The structure of rice storage protein glutelin precursor deduced from cDNA

Fumio Takaiwa, Shoshi Kikuchi and Kiyoharu Oono

Department of Cell Biology, National Institute of Agrobiological Resources, Yatabe, Ibaraki 305, Japan

Received 25 July 1986

The complete amino acid sequence of rice storage protein glutelin was determined by the sequencing of the corresponding cDNA. The deduced glutelin precursor has a 37 amino acid signal peptide sequence at the NH_2 terminus, which is followed by a 269 amino acid acidic subunit (M_r = 32 489) and a 193 amino acid basic subunit (M_r = 19 587). The glutelin precursor sequence is homologous to those of pea legumin and soybean glycinin.

(Oriza sativa) *Storage protein* *Glutelin* *cDNA* *Glutelin precursor*

1. INTRODUCTION

Glutelin is the major seed storage protein of rice, which constitutes about 80% of the total protein in the starchy endosperm. This protein is composed of two disulfide-linked acidic and basic subunits with molecular masses of 19–22 and 30–36 kDa [1–4]. Glutelin has been shown to be synthesized as a large precursor (57 kDa), which is subsequently processed into two subunits by post-translational cleavage [1]. In this paper, the nucleotide sequence of glutelin cDNA was determined, from which the complete amino acid sequence of the glutelin precursor was deduced. When compared with the amino acid sequences of pea legumin [5] and soybean glycinin [6], we found that rice glutelin shows homology to these leguminous 11 S globulins. This is the first complete nucleotide sequence of the rice storage protein gene.

2. MATERIALS AND METHODS

Poly(A) mRNA was prepared from the immature rice endosperm (*Oriza sativa* L. cv. Mangetsumochi) 4–10 days after flowering. Complementary DNA was constructed from this poly(A) mRNA by the method of Gubler and Hoffman [7] and ligated to *Eco*RI linker molecules at the 5′- and 3′-termini. This was inserted into an *Eco*RI site of pBR325 and used for transformation of *E. coli* RR1 cells. Candidate clones were selected by colony hybridization and hybrid-released translation, which were carried out as described [8]. DNA sequence of the obtained glutelin pREE61 cDNA clone was determined by the chemical method [9] and analyzed using the GENETYX program (Software Development Co., Tokyo).

3. RESULTS AND DISCUSSION

One of the cDNA clones which encodes the glutelin sequence, pREE61, had an insert of 1627 nucleotides and a single open reading frame for 499 codons was inferred between positions 1 and 1497 of the DNA sequence (fig.1). This deduced amino acid sequence has a molecular mass of 56230 Da, which is in good agreement with that of the in vitro translation product (57 kDa) of rice glutelin mRNA [1]. This suggests that the pREE61 cDNA codes for the entire glutelin precursor se-

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/86/$3.50

-1
CAACAAC

1 100
ATGGCATCCATAAATCGCCCCATAGTTTTCTTCACAGTTTGCTTGTTCCTCTTGTGCAATGGCTCTCTAGCCCAGCAGCTATTAGGCCAGAGCACTAGTC
M A S I N R P I V F F T V C L F L L C N G S L A Q Q L L G Q S T S

200
AATGGCAGAGTTCTCGTCGTGGAAGTCCAAGAGAATGCAGGTTCGATAGGTTGCAAGCATTTGAGCCAATTCGGAGTGTGAGGTCTCAAGCTGGCACAAC
Q W Q S S R R G S P R E C R F D R L Q A F E P I R S V R S Q A G T T
50
300
TGAGTTCTTCGATGTCTCTAATGAGCAATTTCAATGTACCGGAGTATCTGTTGTCCGTCGAGTTATTGAACCTAGAGGCCTTCTACTACCCCATTACACT
E F F D V S N E Q F Q C T G V S V V R R V I E P R G L L L P H Y T
100
400
AATGGTGCATCTCTAGTATATATCATCCAAGGGAGAGGTATAACAGGGCCAACTTTCCCAGGCTGTCCTGAGTCCTACCAACAACAGTTCCAACAATCAG
N G A S L V Y I I Q G R G I T G P T F P G C P E S Y Q Q Q F Q Q S

