Order Number 9325062

Sulfur-rich 2S proteins in *Lecythidaceae* and their methionine-enriched forms in transgenic plants

Zuo, Wei-Neng, Ph.D.

University of Hawaii, 1993

U·M·I

300 N. Zeeb Rd.
Ann Arbor, MI 48106

SULFUR-RICH 2S PROTEINS IN LECYTHIDACEAE

AND THEIR METHIONINE-ENRICHED FORMS IN TRANSGENIC PLANTS

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAII IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

BOTANICAL SCIENCES
(PLANT PHYSIOLOGY)

MAY 1993

BY

WEI-NENG ZUO

Dissertation Committee:

Samuel S.M. Sun, Chairperson
H. Michael Harrington
Tom Humphreys
Morton Mandel
John I. Stiles

ACKNOWLEDGEMENTS

I would like to express my appreciation to my advisor, Dr. Samuel S.M. Sun, for his guidance, support and training during the development of this dissertation, and to the other members of my committee, Dr. H. Michael Harrington, Dr. Tom Humphreys, Dr. Morton Mandel, and Dr. John Stiles, for their valuable comments and help.

I would like to express my special appreciation to Dr. Helen Tu for her help, especially during the modification of the MP2S gene and the construction of the Phas/MP2S chimeric genes.

I would also like to thank Dr. Piera Sun for her help and advice and Ms. Kathleen Houtchens for her help in cloning MP2S cDNA. Also, thanks to the other members of Dr. Sun's laboratory.

Special thanks to my parents, Xiuqin Chen and Guangyuen Zuo, for their support.

I would like to express my special appreciation to my wife, Dr. Li-Xin Zhang, for her support, comments, help, patience, and understanding.

ABSTRACT

The 2S seed proteins from three members of the Brazil nut family (*Lecythidaceae*), monkeypot, cannonball, and Brazil nut, were isolated, characterized, and compared in regard to physical and chemical properties including subunit structure, amino acid composition, immunoreactivity, N-terminal sequence, and processing. These 2S proteins contain about 24% sulfur amino acids, thus they are sulfur-rich proteins (SRPs). The 2S SRPs are major seed proteins in the three plant species and all consist of two subunits, 9 kD and 3 kD, linked by disulfide bonds. In addition, these 2S SRPs share a high degree of similarity in subunit structure, amino acid composition, immunoreactivity, and N-terminal sequence. However, the precursor processing patterns of these 2S proteins are distinctly different. In monkeypot and Brazil nut, the precursor is cleaved into mature subunit polypeptides in three steps, *i.e.* 18 kD → 15 kD → 12 kD → 9 + 3 kD, while in cannonball, only two steps are detected, *i.e.* 18 kD → 15 kD → 9 + 3 kD.

A total of 9 cDNA clones encoding the 2S SRPs in monkeypot and cannonball have been isolated and characterized. DNA sequence analysis reveals that the 2S SRPs isolated from different members of the Brazil nut family share a high degree of homology in amino acid (>80%) and nucleotide (>90%) sequences. The methionine (Met) residues

are clustered in two areas of the variable region (between the 6th and 7th Cys residues) of the larger subunit. The 2S SRPs are encoded by multigene families and can be classified into two subfamilies.

Native monkeypot 2S SRP (MP2S) contains 16 Mol% Met. The variable region of the MP2S gene was modified through nucleotide sequence alternations to increase the protein's Met content. Eight Met-enriched MP2S genes were engineered to increase the Met varying amounts ranging from 18 to 24 Mol%. To test the effect of these modifications on the stability of the MP2S, chimeric genes containing coding sequences of three modified (19, 21, and 23% Met) and the wild type MP2S cDNA were transferred into tobacco plants via the *Agrobacterium* transformation system. Northern and Western blot analyses demonstrated that the genes for the modified MP2S were expressed in the transgenic tobacco seeds at levels comparable to that of the wild type MP2S gene. Both the wild type and modified MP2S proteins were correctly processed into mature subunits. These results suggest that the wild type as well as the modified MP2S genes are suitable candidates for use in protein quality improvement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AA	amino acid
Arg	arginine
Asn	asparagine
ATP	adenosine triphosphate
BN2S	sulfur-rich 2S seed protein from Brazil nut
bp	base pair
BSA	bovine serum albumin
CaMV	cauliflower mosaic virus
CB2S	sulfur-rich 2S seed protein from cannonball
CB2S-1	component 1 of CB2S
CB2S-2	component 2 of CB2S
cDNA	complementary DNA
cfu	colony forming unit
CIP	calf intestinal phosphatase
CPM	counts per minute
C-SRP	cysteine-rich sulfur-rich protein
Cys	cysteine
CTAB	cetyltrimethylammonium bromide
DAF	day after flowering
DEPC	diethylpyrocarbonate
DMF	N,N-dimethyl formamide
DMSO	dimethylsulfoxide
dNTP	deoxyribonucleotide triphosphate
DTT	dithiothreitol
E. coli	*Escherichia coli*
EDTA	ethylenedinitrilo-tetraacetic acid
ER	endoplasmic reticulum
Fig	figure
Gln	glutamine
Glx	glutamic acid or glutamine
GUS	β-glucuronidase
himet	modified phaseolin with a high content of methionine or DNA fragment a with high content of methionine
kb	kilobase
kD	kilodalton
LB	left border (of T-DNA in pBI121) or Luria-Bertani (medium)
Lys	lysine
β-Me	β-mercaptoethanol
Met	methionine
M-MLV	Moloney murine leukemia virus
MP2S	sulfur-rich 2S seed protein from monkeypot
mRNA	messenger RNA
M_r	relative molecular weight
MS	Murashige and Skoog (medium)
M-SRP	methionine-rich sulfur-rich protein
M.W.	molecular weight
NOS	nopaline synthase

NPTII	neomycin phosphotransferase II
NRB	nuclear resuspension buffer
NRS	negative regulatory sequence
NTP	ribonucleotide triphosphate
$O.D._{660}$	optical density at 660 nm
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PEG	polyethylene glycol
pfu	plaque forming unit
Phas/MP2S	chimeric genes consisting of phaseolin promoter and terminator and the coding regions of non-modified and modified MP2S cDNA
Phas/MPIa	chimeric gene consisting of the phaseolin promoter and terminator and the coding region of MP2S cDNA with Box I replaced with the DNA fragment Ia.
Phas/MPIb	chimeric gene consisting of the phaseolin promoter and terminator and the coding region of MP2S cDNA with Box I replaced with the DNA fragment Ib.
Phas/MPIIa	chimeric gene consisting of the phaseolin promoter and terminator and the coding region of MP2S cDNA with Box II replaced with the DNA fragment IIa.
Phas/MPIIb	chimeric gene consisting of the phaseolin promoter and terminator and the coding region of MP2S cDNA with Box II replaced with the DNA fragment IIb.
Phas/MPIIIaa	chimeric gene consisting of the phaseolin promoter and terminator and the coding region of MP2S cDNA with Boxes I and II replaced with the DNA fragments Ia and IIa, respectively.
Phas/MPIIIab	chimeric gene consisting of the phaseolin promoter and terminator and the coding region of MP2S cDNA with Boxes I and II replaced with the DNA fragments Ia and IIb, respectively.
Phas/MPIIIba	chimeric gene consisting of the phaseolin promoter and terminator and the coding region of MP2S cDNA with Boxes I and II replaced with the DNA fragments Ib and IIa, respectively.
Phas/MPIIIbb	chimeric gene consisting of the phaseolin promoter and terminator and the coding region of MP2S cDNA with Boxes I and II replaced with the DNA fragments Ib and IIb, respectively.
poly(A)	poly-adenylate
RB	right border of T-DNA
RER	rough endoplasmic reticulum
RT	room temperature
RNase	ribonuclease
SDS	sodium dodecyl sulfate
SRP	sulfur-rich protein

TCA	trichloroacetic acid
T-DNA	transfer DNA
Ti plasmid	Tumor-inducing plasmid
Tris	Tris[hydroxymethyl]aminomethane
Trp	tryptophane
UAS	upstream activating sequence
UTR	untranslated region
X-gluc	5-bromo-4-chloro-3-indolyl-β-D-glucuronic acid

CHAPTER I

LITERATURE REVIEW

INTRODUCTION

Plant proteins, especially seed proteins, are one of the major sources of dietary protein for both human beings and livestock. In some developing countries, people often depend heavily on plant proteins as the only or most important protein source (Aykroyd and Doughty, 1964). In developed countries, although plant proteins are not important as a direct source of dietary protein for human consumption, they are a primary source of protein for livestock. One disadvantage in using animal proteins is that the protein production cost is high. Another disadvantage is that animal proteins are often associated with animal lipids, including cholesterol which has been shown to be directly related to cardiovascular disease. Increasing the production and consumption of plant proteins will help resolve the problem of protein deficiency in certain underdeveloped and developing countries and help to reduce the over-reliance on animal proteins by people in developed countries.

Unfortunately, while plant proteins are a good and inexpensive source of dietary protein for human beings and livestock, they are deficient in certain essential amino acids (AAs). This deficiencies lower nutritional value. Generally, in cereals, the seed storage proteins are

deficient in lysine (Lys) and tryptophan (Trp), while in legumes, the seed storage proteins are low in the sulfur AAs, methionine (Met) and cysteine (Cys) (FAO, 1970 and Yamaguchi, 1980). Attention has been given to Met in the diet since it correlates with cancer prevention (Ghoshal and Farber, 1984). Evans and Gridley (1980) estimated that the sulfur AA level in beans should be increased two-fold in order to be nutritionally satisfactory.

SEED STORAGE PROTEINS

Seed storage proteins can be classified into four major classes according to their solubility: albumin, globulin, prolamin and glutelin, which are water, salt, aqueous alcohol, and acid or alkali soluble, respectively (Osborne, 1908), and into 11S, 7S, and 2S proteins according to their sedimentation coefficients. Seed storage proteins function as sources of nitrogen and carbon for the germinating seeds and growing seedlings. Therefore, they are abundant in seeds, contain high levels of arginine (Arg), glutamine (Gln) and asparagine (Asn), and degrade during seed germination.

The biosynthesis and accumulation of seed storage proteins is developmentally regulated, and their genes are expressed in an organ-specific manner. The expression of seed storage protein genes is regulated transcriptionally and post-transcriptionally (Walling *et al.*, 1986, Colot *et al.*, 1987 and Thompson *et al.*, 1989). All seed storage proteins

characterized thus far are encoded by multigene families (Higgins, 1984). For example, there are at least 10 prolamin families in maize (Wienend and Feix, 1980, Hagen and Rubenstein, 1981, Pedersen *et al.*, 1982, Viotti *et al.*, 1979, and Viotti *et al.*, 1982). The globulins of dicotyledonous plants are also the products of multigene families (Thomson and Schroeder, 1978, Casey, 1979, Moreira *et al.*, 1979, Schuler *et al.*, 1982a, Schuler *et al.*, 1982b, Brown *et al.*, 1981, and Dure and Chlan, 1981).

SYNTHESIS, PROCESSING, AND TARGETING OF SEED STORAGE PROTEINS

Seed storage proteins are synthesized on ribosomes attached to the rough endoplasmic reticulum (RER), transported through the Golgi apparatus and then deposited into small vacuoles which form protein bodies. The precursors of seed storage proteins contain a signal peptide (Higgins and Spencer, 1977, Burr *et al.*, 1978, Matthews and Miflin, 1980, and Bollini *et al.*, 1983), which facilitates the translocation of the seed storage protein into the lumen of the ER, the first step in intracellular transport.

Seed storage proteins are often synthesized as precursor molecules which undergo one to several steps of co- and/or post-translational processing into mature proteins (Crouch *et al.*, 1983 and Higgins, 1984). Apparently, one of the processing steps is the co-translational removal of the

signal peptides (Larkins and Hurkman, 1978, Burr and Burr, 1981, Higgins and Spencer, 1981, and Robert and Lord, 1981). Other processing steps may be required for the correct assembly of subunits, the formation of proper disulfide bonds (Steiner, 1977), or protein targeting (Bednarek *et al.*, 1990). Some proteins such as glycinin from soybeans (Sengupta *et al.*, 1981), globulin from oats (Brinegar and Peterson, 1982), and legumin from peas (Croy *et al.*, 1982) do not have extensive propeptide regions, other than a signal peptide, and thus, undergo only one processing step. Napin from oilseeds (Crouch *et al.*, 1983) and the 2S sulfur-rich protein (SRP) from Brazil nut (Sun *et al.*, 1987b) undergo several processing steps to form their mature proteins. The significance of separating the processing into several steps remains unknown. Besides the enzymatic cleavage processing, some seed storage proteins such as phaseolin and vicilin are co-translationally modified by glycosylation (Badennoch-Jones *et al.*,1981, Matthews *et al.*, 1981, and Bollini *et al.*, 1983).

Seed proteins are targeted into storage vacuoles called protein bodies by information present in the precursors. Results from experiments using glycosylation inhibitors or site-directed mutagenesis to destroy glycosylation sites have shown that sugar moieties are not necessary for targeting the protein into the storage vacuoles (Badennoch-Jones *et al.* 1981 and Wilkins *et al.*, 1990). On the other hand, some

studies indicated that certain sequences in the polypeptides contain information for the protein targeting (Bednarek *et al.*, 1990, Hunt and Chrispeel, 1991, Gonzalez, 1991, and Neuhaus *et al.*, 1991). The mechanism of the sorting and targeting of seed proteins is presently receiving much attention.

CHARACTERISTICS OF 2S PROTEINS

2S proteins were recognized as a class of important seed storage proteins only about 10 years ago (Youle and Huang, 1978 and Crouch and Sussex, 1981). Youle and Huang (1981) surveyed the oilseed proteins from several families of plants and found that these seeds all contain a certain amount of 2S protein. Generally speaking, 2S seed storage proteins consist of 2 subunits, a large and a small polypeptide, which are cleavage products of a precursor polypeptide (Crouch *et al.*, 1983, Altenbach *et al.*,1986, Ericson *et al.*, 1986, Higgins *et al.*, 1986, and de Castro *et al.*, 1987).

Another characteristic of the 2S proteins is that they consist of isoforms. The number of isoforms reported in the *Brassicaceae* family ranges from 2 to 5 (Monsalve and Rodriguez, 1990). Data from the nucleotide sequencing of cDNAs (Altenbach *et al.*, 1987, Altenbach *et al.*, 1992a, and de Castro *et al.*, 1987) and the AA sequencing of the proteins (Ampe *et al.*, 1986) suggests that at least 10 different 2S protein isoforms exist in the Brazil nut.

The synthesis of 2S seed storage proteins is regulated at both the transcriptional and translational levels (Krochko *et al.*, 1992). At the transcriptional level, the expression of the 2S protein is regulated by a seed-specific promoter. The promotor from the AT2S1 gene encoding the *Arabidopsis* 2S albumin can correctly direct the expression of this gene in the seeds of heterologous plants such as tobacco and *Brassica napus* (De Clercq *et al.*, 1990a and b).

The 2S proteins are Cys-rich. In oilseed plants, the Cys content of 2S seed proteins is about 8% (Youle and Huang, 1981). In Figure 1, the AA sequences of 2S seed proteins from 7 plant species are compared. It is interesting to note that the position and number of Cys residues (C_1-C_8) in these 2S proteins are highly conserved, as are the four leucine residues (L_1-L_4). In napin from *Brassica napus*, most of the Cys residues are involved in disulfide bridge formation (Lonnedahl and Janson, 1972), implying that the disulfide bridges constitute a major structural element of the 2S storage proteins. At the whole polypeptide level, as shown in Fig. 1, significant homologies exist between the 2S SRP from Brazil nut and cotton seeds (44%; Galau *et al., 1992*), sunflower (34%; Kortt, *et al.*, 1991), castor bean (44%; Sharief and Li, 1982), *Arabidopsis* (33%; Krebbers *et al.*, 1988), rapeseed (21%; Crouch *et al.*, 1983, Ericson *et al.*, 1986) and *lupinus* (26%; Gayler *et al.*, 1990).

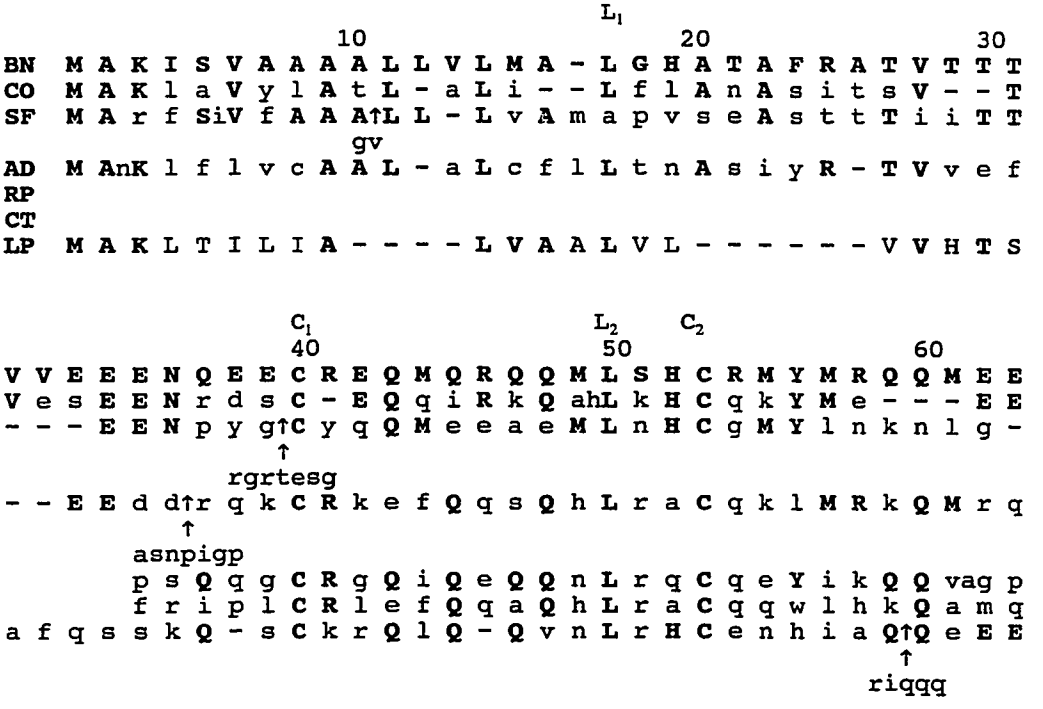
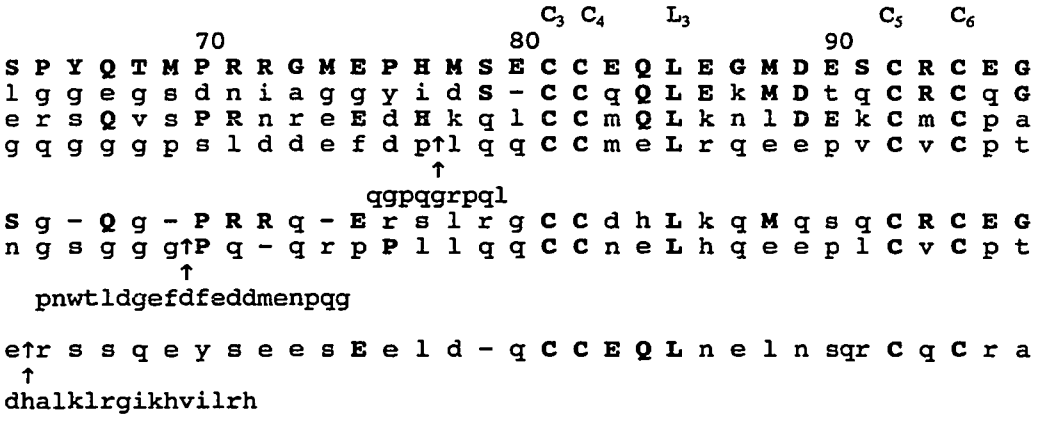
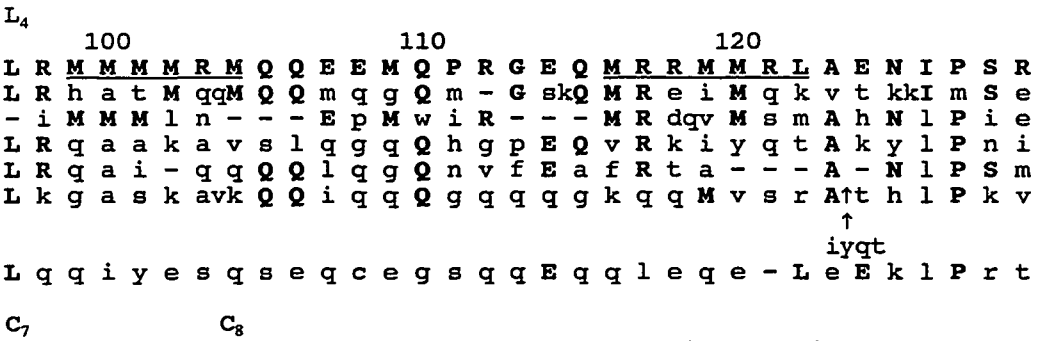
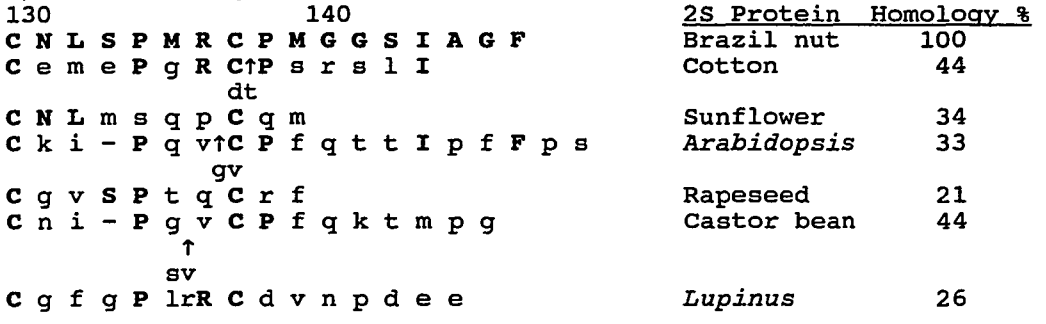


Fig. 1. Comparison of the AA sequences of 2S seed proteins from higher plants.

The one letter AA code is used for these comparisons and the sequences are aligned to maximize homology. Arrows indicate the displaced AAs for this alignment. The AA residues which are identical to those of the BN2S are capitalized and in boldface. The conserved 8 Cys and 4 leucine residues are indicated, C_1 to C_8 and L_1 to L_4, respectively. The percent homology to the BN2S is indicated at the end of the sequence. The Met-rich regions in the BN2S are underlined. BN: Brazil nut; CO: cotton; SF: sunflower; AD: *Arabidopsis*; RP: rapeseed; CT: castor bean; and LP: *lupinus*.







2S Protein	Homology %
Brazil nut	100
Cotton	44
Sunflower	34
Arabidopsis	33
Rapeseed	21
Castor bean	44
Lupinus	26

The 2S proteins constitute approximately 20 to 60% of total seed protein in various oilseed species (Youle and Huang, 1981), and are soluble in water and low salt solutions. These properties allow for their purification by either salt fractionation or dialysis. Both techniques are adaptable to bulk preparations. The abundance and water solubility of the 2S proteins make them good candidates for genetic improvement.

However, since the 2S proteins are small, a few alterations in their AA sequences may greatly affect their structure. Therefore, the site of the alteration and the physical property of the altered sequence may be important to the stability of the resulting 2S protein. By comparing the AA sequences of the 2S proteins from different plant species, the variable and conserved regions of the 2S proteins could be identified. Alternation of some AA residues in the conserved region *in vitro* and subsequent observation of the behavior of the modified protein *in vivo* is an approach to study the structural/functional relationship of a conserved region. On the other hand, the variable region of a 2S protein is more useful for genetic engineering purposes since the AA residues in the variable region may be replaced with nutritionally essential AAs, or since the variable region can be a site for the insertion of sequences encoding essential AAs or specific functional peptides. An *in vitro* assay has been developed to evaluate the effect of structural

modification on the assembly of soybean glycinin (Dickinson *et al.*, 1990). This type of *in vitro* assay system will facilitate protein engineering, since it will circumvent the need for time-consuming *in vivo* tests of engineered proteins.

SULFUR-RICH PROTEINS

Sulfur-rich proteins (SRPs) have been found in a variety of plants (Table 1). They include dicots such as Brazil nut (Ampe *et al.*, 1986, Sun *et al.*,1987, and de Castro *et al.*, 1987), peanut (Bashsa and Pancholy, 1981), pea (Higgins *et al.*, 1986), cotton (Galau *et al.*, 1992) and sunflower (Kortt and Cladwell, 1990) and monocots such as maize (Giannazza *et al.*, 1977, Melcher and Fraij, 1980, Pedersen *et al.*, 1986, and Kirihara *et al.*, 1988a), rice (Masumura *et al.*, 1989) and millet (Naren and Virupaksha, 1990). All these SRPs are small polypeptides with molecular weights ranging from 5 to 22 kD. The SRPs found in dicot plants are mostly albumin, while the SRPs from monocot plants are prolamin.

Based on their Met and Cys contents, the SRPs can be grouped into Met-rich SRP (M-SRP) and Cys-rich SRP (C-SRP). All the M-SRPs found so far are seed storage proteins with Met contents ranging from 10 to 22.5 %. The M-SRPs also contain certain levels of Cys (3.9 to 10 %), while the C-SRPs contain no or low levels of Met. As mentioned earlier, the 2S albumins from various oilseeds consistently contain about 8 % Cys. The highest Cys content was found in a pea 2S

Table 1 Sulfur-Rich Seed Proteins

Protein	Source	Met	Cys	Ref
Met-Rich				
12-kD albumin	Brazil nut	18.0	8.0	1
10-kD albumin	sunflower	16.0	8.0	2
12-kD Mat5-A	cotton	10.0	8.0	3
10.8-kD MRP	soybean	12.1	2.5	4
10-kD zein	maize	22.5	3.9	5
15-kD zein	maize	11.0	4.3	6
10-kD prolamin	rice	20.0	10.0	7
7.9-kD α-setarin	millet	12.7	4.8	8
9.1-kD β-setarin	millet	11.3	4.2	8
17-kD α-coixin	coix	11.6	5.2	9
Cys-Rich				
10.0-kD PA1	pea	0	23.7	10
5.0-kD β-purothionin	wheat	0	17.8	11
14-kD conglutin	lupin	0	7.8	12
2S albumin	oilseeds	1-2	8	13
Trypsin inhibitor	soybean	1.4	19.7	14
α-amylase inhibitor	wheat	2.5	9.0	15

Ref.: 1. Sun *et al.*, 1987a
2. Kortt *et al.*, 1991
3. Galau *et al.*, 1992
4. George and de Lumen, 1991
5. Kirihara *et al.*, 1988
6. Pedersen *et al.*, 1986
7. Masumura *et al.*, 1989
8. Ponnappanaren and Virupaksha, 1990
9. Leite *et al.*, 1992
10. Higgins *et al.*, 1986
11. Mak and Jones, 1976
12. Gayler *et al.*, 1990
13. Youle and Huang, 1981
14. Odani and Ikenaka, 1972
15. Kachlan and Richardson, 1981

albumin, PA1, with 23.7 % Cys (Higgins *et al.*, 1986).

