Nucleotide Sequences of a Soybean Complementary DNA Encoding a 50-Kilodalton Late Embryogenesis Abundant Protein

Yue-ie C. Hsing*, Zuei-ying Chen, and Teh-yuan Chow
Institute of Botany, Academia Sinica, Taipei, Taiwan, Republic of China

We report the sequence of a cDNA clone, pGmPM2, corresponding to a soybean mature seed-abundant mRNA (GmPM2). The characteristics of pGmPm2 cDNA are given in Table I. It was selected from a pod-dried cotyledon cDNA λZAPII library screened by differential hybridization with single-stranded cDNA probes prepared from immature seeds (35 DAF) and pod-dried seeds poly(A+) RNA (7). Figure 1 shows the nt sequences of the 1741-base pair cDNA insert and the derived aa sequence.

Only one long open reading frame was found, and the mol wt of the deduced protein was similar to that predicted by hybrid select translation (7). The deduced protein is very hydrophilic, as are most other Lea proteins (3), and consists of 463 aa residues corresponding to a molecular mass of 50.6 kD. One potential polyadenylation signal, AATAAA, is found in the 3′-noncoding region, 102 nt upstream from the poly(A+) tail. A search for polypeptide homologies in data banks revealed a strong local similarity with several Lea proteins, including embryonic protein DC8 in carrot (4), Lea proteins D7 and D29 in cotton (1), ABA-induced protein PHVA1 in barley (6), and Lea protein 76 in rape (5). These genes have been shown to be expressed during late seed development and also respond to ABA or water stress treatments. Another protein with strong local similarity is the Plasmodium falciparum S antigen present in the sera of some malaria-infected individuals (2). All of these proteins are soluble heat-stable proteins with a series of tandemly repeated aa domains, and most of these proteins have no Trp or Cys residues.

LITERATURE CITED

---

1 This research was supported in part by a grant from the National Science Council to Y.C.H. and T.C.
2 Abbreviations: GmPM, Glycine max physiological mature; Lea, late embryogenesis abundant; aa, amino acid(s); nt, nucleotide; poly(A+), polyadenylate.
Table I. Characteristics of pGmPM2 cDNA

Organism:
Soybean (Glycine max L. Merrell), variety Williams'82.

Function:
Encodes 50-kD soybean seed maturation polypeptides (7).

Clone Type; Designation:
cDNA, full-length, pGmPM2.

Source:
cDNA library in λZAPII vector constructed from the poly(A)’ RNA of 4-d pod-dried 35 DAF soybean cotyledons.

Method of Identification:
Differential hybridization with single-stranded cDNA probes prepared from fresh immature seeds (35 DAF) and 4-d pod-dried seed poly(A)’ RNA, pGmPM2 showed a strong hybridization signal only with the homologous probe.

Sequence Strategy:
Single-strand DNA template; unidirectional deletion subcloning and complete dideoxy sequencing of both strands.

Feature of mRNA Structure:
Transcript of approximately 1800 nts as detected on northern blots; this clone of 1741 nt contains 57 nt 5’ untranslated region, 1389 nt open reading frame, and 294 nt 3’ untranslated region.

Codon Usage:
Codons not present in the cDNA: TGG(W), TGT(C), TGC(C), TTA(L), CGG(R), CCA(A), CCG(G), and CCG(P), and CCT(P).

(G + C) Content:
46.82% along entire length; 50.86% in protein-coding region.

Structural Features of Protein:
463 aa. No Trp or Cys residues are present in the protein. There are internal repeats from aa 129 to 260 and aa 304 to 398, with the consensus sequences of VNKMGEYKDYAAEKGERDAT and TAET.EAA.K, respectively.

Antibody:
Not available.

GenBank Accession Number:
M80664.

---

Figure 1. Nt sequence of the 1741-base pair cDNA insert and the derived aa sequence of pGmPM2 from soybean. Underlining, the start codon, stop codon, and potential polyadenylation sites; asterisks, the putative glycosylation sites.