8S Globulin of Mungbean [Vigna radiata (L.) Wilczek]: Cloning and Characterization of Its cDNA Isoforms, Expression in Escherichia coli, Purification, and Crystallization of the Ma...

ARTICLE *in* JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY · JUNE 2004

Impact Factor: 2.91 · DOI: 10.1021/jf0305938 · Source: PubMed

8 AUTHORS, INCLUDING:

Roberta N Garcia
University of the Philippines Los Baños
9 PUBLICATIONS **94** CITATIONS
SEE PROFILE

Jorge Gil C Angeles
University of the Philippines Los Baños
7 PUBLICATIONS **80** CITATIONS
SEE PROFILE

Akito Kaga
National Institute of Agrobiological Sciences
78 PUBLICATIONS **1,217** CITATIONS
SEE PROFILE

JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

8S Globulin of Mungbean [*Vigna radiata* (L.) Wilczek]: Cloning and Characterization of Its cDNA Isoforms, Expression in *Escherichia coli*, Purification, and Crystallization of the Major Recombinant 8S Isoform

AMY EMILIANA N. BERNARDO,[†] ROBERTA N. GARCIA,[†] MOTOYASU ADACHI,[‡] JORGE GIL C. ANGELES,[†] AKITO KAGA,[§] MASAO ISHIMOTO,[||] SHIGERU UTSUMI,[‡] AND EVELYN MAE TECSON-MENDOZA*,[†]

Institute of Plant Breeding, College of Agriculture, University of the Philippines Los Baños, College, Laguna Philippines 4031, Laboratory of Food Quality Design and Development, Graduate School of Agriculture, Kyoto University, Uji, Kyoto 611-0011, Japan, National Institute of Agrobiological Sciences, 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan, and National Agricultural Research Center for Western Region, 6-12-1 Nishifukatsu, Fukuyama, Hiroshima 721-8514, Japan

Three isoforms of the cDNA of the major 8S globulin of mungbean, 8Sα, 8Sα′, and 8Sβ, were isolated, cloned, and characterized. The cDNA sequences of 8Sα, 8Sα′, and 8Sβ had open reading frames of 1362, 1359 or 1362, and 1359 bp, respectively, which code for 454, 453 or 454, and 453 amino acids corresponding to molecular weights of 51 973, 51 627 or 51 758, and 51 779, respectively. Homology in terms of cDNA and amino acid sequences was 91−92% between 8Sα and 8Sα′, 87% between 8Sα and 8Sβ, and 86−88% between 8Sα′ and 8Sβ. The signal peptide was found to be 1−25, 1−24 or 25, and 1−23 for 8Sα, 8Sα′, and 8Sβ, respectively, using the signalP website (Nielsen, H.; Engelbrecht, J.; Brunak, S.; von Heijne, G. *Protein Eng.* **1997**, *10*, 1−6). The propeptide was determined to be IVHREN. A single site for glycosylation (N-X-S/T) was observed about 90 amino acids from the C terminus. Homology between mungbean 8S isoforms and other 7−8S proteins ranged from 45 to 68% within members of the legume family and 29 to 34% for crops of different species. The major isoform 8Sα was expressed in *Escherichia coli* and purified by successive ammonium sulfate fractionation, hydrophobic interaction, and Mono Q column chromatography. The recombinant 8Sα, but not the native form, was successfully crystallized producing rhombohedral crystals.

KEYWORDS: 8S globulin; cDNA; isoforms; vicilin; amino acid sequence; crystallization; purification; homology analysis; signal peptide; propeptide; mungbean; *Vigna radiata* L.; legumes

INTRODUCTION

Mungbean [*Vigna radiata* (L.) Wilczek] has been utilized as food since ancient times. It has increased its popularity as dry beans and sprouts in Asia, especially in India where it traces its origin, but also in South America, Australia, and the United States. In the United States, 7000−9000 tons of mungbean are consumed annually and 75% of this is imported. Australia's production of mungbean of 12 600−27 000 tons/year is mostly for export. Mungbean is grown widely in about 5.77 million ha in Asia producing 2.9 million tons per year (*1*). For developing countries, mungbean is a major source of protein.

Storage proteins of legumes such as mungbean comprise most of the seed proteins. We have previously reported the purification and characterization of the storage proteins of mungbean, which consist of the basic 7S, vicilin type (8S), and legumin type (11S) globulins at 3.4, 89, and 7.6% levels (*2*). The major storage protein, the 8S, has a native molecular weight of 200 000 for the 8S globulin although it also coeluted with 11S, which has a molecular weight of 360 000. On sodium dodecyl sulfate−polyacrylamide gel electrophoresis (SDS−PAGE), the 8S major globulin resolved into four bands with molecular weights of 60 000, 48 000, 32 000, and 26 000. The 8S globulin of mungbean was found to have no disulfide linkages and was positive for carbohydrate using conjugated peroxidase reaction. Cross-reactivity of soybean β-conglycinin subunits (α, α′, and β) was observed with the first three major bands of the 8S, whereas only the fourth band cross-reacted with the anti-β

* To whom correspondence should be addressed. E-mail: emtm@laguna.net. A.E.N.B. and R.N.G. contributed equally to this paper.
[†] University of the Philippines Los Baños.
[‡] Kyoto University.
[§] National Institute of Agrobiological Sciences.
[||] National Agricultural Research Center for Western Region.

10.1021/jf0305938 CCC: $27.50
Published on Web 04/10/2004

subunit. The 8S globulin could be extracted over a wide range of NaCl concentrations. The N-terminal sequences of the different subunits or fragments had 45−78% homology in 10−14 amino acid overlap with storage proteins of various legumes and other crops. Earlier reports have been made on the storage proteins of mungbean (*3*) and the purification of vicilin (*4*).

Attempts to improve the nutritional quality of mungbean through breeding have largely been unintentional since efforts have primarily focused on improving its yield and pest resistance. However, it has been noted that the protein content of mungbean has steadily increased from 17 to 26%. In spite of the higher protein level, its nutritional quality still remains the same since mungbean, like other legumes, is deficient in sulfur-containing amino acids methionine and cysteine. There are now studies worldwide, which aim at improving nutritional quality and functional properties of proteins of various food crops such as soybean (*5*−*7*), rice (*8*), corn (*9*), etc. through genetic engineering.

In line with our efforts to study the possibility of improving the nutritional quality of mungbean by nonconventional methods such as genetic engineering, we have isolated, cloned, and characterized the cDNAs of three isoforms of the major 8S globulin of mungbean, which we are reporting herein. Unlike other leguminous seed crops, there is a dearth of information on the DNA or protein sequences of the storage proteins of the *Vigna* species, including that of mungbean. This is, therefore, the first report on the molecular and structural aspects of the major storage protein of mungbean. Using various softwares and databases, we determined several important features of the mungbean 8S globulin isoforms such as their signal peptide, propeptide, secondary structures, and phylogenetic relation with other seed storage proteins, which affirmed mungbean's classification in the *Vigna* instead of *Phaseolus*. Toward a detailed study of the structure of the mungbean vicilin, we expressed the major isoform 8Sα in *Escherichia coli* and successfully crystallized the recombinant 8Sα protein.

EXPERIMENTAL PROCEDURES

Materials. The ExAssist helper phage system was purchased from Stratagene. Restriction enzymes were purchased from Takara, Japan. γ^{32}P-(ATP) was purchased from Daiichi Pure Chemicals, Japan. Other reagents of the highest purity were purchased from various vendors.

cDNA Library Construction. The mungbean cDNA library was prepared using the following method. Total RNA was isolated from a mixture of 2−5 mm developing seeds of BC20F4 (cultivar Osaka-ryokuto) resistant isogenic line of mungbean according to the method of Hall et al. (*10*). Poly(A)-RNA was obtained by two cycles of BioMag Oligo(dT)20 chromatography (PerSeptive Biosystems Inc., Japan). First and second strand cDNAs were synthesized using ZAP-cDNA Synthesis kit and oligonucleotides consisting of oligo (dT)18 and Xho I restriction enzyme recognition sequences as a primer (Stratagene, U.S.A.). The cDNAs extended with EcoRI adaptors were digested with XhoI and then directionally ligated into Uni-ZAP XR vector arms. The phages were packaged in vitro, resulting in 4×10^5 plaques. Positive plaques were purified for three more rounds, and then, the inserts were subcloned into pBluescript SK$^-$ by biological rescue as recommended by the supplier (Stratagene).