The 2S SRP (BN2S) and its gene in Brazil nut seeds has been well characterized (Ampe *et al.*, 1986, Sun *et al.*,1987, and de Castro *et al.*, 1987). This 12-kD SRP consists of two subunits and constitutes approximately 30% of the total extractable protein in the Brazil nut seeds. The BN2S contains an exceptionally large amount of the sulfur AAs, 17.9% Met and 8.7 % Cys (Sun *et al.*, 1987a). cDNAs coding for the BN2S have been cloned and sequenced (Altenbach *et al.*, 1987 and Altenbach *et al.*, 1992). The large 9-kD subunit of this protein contains 77 AAs of which 14 are Met and 6 are Cys. Over half of the Met residues in the large subunit are clustered in two regions of the polypeptide where they are interspersed with arginine residues. In the first Met-rich region, Met residues account for 5 out of 6 AAs and four of these Met residues are contiguous. One isoform has a total of six Met residues in a row in this region (Ampe *et al.*, 1986). At least 10 isoforms of the BN2S protein have been observed (Ampe *et al.*, 1986, Altenbach *et al.*, 1987, and de Castro *et al.*, 1987). The BN2S genes contain an intervening sequence located in the coding region of the small subunit (Gander *et al.*, 1991).

In a 2S SRP from sunflower, 7 out of 16 Met residues are found in a region similar to the Met-rich region of the BN2S; other Met residues are scattered throughout the whole peptide. Besides the high level of Met (16%), this single

peptide 2S albumin also contains 8% Cys (Kortt *et al.*, 1991). A 2S SRP (Mat5-A) in cotton seeds contains 10% Met and 8% Cys (Galau *et al.*, 1992). The Mat5-A has a 40% homology with the BN2S. Again, most of the Met residues (7 out of 10) are located in the region similar to the Met-rich region in the BN2S. This information indicates that 2S sulfur-rich albumins contain a common Met-rich region. The water-solubility of albumins might be related to the clustering pattern of the hydrophobic Met residues in the protein.

In the sulfur-rich prolamins, no common Met-rich region can be identified. In the 15-kD zein, the Met residues are clustered in three regions covering approximately 39 AAs. Five of the 7 Cys residues are located in the first half of the protein molecule. In the 10-kD zein (Kirihara *et al.*, 1988a), 17 out of the 29 Met residues are found in the central region, and they are generally present as Met-Met doublets, separated by a spacer of 2 or 3 AA residues, one of which is proline. In the 10-kD sulfur-rich prolamin from rice (Masumura *et al.*, 1989), 22 Met residues are scattered throughout the polypeptide while 5 Met residues are clustered in an 8 AA residue stretch from position 108 to 115. In the α- and β-setarins, the sulfur-rich prolamins found in Italian millet (*Setaria italica* (L.) Beauv.) seeds, the Met residues, 12.7 and 11.3%, respectively, were randomly distributed throughout the polypeptide chains (Naren and Virupaksha, 1990).

Amongst these SRPs, the BN2S is the most abundant, accounting about 30% of the total seed protein. The 2S SRP from sunflower constitutes approximately 7% of the total seed protein, while the PA1 comprises 4.5% of total seed protein. Other sulfur-rich proteins constitute only a small percentage of their total seed proteins.

EVOLUTION OF THE SULFUR-RICH 2S PROTEINS IN BRAZIL NUT

The Brazil nut family (*Lecythidaceae*) is an important woody element of the lowland rainforest areas of the neotropics. The family is best known economically for the edible seeds of Brazil nut (*Bertholletia excelsa*) and monkeypot (or sapucaia nut, or Paradise nut, or cream nut, *Lecythis zabucajo*) (Prance and Mori, 1979). Recently, the exceptionally high sulfur AA content in Brazil nut seed has received much attention. These sulfur AAs in the seed proteins are bio-available (Antunes and Markakis, 1977, and Tao *et al.*, 1987). In addition to having a high level of the sulfur AAs, all other essential AAs are present in the Brazil nut seed proteins (FAO, 1970), indicating that the nutritional value of Brazil nut seeds is high.

Plant proteins usually contain about 1-2% Met (FAO, 1970); thus legume seed storage proteins as well as many other seed storage proteins are Met deficient. The extremely high level of the sulfur AAs, 1.3% by seed weight and 6% by weight of the total seed protein in Brazil nut, is unusual

(FAO, 1970, and Sun *et al.*, 1987a). The high level of the sulfur AAs present in the seeds of Brazil nut raises the question of what functions these 2S proteins play besides supplying nitrogen and energy as other seed storage proteins do. Some seed storage proteins have other biological functions, such as the insecticidal activity of arcelins (Osborn *et al.*, 1988), and the agglutination activity of lectins (Sequeira, 1978). Some 2S proteins are inhibitors of trypsin (Odani *et al.*, 1983) and α-amylase (Kachlan and Richardson, 1981). However, so far, no proteinase inhibitory activity has been found in the BN2S proteins (Ampe *et al.*, 1986).

Ohta and Kimura (1971) found that during natural molecular evolution, the percentage of Met codons in a gene tends to decrease. It is not economical for plants to store extra sulfur AAs in seed storage proteins which serve as a nitrogen source, since it takes a large amount of energy to assimilate inorganic sulfur into organic sulfur. The 2S SRP genes in the Brazil nut family may originate from either a remnant of a Met-rich protein gene or an acquisition of Met codons during evolution. In either case, the sulfur-rich seed storage proteins in the Brazil nut family must have some other special functions during seed development and germination and seedling growth to have persisted against the force of natural selection. Since the soil of the Amazon region where the Brazil nut family originated is sulfur

deficient (Sanchez *et al.*, 1982), it has been suggested that the seeds need to store a large amount of sulfur for seedling growth (Altenbach *et al.*, 1987). Met can be oxidized by peroxides (Stranger, 1984) and both Met and Cys can serve as donors in the detoxification of cyanide in animal bodies (Oke, 1980). Another putative function of the sulfur-rich proteins thus could be as an antioxidant in the seeds or seedlings of Brazil nut. At present, the reason(s) for the exceptionally high Met content in the seeds of Brazil nut remain unknown.

According to the sequence data on the isoform proteins (Ampe *et al.*, 1986) and the cDNAs (Altenbach *et al.*, 1987, Altenbach *et al.*, 1992 and de Castro *et al.*, 1987), the 2S SRPs in Brazil nut are encoded by a multi-gene family with at least 10 functional members. Figure 2 shows the AA sequences of the BN2S isoforms which have been characterized. The sequence homology among different isoforms is very high (88.3 to 97%). Variations in the number of Met residues in the two Met-rich regions are common. A cluster of six contiguous Met residues was observed in one of the isoforms (Ampe *et al.*, 1986). The cDNA sequence analysis indicates that the Met variations are due to a single base change in one codon, resulting in ATG⟷AGG transversion in most cases. Interestingly, the sequence data on the cDNA clones shows that the small subunit has more than one isoform, while the

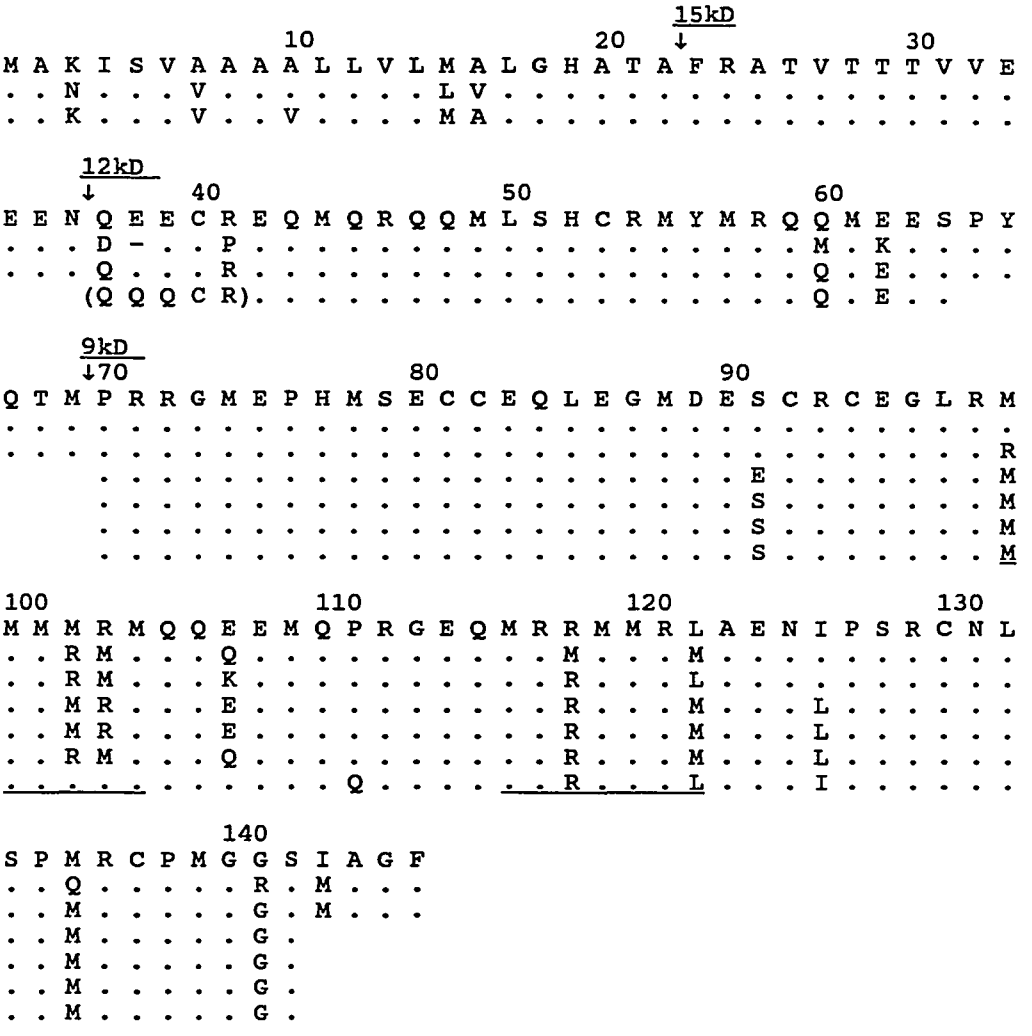


Fig. 2. Comparison of the AA sequences of the BN2S isoforms. The one letter AA code is used for these comparisons. The top three sequences are deduced from the cDNA clones (Altenbach *et al.*, 1991, and Altenbach *et al.*, 1992). The bottom three lines are from AA sequencing (Ampe *et al.*, 1986). Only one sequence of the small subunit was found in AA sequencing. Dots indicate identical residues. Sequence data within the parentheses is not certain. The approximate sites of cleavages involved in the maturation of the BN2S are marked with arrows. The two Met-rich regions are underlined.

AA sequencing analysis revealed only one small subunit isoform.

As discussed earlier, the 2S proteins from Brazil nut, cotton, sunflower, castor bean, *Arabidopsis*, rapeseed and *lupinus* are very similar in their subunit structure, processing, and AA sequences, although these plants are not closely related taxonomically and their sulfur-AA contents are quite different. This information suggests that these 2S seed proteins may have derived from a common ancestral gene.

Recently, a 2S storage protein has been identified in the ostrich fern, *Matteuccia struthiopteris*. This 2S protein, named matteuccin, is synthesized as a large precursor, which undergoes processing into two subunits with molecular weights of 8 kD and 3 kD. The available partial AA sequence of the matteuccin shows that it has a significant homology with some of the 2S seed storage proteins, including the BN2S, napin and SFA8 (Rodin and Rask, 1990). The nuclear DNA of the ostrich fern contains sequences having some homology with the napin cDNA (Templeman *et al.*, 1988). The similarity in the subunit structure, processing and AA sequence suggests that the matteuccin could be evolutionarily related to the 2S proteins of higher plants.

The regions in the 2S proteins with low AA homology are mainly located between the 6th and 7th Cys residues, suggesting that this region is not highly conserved in the 2S proteins. Mutations in this region probably will not have

any deleterious effect on the structure, stability and function of these proteins. Interestingly, the Met-rich region common in the BN2S, sunflower SRP and cotton 2S SRPs, is located in this region. This information could explain the evolution and existence of a large amount of Met in this region of the SRPs, but sheds no light on the possible reason of favoring/selecting these mutations for Met codons.

IMPROVEMENT OF SEED PROTEIN QUALITY

Classic breeding of superior crop varieties has played a great role in increasing the yield of crops for the past several decades (Tollenaa, 1989). In contrast, the progress in enhancing the protein quality of crops is very slow (Payne, 1983 and Bliss and Hall, 1977). A disadvantage has been found relating to the classic breeding for improved protein quantity in cereals: their overall nutritional quality decreases as their total protein content is increased. This is because an increase in the seed storage proteins will lead to a concomitant reduction in the amount of metabolic and structural proteins, which are relatively rich in the essential AAs.

Mutants rich in the essential AAs are important genetic materials when breeding techniques are applied to improve seed protein quality. A mutant line with 30% higher Met content was found in maize (Phillips *et al.*, 1981). This elevation of the Met content resulted from an increase in the

proportion of the Met-rich zeins of 10-kD and 14-kD and a decrease in the Met-poor zeins of 22-kD and 24-kD (Phillips and McClure, 1985). Studies on the utilization of this high-Met maize mutant are on the way.

Phaseolin is the major storage protein in the common bean. It contains the highest Met level amongst all of the seed storage proteins of common bean. It has been found that the amount of phaseolin present is positively correlated with the Met content of the seeds (Gepts and Bliss, 1984). Increasing the amount of phaseolin in the common bean should thus increase the Met content of the seeds. Genetic removal of the lectin (and arcelin) fraction and introduction of two other types of phaseolin in a common bean line resulted in higher phaseolin levels, however, the Met level of the whole seeds remained unchanged (Osborn and Bliss, 1985 and Romero *et al.*, 1986). Breeding techniques have been successfully used to increase the Lys content of maize seeds. A high Lys mutant, Opaque-2, was found in maize (Mertz *et al.*, 1964). Introduction of an Opaque-2 mutant to a maize line reduced the Lys-poor zein, resulting in an increase in the zeins with higher Lys content; consequently, the whole Lys content of the seeds increased. However, some other property changes in this high Lys mutant maize have limited its wider use (Bright and Shewry, 1983). These changes include a 10% lower yield and more mechanical damage due to the softer endosperm of the high Lys variety (Payne, 1983).

Several high-Lys mutants have been identified in barley. In these mutants, higher Lys contents were the result of a reduced accumulation of hordein, the low-Lys major storage protein in barley (Koie and Doll, 1979). Unfortunately, starch accumulation is also reduced in these high-Lys mutants, resulting in shrivelled grains and low crop yields.

In summary, a lack of desired mutants limits the use of classic methods to breed high nutritional crops.

STRATEGIES FOR INCREASING MET LEVELS IN SEED PROTEINS THROUGH BIOTECHNOLOGY

Recent developments in biotechnology provide new opportunities to improve agriculture. The ability to introduce foreign genes into plants represents "one of the most significant developments in a continuum of advances in agricultural technology" (Gasser and Frayley, 1989); and offers a promising strategy for breeding useful plants (Wettstein, 1989). In the area of protein quality, several strategies can be applied to improve the AA quality of seed storage proteins.

Over-expression of homologous genes. This is an approach which aims to increase the expression of a protein component rich in the essential AAs, or to decrease the expression of a component with a low level of the essential AA.

Transfer and expression of heterologous genes. In this approach a foreign gene encoding a protein rich in the essential AAs is introduced into a target crop genome. A gene encoding an essential AA-rich protein is essential in this approach. Such genes can be obtained by isolating natural genes encoding essential AA-rich proteins, or by *in vitro* sequence modification of a gene to alter its AA composition, or even by synthesizing a DNA molecule to code for a polypeptide with a high level of the essential AAs. An advantage of this approach is that through fusion with various regulatory promoters, chimeric genes can be generated, transferred into different crops, and expressed in specific tissues or organs according to their economic benefits. Generally speaking, all of the genes encoding the SRPs listed in Table 1 can be candidate genes for this approach.

Besides a coding sequence, a proper promoter is also critical for seed storage protein improvement through molecular approaches. For example, a monocot maize zein gene under the control of its own promoter could not express correctly in transgenic dicot petunia plants (Ueng *et al.*, 1988). However, a zein gene does express correctly in dicot tobacco plant if it is driven by a dicot promoter (Hoffman *et al.*, 1987 and Williamson *et al.*, 1988). On the other hand, since seed storage proteins are coded by multi-gene families, if being transferred back to the original plant, the

engineered gene should be controlled by a promoter which can express at a high rate so that its protein product can accumulate sufficiently to affect the overall protein composition of the seed. Finally, the promoter should be able to control temporal- and seed-specific expression.

EXPRESSION AND STABILITY OF HETEROLOGOUS PROTEINS IN TRANSGENIC PLANTS

If seed storage proteins were only polymerized AAs devoted to the young seedling's nutrition, a very high mutability could be expected. By comparing the AA sequences of seed storage proteins, it becomes clear that only some regions of the sequences are variable, while others are quite conserved. The conserved regions are probably important for protein targeting, assembly, stability, digestibility as well as for certain biophysical and biochemical characteristics of the seed storage proteins. Thus factors affecting the structure, processing, assembly, function, and stability of a target protein in transgenic plants should be considered, especially when a modified or synthetic gene is involved. The stability of a target protein in transgenic plants is extremely important. Pre-proteins should undergo proper processing and correct assembly in order to accumulate stably in the transgenic plants.

Development of amenable model systems for expressing foreign genes in transgenic tobacco, petunia and other plants

has facilitated the testing of protein targeting, translocation, glycosylation, assembly, and stability in transgenic plants (Fraley, 1986). An *in vitro* assay system to test the behaviors of heterologous or modified polypeptides during assembly is also available (Dickinson *et al.*, 1990).

Naturally occurring protein genes.

With the development of recombinant DNA and plant transformation and regeneration techniques, it is now possible to introduce genes from distant species into crop plants. Even animal protein genes have been expressed efficiently in transgenic tobacco plants (Hiatt *et al.*, 1989). In Table 2, the stability of various seed storage proteins in transgenic plants is summarized.

In a trial to elevate the Met content of plants, Altenbach *et al.* (1989) constructed a chimeric gene by attaching the phaseolin gene regulatory region to the coding sequence of the BN2S cDNA and transferred it into tobacco plants. This gene was expressed in the seeds of the transgenic plants and its product was processed correctly as it is in the Brazil nut, resulting in a 30% increase in the Met content of the transgenic tobacco seeds. When the same chimeric gene was transferred into canola, a similar result was observed (Altenbach *et al.*, 1992b).

Table 2. Stability of Storage Proteins in Transgenic Plants

Gene	Transgenic plant	Stability*	Source
Common bean			
Phaseolin	Sunflower tissue	±	Murai *et al.*,1983
	Tobacco seed	±	Sengupta-Gopalan *et al.*,1985
	Tobacco tissue	±	Chee *et al.*,1986
Lectin	Tobacco seed	+	Voelker *et al.*,1987
	Tobacco seed	+	Sturm *et al.*,1988
Soybean			
Conglycinin	Petunia seed	±	Beachy *et al.*,1985
	Petunia seed	±	Chen *et al.*,1986
	Petunia	+	Lawton *et al.*,1987
	Petunia seed	+	Bray *et al.*,1987
	Tobacco seed	+	Bray *et al.*,1987
Pea			
Legumin	Tobacco seed	+	Ellis *et al.*,1988
Brazil nut			
2S albumin	Tobacco seed	+	Altenbach *et al.*,1989
	Canola seed	+	Altenbach *et al.*,1992
	Potato leaf	+	Sun *et al.*,1992
	Potato stem	+	Sun *et al.*,1992
	Potato tuber	+	Tu and Sun***
Arabidopsis			
2S albumin	Tobacco seed	+	De Clercq *et al.*,1990
Potato			
Patatin	Tobacco leaf	+	Rosahl *et al.*,1987
	Tobacco stem	+	Rosahl *et al.*,1987
Maize			
Zein	Sunflower tumor	**	Goldsbrough *et al.*,1986
	Sunflower calli	**	Matzke *et al.*,1984
	Tobacco seed	+	Hoffman *et al.*,1987
	Petunia seed	±	Williamson *et al.*,1988
	Tobacco seed	-	Ohtani *et al.*,1991

* "+" indicates protein is stable; "-" unstable; "±" both stable protein and degraded smaller products, or less protein.
** There was no detectable zein.
*** Personal communication

Hoffman *et al.* (1987) placed the maize 15-kD sulfur-rich zein coding region under the control of the French bean β-phaseolin promotor and terminator and transferred this chimeric gene into tobacco plants. The transgenic plants synthesized the zein in a tissue specific manner during the later stage of seed development, and the zein was correctly processed and deposited into protein bodies. This result indicates that a monocot gene can be expressed in dicot plants when driven by a dicot promotor.

Many genes coding for seed storage proteins have been be transferred and expressed in transgenic plants. Generally, the heterologous seed proteins can stably accumulate in transgenic plants as shown in Table 2. However, in a few cases, degradation of the introduced proteins in transgenic plants was observed.

A phaseolin gene was expressed properly in the transformed sunflower calli (Murai *et al.*, 1983). However, in addition to a protein of approximately the same molecular weight as the authentic phaseolin (48 kD), polypeptides of 28, 26 and 14 kD were also detected by Western blot analysis. After incubating exogenous phaseolin with sunflower callus tissues or its extracts, a 14-kD product was detected. Both *in vivo* and *in vitro* experiments revealed that specific degradation of the phaseolin occurred in the transformed sunflower calli. When this gene was transferred into a tobacco plant, authentic phaseolin was present in the

developing seeds of the transgenic tobacco plants, and the phaseolin was found glycosylated as it is in bean seeds. Again, small immunoreactive polypeptides of 25, 26, 27 and 29 kD were detected. Since in the nonmembrane system of the sunflower extract the small piece of phaseolin could be detected, the authors concluded that the degradation of phaseolin in transgenic sunflower callus tissue is not due to improper processing or compartmentalization of the foreign protein in the undifferentiated sunflower callus, but due to some other factors which affect the stability of the heterologous phaseolin (Sengupta-Gopalan *et al.*, 1985).

Using the phaseolin gene and its cDNA counterpart, Chee *et al.* (1986) constructed a mutant phaseolin gene lacking the five introns but retaining its natural 5' and 3' regulatory sequences. This mutated gene or "minigene" was transferred into tobacco. The authentic polypeptide of phaseolin was detected immunologically in the transformed tobacco, but some larger (90 kD) and smaller (24 kD) polypeptides were also detected. The authors suggested that the larger polypeptide could be an aggregate of phaseolin, while the smaller immunoreactive polypeptide were the products of phaseolin degradation. This study also indicated that, at least in the case of phaseolin, introns are not necessary for gene transcription and pre-mRNA processing.

The expression of the soybean α-conglycinin gene in petunias resulted in the accumulation of an immunoreactive

protein of the expected size (76 kD) in the transgenic seeds (Beachy *et al.*, 1985). Sedimentation analysis of the extracted proteins from the transgenic petunia seeds indicated that the α-conglycinin protein was properly assembled into the 7S complex of three subunits. However, in addition to the 76-kD polypeptide, some smaller polypeptides of 68, 64, 55, 20 and 18 kD were also detected antigenically in the seeds. When the soybean β-conglycinin gene was expressed in tobacco and petunia (Bray *et al.*, 1985), multiple isoelectric forms of the β-conglycinin were stably accumulated in the transgenic plants, but the level of the protein accumulated in the mature seeds was not correlated with the level of its mRNA. Thus some improper cleavages or degradations must have occurred during translation and/or posttranslation.

Modified and synthetic genes.

Recent developments in site-directed mutagenesis, DNA synthesis, and DNA recombination technology facilitate the alteration of a DNA sequence, the insertion of a fragment of DNA into a gene, and even the synthesis of a designed DNA molecule. Therefore, a modified or novel gene can be obtained.

Some modified genes have been transferred into plants (Table 3). Hoffman *et al.* (1988) inserted a 45-bp synthetic DNA duplex into a β-phaseolin gene. This alteration

Table 3. Stability of Modified Proteins in Transgenic Plants

Gene	Transformed plant	Stability*	Source
Common bean			
Phaseolin	Tobacco seed	-	Hoffman *et al.*,1988
Phytohemag-glutinin	Tobacco seed	±	Voelker *et al.*,1989
Arabidopsis			
2S albumin	*Arabidopsis* seed	+	Vandekerckhove *et al.*, 1989
	Rapeseed seed	+	Vandekerckhove *et al.*, 1989
	Arabidopsis seed	+	De Clercq *et al.*, 1990
	Rapeseed seed	+	De Clercq *et al.*, 1990
	Tobacco seed	+	De Clercq *et al.*, 1990
Maize			
zein	Tobacco seed	-	Ohtani *et al.*,1991
Synthetic			
gene	Potato tuber	-	Yang *et al.*,1989

* "+" indicates that the protein is stable, "-" unstable, "±" both stable proteins and degraded smaller products, or less proteins.

increased the number of Met codons in the phaseolin gene from three to nine. They transferred this modified phaseolin gene (himet) into tobacco plants and found that the himet-phaseolin gene was expressed in a temporal- and organ-specific fashion. At the mRNA level, the mutant gene was expressed at a level similar to that of the unmodified normal phaseolin gene. However, considerably less phaseolin polypeptide (0.2% of the unmodified phaseolin) accumulated in the seeds of the himet expressing plants as compared with the plants expressing normal phaseolin. The himet-phaseolin was synthesized at a rate equivalent to that of normal phaseolin, and it was processed in a similar manner, but the himet phaseolin was turned over in the Golgi secretion vesicles or protein bodies. Although the predicted secondary structure of the inserted peptide region was α-helical, matching the structure of the peptide surrounding the insertion site, the modified protein still was not accumulated in the tobacco seeds. This was probably because the insertion site was located in a region conserved amongst members of the phaseolin gene family (Slightom *et al.*,1985). Later study indicated that the insertion site was located in an important region of a major structural element of the phaseolin trimer. The inclusion of 15 AA residues at this site could distort the protein structure at the tertiary and/or quaternary level (Lawrence *et al.*, 1990) resulting in the instability and degradation of the protein.

