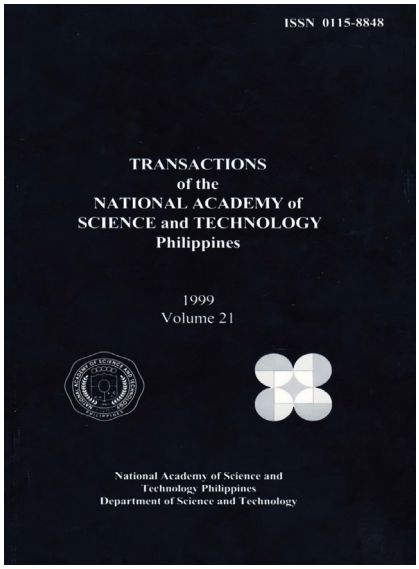
TRANSACTIONSNASTPHL

ISSN 0115-8848 (print)
ISSN 2815-2042 (online)

https://transactions.nast.ph/

Vol. 21 (1999)
doi.org/10.57043/transnastphl.1999.5773

Transactions NAST PHL, is the official journal of the National Academy of Science and Technology Philippines. It has traditionally published papers presented during the Academy's Annual Scientific Meeting since 1979 to promote science-based policy discussions of and recommendations on timely and relevant national issues as part of its functions as a national science academy. Starting in 2021, this journal has been open to contributions from the global scientific community in all fields of science and technology.



A Novel 8 KD Sulfur-Rich Protein in Soybean (Glycine Max) Cotyledon: Purification and Gene Cloning

Ma. Jamela R. Revilleza[1], Alfred Galvez[2], Deanne C. Krenz[2], Benito O. de Lumen[2], and Evelyn Mae T. Mendoza[3]

[1] Institute of Chemistry, University of the Philippines at Los Baños (UPLB), College, Laguna 4031
[2] Dept. of Nutritional Sciences, University of California, Berkeley, CA 94720-3104
[3] Institute of Plant Breeding, UPLB, College, Laguna 4031

Citation
Revilleza MJR et al. 1999. A novel 8 kD sulfur-rich protein in soybean (Glycine max) cotyledon: Purification and gene cloning. Transactions NAST PHL 21: 185-196. doi.org/10.57043/transnastphl.1999.5773

Trans. Natl. Acad. Sci. Tech. Philippines 21: 185-196 (1999).
ISSN 0115-8848

A NOVEL 8 KD SULFUR-RICH PROTEIN IN SOYBEAN (GLYCINE MAX) COTYLEDON: PURIFICATION AND GENE CLONING

MA. JAMELA R. REVILLEZA[1], ALFRED GALVEZ[2], DEANNE C. KRENZ[2], BENITO O. DE LUMEN[2] AND EVELYN MAE T. MENDOZA[3]

[1]*Institute of Chemistry*
University of the Philippines at Los Baños (UPLB)
College, Laguna 4031

[2]*Dept. of Nutritional Sciences*
University of California, Berkeley
CA 94720-3104

[3]*Institute of Plant Breeding, UPLB*
College, Laguna 4031

ABSTRACT

An 8 kD methionine-rich protein (MRP) (2D-1) was purified through a combination of selective enrichment of the albumin fraction through heparin-Sepharose affinity chromatography, 2D-SDS-PAGE and radiolabeling of methionine-containing proteins with [1-14C] iodoacetate. The MRP contains 8.6% methionine, 8.6% cysteine, 11.4% lysine and high amounts of glutamic and aspartic acids. The N-terminus sequence facilitated the synthesis of degenerate oligonucleotide primers for use in Polymerase Chain Reaction (PCR) to amplify cDNA fragments specifically coding for the gene. The PCR product was then used to screen a cDNA library and analyze the DNA sequence which confirmed the high sulfur amino content of 17.2% Southern analysis revealed the presence of two copies of the gene in the genome. This work lays down the groundwork for cloning of the gene which actually encodes the 8 kD MRP and a methionine poor 2.5 kD peptide .

Keywords: Methionine-rich proteins (MRP); soybean *(Glycine max);* Heparin-Sepharose chromatography; Polymerase Chain Reaction (PCR); N-terminus; cDNA library, Southern analysis

INTRODUCTION

Legumes have become important sources of proteins with the increase in the demand for high fiber, low cholesterol food. Their full utilization is however limited by the deficiency in the sulfur amino acids, methionine, an essential amino acid and cysteine, although a non-essential amino acid, imparts a sparing effect on the former. Soybean, the world's single largest cash crop in the US (Soya Bluebook, 1982) contains only 56% as much sulfur amino acids as the nutritionally complete Food and Agriculture (FAO) reference protein (Kelley, 1973).