500
GCCAAGCCCAATTGACCGAAAGTCAAAGCCAAAGTCAAAAGTTCAAGGATGAACATCAAAAGATCCACCGTTTCAGACAAGGAGATGTAATTGCATTGCC
G Q A Q L T E S Q S Q S Q K F K D E H Q K I H R F R Q G D V I A L P
150
600
TGCTGGTGTAGCTCATTGGTGCTACAATGATGGTGAAGTGCCAGTTGTTGCCATATATGTCACTGATCTCAACAACGGTGCTAATCAACTTAGCCCTAGG
A G V A H W C Y N D G E V P V V A I Y V T D L N N G A N Q L D P R
200
700
CAAAGGGATTTCTTGTTAGCTGGAAATAAGAGAAACCCTCAAGCATACAGGCGTGAGGTTGAGGAGCGGTCACAGAACATATTTAGTGGCTTTAGCACTG
Q R D F L L A G N K R N P Q A Y R R E V E E R S Q N I F S G F S T

800
AACTACTTAGCGAGGCTCTTGGCGTAAGCGGCCAAGTGGCAAGGCAGCTCCAATGTCAAAATGACCAAAGAGGAGAAATTGTCCGTGTCGAACACGGGCT
E L L S E A L G V S G Q V A R Q L Q C Q N D Q R G E I V R V E H G L
250
900
CAGTTTGCTGCAGCCATATCCATCATTGCAGGAGCAGGAACAAGGACAAGTGCAATCAAGAGAGCGTTATCAAGAAGGACAATATCAGCAAAGTCAATAT
S L L Q P Y A S L Q E Q E Q G Q V Q S R E R Y Q E G Q Y Q Q S Q Y
300
1000
GGAAGTGGCTGCTCTAACGGTTTGGATGAGACCTTTTGCACCCTGAGGGTAAGGCAAAACATCGATAATCCTAACCGTGCTGATACATACAATCCAAGAG
G S G C S N G L D E T F C T L R V R Q N I D N P N R A D T Y N P R

1100
CTGGAAGGGTTACAAATCTCAACACCCAGAATTTCCCCATTCTCAGTCTTGTACAGATGAGTGCAGTCAAAGTAAATCTATACCAGAATGCACTCCTTTC
A G R V T N L N T Q N F P I L S L I Q M S A V K V N L Y Q N A L L S
350
1200
ACCATTTTGGAACATCAACGCTCACAGCGTCGTGTATATTACTCAAGGCCGTGCCCGGGTTCAAGTTGTCAACAACAATGGAAAGACAGTGTTCAACGGC
P F W N I N A H S V V Y I T Q G R A R V Q V V N N N G K T V F N G
400
1300
GAGCTTCGCCGCGGACAGCTGCTTATTATACCACAACACTACGCAGTTGTAAAGAAGGCACAAAGAGAAGGATGTGCTTACATTGCATTCAAGACCAATC
E L R R G Q L L I I P Q H Y A V V K K A Q R E G C A Y I A F K T N

1400
CTAACTCTATGGTAAGCCACATTGCAGGAAAGAGTTCCATCTTCCGTGCTCTCCCAAATGATGTTCTAGCAAATGCATATCGCATCTCAAGAGAAGAGGC
P N S M V S H I A G K S S I F R A L P N D V L A N A Y R I S R E E A
450
1500
TCAGAGGCTCAAGCATAATAGAGGAGATGAGTTCGGTGCATTCACTCCAATCCAATACAAGAGCTACCAAGACGTTTATAATGCGGCAGAATCCTCTTAG
Q R L K H N R G D E F G A F T P I Q Y K S Y Q D V Y N A A E S S *

