Volume 17 Number 22 1989 Nucleic Acids Research

Nucleotide sequence of a cDNA that encodes a rice glutelin

Lisa Wen, Jenq-Kuen Huang, Barbara H.Johnson and Gerald R.Reeck

Department of Biochemistry, Kansas State University, Manhattan, KS 66506, USA
Submitted August 16, 1989 EMBL accession no. X15833

Glutelin is a major storage protein of rice endosperm. We (1) and others (2-4) have previously reported heterogeneity in rice glutelin genes. We reported here the nucleotide sequence of a cDNA encoding a previously undescribed glutelin polypeptide. The cDNA was isolated from a λgt10 cDNA library prepared from immature rice endosperm RNA and kindly provided by Dr. Susan Wessler from University of Georgia. The nucleotide sequence (1647 bp) and the inferred amino acid sequence (499 residues) are shown below. The amino acid sequence is identical 63% and 80% of the alignment positions when compared to the sequences encoded by clone pREE61 (2,3) and clone λRG21 (4), respectively. The arrow between amino acids 24 and 25 indicates the predicted signal peptide cleavage site. The arrow between amino acids 303 and 304 indicates the proteolytic cleavage site between the acidic and basic subunits. Putative polyadenylation signal sequences are underlined.

```
                                                                                GAATTCCCCATAAGCAAGTACAAATAGCT   29
ATGGCGAGTTCCGTTTTCTCTCGGTTTTCTATATACTTTTGTGTTCTTCTATTATGCCATGGTTCTATGGCCCAGCTATTTAATCCCAGCACAAACCCATGGCATAGTCCTCGGCAAGGA  149
M  A  S  S  V  F  S  R  F  S  I  Y  F  C  V  L  L  L  C  H  G  S  M  A ↑Q  L  F  N  P  S  T  N  P  W  H  S  P  R  Q  G    40
AGTTTTAGGGAGTGTAGATTTGATAGACTACAAGCATTTGAACCACTTCGGAAAGTGAGGTCAGAAGCTGGGGTGACTGAGTACTTCGATGAGAAGAATGAATTATTCCAGTGCACGGGT  269
S  F  R  E  C  R  F  D  R  L  Q  A  F  E  P  L  R  K  V  R  S  E  A  G  V  T  E  Y  F  D  E  K  N  E  L  F  Q  C  T  G    80
ACTTTTGTGATCCGACGTGTCATTCAGCCTCAAGGCCTTTTGGTACCTCGATACACAAATATTCCTGGCGTGGTCTACATCATCCAAGGGAGAGGTTCTATGGGTTTAACCTTCCCCGGT  389
T  F  V  I  R  R  V  I  Q  P  Q  G  L  L  V  P  R  Y  T  N  I  P  G  V  V  Y  I  I  Q  G  R  G  S  M  G  L  T  F  P  G   120
TGCCCTGCGACTTACCAGCAACAATTCCAACAATTTTCATCTCAAGGCCAAAGTCAGAGCCAAAAGTTTAGAGATGAGCACCAAAAGATTCATCAATTTAGGCAAGGAGACATTGTTGCA  509
C  P  A  T  Y  Q  Q  Q  F  Q  Q  F  S  S  Q  G  Q  S  Q  S  Q  K  F  R  D  E  H  Q  K  I  H  Q  F  R  Q  G  D  I  V  A   160
CTCCCAGCTGGTGTTGCACATTGGTTCTACAATGATGGTGATGCACCTATTGTTGCCGTATATGTTTATGACGTAAACAACAACGCCAATCAGCTTGAACCTAGGCAAAAGGAGTTCCTA  629
L  P  A  G  V  A  H  W  F  Y  N  D  G  D  A  P  I  V  A  V  Y  V  Y  D  V  N  N  N  A  N  Q  L  E  P  R  Q  K  E  F  L   200
