JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Gene Families Encoding Isoforms of Two Major Sesame Seed Storage Proteins, 11S Globulin and 2S Albumin†

ERIC S. L. HSIAO,‡ LI-JEN LIN,‡ FENG-YIN LI,§ MIKI M. C. WANG,#
MING-YUAN LIAO,# AND JASON T. C. TZEN*,‡

Graduate Institute of Biotechnology and Department of Chemistry, National Chung Hsing University, Taichung, Taiwan; and Graduate Institute of Biomedicine and Biomedical Technology, National Chi Nan University, Puli, Taiwan

Sesame (*Sesamum indicum* L.) seed has been recognized as a nutritional protein source owing to its richness in methionine. Storage proteins have been implicated in allergenic responses to sesame consumption. Two abundant storage proteins, 11S globulin and 2S albumin, constitute 60−70 and 15−25% of total sesame proteins, respectively. Two gene families separately encoding four 11S globulin and three 2S albumin isoforms were identified in a database search of 3328 expresssed sequence tag (EST) sequences from maturing sesame seeds. Full-length cDNA sequences derived from these two gene families were completed by PCR using a maturing sesame cDNA library as the template. The amino acid compositions of these deduced storage proteins revealed that the richness in methionine is attributed mainly to two 2S albumin isoforms and partly to one 11S globulin isoform. The presence of four 11S globulin and three 2S albumin isoforms resolved in SDS-PAGE was confirmed by MALDI-MS analyses. The abundance of these isoforms was in accord with the occurrence frequency of their EST sequences in the database. A comprehensive understanding of these storage proteins at the molecular level may also facilitate the identification of allergens in crude sesame products that have caused severe allergic reactions increasingly reported in the past decade.

KEYWORDS: 11S globulin; 2S albumin; MALDI-MS; seed storage protein; sesame

INTRODUCTION

Sesame seed has been well recognized as a nutritional protein source owing to its richness in sulfur-containing amino acids, particularly methionine (*1*). Proteins represent approximately 15−25% of the dry weight of sesame seed, or 30−50% mass of the defatted sesame cake. Compared with the standard values recommended for dietary requirement by the Food and Agriculture Organization and the World Health Organization, sesame proteins are slightly lower in lysine but richer in methionine, cystine, arginine, and leucine (*2*). In terms of essential amino acid composition, sesame seed proteins are comparable to those in beef and milk, with the exception of lysine.

Conventionally termed α-globulin and β-globulin, 11S globulin and 2S albumin are the two major seed storage proteins that constitute 60−70 and 15−25% of total sesame proteins, respectively (*3*, *4*). Sesame contains several 11S globulin isoforms that are assembled, in a random combination, to form hexamers of 300−350 kDa (*5*). Each 11S globulin isoform consists of an acidic subunit (30−40 kDa) and a basic subunit (20−25 kDa) linked by a single disulfide bond (*6*). The two subunits are encoded by a single gene producing a precursor protein of 50−60 kDa, which is cleaved post-translationally by a unique asparaginyl endopeptidase after the formation of an interchain disulfide bond between the N-terminal acidic and the C-terminal basic subunits. Similarly, several 2S albumin isoforms are present in sesame seed (*7*); and each 2S albumin isoform consists of a small subunit (4 kDa) and a large subunit (9 kDa) linked by two disulfide bonds. These small and large subunits of 2S albumins occur as specific pairs because they are processed from a single gene product (17 kDa), possibly through cleavage by the same asparaginyl endopeptidase involved in the post-translational processing of 11S globulin.

Cases of allergy to sesame have been increasingly reported in the past decade, possibly due to the inclusion in the everyday diet of sesame seed products in countries or regions where people were not exposed to these items (*8*−*12*). Sesame seed has been added to the list of food allergens in European countries and Canada but not in the United States. Because crushed seed or extracted oil of sesame is occasionally incorporated in food

* Author to whom correspondence should be addressed (telephone 886-4-22840328; fax 886-4-22853527; e-mail TCTZEN@dragon.nchu.edu.tw).
† The nucleotide sequences reported in this paper have been submitted to the GenBank Database under accession no. DQ256292 (2S albumin isoform 3), DQ256293 (11S globulin isoform 3), and DQ256294 (11S globulin isoform 4).
‡ Graduate Institute of Biotechnology, National Chung Hsing University.
§ Department of Chemistry, National Chung Hsing University.
# National Chi Nan University.