Several studies have already been conducted on a number of plant crops with the objective of cloning a gene coding for a sulfur-rich protein. More often a heterologous gene was used to increase the methionine content. Unfortunately, none had been applied successfully to soybean or any legume. More disturbing is the result of a recent study indicating the allergenicity of the Brazil nut MRP gene cloned in soybean. We propose that increasing the biosynthesis of naturally occurring MRP found in very small amounts in soybean is the most feasible genetic engineering strategy to address the deficiency. An endogenous gene coming from soybean itself with a long history of non-allerginicity would be the most likely candidate for cloning. This work lays down the foundation to cloning the soybean MRP gene, thereby enhancing its expression and elucidating the possible biological role in the seed.

This paper presents the purification of a sulfur-rich protein from soybean mature cotyledon using a combination of affinity chromatography and two dimensional SDS-PAGE. The MRP was earlier identified as an 8 kD albumin (Revilleza et al., 1996). In addition, the N-terminal sequence of the protein was determined which made possible the synthesis of oligonucleotide which were used in a polymerase chain reaction (PCR) to amplify DNA fragments coding for the MRP. The PCR product was later used as probe in cDNA library screening and Southern analysis.

MRP DETECTION AND PURIFICATION

Methionine-rich proteins resolved electrophoretically and blotted on Immobilon membrane, were identified based on the selective tagging of methionine residues as reported by Kho and de Lumen (1987). MRPs were identified to be in the low molecular weight albumin fraction of mature soy coteledon. The focus then centered on this fraction which was subjected to a number of chromatographic separations. This resulted in the identification on of heparin-sepharose (affinity) chromatography which facilitated the enrichment of the 8 kD MRP prior to two-dimensional electrophoresis. The single 8 kD band on one dimensional SDS-PAGE was in fact a mixture of three 8 kD proteins where the major component was referred as 2D-I (Fig. 1). The 8 kD MRP was submitted for N-terminal microsequencing.

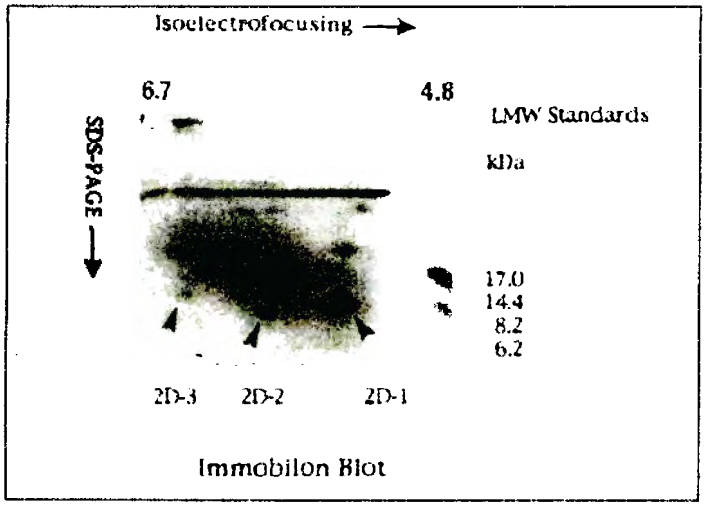


Figure. 1 2D-SDS-PAGE of soybean seed albumin methionine-rich protein enriched fraction eluted from heparin-sepharose affinity column. (Arrow points to the position of 2D-1, the major acidic protein).

cDNA LIBRARY SCREENING AND SEQUENCE ANALYSIS

Degenerate oligonucleotides deduced from the N-terminal sequence were synthesized and used as the gene specific primer for amplifying MRP gene encoding, with a mid maturation seed cDNA pool as template. PCR produced a truncated fragment which encoded the MRP. This was used as probe for Southern analysis and to fish out the cDNA clone harboring the MRP. The cDNA clone was sequenced and characterized (Galvez et al, 1997; de Lumen et al, 1999).