TTAGCCGGCAACAACAATCGGGCTCAACAACAACAAGTATATGGTAGCTCAATTGAGCAACACTCTGGGCAAAACATATTCAGCGGATTTGGTGTTGAGATGCTAAGTGAGGCTTTAGGC  749
L  A  G  N  N  N  R  A  Q  Q  Q  Q  V  Y  G  S  S  I  E  Q  H  S  G  Q  N  I  F  S  G  F  G  V  E  M  L  S  E  A  L  G   240
ATCAACGCAGTAGCAGCAAAGAGGCTACAGAGCCAAAATGATCAAAGAGGAGAGATCATACATGTGAAGAATGGCCTTCAATTGTTGAAACCGACTTTGACACAACAGCAAGAACAAGCA  869
I  N  A  V  A  A  K  R  L  Q  S  Q  N  D  Q  R  G  E  I  I  H  V  K  N  G  L  Q  L  L  K  P  T  L  T  Q  Q  Q  E  Q  A   280
CAAGCACAAGATCAATATCAACAAGTTCAATACAGTGAACGACAGCAAACATCTTCTCGATGGAACGGATTGGAGGAGAACTTTTGCACGATCAAGGTGAGAGTAAACATTGAAAATCCT  989
Q  A  Q  D  Q  Y  Q  Q  V  Q  Y  S  E  R  Q  Q  T  S  S  R  W  N  G ↑L  E  E  N  F  C  T  I  K  V  R  V  N  I  E  N  P   320
AGTCGTGCTGATTCATACAACCCACGTGCCGGAAGGATAACAAGTGTCAATAGTCAGAAGTTCCCCATCCTTAACCTCATCCAAATGAGCGCTACCAGAGTAAACCTATACCAGAATGCT 1109
S  R  A  D  S  Y  N  P  R  A  G  R  I  T  S  V  N  S  Q  K  F  P  I  L  N  L  I  Q  M  S  A  T  R  V  N  L  Y  Q  N  A   360
ATTCTCTCGCCGTTCTGGAACGTCAATGCTCATAGTTTGGTCTATATGATTCAAGGGCGATCTCGAGTTCAAGTCGTTAGTAACTTTGGAAAGACTGTGTTTGATGGTGTCCTTCGCCCA 1229
I  L  S  P  F  W  N  V  N  A  H  S  L  V  Y  M  I  Q  G  R  S  R  V  Q  V  V  S  N  F  G  K  T  V  F  D  G  V  L  R  P   400
GCACAATTATTGATCATTCCGCAACATTATGCTGTCTTGAAGAAAGCAGAGCGTGAAGGATGCCAATATATCGCAATCAAGACAAACGCTAACGCCTTCGTCAGCCACCTTGCAGGGAAA 1349
A  Q  L  L  I  I  P  Q  H  Y  A  V  L  K  K  A  E  R  E  G  C  Q  Y  I  A  I  K  T  N  A  N  A  F  V  S  H  L  A  G  K   440
AACTCAGTATTCCGTGCCTTGCCAGTTGATGTAGTCGCTAATGCGTATCGCATCTCAAGGGAGCAAGCCCGAAGCCTCAAGAACAACAGGGGAGAAGAGCACGGTGCCTTCACTCCTAGA 1469
N  S  V  F  R  A  L  P  V  D  V  V  A  N  A  Y  R  I  S  R  E  Q  A  R  S  L  K  N  N  R  G  E  E  H  G  A  F  T  P  R   480
TTTCAACAACAATACTACCCAGGATTATCGAATGAGTCCGAAAGCGAGACCTCAGAGTAATGTAATTGAGAACTAGTATCGGCGTAGAGTAAAATAAAACACCACAAGTATGACACTTGG 1589
F  Q  Q  Q  Y  Y  P  G  L  S  N  E  S  E  S  E  T  S  E  <     499
TGGTGATTCTGTTCGATATCAGTACTAAATAAAGGTTACAAACTTCTTAAAAAAAAAA  1647
```

Acknowledgment: This work was supported by the Kansas Agricultural Experiment Station (publication number 90-46-J) and by a research grant from The Rockefeller Foundation to GRR.

References:

1. Wang, C-S., Shastri, K., Wen, L., Huang, J-K., Sonthayanon, B., Muthukrishnan, S., and Reeck, G.R. (1987) FEBS Lett. **222**, 135-138.
2. Takaiwa, F., Kikuchi, S., Oono, K. (1987) Mol. Gen. Genet. **208**, 15-22.
3. Takaiwa, F., Kikuchi, S., Oono, K. (1986) FEBS Lett **206**, 33-35.
4. Masumara, T., Kidzu, K., Sugiyama, Y., Mitsukawa, N., Hibino, K., Tanaka, K., and Fujii, S. (1989) Plant Mol. Biol. in press.