10.1021/jf061505x CCC: $33.50
Published on Web 11/16/2006

Table 1. EST Sequences Derived from 11S Globulin cDNA Fragments

isoform	accession no. of EST sequences				
11S-1 (44)[a]	BU668714	BU669471	BU668627	BU667406	BU669376
	BU668764	BU669442	BU667388	BU669080	BU668353
	BU669998	BU670470	BU669369	BU668793	BU669964
	BU667495	BU670660	BU669464	BU668805	BU670175
	BU668279	BU667452	BU667727	BU667581	BU670289
	BU669490	BU670280	BU668314	BU669294	BU669048
	BU667967	BU667543	BU669564	BU669583	BU667976
	BU668534	BU667369	BU667817	BU668982	BU668186
	BU667970	BU667400	BU668356	BU668230	
11S-2 (9)	BU670505	BU670191	BU668616	BU668745	BU670484
	BU668671	BU670353	BU667919	BU669978	
11S-3 (19)	BU669589	BU670389	BU668893	BU670321	BU669440
	BU670452	BU667731	BU667771	BU668417	BU669870
	BU668545	BU667836	BU667582	BU667881	BU667379
	BU669245	BU667783	BU667551	BU667485	
11S-4 (28)	BU669133	BU669545	BU668138	BU669336	BU668261
	BU669222	BU669079	BU670112	BU670249	BU668755
	BU669306	BU667799	BU667984	BU670594	BU669890
	BU670070	BU669093	BU669496	BU669239	BU669540
	BU669963	BU669292	BU669868	BU668423	BU667837
	BU669054	BU668021	BU667525		

[a] Total number of EST sequences of each isoform is indicated in parentheses.

Table 2. EST Sequences Derived from 2S Albumin cDNA Fragments

isoform	accession no. of EST sequences				
2S-1 (25)[a]	BU668750	BU668577	BU670226	BU669307	BU669060
	BU668593	BU669485	BU669382	BU669445	BU670007
	BU669809	BU669435	BU668922	BU667663	BU668300
	BU667410	BU670155	BU669834	BU668803	BU668296
	BU668989	BU669102	BU669413	BU667536	BU668002
2S-2 (46)	BU668361	BU668655	BU668584	BU670262	BU668316
	BU670672	BU669196	BU667764	BU670034	BU669643
	BU669510	BU669229	BU669030	BU668760	BU668351
	BU667619	BU667454	BU668206	BU668388	BU667872
	BU669891	BU668221	BU668031	BU668468	BU669546
	BU668968	BU669554	BU669552	BU668762	BU670198
	BU669804	BU669519	BU668976	BU669885	BU668973
	BU667834	BU667679	BU667712	BU667533	BU668900
	BU670681	BU670228	BU668558	BU669681	BU668215
	BU669063				
2S-3 (13)	BU669448	BU668679	BU667662	BU667500	BU668302
	BU670517	BU667566	BU667458	BU670351	BU668143
	BU667634	BU668566	BU669539		

[a] Total number of EST sequences of each isoform is indicated in parentheses.

products as a minor ingredient without declaration, it may be classified as a "hidden" allergen (*13*). The allergic factors in sesame are frequently ascribed to the two abundant seed storage proteins, 11S globulin and 2S albumin (*14*−*16*). Although cDNA fragments encoding two 11S globulin and two 2S albumin isoforms have been sequenced, the gene families encoding these two storage proteins have not been completely identified.

Recent technical advancement has demonstrated that mass spectrometry is a powerful tool for the structural characterization of proteins (*17*). Massive protein sequences deduced from their corresponding nucleic acid sequences have provided a rich source for protein identification by peptide mass fingerprinting (*18*). For a comprehensive understanding of the two major sesame storage proteins at the molecular level, we intended to identify and sequence all members in the two gene families encoding 11S globulin and 2S albumin isoforms by analyzing a published database of 3328 expressed sequence tag (EST) sequences from maturing sesame seeds (*19*). Corresponding proteins of the identified clones were subsequently confirmed in the extract of sesame seed by matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) analyses. The results of this study may prove to be beneficial for the diagnosis of allergens in pertinent food products.

MATERIALS AND METHODS

Plant Materials. Mature and fresh maturing sesame (*Sesamum indicum* L.) seeds were gifts from the Crop Improvement Department, Tainan District Agricultural Improvement Station. Mature seeds were used for the preparation of crude storage proteins, and maturing seeds approximately 20 days after pollination were used for the construction of a cDNA library.

Database Searching for Gene Families Encoding 11S Globulin and 2S Albumin Isoforms from EST Sequences of Maturing Sesame Seeds. Four cDNA sequences (GenBank accession no. AF091842, AF240004, AF091841, and AF240005) encoding two 11S globulin and two 2S albumin storage proteins, named 11S-1, 11S-2, 2S-1, and 2S-2 in this study, have been sequenced (*6*, *7*). To find out the other members of gene families encoding these two storage proteins, the above four protein sequences were used as entrance queries to search for identical and homologous clones (with sequence identity >90 and 30%, respectively) in a database of 3328 EST sequences from maturing sesame seed (*19*) by the tblastn program (http://www.ncbi.nlm.nih.gov/blast/) at the National Center for Biotechnology Information (*20*). Conservative cysteine residues forming disulfide bonds in 11S globulin and 2S albumin were used to double-check the accuracy of putative clones (*21*). Among the homologous clones found in the database search, three EST groups encoding two 11S globulin and one 2S albumin isoforms, named 11S-3, 11S-4, and 2S-3, were identified. To identify all of the members of each isoform in the 3328 sesame EST sequences, the complete cDNA sequences of all 11S globulin and 2S albumin isoforms were used as entrance queries to search for identical sequences in the database.

