

Plant Gene Register

Nucleotide Sequence of a cDNA Clone Corresponding to the Maize Globulin-2 Gene¹

Nancy H. Wallace² and Alan L. Kriz*

Department of Agronomy, University of Illinois, Urbana, Illinois 61801

Maize globulins represent a group of saline-soluble proteins that accumulate primarily in the developing embryo (4, 6, 7). The most abundant globulin components are the M_r 63,000 GLB1 protein and the M_r 45,000 GLB2 protein, which are encoded by the *Globulin-1* (*Glb1*) and *Globulin-2* (*Glb2*) genes, respectively (4, 7). GLB1 and GLB2 collectively comprise 10 to 20% of the protein present in the mature maize embryo. The amino acid composition and the developmental accumulation pattern of GLB1 and GLB2 suggest that these are embryo storage proteins (4). GLB1 exhibits sequence similarity to 7S seed globulins of both monocots and dicots (2, 3).

A common characteristic of seed storage proteins described to date is that they are encoded by multigene families (8). In contrast, the *Glb1* gene exists as a single copy in the haploid genome as determined by genetic analyses of several electrophoretic variants (7) and as recently confirmed at the molecular level by gene copy-number reconstruction analysis (2). Although we have not observed electrophoretic variants for GLB2, two null alleles for *Glb2* have been identified (Table I). GLB2 also appears to be encoded by a single gene as determined by genetic analyses in which each of the *Glb2* null alleles segregates as a Mendelian recessive trait and by Southern blot analysis of maize DNA in which *Glb2*-specific clones are used as probes. The availability of allelic variants for both of the single-copy maize *Glb* genes allows for detailed analysis of allelic polymorphism in plants at the molecular level.

Recently, we reported (2) the isolation and characterization of a cDNA clone corresponding to the *Glb1-S* allele, which

represents the first monocot 7S globulin to be analyzed at the molecular level. We present here the nucleotide sequence of a full-length cDNA clone corresponding to *Glb2*. The amino acid sequence deduced from the pcGlb2 nucleotide sequence (Fig. 1) indicates that GLB2 is synthesized as a preproprotein, as is the case for GLB1 (2, 5) and other seed storage globulins (8). Alignment of mature GLB1 and GLB2 from their amino termini indicates that the two polypeptides share 28% amino acid sequence identity. Although a slight amount of immunological cross-reactivity has been observed for the GLB2 and GLB1 polypeptides, the two cDNA clones pcGlb2 and pcGlb1S do not cross-hybridize at a T_m of -25°C . Efforts are underway to isolate genomic clones corresponding to *Glb2* as well as the null alleles *Glb2-0* and *Glb2-N1* to determine the molecular basis of the null phenotype in each case.

LITERATURE CITED

1. Abersold RH, Teplow DB, Hood LE, Kent SBH (1986) Electrophoretic blotting onto activated glass: high efficiency preparation of proteins from analytical sodium dodecylsulfate polyacrylamide gels for direct sequence analysis. *J Biol Chem* **261**: 4229–4238
2. Belanger FC, Kriz AL (1989) Molecular characterization of the major maize embryo globulin encoded by the *Glb1* gene. *Plant Physiol* **91**: 636–643
3. Dure L III (1990) An unstable domain in the vicilin genes of higher plants. *New Biol* **2**: 487–493
4. Kriz AL (1989) Characterization of embryo globulins encoded by the maize *Glb* genes. *Biochem Genet* **27**: 239–251
5. Kriz AL, Schwartz D (1986) Synthesis of globulins in maize embryos. *Plant Physiol* **82**: 1069–1075
6. Puckett JL, Kriz AL (1991) Globulin gene expression in *opaque-2* and *floury-2* mutant maize embryos. *Maydica* **36**: 1–7
7. Schwartz D (1979) Analysis of the size alleles of the *Pro* gene in maize—evidence for a mutant protein processor. *Mol Gen Genet* **174**: 233–240
8. Shotwell MA, Larkins BA (1989) The biochemistry and molecular biology of seed storage proteins. In A Marcus, ed, *The Biochemistry of Plants: A Comprehensive Treatise*, Vol 15. Academic Press, New York, pp 297–345
9. von Heijne G (1986) A new method for predicting signal sequence cleavage sites. *Nucleic Acids Res* **14**: 4683–4690

¹ This contribution from the University of Illinois Agricultural Experiment Station was supported in part by a grant to A. L. K. from the U. S. Department of Agriculture Competitive Research Grants Office (No. 88-37262-3427). N. H. W. was recipient of a College of Agriculture Hackett Fellowship.

² Present address: DEKALB Plant Genetics, Eastern Point Road, Groton, CT 06340.

Table I. Characteristics of pcGlb2

Organism:

Zea mays L., inbred line Va26.

