Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 11774–11778, December 1992
Plant Biology

Molecular cloning of a gene encoding a seed-specific protein with nutritionally balanced amino acid composition from *Amaranthus*

ANJANA RAINA AND ASIS DATTA*

Molecular Biology Laboratory, School of Life Sciences, Jawaharlal Nehru University, New Delhi 110 067, India

Communicated by M. S. Swaminathan, July 13, 1992

ABSTRACT **An albumin with a well-balanced amino acid composition and high levels of the essential amino acids was purified to homogeneity from the mature seeds of *Amaranthus hypochondriacus*. The amino acid composition of this protein is comparable to the World Health Organization recommended values for a highly nutritional protein. The protein is a 35-kDa monomer with four isoforms that can be separated by chromatofocusing. Antibodies raised against one of the isoforms, AmA1, cross-reacted with the other three isoforms. Affinity-purified AmA1 antibodies were used to isolate cDNA clones from a developing-seed expression library. The six immunopositive recombinants obtained were found to be related. The cDNA of the largest clone (1.2 kilobases) has a single major open reading frame corresponding to a 304-amino acid polypeptide. The clone was confirmed by hybrid-selected translation and immunoprecipitation. The size of the immunoprecipitated product was identical to the mature protein. Analysis of RNA and protein in developing seeds showed that AmA1 is synthesized during early embryogenesis, reaching a maximum by midmaturation. No RNA was detected in 1-day-old seedlings although the protein showed delayed breakdown on germination. Expression of the AmA1 gene was found to be seed-specific, as no protein or RNA was detected in other plant tissues.**

Seed storage proteins, intended as a source of nitrogen for germinating seedlings, form an important source of dietary protein for human beings. Humans require a diet with a balanced amino acid composition, but often seeds are deficient in at least one of the essential amino acids. For years plant breeders have tried to improve the balance of essential amino acids in important crop plants (1). Molecular approaches for improving the nutritional quality of seed proteins provide alternatives to the conventional approaches. *In vitro* mutagenesis of the coding region of certain seed proteins has been attempted to increase the levels of essential amino acids (1, 2). Another approach is to transfer heterologous storage protein genes that encode proteins with higher levels of limiting amino acids (3). Expressing high levels of a particular amino acid by heterologous gene transfer or by mutation may be detrimental to the normal physiology of seed development. This may also produce seeds with a biased amino acid composition. An alternative approach is, therefore, to express a gene for a heterologous protein with a balanced amino acid composition.

Grain amaranth is a pseudo cereal with a high protein content (17–19% of seed dry weight) as compared to more traditional crops that have an average of ≈10% protein (4). Its protein is rich in essential amino acids such as lysine, tryptophan, and sulfur-containing amino acids (5). It can therefore be used as the source of a gene that encodes a protein of high nutritional value. In spite of good quality and quantity, these proteins have not yet been purified and characterized. In *Amaranthus*, 50% of the total seed proteins at maturity are globulin and albumin (6). As a first step toward developing transgenic plants with balanced protein composition, we report the cloning of a full-length cDNA from *Amaranthus* that encodes a 35-kDa protein rich in essential amino acids.† Amino acid composition of this protein corresponds closely to that of the World Health Organization recommended protein standard for optimum human nutrition. It is developmentally regulated and is expressed only in seeds. Furthermore, nucleotide sequence analysis does not show any apparent homology with other known seed-specific proteins.

MATERIALS AND METHODS

Plant Material. Seeds of *Amaranthus hypochondriacus* were obtained from National Bureau of Plant Genetic Research, Simla, India. Mature seeds were ground to a fine powder and defatted by extraction with ice-cold acetone. Seeds and the defatted seed meal were stored at 4°C under desiccation. Seeds at various developmental stages were kept frozen until use.

Purification and Characterization of the 35-kDa Albumin. Defatted seed meal (1 g) was extracted with 10 ml of buffer A (25 mM Tris acetate, pH 8.5/1 M NaCl/2 mM phenylmethylsulfonyl fluoride). The homogenate was centrifuged and the supernatant was dialyzed against buffer A. The precipitate formed on dialysis was removed by centrifugation and the supernatant was chromatographed on a DEAE-Sepharose column (1 × 25 cm; 25 ml/h), preequilibrated with buffer A. The column was washed with buffer A and the bound protein was eluted with the multicomponent buffer system of Prestidge and Hearn (7) with slight modifications. Fractions (2 ml) were collected and analyzed by SDS/PAGE. Appropriate fractions, after analysis on SDS/PAGE, were concentrated with a Centricon (Amicon) and subjected to further purification on a Sephadex G-75 (1.6 × 30 cm) column equilibrated with buffer A/0.1 M NaCl. Elution was carried out in buffer A/0.1 M NaCl at a flow rate of 10 ml/h. Fractions were analyzed on SDS/PAGE, and pure protein was concentrated and stored at 4°C.