DISCUSSION

Purification of the MRP

The 8 kD MRP is distinct from the first MRP reported in soybean with a size of 10.8 kD and 12% methionine (George and de Lumen, 1991), although it is possible that they belong to a family of MRP. MRP has been reported in cereals and oil seeds such as Brazil nut, sunflower seed, corn and rice, but this is the second time only that a report of another MRP was found in legumes. The initial attempt to clone the 10.8 kD MRP was not successful due to the insufficient amount of protein for further analyses. In the absence of conserved regions from other known MRPs, gene cloning which encoded the soy MRP using heterologous sequences, was not feasible. Achieving the objective through the protein route, although more difficult, was a more logical solution.

Amino Acid and N-Terminal Sequence Analyses

The MRP (2D-1) contained high amounts of acidic amino acids aspartic and glutamic acids which then explained the acidic lpH on 2D-SDS-PAGE. The amino acid profile also yielded high levels of sulfur-rich proteins with 8.6% methionine and 8.6% cysteine. Lysine content was also high at 11.4% These represent the limiting amino acids in legumes and cereals respectively and are of great importance since MRPs from other sources were found to be limiting in lysine (Altenbach, 1987; Kortt et al., 1991). Table 1 gives the amino acid composition of 2D-1, defatted soy meal, Brazil nut MRP and FAO reference protein. If the Brazil nut

Table 1. Amino Acid Composition[a] of 8 kDa Soybean MRP,[b] Defatted Soymeal,[c] Brazil Nut MRP,[d] and FAO Reference Protein[e]

amino acid	8kDa-1D MRP	2D-1 MRP	Brazil nut MRP	defatted soymeal	FAO ref protein
methionine	7.7	8.6	18.8	1.4	3.5 (met + cys)
cysteine	5.0	1.5	7.9	1.3	
lysine	12.7	1.4	0.0	6.0	5.5
tryptophan	nd[f]	nd	0.0	1.2	
threonine	2.3	3.1	0.0	3.7	4.0
isoleucine	4.0	4.0	1.0	4.4	4.0
leucine	7.7	7.7	5.0	6.7	7.0
phenylalanine	1.6	0.3	0.0	4.5	6.0 (phe + tyr)
tyrosine	nd	nd	0.0	4.6	5.0
asparagine	nd	nd	2.0	nd	
aspartic acid	10.3	11.2	1.0	10.4	
serine	4.8	5.6	6.9	4.6	
glutamine	nd	nd	11.9	nd	
glutamic acid	28.8	35.1	14.9	18.4	
proline	2.9	2.8	5.9	5.3	
glycine	2.1	3.1	5.9	3.4	
alanine	2.0	2.1	1.0	3.6	
histidine	1.7	0.0	2.0	2.2	2.2
arginine	6.0	3.4	14.9	7.6	

[a]Values are in g/100 g of protein as determined by amino acid analysis. [b]Amino acid analysis was carried out on 8 kDa protein band on Immobilon membrane after 1D SDS-PAGE (8 kDa-1D) and on 2D-1 after 2D SDS-PAGE of fraction enriched for LMW proteins by heparin-Sepharose chromatography. [c]Rackis (1961). [d]Determined from cloned cDNA encoding mature protein, % of total (Altenbach and Simpson, 1990). [e]FAO/WHO Food and Agriculture Organization/World Health Organization of the United Nations (1973). [f]$_{nd}$, not detected.

MRP is to be overexpressed in a host plant, there is a possibility of the danger in causing an imbalance in essential amino acids. This is so since evidence showed that transgenic MRP increases at the expense of other proteins (Nordlee, 1994).

Protein microsequencing of the 8 kD MRP revealed the following 20 amino acid N-terminal sequence.

E G K D E D E E E E G H M Q K C A T E M

A database search using Blast showed that the N-terminal sequence yielded partial homology to glycinin in soybean and to a sucrose binding protein.

The middle N-terminal sequence of seven amino acids with the least codon degeneracy was used to design oligonucleotide primers for PCR amplification of the DNA coding for the 8 kD MRP.

PCR and Cloning of PCR Product

The PCR reactions using modified RACE protocols, yielded a 500 bp product which was cloned to GEM-T vector. The recombinant plasmid was used to transform an *E. coli* bacterial strain JM 109.

The deduced protein sequence of the 2D-1 PCR fragment which started from residue 9 of the N-terminus sequence, codes for an 8 kD protein containing 8.6 methionine and 8.6% cystein for a total sulfur content of 17.2% (Fig. 2). A stop codon was present after 211 bp. This cut short the open reading frame after 70 amino acids, leaving a large 3' untranslated region.