Yang *et al.* (1989) synthesized a novel gene encoding a polypeptide with 80% essential AA residues, i.e. Lys and Met, and transferred it into potato plants. The accumulation of this engineered polypeptide in bacteria indicated that the chemical structure of the polypeptide was stable. However, when this novel gene was expressed in the transformed potato plants, a low level of accumulation of the polypeptide in potatoes was obtained. The authors concluded that degradation of the translational product was the reason for the low level of polypeptide accumulation, since a higher level of its mRNA was detected in the transgenic plants.

Voelker *et al.* (1989) modified a phytohemagglutinin gene and expressed the engineered gene in tobacco plants. They found that when the glycosylation signal was disrupted by site-directed mutagenesis, the attachment of asparagine-linked glycans to the polypeptide was inhibited. This modification did not have any effect on targeting the polypeptide to protein bodies but the stability of the protein decreased, resulting in a lower protein accumulation.

A question that arises in protein engineering is where in the protein and what kind of AA sequence modification will not adversely affect the stability of the protein *in vivo*? Since the variable regions of a protein might be of less biological importance, they could be potential sites for modification (Doyle *et al.*, 1986). The variable regions can be identified by comparing the AA sequences of different

isoforms of the proteins or homologous proteins from different plants of the same family. The altered AA sequence which matches the physical and chemical properties of the original AA sequence in the insertion region would be considered most likely to work. For synthetic genes, it is even harder to predict the fate of the totally new products in host plants.

Based on the fact that the region between the 6th and 7th Cys residues in the BN2S is highly variable in both length and AA composition (Ampe *et al.*, 1986), Vandekerckhove *et al.* (1989) mutated this region in the *Arabidopsis thaliana* 2S gene (AT2S1) by site directed mutagenesis. Thus six AA residues in this region were changed into a sequence encoding enkephalin, a pentapeptide (YGGFL) displaying opiate activity, flanked by tryptic cleavage sites. This modified gene was transferred back into *Arabidopsis* and into rapeseed. The 2S albumin was isolated from the seeds of the transformed plants and treated with trypsin to recover the enkephalin. A yield of enkephalin of up to 200 nmol per gram of seed was achieved, indicating that the modified protein was stably accumulated in the seeds.

De Clercq *et al.* (1990) inserted some Met-rich sequences into the variable regions between the 6th and 7th Cys residues of the AT2S. The Met-enriched AT2S genes were expressed and the proteins were accumulated in transgenic tobacco, rapeseed, and *Arabidopsis*.

Based on the AA sequences of several 11S globulins (Wright, 1988) and X-ray analysis of the crystals of 11S globulins of soybean and other seed storage proteins, Utsumi (1989) found that some regions of the soybean 11S globulin were variable. Therefore, it appears possible to modify these variable regions so that the engineered gene would code for a protein with a higher content of the essential AAs such as Met and Cys. Since the variable regions of 11S globulin are hydrophilic, but Met is hydrophobic, Utsumi proposed to insert some hydrophilic AAs together with Met in the same site to balance Met's hydrophobicity.

The examples above indicate that inappropriate modifications may have some unknown effects on the biophysical and biochemical properties of the protein, which could result in a failure to accumulate of the modified protein in transgenic plants. Modifications in less conserved regions, such as the variable region of the AT2S and BN2S, would be less likely to affect the protein's stability in plants. From this point of view, the use of naturally occurring proteins has some advantages.

Failure to express heterologous genes in plants.

Although some genes from very distant taxonomical relatives, and even from animals, have been successfully expressed in transgenic plants (Spena and Schell, 1987), failure to express heterologous protein genes in transgenic

plants has been reported (Frayley and Papahadjopoulas,1982 and An, 1986). This lack of expression was due to events at both the transcriptional and post-transcriptional level.

When a monocot gene encoding a 19-kD zein protein was introduced into dicot petunia plants, a very small amount of mRNA was synthesized and no zein could be detected (Ueng *et al.*, 1988). When the coding region of the zein, however, was driven by a dicot phaseolin gene promoter, a high level of the mRNA could be detected, but, a very low level of the zein could be found in the transgenic petunia plants (Williamson *et al.*, 1988). These results indicate that the promotor of the 19-kD monocot zein gene lacks some specific sequences required for efficient transcription; the mRNA might lack some sequences for efficient translation; or the precursor of the zein might lack the information for correct protein targeting, transport, and assembly; and thus, result in the degradation of the zein. An alterative explanation might be that the gene and its products, mRNA and precursor, of the zein were not readily recognized by the host transcriptional, translational, and targeting machinery.

Failure to express a transferred gene in plants also has been observed at the RNA processing level. Odell *et al.* (1985) and Hunt *et al.* (1987) found that the mRNAs of transferred genes were not correctly polyadenylated.

Inhibition of the expression of heterologous genes at the translational level has also been reported. Matzke *et*

al. (1984) transferred a 23-kD zein gene (Z4) into sunflowers. The mRNA of Z4 gene isolated from the transformed sunflower calli could be translated in a wheat germ translation system to yield an immunoprecipitable polypeptide with the expected molecular weight. However, the zein protein or its degraded product could not be found in the transformed sunflower tissue. Similar results were observed when the 19- and 15-kD zein genes were expressed in sunflower tumors (Goldsbrough *et al.*, 1986). The mRNA of a nopaline synthetase-human growth hormone chimeric gene transferred into tobacco and sunflower callus tissues was not translated in transformed tissues (Barta *et al.*,1986). Failure of the translation machinery of dicot plants such as tobacco and sunflower to recognize the monocot or animal mRNA may result in the inhibition of expression of the heterologous genes.

Interestingly, although a monocot zein mRNA could not be translated in transgenic dicot sunflowers, it could be translated in transgenic tobacco (Hoffman *et al.*, 1987), suggesting that the translational machinery may differ in different plants.

Processing and targeting of heterologous polypeptides in transgenic plants.

Although some animal pre-proteins were not correctly processed in transgenic plants (Koncz *et al.*, 1984 and Van

Santen and Spritz, 1987), precursors for a number of seed storage proteins, including those of monocots, can be processed correctly in the seeds of transgenic tobacco or petunia (Altenbach *et al.*, 1989 and De Clercq *et al.*, 1990), indicating that there is a commonality between species in the processing of seed storage proteins. These observations increase the likelihood that the complement of seed storage proteins in other plants might be manipulated using a gene transfer approach.

In transgenic plants, heterologous storage proteins have been observed to be targeted to the correct subcellular compartments. When the phaseolin gene was expressed in tobacco plants, the phaseolin was restricted solely to the protein bodies. The subcellular location of the phaseolin in the seeds of the transformed tobacco plants was similar to that in common beans (Greenwood and Chrispeels, 1985). The AT2S protein of *Arabidopsis* was also targeted into protein bodies in transgenic tobacco seeds (De Clercq *et al.*, 1990).

Fate of heterologous proteins during germination.

Heterologous seed proteins disappeared in germinating transgenic seeds (Sengupta-Gopalan *et al.*, 1985, Voelker *et al.*, 1987, and Altenbach *et al.*, 1989). This indicates that heterologous seed storage proteins function as storage proteins in transgenic plants. In the case of prolamin, however, the accumulated zein did not disappear in the

transformed tobacco seeds during germination. This is probably because proteinases in the tobacco seeds could not efficiently degrade the monocot prolamin, or the physiological condition in the tobacco seeds rendered prolamin insoluble and thus less susceptible to proteolytic degradation.

Summary of expression of heterologous protein genes in transgenic plants.

The above discussion shows that seed storage protein genes, or chimeric genes regulated by the promotor of seed storage protein genes, can be transcribed in transgenic plants in a seed-specific manner. The gene products can be properly processed, glycosylated, targeted and assembled. Generally speaking, molecular transformation is a plausible way to introduce a new gene into plants. This strategy makes it possible to introduce a desirable protein gene into plants for protein quality improvement.

OTHER CONSIDERATIONS IN THE MODIFICATION OF SEED PROTEINS

The AA sequence of a polypeptide determines its primary, secondary and tertiary structure. Therefore, when a protein is to be modified, the effect of modification on its structure, solubility, precursor processing, subunit assembly, intercellular transportation, and degradation, in addition to protein stability, should be considered.

Host plants accepting the genes are also important. A general consideration is that the host plants should contain the necessary *trans*-acting factors and machinery for the expression of the transgene. This problem can be avoided if a proper promoter is used. Amino acid pools could be another important factor limiting the expression of transgenes, especially for those genes encoding proteins particular rich in certain AAs. For example, the Met pool in a host plant might be an important factor limiting the expression of the Met-rich protein gene. The amino acid codon bias of the host plant should also be considered when a target gene is sequence modified.

CHAPTER II. DEVELOPMENT OF HYPOTHESIS

The following conclusions can be drawn from the literature reviewed: 1) The BN2S and some other 2S seed proteins in higher plants are structurally and evolutionarily related; 2) the M-SRPs from Brazil nut, pea, and sunflower share a common Met-rich region that is located in a variable region of the 2S proteins; 3) the introduction of the Brazil nut SRP gene into tobacco can enhance the Met content of the transgenic tobacco seeds; and 4) sequence alterations in the variable region of the 2S proteins are tolerated.

On the basis of the above information, it is proposed to test the hypotheses:

1). The 2S seed proteins from plants in the Brazil nut family are highly related and rich in the sulfur AAs. 2). The sulfur-rich 2S proteins modified for additional Met residues in the variable region can stably accumulate in the transgenic seeds.

CHAPTER III

MOLECULAR CHARACTERIZATION OF THE SULFUR-RICH 2S PROTEINS AND CLONING OF THEIR GENES FROM LECYTHIDACEAE

INTRODUCTION

Among all the SRPs identified so far, only the 2S SRP in Brazil nut is a major seed storage protein. The large amount of the SRP contributes to the high level of Met in Brazil nut seeds.

The gene encoding the Brazil nut sulfur-rich 2S (BN2S) protein is a good candidate for increasing the Met content of transgenic plants through molecular transformation (Altenbach *et al*, 1989, Altenbach *et al*, 1992, and Sun *et al*, 1992). Little is known, however, about the origin of the sulfur-rich protein genes and the reason for the high Met content. Sequence analyses/comparisons of the 2S proteins from diverse plant species suggest that they might share a common ancestral gene (Chapter I). Therefore, it is of interest to get information on the 2S proteins and their genes in plants that are taxonomically closely related to the Brazil nut.

MATERIALS AND METHODS

Chemicals. All chemicals used (Fisher Scientific or otherwise noted) were reagent grade or better.

Experimental plants. Three species in the Brazil nut family (*Lecythidaceae*), including Monkeypot (*Lecythis zabucajo*), Cannonball *(Couroupita quianensis*) and Brazil nut (*Bertholletia excelsa*), were studied. Monkeypot seeds and leaves were collected from the Hilo Nursery Arboretum, Hilo, Hawaii. Cannonball seeds and leaves were collected from the Foster Botanic Garden, Honolulu, Hawaii. The seeds and leaves were harvested and used directly or frozen at -80°C until use. Brazil nut seeds were purchased from Orinda Nuts, California.

Protein isolation. Total protein was directly extracted from the fresh mature embryos with the salt buffer previously described by Sun *et al* (1987). The 2S protein was fractionated from the total protein in a 5 to 30% sucrose gradient (Youle and Huang, 1981).

Amino acid analysis. The acid hydrolyzate of 2S protein was prepared by hydrolysis in constant boiling 5.7N HCl including 2% phenol and 0.01% β-mercaptoethanol in evacuat for 24 hr at 110°C and analyzed in a Beckman 6300 AA amino acid analyzer. The performic acid oxidation procedure was used to determinate the sulfur AAs, (Hirs, 1967).

Amino acid sequence determination. The 2S protein was sequenced using an Applied Biosystem model 477A liquid pulse sequencer equipped with a Model 120 on-line PTH analyzer following the standard protocol of the vender.

SDS polyacrylamide gel electrophoresis (SDS-PAGE). 20% SDS-PAGE of the seed proteins was performed essentially as described by Laemmli (1970). In some experiments, the seed protein samples were treated without β-mercaptoethanol (β-Me) as indicated in the figure legends.

Western blots. Total protein and 2S proteins from the seeds of Brazil nut, monkeypot and cannonball were separated by SDS-PAGE, electrophoretically transferred onto nitrocellulose membranes, and then immunoreacted with monoclonal antibodies specific for the 9-kD subunits of Brazil nut and cannonball 2S protein, respectively. Immunodetection of the 2S protein was performed with the Bio-Rad Immuno-Blot^{R} Goat-Anti- Mouse-Alkaline-Phosphatase (GAM-AP) Conjugate system.

RNA isolation. Total RNA was extracted from the developing embryos and total poly(A)$^{+}$ RNA was prepared by the method of Hall *et al* (1978).

Northern blots. Total RNA (6 μg) or poly (A)$^{+}$ RNA (1 μg) from developing monkeypot and cannonball seeds was separated on a 1.5% agarose/formaldehyde gel. Transfer to NYTRAN nylon membrane, hybridization, and probe removal were performed using the protocols of Schleicher & Schuell (Keene, NH). The *Hinf*I fragment of a pHS-3 cDNA clone containing the BN2S coding sequence was labelled with [α-^{32}P]-dCTP using a nick translation kit [Bethesda Research Laboratories, Life Technologies, Inc (BRL)] or by random priming (Feinberg and

Vogelstein, 1983) and used as probe. Hybridization was carried out in a solution containing 50% formamide, 5X Denhardt's solution, 0.1% SDS, 100 μg/ml sonicated denatured salmon sperm DNA and 5X SSPE at 42°C overnight. The filters were washed once with 2X SSC containing 0.1% SDS at room temperature (RT) for 15 min, once with 0.1X SSC containing 0.1% SDS at RT for 15 min, once with 0.1X SSC containing 0.1% SDS at 65°C for 30 min, and then exposed to Kodak XAR-5 X-ray film at -80°C.

In vitro **translation.** A wheat germ system was used to synthesize polypeptides *in vitro* as described by Sun *et al* (1975). Each translation reaction contained 8 μg of total RNA or 0.4 μg poly $(A)^+$ RNA from developing monkeypot and cannonball seeds. [^{35}S]-Met (ICN Biochemicals, Inc.) was used to label the newly synthesized polypeptides. The incorporation of [^{35}S]-Met into the translational products was quantitated by trichloroacetic acid precipitation and scintillation counting. The reaction mixture was analyzed by 20% SDS-PAGE and autoradiography.

In vivo **pulse-chase labelling experiments.** Freshly-collected developing monkeypot embryos were cut into pieces of about 2mm-thick and were then incubated with 25 μCi [^{35}S]-Met (ICN Biochemicals, Inc.) at RT for 1 hr. After the incubation, the samples were washed 3 times with water to remove the unincorporated [^{35}S]-Met and then incubated with 200μl 10 mM unlabelled Met at 30°C for specific lengths of

time. At each time point, one piece of the embryo was collected and its total protein was extracted with the salt buffer as previously described by Sun *et al* (1987). The same pulse-chase labelling procedure was used for cannonball except that the whole cannonball embryo was used. The incorporation of [^{35}S]-Met into the translational products was measured by trichloroacetic acid precipitation and scintillation counting. Equal amount of radioactive protein counts were analyzed by 20% SDS-PAGE and autoradiography.

DNA isolation. Genomic DNA was isolated from cannonball seeds and leaves, and from monkeypot leaves. Twenty grams of frozen seeds or leaves were ground into a powder using a mortar and pestle in the presence of liquid nitrogen. The powder was then extracted with 150 ml of extraction buffer (15% sucrose, 50 mM Tris/HCl, 50 mM EDTA and 0.25 M NaCl, pH 8.0) and centrifuged at 4,000 rpm for 15 min to collect the crude nuclear pellet. The nuclear pellet was resuspended in 30 ml buffer (15% sucrose, 50 mM Tris/HCl (pH 8.0) and 50 mM EDTA) and 3.2 ml of 10% sarkosyl was added to release the genomic DNA from the nuclei. The nuclear debris was removed by centrifuging at 20,000g for 20 min. After CsCl centrifugation and ethanol precipitation, the DNA pellet was dissolved in TE buffer (pH 7.5).

Southern blots. Genomic DNA (10 μg) from monkeypot, cannonball and Brazil nut was digested with *Eco*RI (Promega). The digestion products were separated on a 0.7% agarose/TBE

(89 mM Tris-borate, 89 mM Boric acid and 20 mM EDTA) gel containing 0.5 mg/ml EtBr. Transfer and hybridization were performed following the protocols of Schleicher & Schuell. The *Hinf*I fragment of the pHS-3 cDNA clone containing the BN2S coding sequence was labelled with [α-^{32}P]-dCTP by nick translation using a Bethesda Research Laboratories kit (BRL) or by random priming (Feinberg and Vogelstein, 1983) and used as probe. Hybridization was carried out in a solution containing 5X Denhardt's solution, 0.5% SDS, 10 mM EDTA, 100 μg/ml sonicated denatured salmon sperm DNA and 6X SSC at 65°C overnight. The filters were washed once with 2X SSC/0.5% SDS once at RT 15 min, 0.1X SSC/0.1% SDS once at RT for 15 min, 0.1X SSC/0.5% SDS once at 65°C for 30 min, and than exposed to Kodak XAR-5 X-ray film at -80°C.

cDNA library construction. Two methods were used to construct cDNA libraries.

A vector-primer cDNA cloning system developed by Alexander *et al* (1984) was used to construct a cannonball seed cDNA library. In this library, the cloned cDNAs could easily be released by digesting the plasmid DNA with restriction endonuclease *Xba*I.

The cDNAs for cannonball and monkeypot were also synthesized with the cDNA Synthesis System (BRL). After second strand DNA synthesis, *Eco*RI linker was added to both ends of the double stranded cDNA and then cloned into the bacteriophage λgt11 vector (BRL).

The *Hinf*I fragment containing the Brazil nut SRP coding sequence from the cDNA pHS-3 was used as a probe to screen the libraries. All of the positive clones, either from the vector-primer cloning or from the λgt11 cloning library, were subcloned into the pVZ-1 (derived from Stratagene's pBluescript but with a modified polylinker) or pBluescript SK+ (Strategene) vector for further study. The positive cDNA clones encoding the cannonball or monkeypot 2S protein were named pCB2S and pMP2S, respectively.

DNA sequencing. Dideoxy-sequencing was conducted using the Sequenase Version 2.0 kit and protocol (United States Biochemical). Both dGTP and dITP reactions were applied to the same DNA templates. Double stranded plasmid DNA was prepared using the boiling method (Sambrook *et al*, 1989). The DNA was polyethylene glycol (PEG) precipitated, denatured with NaOH and used as a template for sequencing. Both T3 and T7 promotor primers (Promega) and two internal primers based on the cannonball 2S protein sequence (primer 1: 5'-CTGCTGTCTCTCCATCTGCT-3' and primer 2: 5'-GGGACTGAGGTTGCAGCGGG-3') were used for sequencing.

RESULTS

2S proteins in the Brazil nut family.

Protein profiles The total and 2S seed proteins were isolated from 3 members of the Brazil nut family, namely Brazil nut, cannonball, and monkeypot. These 2S seed proteins are designated BN2S, CB2S and MP2S, respectively. Figure 3 shows the SDS-PAGE profiles of the total and 2S proteins from Brazil nut, cannonball and monkeypot. Based on the densitometric scanning of the stained SDS-PAGE gels, the BN2S, CB2S and MP2S represent about 30%, 50% and 15% of the total extractable seed proteins, respectively. All the BN2S, MP2S and CB2S consist of two polypeptides with molecular weights of about 9 kD and 3 kD.

Subunit structure Further analysis on the subunit structure of the 2S proteins revealed a more complex profile. As shown in Fig. 4 lane 2, when CB2S was separated in a SDS-polyacrylamide gel without β-mercaptoethanol (β-Me) treatment, two bands were detected: a major component (CB2S-1) which migrated slower and a minor component (CB2S-2) which migrated faster. However, when treated with β-Me, three bands were detected in the SDS-PAGE gel, two about 9 kD and one about 3 kD (Fig. 4, lane 3). If the protein sample was first treated with β-Me and then dialyzed to remove the β-Me, the two components were again observed (Fig. 4, lane 1). When the CB2S-1 and CB2S-2 bands were individually isolated

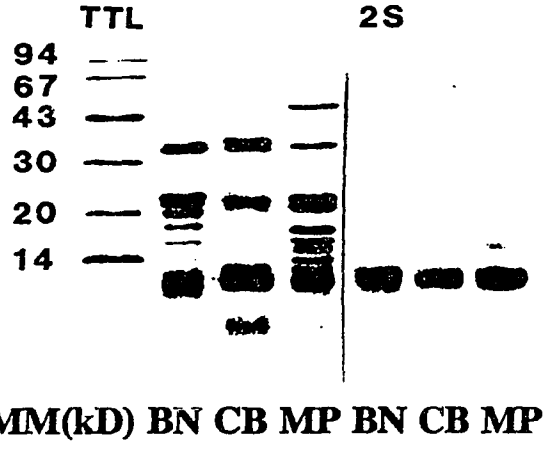


Fig. 3. SDS-PAGE profiles of the total and 2S proteins from seeds of the Brazil nut family.

MM, molecular weight marker; BN, Brazil nut; CB, cannonball; MP, monkeypot; TTL, total salt buffer extractable proteins; 2S, 2S proteins purified by sucrose-gradient centrifugation.

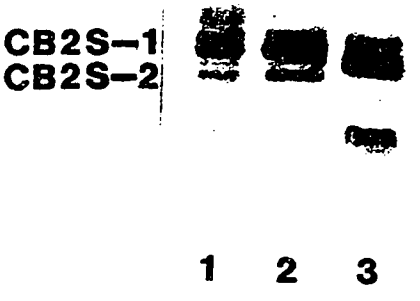


Fig. 4. Components and subunit structure of the CB2S protein.

The Purified CB2S was separated in SDS-PAGE. Lane 1, CB2S protein treated with β-mercaptoethanol (β-Me) then dialyzed to remove the β-Me; lane 2, CB2S protein without β-Me treatment; lane 3, CB2S protein treated with β-Me.

from the gel, treated with β-Me, and separated by SDS-PAGE, both components contained two polypeptides of about 9 kD and 3 kD (Fig. 5). While the size of the two large polypeptides from the CB2S-1 and CB2S-2 is slightly different, the two small polypeptides have very similar, if not identical, molecular weight, and are not separated in a 20% SDS-PAGE gel. Therefore, only three polypeptide bands were observed (Fig. 4, lane 3). Similar changes in the polypeptide patterns upon β-Me treatment were also observed for the BN2S and MP2S. In the absence of β-Me, two bands were observed for the BN2S (Fig. 6, lane 1). After β-Me treatment, however, three bands appeared, two around 9 kD and one about 3 kD (Fig. 6, lane 2). In MP2S, treatment with β-Me resulted in a change from one (or two) band (Fig. 6, lane 5) to three bands of about 15, 9 and 3 kD (Fig. 6, lane 6).

Amino acid composition The 2S proteins purified by sucrose gradient centrifugation from monkeypot and cannonball seeds contain high levels of the sulfur amino acids (Met plus Cys), 18.3% in the MP2S and 24.4% in the CB2S proteins as determined by amino acid analysis (Table 4). These total sulfur AA contents are slightly lower than the 25.8% in the BN2S (Sun *et al*, 1987). The CB2S and MP2S are also rich in Arg and Glx, 15.5% and 27.4% in CB2S and 10.6% and 23.3% in MP2S, respectively.

Immuno-reactivity By Western blot analysis, the BN2S, MP2S and CB2S proteins could be detected immunologically

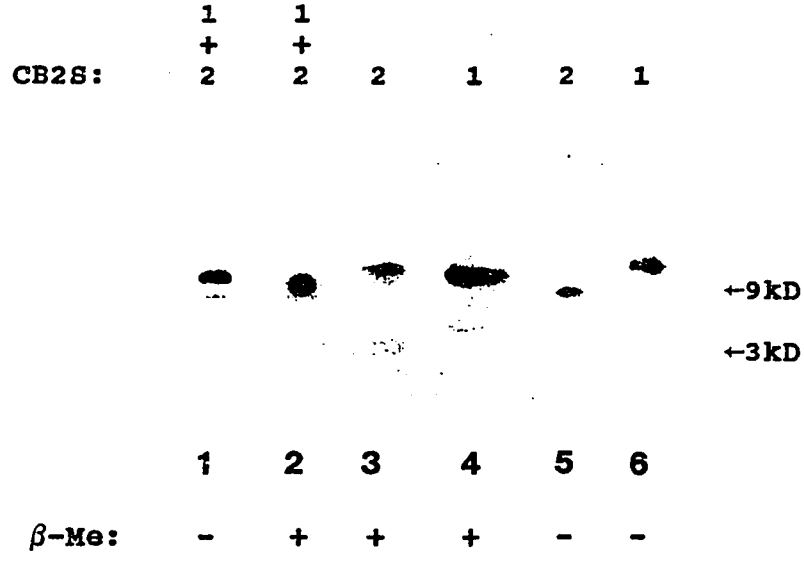


Fig. 5. Subunit structural analysis of the two protein components of the CB2S.

The CB2S-1 and CB2S-2 protein components were isolated from SDS-PAGE, treated with or without β-Me, and analyzed on SDS-PAGE. Molecular weights are marked with arrows. 1+2, total CB2S; 1, CB2S-1; 2, CB2S-2; +, treated with β-Me; -, without β-Me treatment.