SDS/PAGE and Nonequilibrium pH-Gradient Gel Electrophoresis (NEPHGE). Purity of the sample was routinely checked by SDS/PAGE on 12% separating and 4% stacking gels as described by Laemmli (8). The gels were stained with Coomassie brilliant blue R-250. NEPHGE was performed by the procedure of O'Farrell (9). Purified protein or a crude albumin fraction was separated by NEPHGE in the first dimension by using Pharmalyte, pH 3–10 (Pharmacia). The gel was equilibrated in SDS sample buffer prior to separation in the second dimension by SDS/PAGE.

Antisera Preparation and Immunodetection. Antibodies against purified AmA1 were raised in rabbits. They were

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: NEPHGE, nonequilibrium pH gradient gel electrophoresis.

*To whom reprint requests should be addressed.

†The sequence reported in this paper has been deposited in the GenBank data base (accession no. Z11577).

affinity-purified essentially by the procedure of Elledge and Davis (10). Western blot analysis was done as described by Towbin *et al.* (11). Protein bands were visualized by staining with ponceau S (12) prior to immunodetection. Antibodies were used at a dilution of 1:10,000 (antiserum) or 1:5000 (affinity-purified antibodies). Alkaline phosphatase-conjugated anti-rabbit IgG (Promega) was used as the secondary antibody.

Amino Acid Analysis. The amino acid content of the purified protein was determined by using a LKB α Plus 4151 amino acid analyzer. Samples were hydrolyzed with 6 M HCl under vacuum at 105°C for 22 h prior to analysis.

RNA Isolation. Total RNA was isolated by the phenol/chloroform extraction and LiCl precipitation procedure of Ausubel *et al.* (13). Poly(A)$^+$ RNA was isolated from total RNA by two rounds of selection on an oligo(dT)-cellulose column by the procedure of Okayama *et al.* (14).

***In Vitro* Translation and Immunoprecipitation.** RNA (2 μg) or poly(A)$^+$ RNA (0.5 μg) was translated in a rabbit reticulocyte lysate system (Promega) and immunoprecipitated by the procedure of Anderson and Blobel (15) except that antibodies were bound to protein A-Sepharose beads in 0.1 M sodium phosphate (pH 8.0) prior to immunoprecipitation. The immunoprecipitated product was boiled in SDS sample buffer and electrophoresed. The gel was stained, fluorography was done (16), and the gel was exposed to x-ray film at −70°C. Hybrid-selected translation of AmA1 mRNA was done by the procedure of Ricciardi *et al.* (17).

cDNA Construction and Screening. Poly(A$^+$) RNA isolated at the stages when AmA1 mRNA was high (stages III and IV, Table 1) was used as a template for cDNA synthesis. cDNA was synthesized by the method of Gubler and Hoffman (18) and cloned in λgt11 according to Young and Davis (19). Briefly, first-strand cDNA synthesis was carried out using reverse transcriptase and oligo(dT) as the primer (Amersham cDNA synthesis kit). cDNA was sequentially treated with S1 nuclease, *Eco*RI methylase, and the Klenow fragment of DNA polymerase I prior to blunt-end ligation to *Eco*RI linkers. Linkered cDNA was ligated to λgt11 containing *Eco*RI arms (Amersham cDNA cloning kit). The resulting DNA was packaged *in vitro* and used to infect *Escherichia coli* Y1090. About 17,000 recombinant phages were plated on two 150-mm plates and induced with isopropyl β-D-thiogalactoside, and fusion proteins were detected by immunoscreening with AmA1 antibodies.

Subcloning of Insert DNA. λ DNA was purified by the hexadecyltrimethylammonium bromide (CTAB) procedure of Manfioletti and Schneider (20). The insert was purified from the *Eco*RI-digested recombinant λ DNA and ligated to *Eco*RI-cut pTZ18U and M13mp18 vector DNA. *E. coli* JM101 and JM109 served as hosts for plasmid and phage vectors, respectively. Plasmid and M13 DNA was purified by standard protocols (21).

cDNA Sequence Analysis. Sequencing was done in M13mp18 DNA in both orientations by the dideoxynucleotide chain-termination method using Sequenase Version 2.0 (United States Biochemical). To read the complete cDNA sequence, deletions were generated in both orientations by exonuclease III and S1 nuclease (21). Orientation of the insert in the λgt11 clone was directly determined by using λgt11 primers.