Probes. The oligonucleotide probes Mung8S1 [ATHGAYGCNGC-NGAAGT] and Mung8S2 [ATHGAYGCNGCNGAGGT] were designed based on the results of the N-terminal amino acid sequence of the major 8S band 2 (*2*).

cDNA Library Screening. About 100 000 pfu/plate of immature mungbean seed cDNA library packaged in λZAPI (Stratagene) were screened on Hybond N$^+$ membranes (Amersham). Prehybridization in 6xSSC, 20 mM $NaH_2PO_4 \cdot 2H_2O$, 0.4% SDS, and 10 mg of salmon sperm DNA was done at 42 °C for at least 2 h. Hybridization was carried out at 42 °C overnight in 10 mL of prehybridization solution plus 20 mg of salmon sperm DNA and 200 μg of γ^{32}P-(ATP) 5′-end-labeled probes. The membranes were washed three times for 20 min at room temperature in 6xSSC/0.1% SDS followed by sequential washing at 42, 47, 52, and 57 °C in 6xSSC/0.1% SDS for 20 min each time. Membranes were exposed to the imaging plate, and the hybridized probe was detected by BAS2000 imaging analyzer (Fuji). Sixteen plaques corresponding to clones that exhibited a strong signal were cored from the agar plates, replated, and underwent second screening. Seven positive clones were selected and excised in vivo with the ExAssist helper phage system (Stratagene).

DNA Sequencing. The polymerase chain reaction (PCR) was conducted on the full length 8S cDNAs in pBluescript plasmid using T3 or T7 sequencing primers and some internal primers. The PCR conditions were as follows: 96 °C for 20 s, 25 cycles of 96 °C for 10 s, 50 °C for 5 s, and 60 °C for 4 min. The PCR products were sequenced using the Applied Biosystems automated sequencer (model 310A) according to the manufacturer's protocol.

Analysis of DNA and Amino Acid Sequences. The cDNA sequences of the different 8S isoforms were aligned with each other and with sequences of 7S-8S seed proteins using the ClustalW algorithm of the AlignX module of VectorNTI (*11*). The cDNA sequences of 8S isoforms were also compared with sequences found in the database using the BLASTN (*12*).

The deduced amino acid sequences of the different 8S isoforms were analyzed using the BLASTP 2.2.2 module (*12*). The conserved domains of the protein were analyzed using RPS-BLAST 2.2.5 (*12*). Multiple alignment analysis of the amino acid sequences of the isoforms with those of 7S-8S seed proteins was done using the ClustalW algorithm of the AlignX module of Vector NTI (*11*). The signal peptide was determined using the SignalP website (*13*).

Expression of Recombinant 8Sα in *E. coli*. The 8Sα cDNA was inserted into pET 21d (Novagen Inc., U.S.A.) expression vector using the following method. The vector was first cut with NcoI and then blunt-ended resulting in CCATG-3′. The vector was dephosphorylated and then restricted with EcoRI to generate an open vector with one blunt end and one sticky end. The 8Sα was amplified using KOD DNA polymerase (TOYOBO, Japan) with primers 8SN1 5′-ATT GAT GGC GCA GAA GTG TCC GTT TCA AG-3′ (designed based on the N-terminal protein sequence IDGAEVSVS) (from this paper) and 8SC1 5′-GGA ATT CTT CAG GTT GGG CTC ATT TAT TG-3′ (introducing an EcoRI site next to the stop codon TGA). The 5′-end of the PCR product was phosphorylated and then cut with EcoRI. The PCR product and expression vector were gel purified using Gene Clean (Bio 101) and ligated.

The pET21d containing the 8Sα insert was transformed in *E. coli* strain HMS174(DE3), JM109(DE3), and BL21(DE3). The bacteria were grown on LB medium containing carbenicillin, and when the culture reached A_{600} of 0.8, isopropyl-thio-β-D-galactopyranoside (IPTG) was added to the medium to induce expression of the protein. The culture continued to be grown for 72 h at 20 °C. Then, the cells were harvested by centrifugation at 6000 rpm for 20 min at 4 °C. The cells were resuspended in buffer A [35 mM potassium phosphate (K-Pi) buffer, pH 7.6, 0.4 M NaCl, 1 mM EDTA, 0.1 M *p*-amidinophenyl-methylsulfonyl fluoride (APMSF), 1 μg/mL leupeptin, and 1 μg/mL pepstatin A] and sonicated. The protein contents of the crude extract, supernatant, and precipitate resuspended in extraction buffer were determined using Bradford method (*14*) and subjected to SDS−PAGE using 11% acrylamide. Electrophoresis was run at 200 V for 35 min for samples in low salt concentration and 150 V and 55 min for samples in high salt concentration, using Tris-glycine buffer at pH 8.3 (*15*). Six micrograms of protein was applied in each well. The recombinant product was detected based on its expected size.

Purification of Recombinant 8Sα. For large scale purification of the 8Sα globulin, a total of eight liters (16×0.5 L) of *E. coli* HMS174-(DE3) carrying the pET expression vector with 8Sα gene insert were grown on LB medium containing carbenicillin. At A_{600} of 0.8, IPTG was added to induce expression of the protein. After 72 h, the bacterial cells were harvested by centrifugation at 6000 rpm ($\times$ 5000*g*) for 20 min at 4 °C. The cells were resuspended in buffer A and sonicated.

Ammonium sulfate was added to the crude protein extract to 35% saturation, and then, the sample was subjected to hydrophobic interac-

tion chromatography (HIC) using butyl toyopearl. The sample was eluted with ammonium sulfate gradient of 30 to 0% in buffer A. Fractions were run on SDS−PAGE, and those that contained the 8Sα protein were collected and dialyzed against buffer A containing 0.1 M NaCl.

The pooled and dialyzed HIC fractions were further purified using Mono Q HR (Pharmacia Biotech) column developed at a flow rate of 1.5 mL/min with a gradient of 0.1−0.4 M NaCl in buffer A. The 8Sα fractions resolved by SDS−PAGE were pooled and dialyzed against 5 mM KPi buffer, pH 7.6, 0.30 M NaCl, 1 mM EDTA, 0.10 mM APMSF, 1 μg mL leupetin, and 1 μg/mL pepstatin A. After dialysis, the sample was concentrated to 1.4 mL using Centricon (Millipore). The protein content of the samples was determined using WAKO Protein Assay Rapid Kit (WAKO) with bovine serum albumin as standard.

Purification of Native Mungbean 8Sα. Native mungbean 8S was purified from defatted seed meal following the procedures of Tecson-Mendoza et al. (*2*) except that buffer A was used for dialysis instead of water and 10 mM β-mercaptoethanol solution.

Crystallization of the Recombinant and Native Mungbean 8Sα. The hanging drop method (*16*) was used to crystallize the mungbean 8S globulins. Various buffers and precipitating agents were used such as 0.1 M HEPES, pH 4.5 and 5.0; 0.1 M Tris, pH 7.0 and 7.5; 0.1 M sodium acetate buffer, pH 4.6, 5.0, and 5.38; and 0.10 M MES, pH 6.0, with ammonium sulfate (50, 52, 54, 56, 58, and 60%). PEG 1000 and PEG 4000 (8, 10, 12, and 14%) were used with increasing concentrations of NaCl in 35 mM KPi buffer, pH 7.6. The protein concentration was 5 mg/mL, and crystallization was allowed to proceed at 20 °C.

N-Terminal Amino Acid Sequence Analysis. The N-terminal amino acid sequence of the recombinant mungbean 8S globulin was determined using a Procise 492 Protein Sequencer. The sample was transferred onto PVDF membrane using the Prosorb Sample Preparation Cartridge by Perkin-Elmer. Briefly, the PVDF membrane at the bottom of the sample reservoir was washed with 10 μL of methanol. A 1 μL protein sample was diluted to 100 μL with 0.1% TFA, applied to the sample reservoir supplied with absorbent material, and washed with 750 μL 0.1% TFA. The membrane was punched out, washed with 4 μL of methanol, allowed to dry, and kept until use.

RESULTS AND DISCUSSION

Screening of cDNA Library. Two rounds of screening resulted in the selection of seven positive clones. On the basis of the DNA sequences of these seven clones (data not shown), three types of 8S cDNAs from mungbean are evident as follows: 8Sα, consisting of four clones; 8Sα′, consisting of two clones; and 8Sβ, with one clone. This indicates the presence of isoforms of the 8S globulin, with the 8Sα being the major isoform. The presence of isoforms in storage proteins has been shown for soybean β-conglycinin with its α, α′, and β forms (*17*, *18*); phaseolin with the α, β, and C types (*19*, *20*); and pea vicilin with its α and β types (*21*), among others.

cDNA Sequences of Isoforms of 8S Mungbean Globulin. The nucleotide sequences of 8Sα, 8Sα′, and 8Sβ cDNAs are shown in **Figure 1**. The 8S cDNAs had open reading frames of 1362 for 8Sα, 1359 or 1362 for 8Sα′, and 1359 bp for 8Sβ. The tandem initiation codon of ATGATG was observed in 8Sα′.

The 8Sα cDNA sequence was found to be highly homologous with 8Sα′ at 91% and was less homologous with 8Sβ at 87%. The 8Sα′ cDNA sequence was 86% homologous with 8Sβ.

