Chem. Soc. 107: 7684–7689.

Degrado, W.F., Wasserman, Z.R., and Lear, J.D. 1989. Protein design, a minimalist approach. Science 241: 622–628.

Esen, E. 1986. Separation of alcohol-soluble proteins (zeins) from maize into three fractions by differential solubility. Plant Physiol. 80: 623–627.

Fasman, G. 1989. Protein conformational prediction. Trends in Biochem. 14: 295–299.

Finley, D. and Varshavsky, A. 1985. The ubiquitin system:functions and mechanisms. Trends Biochem. Sci. 10: 343–346.

Forde, B.G., Kreis, M., Williamson, M.S., Fry, R.P. and Pywell, J. 1985. Short tandem repeats shared by B- and C-hordein cDNAs suggest a common evolutionary origin for two groups of cereal storage protein genes. The EMBO J. 4: 9–15.

Goldberg, A.L., and St John, A.C. 1976. Intracellular protein degradation in mammalian and bacterial cells: part 2. Annu. Rev. of Biochem. 45:747–803.

Greenwood, J. S., and Chrispeels, M.J. 1985. Correct targeting of the bean storage protein phaseolin in the seeds of transformed tobacco. Plant Physiol. 79: 65–71.

Gross, D.S., and Garrard, W.T. 1987. Poising chromatin for transcription. Trends in Biochem. 12: 293–296.

Ho, S.P., and Degrado, W.F. 1987. Design of a 4–helix bundle protein: synthesis of peptides which self-associate into helical protein. J. Am. Chem. Soc. 109: 67516758.

Hoffmann, L.E., Donaldson, D.D., and Herman, E.M. 1988. A modified storage protein is synthesized. processed, and degraded in the seeds of transgenic plants. Plant Mol. Biol. 11: 717–729.

Hoffmann, L.E. Donaldsopn, D.D. Bookland, R. Rashka, K and Herman, E.M. 1987. Synthesis and protein body deposition of maize 15–kd zein in transgenic tobacco seeds. The EMBO J. 6: 3213–3221.

Hol, W, G., and Sander, H.C. 1981. Dipole of the α–helix and β–sheet: their role in protein folding. Nature. 294: 532–536.

Horsch, R.B., Fry, J., Hoffmann, N., Neidermeyer, J., Rogers, S.G. and Fraley, R.T. 1988. *In:* Plant Mol. Biol. Manual ed. S.B. Gelvin and R.A. Schilperoort, Dordrecht: Kluwer Academic.

Jaynes, J.M., Nagpala, P., Destefano, L., Denny, T., Clark, C., and Kim, J-H. 1990. Expression of a de novo designed peptide in transgenic tobacco plants confers enhanced resistance to *Pseudomonas solanacearum* infection. Plant Science: 89, 43–53.

Jaynes, J.M., Yang, M.S., Espinoza, N.O., and Dodds, J.H. 1986. Plant protein improvement by genetic engineering: use of synthetic genes. Trends in Biotechnol. 4: 314–320.

Jones, J.D.G., and Gilbert, D.E. 1987. T-DNA structure and gene expression in petunia plants transformed by *Agrobacterium tumefaciens* C58 derivatives. Mol. Gen. Genet. 207: 478–485.

Kabsch, W., and Sander, C. 1983. How good are predictions of protein structure? FEBS lett. 155: 179–182.

Kane, J.F. and Hartley, D.L. 1988. Formation of recombinant protein inclusion bodies in *Escherichia coli*. Trends in Biotechnol. 6: 95–101.

Kasarda, D.D., Okita, T.W., Bernardin, J.E., Baecker, P.A., and Nimmo, C.C. 1984. DNA and amino acid sequences of alpha and gamma gliadins. Proc. Natl. Acad. Sci. U.S.A. 81: 4712–4716.

Keris, M., Shewry, P.R., Forde, B.G., Forde, G. and Miflin, J. 1985. Structure and evolution of seed storage proteins and their genes with particular reference to thos of wheat, barley and rye. Oxford Survey of Plant Mol. and Cell Biol. 2: 253–317.

Komoriya, A., and Chaiken, J.M. 1982. Sequence modeling using semisynthetic Ribonuclease S.J. Biol. Chem. 257: 2599–2604.

Larkins, B.A. 1983. Genetic engineering of seed storage protein. *In:* Genetic Engineering of Plants B. A. Larkins (ed.), pp. 93–120. New York: Prenum.

Larkins, B.A., Pederson, K, Mark, M.D., and Wilson, D.R. 1984. The zein protein of maize endosperm. Trends in Biochem. 9: 306–308.

Lawrence, M.C., Suzuki, E., Varghes, J.N., Davis, P.C., Van Donkelaar, A. Tulloch, P.A. and Collman, P.M. 1990. The three-dimensional structure of the seed storage protein phaseolin at 3 Å resolution. The EMBO J. 9: 9–15.

Lear, J.D., Wasserman, Z.R. and Degrado, W.F. 1988. Synthetic apmphiphilic peptide model for protein ion channels. Science. 240: 1177–1181.

Lending, C.R., Kriz, A., Larkins, B.A. and Bracker, C.E. 1988. Structure of maize protein bodies and immunocytochemical localization of zeins. Protoplasma 143: 51–62.

Lycett, G.W., Cory, R.D., Shirsat, A.H., Richards, D.M., and Boulter, D. 1985. The 5'-flanking regions of three pea legumin genes: comparison of DNA sequences. Nucleic Acids Res. 13: 6733–6743.

Marqusee, S. and Baldwin, R. 1987. Helix stabilization by GLU-LYS salt bridges in short peptides of de novo design. Proc. Natl. Acad. Sci. U.S.A. 84: 8898–8902.