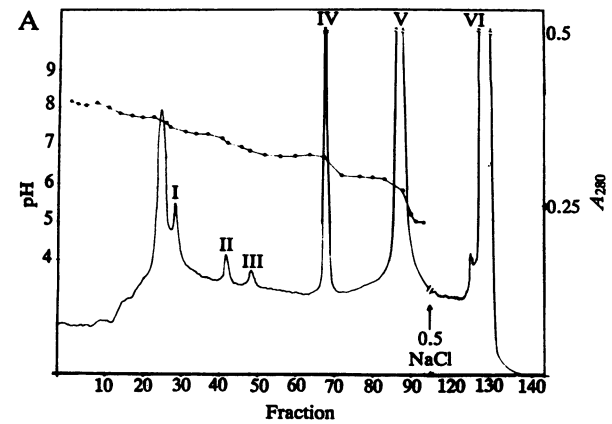
Table 1. Average weight of seeds at various stages of development

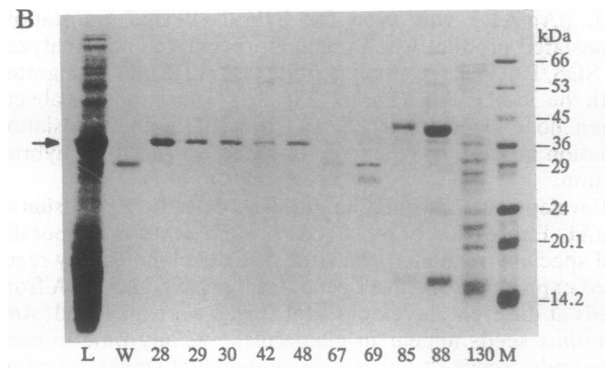
Stage	Seed weight, mg
I	0.1
II	0.2
III	0.3
IV	0.4
V (mature)	0.8

Northern Blot Analysis. Total cellular RNA (10 μg) was denatured with glyoxal and separated by electrophoresis on a 2% agarose gel containing glyoxal. The amount of RNA and the integrity of rRNA was confirmed by ethidium bromide staining of a duplicate gel. Gel was blotted onto GeneScreen-*Plus* (DuPont) and probed with AmA1 cDNA (specific activity, 3 × 10^8 cpm/μg of DNA) (22).

RESULTS

Purification of 35-kDa Protein and Immunoreactivity. The albumin fraction obtained on dialysis of the crude extract was





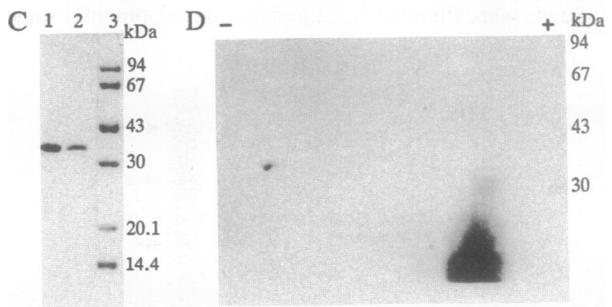


FIG. 1. Purification of the 35-kDa protein. (*A*) Elution profile of the chromatofocusing column: ●, A_{280}; ——, pH; I–VI, protein peaks. The unnumbered peak before peak I did not contain protein. (*B*) SDS/PAGE analysis of column fractions. Lanes: L, material loaded; W, wash; M, molecular mass standards. Numbers indicate the fractions analyzed. Arrow indicates the 35-kDa protein. Equal volumes of each fraction were analyzed. (*C*) Purity of AmA1 by one-dimensional gel electrophoresis. The purified protein [10 μg (lane 1) and 5 μg (lane 2)] was analyzed by SDS/PAGE on a 12% gel and stained with Coomassie blue. Lane 3 shows molecular mass standards. (*D*) Purity of AmA1 by two-dimensional gel electrophoresis. Protein (5 μg) was analyzed in the first dimension by NEPHGE and in the second dimension by SDS/PAGE on a 10% gel and stained with Coomassie blue. Dark patch at the bottom of the gel contains the ampholytes.

chromatographed on a chromatofocusing column. A 35-kDa albumin protein was eluted in four peaks at pH 7.4, 7.1, 6.8, and 6.7 (a faint band but clearly visible on immunoblot analysis; Fig. 1 *A* and *B*); the other two peaks did not have any 35-kDa albumin protein. Protein eluting at pH 7.4, AmA1 (peak I), was further purified on a gel-filtration column (data not shown). Fig. 1 *C* and *D* shows the purity of the protein by one- and two-dimensional gel electrophoresis. Antibodies raised against purified peak I protein, showed immunoreactivity with the 35-kDa polypeptide present in other peaks, indicating that the protein may have at least four isoforms (Fig. 2). When albumin fraction was analyzed on an isoelectric focusing gel, three distinct bands and one faint band were visible on immunostaining (Fig. 2, lane C), to further confirm the existence of four isoforms.