Analysis of the Derived Amino Acid Sequences. The 8S globulin cDNAs had open reading frames coding for 454, 453 or 454, and 453 amino acids for 8Sα, 8Sα′, and 8Sβ, respectively, corresponding to molecular weights of 51 973, 51 627 or 51 758, and 51 779, respectively. On the basis of the cDNA sequence, it is possible for 8Sα′ to have another met after the initial met.

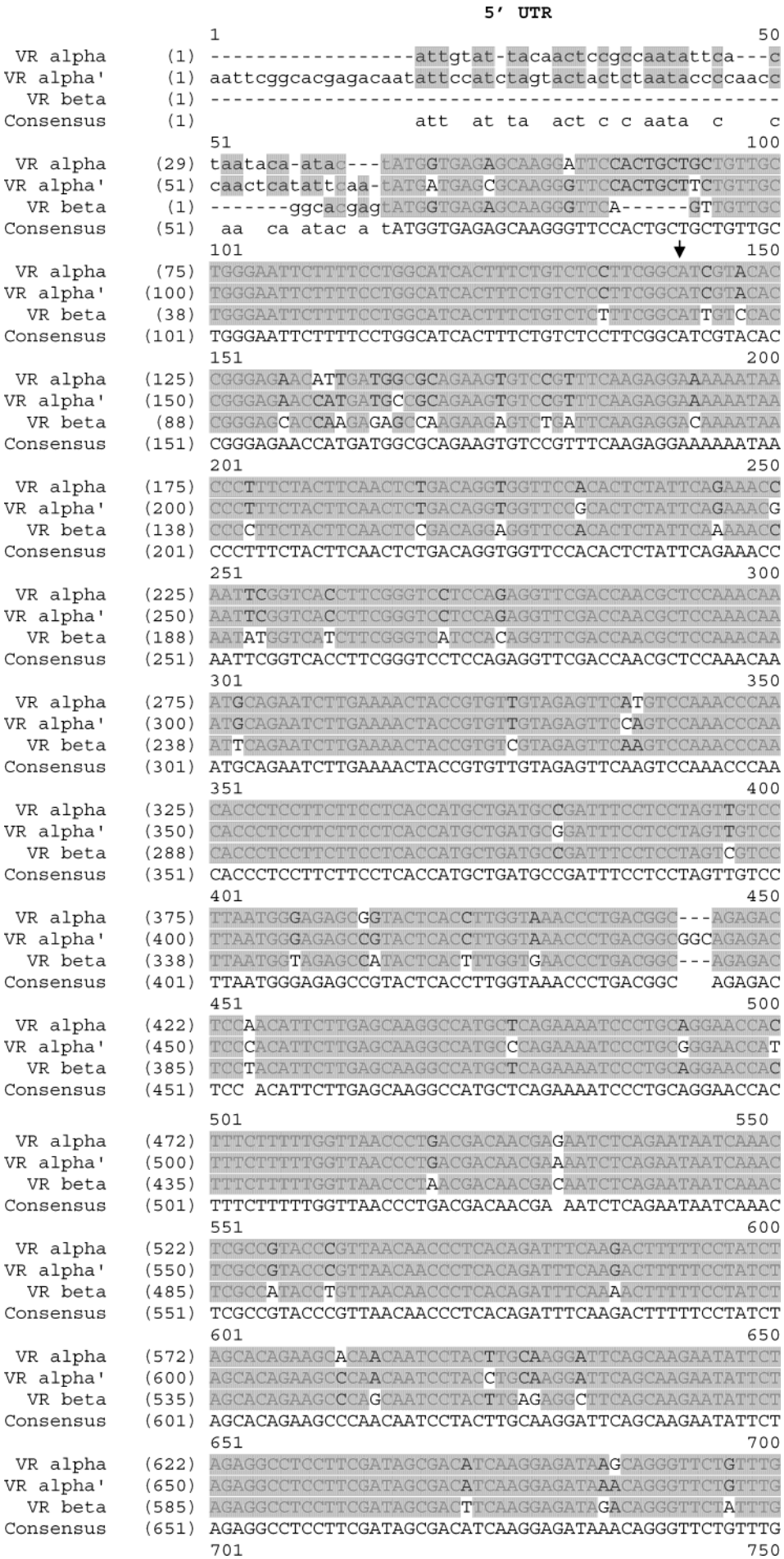
The deduced amino acid sequences for 8Sα, 8Sα′, and 8Sβ are shown and compared in **Figure 2**. As with the cDNA sequence comparison, the 8Sα amino acid sequence had higher homology with 8Sα′ (92%) than with 8Sβ (87%), while 8Sα′ was 88% homologous with 8Sβ. The detection of the three isoforms of 8S globulin indicates that the native 8S with a molecular weight of 200 000 (*2*) consists of heterotrimers. However, the similarities between the individual subunits cloned in this study are much higher than in the case of the β-conglycinin subunits of soybean, which had homologies of around 75.5−90.4% (*22*).

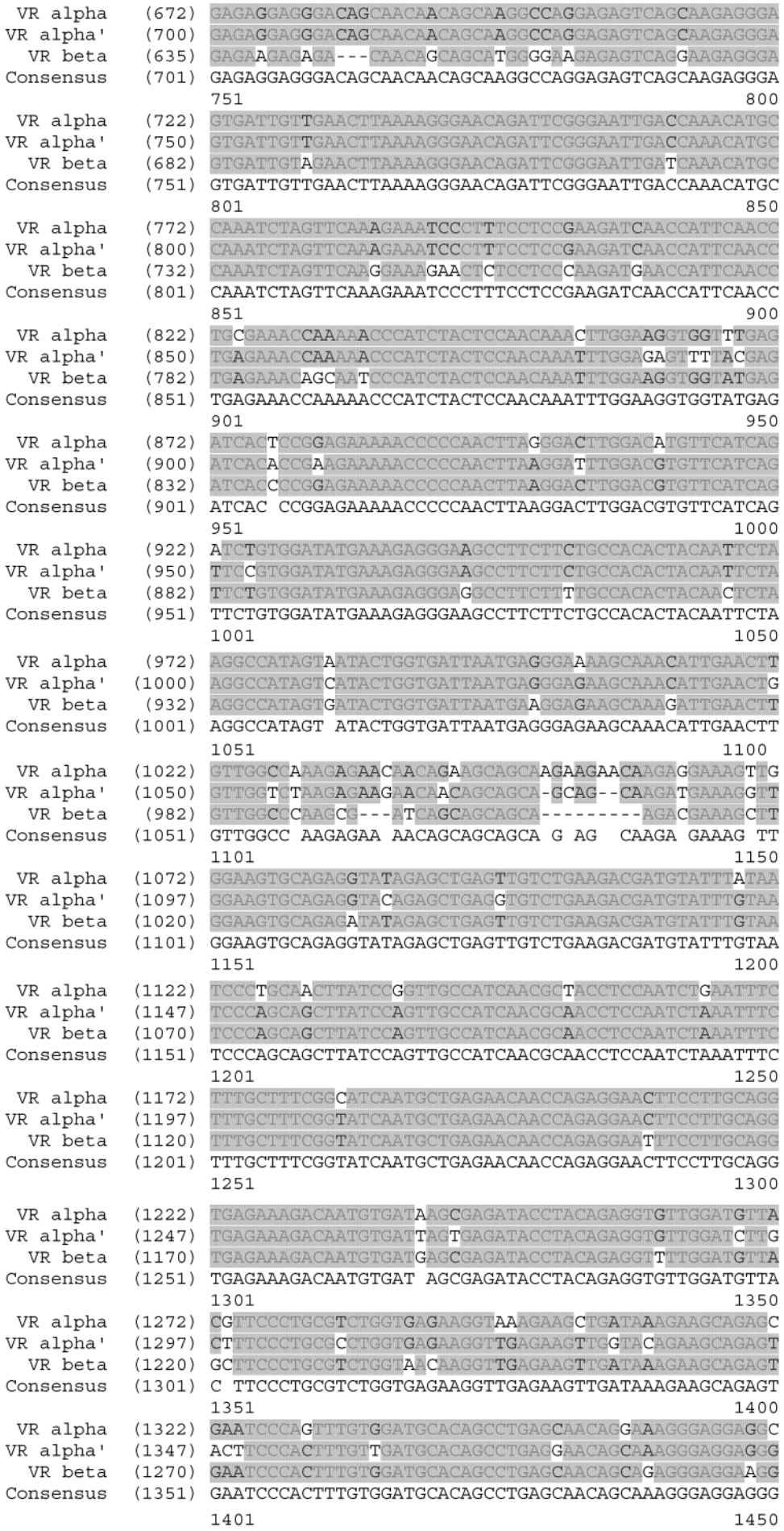
Analysis of the amino acid sequences using BLASTP (*12*) shows that the mungbean 8S globulins belong to the 7S seed storage family. These proteins have two domains with similar double-stranded β-helix folds and are a member of the "cupin" (the Latin term for barrel) superfamily on the basis of their conserved barrel domain (*23*, *24*). At the N-terminal domain, 44−244 sequence of mungbean 8S α, α′, and β was found to be 99.0% aligned with the corresponding sequence in the 7S seed storage family. Similarly, at the C-terminal domain, the sequence 260−433 was 99.4% aligned with the 7S seed storage family. The amino acid sequence from 88 to 146 was found to be 98.5% aligned with the domain involved in carbohydrate binding and protein−protein interactions.