Southern Analysis and Library Screening

Southern hybridizations using 2D-1 PCR amplified product as probes revealed at least two bands in the autoradiograph. This implied that there could be at least two copies in the genome. Most of the genomic digests showed 2 restriction bands on samples digested with 6 base cutters and a single band on a 4 base cutter, Hinf I (Fig. 3).

A mid maturation seed cDNA library was probed with 2D-1 Ten positive clones were plaque purified after screening 6 x 10^5 cDNA clones

Characterization of the cDNA Clone Encoding the MRP

The cDNA encoding 2D-1, renamed Gm2S-1, has been completely sequenced and characterized (de Lumen, 1999). The number and position of the cysteine residues are conserved so they play a role in the formation of disulfide bridges, one of the determinants in the hierarchy of protein structure. Studies from MRP of other oilseeds suggested that the protein is synthesized as a large single-chain precursor which undergoes post-translational processing to yield 2 chain structures of the mature protein. In the case of Gm2S-1 cDNA, another low molecular weight protein (2.5 kD), was coded by the cDNA clone. Although this was

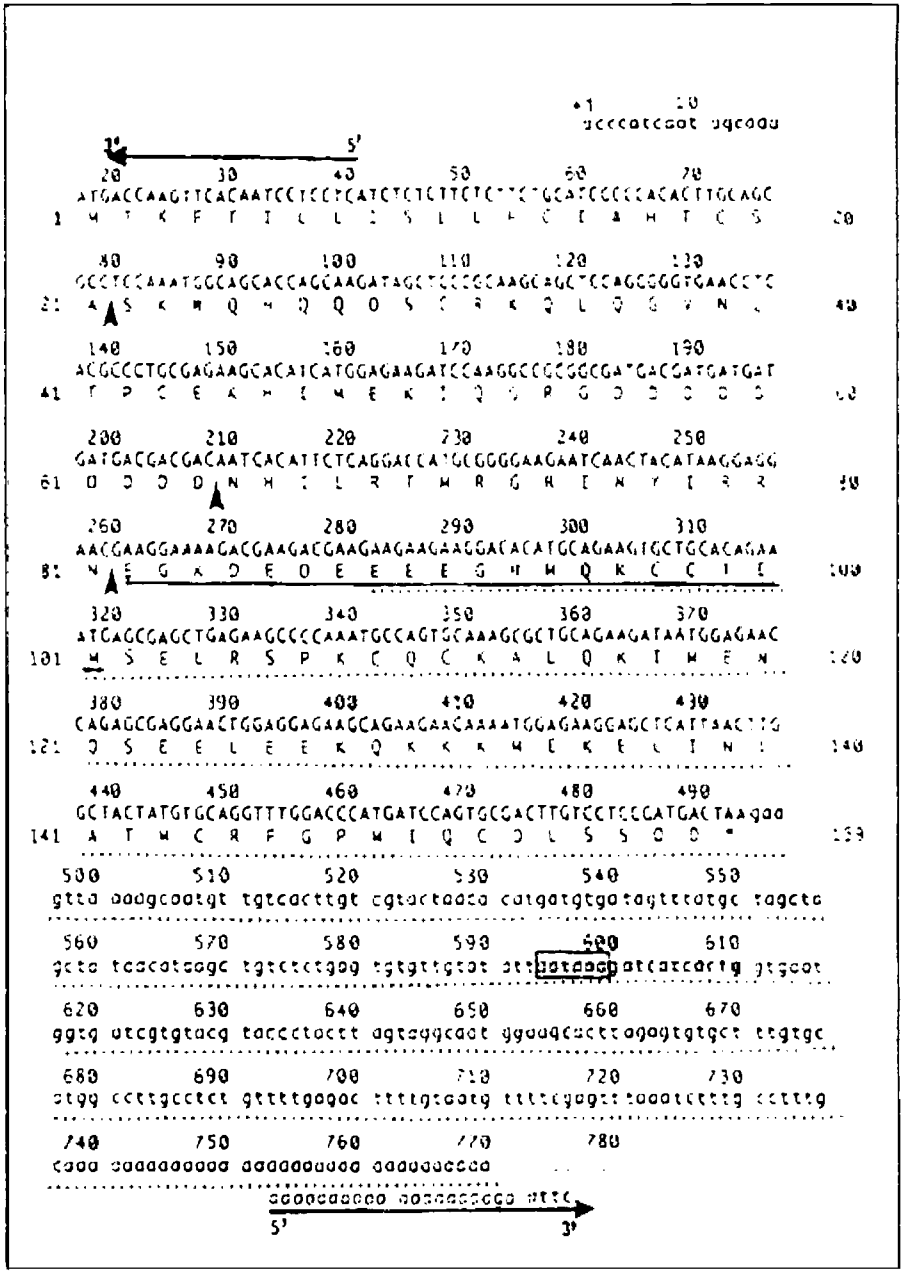


Figure. 2 Deduced DNA sequence of 2D-1 PCR generated fragment (Arrows flank coding region of 2D-1 which terminates after 211bb).