Molecular Cloning of AmA1 cDNA. AmA1 cDNA clones were isolated from a cDNA library constructed in a λgt11 expression vector. The level of AmA1 mRNA in developing seeds was analyzed by *in vitro* translation and immunoprecipitation. Poly(A)$^+$ RNA, from stages with a high level of AmA1 mRNA, was used as the template for cDNA synthesis. A total of 35,000 plaques were obtained, of which 6 were immunopositive. Positive plaques were selected and subjected to successive rounds of phage titering and screening to obtain clonally pure recombinant plaques. DNA was isolated and Southern blot hybridization was done to determine the relatedness of all the clones. Three clones (AmA1.2, AmA1.3, and AmA1.5) were subcloned in plasmid vector pTZ18U. pAmA1.2 and pAmA1.3 had large inserts of 1.2 kilobases and pAmA1.5 had an insert of 0.25 kilobase (Fig. 3*A*). pAmA1.3 was used for hybrid-selected translation. Translated product when immunoprecipitated and analyzed by SDS/PAGE gave a polypeptide of 35 kDa that comigrated with the 35-kDa AmA1 band (Fig. 3*B*). This band was absent when no exogenous RNA was present in the translation reaction mixture or when vector alone was used for hybridization.

Developmental Regulation and Seed-Specific Expression of AmA1. Expression of most seed protein genes is temporally and spacially regulated. To study the developmentally regulated expression of AmA1 gene, total protein and RNA from seeds at different developmental stages were analyzed. *Amaranthus* seeds appear in glomerules. At any time in each glomerule, seeds at various developmental stages are present. Seeds were therefore grouped into developmental stages

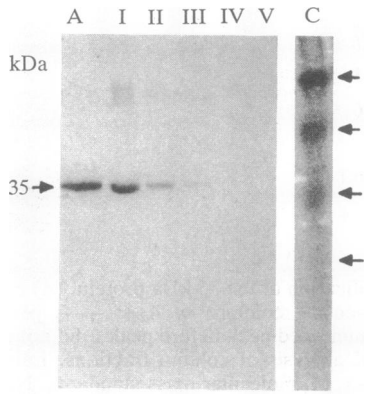


FIG. 2. Western blot analysis of peak fractions from a chromatofocusing column probed with AmA1 antibodies. Equal volumes of sample from fractions obtained after chromatofocusing were resolved by SDS/PAGE on a 12% gel and transferred to nitrocellulose. Immunoreactive polypeptides were detected with AmA1 antibodies. Lanes: A, albumin fraction; I–V, peak fractions (Fig. 1*A*) analyzed by SDS/PAGE; C, albumin fraction analyzed on an isoelectric focusing gel prior to immunodetection. Arrows indicate the various isoelectric forms. Fourth band is very faint and thus not very clear.

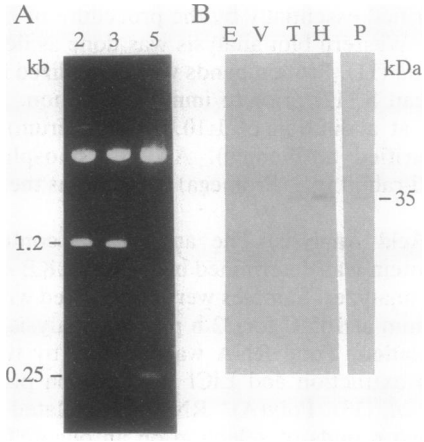


FIG. 3. Analysis of subclones. (*A*) Inserts from λ clones 2, 3, and 5 were subcloned in pTZ18U vector. All three clones were digested with *Eco*RI and analyzed on a 2% agarose gel. Clone numbers are shown above the lanes. (*B*) Hybrid-selected translation product of subclone 3 was immunoprecipitated with AmA1 antibodies bound to protein A-Sepharose and analyzed by SDS/PAGE on a 12% gel. The ^{35}S-labeled and immunoprecipitated 35-kDa protein was detected by autoradiography.

by their weight (Table 1). By SDS/PAGE and Western blot analyses, AmA1 was found to be synthesized in very early embryogenesis (Fig. 4 *A* and *B*). Total RNA when subjected to Northern blot analysis also showed the presence of AmA1 mRNA during early embryogenesis (Fig. 4 *C* and *D*). Mature seeds showed low level of AmA1 RNA and no RNA was detected in 1-day-old seedlings. The protein increased in proportion to the RNA until stage IV (Table 1 and Fig. 4) after which there was no further increase in protein inspite of the presence of AmA1 mRNA. AmA1 was not detected in other plant tissues (data not shown). Northern blot analysis did not show any AmA1 mRNA in leaves or roots (Fig. 5). Therefore, the expression of AmA1 is seed-specific.