1600
GTCGGCTTGCGGATAAAGAATAACTAAATAAATAAATTGCAAGCAATTGTTTTGCTGCTATGTACTGTCCAGTCTTTCGACTAATGATGATAAAGCCTCT

CTTTATCCTTAAAAAAAAA

quence. There is a hydrophobic region characteristic of a signal peptide at the NH_2 terminal region. We assigned the cleavage site for the signal peptide between the 37th and 38th serine according to a common format for the cleavage sites [10], because the NH_2 terminal sequence of the mature glutelin acidic subunit has not yet determined. On the other hand, the position of the NH_2 terminus of the mature basic subunit was identified as the 307th glycine by comparison with the partial terminal sequence of the corresponding peptide [2]. Therefore, the coding region for mature acidic and basic subunits could be defined between 38th serine and 306th asparagine (269 amino acids) and between 307th glycine and 499th serine (193 amino acids), respectively. The molecular masses of both mature subunits, calculated from the deduced amino acid sequences, are 32489 and 19587 Da, which agrees well with those estimated from SDS-polyacrylamide gel. The predicted amino acid compositions also show similarity with the values obtained from these purified subunits [4].

The 3′-nontranslated region of pREE61 cDNA is 110 nucleotides long. Two putative polyadenylation signals (AATAAATAAA and GATAAA) are observed 27 and 89 nucleotides downstream from the TAG stop codon. The presence of two or three polyadenylation signals is a general feature of plant mRNA [11].

Rice glutelin shares several characters in common with leguminous 11 S globulin. When the deduced amino acid sequence of the glutelin precursor was compared with those of pea legumin [5] and soybean glycinin [6], it showed 38 and 37% homology to them. The sequence of the acidic subunit is less conserved than that of the basic subunit. This is mainly due to the remarkable divergence found in the COOH terminal region of the former. The cleavage site between the acidic and basic subunits of 11 S globulin is known to exist between asparagine and glycine residues. Such a sequence is also strictly conserved in rice glutelin. The positions of cysteine residues which are involved in linking acidic and basic subunits have been identified in soybean glycinin [12]. Rice glutelin has cysteine residues at the homologous positions (i.e. 112th cystine and 313th cystine), whereas most of the other cysteine residues found in rice glutelin diverged in soybean glycinin. These findings suggest that rice glutelin and leguminous 11 S globulin have evolved from a common ancestral gene.

ACKNOWLEDGEMENT

This work was supported in part by a project grant from the Ministry of Agriculture, Forestry and Fisheries (Japan).

REFERENCES

[1] Yamagata, H., Sugimoto, T., Tanaka, K. and Kasai, Z. (1982) Plant Physiol. 70, 1094–1100.
[2] Zhao, W.-M., Gatehouse, J.A. and Boulter, D. (1983) FEBS Lett. 162, 96–102.
[3] Robert, L.S., Nozzolillo, C. and Altosaar, I. (1985) Biochim. Biophys. Acta 829, 19–26.
[4] Wen, T.-N. and Luthe, D.S. (1985) Plant Physiol. 78, 172–177.
[5] Lycett, G.W., Croy, R.R.D., Shirsat, A.H. and Boulter, D. (1984) Nucleic Acids Res. 12, 4493–4506.
[6] Negoro, T., Momma, T. and Fukazawa, C. (1985) Nucleic Acids Res. 13, 6719–6731.
[7] Gubler, U. and Hoffman, B.J. (1983) Gene 25, 263–269.
[8] Maniatis, T., Fritsh, E.F. and Sambrook, J. (1982) in: Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
[9] Maxam, A. and Gilbert, W. (1980) Methods Enzymol. 65, 449–559.
[10] Von Heijne, G. (1983) Eur. J. Biochem. 133, 17–21.
[11] Messing, J., Geraghty, D., Heidecker, G., Hu, N.T., Kridl, J. and Rubenstein, I. (1983) in: Genetic Engineering of Plants (Kosuge, T. et al. eds) pp.211–227, Plenum press, New York.
[12] Staswick, P.E., Hermodson, M.A. and Nielsen, N.C. (1984) J. Biol. Chem. 259, 13431–13435.

Fig.1. Complete nucleotide sequence of pREE61 cDNA insert and deduced amino acid sequence. The arrow between the 37th and 38th serine denotes the predicted signal peptide cleavage site. The arrow between the 306th asparagine and 307th glycine indicates the proteolytic cleavage site between the acidic and basic subunits. Putative polyadenylation signal sequences are underlined. Cysteine residues which are involved in the disulfide bond between acidic and basic subunits are circled.