Locus:

Glb2; single copy gene by segregation analysis (4) and gene copy number reconstruction analysis. Two null alleles known: *Glb2-0* (4) and *Glb2-N1* (AL Kriz, unpublished data).

Function:

Encodes vicilin-like 7S embryo storage protein (4).

Clone Type; Designation:

cDNA, full-length; pcGlb2.

Source:

Expression library in LambdaZAP vector constructed from 27 days after pollination embryo poly(A) RNA (2).

Method of Identification:

Library was screened with antiserum to whole embryo globulin (6) to identify immunoreactive clones that were subsequently screened with a radiolabeled fragment from pcGlb1S (2) to eliminate clones corresponding to *Glb1*. Full-length *Glb2* clones were obtained by rescreening the library with a restriction fragment from the 5' region of the longest immunoreactive *Glb2*-specific clone.

Sequencing Strategy:

M13 ssDNA templates; unidirectional deletion subcloning and complete dideoxy sequencing of both strands as previously described (2).

Confirmation:

Amino-terminal sequence of the mature GLB2 protein, as determined by microsequence analysis (1, 4), is underlined in Figure 1. Molecular weight of deduced mature polypeptide (*M*_r 47,377) is in good agreement with that estimated from SDS-PAGE for mature GLB2 protein (*M*_r 45,000).

Features of mRNA Structure:

Transcript of ~1700 nucleotides (n) as detected on Northern blots; this clone of 1637 n contains 47 n 5' untranslated region, 1350 n open reading frame, and 240 n 3' untranslated region.

Codon Usage:

89% XXC/G.

(G + C) Content:

64% along entire length; 66% in protein coding region.

Structural Features of Protein:

450 amino acids (aa), including 19 aa signal sequence and 427 aa mature protein. Signal sequence cleavage site predicted by method of von Heijne (9). The four aa between this site and N terminus of mature GLB2 (Fig. 1) is presumably removed in a posttranslational manner. Presence of a consensus N-linked glycosylation site (boxed in Fig. 1) is consistent with observation that GLB2 binds Con A (4).

Antibodies:

Antisera to whole embryo globulin available.

Subcellular Location:

Present in ER and protein storage vacuoles (protein bodies) through cell fractionation studies.

EMBL Accession Number:

X53715 Z. MAYS GLB2 CDNA.

```

-47  TCGCACTCGCACGTACACCACACTAGTTACCACAGACGACGGGCGCC  ATG  AAA  GTC  CCG  GTG  CTT  CTC  CTC  CTG  GTC  TCC  CTG  TGC  TTC  TCG  CTC  GCG  CTC
      M  K  V  P  V  L  L  L  L  V  S  L  C  F  S  L  A  L
      ↓
55  GCG TGG CAA ACG GAC ACG GAA TCG GGC TCA GGC AGG CCG TAC CAC TAC GGC GAG GAG AGT TTC CGG CAC TGG ACG CGC TCC CGG CAG GCG
    A  W  Q  T  D  T  E  S  G  S  G  R  P  Y  H  Y  G  E  E  S  F  R  H  W  T  R  S  R  Q  G

145  CGG TTC AGG GTG CTG GAA CGG TTC ACC CAC GAG CTG CTG GAG GAC GCC GTC GGC AAC TAC CGC GTC GCC GAG CTG GAG GCC GCG CCG CGC
    R  F  R  V  L  E  R  F  T  H  E  L  L  E  D  A  V  G  N  Y  R  V  A  E  L  E  A  A  P  R

235  GCG TTC CTG CAG CCC AGC CAC TAC GAT GCC GAC GAG GTG ATG TTC GTG AAG GAA GGC GAG GGC GTC ATC GTG CTG CTC CGC GGC GGG AAG
    A  F  L  Q  P  S  H  Y  D  A  D  E  V  M  F  V  K  E  G  E  G  V  I  V  L  L  R  G  G  K

325  AGG GAG TCG TTC TGC GTC AGG GAG GGC GAC GTC ATG GTC ATC CCC GCG GGC GCC GTC GTG TAC TCG GCC AAC ACG CAC CAG TCG GAG TGG
    R  E  S  F  C  V  R  E  G  D  V  M  V  I  P  A  G  A  V  V  Y  S  A  N  T  H  Q  S  E  W

415  TTC CGC GTC GTC ATG CTC CTC AGC CCC GTC GTC TCC ACG TCT GGA CGC TTC GAG GAG TTC TTC CCC ATC GGA GGC GAG AGC CCC GAG TCC
    F  R  V  V  M  L  L  S  P  V  V  S  T  S  G  R  F  E  E  F  F  P  I  G  G  E  S  P  E  S

505  TTC CTC AGC GTC TTC AGC GAC GAC GTT ATC CAG GCG TCG TTC AAC ACT CGC CGG GAG GAG TGG GAG AAA GTG TTC GAG AAG CAG AGC AAG
    F  L  S  V  F  S  D  D  V  I  Q  A  S  F  N  T  R  R  E  E  W  E  K  V  F  E  K  Q  S  K

595  GGG GAG ATC ACG ACG GCG TCC GAG GAA CAG ATC CGG GAG CTG AGC AGG TCC TGC TCA CGT GGT GGC CGC AGC AGC CGC AGC GAG GGT GGC
    G  E  I  T  T  A  S  E  E  Q  I  R  E  L  S  R  S  C  S  R  G  G  R  S  S  R  S  E  G  G

685  GAC TCG GGT TCT TCT TCT TCC AAG TGG GAG ATC AAG CCG TCC AGC CTC ACG GAC AAG AAG CCG ACC CAT TCC AAC AGC CAC GGC AGG CAC
    D  S  G  S  S  S  S  K  W  E  I  K  P  S  S  L  T  D  K  K  P  T  H  S  N  S  H  G  R  H

775  TAC GAG ATC ACC GGC GAC GAG TGC CCG CAC CTC CGG CTG CTC GAC ATG GAC GTC GGC CTG GCA AAC ATC GCC CGC GGT TCT ATG ATG GCG
    Y  E  I  T  G  D  E  C  P  H  L  R  L  L  D  M  D  V  G  L  A  N  I  A  R  G  S  M  M  A

865  CCG AGC TAC AAC ACC CGC GCC AAC AAG ATA GCT ATC GTC CTC AAG GGC CAG GGC TAC TTC GAG ATG GCG TGC CCG CAC GTC TCC GGC GGC
    P  S  Y  N  T  R  A  N  K  I  A  I  V  L  K  G  Q  G  Y  F  E  M  A  C  P  H  V  S  G  G

955  CGG TCG TCG CCG CGC CGG GAG CGT GGG CAC GGC CGA GAG GAA GAG GAG GAG CGG GAA GAG GAG CAA GGC GGT GGC GGT GGG CAA AAG TCG
    R  S  S  P  R  R  E  R  G  H  G  R  E  E  E  E  E  R  E  E  E  Q  G  G  G  G  G  G  Q  K  S

1045 AGG AGC TAC AGG CAG GTG AAA TCC CGC ATC AGA GAG GGC TCC GTC ATC GTG ATC CCG GCG GGG CAC CCG ACC GCG CTG GTC GCC GGC GAA
    R  S  Y  R  Q  V  K  S  R  I  R  E  G  S  V  I  V  I  P  A  G  H  P  T  A  L  V  A  G  E

1135 GAC AAG AAC CTC GCC GTC CTC TGC TTC GAA GTG AAC GCC AGC TTC GAC GAT AAG GTG TTC CTC GCG GGC ACG AAC AGC GCG CTT CAG AAG
    D  K  N  L  A  V  L  C  F  E  V  N  A  S  F  D  D  K  V  F  L  A  G  T  N  S  A  L  Q  K

1225 ATG GAT AGG CCG GCG AAG CTG CTG GCG TTC GGG GCG GAC GAG GAG CAG CAG GTG GAC CGT GTC ATA GGA GCA CAG AAG GAC GCC GTC TTC
    M  D  R  P  A  K  L  L  A  F  G  A  D  E  E  Q  Q  V  D  R  V  I  G  A  Q  K  D  A  V  F

1315 CTG CGC GGC CCC CAG AGC CAC AGG GTC TCG TCG GTG TAA GTAGTGCATGCTTGAGATGTAGCGGCCGCTCATCGGCCTCGAGACCTCGACGGTCGCGGCCACCCCG
    L  R  G  P  Q  S  H  R  V  S  S  V  *

1421 GTGGTGAGGTTAAGAAGTGAATAAAGTAGGTGATGGCTCGCCGCGTCGCGTTACGTATGAACAGTAGTATGTACGTCCAGTGAGGGTGACGCCCGGCCGTACGGGAGTATGTAATC

1540 TACCTACCTACCTATGCAGGTGAAATAACGTTGTTGAGACTTGAGAAAAAA

```

Figure 1. Nucleotide sequence and deduced amino acid sequence of pGlb2. The open reading frame extends from positions +1 to +1350 and encodes a polypeptide of 450 amino acids. The stop codon TAA is indicated by an asterisk. The site of cleavage of the putative 19 amino acid signal peptide, as predicted by the weighted matrix method of von Heijne (9), is indicated by an arrow. A sequence of 15 amino acids (underlined) represents the amino-terminal portion of GLB2 determined by direct protein sequence analysis. One putative glycosylation site, Asn-X-Ser, is indicated by a box. A potential polyadenylation site at nucleotide position +1562 is indicated by an overline.