Vol. 193, No. 1, 1993 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
May 28, 1993 Pages 460-466

NUCLEOTIDE SEQUENCE OF AN ABUNDANT RICE SEED GLOBULIN: HOMOLOGY WITH THE HIGH MOLECULAR WEIGHT GLUTELINS OF WHEAT, RYE AND TRITICALE

Hari B. Krishnan* and Steven G. Pueppke

Department of Plant Pathology, 108 Waters Hall, University of Missouri, Columbia, MO 65211

Received April 16, 1993

A cDNA clone corresponding to a 19 kD salt-soluble globulin of rice (*Oryza sativa* L.) was isolated by screening a λgt11 expression library of endosperm mRNA with antibodies raised against the purified rice seed α-globulin. The cDNA contained a single large open reading frame encoding a putative globulin precursor of molecular weight of 21 kD. The polypeptide consists of 182 amino acids and is devoid of lysine residues. Computer analysis of the NH_2-terminal sequence of the globulin precursor protein indicated the presence of an 18 amino acid signal peptide. Segments of the amino acid sequence of the rice globulin were homologous with sequences in high molecular weight glutelins of wheat. Antibodies specific for rice seed globulin also cross-reacted with seed proteins from wheat, rye and triticale.

The major storage proteins of rice (*Oryza sativa* L.) are the glutelins, which may comprise up to 80% of the total seed proteins (1,2). The alcohol-soluble prolamins, which are the predominant storage proteins of most cereals, account for only 5-10% (1,2). Both of these storage proteins are localized in specialized membrane bound organelles called protein-bodies (3). Rice seeds contain two types of protein bodies: irregular and spherical shaped. The former store the glutelin and the latter the prolamins (4). In addition to the glutelins and prolamins, rice endosperm also stores significant amounts of salt-soluble globulins (5,6,7). Immunocytochemical localization studies have indicated that the globulins are deposited in discrete zones within irregular-shaped protein bodies (7).

The rice seed globulins are composed of two abundant proteins having molecular weights of 25-20 kD (referred as α-globulin) and 16 kD. Immunoprecipation of *in vitro* translation products with antibodies raised against purified rice globulins has revealed that the globulins are synthesized as precursors having signal sequences (7).

* To whom correspondence should be addressed.

0006-291X/93 $4.00

Here we report the isolation and sequencing of a cDNA clone encoding the α-globulin. This protein is synthesized as a preprotein containing a signal sequence of about 2 kD. The rice globulin is rich in sulfur-containing amino acids but totally devoid of lysine, an essential amino acid for human nutrition. Analysis of the derived primary sequence of this protein revealed homology with high molecular weight glutelins of wheat, rye and triticale. This observation was substantiated directly by Western blot analysis.

MATERIALS AND METHODS

Rice (*Oryza sativa* L. cv. Lamont) plants were grown at the the University of Missouri Delta Research Center, Portageville. Ears were tagged at anthesis and developing rice caryopses were harvested at 15 and 20 days after flowering. Poly(A)$^+$ RNA from seeds was extracted essentially as described (8). A cDNA library of rice endosperm poly(A)$^+$ RNA was constructed in the vector λ gt11 (9). The library was screened with polyclonal antibodies prepared against the purified α- globulin (7). Five recombinant antigen-producing bacteriophage were obtained, and the clone containing the largest insert was subcloned into pBluescript SK+ (Stratagene, La Jolla, CA). The nucleotide sequence was determined by the dideoxy chain termination method using Sequenase (US Biochemical Co., Cleveland, OH). Northern blot analysis was performed according to Sambrook, Fritsch and Maniatis (10). The blots were probed with randomly primed cDNA insert. SDS-polyacrylamide gel electrophoresis and Western blotting was carried out as described earlier (7).