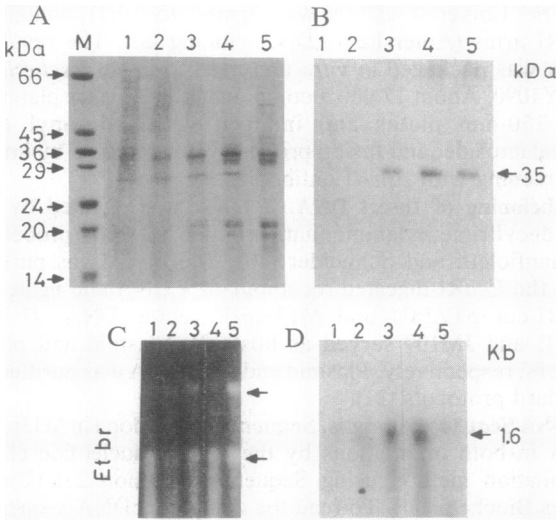


FIG. 4. Expression of AmA1 gene during seed development. (*A* and *B*) Crude extract (20 μg of protein) of seeds at various stages of development, as in Table 1, were resolved by SDS/PAGE on a 12% gel in duplicate. Protein bands in one gel were stained with Coomassie blue (*A*) and in the other were subjected to Western blot analysis (*B*). (*C* and *D*) Total RNA was extracted from developing seeds and 2-μg samples were separated on a 1.2% agarose gel containing glyoxal in duplicate. RNA integrity was checked by ethidium bromide staining (*C*). AmA1-specific mRNA was identified by Northern blot analysis. RNA was transferred to GeneScreen*Plus* and probed with labeled AmA1 cDNA (*D*).

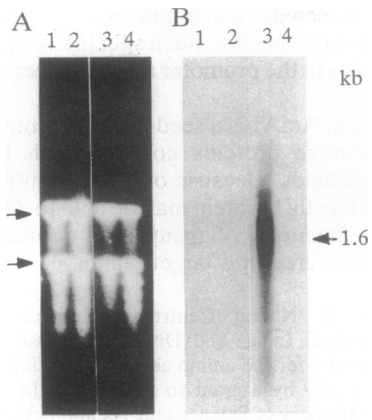


FIG. 5. Seed-specific expression of AmA1 gene. Total RNA was isolated from the following tissues. (*A*) Lanes: 1, seedling; 2, seed; 3, root; 4, leaf. (*B*) Lanes: 1, leaf; 2, root; 3, seed; 4, 1-day-old seedling. (*A*) Ethidium bromide-stained gel to show the integrity of rRNA bands. Arrows indicate the rRNA bands. (*B*) Northern blot probed with labeled AmA1 cDNA.

Sequence Analysis of AmA1 cDNA. The largest insert of 1.2 kilobases (from pAmA1.3) was subcloned in M13mp18 in both orientations to obtain single-stranded DNA. The sequencing strategy is summarized in Fig. 6. The sequenced cDNA (Fig. 7*A*) is 1183 base pairs (bp) long with an open reading frame of 912 bp and noncoding 5′ and 3′ flanking sequences. The open reading frame encodes a 35-kDa protein of 304 amino acids with a pI value of 6.8. Amino acid sequence analysis indicates a hydrophilic protein with a small stretch of hydrophobic amino acids at the N terminus (Fig. 7*B*). The derived amino acid composition shows high levels of the essential amino acids that is similar to the composition obtained by protein hydrolysis (Table 2). The differences observed are due to technical limitations with Picotag system (Waters) of amino acid analysis as it normally shows reduced levels of sulfur amino acids and high level of glycine. A poly(A) stretch is missing from the sequence although two putative polyadenylylation signals are located 50 and 194 bp downstream from the stop codon (Fig. 7*A*). No homology was seen between the deduced protein sequence of AmA1 cDNA and other seed-specific proteins.

DISCUSSION

We report the isolation and characterization of a full-length cDNA clone that encodes a nutritionally balanced 35-kDa protein. The protein was purified to homogeneity and found to be developmentally regulated and seed-specific. It shows

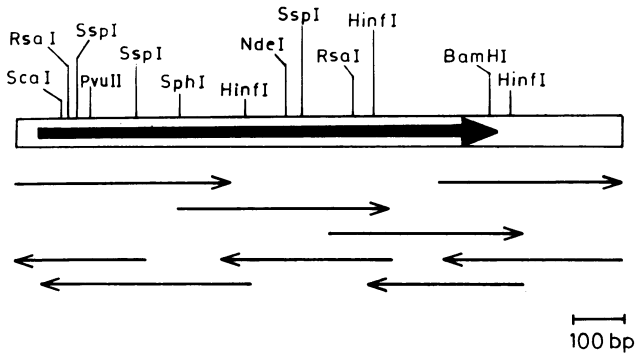


FIG. 6. Sequencing strategy and the restriction map of AmA1 cDNA. The direction of transcription and length of the open reading frame is indicated by the solid thick arrow. The short thin horizontal arrows indicate the sequencing strategy used in both orientations.