The signal peptide was found to be 1−25, 1−24 or 25, and 1−23 for 8Sα, 8Sα′, and 8Sβ, respectively, using the SignalP website (*13*). The most likely cleavage site is at a conserved site between SFG and IV for the three isoforms. A highly conserved region in the signal peptide covers 15 amino acids. Considering that the N-terminal sequence earlier reported for mungbean 8S globulin was IDAAE (*2*), the propeptide, therefore, consists of six amino acids (IVHREN). The signal peptide and propeptide comprise 3449, 3266 or 3397, and 3262 Da for 8Sα, 8Sα′, and 8Sβ, respectively, and this region is highly hydrophobic with 55−58% frequency. The mature polypeptide contains, on the other hand, 32−34% hydrophobic amino acids. The signal peptide sequence is predicted to be 20−30 amino acids and usually consists of a conserved core of hydrophobic amino acid residues followed by a region of variable amino acids where cleavage occurs (*20*, *25*). However, we cannot rule out the possibility that 8Sβ does not contain propeptide since the amino acid just after IVHRE is not asparagine whose C-terminal side would be cleaved by asparagynylendopeptidase.

The 8S isoforms reported herein correspond to the major band (no. 2) reported previously (*2*) with a molecular weight of 48 000, representing the mature polypeptide. The signal peptide and propeptide comprise about 3 kDa, which with the mature polypeptide will be around 51 000 similar to the size obtained from the derived amino acid sequence. The 8S band no. 3, which has an N-terminal sequence found past midway the amino acid sequence of the whole polypeptide, probably represents the C-terminal domain. On the other hand, 8S band no. 4 represents the fragment of the N-terminal domain.

However, none among the isolated 8S isoforms exhibited the N-terminal sequence of 8S band no. 1 (EDKEEQ), which was noted to be similar to part of the extension region of α and α′ subunits of β-conglycinin. This observation indicates that there is a convicilin type of protein in mungbean. The vicilin fraction of pea (*Pisum sativum*) was first reported to contain major polypeptides of 71 000, 50 000, and 33 000 (*26*). The polypeptide of M_r 71 000 was found to be the third storage protein of pea termed convicilin with a molecular weight of 290 000 in its native form (*27*). The pea convicilin differs from vicilin by the insertion of a 121 amino acid sequence near the N terminus


5′ UTR
1 50
VR alpha (1) ------------------attgtat-tacaactcgccaatattca---c
VR alpha' (1) aattcggcacgagacaatattccatctagtactactctaataccccaacc
VR beta (1) ----------
Consensus (1) att at ta act c c aata c c
51 100
VR alpha (29) taataca-atac---tATGGTGAGAGCAAGGATTCCACTGCTGCTGTTGC
VR alpha' (51) caactcatatt aa-tATGATGAGCGCAAGGGTTCCACTGCTTCTGTTGC
VR beta (1) -------ggcacgagtATGGTGAGAGCAAGGGTTCA------GTTGTTGC
Consensus (51) aa ca atac a tATGGTGAGAGCAAGGGTTCCACTGCTGCTGTTGC
101 150
VR alpha (75) TGGGAATTCTTTTCCTGGCATCACTTTCTGTCTCCTTCGGCATCGTACAC
VR alpha' (100) TGGGAATTCTTTTCCTGGCATCACTTTCTGTCTCCTTCGGCATCGTACAC
VR beta (38) TGGGAATTCTTTTCCTGGCATCACTTTCTGTCTCTTTCGGCATTGTCCAC
Consensus (101) TGGGAATTCTTTTCCTGGCATCACTTTCTGTCTCCTTCGGCATCGTACAC
151 200
VR alpha (125) CGGGAGAACATTGATGGCGCAGAAGTGTCCGTTTCAAGAGGAAAAATAA
VR alpha' (150) CGGGAGAACCATGATGCCGCAGAAGTGTCCGTTTCAAGAGGAAAAATAA
VR beta (88) CGGGAGCACCAAGAGAGCCAAGAAGAGTCTGATTCAAGAGGACAAAATAA
Consensus (151) CGGGAGAACCATGATGGCGCAGAAGTGTCCGTTTCAAGAGGAAAAAATAA
201 250
VR alpha (175) CCCTTTCTACTTCAACTCTGACAGGTGGTTCCACACTCTATTCAGAAACC
VR alpha' (200) CCCTTTCTACTTCAACTCTGACAGGTGGTTCCGCACTCTATTCAGAAACG
VR beta (138) CCCCTTCTACTTCAACTCCGACAGGAGGTTCCACACTCTATTCAAAACC
Consensus (201) CCCTTTCTACTTCAACTCTGACAGGTGGTTCCACACTCTATTCAGAAACC
251 300
VR alpha (225) AATTCGGTCACCTTCGGGTCCTCCAGAGGTTCGACCAACGCTCCAAACAA
VR alpha' (250) AATTCGGTCACCTTCGGGTCCTCCAGAGGTTCGACCAACGCTCCAAACAA
VR beta (188) AATATGGTCATCTTCGGGTCATCCACAGGTTCGACCAACGCTCCAAACAA
Consensus (251) AATTCGGTCACCTTCGGGTCCTCCAGAGGTTCGACCAACGCTCCAAACAA
301 350
VR alpha (275) ATGCAGAATCTTGAAAACTACCGTGTTGTAGAGTTCATGTCCAAACCCAA
VR alpha' (300) ATGCAGAATCTTGAAAACTACCGTGTTGTAGAGTTCCAGTCCAAACCCAA
VR beta (238) ATTCAGAATCTTGAAAACTACCGTGTCGTAGAGTTCAAGTCCAAACCCAA
Consensus (301) ATGCAGAATCTTGAAAACTACCGTGTTGTAGAGTTCAAGTCCAAACCCAA
351 400
VR alpha (325) CACCCTCCTTCTTCCTCACCATGCTGATGCCGATTTCCTCCTAGTTGTCC
VR alpha' (350) CACCCTCCTTCTTCCTCACCATGCTGATGCGGATTTCCTCCTAGTTGTCC
VR beta (288) CACCCTCCTTCTTCCTCACCATGCTGATGCCGATTTCCTCCTAGTCGTCC
Consensus (351) CACCCTCCTTCTTCCTCACCATGCTGATGCCGATTTCCTCCTAGTTGTCC
401 450
VR alpha (375) TTAATGGGAGAGCGGTACTCACCTTGGTAAACCCTGACGGC---AGAGAC
VR alpha' (400) TTAATGGGAGAGCCGTACTCACCTTGGTAAACCCTGACGGCGGCAGAGAC
VR beta (338) TTAATGGTAGAGCCATACTCACTTTGGTGAACCCTGACGGC---AGAGAC
Consensus (401) TTAATGGGAGAGCCGTACTCACCTTGGTAAACCCTGACGGC AGAGAC
451 500
VR alpha (422) TCCAACATTCTTGAGCAAGGCCATGCTCAGAAAATCCCTGCAGGAACCAC
VR alpha' (450) TCCCACATTCTTGAGCAAGGCCATGCCAGAAAATCCCTGCGGGAACCAT
VR beta (385) TCCTACATTCTTGAGCAAGGCCATGCTCAGAAAATCCCTGCAGGAACCAC
Consensus (451) TCC ACATTCTTGAGCAAGGCCATGCTCAGAAAATCCCTGCAGGAACCAC
501 550
VR alpha (472) TTTCTTTTTGGTTAACCCTGACGACAACGAGAATCTCAGAATAATCAAAC
VR alpha' (500) TTTCTTTTTGGTTAACCCTGACGACAACGAAAATCTCAGAATAATCAAAC
VR beta (435) TTTCTTTTTGGTTAACCCTAACGACAACGACAATCTCAGAATAATCAAAC
Consensus (501) TTTCTTTTTGGTTAACCCTGACGACAACGA AATCTCAGAATAATCAAAC
551 600
VR alpha (522) TCGCCGTACCCGTTAACAACCCTCACAGATTTCAAGACTTTTTCCTATCT
VR alpha' (550) TCGCCGTACCCGTTAACAACCCTCACAGATTTCAAGACTTTTTCCTATCT
VR beta (485) TCGCCATACCTGTTAACAACCCTCACAGATTTCAAAACTTTTTCCTATCT
Consensus (551) TCGCCGTACCCGTTAACAACCCTCACAGATTTCAAGACTTTTTCCTATCT
601 650
VR alpha (572) AGCACAGAAGCACAACAATCCTACTTGCAAGGATTCAGCAAGAATATTCT
VR alpha' (600) AGCACAGAAGCCCAACAATCCTACCTGCAAGGATTCAGCAAGAATATTCT
VR beta (535) AGCACAGAAGCCCAGCAATCCTACTTGAGAGGCTTCAGCAAGAATATTCT
Consensus (601) AGCACAGAAGCCCAACAATCCTACTTGCAAGGATTCAGCAAGAATATTCT
651 700
VR alpha (622) AGAGGCCTCCTTCGATAGCGACATCAAGGAGATAAGCAGGGTTCTGTTTG
VR alpha' (650) AGAGGCCTCCTTCGATAGCGACATCAAGGAGATAAACAGGGTTCTGTTTG
VR beta (585) AGAGGCCTCCTTCGATAGCGAGTTCAAGGAGATAGACAGGGTTCTATTTG
Consensus (651) AGAGGCCTCCTTCGATAGCGACATCAAGGAGATAAACAGGGTTCTGTTTG
701 750