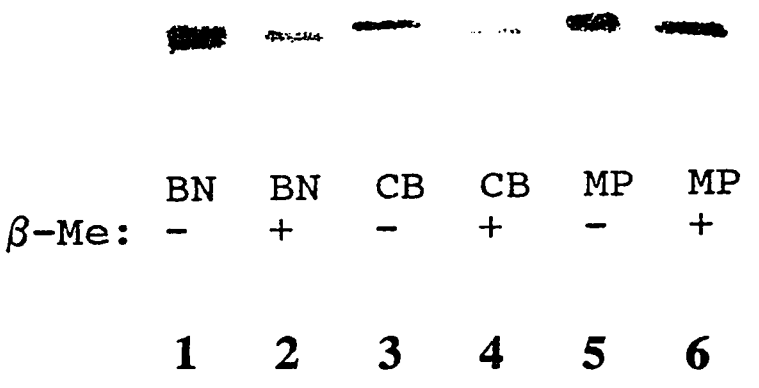


Fig. 6. Comparison of protein components in the 2S proteins of the Brazil nut family.

The 2S proteins were treated with or without β-Me and then separated on a SDS-PAGE. BN, Brazil nut; CB, cannonball; MP, monkeypot; "+", with β-Me; "-", without β-Me.

Table 4. Amino Acid Compositions of the MP2S, CB2S and BN2S

Amino acid	MP2S Mol%	CB2S Mol%	BN2S Mol%*
Lys	2.8	0.3	1.6
His	1.9	1.7	1.9
Arg	10.6	15.5	12.1
Asx	4.9	2.8	4.4
Thr	3.3	--	0.5
Ser	6.4	6.3	5.6
Glx	23.3	27.4	23.1
Pro	7.7	8.1	6.5
Gly	7.2	4.7	7.5
Ala	4.2	1.8	1.8
Cys	4.3	4.5	9.7
Val	1.8	0.3	0.8
Met	14.0	19.9	16.1
Ile	1.1	0.2	0.7
Leu	5.7	5.6	5.7
Phe	0.9	0.8	.4

* Data from Sun *et al*, 1987

using the monoclonal antibody specific for the large subunit of BN2S (Fig. 7A) or of CB2S (Fig. 7B). This suggests that the BN2S, MP2S and CB2S proteins share common epitopes and are immunologically related.

Amino-terminal Sequence An 11-amino acid sequence from the amino terminus of the large subunit of CB2S protein was determined. This sequence shows 54.5% homology with that of the BN2S protein (Fig. 8). The N-terminus AA sequence of the CB2S contain two less Met residues as compared with the BN2S.

Southern and Northern blot analyses.

Genomic DNA from cannonball was digested with *Eco*RI or *Hind*III and probed with [^{32}P]-labelled BN2S cDNA pHS-3. A 5.9-kb *Eco*RI fragment and a 3.6-kb *Hind*III fragment hybridized with the BN2S probe (Fig. 9A). The results of monkeypot genomic Southern blot are shown in Fig. 9B. Two fragments of *Xba*I digestion products, 3.8 and 3.5 kb, and two fragments of *Eco*RI digestion products, 5.0 and 3.6 kb, respectively, hybridized with the BN2S probe.

Poly(A)$^+$ RNA was isolated from developing seeds of monkeypot and cannonball. When the cannonball poly(A)$^+$ RNA was electrophoresed on a denaturing agarose gel, a major band of 0.7 kb was observed by ethidium bromide staining (Fig. 10 A1). Northern blot analysis using BN2S cDNA as a probe showed that this 0.7 kb poly(A)$^+$RNA had a high homology with the BN2S cDNA sequence (Fig. 10 A2). Similarly, a 0.7-kb

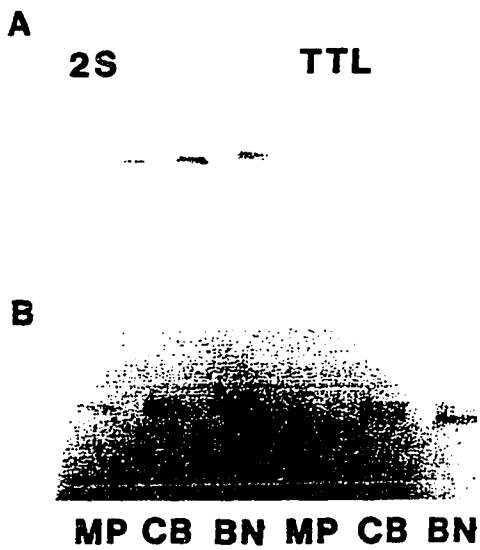


Fig. 7. Immunoblot analysis of the 2S proteins in the Brazil nut family.

Total and 2S proteins from seeds of Brazil nut (BN), cannonball (CB) and monkeypot (MP) were electrophoretically separated, transferred to membrane, reacted with antibody, and then detected by enzyme-linked reaction. Panel A: treatment with monoclonal antibody specific for the large subunit of BN2S; B: treatment with monoclonal antibody specific for the large subunit of CB2S.

BN2S: NH-**Pro Arg Arg** Gly Met **Glu** Pro **His** Met **Ser** Glu...

CB2S: NH-**Pro Arg Arg** Pro Glu **Glu** Ser **His** Leu **Ser** Gln...

Fig. 8. Comparison of the amino acid sequences at the amino-terminus of the 9-kD subunits of BN2S and CB2S.
Identical amino acids are shown in bold.

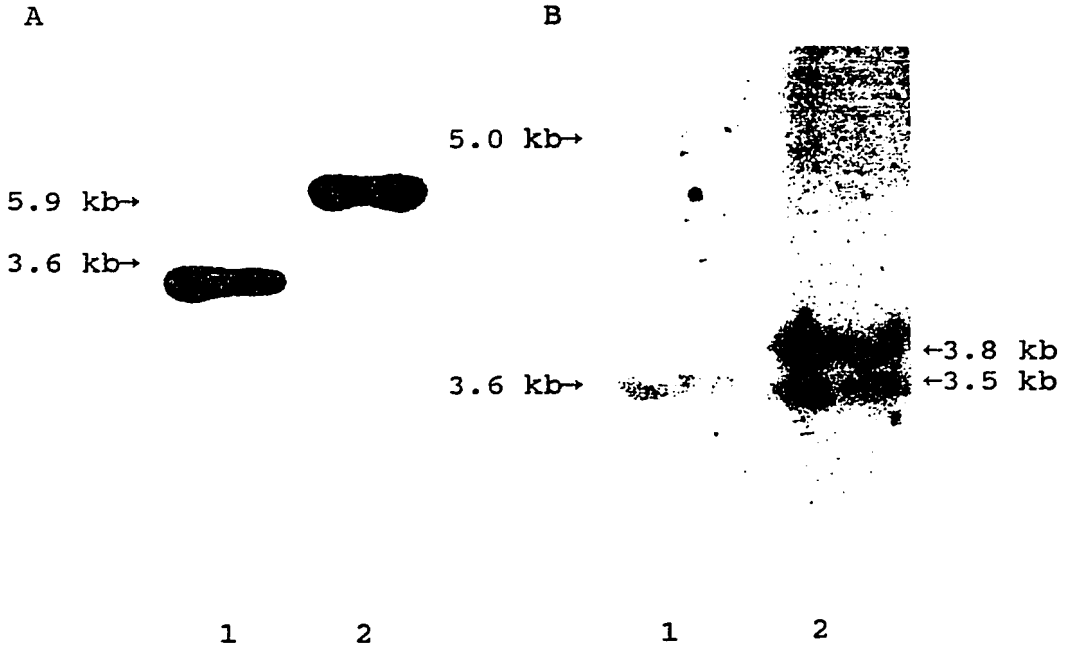


Fig. 9. Southern blot analysis of cannonball and monkeypot genomic DNA.

Genomic DNA from cannonball (A) was digested with *Eco*RI (lane 1) or *Hind*III (lane 2) and from monkeypot (B) with *Eco*RI (lane 1) or *Xba*I (lane 2). After separation by electrophoresis and transfer to membrane, the DNA was hybridized with the BN2S cDNA probe. The size of the fragment was as indicated by the arrow.

RNA in the total RNA of monkeypot seeds hybridized with the BN2S cDNA (Fig. 10B). A 0.7-kb mRNA is sufficient to encode the mature 2S protein (12 kD). This information suggests that the mRNAs encoding for the CB2S and MP2S are abundant in the developing seeds and have a size of 0.7 kb.

Processing of the 2S Proteins in Cannonball and Monkeypot Seeds.

When the cannonball poly(A)$^+$ RNA was translated in a wheat germ *in vitro* system, one polypeptide of about 18 kD was heavily labelled by [^{35}S]-Met (Fig. 11, lane 1). However, *in vivo* labelling of cannonball seeds at the same developing stage showed that most of the [^{35}S]-Met was incorporated into a 15-kD polypeptide. After chasing the [^{35}S]-Met-labelled seeds with nonradioactive Met for 15 min to 12 hr at 30°C, the amount of radioactivity in the 15-kD polypeptide decreased while the amount of radioactivity in the 9-kD and 3-kD polypeptides appeared and increased gradually (Fig. 11, lanes 2 through 10).

In monkeypot seeds, the major cell-free translational product of the monkeypot poly(A)$^+$ RNA was an 18-kD polypeptide (Fig. 12, lane 1). *In vivo* labelling of developing monkeypot seeds indicated that most of the [^{35}S]-Met was incorporated into a 15-kD polypeptide. Chasing the labelled monkeypot seeds with cold Met at 30°C showed that the amount of radioactivity of the 15-kD polypeptide decreased

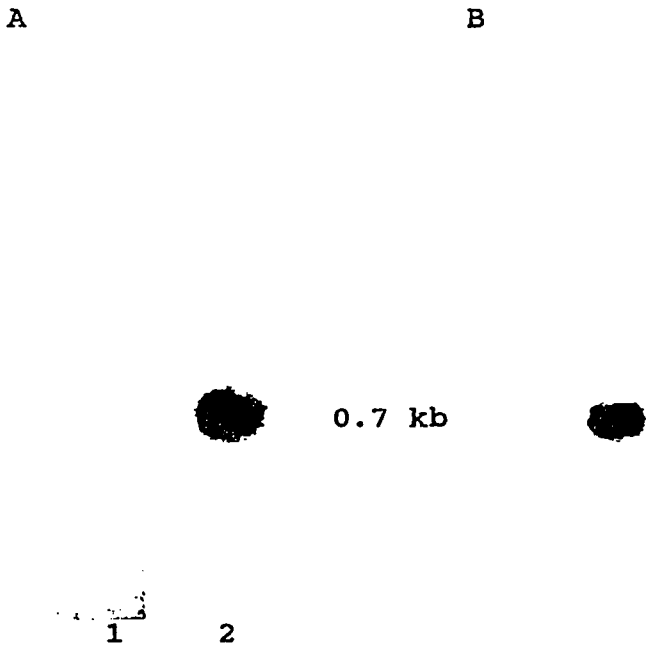


Fig. 10. Northern blot analysis of mRNAs from the immature seeds of cannonball and monkeypot.

A, Poly $(A)^+$ RNA from cannonball seeds, Lane 1, EtBr staining; lane 2, hybridized with the BN2S cDNA. B, total RNA from monkeypot seeds hybridized with the BN2S cDNA probe.

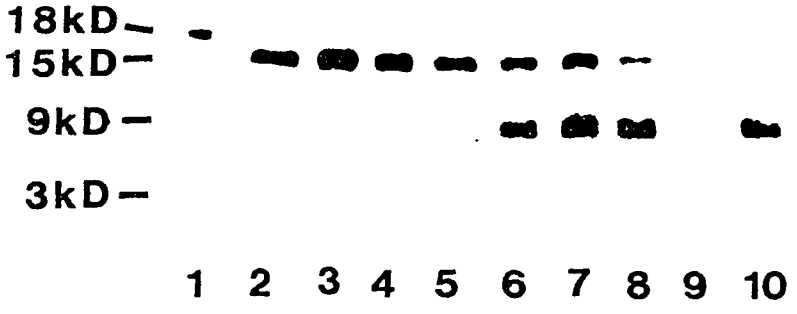


Fig. 11. <u>In vitro</u> cell-free translation and <u>in vivo</u> pulse-chase labeled products from developing cannonball seeds.

Cannonball seeds were pulse labeled with [^{35}S]-methionine for 1 hr, then chased with cold methionine at 30°C for various time periods. The sizes of the major labeled polypeptides are indicated with bars. Lane 1, *In vitro* translation products; lanes 2, 3, 4, 5, 6, 7, 8, 9, and 10, chased for 0, 15 min, 0.5, 1, 3, 5, 8, 10, and 12hr, respectively.

Fig. 12. In vitro cell-free translation and in vivo pulse-chase labeled products from developing monkeypot seeds.

Monkeypot seeds were pulse labeled with [^{35}S]-methionine for 1 hr, then chased with cold methionine at 30°C for various time periods. The sizes of the major labeled polypeptides are indicated with bars. Lane 1, *In vitro* translation products; lanes 2, 3, 4, 5, 6, 7, 8, 9, 10, and 11, chased for 0, .5, 1, 2, 3, 5, 7, 9, 12, 15 hr, respectively.

gradually, while the amount of radioactivity of a 12-kD polypeptide appeared after chasing the seeds for 1 hr. The level of the 12-kD polypeptide diminished slowly with longer chasing periods, and the amount of radioactivity gradually accumulated into the 9- and 3-kD polypeptides (Fig. 12, lanes 2 to 11).

Interestingly, the changes in labeling patterns of the monkeypot and cannonball seed proteins were observed only when chasing at 30°C. At room temperature (about 25°C), the labelled polypeptides remained unchanged (Fig. 13).

cDNAs encoding the 2S proteins in monkeypot and cannonball.

A vector-primer cDNA library was constructed from the cannonball seed poly $(A)^+$ RNA. By screening the library using the BN2S cDNA, pHS-3, as a probe, two positive clones, pCB2S-3 and pCB2S-4, were identified.

Complementary DNAs were also synthesized from poly $(A)^+$ RNA of cannonball and monkeypot seeds and inserted into the bacteriophage vector λgt11. The resulting libraries were also screened with the BN2S cDNA probe. Four CB2S (pCB2S-1, pCB2S-2, pCB2S-5 and pCB2S-6) and three MP2S (pMP2S-1, pMP2S-2, and pMP2S-56) clones were identified. The insert size of these positive clones ranges from 600 to 850 bp.

MP2S cDNA clones. The NT sequences of all the MP2S positive clones were determined as shown in Fig. 14. In

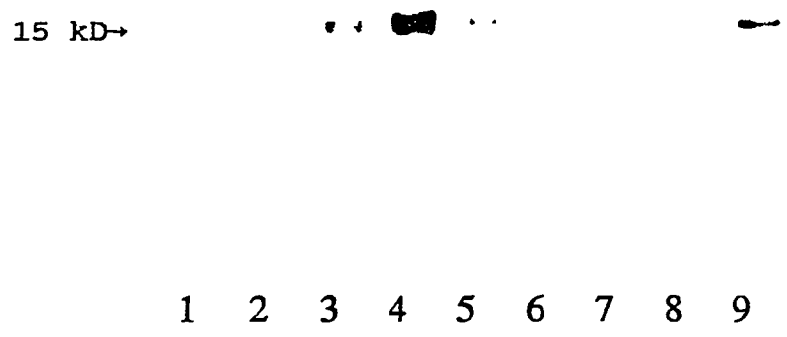


Fig. 13. <u>In vivo</u> pulse-chase labelled products of developing cannonball seeds at 25°C.

The cannonball seeds were labelled with [^{35}S]-methionine for 1 hr, then chased with cold methionine at room temperature (about 25°C) for 0, .25, .5, 1, 1.5, 2, 3, 6.5, 12.5, and 18 hr (lanes 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10, respectively).

the coding regions, there is only one nucleotide difference (A/C) between pMP2S-1 and pMP2S-56. pMP2S-1 contains a poly (A) tail while pMP2S-56 did not. pMP2S-2 has a lower sequence identity with pMP2S-1 and pMP2S-56, and a shorter 3' untranslated region (UTR) than pMP2S-1. Between translation start site and the poly (A) tail, the pMP2S-2 has 71% sequence identity with the pMP2S-1 and pMP2S-56. However, it contains some insertions and deletions as well as point mutations that are absent from the other two clones.

A typical translation start codon ATCATGGC is present in all three MP2S clones. In both pMP2S-1 and pMP2S-56 clones, a translational stop codon is located 429 bp downstream of the translation start codon. The open reading frames (ORF) of both pMP2S-1 and pMP2S-56 encode a polypeptide of 144 AAs, about the size of the MP2S precursor. In pMP2S-1, following the translation stop codon (TGA) is a 3' UTR of 126 bp before reaching the poly (A) tail, while 111 bp in pMP2S-56 before reaching the cloning site. An additional translation stop codon is present in the 3' UTR of these two clones. Two putative polyadenylation signals, AATAA and AATAATA, are located at 93 bp and 26 bp, respectively, upstream of the poly (A) sequence in the pMP2S-1. These two putative polyadenylation signals can be found at the same positions in the pMP2S-56 clone, but, there is no poly (A) sequence in this region before reaching the cloning site.

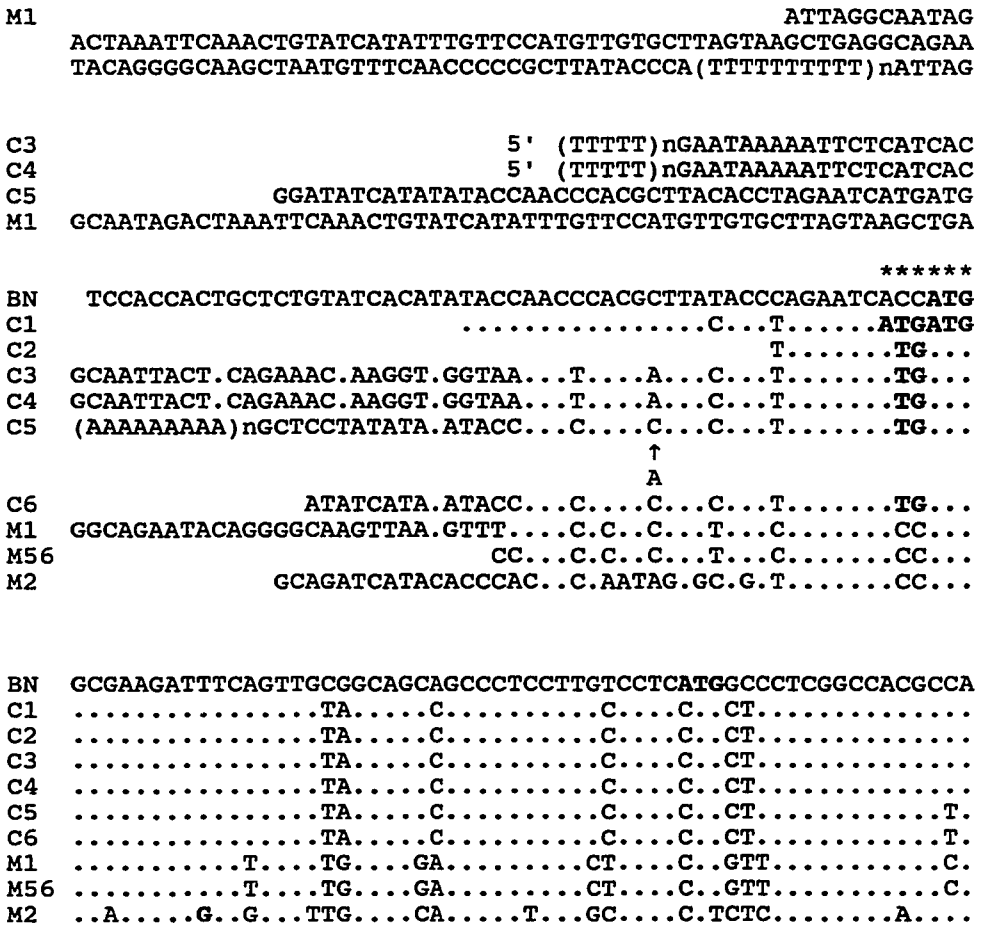


Fig. 14. Comparison of nucleotide sequences of the cDNA clones of BN2S, CB2S, and MP2S.

BN represents BN2S cDNA clone pHS-3; C1 to C6 represent CB2S cDNA clones, pCB2S-1 to -6; M1, M56 and M2 represent MP2S cDNA clones, pMP2S-1, pMP2S-56, and pMP2S-2, respectively. The sequences are aligned to maximize homology. Arrows with NTs indicate sequence displaced by the alignment. The putative translation start and stop codons are indicated by "*"s. The methionine codons are bolded. The two methionine rich regions are boxed. The putative polyadenylation signals are underlined. (A)n and (T)n indicate long A or T sequences.

```
BN   CCGCCTTCCGGGCCACCGTCACCACCACAGTGGTGGAGGAGGAGAACCAGGAGGAGTG
C1   ..............G.....................................C.......
C2   ..............G.....................................C.......
C3   ..............G.........G...........................C.......
C4   ..............G.........G...........................C.......
C5   ..............G.........A...........................C.......
C6   ..............G.........A...........................C.......
M1   ..............C.........A...........................G.T---..
M56  ..............C.........A...........................G.T---..
M2   .T......A..A...CT.........A.GT...A.................C.GC....
                                          ↑           ↑
                                       GGAGCA   CCCGCCAACAGTCCG

BN   TCGCGAGCAGATGCAGAGACAGCAGATGCTCAGCCACTGCCGGATGTACATGAGACAG
C1   C............G.....................................................
C2   C............G.....................................................
C3   C............G.....................................................
C4   C............G.....................................................
C5   C............G.....................................................
C6   C............G.....................................................
M1   C............C.................................A.....................
M56  C............C.................................A.....................
M2   C............G....G......CA.....A.......A.A.T....TTA......

BN   CAGATGGAGGAGAGCCCGTACCAGACCATGCCCAGGCGGGGAATGGAGCCGCACATGA
C1   ............G......T.......C........AAT.A.GGA....T.....C...
C2   ............G......T.......C........AAT.A.GGA....T.....C...
C3   ............G......T.......C........GCG.A.GGA....T.....C...
C4   ............G......T.......C........GCG.A.GGA....T.....C...
C5   ............G......T.......C........GCG.A.GGA....T.....C...
C6   ............G......T.......C........GCG.A.GGA....T.....C...
M1   ............A......A.......A........GCG.G.AAT....C.....A...
M56  ............A......A.......A........GCG.G.AAT....C.....A...
M2   ....G.......A......A.......GG....C.GAG.G.GGA....C....TC..G
                               ↑             ↑
                               TTG           CGG

BN   GCGAGTGCTGCGAGCAGCTGGAGGGGATGGACGAGAGCTGCAGATGCGAAGGCTTAAG
C1   .....................................................................
C2   .....................................................................
C3   .....................................................................
C4   .....................................................................
C5   .....................................................................
C6   .....................................................................
M1   .....................................................................
M56  .....................................................................
M2   A.....................................TG........T.......-...

BN   G|ATGATGATGATGAGGATG|CAACAGGAGGAGATGCAACCCCGAGGGGAGCAG|ATGCGA
C1   .|.........---....G.|....G..........G..........C.......|......
C2   .|.........---....G.|....G..........G..........C.......|......
C3   .|.........---....G.|....G..........G..........C.......|....A.
C4   .|.........---....G.|....G..........G..........C.......|....A.
C5   .|.........---....G.|....G..........G..........C.......|....A.
C6   .|.........---....G.|....G..........G..........C.......|....A.
M1   .|..........G..T..T.|....G........---C.........AG......|.A..A.
M56  .|..........G..T..T.|....G........---C.........AG......|.A..A.
M2   .|.........---....G.|....G..C.......G.G.TT.AG.GC......|G...AG
```

```
BN   AGGATGATGAGGCTG|GCCGAGAATATCCCTTCCCGCTGCAACCTCAGTCCCATGAGAT
C1   ...........A..|..T......C.........................CA.....
C2   ...........A..|..T......C.........................CA.AA..
C3   ...........A..|..C......C.........................CA.GG..
C4   ...........A..|..C......C.........................CA.GG..
C5   ...........A..|..C......C.........................CA.AG..
C6   ...........A..|..C......C.........................CA.AG..
M1   ...........A..|..C......A.........................CA.AG..
M56  ...........A..|..C......C.........................CA.AG..
M2   ...........A..|..C....GCC...T................C..GCA.AG..

                                    ***
BN   GCCCCATGGGTGGCTCCATTGCCGGGTTCTGAATCTGCCACTAGCCAGTGCTGTAAAT
C1   ..T.....A.AG.....G.G.....T.........A......CAT--.CTG..--...
C2   ..T.....A.AG.....G.G.....T.........A......CAT--.CTG..--...
C3   ..T.....A.AA.....G.G.....T.........A......AAT--.TTG..--...
C4   ..T.....A.AA.....G.G.....T.........A......AAT--.TTG..--...
C5   ..T.....A.AA...T.G.G.....T.........A......AAT--.TTG..--...
C6   ..T.....A.AA...T.G.G.....T.........A......AAT--.TTG..--...
M1   ..C.....G.TA...C.A.G.....C.........A......ACTT..TGA..TG...
M56  ..C.....G.TA...C.A.G.....C.........A......ACTT..TGA..TG...
M2   ........G.TGA..A.A.GA..T.G.......C.A..---.AGCC-CTGC-.CAT.C

BN   GTTAATAAGGCTCTCACAAACTAGCTCTTTGTTGGCTTTTGGCCGGAGACTAGGGTGT
C1   ...............C..............C............TG.............
C2   ...............C..............C............TG.............
                     ↑
                     G
C3   ...............C..............C............TA..........C..
C4   ...............C..............C............TA..........C..
C5   ...............C..............C............TA..........C..
C6   ...............C..............C............TA..........C..
M1   A...........G..A......T.....C.CC...........CG..........T..
M56  A...........G..A......T.....C.CC...........CG..........T..
M2   G.....AT.AA-.A.CG....C..G...--.....GCA.AGGTT..GG.GA.AT-.

BN   GGGATTAATAATAATAGCACACTATCGTGTGTTCTCAGCTTC(A)n   3'
C1   ..............A......T...........T........(A)n
C2   ..............A......T...........T........(A)n
C3   ..............A......T...........T........(A)n
C4   ..............A......T...........T........(A)n
C5   ..............A......T...........T........(A)n
C6   ..............A......T...........T...(A)n
M1   ..............G......T...........T........(A)n
M56  ..............G......T.....
M2   AAT.A..G.....-.....T...........TC(A)n
```

For the pMP2S-2 clone, the ORF encodes a polypeptide of 153 AAs as compared to the 144-AA polypeptide encoded by the other two MP2S clones. Two putative polyadenylation signals, AATAAA and AATAATAGTAATAA, are located at 80 and 19 bp upstream of the poly (A) sequence.