Isolation of Poly(A)$^+$ RNA and cDNA Library Construction. Total RNA was extracted from maturing seeds (24 days after flowering) and ground in liquid nitrogen using the phenol/SDS method (*22*). Poly(A)$^+$ RNA, isolated with Dynabeads (Dynal) following the manufacturer's instructions, was dissolved in diethyl pyrocarbonate-treated water and then quantified as the absorbance at 260 nm with a spectrophotometer. cDNA was synthesized from poly(A)$^+$ RNA according to the protocol described in the manufacturer's instructions (cDNA synthesis, ZAP-cDNA synthesis, and ZAP-cDNA Gigapack III Gold Cloning kits purchased from Stratagene). A cDNA library of approximately 10^6 plaques was constructed with 5 μg of poly(A)$^+$ RNA.

PCR Cloning of cDNA Fragments Encoding 11S-3, 11S-4, and 2S-3. Two pairs of specific primers were used to obtain upstream and downstream overlapping fragments of a cDNA sequence encoding 11S-3 by PCR amplification using the excised phagemids from the cDNA library as templates. For the upstream fragment, a 24-nucleotide primer (5′-TATTCAGGCAAGACCAATCTGAGC-3′) and an 18-nucleotide primer (5′-AAGCAATCACCTCCATGG-3′) were separately designed according to two putative 11S-3 EST clones (BU669440 and BU668417). For the downstream fragment, a T7 primer from the phagemid vector (5′-GTAATACGACTCACTATAGGGC-3′) and a 23-nucleotide primer (5′-CAACCCAGAGGCGAGGGCAGGAG-3′) designed according to a putative 11S-3 EST clone (BU669440) were used. The two PCR fragments were harvested, ligated into the pGEM-T Easy Vector (Promega, Madison, WI), and subjected to sequencing using the core facility at the Biotechnology Service Center of National Chung-Hsing University, Taiwan. The upstream and downstream fragments were 877 and 1007 bp, respectively.

The 11S-4 sequence was also obtained by PCR cloning using the same strategy as described above for 11S-3. The paired primers used for cloning the upstream or downstream fragment were a 24-nucleotide primer (5′-ATCGCGGCGGCCGTAATATGGCTC-3′) and a 28-nucleotide primer (5′-CTAAGTGTTTCCTCCACCATGGCTAAGC-3′)