VR alpha (672) GAGAGGAGGGACAGCAACAACAGCAAGGCCAGGAGAGTCAGCAAGAGGGA
VR alpha' (700) GAGAGGAGGGACAGCAACAACAGCAAGGCCAGGAGAGTCAGCAAGAGGGA
VR beta (635) GAGAAGAGA---CAACAGCAGCATGGGAAGAGAGTCAGGAAGAGGGA
Consensus (701) GAGAGGAGGGACAGCAACAACAGCAAGGCCAGGAGAGTCAGCAAGAGGGA
751 800
VR alpha (722) GTGATTGTTGAACTTAAAAGGGAACAGATTCGGGAATTGACCAAACATGC
VR alpha' (750) GTGATTGTTGAACTTAAAAGGGAACAGATTCGGGAATTGACCAAACATGC
VR beta (682) GTGATTGTAGAACTTAAAAGGGAACAGATTCGGGAATTGATCAAACATGC
Consensus (751) GTGATTGTTGAACTTAAAAGGGAACAGATTCGGGAATTGACCAAACATGC
801 850
VR alpha (772) CAAATCTAGTTCAAAGAAATCCCTTTCCTCCGAAGATCAACCATTCAACC
VR alpha' (800) CAAATCTAGTTCAAAGAAATCCCTTTCCTCCGAAGATCAACCATTCAACC
VR beta (732) CAAATCTAGTTCAAGGAAAGAACTCTCCTCCCAAGATGAACCATTCAACC
Consensus (801) CAAATCTAGTTCAAAGAAATCCCTTTCCTCCGAAGATCAACCATTCAACC
851 900
VR alpha (822) TGCGAAACCAAAAACCCATCTACTCCAACAAACTTGGAAGGTGGTTTGAG
VR alpha' (850) TGAGAAACCAAAAACCCATCTACTCCAACAAATTTGGAGAGTTTTACGAG
VR beta (782) TGAGAAACAGCAATCCCATCTACTCCAACAAATTTGGAAGGTGGTATGAG
Consensus (851) TGAGAAACCAAAAACCCATCTACTCCAACAAATTTGGAAGGTGGTATGAG
901 950
VR alpha (872) ATCACTCCGGAGAAAAACCCCCAACTTAGGGACTTGGACATGTTCATCAG
VR alpha' (900) ATCACACCGAAGAAAAACCCCCAACTTAAGGATTTGGACGTGTTCATCAG
VR beta (832) ATCACCCGGAGAAAAACCCCCAACTTAAGGACTTGGACGTGTTCATCAG
Consensus (901) ATCAC CCGGAGAAAAACCCCCAACTTAAGGACTTGGACGTGTTCATCAG
951 1000
VR alpha (922) ATCTGTGGATATGAAAGAGGGAAGCCTTCTTCTGCCACACTACAATTCTA
VR alpha' (950) TTCCGTGGATATGAAAGAGGGAAGCCTTCTTCTGCCACACTACAATTCTA
VR beta (882) TTCTGTGGATATGAAAGAGGGAGCCTTCTTTGCCACACTACAACTCTA
Consensus (951) TTCTGTGGATATGAAAGAGGGAAGCCTTCTTCTGCCACACTACAATTCTA
1001 1050
VR alpha (972) AGGCCATAGTAATACTGGTGATTAATGAGGGAAAGCAAACATTGAACTT
VR alpha' (1000) AGGCCATAGTCATACTGGTGATTAATGAGGGAGAAGCAAACATTGAACTG
VR beta (932) AGGCCATAGTGATACTGGTGATTAATGAAGGAGAAGCAAAGATTGAACTT
Consensus (1001) AGGCCATAGT ATACTGGTGATTAATGAGGGAGAAGCAAACATTGAACTT
1051 1100
VR alpha (1022) GTTGGCCAAAGAGAACAACAGAAGCAGCAAGAAGAACAAGAGGAAAGTTG
VR alpha' (1050) GTTGGTCTAAGAGAAGAACAACAGCAGCA-GCAG--CAAGATGAAAGGTT
VR beta (982) GTTGGCCCAAGCG---ATCAGCAGCAGCA---------AGACGAAAGCTT
Consensus (1051) GTTGGCC AAGAGAA AACAGCAGCAGCA G AG CAAGA GAAAG TT
1101 1150
VR alpha (1072) GGAAGTGCAGAGGTATAGAGCTGAGTTGTCTGAAGACGATGTATTTATAA
VR alpha' (1097) GGAAGTGCAGAGGTACAGAGCTGAGGTGTCTGAAGACGATGTATTTGTAA
VR beta (1020) GGAAGTGCAGAGATATAGAGCTGAGTTGTCTGAAGACGATGTATTTGTAA
Consensus (1101) GGAAGTGCAGAGGTATAGAGCTGAGTTGTCTGAAGACGATGTATTTGTAA
1151 1200
VR alpha (1122) TCCCTGCAACTTATCCGGTTGCCATCAACGCTACCTCCAATCTGAATTTC
VR alpha' (1147) TCCCAGCAGCTTATCCAGTTGCCATCAACGCAACCTCCAATCTAAATTTC
VR beta (1070) TCCCAGCAGCTTATCCAGTTGCCATCAACGCAACCTCCAATCTAAATTTC
Consensus (1151) TCCCAGCAGCTTATCCAGTTGCCATCAACGCAACCTCCAATCTAAATTTC
1201 1250
VR alpha (1172) TTTGCTTTCGGCATCAATGCTGAGAACAACCAGAGGAACTTCCTTGCAGG
VR alpha' (1197) TTTGCTTTCGGTATCAATGCTGAGAACAACCAGAGGAACTTCCTTGCAGG
VR beta (1120) TTTGCTTTCGGTATCAATGCTGAGAACAACCAGAGGAATTTCCTTGCAGG
Consensus (1201) TTTGCTTTCGGTATCAATGCTGAGAACAACCAGAGGAACTTCCTTGCAGG
1251 1300
VR alpha (1222) TGAGAAAGACAATGTGATAAGCGAGATACCTACAGAGGTGTTGGATGTTA
VR alpha' (1247) TGAGAAAGACAATGTGATTAGTGAGATACCTACAGAGGTGTTGGATCTTG
VR beta (1170) TGAGAAAGACAATGTGATGAGCGAGATACCTACAGAGGTTTGGATGTTA
Consensus (1251) TGAGAAAGACAATGTGAT AGCGAGATACCTACAGAGGTGTTGGATGTTA
1301 1350
VR alpha (1272) CGTTCCCTGCGTCTGGTGAGAAGGTAAGAAGCTGATAAAGAAGCAGAGC
VR alpha' (1297) CTTTCCCTGCGCCTGGTGAGAAGGTTGAGAAGTTGGTACAGAAGCAGAGT
VR beta (1220) GCTTCCCTGCGTCTGGTAACAAGGTTGAGAAGTTGATAAAGAAGCAGAGT
Consensus (1301) C TTCCCTGCGTCTGGTGAGAAGGTTGAGAAGTTGATAAAGAAGCAGAGT
1351 1400
VR alpha (1322) GAATCCCAGTTTGTGGATGCACAGCCTGAGCAACAGGAAAGGGAGGAGGC
VR alpha' (1347) ACTTCCCACTTTGTTGATGCACAGCCTGAGGAACAGCAAAGGGAGGAGGG
VR beta (1270) GAATCCCACTTTGTGGATGCACAGCCTGAGCAACAGCAGAGGGAGGAAGG
Consensus (1351) GAATCCCACTTTGTGGATGCACAGCCTGAGCAACAGCAAAGGGAGGAGGG
1401 1450