RESULTS AND DISCUSSION

Antibodies raised against the purified rice seed α-globulin were used to screen a λ gt11 expression library. Five putative globulin clones were obtained, and the clone with the longest insert was chosen for further analysis. It contained a sequence of approximately 820 bp and was subcloned into the *Eco*RI site of pBluesript SK+ as pLV1. The identity of the insert in pLV1 was verified by generating RNA transcripts with an *in vitro* transcription system (Promega, Madison, WI) that relies on T3 RNA polymerase. The generated RNA was translated *in vitro* with a rabbit reticulocyte system in the presence of dog microsomal membranes. The reaction products were immunoprecipated with antiserum specific for the rice α-globulin (Fig.1). Three proteins of 21, 19 and 16 kD, were specifically immunoprecipated by the antiserum (lane A), but no proteins were immunoprecipated when preimmune serum was used (lane B). The 21 kD protein appears to be the globulin precursor, which presumably undergoes processing to a mature protein of 19 kD. This is consistent with our earlier report of a presence of a 2 kD signal sequence for the rice globulin (7).

The above results indicate that pLV1 encodes for the 19 kD globulin. However, the possibility that the 19 and 16 kD proteins may the products of translation initiation

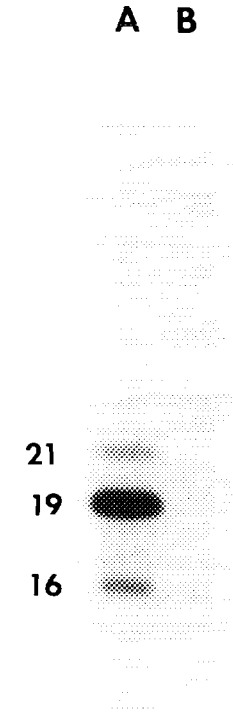

Figure 1. Characterization of cloned cDNA by *in vitro* transcription and translation. RNA transcripts were generated *in vitro* with T3 RNA polymerase after linearizing the plasmid pLV1 with *ClaI*. The RNA transcripts were translated *in vitro* with a rabbit reticulocyte lysate system in the presence of canine pancreatic microsomal membranes. The *in vitro* translation products were immunoprecipated with antiserum specific for the 25 kD globulin and resolved by electrophoresis in a 12.5% SDS-PAGE gel. Lanes A and B are the immunoprecipated products with rice globulin antibodies and preimmune serum, respectively. Values to the left indicate the molecular weights in kilodaltons.

from other methionine sites that are present in the coding region cannot be overlooked. It is also possible that the two smaller bands represent the proteolytic products of the 21 kD protein. There is also an apparent difference in the molecular weight of globulin encoded by the cDNA and the value obtained by SDS-PAGE, 25 kD. This size difference could be due to post-translational modifications. Moreover, the molecular weight of the α-globulin has been reported to vary from 25 to 20 kD in different rice cultivars (5,6,7). It will be interesting to test the possibility that the α-globulin could be glycosylated, a modification that is known to cause proteins to migrate anamolusly on SDS-PAGE.