CB2S cDNA clones. A translation start site with an ATG-doublet (ATC**ATGATG**GC) is present in all six CB2S cDNA clones (Fig. 14). Between this translation start codon and the poly (A) tail, these clones share over 98% sequence identity. All the six clones contain an ORF of 435 bp, encoding a 145-AA polypeptide, about the size of the 18 kD precursor of the CB2S. Following the translation stop codon, there are 122 bp of 3' UTR in clones pCB2S-1, 3, 4, and 5, 123 bp in pCB2S-2, and 114 bp in pCB2S-6. Translation stop codons are also present in the 3' UTR at 9 bp (TAA) and 24 bp (TAATAA) downstream of the first stop codon (TGA). The AATAA sequence in the TAA-doublet stop codon could also serve as polyadenylation signal. Another putative polyadenylation signal, AATAATAAAA, is located at 26 bp upstream of the poly (A) sequence. In comparison with other clones, pCB2S-5 has a single nucleotide (A) insertion in its 5' UTR, while pCB2S-2 has a single nucleotide (G) insertion in its 3' UTR.

Sequence Comparison of the MP2S and CB2S cDNA clones with the BN2S cDNA clone pHS-3. Figure 14 shows a comparison of NT sequences of the pCB2S-1, -2, -3, -4, -5, -

6, and pMP2S-1, -56, and -2 with that of the BN2S cDNA clone pHS-3. The translation start and stop codons in all the CB2S and the pMP2S-1 and -56 clones are located at the same position as in the BN2S pHS-3, resulting in very similar ORFs for these clones. The CB2S clones have only one less and the MP2S clones have two less codons than the BN2S pHS-3 clone. All the MP2S clones share the same translation start sequence (ATCACC**ATG**GC) with the BN2S clone. However, the translation start sequences of all the CB2S clones contain an ATG-doublet (ATC**ATGATG**GC). The ORFs of all these 2S cDNAs terminate at the same stop signal, TGA.

At the NT level, between the translation start site and the poly (A) tail, there is 90% homology between all of the CB2S cDNA clones with the BN2S cDNA (pHS-3) and 94% homology between pMP2S-1, pMP2S-56 and BN2S pHS-3 (Table 5). At the AA level, CB2S has 82-84 and 77-80% homologies with BN2S and MP2S, respectively, and MP2S shares 84% homology with the BN2S. pMP2S-2 has a lower homology with other 2S protein clones, about 72% at the NT level and 74% at the AA level.

The Met and Cys content of all of the CB2S and MP2S isoforms are shown in Table 6. The CB2S isoforms contain 17 to 18 Met and 8 Cys residues, and the isoforms, derived from pMP2S-1 and pMP2S-56, contain 19 Met and 8 Cys residues. In their putative mature polypeptides, isoforms of CB2S contain 15 to 16 Met and 8 Cys; pMP2S-1 and pMP2S-56, 16 Met and 8 Cys; and BN2S pHS-3, 19 Met and 8 Cys residues. pMP2S-2

Table 5. Nucleotide and Amino Acid Sequence Homologies of the 2S SRPs in the Brazil nut family

% Homology (Nucleotide / Amino Acid)

	BRAZIL NUT	MONKEYPOT
MONKEYPOT*	94 / 84	
CANNONBALL	90 / 82-84	91 / 77-80

* pMP2S-2 clone is not included in these comparisons.

Table 6. Methionine and Cysteine Contents of the Different Isoforms of CB2S, MP2S and BN2S

Clones	# of Met			# of Cys		
	TTL*	9kD	3kD	TTL	9kD	3kD
BRAZIL NUT pHS-3	22	14	5	8	6	2
CANNONBALL						
pCB2S-1	18	11	5	8	6	2
2	18	11	5	8	6	2
3	17	10	5	8	6	2
4	17	10	5	8	6	2
5	17	10	5	8	6	2
6	17	10	5	8	6	2
MONKEYPOT pMP2S-1	19	12	4	8	6	2
56	19	12	4	8	6	2
2	13	9	2	8	6	2

* TTL denotes the whole precursor; 9 kD and 3 kD represent the two subunits.

encodes only 13 Met codons in its ORF. In its putative mature polypeptides, pMP2S-2 contains 11 Met and 8 Cys residues.

The 5' UTR from position -1 to -26 bp is very similar for all of the CB2S, the two MP2S clones, pMP2S-1 and pMP2S-56, and the BN2S pHS-3 clones. Beyond this region, however, the homology becomes very low. Unexpectedly, a long T sequence is present in the upstream region of the 5' UTRs of the pCB2S-3, pCB2S-4, and pMP2S-1, while a long A sequence is present in the pCB2S-5 clone. The size of these poly (T) or (A) sequences is longer than 50 bp. There are additional sequences 5' beyond these poly A's or T's before reaching the vector sequences in the pCB2S-5 and pMP2S-1 clones. In the case of pCB2S-5, up to 40 bp of this 5' upstream (of the A's) sequence are identical to its 5' UTR except for one A. However, in the pMP2S-1 clone, such a sequence is not homologous to any sequence in all of the clones. pCB2S-3 and pCB2S-4 do not contain these unusual sequences.

DISCUSSION

Structure of the CB2S and MP2S.

Like most other 2S seed proteins, BN2S, MP2S and CB2S from the Brazil nut family contain two subunits, one large (9 kD) and one small (3 kD). The fact that the two subunits migrate together in their oxidized forms and migrate separately in their reduced forms on a SDS-PAGE suggests that they are linked by disulfide bonds. This conclusion is confirmed by the β-Me treatment study and by their derived AA sequences. Eight Cys residues are present in the two polypeptides which can form four pairs of disulfide bonds. The reversible association and dissociation of the subunit polypeptides of the CB2S protein under oxidation and reduction conditions is illustrated in Fig. 15. The results from the β-Me treatment study indicate that CB2S-1 is smaller than CB2S-2 in size. However, in their deoxidized forms, CB2S-1 migrates slower than CB2S-2 in SDS-PAGE, suggesting that the structure of CB2S-1 might be slightly different from that of CB2S-2.

2S proteins as storage proteins in monkeypot and cannonball.

The 2S proteins are one of the major seed proteins in Brazil nut, monkeypot and cannonball. The CB2S protein accounts for more than half of the total seed protein in cannonball. The BN2S, MP2S and CB2S proteins contain high

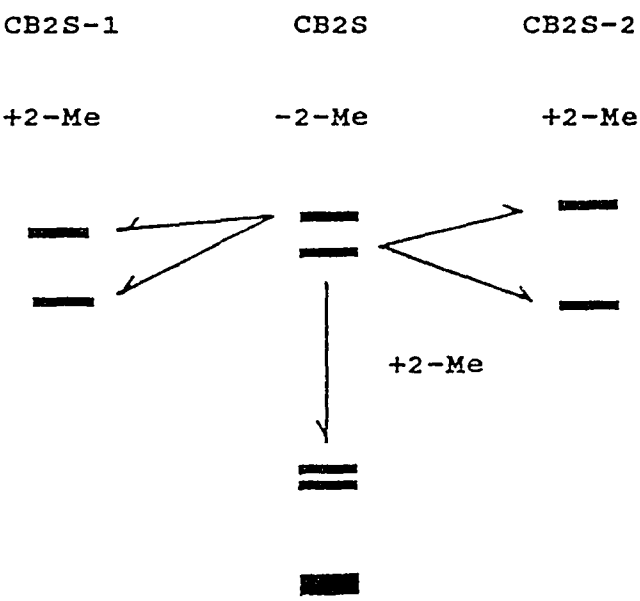


Fig. 15. Reversible association and dissociation of the subunit polypeptides of the CB2S protein.

levels of Arg and Glx, a common feature of seed storage proteins. The 2S proteins and their mRNAs can only be detected at specific stages of seed development, suggesting that the 2S protein genes are expressed temporally in seeds. The results of immunological reaction (Fig. 7) indicate that the MP2S, CB2S and BN2S proteins share some antigenic similarities. Sequence data reveal that BN2S, MP2S and CB2S are highly related. It has been shown that the Brazil nut 2S SRP disappears during seed germination, suggesting that BN2S is a seed storage protein. All this information strongly suggests that the MP2S and CB2S proteins are also seed storage proteins.

MP2S and CB2S as SRPs.

The 2S seed proteins from the Brazil nut family are rich in sulfur amino acids. AA composition analysis reveals that MP2S, CB2S and BN2S contain 18.3%, 24.4%, 25.8% sulfur AA residues, respectively. cDNA sequence data further confirm that MP2S contains 18%, CB2S, 16%, and BN2S 19% Met. The discrepancy between the sulfur AA content as derived from AA composition analysis and from cDNA sequence data could be due to that the 2S protein fraction obtained by sucrose-gradient centrifugation is a mixture of 2S proteins, therefore, its AA composition represents an average of all these protein species. For example, in the case of MP2S protein, another polypeptide with a molecular weight of about 15 kD could be

observed in the 2S fraction (Fig. 3, lane 6). This polypeptide may have a higher or lower sulfur AA content than MP2S. Another explanation could be that there are some isoforms of the 2S SRP that contain less Met residues, for example, pMP2S-2. Finally, there could be an inaccuracy in the AA composition analysis.

Since the 2S protein in the Brazil nut family is a major seed protein, its abundant Met and Cys residues contribute significantly to the overall sulfur AA level of the total seed protein. Thus, the Brazil nut total seed protein contains 6% sulfur amino acids (FAO, 1970). In cannonball, the CB2S protein constitutes more than 50% of the total seed protein. Since it contains 24.4% sulfur AAs, a content of about 12% sulfur AAs can be expected for the cannonball total seed protein.

The high level of the 2S SRP accumulated in the seeds suggests that this protein has a low potential to interfere with the normal biological function of the cells. This characteristic is an advantage for introducing the sulfur-rich protein gene from Brazil nut, monkeypot or cannonball into other plants for protein quality improvement.

Processing of the MP2S and CB2S.

The results from *in vivo* and *in vitro* labelling studies on monkeypot seeds (Fig. 12) suggest that MP2S is synthesized as an 18-kD precursor (*in vitro* translation product), from

which a signal peptide of about 3 kD is co-translationally cleaved to give a 15-kD precursor (1st polypeptide observed in the *in vivo* labelling products). This polypeptide undergoes further processing into a 12-kD intermediate polypeptide, and, finally, this 12-kD polypeptide is cleaved to form the two mature subunits. This processing pattern is the same as that of BN2S (Fig. 16). The processing of CB2S is similar to those of BN2S and MP2S; the sole difference is that in CB2S, the 15-kD precursor appears to give rise directly to the 9-kD and 3-kD subunits in the absence of the 12 kD intermediate precursor (Fig. 11). The reason for this difference is not clear. The AA sequence comparison indicates that around the N-terminus of the large subunit, which is the final cleavage site, there are some AAs conserved in this region between BN2S and MP2S, but not CB2S. This difference might affect the processing pattern. The processing enzymes or the cytoplasmic conditions might also contribute to the different processing patterns. For example, when the Brazil nut SRP gene was expressed in yeasts, the precursor was processed only to the 12-kD intermediate polypeptide, but not the mature polypeptides of 9 kD and 3 kD (Sun *et al*, 1988). An interesting observation is that processing of the 2S proteins in these three plants occurs only at 30°C or above, but not at 25°C. At 25°C, the 15-kD precursors remain in the seeds without further processing (Fig. 13). The requirement

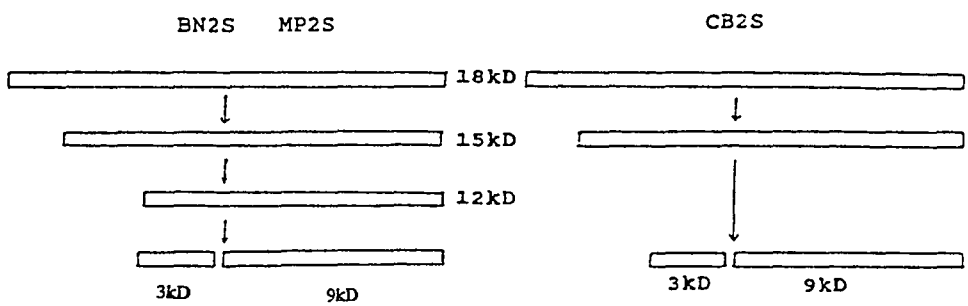


Fig. 16. Processing patterns of the BN2S, MP2S and CB2S. The molecular weights of the precursors and the mature subunits are indicated.

of higher temperatures for cleavages may be related to adaptation of these plants to their tropical climate.

In BN2S, the processing sites for the 15-kD precursor, the 12-kD intermediate precursors, and the 9-kD and 3-kD subunits have been determined (Ampe *et al*, 1986 and Altenbach *et al*, 1987). N-terminal protein sequencing of the first eleven AA residues of CB2S revealed that the large subunit of CB2S starts at the same site as BN2S. The high degree of AA and NT sequence homology between BN2S, MP2S and CB2S and the similarity of their mature subunit polypeptides suggest that the processing sites of precursors of MP2S and CB2S are very similar to, if not the same as, those of BN2S.

The two major components of the 2S proteins (Fig. 6) might have resulted from different processing sites (Altenbach *et al*, 1987). However, in the pulse-chase labelling experiments, we could only detect one component. It is possible that the two components are coded for by different members of the gene family which are expressed at different seed developmental stages. Alternatively, the processing may vary at different developing stages, or additional cleavage and/or modifications may occur later in the protein bodies.

The removal of a signal peptide and post-translational cleavage of the resulting polypeptide is a common feature of seed storage proteins. Endoproteolytic cleavage may be accompanied by terminal peptidase activity during processing

(Higgins, 1984). Except for the first step of co-translational removal of the signal peptide from the precursor, the function of post-translational cleavage is unknown. The propeptide in the precursor molecules may play a role in the determination of a protein structure that is compatible with packaging and storage, such as helping to form the proper disulfide bonds. For the 2S proteins, it is unclear why the cleavage has to involve several steps instead of occurring at the same time. The different processing patterns we observed between the CB2S and the MP2S and BN2S (Fig. 16) may offer a system to study this question.

Cloning and sequence comparison of the CB2S and MP2S cDNAs.

Some of the MP2S and CB2S clones are larger than their 0.7-kb mRNAs. This is mainly due to the presence of the poly A's or T's and the sequences further upstream in these clones. The additional sequences are most likely due to cloning errors.

All the CB2S clones and the pMP2S-1 and pMP2S-56 clones contain a major ORF starting and ending at the same position. These ORFs code for a polypeptide containing a high level of Met and Cys. The sequence homology between these clones and the BN2S cDNAs is over 90% at the NT level and 80% at the AA level, indicating that these clones code for the 2S SRPs.

Sequence analysis indicates that the BN2S, MP2S and CB2S genes are very similar. However, some minor sequence

differences can be observed in different clones, indicating that they come from different members of the gene family. More than 10 isoforms have been identified for BN2S. In this study, five isoforms of CB2S and 3 isoforms of MP2S have been identified through cDNA cloning and sequence analysis. Since the genome size of these plants is not available at the present time, the copy number of the 2S SRP genes in Brazil nut, monkeypot and cannonball can not be determined.

At the NT level, the pMP2S-2 has only 71.5% homology with pMP2S-1 and pMP2S-56. This homology is substantially lower in comparison to the over 90% homology among the other clones of MP2S, BN2S and CB2S cDNA. "Insertions" and "deletions" in pMP2S-2 are two important causes for this lower homology. Interestingly, pMP2S-2 has a slightly higher homology (74%) to BN2S and CB2S than to pMP2S-1 and pMP2S-56 (71.5%). Furthermore, pMP2S-2 is very similar to a BN2S genomic clone, BE2S2 (Gander *et al*, 1991), which codes for an "irregular" BN2S protein. Their coding sequences share 91% homology (Fig. 17). This result indicates that the gene family of the 2S proteins in Brazil nut family can be classified into subfamilies. The fact that BE2S2 was not among the several BN2S protein isoforms that were purified and sequenced (Ampe *et al*, 1986, Altenbach *et al*, 1987, Altenbach *et al*, 1992, and de Castro *et al*, 1987) indicates that it probably codes for a minor 2S protein. Since pMP2S-2 is similar to the BE2S2 and the size of MP2S mRNA is closer

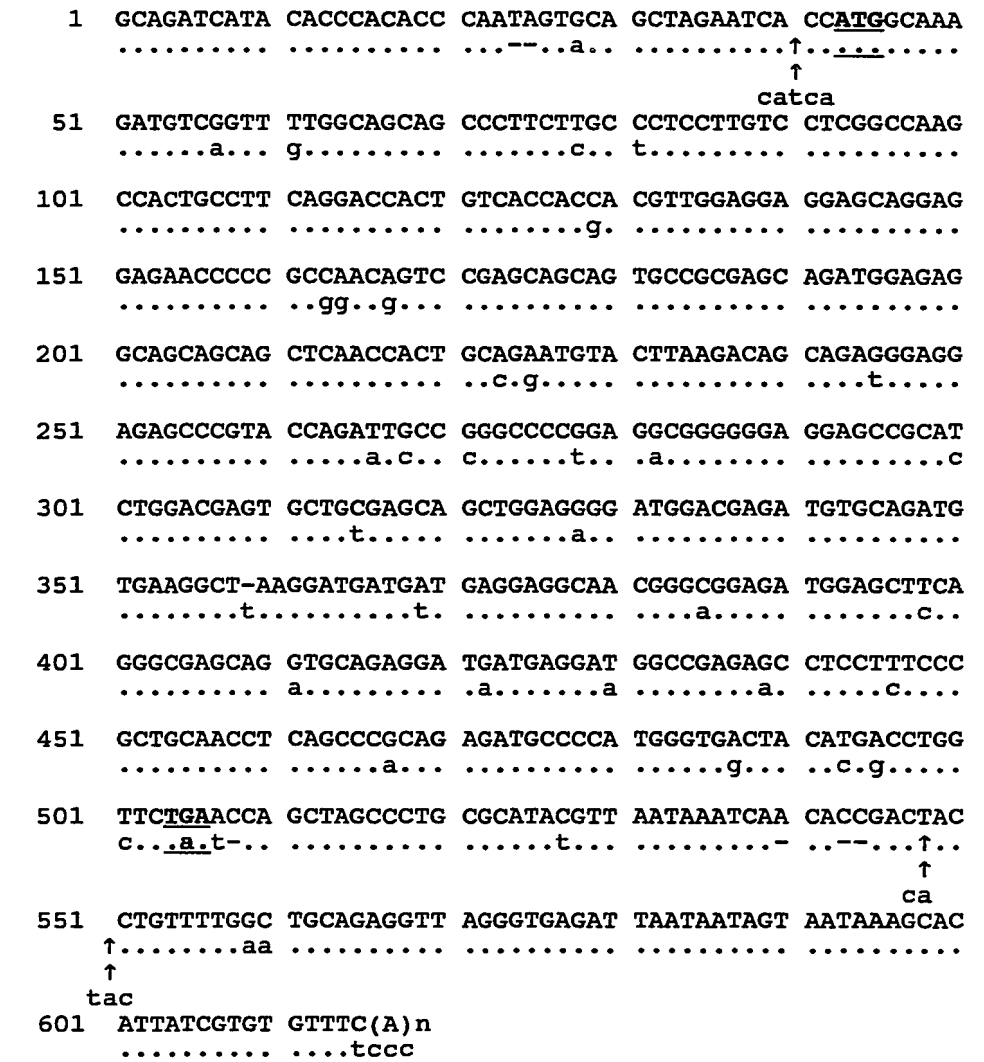


Fig. 17. Comparison of the nucleotide sequences of the MP2S cDNA clone, pMP2S-2, and the BN2S genomic clone, BE2S2.

The top line shows the nucleotide sequence of pMP2S-2 and the bottom line, the BE2S2. The two sequences were aligned to maximize homology. Dots indicate identical NTs. Arrows indicate insertions. Translation start site and stop site are in bold and underlined.

to pMP2S-1 and -56, it is likely that pMP2S-2 comes from a minor subfamily, while pMP2S-1 and pMP2S-56 belong to a major subfamily.

All of the cDNA clones contain a translation start sequence consistent with Kozak's rule for an initiation codon (ANNATGGC, Kozak, 1984). The translation start sequence of the CB2S (ACCATGATGGC), even with two ATG's, still fits the rule of ANNATGGC. This ATG-doublet at the translation start site can also be found in the phaseolin gene (Sun *et al*, 1981).

Although its 3' UTR has a high homology with other clones and its polyadenylation signal is present at the same position as those in the other clones, pCB2S-6 has a shorter 3'UTR. In the pMP2S-2 clone, although its 3' UTR has a lower homology with the other clones and its two poly (A) additional signals are located at positions similar to the other clones, its poly (A) tail starts at a position different from the other clones. These observations imply that the poly (A) sequence in the mRNAs of the 2S proteins is added imprecisely although some positions are preferred.

At the AA level, the deduced 2S protein sequences of the various cDNAs are also very similar except for pMP2S-2. They all contain a high level of Met and their Met residues cluster in two regions, assigned as Box I and Box II. All the CB2S and MP2S protein sequences contain 8 Cys residues and the position of these residues are conserved. These Cys

residues are involved in the formation of disulfide bonds which in turn play an important role in the secondary structure of the 2S SRPs. The cDNA sequence data shows that no Lys codon is present in the BN2S or CB2S clones, while 2 Lys codons (2%) can be found in pMP2S-1 and pMP2S-56.

Evolution of SRP in the Brazil nut family.

As discussed in Chapter I, the 2S SRPs in the Brazil nut family and the 2S proteins from sunflower, castor bean, rape seed and other plants might share a common ancestral gene. The 2S matteuccin gene of ostrich fern could be one of their intermediate ancestral genes. However, matteuccin contains a low level of Met (3%). Amino acid composition analysis (Table 7) reveals that although the Met content of matteuccin is low compared to that of the 2S SRPs from the Brazil nut family, the relative amount of Met plus Arg among these proteins is quite similar. Unfortunately, the entire AA or DNA sequence of matteuccin is not available at this time. Comparison of their Arg and Met residues/codons with those in the Met-rich regions of the SRPs from the Brazil nut family should provide useful information and insight into the evolutionary relationship between these genes. The Met-Arg conversions can be observed among different isoforms of the BN2S, MP2S and CB2S cDNAs (Fig. 18). These conversions occur frequently in the Met-rich regions and are mainly due to a switching of a single base from G to T in their genes,

Table 7. Comparison of the Met and Arg content of the matteuccin and the 2S SRPs from Brazil nut family

	Met mol%	Arg mol%	Met+Arg mol%
Matteuccin	3.0	26.9	29.9
BN2S	19	14	33
MP2S	16	14	30
CB2S	15	17	32

A.

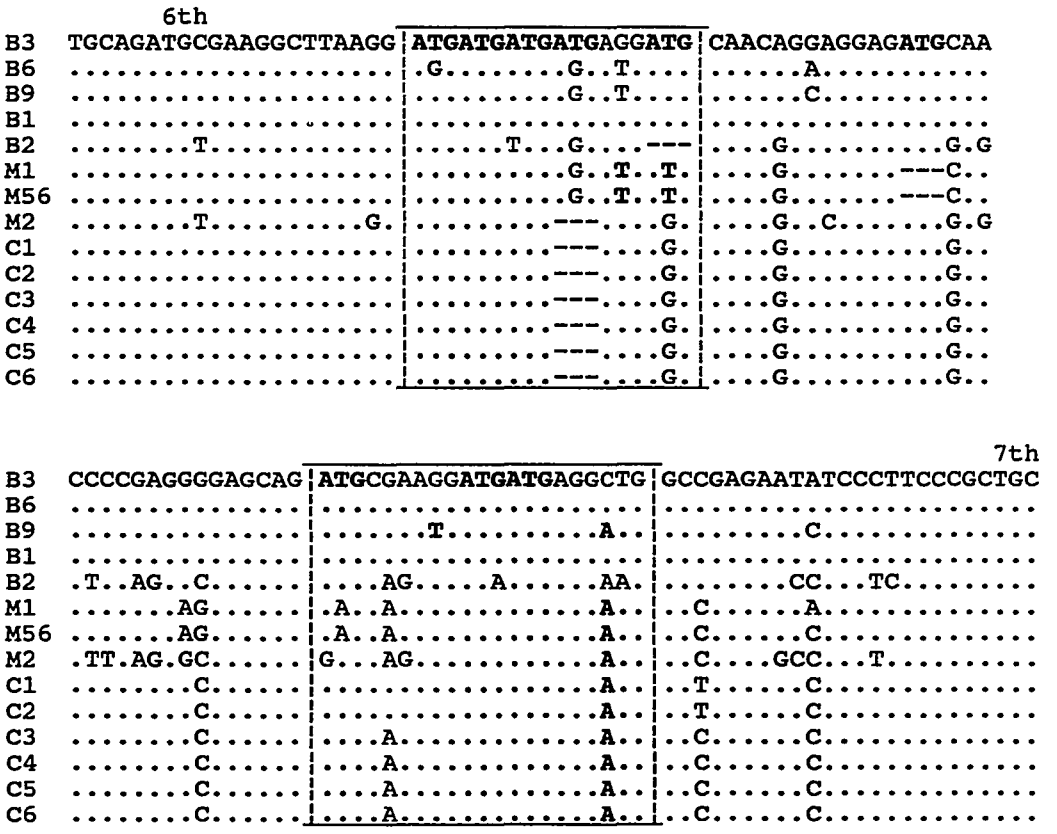
	6				7
pHS-3	ESCRCEGLR	MMMMRM	QQEEMQPRGEQ	MRRMMRL	AENIPSRC
pHS-6	.....	...RM.	..Q.....	..M...M	...L....
pHS-9	.....	R..RM.	..K.....	.....	.....
BE2S1	.....	.....	..E..E.....	.....	.....
BE2S2	.N.....	.-.L..	.....ELQ...	.Q.I..K	...LL...
pMP2S-1	.....	...RM.	....-E..E..	KQ....M	.....
pMP2S-56	.....	...RM.	....-E..E..	KQ....M	...L....
pMP2S-2	.M.....	.-....	..A..ELQ...	VQ....M	..SLL...
pCB2S-1	.....	.-...R	.R...E.....	.R....M	...L....
pCB2S-2	.....	.-...R	.R...E.....	.R....M	...L....
pCB2S-3	.....	.-...R	.R...E.....	.Q....M	...L....
pCB2S-4	.....	.-...R	.R...E.....	.Q....M	...L....
pCB2S-5	.....	.-...R	.R...E.....	.Q....M	...L....
pCB2S-6	.....	.-...R	.R...E.....	.Q....M	...L....