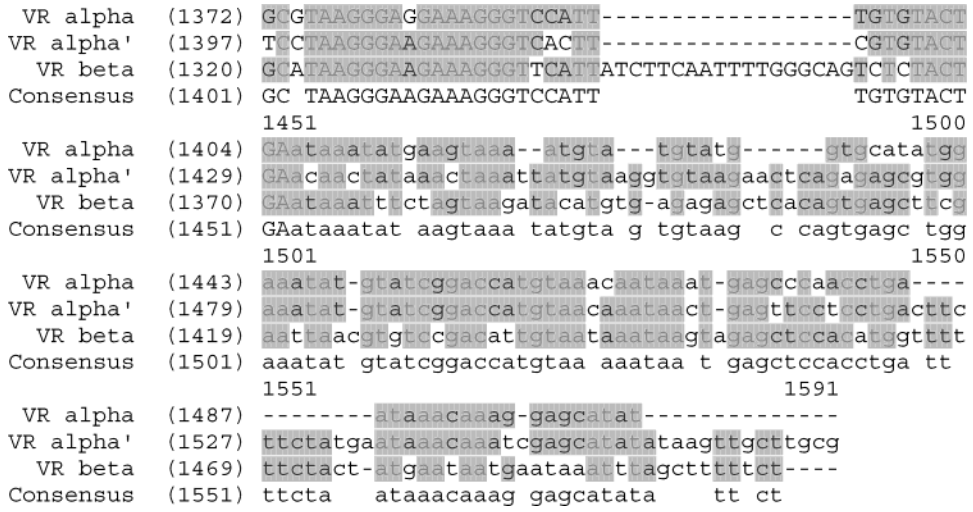


Figure 1. DNA sequences of three isoforms of mungbean 8S globulin. The untranslated regions (5′UTR and 3′UTR) are in lowercase letters while the translated regions are in uppercase; the propeptide site is marked with an arrowhead; VR, *Vigna radiata*.

characterized by a large proportion of charged and hydrophilic residues (*28*). The α and α′ subunits of soybean β-conglycinin contain extension regions (α, 125 residues; α′, 141 residues) at the N termini, which are rich in acidic amino acids and which exhibit high (>75%) homologies (*22*).

A single site for glycosylation (N-X-S/T) was observed within the highly conserved region about 90 amino acids from the C terminus. Similarly, the α and α′ subunits of soybean β-conglycinin contain two high-mannose glycans in the core regions while the β subunit contains one (*7*).

Homology Analysis of the Mungbean 8S Isoforms. Homology analysis was done on both the cDNA and the amino acid sequences. The cDNA sequences of the three isoforms compared with those from other plants using BLASTN (*12*) showed high (81−84%) homology with the α, α′, and β subunits of soybean β-conglycinin over 900−1000 bases (data not shown). Homology with *Phaseolus vulgaris* and *Phaseolus lunatus* vicilin type genes was even higher at 86−89% over 600−900 bases. With other legumes, there are short segments (100−300 bases) in their DNA sequences, which showed high homology (84−91%) with corresponding short segments in the mungbean 8S α, α′, and β sequences. It was noted that the sequence 774−788 of 8Sα and the corresponding sequences in 8Sα′ and 8Sβ are the same as repeat regions of 711−725 in phaseolin of *P. vulgaris* (*19*). However, the repeat region of 726−740 in phaseolin is not exactly the same in mungbean where there are now inserted nucleotides within this region.

Figure 3 shows the phylogenetic tree analysis of the cDNA sequences of the mungbean 8S isoforms and vicilin type proteins of soybean (*Glycine max*), kidney bean (*P. vulgaris*), and lima bean (*P. lunatus*) using the Vector NTI (*11*). The mungbean 8S isoforms are clustered together and are separate from, although quite close to, the *Phaseolus*. Previously, mungbean was classified under the *Phaseolus* family as *Phaseolus aureus.* Goel et al. (*33*) recently affirmed the recent classifications of the *Vigna* and the *Phaseolus* based largely on morphological, biochemical, cytogenetical, and palynological features by using internal transcribed spacer (ITS) sequences. The mungbean 8S isoforms are also farthest from the vicilin type of proteins from sesame, wheat, and oil palm.

Comparison of the amino acid sequences of the three isoforms of mungbean 8S globulins with those of 7−8S type of proteins using BLASTP (*12*) shows homology ranging from 45 to 68% within the legume family and a lower 29 to 34% for crops of different species (data not shown). Both the α and the β subunits of soybean β-conglycinin showed the highest homology of 58−61% with 8Sα′, 59−61% with 8Sα, and 66−68% with 8Sβ. Phaseolin of *P. vulgaris* also showed high homology of 50−55% with the mungbean 8S globulins. Vicilin of pea (*Pisum sativum*) and faba bean (*Vicia faba*) was 50−52%, and canavalin from jackbean and swordbean was 46 to 52% homologous with the 8S mungbean globulins. Similar types of proteins from *Sesamum indica*, *Cucurbita maxima*, and *Macademia* were least homologous at 31−33%.

Expression in *E. coli* and Purification of Recombinant Mungbean 8Sα. Before expression experiments were conducted, the plasmid vector containing the 8Sα gene was purified and sequenced to check the fidelity of the inserted gene.

The mungbean 8Sα globulin was expressed in three *E. coli* strains, namely, HMS174(DE3), JM109(DE3), and BL21(DE3). Expression of 8Sα was highest in HMS174(DE3). Thus, this strain was used for preliminary and large scale expression of the 8Sα protein.

On the basis of preliminary small scale purification, the following steps were found to efficiently purify recombinant mungbean 8Sα globulin from *E. coli*: 0−35% ammonium sulfate fractionation followed by hydrophobic interaction chromatography with a gradient of 30−0% ammonium sulfate and by chromatography on Mono Q. **Figure 4** shows the purification of 8Sα with a molecular weight of 48 000 as monitored by using SDS gel electrophoretic patterns. Protein recovery was 75.8% (23.6 mg).

The first six amino acids of the recombinant mungbean 8Sα were MIDG(A)AE. The presence of methionine signifies bacterial translation. The next five amino acids are similar to the first five amino acids of the 8Sα obtained in this paper. It was noted though that the fourth amino acid is a mixture of G and A as evident in the chromatogram (not shown). This mixture could be due to mistranslation of G or recombinant protein whose initial methionine was removed.

Because of the molecular heterogeneity of most of seed storage globulins, it has been difficult to study the structure−physicochemical function relationships of constituent subunits. To obtain large amounts of homogeneous subunit species, specific subunit species has been expressed in *E. coli* systems (*22*, *34*) or purified from mutant soybean cultivars containing only specific subunit species of β-conglycinin (*35*, *36*). Previous studies using *E. coli* expression system showed that neither the