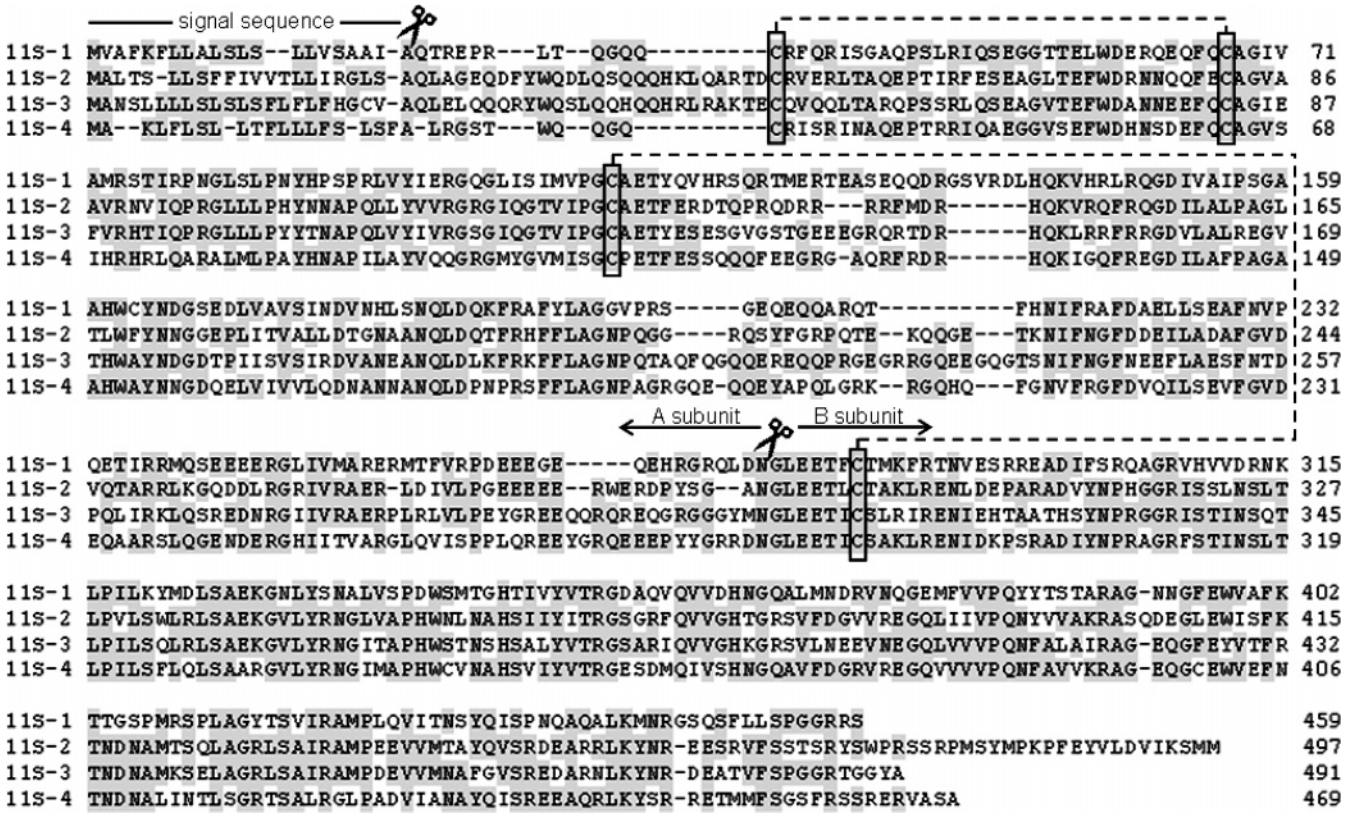

Figure 1. Sequence alignment of the precursor polypeptides of four sesame 11S globulin isoforms. The amino acid number for the last residue in each line is listed on the right for each isoform. Broken lines in the sequences represent gaps introduced for best alignment. The cleavage site of the putative N-terminal signal sequence and the consensus asparaginyl cleavage site for splitting these polypeptides into A (acidic) and B (basic) subunits are indicated by scissors symbols, individually. Four conserved cysteine residues are boxed and predicted to form an intrachain disulfide bond within A subunit and an interchain disulfide bridge between A and B subunits.

separately designed according to two putative 11S-4 EST clones (BU670249 and BU670594) or the T7 primer from the phagemid vector and a 24-nucleotide primer (5′-CAGGTTATCAGCCCACCTCTCC-AG-3′) designed according to a putative 11S-4 EST clone (BU670249). The upstream and downstream fragments were 858 and 823 bp, respectively.

For 2S-3 cloning, a pair of specific primers, the T7 primer from the phagemid vector and a 24-nucleotide primer (5′-CACTCATATA-TACAACTGTAGATG-3′) designed according to a putative 2S-3 EST clone (BU669448), were used for PCR amplification. A PCR fragment of approximately 635 bp was harvested and subjected to sequencing.

Differential Solubility Extraction of Sesame Seed Proteins. Proteins of mature sesame seeds were extracted with a medium containing 0.6 M sucrose and 10 mM sodium phosphate buffer, pH 7.5, by homogenization using a Polytron at 4 °C (*23*). The homogenate was filtered through three layers of cheesecloth. After filtration, the crude extract was separated into three fractions (pellet, supernatant, and oil layer) by centrifugation at 10000*g* for 15 min. To avoid the interference of the abundant oil, the pellet containing both 11S globulin and 2S albumin was collected for further analyses.

Separation of 11S Globulin and 2S Albumin Isoforms by SDS-PAGE. For SDS-PAGE analysis, the pellet was extracted with the sample buffer containing 62.5 mM Tris-HCl, pH 6.8, 2% SDS, 0.02% bromophenol blue, and 10% glycerol without β-mercaptoethanol according to the Bio-Rad instruction manual. To resolve 11S globulin isoforms, the separating gel was composed of 12.5% polyacrylamide, and the electrophoresis was performed under 200 V for 120 min. A separating gel of 18% polyacrylamide was used to resolve 2S albumin isoforms, and the electrophoresis was performed under 200 V for 70 min. Following electrophoresis, the gels were stained with Coomassie Blue R-250.

MALDI-MS Identification of 11S Globulin and 2S Albumin Isoforms. Four and two protein bands of putative 11S globulin and 2S albumin isoforms resolved, respectively, in two different SDS-PAGE gels were manually excised from the ground into pieces. The gel pieces were washed twice with 50% acetonitrile and 50% acetonitrile/25 mM ammonium bicarbonate. The proteins in gels were then reduced and alkylated at 56 °C for 45 min in 10 mM dithiothreitol and 55 mM iodoacetamide in 25 mM ammonium bicarbonate, followed by in-gel digestion overnight at 37 °C with 0.1 μg of TPCK-treated modified porcine trypsin (Promega) in the same buffer. The supernatant containing the resulting tryptic peptides was combined with that of the subsequent two extractions from the gel by 50% acetonitrile/5% formic acid and subjected to MALDI-MS analysis using the quadrupole time-of-flight (Q-TOF) mass spectrometer (Micromass Q-Tof Ultima, Manchester, U.K.) in the Proteomics Research Core Laboratory at National Cheng-Kung University, Taiwan.

RESULTS

Identification of Two Gene Families Encoding 11S Globulin and 2S Albumin from EST Eequences of Maturing Sesame Seeds. Four groups of EST sequences putatively encoding 11S globulin isoforms were identified in a database search of 3328 EST sequences from maturing sesame seeds (**Table 1**). These four groups were composed of 44, 9, 19, and 28 EST sequences, respectively. The four putatively encoded 11S globulin isoforms, including the two sequences reported previously (*6*, *7*), were subsequently named 11S-1, 11S-2, 11S-3, and 11S-4, according to the time they were documented in the literature. In a similar database search, three groups of EST sequences putatively encoding 2S albumin isoforms were

Figure 2. Sequence alignment of the precursor polypeptides of three sesame 2S albumin isoforms. The amino acid number for the last residue in each line is listed on the right for each isoform. Broken lines in the sequences represent gaps introduced for best alignment. The cleavage site of the putative N-terminal signal sequence and that for splitting these polypeptides into small and large subunits are indicated by scissors symbols, individually. Eight conserved cysteine residues are boxed and predicted to form two intrachain disulfide bonds within the large subunit and two interchain disulfide bridges between the small and large subunits.