The complete nucleotide sequence of pLV1 was determined by sequencing both strands (Fig. 2). The insert is composed of 825 bp and contains a single partial open reading frame (ORF) encoding 182 amino acids. There are no other ORF's of more than 200 bases in other reading frames. The longest open reading frame in pLV1 encodes a

```
GTCGTCTTCTTCGCGGCGGCGCTCATGGCGGCCATGGTGG     40
 V  V  F  F  A  A  A  L  M  A  A  M  V

CCATCTCCGGCGCGCACGTGAGCGAGTCGGAGATGAGGTT     80
A  I  S  G  A  H  V  S  E  S  E  M  R  F

CAGGGACAGGCAGTGCCAGCGGGAGGTGCAGGACAGCCCG    120
  R  D  R  Q  C  Q  R  E  V  Q  D  S  P

CTGGACGCGTGCCGGCAGGTGCTCGACCGGCAGCTCACCG    160
 L  D  A  C  R  Q  V  L  D  R  Q  L  T

GCCGGGAGAGGTTCCAGCCGATGTTCCGCCGCCCGGGCGC    200
G  R  E  R  F  Q  P  M  F  R  R  P  G  A

GCTCGGCCTGCGGATGCAGTGCTGCCAGCAGCTGCAGGAC    240
  L  G  L  R  M  Q  C  C  Q  Q  L  Q  D

GTGAGCCGCGAGTGCCGCTGCGCCGCCATCCGCCGGATGG    280
 V  S  R  E  C  R  C  A  A  I  R  R  M

TGAGGAGCTACGAGGAGAGCATGCCGATGCCCCTGGAGCA    320
V  R  S  Y  E  E  S  M  P  M  P  L  E  Q

AGGCTGGTCGTCGTCGTCGTCGGAGTACTACGGCGGCGAG    360
  G  W  S  S  S  S  S  E  Y  Y  G  G  E

GGGTCGTCGTCGGAGCAGGGGTACTACGGCGAGGGGTCGT    400
 G  S  S  S  E  Q  G  Y  Y  G  E  G  S

CGGAGGAGGGCTACTACGGCGAGCAGCAACAGCAGCCGGG    440
S  E  E  G  Y  Y  G  E  Q  Q  Q  Q  P  G

GATGACCCGCGTGAGGCTGACCAGGGCGAGGCAGTACGCG    480
  M  T  R  V  R  L  T  R  A  R  Q  Y  A

GCGCAGCTGCCGTCGATGTGCCGGGTTGAGCCCCAGCAGT    420
 A  Q  L  P  S  M  C  R  V  E  P  Q  Q

GCAGCATCTTCGCCGCCGGCCAGTACTAGGCTAGCTGGCT    560
C  S  I  F  A  A  G  Q  Y  *

TGCCTCATAGCTAGTAGGATCACACCGTTGCTGTGTTGCA    600
TGCAACGAGATCCATTCCTCCAAACACAATTTCACCACCT    640
TCTCCTTCATCAGAATGACGGTTGAACCAGAAGGAGAAAA    680
CAGGCTCACTGACAAGACCTTGCTCTACCATCTTATACCA    720
CACAGGTACTGCATCTCCAACGGATATTTCCTGAAATCCA    760
AGCCCAAGAATGCCATCGAATTTTGCAACCATGAAAGTAA    800
(2) GACCTGGCTCCTTGGTAGCTTCAAT               825
```

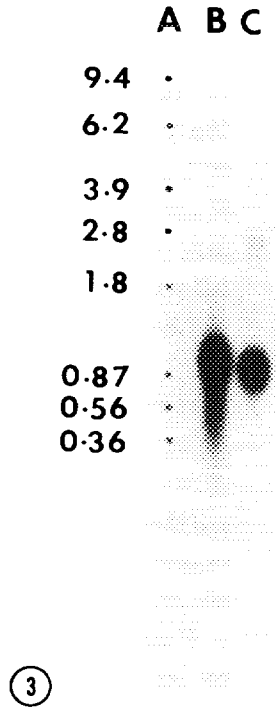

Figure 2. Nucleotide and deduced amino acid sequences of the rice 19 kD globulin cDNA. The arrowhead indicates the putative cleavage site for the signal sequence. This sequence appears in the EMBL/GeneBank/DDBJ Nucleotide Sequence Data Libraries under the accession number L12252.

Figure 3. Northern blot analysis of rice endosperm mRNA. Fifteen (lane B) and ten (lane C) μg of total RNA from rice endopem (20 days after flowering) were resolved on a formaldehyde agarose gel and transferred to nitrocellulose. The blot was probed with random-primed rice globulin cDNA insert. Lane A contains RNA markers, and the numbers to the left are the sizes of the markers in kilobases.

protein of 21 kD. Northern blot analysis of rice seed RNA probed with ^{32}P- labeled pLV1 insert showed hybridization to a transcript of approximately 870 bp (Fig. 3). The length of the cDNA thus is close to the size of the mRNA. The first 18 N-terminal

residues are rich in hydrophobic amino acids which are characteristic of signal sequences (11). Cleavage of the signal peptide presumably occurs between alanine (residue 18) and histidine (residue 19) as calculated by Von Heijne's rules (11). Thus the deduced mature protein has a molecular weight of 19 kD. The deduced amino acid composition of this protein is very similar to the published composition of the α-globulin protein (5). Glutamic acid and arginine are the two predominant amino acids, but the protein is devoid of lysine (Fig. 2).