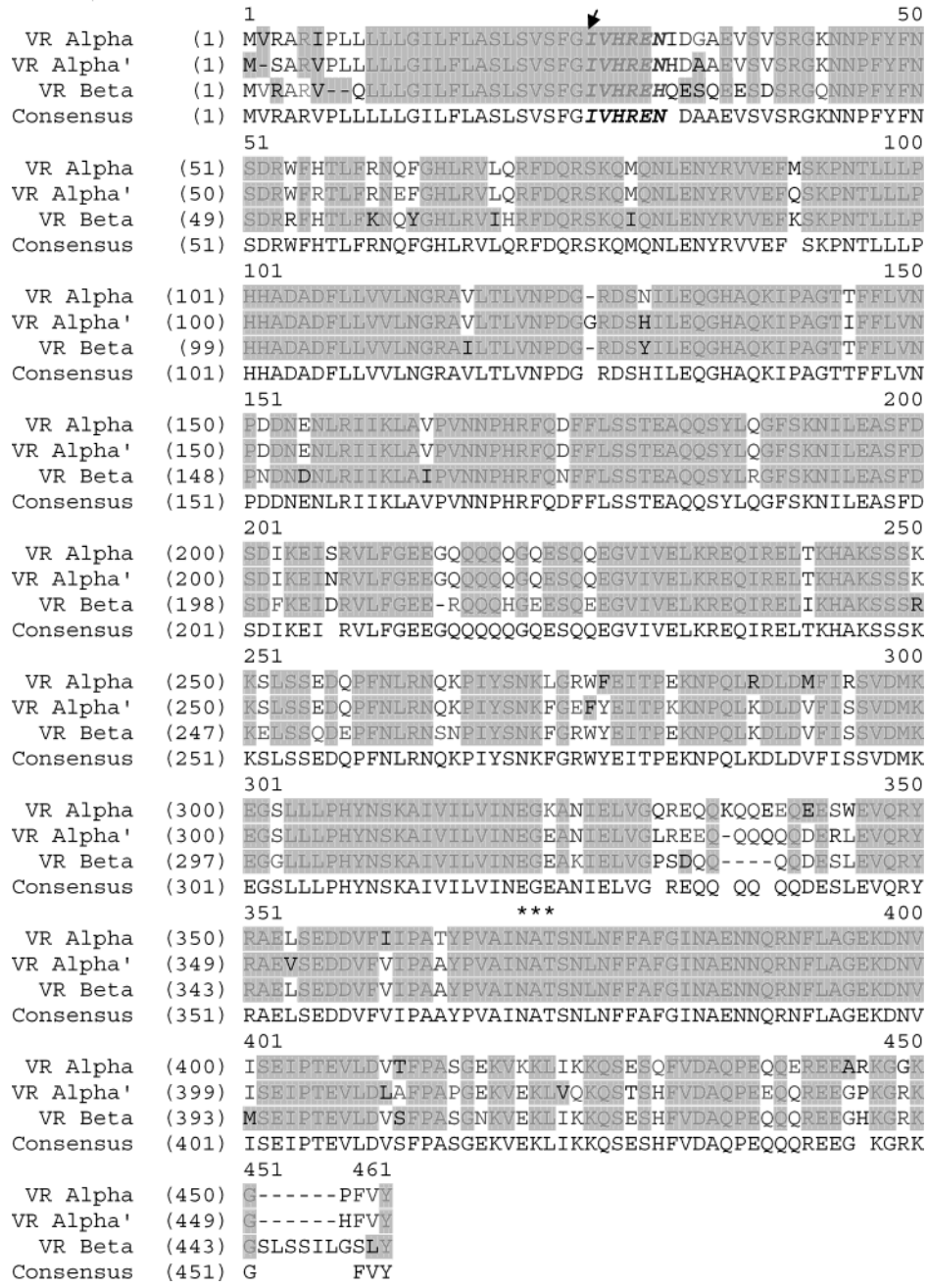


Figure 2. Comparison of the deduced amino acid sequences of the three isoforms of 8S mungbean globulins showing the cleavage site of the signal peptide (arrowhead), propeptide (italicized bold-faced letters), and glycosylation site (asterisks); VR, *Vigna radiata*.

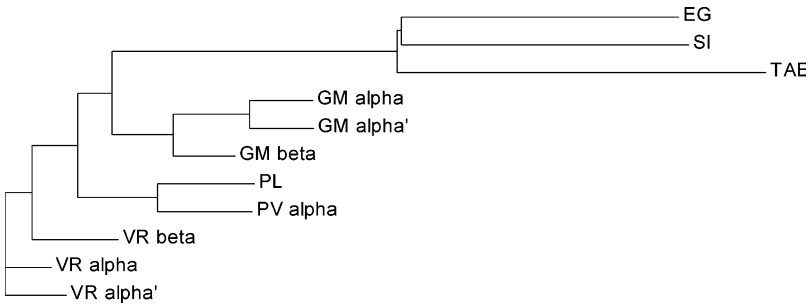


Figure 3. Phylogenetic analysis of the DNA sequences of the mungbean 8S isoforms and vicilin type cDNA from other crops using the Vector NTI software (*11*). Legends: *Vigna radiata* (VR) 8S α, α′, and β, sequences from this paper; *P. vulgaris* (PVα) α type (X02980), ref *19*; *P. lunatus* (PLα) α type (U01122), ref *29*; *G. max* (GMα) α, β-conglycinin (AB008678); *G. max* (GMα′) α′, β-conglycinin (AB008680); *G. max* (GMβ) β, β-conglycinin (AB008679) (*22*); *T. aestivum* (TAE) (AAA34269.1), ref *30*; *Sesame indicum* (SI) (AAK15089) (*31*); *Elaeis guineensis* (EG) (AAK28402.1) (*32*).

propeptide region, the extension region, nor the glycans are essential for the folding and self-assembly of β-conglycinin (*22*, *34*). By comparing homotrimers purified from mutant soybean and recombinant homotrimers, which lack the N-linked glycans,

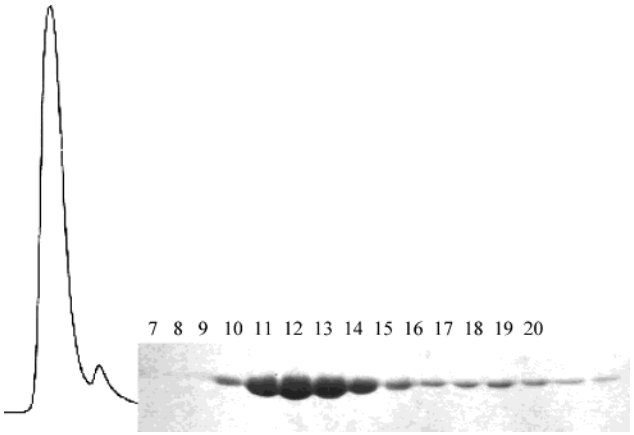


Figure 4. MONO Q chromatogram and SDS–PAGE profile of fractions from the main peak of the hydrophobic interaction chromatography step. Fractions 9–13 were dialyzed and concentrated. Protein yield was 16.9 mg/mL. The pure protein was used for crystallization experiments.

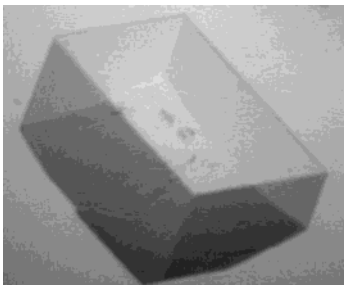


Figure 5. Sample crystal of mungbean r8Sα globulin, which developed under the following conditions: 0.1 M MES buffer, pH 6.0, containing 12% PEG 1000 and 0.20 M NaCl. The crystallization was allowed to proceed at 20 °C.

it was shown that such modification contributes to solubilities at low ionic strength, to emulsifying ability, and also to prevent heat-induced associations of the native α and α′ homotrimers but not to the thermal stability of β-conglycinin (*37*).

Crystallization of Recombinant and Native Mungbean 8Sα Globulins. Recombinant 8Sα crystals were observed under conditions of 0.1 M MES, pH 6.0, 12% PEG 1000, and 20 °C (**Figure 5**). The crystals belong to the rhombohedral space group and measure 0.4 mm × 0.3 mm × 0.2 mm. Diffractions were obtained at 2.6 Å resolution. The fact that the recombinant 8Sα forms crystals that are suitable for X-ray analysis indicates that the recombinant 8Sα is forming the correct structure identical to that of the native 8Sα. On the other hand, no crystals were obtained from the native form. Heterogeneity of the native 8S globulin might have deterred the formation of crystals.

The crystal structure of soybean storage proteins has recently been elucidated using crystals of recombinant protein or protein purified from mutant soybean lines (*34*, *38*, *39*). Earlier, the structures of French bean phaseolin (*40*) and jackbean canavalin (*41*) had been determined by X-ray crystallography using crystals obtained from native proteins.

This paper reported the successful isolation and cloning of three isoforms of the mungbean 8S globulin genes. Their DNA and derived amino acid sequences were analyzed and characterized using databases and appropriate softwares. The major 8Sα isoform was expressed in *E. coli*, and the recombinant protein was purified and successfully crystallized. The characterization of the physicochemical and functional properties of the recombinant 8Sα mungbean globulin and the detailed analysis of its crystal structure by X-ray diffraction will be reported in separate papers. The latter paper will also discuss the differences in the properties of mungbean 8Sα and soybean β-conglycinin based on their crystal structures.

ACKNOWLEDGMENT

The assistance of Kimiko Inui, Yumiko Kaneda, and Asako Hasimoto in the conduct of the experiments is much appreciated.

LITERATURE CITED

(1) Singha, S. K., Paroda, R. S., Eds. *Production of Pulse Crops in Asia and the Pacific Region*; RAFA, FAO: Bangkok, 1995; 248 pp.

(2) Tecson-Mendoza, E. M.; Adachi, M.; Bernardo, A. E. N.; Utsumi, S. Mungbean [*Vigna radiata* (L.) Wilczek] globulins: purification and characterization. *J. Agric. Food Chem*. **2001**, *49*, 1552−1558.

(3) Ericson, M. C.; Chrispeels, M. J. Isolation and characterization of glucosamine-containing storage glycoproteins from the cotyledons of *Phaseolus aureus*. *Plant Physiol*. **1973**, *52*, 98−104.

(4) Bhattacharya, S. P.; Biswas, B. B. Purification and characterization of high salt-soluble vicilin from mungbean (*Vigna radiata*). *Biochem. Int.* **1990**, *21*, 667−675.

(5) Kim, C. S.; Kamiya, S.; Utsumi, S.; Kito, M. Improvement of nutritional value and functional properties of soybean glycinin by protein engineering. *Protein Eng.* **1990**, *3*, 725−731.

(6) De Lumen, B. O. Molecular approaches to improving the nutritional and functional properties of plant seed as food sources: developments and comments. *J. Agric. Food Chem.* **1990**, *38*, 1779−1788.