Table 3. Amino Acid Composition of Sesame 11S Globulin and 2S Albumin Isoforms (Calculated from the Predicted Mature Processed Proteins)

	isoform						
amino acid	11S-1	11S-2	11S-3	11S-4	2S-1	2S-2	2S-3
Ala	6.62	7.16	6.84	8.91	3.64	2.61	6.19
Cys	1.14	0.84	0.85	1.34	9.09	6.96	8.85
Asp	4.11	5.26	2.78	3.79	0.91	3.48	0.88
Glu	7.53	7.16	9.19	7.35	11.82	7.83	11.50
Phe	3.65	4.21	3.63	4.45	3.64	2.61	2.65
Gly	7.53	7.79	9.62	9.13	4.55	6.96	3.54
His	2.51	1.68	2.14	2.45	2.73	0.00	1.77
Ile	4.34	4.00	4.70	5.35	0.91	1.74	0.88
Lys	1.83	2.53	1.92	1.34	0.91	0.00	1.77
Leu	6.62	9.05	7.69	6.24	3.64	5.22	4.42
Met	3.20	1.68	0.85	1.56	13.64	1.74	10.62
Asn	4.79	4.84	5.56	5.79	1.82	2.61	1.77
Pro	4.34	4.63	3.63	4.01	2.73	2.61	2.65
Gln	8.68	7.16	8. 97	8.46	14.55	24.35	12.39
Arg	9.82	10.32	10. 47	10.02	12.73	14.78	13.27
Ser	7.53	5.47	5.98	6.68	5.45	6.96	5.31
Thr	5.25	4.84	5.56	2.67	0.91	0.87	0.88
Val	6.85	6.53	5.77	6.46	0.91	5.22	4.42
Trp	0.91	1.68	0.85	1.11	1.82	0.00	2.65
Tyr	2.74	3.16	2.99	2.90	3.64	3.48	3.54

identified (**Table 2**). These three groups were composed of 25, 46, and 13 EST sequences, respectively. The three putatively encoded 2S albumin isoforms, including the two sequences reported previously (*6*, *7*), were subsequently named 2S-1, 2S-2, and 2S-3 according to the time they were documented in the literature.

Completing Full-Length cDNA Sequences Encoding 11S-3, 11S-4, and 2S-3. Full-length cDNA sequences encoding 11S-3, 11S-4, and 2S-3 were obtained by PCR cloning using primers designed according to their available EST sequences. The cDNA sequence encoding 11S-3 (GenBank accession no. DQ256293) comprises 1687 nucleotides, consisting of a 14-nucleotide 5′ untranslated region, an open reading frame of 1476 nucleotides, and a 197-nucleotide 3′ untranslated region. The cDNA sequence encoding 11S-4 (GenBank accession no. DQ256294) comprises 1606 nucleotides, consisting of an 18-nucleotide 5′ untranslated region, an open reading frame of 1410 nucleotides, and a 178-nucleotide 3′ untranslated region. The cDNA sequence encoding 2S-3 (GenBank accession no. DQ256292) comprises 635 nucleotides, consisting of a 21-nucleotide 5′ untranslated region,

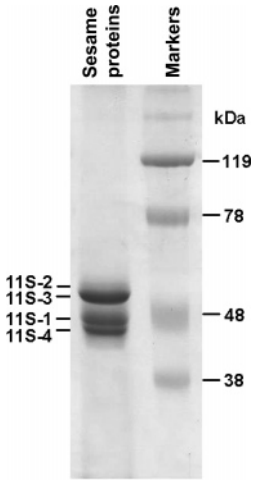

Figure 3. Sesame 11S globulin isoforms resolved in SDS-PAGE. The pellet fraction of crude sesame extract was resolved in a 12.5% polyacrylamide gel under 200 V for 120 min. Four protein bands corresponding to 11S globulin isoforms were confirmed by MALDI-MS analysis (**Table 4**).

an open reading frame of 453 nucleotides, and a 161-nucleotide 3′ untranslated region.

Sequence Analyses of the Deduced 11S Globulin and 2S Albumin Isoforms. Sequence alignment of the four deduced 11S globulin isoforms suggests that all precursor polypeptides contain an N-terminal signal sequence responsible for ER targeting via a signal recognition particle (SRP) dependent pathway, a conservative processing site for splitting each polypeptide into two segments (acidic and basic subunits) via proteolytic cleavage, and four conservative cysteine residues responsible for forming an intrachain disulfide bond within the acidic subunit and an interchain disulfide bond linking both subunits in their mature proteins (**Figure 1**). Similarly, sequence alignment of the three deduced 2S albumin isoforms suggests that all precursor polypeptides contain an N-terminal signal sequence for ER targeting, a conservative processing site for splitting into two segments (small and large subunits), and eight conservative cysteine residues responsible for forming two intrachain disulfide bonds within the large subunit and two interchain disulfide bonds linking both subunits in their mature proteins (**Figure 2**). Amino acid compositions of these deduced 11S globulin and 2S albumin isoforms (**Table 3**) indicate that the nutritional importance of sesame proteins, that is, richness