The amino acid sequence of the 19 kD globulin is perfectly homologous to the previously reported α-globulin gene from Japonica rice (12). A comparison of the two sequences indicate that pLV1 lacks the first four N-terminal amino acids. The 3'-untranslated region of the globulin mRNA contained 276 bp. In contrast to the 19 kD globulin cDNA from Japonica rice, no putative polyadenylation signals or a poly $(A)^+$ tail were located in the 3'-untranslated region of the Indica globuin cDNA (Fig. 2). The 3'-untranslated regions of these two globulin cDNA's show significant differences, and alignment of the two sequences indicates only 59% homology in the 3'-untranslated region. The 19 kD globulin from Japonica rice is encoded by a single-copy gene (12). The divergence of sequences in the 3'-untranslated regions observed in our study imply that the α-globulin may be encoded by more than a single gene or they may be due to differences in the cultivars used.

A computerized search of the Swiss-Prot 23 data bank revealed regions of strong sequence similarity between the 19 kD globulin and the high molecular weight glutelins of wheat (13,14). Figure 4 is a comparison of the homology between rice α-globulin and the wheat glutenin. We obtained additional corroborative evidence for the existence of homology by performing Western blot analysis (Fig. 5). Rice α-globulin antisera cross reacted with several proteins from wheat, rye and triticale (Fig. 5, lanes A, B and C). The molecular weights of these proteins ranged between 65 to 80 kD. Earlier it had been shown that the high molecular weight glutenin also shares sequence homology to the acidic glutelin subunit of rice (15). This homology was restricted to a conserved peptide with an Arg-Gln-Leu-Gln-Cys motif (15). These results confirm that the rice seed α-globulin and the acidic glutenin subunit share some unique regions with the high molecular weight glutenins of wheat.

We have earlier shown that the rice globulins are deposited in irregular shaped protein bodies during the seed development (7). The same protein bodies also accumulate the glutelins, the major storage proteins of the rice seed. These proteins, which are synthesized on the rough ER, are transported to the protein bodies via the Golgi apparatus (7). Interestingly, there appears to be no significant homology between these two proteins. However, blocks of three amino acids, Met-Ala-Ser and Val-Phe-Phe, are conserved in the signal sequences of the rice α-globulin and the glutelins (12,16).

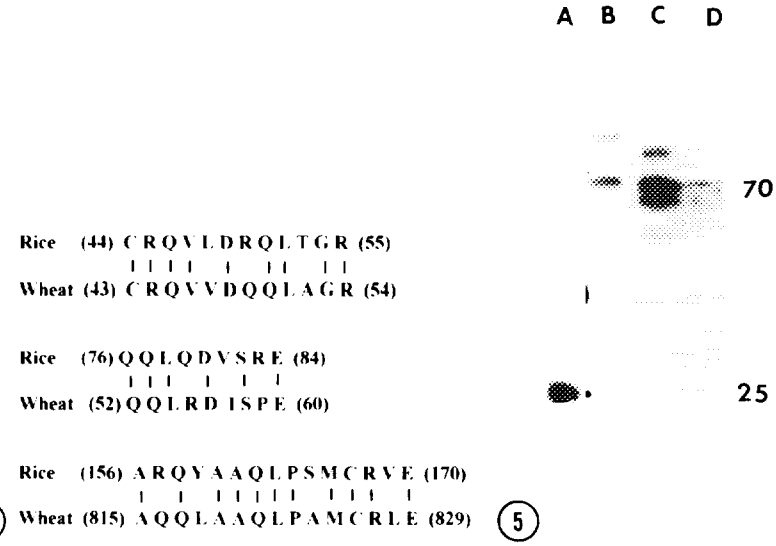

Figure 4. Amino acid sequence homology between the rice 19 kD globulin and the high molecular weight wheat seed glutenin.