(7) Utsumi, S.; Katsube, T.; Ishige, T.; Takaiwa, F. Molecular design of soybean glycinins with enhanced food qualities and development of crops producing such glycinins. In *Food Proteins and Lipids*; Damodaran, S., Ed.; Plenum Press: New York; 1997; pp 1−15.

(8) Katsube, T.; Kurisawa, N.; Ogawa, M.; Ohtsuka, R.; Utsumi, S.; Takaiwa, F. Accumulation of soybean glycinin and formation of its assembly with glutelins in rice. *Plant Physiol.* **1999**, *120*, 1063−1073.

(9) Sun, X.; Ao, G.; Yu, J.; Zhao, Q. Transfer of a lysine-rich gene into maize inbred lines and the detection of transgenic plants. *J. Agric. Biotechnol.* **2001**, *9*, 156−158.

(10) Hall, T. C.; Ma, Y.; Buchbinder, B. W.; Pyne, J. W.; Sun, S. M.; Bliss, F. A. Messenger RNA for G1 protein of French bean seeds: Cell-free translation and product characterization. *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 3196−3200.

(11) Informax Inc. *Vector NTI Suite*, Version 5.1; MD USA, 1998.

(12) Altschul, S. F.; Madden, T. L.; Schäffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. Gapped BLAST and PSI−BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389−3402.

(13) Nielsen, H.; Engelbrecht, J.; Brunak, S.; von Heijne, G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **1997**, *10*, 1−6.

(14) Bradford, M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Chem.* **1976**, *72*, 248−254.

(15) Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (London)* **1970**, *227*, 680−685.

(16) Rhodes, G. *Crystallography Made Crystal Clear*, 2nd ed.; Academic Press: CA, 2000.

(17) Thanh, V. H.; Shibasaki, K. Heterogeneity of β-conglycinin. *Biochim. Biophys. Acta* **1976**, *439*, 326−338.

(18) Thanh, V. H.; Shibasaki, K. Major proteins of soybean seeds. Subunit structure of β-conglycinin. *J. Agric. Food Chem.* **1978**, *26*, 692−695.

(19) Slightom, J. L.; Drong, R. F.; Klassy, R. C.; Hoffman, L. M. Nucleotide sequences from phaseolin cDNA clones: the major storage proteins from *Phaseolus vulgaris* are encoded by two unique gene families. *Nucleic Acids Res.* **1985**, *13*, 6483−6498.

(20) Doyle, J. J.; Schuller, M. A.; Godette, W. D.; Zenger, V.; Beachy, R. N.; Slightom, J. L. The glycosylated seed storage proteins of *Glycine max* and *Phaseolus vulgaris*. Structural homologies of genes and proteins. *J. Biol. Chem.* **1986**, *261*, 9228−9238.

(21) Lycett, G. W.; Delauney, A. J.; Gatehouse, J. A.; Gilroy, J.; Croy, R. R.; Boulter, D. The vicilin gene family of pea (*Pisum sativum* L.): a complete cDNA coding sequence for preprovicilin. *Nucleic Acids Res.* **1983**, *11*, 2367−2380.

(22) Maruyama, N.; Katsube, T.; Wada, Y.; Oh, M. H.; Barba dela Rosa, A. P.; Okuda, E.; Nakagawa, S.; Utsumi, S. The roles of the N-linked glycans and extension regions of soybean β-conglycinin in folding, assembly and structural features. *Eur. J. Biochem.* **1998**, *258*, 854−862.

(23) Dunwell, J. M. Cupins: a new superfamily of functionally diverse proteins that include germins and plant storage proteins. *Biotechnol. Genet. Eng. Rev.* **1998**, *15*, 1−32.

(24) Dunwell, J. M.; Khuri, S.; Gane, P. J. Microbial relatives of the seed storage proteins of higher plants: conservation of structure and diversification of function during evolution of the cupin superfamily. *Microbiol. Mol. Biol. Rev.* **2000**, *64*, 153−179.

(25) Nielsen, N. C. The structure and complexity of the 11S polypeptides in soybeans. *J. Am. Oil Chem. Soc.* **1985**, *62*, 1680−1686.

(26) Croy, R. R. D.; Gatehouse, J. A.; Evans, I. M.; Boulter, D. Characterisation of the storage protein subunits synthesised in vitro by polyribosomes and RNA from developing pea (*Pisum sativum* L.). *Planta* **1980a**, *148*, 57−63.

(27) Croy, R. R. D.; Gatehouse, J. A.; Tyler, M.; Boulter, D. The purification and characterization of a third storage protein (convicilin) from the seeds of pea (*Pisum sativum* L.). *Biochem. J.* **1980b**, *19*, 509−516.

(28) Bown, D.; Noel Ellis, T. H.; Gatehouse, J. A. The sequence of a gene encoding convicilin from pea (*Pisum sativum* L.) shows that convicilin differs from vicilin by an insertion near the *N*-terminus. *Biochem. J.* **1988**, *251*, 717−726.

(29) Kami, J. A.; Gepts, P. Phaseolin nucleotide sequence diversity in Phaseolus. I. Intraspecific diversity in *Phaseolus vulgaris*. *Genome* **1994**, *37*, 751−757.

(30) Litts, J. C.; Colwel, G. W.; Chakerian, R. L.; Quatrano, R. S. Genomic nucleotide sequence of a *Triticum aestivum* 7S globulin (Gbl1) storage protein gene. Unpublished, 1991; http://www.ncbi.nlm.nih.gov.entrez.

(31) Tai, S. S. K.; Tzen, J. T. C. Molecular Cloning of three storage proteins in sesame. Unpublished, undated; http://www.ncbi.nlm.nih.gov.entrez.

(32) Morcillo, F.; Hartmann, C.; Duva, Y.; Tregear, J. Regulation of 7S globulin gene expression in zygotic and somatic embryos of oil palm. Unpublished, undated; http://www.ncbi.nlm.nih.gov.entrez

(33) Goel, S.; Raina, S. N.; Ogihara, Y. Molecular evolution and phylogenetic implications of internal transcribed spacer sequences of nuclear ribosomal DNA in the *Phaseolus-Vigna* Complex. *Mol. Phylogenet. Evol.* **2002**, *22*, 1−19.

(34) Maruyama, N.; Adachi, M.; Takahashi, K.; Yagasaki, K.; Kohno, M.; Takenaka, Y.; Okuda, E.; Nakagawa, A.; Mikami, B.; Utsumi, S. Crystal structures of recombinant and native soybean β-conglycinin homotrimers. *Eur. J. Biochem.* **2001**, *268*, 3595−3604.

(35) Takahashi, K.; Banba, H.; Kikuchi, A.; Ito, M.; Nakamura, S. An induced mutant line lacking the α subunit of β-conglycinin in soybean (*Glycine max* (L.) Merril). *Breed Sci.* **1994**, *44*, 65−66.

(36) Takahashi, K.; Mizuno, Y.; Yumoto, S.; Kitamura, K.; Nakamura, S. Inheritance of the α-subunit deficiency of β-conglycinin in soybean (*Glycine max* (L.) Merril) line induced by γ-ray irradiation. *Breed Sci.* **1996**, *46*, 251−255.

(37) Maruyama, N.; Mohamad Ramlan, M. S.; Takahashi, K.; Yagasaki, K.; Goto, H.; Hontani, N.; Nakagawa, S.; Utsumi, S. Structure-physicochemical function relationships of soybean β-conglycinin heterotrimers. *J. Agric. Food Chem.* **2002**, *50*, 4323−4326.

(38) Adachi, M.; Takenaka, Y.; Gidamis, A. B.; Mikami, B.; Utsumi, S. Crystal structure of soybean proglycinin A1aB1b homotrimer. *J. Mol. Biol.* **2001**, *305*, 291−305.

(39) Adachi, M.; Kanamori, J.; Masuda, T.; Yagasaki, Y.; Kitamura, K.; Mikami, B.; Utsumi, S. Crystal structure of soybean 11S globulin: Glycinin A3B4 homohexamer. *Proc. Natl. Acd. Sci. U.S.A.* **2001**, *100*, 135−1400.

(40) Lawrence, M. C.; Izard, T.; Beuchat, M.; Blagrove, R. J.; Colman, P. M. Structure of phaseolin at 2.2 A resolution. Imiplication for a common vicilin/legumin structure and the genetic engineering of seed storage proteins. *J. Mol. Biol.* **1994**, *238*, 748−776.

(41) Ko, T. P.; Ng, J. D.; McPherson, A. The three-dimensional structure of canavalin from jack bean (*Canavalia ensiformis). Plant Physiol.* **1993**, *101*, 729−744.

Received for review August 11, 2003. Revised manuscript received February 3, 2004. Accepted February 3, 2004. The study grant awarded to A.E.N.B. and R.N.G. by Heiwa Nakajima through Prof. S. Utsumi of Kyoto University is gratefully acknowledged.

JF0305938