Table 4. Fragments of 11S Globulin Isoforms Identified by MALDI-MS Analyses

isoform	residue	sequence
11S-1	48–61	IQSEGGTTELWDER
	62–74	QEQFQCAGIVAMR
	75–92	STIRPNGLSLPNYHPSPR
	99–118	GQGLISIMVPGCAETYQVHR
	192–201	AFYLAGGVPR
	211–218	QTFHNIFR
	219–238	AFDAELLSEAFNVPQETIRR
	239–253	MQSEEEERGLIVMAR
	256–271	MTFVRPDEEEGEQEHR
	329–353	GNLYSNALVSPDWSMTGHTIVYVTR
	354–372	GDAQVQVVDHNGQALMNDR
	373–390	VNQGEMFVVPQYYTSTAR
	391–402	AGNNGFEWVAFK
	410–420	SPLAGYTSVIR
	421–442	AMPLQVITNSYQISPNQAQALK
	446–457	GSQSFLLSPGGR
11S-2	63–75	FESEAGLTEFWDR
	76–89	NNQQFECAGVAAVR
	116–136	GIQGTVIPGCAETFERDTQPR
	198–209	HFFLAGNPQGGR
	267–279	LDIVLPGEEEEER
	387–401	EGQLIIVPQNYVVAK
	402–415	RASQDEGLEWISFK
	434–448	AMPEEVVMTAYQVSR
11S-3	188–199	DVANEANQLDLK
	202–220	KFFLAGNPQTAQFQGQQER
	283–295	LVLPEYGREEQQR
	318–334	IRENIEHTAATHSYNPR
	338–353	ISTINSQTLPILSQLR
	364–383	NGITAPHWSTNSHSALYVTR
	421–432	AGEQGFEYVTFR
	451–472	AMPDEVVMNAFGVSREDARNLK
	473–486	YNRDEATVFSPGGR
11S-4	78–97	ALMLPAYHNAPILAYVQQGR
	98–122	GMYGVMISGCPETFESSQQQFEEGR
	182–192	SFFLAGNPAGR
	193–205	GQEQQEYAPQLGR
	208–217	GQHQFGNVFR
	237–253	SLQGENDERGHIITVAR
	254–264	GLQVISPPLQR
	265–278	EEYGRQEEEPYYGR
	377–393	VREGQVVVVPQNFAVVK
	425–439	GLPADVIANAYQISR
	451–460	ETMMFSGSFR

in methionine, is attributed mainly to two 2S albumin isoforms (2S-1 and 2S-3) and partly to one 11S globulin isoform (11S-1). Lysine content is low in all 11S globulin and 2S albumin isoforms.

Separation and Confirmation of 11S Globulin and 2S Albumin Isoforms in Sesame Extract. To confirm the presence of the deduced 11S globulin and 2S albumin isoforms in sesame, crude proteins extracted from sesame seeds were subjected to SDS-PAGE analyses. Four candidate protein bands of 11S globulin isoforms were resolved in a 12.5% polyacrylamide gel (**Figure 3**). MALDI-MS analysis revealed that these four bands corresponded to the four deduced 11S globulin isoforms (**Table 4**). The abundance of these four isoforms was in accord with the occurrence frequency of their EST sequences found in the database search (**Table 1**), with 11S-2 as the least abundant among the four 11S globulin isoforms. In contrast, two candidate protein bands of 2S albumin isoforms were resolved in an 18% polyacrylamide gel (**Figure 4**). MALDI-MS analysis revealed that the upper band corresponded to 2S-2 and that the lower band contained the two methionine-rich isoforms, that is, 2S-1 and 2S-3, which were highly homologous (70% identity) and could not be resolved in our electrophoresis condition (**Table 5**). The abundance of 2S albumin isoforms was in accord with the occurrence frequency of their EST sequences found in the database search (**Table 2**). Similar profile and abundance of 11S globulin and 2S albumin isoforms resolved in SDS-PAGE were observed for crude proteins extracted from different varieties of black, brown, and white sesame seeds (data not shown).

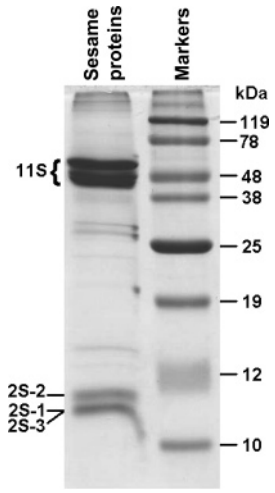

Figure 4. Sesame 2S albumin isoforms resolved in SDS-PAGE. The pellet fraction of crude sesame extract was resolved in an 18% polyacrylamide gel under 200 V for 70 min. Two protein bands corresponding to 2S albumin isoforms were confirmed by MALDI-MS analysis (**Table 5**).