A

```
                                                  M  A  G  L  P
ATCAGATTAACATAATTTCACAATAAAAAAAAAAAAAAAGAGCTTAAATGGCGGGATTACC      60

 V  I  M  C  L  K  S  N  N  N  Q  K  Y  L  R  Y  Q  S  D  N
AGTGATTATGTGCCTAAAATCAAATAACAACCAGAAGTACTTAAGATATCAAAGTGATAA     120

 I  Q  Q  Y  G  L  L  Q  F  S  A  D  K  I  L  D  P  L  A  Q
TATTCAACAATATGGTCTTCTTCAATTTTCAGCTGATAAGATTTTAGATCCATTAGCTCA     180

 F  E  V  E  P  S  K  T  Y  D  G  L  V  H  I  K  S  R  Y  T
ATTTGAAGTCGAACCTTCCAAGACTTATGATGGTCTTGTTCACATCAAATCTCGCTACAC     240

 N  K  Y  L  V  R  W  S  P  N  H  Y  W  I  T  A  S  A  N  E
TAACAAATATTTGGTTAGGTGGTCTCCCAATCATTATTGGATTACAGCATCAGCCAATGA     300

 P  D  E  N̄  K  S  N  W  A  C  T  L  F  K  P  L  Y  V  E  E
ACCAGATGAAAATAAAAGCAATTGGGCATGCACATTATTCAAACCACTTTACGTAGAAGA     360

 G  N  M  K  K  V  R  L  L  H  V  Q  L  G  H  Y  T  E  N̄  Y
AGGTAACATGAAAAAGGTTCGACTTTTGCACGTCCAATTAGGTCATTATACAGAAAATTA     420

 T  V  G  G  S  F  V  S  Y  L  F  A  E  S  S  Q  I  D  T  G
TACCGTTGGTGGGTCCTTCGTATCATACTTATTTGCCGAATCAAGTCAAATTGATACCGG     480

 S  K  D  V  F  H  V  I  D  W  K  S  I  F  Q  F  P  K  T  Y
CTCTAAAGACGTATTCCATGTCATAGATTGGAAATCAATCTTTCAATTTCCCAAAACATA     540

 V  T  F  K  G  N  N  G  K  Y  L  G  V  I  T  I  N  Q  L  P
TGTCACATTTAAAGGAAATAATGGAAAATATTTAGGGGTTATCACAATTAATCAACTTCC     600

 C  L  Q  F  G  Y  D  N  L  N  D  P  K  V  A  H  Q  M  F  V
ATGTCTACAATTTGGGTATGATAATCTTAATGATCCAAAGGTGGCTCATCAAATGTTTGT     660

 T  S  N̄  G  T  I  C  I  K  S  N  Y  M  N  K  F  W  R  L  S
CACTTCTAATGGTACTATTTGCATTAAATCCAATTATATGAACAAGTTTTGGAGACTCTC     720

 T  D  N  W  I  L  V  D  G  N  D  P  R  E  T  N  E  A  A  A
TACGGATAATTGGATATTAGTCGATGGGAATGATCCTCGCGAAACTAATGAAGCTGCTGC     780

 L  F  R  S  D  V  H  D  F  N  V  I  S  L  L  N  M  Q  K  T
GTTGTTTAGGTCGGATGTGCATGATTTTAATGTGATTTCGCTTTTGAACATGCAAAAAAC     840

 W  F  I  K  R  F  T  S  G  K  P  E  F  I  N  C  M  N  A  A
TTGGTTTATTAAGAGATTTACGAGTGGTAAGCCTGAGTTTATAAATTGTATGAATGCAGC     900

 T  Q  I  V  D  E  T  A  I  L  E  I  I  E  L  G  S  N  N
TACTCAAATTGTTGATGAAACTGCTATTTTAGAGATAATAGAATTGGGATCCAACAACTA     960

ATATATTGGATTGCTTTTAAGATTCAAATTAAAGTCTAGTTGTTAATGTAAGGAATAAAA    1020

CGTTGTAAGTCGTCTCTTTGGAAACAAGAGGGTTCTTCCTTGTATCATATCTCTATGGTC    1080

TCTTTCAGATTTTGACCATAAGATTACTATTAAATACTTGTAATGTGTTTGTCTGTGATG    1140

ATTACTCTTTGTTGGAATAAAATAATTGTTAGAATTATATTAC                     1183
```