Marries, C., Gallois, P., Copley, J. and Keris, M. 1988. The 5' flanking region of a barley B hordein gene controls tissue and developmental specific CAT expression in tobacco plants. Plant Mol. Biol. 10: 359–366.

Mutter, M. 1988. NatureUs rules and chemistUs tools: a way for creating novel proteins. Trends in Biochem. 13: 260–264.

Pace, C.N. and Barret, A.J. 1984. Kinetics of tryptic hydrolysis of the arginine-valine bond in folded and unfolded ribonuclease T1. Biochem. J. 219: 411–417.

Pakula, A.A. and Sauer, R.T. 1986. Bacteriophage 1 Cro mutation: effect on activity and intracellular degradation. Proc. Natl. Acad. Sci. U.S.A. 82: 8829–8833.

Pakula, A.A. and Sauer, R.T. 1989. Amino acid substitutions that increase the thermal stability of the l Cro protein. Proteins. 5: 202–210.

Parasell, D.A. and Sauer, R.T. 1989. The structural stability of a

protein is an important determinant of its proteolytic susceptibility in *Escherichia coli*. J. of Biol. Chem. 264: 7590–7595.
Pederson, K., Agros, P., Naravana, S.V.L., and Larkins, B.A. 1986. Sequence analysis and characterization of a maize gene encoding a high-sulfur zein protein of Mw 15,000. J. Biol. Chem. 201: 6279–6284.
Pernollet, J.C. and Mosse, J. 1983. Structure and location of legume and cereal seed storage protein. Seed Proteins. Phytochemical Soc. of Eur. Sym. Series 20: 155–187.
Pontremoli, S. and Melloni, E. 1986. Extralysosomal protein degradation. Annu. Rev. Biochem. 55: 455–481.
Presnell, S.R., and Cohen, F.E. 1989. Topological distribution of a four–β–helix bundle. Proc. Natl. Acad. Sci. U.S.A. 86: 6592–6596.
Presta, L.G. and Rose, G.D. 1988. Helix signals in proteins. Science 240: 1632–1641.
Rafalski, J.A. Scheets, K., Metzler, M., and Peterson, D.M. 1984. Developmentally regulated plant genes: the nucleotide sequence of a wheat gliadin genomic clone. The EMBO J. 3: 1409–1415.
Richardson, J.S. and Richardson, D.C. 1988. Amino acid preferences for specific locations at the ends of α–helices. Science. 240:1648–1652.
Richardson, J.S. and Richarson, D.C. 1989. The de novo design of protein structures. Trends in Biochem. 14: 304–309.
Sanders, P.R., Winter, J.A., Barnason, A.R. and Rogers, S.G. 1987. Comparison of cauliflower mosaic virus 35S and nopaline synthetase promoters in transgenic plants. Nucleic Acids Res 15: 1543–1558.
Scheraga, H. 1978. Use of random copolymers to determine helix-coil stability constants of the naturally occuring amino acids. Pure. Appl. Chem. 50: 315–324.
Scheraga, H.A. 1985. Effect of side chain-backbone electrostatic interaction on the stability of α–helices. Proc. Natl. Acad. Sci. U.S.A. 82:5585–5587.
Scott, R.J., and Draper, J. 1987. Transformation of carrot tissue derived from proembryogenic suspension cells: a useful model system for gene expression studies in plants. Plant Mol. Biol. 8: 265–274.
Sengupta, G.C., Reichert, N.A., Baker, R.F., Hall, T.C. and Kemp, J.D. 1985. Developmentally regulated expression of the bean β–phaseolin gene in tobacco seed. Proc. Natl. Acad. Sci. U.S.A. 82: 3320–3324.
Shen, S-H. 1984. Multiple joined genes prevent product degradation in *E. coli*. Proc. Natl Acad. Sci. U.S.A. 81: 4627–4631.
Slightom, J.L., Sun, S.M., and Hall, T.C. 1983. Complete sequence of french bean storage protein gene: phaseolin. Proc. Natl. Acad. Sci. U.S.A. 80: 1897–1901.
Staswick, P.E. 1989. Preferential loss of an abundant storage protein from soybean pods during seed development. Plant Physiol. 90: 1251–1255.
Stockhaus, J., Eckes, P., Blau, A., Schell, J., and Willmitzer, L. 1987. Organ-specific and dosage-dependent expression of a leaf/stem specific gene from potato after tagging and transfer into potato and tobacco plants. Nucleic Acids Res. 15: 3479–3491.
Sueki, M., Lee, S., Power, S.P., Denton, J.B., Konishi, Y., and Scheraga, H. 1984. Helix-coil stability constants for the naturally occuring amino acids in water. Macromolecules. 17: 148–155.

Twell, D. and Ooms, G. 1987. The 5' flanking DNA of a patatin gene directs tuber specific expression of a chimeric gene in potato. Plant Mol. Biol. 9: 365–375.

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.

Wenzler, H.C., Mignery, G.A. Fisher, L.M., and Park, W.D. 1989. Analysis of a chimeric class I potatin-GUS gene in transgenic potato plants: high level expression of tubers and sucros-inducible expression in cultured leaf and stem explants. Plant Mol. Biol. 12: 41–50.

Yang, M.S., Espinoza, N.O., Dodds, J.H., and Jaynes, J.M. 1989. Expression of a synthetic gene for improved protein quality in transformed potato plants. Plant Science. 64: 99–111.

Zimm, B.H. and Bragg, J.R. 1959. Theory of the phase transition between helix and random coil in polypeptide chains. J. Chem. Phys. 31:526–535.