

## Plant Gene Register

# Nucleotide Sequence of a cDNA Clone Encoding $\gamma$ -Kafirin Protein from *Sorghum bicolor*

Everaldo G. de Barros, Koichi Takasaki, Allen W. Kirleis, and Brian A. Larkins\*

Department of Plant Sciences, University of Arizona, Tucson, Arizona 85721 (E.G.d.B., K.T., B.A.L.), and  
Department of Food Science, Purdue University, West Lafayette, Indiana 47907 (A.W.K.)

The storage protein in sorghum (*Sorghum bicolor* [L.] endosperm consists of a group of alcohol-soluble proteins, the kafirins (5). Separation by SDS-PAGE demonstrated that kafirins are a heterogeneous mixture of polypeptides (9). Much attention has been focused on the  $\gamma$ -kafirin (2, 7–9). This protein, also called reduced-soluble protein (8), is soluble in water or aqueous alcohol plus a reducing agent and resembles very closely the 27-kD  $\gamma$ -zein from maize (9). The  $\gamma$ -kafirin has been purified and partially characterized (2, 8, 9). The  $\gamma$ -kafirin shows slightly faster mobility than 27-kD  $\gamma$ -zein in SDS polyacrylamide gels (9). Amino acid analysis demonstrated that its amino acid composition is very similar to that of the 27-kD  $\gamma$ -zein protein, and the first 11 amino acids of the N-terminus are almost identical between the two proteins (8). Additionally, the fact that antibodies raised against the 27-kD  $\gamma$ -zein cross-react with the  $\gamma$ -kafirin provides additional evidence of the similarity between the two proteins (9).

Based on the resemblance of the two proteins, we reasoned that cDNA(s) encoding the  $\gamma$ -kafirin would readily hybridize with a 27-kD  $\gamma$ -zein gene. A cDNA library prepared with poly(A) mRNA from 20-d post-pollination sorghum kernels was screened with a genomic clone corresponding to the coding region of the 27-kD  $\gamma$ -zein. Three of the hybridizing clones were completely sequenced. Clones sorgI (648 base pairs [bp], covering +112 to +759 of sorgH) and sorgM (625 bp, covering +134 to +758 of sorgH) were not full-length. However, clone sorgH (848 bp) starts 41 bp 5' of the translation initiation codon and ends 171 bp 3' of the stop codon (Table I, Fig. 1). The deduced amino acid sequence from the sorgH cDNA suggests that this clone encodes 212 amino acids with a signal peptide of 19 amino acids (Fig. 1). Furthermore, 26 out of the first 28 amino acids coincide with the previously described N-terminal amino acid sequence of the  $\gamma$ -kafirin (8). The mature  $\gamma$ -kafirin protein has 71% and 67% identity at the nucleotide and amino acid levels, respectively, with mature 27-kD  $\gamma$ -zein protein (1, 6). The 27-kD  $\gamma$ -zein appears to play an important role in protein body formation and endosperm structure (3, 4). Studies are in progress to determine if the  $\gamma$ -kafirin has similar functions in sorghum endosperm.

**Table I.** Characteristics of sorgH

Organism:	<i>Sorghum bicolor</i> (L.) Moench cultivar P721N.
Locus:	Unknown.
Function:	Encodes $\gamma$ -kafirin.
Clone Type; Designation:	cDNA, full-length; sorgH.
Source:	Expression library in LambdaZAP vector constructed with poly(A) RNA from 20-d post-pollination kernels.
Method of Identification:	Library was screened with maize genomic clone encoding 27-kD $\gamma$ -zein.
Sequencing Strategy:	Plasmid sequencing with restriction fragment subcloning and complete dideoxy sequencing of both strands.
Confirmation:	Comparison of deduced N-terminal amino acid sequence with previously described N-terminal amino acid sequence (8); comparison of deduced amino acid composition (mole %) with chemically determined amino acid composition (2, 9); slightly smaller size of mature protein (193 amino acids) than 27-kD $\gamma$ -zein (204 amino acids) is in good agreement with electrophoretic mobility of these polypeptides by SDS-PAGE (9).
Features of mRNA Structure:	This clone of 848 nucleotides(n) contains 41n 5'-untranslated region, 636n open reading frame, and 171n 3'-untranslated region.
Codon Usage:	72% XXC/G.
(G + C) Content:	60.3% along entire length; 67.5% in protein coding region.
Structural Features of Protein:	212 amino acids (aa) ( <i>M</i> , 22,087), including 19 aa signal sequence (Fig. 1) and 193 aa mature protein ( <i>M</i> , 20,278); mature protein contains four repeats of Pro-Pro-Pro-Val-His-Leu/Val (Fig. 1) found in maize 27-kD $\gamma$ -zein (1); mature protein contains high levels of Pro, Cys, His and no Lys, Asn, Asp, Trp.
Antibodies:	Available.
Subcellular Location:	Present in protein storage vacuoles (protein bodies).
GenBank Accession No.:	M73688.