Table 5. Fragments of 2S Albumin Isoforms Identified by MALDI-MS Analyses

isoform	residue	sequence
2S-1 and 2S-3	80–86	FEEEHLR
	111–117	MMQEYGR
	118–127	EQEMQQMMQK
	128–133	AEYLPR
	134–144	MCGMSYPTQCR
2S-2	113–131	QQQQEGGYQEGQSQQVYQR

DISCUSSION

In this study, four and three EST groups separately derived from two gene families encoding 11S globulin and 2S albumin isoforms were identified from a database search, and the presence of these seven gene products was confirmed by MALDI-MS analyses. All of the protein bands of 11S globulin and 2S albumin isoforms resolved in SDS-PAGE were found corresponding to different genes. The results indicated that all of the 11S globulin and 2S albumin isoforms present in sesame seed were encoded by their corresponding genes individually rather than derived from post-translational processing or modification of one or a few gene products. Of course, possible processing or modification that did not cause migration shift of these proteins in SDS-PAGE under our electrophoresis conditions could not be ruled out in our analyses. In addition, complete identification of each protein band of 11S globulin and 2S albumin isoforms resolved in SDS-PAGE provided detailed information of their calculated amino acid compositions and abundance in total sesame proteins, thus revealing all of the major components responsible for the nutritional importance of methionine richness in sesame seed.

Sesame products in the form of whole or crushed seeds, oil, and meal have been recognized as nutritious foods in numerous

countries, as the seeds contain excellent quality of oil (40−60%) and protein (15−25%). However, the incidence of sesame allergy has been increasingly reported in the past decade (*12*). Interestingly, sesame allergy has been mostly reported in subjects from Europe, Israel, the United States, and Australia, but not in subjects from countries that produce sesame seeds. So far, major sesame allergens identified include the storage proteins (11S globulin, 7S globulin, and 2S albumin) accumulated in protein bodies and the structural proteins (two oleosin isoforms) found on the surface of oil bodies (*24*). The contents of 11S globulin, 7S globulin, 2S albumin, and oleosin isoforms represent 60−70, 1−2, 15−25, and 1−2% of total sesame proteins, respectively (*25*). Among these allergic proteins, 2S albumin isoforms, including 2S-1 and 2S-2, seem to be most frequently identified in the cases of sesame allergen (*13*−*15*). Because 2S-1 and 2S-3 are highly homologous with 70% identity, it will not be surprising if 2S-3 reported in this study is identified as an allergen in the future.

The results obtained in this study will benefit the diagnosis of allergens in sesame food products as suggested in the following ways. Along with total sesame seed proteins as a standard, candidate allergen proteins can be resolved in SDS-PAGE and roughly identified under the electrophoresis conditions described in **Figures 3** and **4**. Furthermore, candidate allergen proteins either freshly extracted from sesame food products or excised from SDS-PAGE gels as described above can be subjected to MALDI-MS analyses (commercially available now) for protein identification by comparison with the published sequences of sesame proteins.

Completion of the two gene families encoding 11S globulin and 2S albumin isoforms in the current study was assisted by the available database of 3328 EST sequences from maturing sesame seeds (*19*). However, no EST sequences derived from cDNA fragments encoding 7S globulin isoforms were found in these sesame EST sequences, presumably due to their low expression levels during seed maturation. Although two putative 7S polypeptides were found in sesame proteins with one of the corresponding cDNA fragments cloned (*7*), the number of 7S globulin genes present in sesame genome remains uncertain. Apparently, a random sequencing of 3000 EST clones is insufficient for functional proteomics analysis of most minor proteins possibly translated by a lower level of corresponding mRNA in maturing sesame seeds. For a comprehensive understanding of sesame seed, many more EST sequences or a complete genomic sequencing is indispensable.

ABBREVIATIONS USED

EST, expressed sequence tag; MALDI-MS, matrix-assisted laser desorption/ionization mass spectrometry; SRP, signal recognition particle.

ACKNOWLEDGMENT

We thank the Proteomics Research Core Laboratory at National Cheng-Kung University for assistance in mass spectrometry analysis and Dr. Tien-Joung Yiu of the Crop Improvement Department, Tainan District Agricultural Improvement Station for supplying sesame seeds.

LITERATURE CITED

(1) Namiki, M. The chemistry and physiological functions of sesame. *Food Rev. Int.* **1995**, *11*, 281−329.

(2) Hasegawa, K.; Murata, M.; Fujino, S. Characterization of subunits and temperature-dependent dissociation of 13S globulin of sesame seed. *Agric. Biol. Chem.* **1978**, *42*, 2291−2297.

(3) Nishimura, N.; Okubo, K.; Shibasaki, K. Chemical and physical properties of 13S globulin, the major protein in sesame seeds. *Cereal Chem.* **1979**, *56*, 239−242.

(4) Prakash, V.; Narasinga Rao, M. S. Hydrogen ion equilibria of α-globulin from *Sesamum indicum* L. *J. Agric. Food Chem.* **1990**, *38*, 612−615.

(5) Plietz, P.; Damaschun, G.; Zirwer, D.; Gast, K.; Schwenke, K. D.; Prakash, V. Shape and quaternary structure of alpha-globulin from sesame (*Sesamum indicum* L.) seed as revealed by small angle X-ray scattering and quasi-elastic light scattering. *J. Biol. Chem.* **1986**, *261*, 12686−12691.

(6) Tai, S. S. K.; Wu, L. S. H.; Chen, E. C. F.; Tzen, J. T. C. Molecular cloning of 11S globulin and 2S albumin, the two major seed storage proteins in sesame. *J. Agric. Food Chem.* **1999**, *47*, 4932−4938.