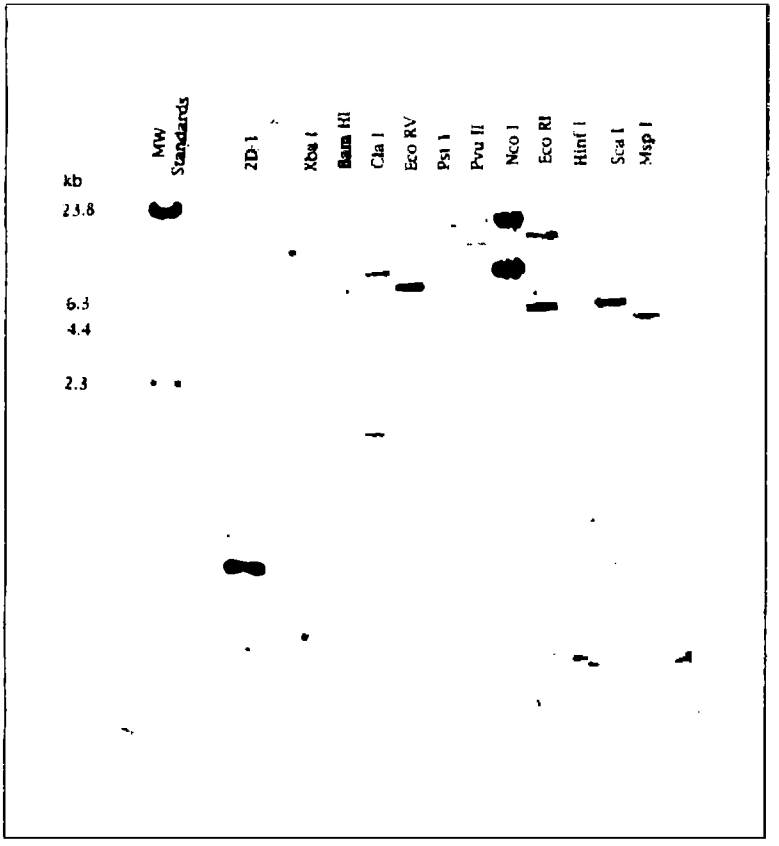


Figure. 3 Southern blot analysis of soybean genomic DNA with the PCR generated fragment 2D-1 as probe. (Ten micrograms of DNA was digested separately by 11 enzymes as labeled in each lane. The probe hybridized to at least two fragments in the genome as indicated by the number of bands the autoradiogram).

methionine- "poor", initial studies suggested that it possesses anti-mitotic property, a potential application in cancer research (Galvez and de Lumen, 1999).

The full length sequence of the 2.5 kb clone revealed a chimera of different transcripts, a glycinin subunit, a fragment homologous to a serine-rich protein from rice callus, a soybean oleosin gene, the complete coding region of Gm2S-1, and an 18 rRNA cDNA. The coding region of Gm2S-1 is within the 1.5 kb EcoR1-bAMH1 fragment between the oleosin fragment and the 18 S rNA cDNA (Fig. 4).

The open reading frame of Gm2S-1 encodes for 158 amino acids where 69 of the amino acids revealed by PCR generated a fragment located near the 3' end