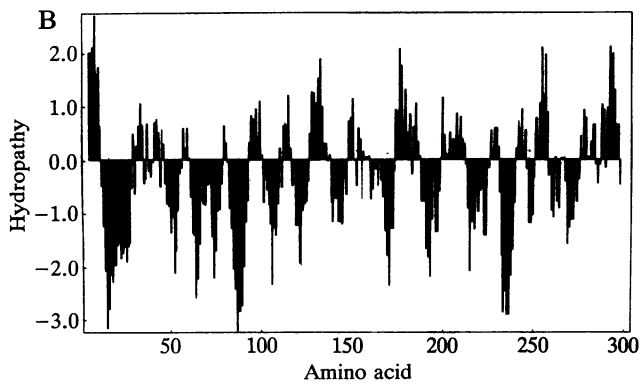


FIG. 7. (*A*) Nucleotide sequence of the 1183-bp AmA1 cDNA and the derived amino acid sequence of the predicted polypeptide. Amino acids (single-letter code) are indicated above the first base of each codon. Possible glycosylation sites are overlined. A poly(A) tract was not found although poly(A) addition signals were present and are underlined. (*B*) Hydropathic plot of the AmA1 protein. The 304-amino acid sequence of the predicted AmA1 protein was analyzed by the method of Kyte and Doolittle (23) with a window setting of 7.

a number of characteristics that distinguish it from other seed storage proteins.

AmA1 is synthesized during early embryogenesis and stored until maturation. On germination, the level of AmA1 does not immediately go down but by the third day of germination it completely disappeared (data not shown). By then the seed also has disintegrated completely. However, when we looked at two globulins of 54 and 33 kDa, their level was considerably reduced by the first day (data not shown). So the function of this protein remains unclear. Albumins from other seeds are known to have storage functions (24). Higgins *et al.* (25) have reported a similar albumin from pea that may have storage function; the delayed breakdown of this protein could be due to its cytoplasmic localization. Proteolytic enzymes are not released into the cytosol until the final stages of cellular disorganization. A similar mechanism may also regulate AmA1.

Table 2. Amino acid composition of AmA1

Amino acid	Residues, mol %		Amino acid	Residues, mol %	
	Purified protein	cDNA sequence		Purified protein	cDNA sequence
Phe	5.6	5.6	Thr	5.3	5.9
Tyr	4.8	4.9	Pro	3.6	3.6
Leu	7.6	8.6	Gly	12.0	5.3
Ile	5.0	6.6	Glx	8.9	9.2
Val	5.6	6.2	Asx	16.6	14.5
Met	1.6	2.3	Ser	6.6	6.9
Cys	0.7	1.6	Arg	5.3	2.6
Ala	5.3	4.6	Lys	6.6	6.9
His	3.3	2.3	Trp	*	2.3

*Not determined.

The amino acid composition obtained from purified protein and the derived amino acid sequence shows a high proportion of essential amino acids such as lysine, leucine, threonine, phenylalanine, valine, and sulfur amino acids (Table 3) that are otherwise deficient in the major seed proteins of legumes and cereals. It has relatively low level of glutamine compared to other seed storage proteins (26). Interestingly, the AmA1 amino acid composition closely matches the values recommended by the World Health Organization, making it more important nutritionally.

Seed storage proteins are localized in protein bodies and are often glycosylated. AmA1 was found to be present in the cytosol and not localized in protein bodies. This was confirmed by cDNA analysis. The putative polypeptide encoded by the open reading frame of AmA1 cDNA is 35 kDa. In addition, hybrid-selected translation of AmA1 mRNA gave a 35-kDa polypeptide, the size of the purified protein (Fig. 3*B*). A hydropathy plot of the deduced protein sequence indicates the presence of a small stretch of hydrophobic amino acids near the N terminus (Fig. 7*B*). The functional relevance of this region is not yet established. This does not appear to be a signal peptide as the amino acids after this stretch do not meet the requirement of a signal sequence (27, 28). Three putative glycosylation sites are also present in the sequence although the protein was not glycosylated as indicated by periodic acid/Schiff reagent staining (data not shown).

The AmA1 gene is expressed during early embryogenesis. Mature seeds even after 1 year of storage contained AmA1 mRNA, although at a reduced level, suggesting that it is very stable. Germinated seeds did not contain any AmA1 mRNA.