(7) Tai, S. S. K.; Lee, T. T. T.; Tsai, C. C. Y.; Yiu, T. J.; Tzen, J. T. C. Expression pattern and deposition of three storage proteins, 11S globulin, 2S albumin and 7S globulin in maturing sesame seeds. *Plant Physiol. Biochem.* **2001**, *39*, 981−992.

(8) Kimura, S. Positive ratio of allergen specific ige antibodies in serum, from a large scale study. *Rinsho Byori* **2001**, *49*, 376−380.

(9) Dalal, I.; Binson, I.; Levine, A.; Somekh, E.; Ballin, A.; Reifen, R. The pattern of sesame sensitivity among infants and children. *Pediatr. Allergy Immunol.* **2003**, *14*, 312−316.

(10) Kusunoki, T.; Miyanomae, T.; Inoue, Y.; Itoh, M.; Yoshioka, T.; Okafuji, I.; Nishikomori, R.; Heike, T.; Nakahata, T. Changes in food allergen sensitization rates of Japanese allergic children during the last 15 years. *Arerugi* **2004**, *53*, 683−688.

(11) Gangur, V.; Kelly, C.; Navuluri, L. Sesame allergy: a growing food allergy of global proportions? *Ann. Allergy Asthma Immunol.* **2005**, *95*, 4−11.

(12) Sporik, R.; Hill, D. Allergy to peanut; nuts; and sesame seed in Australian children. *Biol. Mol. J.* **1996**, *313*, 1477−1478.

(13) Beyer, K.; Bardina, L.; Grishina, G.; Sampson, H. A. Identification of sesame seed allergens by 2-dimensional proteomics and Edman sequencing: seed storage proteins as common food allergens. *J. Allergy Clin. Immunol.* **2002**, *110*, 154−159.

(14) Wolff, N.; Cogan, U.; Admon, A.; Dalal, I.; Katz, Y.; Hodos, N.; Karin, N.; Yannai, S. Allergy to sesame in humans is associated primarily with IgE antibody to a 14 kDa 2S albumin precursor. *Food Chem. Toxicol.* **2003**, *41*, 1165−1174.

(15) Pastorello, E. A.; Varin, E.; Farioli, L.; Pravettoni, V.; Ortolani, C.; Trambaioli, C.; Fortunato, D.; Giuffrida, M. G.; Rivolta, F.; Robino, A.; Calamari, A. M.; Lacava, L.; Conti, A. The major allergen of sesame seeds (*Sesamum indicum*) is a 2S albumin. *J. Chromatogr. B: Biomed. Sci. Appl.* **2001**, *756*, 85−93.

(16) Derby, C. J.; Gowland, M. H.; Hourihane, J. O. Sesame allergy in Britain: a questionnaire survey of members of the anaphylaxis campaign. *Pediatr. Allergy Immunol.* **2005**, *16*, 171−175.

(17) Hernandez, P.; Muller, M.; Appel, R. D. Automated protein identification by tandem mass spectrometry: issues and strategies. *Mass Spectrom. Rev.* **2006**, *25*, 235−254.

(18) Johnson, R. S.; Davis, M. T.; Taylor, J. A.; Patterson, S. D. Informatics for protein identification by mass spectrometry. *Methods* **2005**, *35*, 223−236.

(19) Suh, M. C.; Kim, M. J.; Hur, C. G.; Bae, J. M.; Park, Y. I.; Chung, C. H.; Kang, C. W.; Ohlrogge, J. B. Comparative analysis of expressed sequence tags from *Sesamum indicum* and *Arabidopsis thaliana* developing seeds. *Plant Mol. Biol.* **2003**, 52, 1107−1123.

(20) Altschul, S. F.; Warren, G.; Webb, M.; Eugene, W. M.; David, J. L. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403−410.

(21) Shewry, P. R. Plant storage protein. *Biol. Rev.* **1995**, *70*, 375−426.

(22) Wilkins, T. A.; Smart, L. B. Isolation of RNA from plant tissue. In *A Laboratory Guide to RNA*; Krieg, P. A., Ed.; Wiley-Liss: New York, 1996; Vol. 2, pp 21−41.

(23) Peng, C. C.; Tzen, J. C. T. Analysis of the three essential constituents of oil bodies in developing sesame seeds. *Plant Cell Physiol.* **1998**, *39*, 35−42.

(24) Leduc, V.; Moneret-Vautrin, D. A.; Tzen, J. T. C.; Morisset, M.; Guerin, L.; Kanny, G. Identification of oleosins as major allergens in sesame seed allergic patients. *Allergy* **2006**, *61*, 349−356.

(25) Tzen, J. T. C.; Wang, M. M. C.; Tai, S. S. K.; Lee, T. T. T.; Peng, C. C. The abundant proteins in sesame seed: storage proteins in protein bodies and oleosins in oil bodies. *Adv. Plant Physiol.* **2003**, *6*, 93−105.

Received for review May 29, 2006. Revised manuscript received September 25, 2006. Accepted October 6, 2006. The work was supported by a grant from the National Science Council, Taiwan, ROC (NSC 94-2752-B-005-008-PAE to J.T.C.T.).

JF061505X