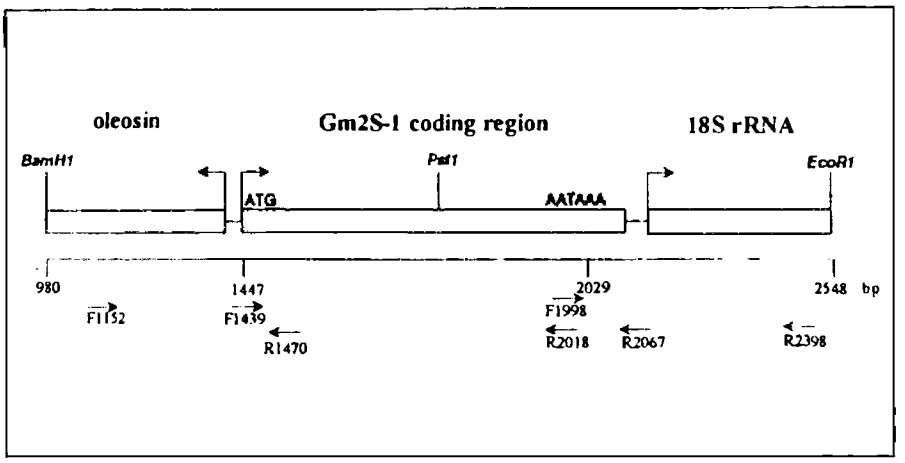


Figure. 4 BamH1 -EcoR1 Fragment containing the Gm2S-1 coding region

suggesting a post translational processing as verified by N-terminal sequence which did not start with a methionine residue. The preproprotein consists of a signal peptide (21 amino acids), a small subunit (43 amino acids), a linker polypeptide (17 amino acids) and a large subunit which is the 8 kD MRP (77 amino acids) (Fig. 5). The MRP contains 8% methionine, 13% lysine and 8% cysteine. The prepotein is highly hydrophilic with a predicted IpH of 5.6 while the mature MRP had an IpH of 4.9 (de Lumen et al, 1999).

Possible Biological Roles of Sulfur-Rich Proteins

The methionine and cysteine-rich proteins represent storage spaces for sulfur in the seed. During germination, the sulfur is mobilized and used not only for amino acid and protein structure, but also for the synthesis of cofactors and coenzymes. In addition the methionine is important in the synthesis of S-adenosyl methionine, a good alkylating agent and polyamines for DNA packaging (Lehninger et al., 1999).

Comparison of the 2D1-PCR DNA sequence with the Genbank data base showed 53% homology with a DNA binding protein from *Arabidopsis thaliana* (Khun et al., 1993) and 60% homology with a cysteine-rich protein from lupine (Gayler et al., 1990). Interestingly, the region of homology with the Gt-2 protein was in the acidic region near the carboxyl terminus. Acidic regions of many NA-binding proteins have been shown to function as transcriptional activators (Saha et al., 1993).

A survey of protein data bases on proteins with high levels of sulfur-rich proteins and enzymes with sulfur-rich motifs revealed some exciting results. Certain cysteine-rich proteins in animal cells have regulatory roles, including differ-

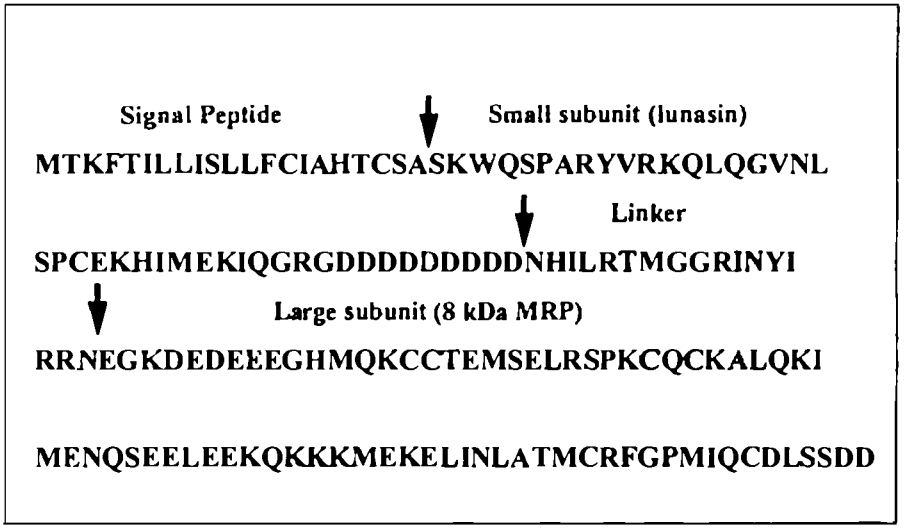


Figure 5. Complete protein encoded by the GM2S-1 cDNA clone. (The protein consists of a hydrophobic signal peptide, a small subunit (SSU) now called lunasin, a linker peptide and the large subunit (LSU) 8 kDa MRP. The arrows indicate proposed posttranslational proteolytic cleavate sites).

entiation, which makes one wonder if these plant protein counterparts play any regulatory role in seed development. Animal metallothioneins (MTs) are cysteine-rich, low-molecular-weight proteins that bind to heavy metals and are believe to play a role in metabolism and detoxification. Genes encoding MT-like proteins have been isolated in a number of plants although their function remains to be educated (Foley and Singh, 1994).