Table 3. Percentage of essential amino acids of AmA1 in comparison to the World Health Organization recommended values

Amino acid	% of total amino acids		
	Amaranth* (total protein)	AmA1†	WHO*
Trp	1.4	3.6	1.0
Met/Cys	4.4	3.9	3.5
Thr	2.9	5.1	4.0
Ile	3.0	6.1	4.0
Val	3.6	5.2	5.0
Lys	5.0	7.5	5.5
Phe/Tyr	6.4	13.7	6.0
Leu	4.7	9.2	7.0

WHO, World Health Organization.
*From ref. 5.
†Calculated by considering total residue number of each amino acid from the sequence and their respective molecular weights.

No RNA was detected in leaves and roots, suggesting that the expression is tissue-specific. Such specificity of expression is due to sequences in the promoter and enhancer regions of the gene.

To summarize, AmA1 is a seed-specific protein that unlike many seed storage proteins contains high levels of the essential amino acids. Because of its high nutritional value, the gene encoding this protein may be able to compensate for amino acid deficiencies of many seed proteins once it is genetically engineered into target plants.

We thank Dr. R. Nagraj (Centre for Cellular and Molecular Biology, Hyderabad, India) and Dr. R. K. Prusti (University of Washington, Seattle) for the amino acid composition. This research was supported partly by a grant to A.D. from the Department of Biotechnology, Ministry of Science and Technology, Government of India, as part of the Centre for Plant Molecular Biology.

1. Larkins, B. A. (1983) *Genetic Engineering of Plants: An Agricultural Perspective* (Plenum, New York), pp. 93–118.
2. Hoffman, L. M., Donaldson, D. D. & Herman, E. M. (1988) *Plant Mol. Biol.* **11**, 717–729.
3. Guerche, P., De Almeida, E., Schwarztein, M. A., Gander, E., Krebbers, E. & Pelletier, G. (1990) *Mol. Gen. Genet.* **221**, 306–314.
4. Larkins, B. A. (1981) *The Biochemistry of Plants* (Academic, New York).
5. Senft, J. P. (1980) *Proceedings of the 2nd Amaranth Conference* (Rodale, Emmaus, PA), pp. 43–47.
6. Paredes-Lopez, O., Mora-Escobedo, R. & Ordorica-Falonir, C. (1988) *Lebensm.-Wiss. Technol.* **21**, 59–61.
7. Prestidge, R. L. & Hearn, M. T. W. (1979) *Anal. Biochem.* **97**, 95–102.
8. Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685.
9. O'Farrell, P. H. (1975) *J. Biol. Chem.* **250**, 4007–4021.
10. Elledge, S. J. & Davis, R. W. (1987) *Mol. Cell. Biol.* **7**, 2783–2793.
11. Towbin, H., Staehelin, T. & Gordon, J. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 4350–4354.
12. Salvinovich, O. & Montelaro, R. (1986) *Anal. Biochem.* **156**, 341–347.
13. Ausubel, F. M., Kingston, B. R., Moore, D. D., Seidman, J. G., Smith, A. J. & Struhl, K. (1987) *Current Protocols in Molecular Biology* (Greene/Wiley-Interscience, New York), pp. 4.3.1–4.3.3.
14. Okayama, H., Kawaichi, M., Brownstein, M., Lee, F., Yokata, T. & Arai, K. (1987) *Methods Enzymol.* **154**, 3–28.
15. Anderson, D. J. & Blobel, G. (1983) *Methods Enzymol.* **96**, 111–120.
16. Chamberlain, J. P. (1979) *Anal. Biochem.* **98**, 132–135.
17. Ricciardi, R. P., Miller, J. S. & Roberts, B. E. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 4927–4931.
18. Gubler, U. & Hoffman, B. J. (1983) *Gene* **25**, 263–269.
19. Young, R. A. & Davis, R. W. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1194–1198.
20. Manfioletti, G. & Schneider, C. (1988) *Nucleic Acids Res.* **60**, 2873–2884.
21. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
22. Feinberg, A. & Vogelstein, B. (1984) *Anal. Biochem.* **137**, 266–267.
23. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132.
24. Croy, R. R. D., Hoque, M. S., Gatehouse, J. A. & Boulder, D. (1984) *Biochem. J.* **218**, 795–803.
25. Higgins, T. J. V., Beach, L. R., Spencer, D., Chandler, P. M., Randall, P. J., Blagrove, R. J., Kortt, A. A. & Guthrie, R. E. (1987) *Plant Mol. Biol.* **8**, 37–45.
26. Higgins, T. J. V. (1984) *Annu. Rev. Plant Physiol.* **35**, 191–221.
27. Perlman, D. & Halvorsen, H. O. (1983) *J. Mol. Biol.* **167**, 391–409.
28. Von Heijne, G. (1984) *J. Mol. Biol.* **173**, 243–251.