Figure 5. Western Blot analysis reveals cross-reactivity of the rice globulin antibody to high molecular weight glutenin. Total proteins from rice (lane A), wheat (lane B), rye (lane C) and triticale (lane D) were fractionated on a 12.5% SDS-PAGE and electrophoretically transfered to a nitrocellulose membrane. Proteins cross-reacting with the rice seed globulin antiserum were identified by incubating the nitrocllulose with ^{125}I-labeled protein A followed by autoradiography. The numbers on the right refer to the molecular weights of the proteins in kilodaltons.

The role of these conserved sequences in the transport of these proteins to the irregular protein bodies warrants exploration. The targeting and translocation of most plant secretory proteins from the cytosol into the ER are controlled by N-terminal signal sequences (14). Vacuolar proteins also contain direct recognition elements within the polypeptide sequence that direct their transport into the vacuoles (17). It would be interesting to investigate the molecular signals contained within the rice globulins and glutelins that enable them to reach the same destination.

ACKNOWLEDGMENT

This work was supported by the Food for the 21st Century Program, University of Missouri, and is Journal Series No. 11917 of the Missouri Agricultural Experiment Station.

REFERENCES

1. Tecson, E.M.S., Esmana, B.V., Lontok, L.P., and Juliano, B.O. (1971) Cereal Chem. 48, 186-181.
2. Villareal, R.M., and Juliano, B.O. (1978) Phytochemistry 17, 177-182.

3. Tanaka, K., Sugimoto, T., Ogawa, M., and Kasi, Z. (1980) Agric. Biol. Chem., 44, 1633-1639.
4. Krishnan, H.B., Franceschi, V.R., and Okita, T.W. (1986) Planta 169, 471-480.
5. Houston, D.F., and Mohammad, A. (1970) Cereal Chem. 47, 5-12.
6. Perdon, A.A., and Juliano, B.O. (1978) Phytochemistry 17, 351-353.
7. Krishnan, H.B., White, J.A., and Pueppke, S.G. (1992) Plant Science 81, 1-11.
8. Krishnan, H.B., and Pueppke, S.G. (1989) J. Plant Physiol. 135, 253-256.
9. Huynh, T.V., Young, R.A., and Davis, R.W. (1985) In DNA Cloning A Practical Approach Vol.1, pp. 49-78. IRL Press, Oxford, England.
10. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, p.202-203. Cold Spring Harbor Laboratory Press, New York.
11. Von Heijine, H.G. (1986) Nucl. Acids Res. 14, 4683-4690.
12. Shorrosh, B.S., Wen, L., Zen, K., Huang, J., Pan, J.S., Hermodson, M.A., Tanaka, K., Muthukrishnan, M., and Reeck, G.R. (1992) Plant Mol. Biol. 18, 151-154.
13. Sugiyama, T., Rafalski, A., Peterson, D., and Soll, D. (1987) Nucl. Acids Res. 13, 8729-8739.
14. Anderson, O.D., Greene, F.C., Yip, R.E., Halford, N.G., Shewry, P.R., and Malpica- Romero, J.M. (1989) Nucl. Acids Res. 17:461-462.
15. Okita, T.W., Krishnan, H.B., and Kim, W.T. (1988) Plant Science 57, 103-111.
16. Okita, T.W., Hwang, Y.S., Hnilo, J., Kim, W.T., Aryan, A.P., Larson, R., and Krishnan, H.B. (1989) J. Biol. Chem. 264, 12573-12581.
17. Chrispeels, M.J. (1991) Annu. Rev. Plant Physiol. Plant Mol. Biol. 42, 21-53.