An important new family of proteins has recently been described which carries a novel cysteine-rich zinc-binding domain called the LIM domain. This protein family is present in mammals, amphibians, flies, worms and plants and its main function is in developmental regulation. Although a role in protein-protein interaction seems likely, intriguing similarities to GATA zinc fingers imply that are LIM domain may also be involved in binding to specific nucleic acids (Sanchez-Garcia and Rabbitts, 1994).

One role of MRP is suggested by the hypothesis that the conserved methionine-rich region of a nuclear protein from *Drosophila* and rat is proposed to play an important indirect role in its DNA-binding property as a transcription factor that controls terminal differentiation in *Drosophila* embryo (Weigel and Jackle, 1990). The 10.2 kD soybean protein (4.7% cysteine and 4.7% methionine) that binds to the regulatory sequence in leghemoglobin C^3 gene is proposed to have a regulatory role during soybean nodule differentiation (Bergmann et al, 1991). Interest-

ingly, the N-terminus sequence of one of the two methionine-cysteine-rich proteins identified in Acha, an African underutilized cereal is at least 46% homologous with 3 eukaryotic transcription factors (de Lumen et al, 1993). Two properties of the 8kD soybean albumin that makes it a transcript factor role: it is non-abundant, was detected early in seed development and maintained up to maturity. The length of the 3' untranslated region of the 2D-1 PCR product might be of significance concerning the stability of the transcript which explains its presence in the seed at a very low concentration, 3 μg/g. The long tail could have prevented the digestion of the coding region, thereby conferring stability. It is also possible that 2D-1 is not a typical seed storage protein and may have some biological function in seed development. In addition, 2D-1 may function as a trans-acting factor involved in activating the desiccation machinery in maturing seeds.

Since sources of the sulfur-rich proteins reported so far are oilseeds, it is possible that they play a role in lipid biosynthesis. It is noteworthy that the N-terminus sequence to each of the proteins isolated indicate have been post translationally processed.

SUMMARY AND CONCLUSION

With minimum number of steps, this study permitted the identification and purification of LMW sulfur-rich proteins from soybean albumin. These steps are central in the cloning and characterization of the gene encoding the MRP since there are no continuous conserved regions among reported MRPs from several sources. The use of heterologous probes for gene identification was therefore ruled out. Thus, the determination of the N-terminal sequence facilitated the design of primers for use in PCR which generated a truncated DNA fragment partially encoding the MRP. This served as probe for initial characterization studies and cDNA clone identification.

The genes encoding methionine-rich proteins (MRP) in seeds are candidates for overexpression to enhance the nutritional quality of legume proteins, which are relatively deficient in methionine. This is significant because of the findings in transgenic soybean. The transformation of soybean with Brazil nut MRP cDNA driven by a β-phaseolin promoter leads to the accumulation of the Brazil nut MRP up to 8% of the total protein, which is equivalent to a 26% increase in methionine content (Nordlee et al., 1994; Townsend et al., 1992). It remains to be seen if this increase is nutritionally significant. Using radioimmune allergosorbant test (RAST), the Brazil nut MRP in the transgenic soybean binds human IgE from sera of individuals allergic to Brazil nut whereas no binding is observed with comparable amounts of protein extracted from a genetically equivalent line of non-transformed soybean (Nordlee, 1994). Further, there are no extremely new cases of allergic responses to soy proteins inspite of their long history and widespread use as human food.

Overexpression of these sulfur-rich genes from soybean provides a viable means of increasing the methionine content of seed proteins in soybean and other legumems through genetic engineering, resulting in the enhancement of their nutritional and economic values as human food and animal feed, and in elucidating their biological roles.

LITERATURE CITED

Altenbach, S.B. and R.B. Simpson. 1990. Manipulation of the methionine-rich protein genes in plant seeds. *Trends Biotechnology.* 8:156-160.

Altenbach, S.B.; Pearson, K.W.; Leung, F.W. and S.S.M. Sun. 1987. Cloning and sequence analysis of a cDNA encoding a Brazil nut protein exceptionally rich in methionine. *Plant Mol. Biol.* 8: 239-250.