Fig. 18. Comparison of the amino acid and nucleotide sequences between the 6th and 7th Cys residues of different isoforms of the BN2S, MP2S and CB2S.

A. Amino acid sequences (derived from cDNA clones); one letter AA codes are used ; the sequences have been aligned to maximize homology; and the two Met-rich regions are boxed.

b. Nucleotide sequences. B3, B6, B9, B1 and B2 represent BN2S cDNA clones of pHS-3, pHS-6, pHS-9, and genomic clones, BE2S1 and BE2S2, respectively; M1, M56 and M2 represent MP2S cDNA clones pMP2S-1, pMP2S-56, and pMP2S-2, respectively; C1 to C6 represent CB2S cDNA clones pCB2S-1 to 6. The two methionine rich regions were boxed.

B.


6th
7th
B3 TGCAGATGCGAAGGCTTAAGG ATGATGATGATGAGGATG CAACAGGAGGAGATGCAA
B6 .....|.G.....G..T....|.....A.....
B9 .....|.....G..T....|.....C.....
B1 .....|.....|.....
B2 .....T.....|.....T...G....---|....G.....G.G
M1 .....|.....G..T..T.|....G.....---C..
M56 .....|.....G..T..T.|....G.....---C..
M2 .....T.....G.|.....---....G.|....G..C.....G.G
C1 .....|.....---....G.|....G.....G..
C2 .....|.....---....G.|....G.....G..
C3 .....|.....---....G.|....G.....G..
C4 .....|.....---....G.|....G.....G..
C5 .....|.....---....G.|....G.....G..
C6 .....|.....---....G.|....G.....G..
B3 CCCCGAGGGGAGCAG ATGCGAAGGATGATGAGGCTG GCCGAGAATATCCCTTCCCGCTGC
B6 .....|.....|.....
B9 .....|.....T.....A..|.....C.....
B1 .....|.....|.....
B2 .T..AG..C.....|....AG.....A.....AA.|.....CC...TC.....
M1 .....AG.....|.A..A.....A..|..C.....A.....
M56 .....AG.....|.A..A.....A..|..C.....C.....
M2 .TT.AG.GC.....|G...AG.....A..|..C....GCC...T.....
C1 .....C.....|.....A..|..T.....C.....
C2 .....C.....|.....A..|..T.....C.....
C3 .....C.....|....A.....A..|..C.....C.....
C4 .....C.....|....A.....A..|..C.....C.....
C5 .....C.....|....A.....A..|..C.....C.....
C6 .....C.....|....A.....A..|..C.....C.....

resulting in a change from AGG (Arg codon) to ATG (Met codon). This finding suggests that point mutation, especially G to T transition, might play an important role in the accumulation of Met codons in the 2S SRP genes during evolution.

In the three 2S SRPs studied, Met residues are found mainly in two Met-rich regions, which in turn are located in a variable region. These findings reveal that the mutations leading to the accumulation of Met codons in this region were tolerated during evolution of the SRP gene in the Brazil nut family. The high content of Met residues in the SRPs is in contrast with the molecular theory on Met that favors the decrease of Met codons in a gene during evolution (Ohta & Kimura, 1971). Also, the accumulation of a large amount of SRP in seeds is an energy-demanding process. The 2S SRPs in the Brazil nut family, therefore, must have certain specific function(s), besides as the source of carbon and nitrogen, which render them advantageous in the evolutionary selection process. This speculation can be further supported by the fact that the Met-richer 2S protein subfamily (pMP2S-1, pMP2S-56 and pHS-3) instead of the Met-poorer subfamily (pMP2S-2 and BE2S2) is the dominant protein fraction in the seeds, even though they grow under the stress of sulfur deficiency in the soil (Sanchez *et al*, 1982). However, more research is needed to understand the function(s) of the 2S SRP in the Brazil nut family.

CHAPTER IV

SEQUENCE MODIFICATION OF THE MP2S GENE AND ACCUMULATION OF THE MODIFIED MP2S IN TRANSGENIC TOBACCO SEEDS

INTRODUCTION

Through the transfer and expression of the Met-rich BN2S gene in transgenic tobacco seeds, Altenbach *et al* (1989 and 1992b) demonstrated that it is feasible to use this approach to increase the Met content in seed proteins by up to 30%. Although large quantities (3 to 8% as total seed protein) of the Met-rich BN2S were synthesized and accumulated in the transgenic tobacco seeds, the total amount of 2S proteins related to the 7S and 11S proteins did not increase, thus suggesting that the expression of the BN2S gene resulted in a reduction of other 2S proteins in the seeds (Altenbach *et al*, 1989). Therefore, it may not be a good approach to further increase the Met content by enhancing the expression of the BN2S gene, as over expression of the BN2S gene might lead to further reduction of other 2S seed proteins which could be functionally important. Modification of the candidate genes to contain more Met codons might be a preferable strategy to modulate the Met content in transgenic plants.

To test this approach, the cDNA encoding the MP2S will be mutagenized to contain more Met codons and the stability

of the modified proteins will be studied in transgenic tobacco seeds.

MATERIALS AND METHODS

Chemicals. All chemicals used (Fisher Scientific or otherwise noted) were reagent grade or better.

Plant materials. Seeds of tobacco (*Nicotiana tabacum* cv Xanthi.nc) were sterilized by treatment with 10% SDS and 10% Chlorox for 30 min. After rinsing with sterile water three times, the sterilized seeds were germinated on 1/2 MSO medium (0.5X macro- and 1X micro-salts of Murashige and Skoog salts, 0.1 g/L *myo*-inositol, 0.4 mg/L thiamine·HCl, 0.5 mg/L nicotinic acid, 0.5 mg/L pyridoxine·HCl, and 2 g/L phytagel, Sigma) and then grown in a growth chamber (Percival I-37L) with continuous light at 28°C. These plants were used as explant materials for transformation. Some of these plants were transferred into soil, grown in a greenhouse and used as control plants for comparison.

In the transformation experiments, the rooted putative transgenic plants were transplanted from growth chambers into soil and grown in a greenhouse. Leaves from transgenic tobacco plants were collected for genomic DNA preparation. Developing seeds from transgenic tobaccos were collected at 10, 15 and 20 days after flowering (DAF) and stored at -70°C for RNA isolation. Mature seeds were collected for protein isolation.

Bacterial strains. Plasmid DNAs were manipulated mostly in *E. coli* DH5α (BRL). For site-directed mutagenesis, *E. coli* CJ236 (*dut*⁻ and *ung*⁻) and MV1190 (Bio-Rad) were used. *E. coli* HB101 containing the helper plasmid pRK2013 and *Agrobacterium tumefacien* LBA4404 (Clontech) were used for plant transformation.

Site-directed mutagenesis. A 600 bp *Hinf*I/*Eco*RI fragment of pMP2S-1 containing the MP2S coding region and 3' UTR was filled-in and ligated with a *Xba* I digested/filled-in/dephosphorylated pTZ19U. The resulting subclone of pMP2S-600 was used for site-directed mutagenesis by following the protocol of the *Muta-Gene Phagemid in vitro Mutagenesis* kit (Bio-Rad). An oligomer, pGGAGATGCAACCTCGAGAGAAGCA, was used to create a *Xho*I site between two Met-rich regions (Box I and Box II). Another oligomer, pGCCGAGAATATTCCTTCCC, was used to create a *Ssp*I site downstream of Box II as well as to destroy a *Hinf*I site in the coding region downstream of Box II. The clone with *Xho*I and *Ssp*I sites was named pMP2S-603.

Plasmid construction. The 440-bp *Hinf*I/filled-in fragment of pMP2S-603, which contains the cDNA fragment between the 7th bp upstream of the translation start codon and the translation stop codon, was subcloned into a *Xba* I/filled-in pHT812 (*Pst* I⁻, *Xho*I⁻ and *Ssp*I⁻, derived from pUC19 by Tu, personal communication) and named pMP2S-440. A 5-kb fragment of the *Pst* I/*Xho*I pMP2S-440 was isolated and ligated with a mixture of two double stranded synthetic oligomers, Ia

and Ib (see texture below). This subcloning resulted in two mutants, pWNIa and pWNIb, with Box I replaced by the Ia and Ib synthetic oligomer, respectively. A 5-kb fragment of the *Xho*I/*Ssp*I pMP2S-440 was isolated and ligated with a mixture of two synthetic oligomers, IIa and IIb (see texture below). This subcloning resulted in two mutants, pWNIIa and pWNIIb, with Box II replaced by the IIa and IIb synthetic oligomer, respectively. The 150-bp *Xho*I/*Hind*III pWNIIa or pWNIIb was ligated to a 3-kb *Xho*I/*Hind*III fragments of pWNIa or pWNIb, resulting in four mutants, pWNIIIaa, pWNIIIab, pWNIIIba and pWNIIIbb, with Boxes I and II replaced by the respective oligomers.

The 440-bp *Hinf*I/filled-in fragments from the pMP2S-1, the pWNI, the pWNII and the pWNIII series were ligated with a 5-kb *Acc*I/filled-in/dephosphorylated fragment of phaseolin genomic clone (pD3-8 in pTZ19U) containing the phaseolin promoter and terminator (Altenbach *et al*, 1989). These recombinations resulted in chimeric genes which contain the coding regions of the unmodified wild type or modified MP2S genes between the promoter and terminator of the phaseolin gene. These chimeric genes were designed pWNPM1-39, pWNPIa-2, pWNPIb-8, pWNPIIa-1, pWNPIIb-7, pWNPIIIaa-8, pWNPIIIab-12, pWNPIIIba-3, and pWNPIIIbb-0 according to the replacement of Box I, Box II or both Boxes.

A *Hind*III site was generated by inserting a *Hind*III linker into the *Sma*I site on the pTZ-19U vector of pWNPM1-39

and pWNPI, II and III series, resulting in two *Hind*III sites flanking the chimeric genes, phaseolin promoter/unmodified or modified MP2S coding region/phaseolin terminator. The resulting clones were designated pWNHM1-39-2, pWNHIa-2-5, pWNHIb-8-4, pWNHIIa-2-1, pWNHIIb-7-1, pWNHIIIaa-8-2, pWNHIIIab-12-3, pWNHIIIba-3-1, and pWNHIIIbb-0-6. The 2.8-kb *Hind*III fragments of pWNHM1 and pWNHI, II and III clones containing the chimeric genes were subcloned into the single *Hind*III cloning site on the binary vector pBI121 (Clontech, Jefferson *et al*, 1987). The resulting clones are designated pPhas/MPM1-39-2-34, pPhas/MPIa-2-5-2, pPhas/MPIb-8-4-5, pPhas/MPIIa-2-1-9, pPhas/MPIIb-7-1-2, pPhas/MPIIIaa-8-2-2, pPhas/MPIIIab-12-3-2,pPhas/MPIIIba-3-1-4,pPhas/MPIIIbb-0-6-2. The insertion of the chimeric genes is located between the NPTII and GUS genes on the pBI121 vector.

Triparental mating between <u>Agrobacterium</u> and <u>E. coli</u>. This method was originally described by Margossian (ARCO, PCRI) and modified by Sun (personal communication). The donor cells, DH5α harboring the pBI121 or the pPhas/MP2S series, and the helper, HB101 harboring pRK2013, were grown in LB with 50 mg/L of kanamycin at 30°C. The recipient *Agrobacterium*, LBA4404, was grown at 30 °C in Min A/sucrose medium (10.5 g/L K_2HPO_4, 4.5 g/L K_2HPO_4, 1 g/L $(NH_4)_2SO_4$, 0.5 g/L Na-citrate, 15 g/L Difco agar, 1 mM $MgSO_4 \cdot 7H_2O$ and 0.2% sucrose). Overnight cultures (late log phase, O.D. 660=1.2 to 1.8) from single colonies of the donor, helper and recipient

cells were harvested by centrifugation at 4,000 rpm for 5 min and resuspended in drug free media to an equivalent $O.D._{600}$. Mating was performed by mixing 200 μl each of the 3 cell cultures and incubating the mixtures at RT for at least 15 min. Two 150 μl aliquots of the mating mixture were spotted onto two pieces of 0.45 μm nitrocellulose filters (2 X 2 cm^2, MSI) which were placed on dry LB plates. Each plate also contained a filter with the donor cells alone and a filter with the recipient cells alone. After drying at RT, the plates were incubated at 30°C overnight with the cover side up. The *Agrobacterium* cells were suspended in 2 ml of 0.9% NaCl by vortexing vigorously. The suspensions were diluted with 0.9% NaCl, and the 10^{-5}, 10^{-6}, and 10^{-7} dilutions were plated on ABS plates (0.013 g/L $CaCl_2 \cdot 2H_2O$, 0.31 g/L $MgSO_4 \cdot 7H_2O$, 2.5 mg/L $FeSO_4 \cdot 7H_2O$, 5 g/L sucrose, 20 g/L Difco agar, 3.93 g/L $K_2HPO_4 \cdot 3H_2O$, 1 g/L NaH_2PO_4, 1 g/L NH_4Cl, and 1.5 g/L KCl) containing 50 mg/L kanamycin. The 10^{-7} and 10^{-8} dilutions were also plated on LB plates without antibiotics for viable counting. The colonies of LBA4404 containing pBI121 or the Phas/MP2S series appeared after a 2 to 3 day incubation whereas *E. coli* colonies were formed only after 1 day incubation. In general, the colony counts on the LB plates were 10 to 40 times higher than those on the ABS/kanamycin plates. The LBA4404 colonies were re-streaked twice and then verified by boiling mini plasmid preparation and restriction digestion for the presence of the vector.

Tobacco transformation and regeneration. The method originally reported by Horsch *et al* (1985) and modified by Chen and Kunhle (personal communication) was used. A single colony of LBA4404 harboring pBI121 or the derivatives was cultured in 5 ml of LB containing 50 mg/L kanamycin at 28°C with shaking (220 rpm). On the second day, 2 μl of 0.5 M acetosyringone (in DMSO) was added and the culture was allowed to incubated under the same condition for another day. The culture (1 to 8 X 10^9 cfu/ml) was then diluted 10 times with LB. Tobacco leaves from 6 to 8 week old *in vitro* grown plants were cut into 0.5 X 1 cm^2 pieces, submerged in the bacterial suspension, blotted on sterile paper towel, and then placed on 1/2MSO plates. After co-cultivated at room temperature in the dark for 2 days, the explants were transferred to 1/2MS104 plates (1/2MSO with 1 mg/L N^6-benzyladenine and 0.1 mg/L α-naphthaleneacetic acid) containing ampicillin (500 mg/L) and kanamycin (300 mg/L). The plates were incubated in the growth chamber (see Plant materials) and the explants were transferred to fresh media every 3 weeks. After 6 to 8 weeks, the apices (with 2 to 3 expanded leaves) of the regenerated shoots were transferred to 1/2MSO medium with 500 mg/L ampicillin and 100 mg/L kanamycin in Magenta boxes. The shoots were transferred again to 1/2MSO with 100 mg/L kanamycin but no ampicillin. Rooted plantlets without bacterial contamination were then transplanted to 4" pots with soil (soil:vermiculite:perlite

= 3:1:1), covered with transparent plastic cups, and grown in a greenhouse. After 2 to 3 days, the plastic cups were removed. About 10 days later, the plants were transplanted to 1 gallon pots. The plants were fertilized on the second day and about a month after transferring to the 1 gallon pots.

GUS assay. GUS assay was performed following the standard procedure. Leaves from *in vitro* grown plants were cut into approximately 0.3 X 0.3 cm^2 pieces. Two to 3 pieces from each plant were placed into a white sterile microfuge tube containing GUS staining buffer (100 mM Na-phosphate, pH 7, 0.1% Triton X-100, 0.5 mM K-ferricyanide ($K_3 \cdot Fe(CN)_6$), 0.5 mM K-ferrocyanide ($K_4 \cdot Fe(CN)_6$), 0.05% X-glucuronide (X-gluc, dissolved in DMF first), and 1 mM EDTA) and incubated at 37°C. After overnight incubation, 300 μl of 100% ethanol was added to eliminate the interference of chlorophyll.

DNA isolation. Genomic DNA was isolated from tobacco leaves, using a CTAB protocol (Doyle *et al*, 1989) with modifications. The steps of DNA treatment and with RNase and the precipitation with ethanol were omitted. The DNA was suspended in 300 μl TE. RNase A treatment was performed during restriction enzyme digestion. DNA concentrations were determined with a fluorometer (Hoefer Scientific Instruments, Model TKO 100).

Southern blots. Ten micrograms of genomic DNA were digested with *Hind*III, separated on 0.7% agarose gels and

then transferred to "Hybond-N+" membrane (Amersham). Hybridization was performed according to the method of Church and Gilbert (1984). The 440-bp *Xba*I fragment of the pMP2S-440 plasmid DNA was labeled with [^{32}P]-dCTP by random priming (see Chapter III) and used as a probe.

RNA isolation. RNA from developing seeds was isolated by the method of Altenbach *et al* (1989). Seeds harvested 10, 15 and 20 DAF from each plant were used.

Northern blots. Five μg of total RNA isolated from the tobacco seeds were separated on 1.5% agarose/formaldehyde gels. The blotting and hybridization were the same as described in Chapter III. The 440-bp *Xba*I fragment of the pMP2S-440 plasmid DNA was labeled with [^{32}P]-dCTP by random priming (see Chapter III) and used as a probe.

Protein extraction and SDS-PAGE. Total seed protein was extracted from tobacco seeds and separated by 20% SDS-PAGE following the methods described in Chapter III.

Western blots. Total protein was separated by a 20% SDS-PAGE gel, electrophoretically transferred to nitrocellulose filter (Schleicher & Schuell, Inc.) and probed with an monoclonal antibody against the large subunit of the BN2S protein. Immunodetection was performed by following the protocol of ECL Western Blotting (Amersham).

RESULTS

Sequence Modification of MP2S.

Generation of *XhoI* and *SspI* sites. Sequence analysis revealed that there is an existing *Pst*I site upstream of Box I. Between Boxes I and II, there is the sequence CCCGAG that could easily be changed to CTCGAG (*Xho*I site) by site-directed mutagenesis. Likewise, downstream of Box II, there is the sequence AATCTC that could be modified to AATATT (*Ssp*I site). The generation of the *Ssp*I site would also eliminate a *Hinf*I site in the coding sequence so that the MP2S coding sequence can be excised by its two flanking *Hinf*I sites. A 24-mer oligonucleotide was synthesized and used to generate the *Xho*I site, while a 19 oligomer was for the *SspI* site. Because the 24-mer oligomer was designed based on the sequence of the BN2S cDNA, it contained 5 NTs that were different from the template DNA of the MP2S (Fig. 19A). Between the 19 oligomer and the template DNA, there were 2 NT differences (Fig. 19B). By decreasing the annealing temperature, increasing the reaction volume and using a single strand DNA binding protein (D20) in the reaction, the *XhoI* and *SspI* sites were successfully generated in the MP2S sequence (Fig. 20 A and B).

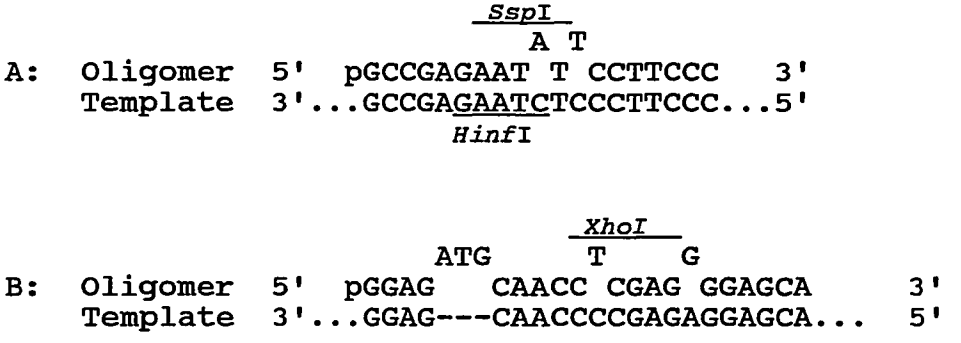


Fig. 19. Comparison of the oligomers and the template DNAs. The restriction sites present in the oligomers are indicated above the sequences. The restriction site presents in the template DNA is indicated under the sequences.

A, 19-mer: two nucleotides difference between the 19mer and the template DNA,

B, 24-mer: two nucleotides difference between the 24mer and the template DNA and three extra nucleotides in the oligomer.

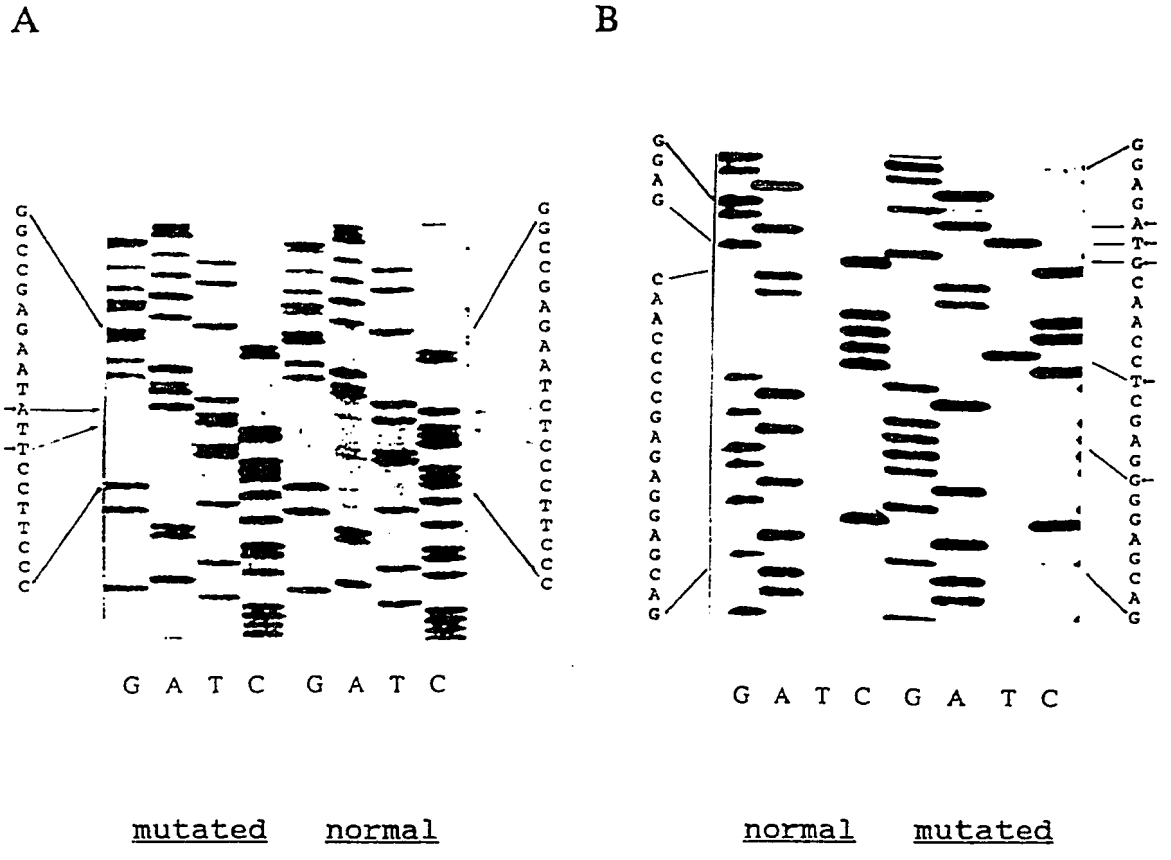


Fig. 20. DNA sequences of the mutated cDNAs.
The mutated nucleotides are marked with arrows.
A: Mutagenesis of C to A and C to T using the 19mer.
B: Mutagenesis of C to T and A to G and an insertion of ATG using the 24mer.

Replacement of the Boxes I and II with synthetic himet fragments.

After engineering *XhoI* site between Boxes I and II and the *SspI* site downstream of Box II, Box I is flanked by unique sites, *Pst*I and *Xho*I, and Box II by *Xho*I and *SspI*. Therefore, Box I and/or Box II can easily be cassette-replaced with DNA fragments containing the same restriction sites.

The regions of Boxes I and II were replaced with Met-rich (himet) DNA fragments. Four himet DNA fragments Ia, Ib, IIa, and IIb, as shown below, were used.

Ia C R C E G L R M M M R M M M M M M Q Q E M M Q P R

Ib C R C E G L R M M M R M M M M R M Q Q E M M Q P R

IIa P R G E M M R M M M Q L A E N I P

IIb P R G E R M R M M M Q L A E N I P

These fragments were designed with a major consideration of preserving the secondary structure of the 2S protein. The Peptide Sort and Plot structure program of the Genetics Computer Group (GCG) was used to monitor the possible change in the secondary structure of the protein as a result of sequence modifications.

The Ia or Ib was used to replace Box I, while the IIa or IIb, Box II. These replacements generate eight possible

combinations and the plasmids containing the genes with these modifications are designated pWNIa, pWNIb, pWNIIa, pWNIIb, pWNIIIaa, pWNIIIab, pWNIIIba, and pWNIIIbb. The AA sequences of the region between the 6th and 7th Cys residues containing these modified MP2S genes are showed in Fig. 21. Besides the increases in Met residues, changes in other AA residue also occur in the region of the modified genes (Table 8). Some of these changes are due to the difference in sequences between the BN2S and MP2S, while others resulted from the creation of restriction enzyme sites for the sequence replacements and for the confirmation of the replacements.

In the wild type MP2S, the Box I and II regions contain 31% (5/16) and 19% (3/16) of the total Met in the mature protein, respectively. The mutant Ia and Ib fragments will add 6 and 5 Met residues in Box I, respectively. The IIa and the IIb fragments will increase the Box II Met content by 2 and 1, respectively. An extra Met residue will be gained in Box I as a result of the creation of a *Xho*I site in the region. By replacing Box I and II with the corresponding himet DNA fragments, Ia, Ib, IIa, or IIb, an increase of 2 to 8 Met residues in these regions will be achieved and a change in AA composition ranging from 10 to 17 AA residues (29 to 50% of the total AA) in this region will occur.