Bergmann, J.E.; Preddie, E.; Cortes, L. and R. Broseau. 1991. A protein Srp90 encoded on the leftward strand of soybean nodule urate oxidase cDNA binds to a regulatory sequence in loghemoglobin C3 gene. *Nucleic Acids Research.* 19: 338.

de Lumen, B.O. and C.J. Kho. 1987. Identification of methionine containing proteins and quantitation of their methionine contents. *J. Agric. Food Chem.* 35: 688-691.

de Lumen, B.O. 1990. Molecular approaches to improving the nutritional and functional properties of plant seeds as food sources: developments and comments. J. Agric. Food Chem. 38: 1779-1788.

de Lumen, B.O.; Krenz D.C. and M.J. Revilleza. 1997. Molecular strategies to improve the protein quality of legumes. *Food Tech.* 51:67-70.

de Lumen, B.O.; Galvez A.F.; Revilleza M.J. and D.C. Krenz. 1999. Molecular strategies to improve the nutritional quality of legume proteins in *Chemicals vis Higher Plant Engineering.* Kluwer Academic/Plenum Publishers NY 117-126.

Foley, R.C. and K. B. Singh. 1994. Isolation of a *Vicia faba* metallothionein-like gene: expression in foliar trichomes. *Plant Mol. Biol.* 26(1): 435-444.

Galvez, A.F.; Revilleza, M.J.; de Lumen, B.O. and D.C. Krenz. 1997. Enhancing the biosynthesis of endogenous methionine-rich proteins (MRP) to improve protein quality of legumes via genetic engineering. *Proceedings of Food for Health in the Pacific Rim. 3rd International Conference of Food Science and Technology.*

Galvez, A.F.; Revilleza, M.J. and B.O. de Lumen. A novel methionine-rich protein from soybean *(Glycine max)* seed: cloning and characterization of cDNA (Accession NO. AF005030) *Plant Physiol.* 114:1567.

Gayler, K.R.; Kolivas, S.; Macfarlane, A.J.; Lilley, G.g.; Baldi, Blagrove, R.J.; and E.D. Johnson. 1990. Biosynthesis, cDNA and amino acid sequences of a precursor of conglutin delta, a sulphur-rich protein from *Lupinus angustifolius. Plant Mol. Biol.* 15: 879-893.

George, A. and B.O. de Lumen. 1990. A novel methionine rich protein in soybean and; identification and amino acid composition and N-terminus sequence. *J. Agric. Food Chem.* 39:224-227.

Khun, R.M. 1993. DNA binding factor GT-2 from *Arabidopsis. Plant Mol. Biol.* 23:337-348.

Kortt, A.A.; Caldwell, J.B.; Lilley, G.G. and T.J.V. Higgins. 1991. Amino acid and cDNA sequences of a methionine-rich 2S protein from sunflower *(Helianthus annuum L). Eur. J. Biochem.* 195: 329-379.

Lehninger, A.L.; Nelson, D.L.; and M. Cox. 1993. Principles of Biochemistry. 2nd ed. Worth publishers.

Nordlee, J.A.; Taylor, S.L.; Townsend, J.A.; Thomas, L.a. and R. Townsend. 1994. Investigations of the allgergenicity of Brazil nut 2S seed storage protein in transgenic soybean. OECD Workshop on the Evaluation of new Foods. Oriel College Oxford, UK.

Revilleza, M.J.R.; Galvez, A.F.; Krenz, D.C.; and B.O. de Lumen. 1996. An 8 kDa methionine-rich protein from soybean (Glycine max) cotyledon: identification, purification and N-terminal sequence, J. Agric. and Food Chem. 44(9): 2930-2935.

Saha, S.; Brickman, J.M.; Lehming, N.; and M. Ptashne. 1993. New eukaryotic transcription repressors. *Nature*. 363(6430):648-52.

Sanchez-Gracie, I. and T.H. Rabbits. 1994. The LIM domain: a new structural motif found in zinc finger-like protein. *Trends in Genetics*. 10(9):315-320.

Townsend, J.A.; Thomas, L.A.; Kulisek, E.S.; Daywalt, M.J.; K.R.K. Altenback. 1992. Improving the quality of seed proteins in soybean. *Proceedings of th 4th Biennial Conference of Molecular and Cellular Biology of Soybean*, p.4.

Weigle, D. and J. Jackle. 1990. The fork domain: a novel DNA binding motif of eukaryotic transcription factors? *Cell* 63:455-456.