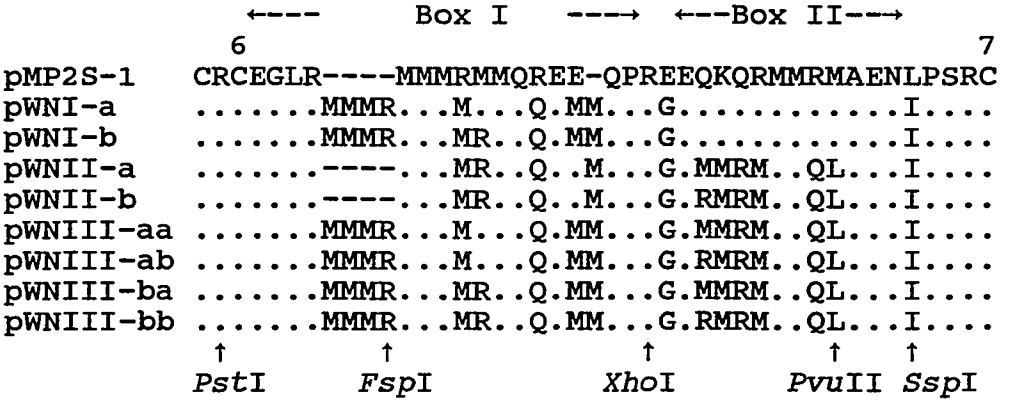


Fig. 21. Comparison of the amino acid sequences between the 6th and 7th Cys residues of the wild type and the modified MP2S.

One letter AA code is used in this comparison. Restriction enzymes *Pst*I, *Xho*I and *Ssp*I can be used for Boxes I and II sequence replacement and *Fsp*I and *Pvu*II for confirming the replacement.

Table 8. Methionine Content and Amino Acid Changes in The Region Between the 6th and 7th Cys Residues

	Box I		Box II		Box I+II		TTL AA* Changed	
	# of Met	% ↑	# of Met	% ↑	# of Met	% ↑	# of AA	%
pMP2S-1	5	0	3	0	8	0	0	0
pWNIa	11	120					10	29
pWNIb	10	100					11	32
pWNIIa	6	20	5	67	11	38	12	35
pWNIIb	6	20	4	33	10	25	12	35
pWNIIIaa					16	100	16	47
pWNIIIab					15	88	16	47
pWNIIIba					15	88	17	50
pWNIIIbb					14	75	17	50

*: # of AA between #6 and #7 cysteine residues is 34.

Tobacco transformation and regeneration.

To test the expression and accumulation of the modified MP2S proteins in transgenic seeds, three of the modified MP2S genes under control of the phaseolin promoter and terminator, Phas/MPIb-2, Phas/MPIIa-4 and Phas/MPIIIba-1, were used to transform tobacco plants. They were modified at Box I for 5, Box II for 3 and both boxes for 7 additional Met codons, resulting in a 31%, 18.8% and 43.8% increase in the Met content over the normal MP2S protein, respectively. The AA residues altered in the region between the 6th and 7th Cys residues are 32, 35 and 50%, respectively, of the total AA's of this region. The wild type MP2S gene under control of the same promoter, Phas/MPM1-4, and a pBI121 vector were also transformed to serve as controls.

A total of 138 shoots were regenerated from 343 leaf explants after selection on antibiotic (kanamycin) media (Table 9). Fifty to 65 % of these shoots showed positive GUS staining. Most of the GUS positive and some of the GUS negative plants were transferred to the greenhouse. The plants started to flower after about two months in the greenhouse.

Generally speaking, the morphology of most of the transformants was similar to that of the untransformed tobacco plants. Some transformants were smaller than other untransformed and transformed plants. Two plants, # 4 of pBI121 vector transformants and # 3 of the Phaseo/MPIIIba

transformants, had smaller leaves and did not form seeds. One Phas/MPIIa transformant grew for 4 months and died without flowering. These abnormalities are possible since the insertion of foreign DNA fragments into the genome may interrupt genes controlling normal development of the plants.

Table 9. Summary of the Tobacco Transformation

Construct	Explant	Shoot	GUS^+	In Green-house
pBI121	40	8	4	4
Phas/ MPM1-4	84	43	24	20
Phas/ MPIb-2	68	29	15	11
Phas/ MPIIa-4	61	31	20	18
Phas/ MPIIIba-1	90	28	17	15

Transformation using tobacco stems as explants was also performed. However, no shoot was recovered although calli appeared on the stems earlier than on the leaf discs. It is probable that the media used for regeneration of plants from leaves are not suitable for stems.

Copy number of transgenes in transgenic plants.

The integration of the chimeric genes into the tobacco genome was confirmed by Southern blot analysis of DNA isolated from tobacco leaves. Using the coding sequence of MP2S cDNA as a probe, a *Hind*III fragment of the expected size (2.8 kb) was found to hybridize with genomic DNA from most of the GUS positive plants, indicating the presence of the phas/MPM1 chimeric genes in their genomes (Fig. 22; Table 10). Neither the non-transformed tobacco nor pBI121 vector transformants showed any DNA fragment hybridizing with the probe. However, some of the transformants showed extra DNA fragments, in addition to the 2.8-kb DNA band, that hybridized with the probe. A GUS negative transformant, Phas/MPIb-1, did not have the 2.8-kb band, but had two bands of 3.5 and 4.3 kb (Fig. 23). These observations suggest that DNA rearrangements occurred in these plants.

Southern analysis revealed that most of the transformants contained a single copy of the Phas/MP2S chimeric gene (Table 10). However, some of the transformants

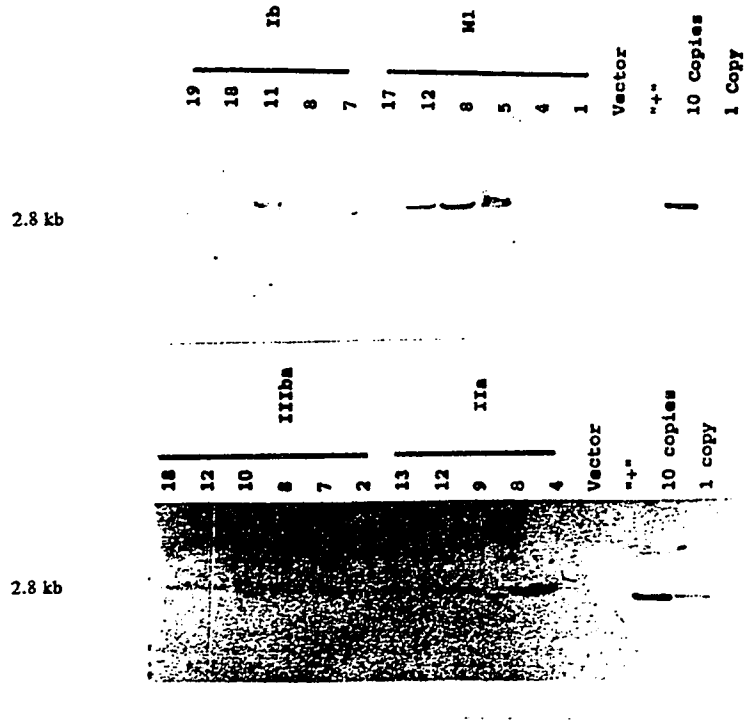


Fig. 22. Southern Blot Analysis of Transgenic Tobacco Genomic DNA.

Genomic DNA isolated from tobacco leaves was digested with *Hind*III, separated in 0.7% agarose/TBE gel, transferred to nitrocellulose filter and hybridized with a probe containing the coding region of MP2S. M1, Phas/MPM1 transformants; Ib, Phas/MPIb transformants; IIa, Phas/MPIIa transformants; IIIba, Phas/MPIIIba transformants; "+", wild type; and vector, pBI121 vector transformant.

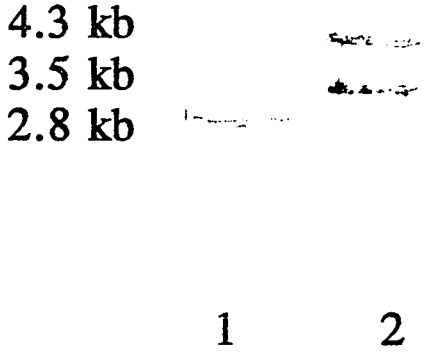


Fig. 23. Southern Blot Analysis of Transformant Phas/MPIb-1. Genomic DNA from leaves of transformant Phas/MPIb-1 (lane 2) was digested with *Hind*III and hybridized with the labelled coding region of the pMP2S-1. The plasmid DNA of Phas/MPIb-2 digested with the same enzyme was used as a control (lane 1).

Table 10. Copy Number of The Wild Type or Modified MP2S Genes in Transgenic Tobacco Plants

Plants	Copy #	Extra* Bands	Plants	Copy #	Extra Bands
Phas/MPM1-1	3	+	Phas/MPIIa-1	3	-
4	1	-	4	1	-
5	1	-	6	1	-
6	1	+	8	2	-
7	1	-	9	5	-
8	2	+	10	1	-
12	1	-	12	1	-
13	1	-	13	4	-
14	1	-	16	2	-
15	1	-	18	1	-
16	1	-			
17	1	-			
18	1	-			
33	1	-			
Phas/MPIb-1	0	+	Phas/MPIIIba-2	1	-
7	2	+	3	5	-
8	1	-	7	1	+
9	1	-	8	1	-
11	1	-	10	3	-
13	1	-	12	1	-
14	1	-	16	1	-
18	2	+	18	4	-
19	1	-	23	1	-
			24	1	-

* Expected size of the transgene on a *Hind*III fragment is 2.8 kb; extra bands indicate other DNA fragments of different lengths hybridizing to the gene probe.

contain multiple copies of the Phas/MP2S genes. For example, Phas/MPM1-1 contained 3 copies of the chimeric gene, while Phas/MPIb-7 and -18 have 2 copies in their genomes. The highest copy number, *i.e.* 5 copies, was found in the genomes of Phas/MPIIa-9 and Phas/MPIIIba-3.

Expression of the Phas/MP2S chimeric constructs in transgenic tobacco.

Total RNA was prepared from the immature seeds of transgenic tobacco plants with single copy of the chimeric gene. As expected, Northern blot analysis revealed a mRNA species of 0.7 kb in the transformants which hybridized to the MP2S coding sequence probe (Fig. 24). No RNA from the untransformed tobacco or the pBI121 transformants hybridized to the probe. The level of the 0.7-kb mRNA varied in different transformants. Transformants Phas/MPM1-8 (Fig. 24, lane 4), Phas/MPIb-9 (Fig. 24, lane 6) and Phas/MPIIIba-7 (Fig. 24, lane 9) contain relatively low levels of this mRNA in their 15-DAF-old seeds, while at the same stage, higher levels of this mRNA were observed in seeds of Phas/MPM1-13 (Fig. 24, lane 3), Phas/MPIb-10 (Fig. 24, lane 5), Phas/MPIIa-9 and -10 (Fig. 24, lane 7 and 8) and Phas/MPIIIba-10 (Fig. 24, lane 10).

Fig 25 shows the expression of the Phas/MP2S chimeric genes at different stages of seed development. MP2S mRNA in

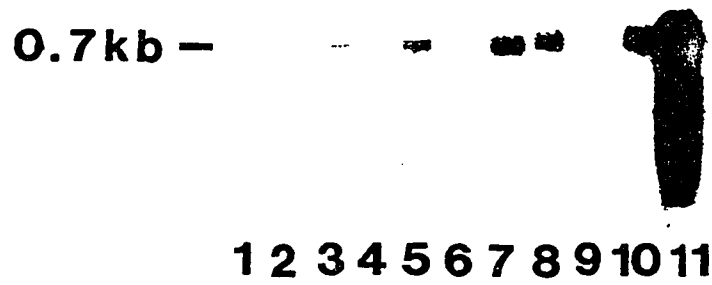


Fig. 24. Comparison of mRNA level in 15-DAF seeds of transgenic tobaccos.

Two micrograms total RNA from 15 DAF seeds of Phas/MPM1-12 and -8 (lanes 3 and 4), Phas/MPIb-19 and -18 (lanes 5 and 6), Phas/MPIIa-10 and -9 (lane 7 and 8), and Phas/MPIIIba-8 and -10 (lanes 9 and 10), respectively, were separated by electrophoresis, Northern blotted, and hybridized to the labelled coding region of the pMP2S-1. Lane 11 was 2 μg of total RNA from immature cannonball seeds. The first two lanes contained 2 μg total RNA from the seeds of untransformed and pBI121 transformed plants, respectively.

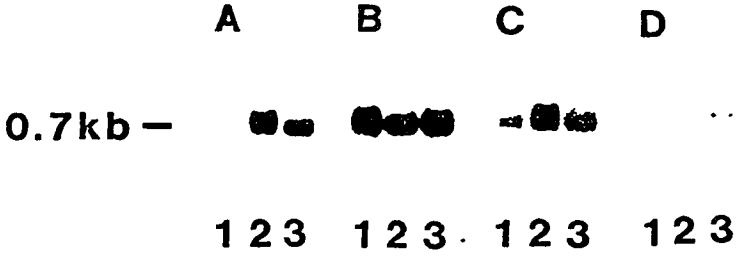


Fig. 25. Northern Blot Analysis of mRNAs from developing tobacco seeds.

Two micrograms of RNA isolated from developing seeds at 10, 15, and 20 DAF (lane 1, 2, and 3, respectively) were separated by electrophoresis, Northern blotted, and hybridized to the labelled coding region of the pMP2S-1. Panel A: Phas/MPM1-12; B, Phas/MPIb-19; C, Phas/MPIIa-10; and D, Phas/MPIIIba-8 transformants.

the seeds of Phas/MPM1-12 could be detected at 10 DAF; reached the highest level at 15 DAF; and decreased at 20DAF (Fig. 25 A). A similar expression pattern could be observed for Phas/MPIIa-10 (Fig. 25 C). However, the mRNA remained at similar levels during the period of 10 to 20 DAF in the Phas/MPIb-19 transformant (Fig. 25 B), while the mRNA could not be detected until 20 DAF in the Phas/MPIIIba-9 transformant (Phas/MPIIIba-8, Fig. 25 D).

Accumulation of the MP2S and modified MP2S proteins in tobacco seeds.

Tobacco plants transformed with the wild type and modified MP2S genes were analyzed for accumulation of their protein products. Total seed protein was extracted from the mature seeds of GUS positive transformed plants and analyzed by Western blot/immunodetection. The monoclonal antibody against the 9-kD subunit of the BN2S was used for the reaction. Results revealed that a 9 kD polypeptide could be detected in seeds from plants transformed with the wild type (Phas/MPM1) and modified (Phas/MPIb, Phas/MPIIa, and Phas/MPIIIba) MP2S genes (Fig. 26). No protein could be detected by the antibody in the seeds of untransformed or pBI121-transformed plants. The quantity of immunoreactive protein accumulated in seeds varied in different transformants containing the same transgene (Fig. 27) or different transgenes (Fig. 26). By comparison with known

Fig. 26. Detection of the unmodified and the modified MP2S proteins in the seeds from transgenic tobacco.

Ten micrograms of total protein extracted from mature seeds were separated in a 20% SDS-PAGE gel, electrophoretically transferred to nitrocellulose filter, and immunoreacted with a monoclonal antibody against the 9-kD subunit of the BN2S. Lanes a, b, and c, contained 1, 0.5, and 0.1 μg purified BN2S protein, respectively; lanes d and e, Phas/MPIIIba-23 and -10; lanes f and g, Phas/MPIIa-10 and -9; lanes h and i, Phas/MPIb-7 and -9; lanes j and k, Phas/MPM1-5 and -4; lane l, pBI121 vector control transformant; lane m, untransformed tobacco.

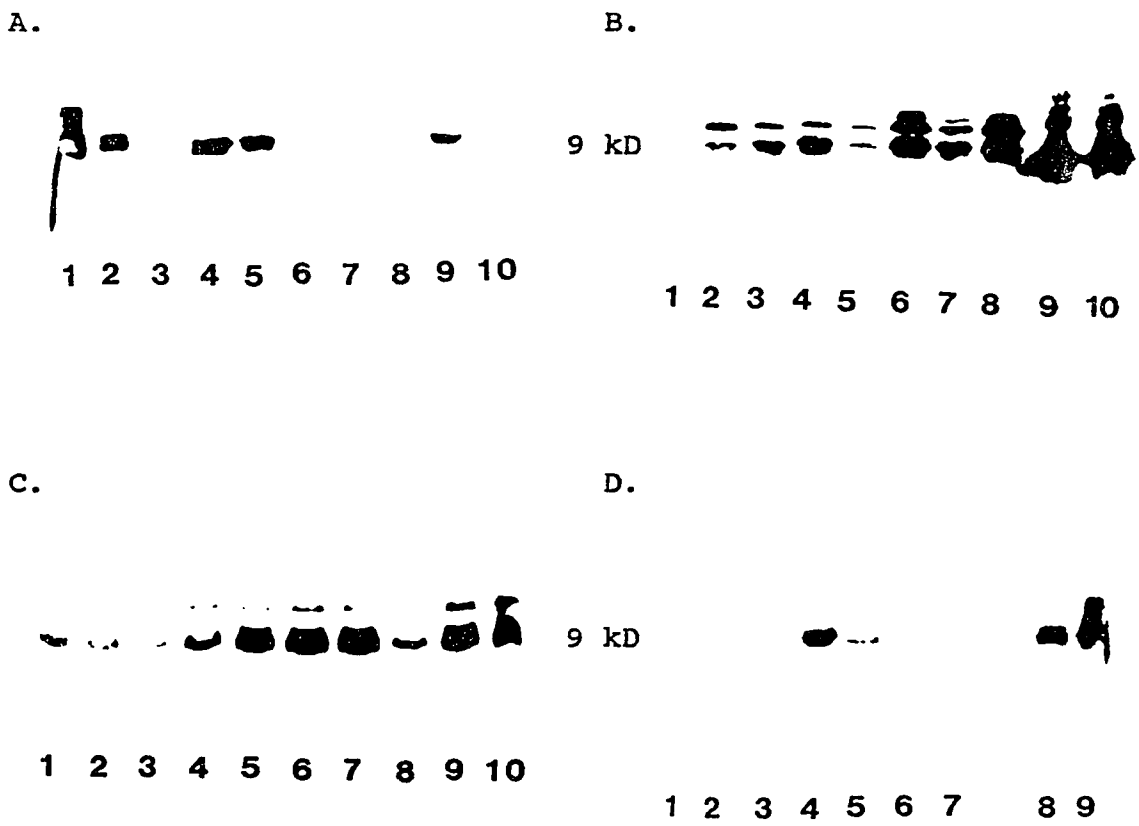


Fig. 27. Variation in the accumulation of the MP2S protein in transgenic tobacco seeds.

Ten micrograms of total seed protein were used for the comparison. Panel A: Phas/MPM1 transformants. Lane 1, 0.5 μg BN2S; 10, untransformed tobacco; 2 to 9, transformants # 16, 14, 13, 12, 8, 5, and 4, respectively. B, Phas/MPIb transformants. Lanes 1, untransformed tobacco; 9 and 10, 0.5 and 1 μg BN2S, respectively; 2 to 8, transformants # 6, 7, 8, 9, 11, 14, and 16, respectively. C, Phas/MPIIa transformants. Lanes 1 to 9, transformants # 1, 4, 8, 9, 10, 13, 16, 18, and 19, respectively; 10, 0.5 μg BN2S. D, Phas/MPIb transformants. Lanes 1, untransformed tobacco; 2 to 7, transformants # 2, 7, 10, 12, 16, and 23; 8 and 9, 0.1 and 0.5 μg BN2S, respectively.

amounts of purified BN2S on Western blots, it was estimated that the amounts of the immunodetectable 9-kDpolypeptides varied from trace amount to 1% in transformants of the unmodified MP2S gene (Fig. 27 A), trace to 3% in the Phas/MPIb transformants (Fig. 27 B), 0.1 to 5% in the Phas/MPIIa transformants (Fig. 27 C), and trace to 1% in the Phas/MPIIIba transformants (Fig. 27 D).

DISCUSSION

The sequence modifications in MP2S result in up to 47% AA alternation in the region between the 6th and the 7th Cys residues. The modified 2S proteins can accumulate at levels comparable to the unmodified wild type 2S protein, (Fig. 26 and 27), indicating that the modifications do not destabilize the protein. These results further confirm that the variable region is a suitable site for modification of 2S proteins.

The immunodetectable polypeptide in plants transformed with the unmodified wild type or modified MP2S genes co-migrates with the 9-kD subunit of BN2S, indicating that the protein products of the transferred genes include a 9-kD polypeptide, which in turn suggests that the 18-kD precursor of MP2S can be correctly processed in the transgenic tobacco seeds into its mature 9-kD and 3-kD subunit polypeptides. Earlier studies have shown the precursor of the Brazil nut sulfur-rich 2S protein can be correctly processed to its mature subunits in tobacco seeds (Altenbach *et al*, 1989), canola seeds (Altenbach *et al*, 1992), and potato leaves, stems (Sun *et al*, 1992) and tubers (Tu and Sun, personal communication). Here we show that, in addition to the untransformed wild type MP2S, the modified forms of MP2S can also be correctly processed in tobacco seeds. The sequence modifications at the variable region between the 6th and 7th Cys residues do not appear to affect the cleavage of MP2S precursor.

It has been demonstrated that up to 8% of the total seed protein was the BN2S when the phaseolin promoter was used to regulate the expression of the BN2S gene in tobacco (Altenbach *et al*, 1989). In this study, a 5% accumulation of MP2S was detected for Phas/MPIIa-10, -13, -16 and -19 transformants (Fig. 27, lanes 5, 6, 7, and 9). This expression level is lower than that observed for the BN2S in tobacco, but we have not exhaustively searched for transgenic plants with higher expression levels. Variations in protein accumulation in transgenic tobacco seeds were observed when the Phas/BN2S chimeric gene was expressed in tobacco seeds (Altenbach *et al* (1989). These variations are most likely due to differences in the transcription rate of the transgenes involved. The copy number and chromosomal integration position of the transgene may play a key role in these observed variations as well.

Variations in the accumulation of unmodified wild type and modified MP2S proteins in transgenic tobacco is consistent with their mRNA levels. For example, transgenic seeds of Phas/MPM1-8 (Fig. 24, lane 4), Phas/MPIb-9 (Fig. 24, lane 6) and Phas/MPIIIba-7 (Fig. 24, lane 9) contain lower levels of the 0.7-kb mRNA; and, the 9-kD protein accumulated in these seeds are also of less quantities (Fig. 27, A7; B5; and D3). Since significant amounts of 2S protein, and its mRNA, were present in wild type as well as modified MP2S transformants, and there were corresponding variations in the

level of 2S protein and its mRNA between the wild type and modified MP2S transformants, it is unlikely that the modifications affected the stability of the protein products or their mRNAs. In summary, the results from this study indicate that the modified MP2S containing additional residues of Met can stably accumulate in a model plant system. Therefore, both the unmodified wild type and the Met-enriched MP2S protein genes can be used for plant protein quality improvement. The modified gene encoding a protein with 23% Met and 8% Cys would be a good candidate for this improvement.

CHAPTER V

CONCLUSION

In this Ph.D dissertation research, I have isolated and characterized 2S SRPs from two members of the Brazil nut family, namely cannonball and monkeypot; cloned the cDNAs encoding these 2S SRPs; constructed eight Met-enriched 2S protein (himet) genes through sequence modifications of one of the clones; and successfully transferred three of the modified genes into tobacco and studied their expression. The following conclusions can be drawn:

1. Plants in the Brazil nut family contain the sulfur-rich 2S protein as their major seed storage proteins. The 2S proteins can account for up to 50% of the total seed protein. The methionine content of the 2S protein in the Brazil nut family ranges from 15 to 19%.

2. The sulfur-rich 2S proteins from different family members share high degrees of similarity in their subunit structure, amino acid composition, synthesis, processing, immunoreactivity, and amino acid and nucleotide sequences.

3. The patterns of the 2S protein cleavage in different members of the Brazil nut family are very similar. However, the precursors of BN2S and MP2S undergo three stepwise processing to form their mature subunits (18 kD → 15 kD → 12 kD → 9 +3 kD), while the precursor of CB2S undergoes only two steps (18 kD → 15 kD → 9 + 3 kD).

4. Data from cDNA cloning and sequence analysis indicate that the 2S protein in the Brazil nut family is encoded by a multigene family. Based on their AA sequences, the 2S protein gene family can be classified into two subfamilies.

5. The high degree of homology in both the AA (over 80%) and NT (over 90%) sequences among the 2S proteins suggests that their genes might have originated from a common ancestral gene. Point mutations in the variable region between the 6th and 7th Cys residues might be an important mechanism for the accumulation of Met codons in the 2S protein genes.

6. Eight genes encoding modified MP2S with 18 to 24% Met residues have been obtained by engineering the variable region between the 6th and 7th Cys residues.

7. Tobacco plants can correctly process the precursor of the MP2S into its mature subunits.

8. The sequence modifications in the variable region of the MP2S do not affect the stability and processing of the protein products. The modified himet MP2S genes could serve as alternative transgene candidates for protein quality improvement through molecular transformation.

REFERENCES

Alexander DC, McKnight TD and William BG (1984) An efficient vector-primer cDNA cloning system. Gene, 31:79-89

Altenbach SB, Pearson KW, Leung FW and Sun SSM (1986) The stepwise processing of a sulfur-rich seed protein from Brazil nut (*Bertholletia excelsa* H.B.K.). in Shannon,LM and Chrispeels, MJ(ed): Molecular Biology of seed storage proteins and lectins. Waverly Press, Baltimore, Maryland, 137-146

Altenbach SB, Pearson KW, Leung FW and Sun SSM (1987) Cloning and sequence analysis of a cDNA encoding a Brazil nut protein exceptionally rich in methionine. Plant Mol. Biol. 8:239-250

Altenbach SB, Pearson KW, Meeker G, Staraci C and Sun SSM (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 Mol. Biol., 13:513-522

Altenbach SB, Pearson KW, and Sun SSM (1992a) Nucleotide sequences of cDNAs encoding two members of the Brazil nut methionine-rich 2S albumin gene family. Plant Physiol., 98:1520-1522

Altenbach SB, Kuo C-C, Staraci LC, Pearson KW, Wainwright C, Georgescu A and Townsend J (1992b) Accumulation of a Brazil nut albumin in seeds of transgenic canola results in enhanced levels of seed protein methionine. Plant Mol. Biol. 18:235-245

Ampe C, Van Damme J, de Castro JAB, Sampaio MAM, Van Montagu M, and Vandekehove J (1986) The amino-acid sequence of the 2S sulfur-rich proteins from seeds of Brazil nut (*Bertholletia excelsa* H.B.K.). Eur. J. Biochem, 159:597-600

Antunes AJ and Markakis P (1977) Protein supplementation of navy beans with Brazil nuts. J. Agric. Food Chem., 25(5):1096-1098

An G (1986) Development of plant promotor expression vectors and their use for analysis of differential activity of nopaline synthetase promoter in transformed tobacco cells. Plant Physiol., 81:86-91

Aykroyd WR and Doughty J (1964) Legume in human nutrition. FAO, Rome

Badennoch-Jones J, Spencer D, Higgins TJV and Millerd A (1981) The role of glycosylation in storage-protein synthesis in developing pea seeds. Planta, 153:201-209

Basha S-M and Pancholy SK (1981) Identification of methionine-rich polypeptide in peanut (*Arachis hypogaea* L.) seed. J. Agric. Food Chem., 29:331-335

Barta A, Sommergruber K, Thompson D, Harmuth K, Matzke MA and Matzke AJM (1986) The expression of a nopaline synthetase-human growth hormone chimeric gene in transformed tobacco and sunflower callus tissue. Plant Mol. Biol., 6:347-357

Beachy RN, Chen Z-L, Horsch RB, Rogers SG, Hoffman NJ and Fraley RT (1985) Accumulation and assembly of soybean β-conglycinin in seeds of transformed petunia plants. EMBO J., 4:3047-3053

Bednarek SY, Wilkins TA, Dombrowski JE and Raikel NV (1990) A carboxyl-terminal propeptide is necessary for proper sorting of barley lectin to vacuoles of tobacco. Plant Cell, 2:1145-1155

Benner MS, Phillips RL, Kirihara JA and Messing JW (1989) Genetic-analysis of methionine-rich storage protein accumulation in maize. Theor. Appl. Gen., 78(6):761-767

Bliss FA and Hall TC (1977) Food legumes-compositional and nutritional Changes induced by breeding. Cereal Food World, 22(3): 106-113

Bollini R, Vitale A and Chrispeel MJ (1983) *In vivo* and *in vitro* processing of seed reserve protein in the endoplasmic reticulum: evidence for two glycosylation steps. J. Cell Biol., 96:999-1007

Bray EA, Naito S, Pan N-S, Anderson E Dube and Beachy RN (1987) Expression of the β-subunit of β-conglycinin in seeds of transgenic plants. Planta, 172:364-370

Bright SWJ and Shewry PR (1983) Improvement of protein quality in cereals. CRC Critical Rev. in Plant Sci., 1:49-93

Brinegar AC and Peterson DM (1982) Synthesis of oat glubulin precursors. Analogy to legume 11S storage protein synthesis. Plant Physiol., 70:1767-171769

Brown JWS, Ma Y, Bliss FA and Hall TC (1981) genetic variation in ten subunits of globulin-1 storage protein of French bean. Theor. Appl. Genet., 59:83-88

Burr B, Burr FA, Rubenstein I and Simon MN (1978) Purification and translation of zein messenger RNA from maize endosperm protein bodies. Proc. Natl. Acad. Sci. USA, 75:696-700

Burr FA and Burr B (1981) *In vitro* uptake and processing of prezein and other maize preproteins by maize membranes. J. Cell Biol., 90:427-434

Casey R (1979) Genetic variability in the structure of the α-subunits of legumin from *Pisum*. A two dimensional gel electrophoresis study. Heredity, 43:265-272

Castro LAB de, Lacerda Z, Aramoyo RA, Sampaio AM and Gander ES (1987) Evidence for a precursor molecule of Brazil nut 2S seed proteins from biosynthesis and cDNA analysis. Mol. Gen. Genet., 206:338-343

Chee PP, Klassy RC and Slightom JL (1986) Expression of a bean storage protein 'phaseolin minigene' in foreign plant tissue. Gene 41:47-57

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

Chrispeel MJ, Higgins TJV and Spencer D (1982) Assembly of storage protein oligomers in the endoplasmic reticulum and processing of the polypeptides in the protein bodies of developing pea cotyledons. J. Cell Biol., 93:306-313

Church GM and Gilbert W (1984) Genomic sequencing. PNAS, U.S.A., 81:1191-1195.

Colot V, Robert LS, Kavanagh TA, Bevan MW and Thompson RD (1987) Localization of sequences in wheat endosperm protein genes which confer tissue-specific expression in tobacco. EMBO J., 6:3559-4564

Crouch ML and Sussex IM (1981) Development and storage-protein synthesis in *Brassica napus* L. embryos *in vivo* and *in vitro*. Planta, 153:64-74

Crouch ML, Tenbarge KM, SImon AM and Ferl R (1983) cDNA clones for *Brassica napus* seed storage proteins: evidence from nucleotide sequence analysis that both subunits of napin are cleaved from a precursor polypeptide. J. Mol. Appl. Genetics, 2:273-283

Croy RRD, Lycett GW, Gatehause JA, Yarwood JN and Boulter D (1982) Cloning and analysis of cDNAs encoding plant storage protein precursors. Nature, 295:76-79

De Clercq A, Vandewiele M, De Rycke R, Van Damme J, Van Montagu M, Krebbers E and Vandekerckhove (1990) Expression and processing of an *Arabidopsis* 2S albumin in transgenic tobacco. Plant Physiol., 92:899-907

Dickinson CD, Scott MP, Hussein EHA, Argos P, and Nielsen NC (1990) Effect of structural modification on the assembly of a glycinin subunit. Plant Cell 2:403-413

Doyle JF, Schuler MA, Godette WD, Zenger V and Beachy RN (1986) The glycosylated seed storage proteins of *Glycine max* and *Phaseolus vugaris*. J. Biol. Chem., 261(20):9228-9238

Doyle JJ, Doyle JL and Hortoium LHB (1989) Isolation of plant DNA from fresh tissue. Focus 12(1):13-15

Dure LS III and Chlan C (1981) Developmental biochemistry of cottonseed embryogenesis and germination. XII. Purification and properties of principal storage proteins. Plant Physiol., 68:180-186

Ellis JR, Shirsat AH, Hepher A, Yarwood JN, Gatehouse JA, Croy RRD and Boulter D (1988) Tissue-specific expression of a legumin gene in seeds of *Nicotiana plumbaginifolia*. Plant Mol. Biol., 10:203-214

Ericson ML, Rodin J, Leenman M, Glimelius K, Josefsson LG and Rask L (1986) Structure of the rapeseed 1.7 S storage protein, napin, and its precursor. J. Biol. Chem., 261(31):14576-14581

Evans AM, and Gridley HE (1980), in Smith H (ed) : Commentaries in plant science, Vol 2., Pregamo, New York, 143-159

FAO (1970) Amino acid content of foods, in FAO nutritional studies, No.24. Rome

Fraley RT (1986) Genetic transformation in higher plants. Critical Rev. Plant Sci., 4(1):1-46

Frayley RT and Papahadjopoulas D (1982) Liposomes: The development of a new carrier system for introducing nucleic acids into plant and animal cells. Curr. Tech. Microbiol., 96:171-191

Galau GA, Wang, HYC and Hughes DW, (1992) Cotton *Mat*5-A (C164) gene and *Mat*5-D cDNAs encoding methionine-rich 2S albumin storage proteins. Plant Physiol., 99: 779-782

Gander ES, Holmstroem K-O, De Paiva GR, De Castro LAB, Carneiro M, and Grossi de Sa M-F (1991) Isolation, characterization and expression of a gene coding for a 2S albumin from *Bertholletia excelsa* (Brazil nut). Plant Mol. Biol., 16:437-448

Gasser CS and Frayley RT (1989) Genetically engineering plants for crop improvement. Science, 244:1293-1299

Gayler KR, Kolivas SK, KacFarlane AJ, Lilley GG, Baldi M, Blagrove J, and Johnson ED (1990) Biosynthesis, cDNA and amino acid sequence of a precursor of conglutin δ, a sulfur-rich protein from *Lupinus angustifolius*. Plant Mol. Biol., 15:879-893

George AA and de Lumen BO (1991) A novel methionine-rich protein in soybean seed:identification, amino acid composition, and N-terminal sequence. J. Agric. Food Chem., 39:224-227

Gepts P and Bliss FA (1984) Enhanced available methionine concentration associated with higher phaseolin levels in common bean seeds. Theor. Appl. Genet., 69:47-53

Ghoshal AK and Farber E (1984) The induction of liver cancer by dietary deficiency of choline and methionine without added carcinogens. Carcinogenesis, 5(10): 1367-1370

Giannazza E, Viglienghi V, Righetti PG, Salamini F and Soave C (1977) Amino acid composition of zein molecular components. Phytochemistry, 16:315-317

Goldsbrough PB, Gelvin SB and Larkins BA (1986) Expression of zein genes in transformed sunflower cells. Mol. Gen. Genet., 202:374-381

Gonzalez E (1991) The C-terminal domain of plant catalases-Implications for a glyoxysome targeting sequence. Eur. J. Biochem., 199:211-215

Greenwood JS and Chrispeels MJ (1985) Correct targeting of the bean storage protein phaseolin in the seeds of transformed tobacco. Plant Physiol.,79:65-71

Hagen G and Rubenstein I (1981) Complex organization of zein genes in maize. Gene, 13:239-249

Hiatt A, Cafferkey R and Bowdish K (1989) Production of antibodies in transgenic plants. Nature, 342:76-78

Higgins TJV and Spencer D (1977) Cell-free synthesis of pea seed proteins. Plant Physiol., 60:655-661

Higgins TJV and Spencer D (1981) Precursor forms of pea vicilin subunits:Modification by microsomal membranes during cell-free translation. Plant Physiol., 67:205-211

Higgins TJV, Chandler PM, Randall PJ, Spencer D, Beach LR, Blagrove RJ, Kortt AA, and Inglis AS (1986) Gene structure, protein structure, and regulation of the synthesis of a sulfur-rich protein in pea seeds. J. Biol. Chem., 261(24):1124-1130

Higgins TJV (1984) Synthesis and regulation of major proteins in seeds. Ann. Rev. Plant Physiol., 35:191-221

Hoffman LM, Donaldson DD, Bookland R, Rashka K and Herman EM (1987) Synthesis and protein body deposition of maize 15-kd zein in transgenic tobacco seed. EMBO J., 6(11):3213-3221

Hoffman LM, Donaldson DD and Herman EM (1988) A modified storage protein is synthesized, processed, and degraded in the seeds of transgenic plants. Plant Mol. Bio. 11:717-729

Horsch RB, Fry JE, Hoffermann NL, Eichholtz D, Rogers SG, and Fraley RT (1985) A simple and general method for transferring genes into plants. Science, 227-1229-1231

Hunt DC and Chrispeel MJ (1991) The signal peptide of a vacuolar protein is neccessary and sufficient for the efficient secretion of a cytosolic protein. Plant Physiol., 96:18-25

Hunt AG, Chu NM, Odell JT, Nagy F and Chua NH (1987) Plant cells do not properly recognize animal gene polyadenylation sequences. Plant Mol. Biol., 8:23-35

Jefferson RA, Kavanaugh TA, and Bevan MW (1987) GUS fusion: β-glucuronidase as a sensitive and versatile gene fusion marker in higher plant. EMBO J., 6:3901

Joshi CP, (1987) Putative polyadenylation signals in nuclear genes of higher plants: a compilation and analysis. Nuc. Acids R., 15(23):9627-9640

Kachlan N and Richardson M (1981) The complete amino acid sequence of a major wheat protein inhibitor of α-amylase. Phytochem., 20:1781-1784

Kirihara JA, Petri JB and Messing J (1988a) Isolation and sequence of a gene encoding a methionine-rich 10-kDa zein protein from maize. Gene, 71:359-370

Kirihara JA, Hunsperger JP, Mahoney WC and Messing J (1988b) Differential expression of a gene for a methionine-rich storage protein in maize. Mol. Gen. Genet., 211:477-484

Koie B and Doll H (1979) in: Seed protein improvement in cereals and grain legumes. International Atomic Energy Agency, Vienna, Vol 1, pp. 205-214

Koncz C, Kreuzaler K, KAlman ZS and Schell J (1984) A simple method to transfer, integrate and study expression of foreign gene, such as chicken ovalbumin and alphaactin in plant tumors. EMBO J., 3:1029-1037

Kortt AA and Caldwell JB (1990) Low molecular weight albumins from sunflower seed: identication of a methionine-rich albumin. Phytochemistry, 29(9):2805-2810

Kortt AA, Caldwell JB, Lilley GG and Higgins TJV (1991) Amino acid and cDNA sequences of a methionine-rich 2S protein from sunflower seed (*Helianthus annuus* L.). Eur. J. Biochem. 193: 329-334

Kozak, M (1984) Nucleic Acids Res 12:857-872

Krebbers E, Herdies L, De Clercq A, Seurink J, Leemans J, Vandamm J, Segura M, Gheysen G, Van Montagu M and Vandekerckhove J (1988) Determination of the processing sites of an *Arabidopsis* 2S albumin and characterization of the complete gene family. Plant Physiol., 87:859-866

Krochko JE, Prmanik SK, and Bewley JD (1992) Contrasting storage protein synthesis and messenger RNA accumulation during development of zygotic and somatic embryos of alfafa (*Medicago sativa* L.). Plant Physiol., 99(1): 46-53

Lawrence MC, Suzuki E, Varghese, Davis PC, Van Donkelaar, Tulloch PA and Colman PM (1990) The three-dimensional structure of the seed storage protein phaseolin at 3 A° resolution. EMBO J., 9(1): 9-15

Larkins BA and Hurkman WJ (1978) Synthesis and deposition of zein in protein bodies of maize endosperm. Plant Physiol., 62:256-263

Lawton MA, Tiemey MA, Nakamura I, Anderson E and Komeda Y (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 Mol. Biol., 9:315-324

Leite A, Yunes JA, Turcinelli SR, and Arruda P (1992) Cloning and characterization of a cDNA encoding a sulfur-rich coxin. Plant Mol. Biol., 18:171-174

Lonnerdal BO and Janson JC, (1972) Studies on *Brassica* seed proteins I. the low molecular weight proteins in rapeseed: isolation and characterization. Biochem. Biophys. Acta, 278:175-183

Mak AS and Jones BL (1976) The amino acid sequence of wheat β-purothionin. Can. J. Biochem., 22:835-842

Masummura T, Shibata D, Hibino T, Kato T, Kawabe K, Takeba G, Tanaka K and Fujii S (1989) cDNA cloning of an mRNA encoding a sulfur-rich 10 kDa prolamin polypeptide in rice seeds. Plant Mol. Biol. 12:123-130

Matthews JA, Brown Jws and Hall TC (1981) Phaseolin mRNA is translated to yield glycosylated polypeptides in *Xenopus oocytes*. Nature, 294:175-176

Matthews JA and Miflin BJ (1980) *In vitro* synthesis of barley storage proteins. Planta, 149:262-268

Matzke MA, Susani M, Binns AN Lewis ED Rubenstein I and Matzke AJM (1984) Transcription of a zein gene introduced into sunflower using a Ti plasmid vector. EMBO J., 3:1525-1531

Melcher U and Fraij B (1980) methionine-rich protein fraction prepared by cryoprecipitation from extracts of corn meal. J. Agric. Food Chem., 28:1134-1136

Mertz ET, Bates LS and Nelson OE (1964) Mutant gene that change protein composition and increase lysine content of maize endosperm. Science, 145:279-280

Monsalve RI and Rodriguez R (1990) Purification and characterization of proteins from the 2S fraction from seeds of the *Brassicaceae* family. J. Exp. Botany, 41(222):89-94

Moreira MA, Hermodson MA, Larkins BA and Nielson NC (1979) Partial characterization of the acidic and basic polypeptides of glycinin. J. Biol. Chem, 254:9921-9926

Murai N, Sutton DW, Murray MG, Slightom JL, Merlo DJ, Reichert NA, Sengupta-Gopalan C, Stock CA, Barker RF, Kemp JD and Hall TM (1983) Phaseolin gene from bean is expressed after transfer to sunflower via tumor-inducing plasmid vectors. Science, 222:476-482

Naren AP and Virupaksha TK (1990) Alpha- and beta-setarins: methionine-rich proteins of Italian millet (*Setaria italica* (L.) Beauv.). Cereal Chem., 67(1):32-34

Neuhaus J-M, Sticher L, Meins F and Boller T (1991) A short C-terminal sequence is necessary and sufficient for the targeting of chitinases to the plant vacuole. PNAS, USA, 88:10362-10366

Odani S and Ikenaka T (1972) Studies on soybean trypsin inhibitor: IV. Complete amino acid sequence and the anti-proteinase sites of Bowman-birk soybean proteinase inhibitor. J. Biochem., 71:839-848

Odell JT, Nagy F and Chua NH (1985) Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter. Nature, 313(28):810-812

Ohta T and Kimura M (1971) Amino acid composition of protein as a product of molecular evolution. Science, 174:150-153

Ohtani T, Galili G, Wallace JC, Thompson GA and Larkins BA (1991) Normal and lysine-containing zeins are unstable in transgenic tobacco. Plant Mol. Biol., 16:117-128

Oke OL (1980-81) Toxicity of cyanogenic glucosides. Food Chem., 6:97-109

Osborn TC and Bliss FA (1985) Effects of genetically removing lectin seed protein on horticultural and seed characteristics of common bean. J. Am. Soc. Hort. Sci., 110:484-488

Osborne TB (1908) Our present knowledge of plant proteins. Science, 28:417-427

Payne PI (1983) Breeding for protein quality and protein quality in seed crop. in Daussant J, Mosse J and Vaughan J(ed): Seed proteins. Academic Press, London. pp. 223-252

Pedersen K, Argos P, Naravana SVL, and Larkins BA (1986) Sequence analysis and characterization of a maize gene encoding a high-sulphur zein protein of Mr 15,000. J. Biol. Chem.,261(14): 6279-6284

Pedersen K, Devereux J, Wilson D, Sheldon E and Larkins BA (1982) Cloning and sequence analysis reveal structural variation among related zein genes in maize. Cell 29:1015-1026

Phillips RL, Morris PR, Wold F, and Gengebach BG (1981) Seedling screening for lysine-plus-threonine resistant maize. Crop Sci., 21:601-607

Phillips RL and MeClure BA (1985) Elevated protein-bound methionine in seeds of a maize line resistant to lysine plus threonine. Cereal Chem., 62(3):213-218

Ponnappanaren A and Virupaksha TK (1990) α- and β-setarins: methionine-rich proteins of Italian millet (*Setaria italica*(L.) Beauv.). Cereal Chem, 67:32-34

Prance GT and Mori SA (1979) Flora neotropics, monograph No.21 *Lecythidaceae*-Part I. The New York Botanical Garden, New York, PP.1 & 107

Robert LM and Lord JM (1981) The synthesis of *Ricinus communis* agglutinin: cotranslational and post-translational modification of agglutinin polypeptides. Eur. J. Biochem., 199:31-41

Rodin J and Rask L (1990) Characterization of matteuccin, the 2S storage protein of the ostrich fern. Evolutionary relationship to angiospern seed storage proteins. Eur. J. Biochem, 192:101-107

Romero AJ, Yandell BS and Bliss FA (1986) Bean arcelin 1: inheritance of a novel seed protein of *Phaselus vugaris* L. and its effect on seed composition. Theor. Appl. Genet., 72:123-128

Rosahl S, Schell J and Willmitzer L (1987) Expression of a tuber-specific storage protein in transgenic tobacco pants: demonstration of an esterase activity. EMBO J., 6:1155--11159

Sanchez PA, Bandy DE, Villachica JH and Nicholaides JJ (1982) Amazon basin: management for continuous crop production. Science, 216:821-827

Schuler MA, Ladin BF, Polacco JC, Freyer G, and Beachy RN (1982) Structural sequences are conserved in the genes coding for the alpha, alpha' and β-subunits of the soybean 7S seed storage protein. Nucleic Acids Res., 10:8245-8261

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

Sengupta C, Deluca V, Bailey DS and Verma DPS (1981) Plant Mol. Biol. 1:42-49

Sengupta-Gopalan C, Reichert NA, Barker RF, Hall TC and Kemp JD (1985) Developmentally regulated expression of the bean β-phaseolin gene in tobacco seed. Proc. Natl. Acad. USA, 82:3320-3324

Sharief FS and Li SSL (1982) Amino acid sequence of small and large subunits of seed storage protein from *Ricinus communis*. J. Biol. Chem.,257(24):14753-14759

Slightom JL, Drong RF, Klassy RC and Hoffman LM (1985) Nucleotide sequences from phaseolin cDNA clones: the major storage proteins from *Phaseolus vulgaris* are encoded by two unique gene families. Nucleic Acids Research, 13(18):6483-6498

Spena A and Schell J (1987) The expression of a heat-inducible chimeric gene in transgenic tobacco plants. Mol. Gen. Genet., 206:436-440

Stranger ED (1984) Oxidation of methionine in model system. J. Agric. Food Chem., 32:358-363

Sturm A, Voelker TA, Herman EM and Chrispeel MJ (1988) Correct glycosylation, Golgi-processing, and targeting to protein bodies of the vacuolar protein phytohemagglutinin in transgenic tobacco. Planta, 175:170-183

Sun, S.M., Slightom, J.L., & Hall, T.C. (1981) Intervening sequences in a plant gene-comparison of the partial sequence of cDNA and genomic DNA of French bean phaseolin. Nature, 289(5793):37-41.

Sun SSM, Leung FW and Tomic JC (1987a) Brazil nut (Bertholletia excelsa H.B.K.) protein: fraction, composition, and identification of a sulfur-rich protein. J. Agric. Food Chem., Vol. 35(2): 232-235

Sun SSM, Altenbach SB, and Leung FW (1987b) Properties, biosynthesis and processing of a sulfur-rich protein in Brazil nut (*Bertholletia excelsa* H.B.K.). Eur. J. Biochem. 162:477-483

Sun SSM, Altenbach SB, Leung FW, Meeker LG and Staraci L (1988) Quality improvement of crop proteins through genetic engineering. in Chang et al (ed): Recent Advance in Biotechnology and Applied Biology, The Chinese University Press, Hong Kong, 621-628

Sun SSM, Zuo W-N, and Tu HM (1992) Molecular approaches for enhancing amino acid quality of plant proteins. Pro. 1992 Penn State Sym. in Plant Physiol., May, 28-30. in press.

Tao S-H, Fox MRS, Fry BE, John ML, Lee YH, Tomic JC and Sun SSM (1987) methionine bioavailability of a sulfur-rich protein from Brazil nut. Fed. Proc., 46:891

Templeman TS, Stein DS and DeMaggio AE (1988) A fern storage protein is genetically similar to the 1.7S storage protein of *Brassica napus*. Biochemical Genetics, 26(9/10):595-603

Thompson AJ, Evans IM, Boulter D, Croy RRD and Gatehouse JA (1989) Transcriptional and posttranscriptional regulation of seed storage-protein gene expression in pea (*Pisum sativum* L.). Planta, 179:279-287

Thomson JA and Schroeder HE (1978) Cotyledonary storage proteins in *Pisum sativum* II. Hereditary variation in components of the legumin and vicilin reactions. Aust. J. Plant Physiol., 5:281-294

Tollenaa M (1989) Genetic improvement in grain-yield of commercial maize hybrids grown in Ontario from 1959 to 1989. Crop Sci.,29(6):1365-1371

Ueng P, Galili G, Sapanara V, Goldsbrough PB, Dube P, Beachy RN and Larkins B (1988) Expression of a maize storage protein gene in petunia plants is not restricted to seeds. Plant Physiol., 86:1281-1285

Utsumi S (1989) Studies on improvement of quality of seed storage proteins by gene manipulation. Nippon Nogeikagaku Kaishi, 63(9): 1471-1478

Van Santen VL and Spritz RA (1987) Splicing of plant pre-mRNA in animal system and *vice versa*. Gene, 56:253-265

Vandekerckhove J, Van Damme J, Van Lijsebettens M, Botterman J, De Block M, Vandewiele M, De Clercq A, Leemans J, Van Montagu M and Krebbers E (1989) Enkephalin produced in transgenic plants using modified 2S seed storage proteins. Biotechnology, 7:929-932

Viotti A, Abildsten D, Pogna N, Sala E and Pirrotta V (1982) Multiplicity and diversity of cloned zein cDNA sequences and their chromosome location. EMBO J., 1:53-58

Viotti A, Sala E, Marotta R, Albeii P, Balducci C and Soave C (1979) Genes and mRNAs coding for zein polypeptides in *Zea mays*. Eur. J. Biochem., 102:211-222

Voelker T, Sturm A and Chrispeels MJ (1987) Differences in expression between two seed lectin alleles obtained from normal and lectin-deficient beans are maintained in transgenic tobacco. EMBO J., 6:3571-3577

Voelker TA, Herman EM and Chrispeels MJ (1989) *In vitro* mutated phytohemagglutinin genes expressed in tobacco seeds: role of glycans in protein targeting and stability. Plant Cell, 1:95-104

Walling L, Drews GN, and Goldberg RB (1986) Transcriptional and post-transcriptional regulation of soybean seed protein mRNA levels. PNAS, USA, 83: 2123-2127

Wettstein D von (1989) Perspectives for the genetic engineering of plants for agriculture, horticulture and industry. Plant Mol. Biol., 13:313-317

Williamson JD, Galili G, Larkins BA and Gelvin SB (1988) The synthesis of a 19 kilodalton zein protein in transgenic *petunia* plants. Plant Physiol., 88:1002-1007

Wienend U and Feix G (1980) Zein specific restriction enzyme fragments of maize DNA. FEBS Lett., 116:14-16

Wilkins TA, Bednarek SY and Raikhel NV (1990) Role of propeptide glycan in post-translational processing and transport of barley lectin to vacuoles in transgenic tobacco. Plant Cell, 2:301-313

Wright DJ (1988) The seed globulins-Part II. in Hudson BJF(ed): Developments in food proteins, Vol.6, Elsevier Applied Science, 119-178

Yamaguchi M (1980) World vegetable. AVI, Westport, CT.,pp.395

Yang MS, Espinoza NO, Nagpala PG, Dodds JH, White FF, Schnorr KL and Jaynes JM (1989) Expression of a synthetic gene for improved protein quality in transformed potato plants. Plant Science, 64:99-111

Youle RJ and Huang AH (1978) Albumin storage protein in the protein bodies of Castor bean. Plant Physiol., 61:13-16

Youle RJ and Huang AH (1981) Occurrence of low molecular weight and high cysteine containing albumin storage proteins in oilseeds of diverse species. Amer. J. Bot. 68(1): 44-48