Organization and Primary Sequence of Multiple Genes Encoding Type II mRNA Species of Rice Prepro-glutelin

Keiko Abe, Yasufumi Emori,* Hiroshi Kawasaki,* Hiroto Kondo, Koichi Suzuki* and Soichi Arai

Department of Agricultural Chemistry, The University of Tokyo, Yayoi-cho, Bunkyo-ku, Tokyo 113, Japan
*Department of Molecular Biology, The Tokyo Metropolitan Institute of Medical Science, Honkomagome, Bunkyo-ku, Tokyo 113, Japan

Received June 7, 1989

We have isolated four independent genomic DNA fragments encoding a rice storage protein, glutelin, by using its cDNA as a probe. Restriction mapping and sequencing analyses showed that all four clones encode type II mRNA of glutelin and have very similar nucleotide sequences. However, several differences were found among the four clones with respect to the nucleotide sequences of their coding and 5' upstream regions, some of which confirmed corresponding changes in the amino acid sequences of the encoded proteins. These results indicate that the genes encoding type II mRNA species of glutelin constitute a multi-gene family of closely related members.

Glutelin is a major storage protein of rice, Oryza sativa L.1) Rice glutelin, as is the case with other storage proteins such as soybean glycinin,2) pea legumin,3) and rape crucifelin,4) is synthesized as a pre-protein which then undergoes post-translational processing. This leads to a mature protein with an acidic subunit-basic subunit association linked by disulfide bonding.5)

Three types of cDNA clones encoding prepro-glutelin have been isolated.6,7) Recently, a genomic DNA segment corresponding to one type of cDNA (type I) was cloned and sequenced.8) In this study, we isolated four independent genomic clones encoding type II mRNA of rice glutelin; these differ in some respects from the previously cloned gene.9) Here we report the results of comparative sequencing studies on our four genomic clones that demonstrate that the type II mRNA species of rice glutelin is encoded by multiple genes.

Materials and Methods

Materials. A middle ripening variety of mature rice, the cultivar Nipponbare, was used. A cDNA clone encoding rice glutelin, AOG12, was previously isolated.9)

Isolation of genomic clones. A rice genomic library was constructed as described previously.10) High-molecular-weight rice DNA was digested with the restriction enzyme Sau 3A1 and fractionated by centrifugation on a 10~40% linear sucrose gradient. DNA ranging from 15~20 kbp was ligated to a λEMBL3 vector and packaged using a commercial packaging kit (Giga pack, U.S.A.). A library of approximately 2 x 106 clones was constructed for screening with a cDNA probe of type I glutelin mRNA (AOG12).9)

Nucleotide sequence analysis. DNA fragments generated by digestion with restriction enzymes were subcloned into the plasmid vector pUC18 and sequenced by the chain termination method.11) As the primers for determination of the sequences around exon-intron junctions, the eight oligonucleotides, 5'-GATAGGTTGCAAGCATT-3' (145 to 161), 5'-TGTGCCAGCTTGAGACC-3' (198 to 182), 5'-AAGTTCAAGGATGAACA-3' (528 to 544), 5'-TAACATCTCCTTGTCG-3' (579 to 563), 5'-TCAT-TGCAAGAGCAGGA-3' (1015 to 1031), 5'-TTCTTGATA-TGCTCTC-3' (1068 to 1052), 5'-TGCAGGAAAGATTCCA-3' (1598 to 1614), 5'-TTTGCTAG-...
K. Abe et al.

AACATCAGT-3' (1649 to 1632), were used. These correspond to the conserved regions of the glutelin cDNAs,6,7 where the numbers in brackets denote the nucleotide numbering (see Fig. 2).

Results and Discussion

Restriction maps and sequencing strategy for the glutelin genes

Four positive clones were finally isolated. Figure 1 shows restriction enzyme maps for the four clones, λOG1, λOG11, λOG15, and λOG17. The lengths of their inserts ranged from 6kbp to 11kbp. Clones λOG1, λOG15, and λOG17 had mutually similar maps, but clone λOG11 was quite different from the others in the 5'-upstream region.

Nucleotide and deduced amino acid sequences for glutelin genes

Of the four clones, λOG15 was selected for

---

**Fig. 1.** Restriction Maps and Sequencing Strategy for the Glutelin Genes.

Restriction sites of the glutelin genes, λOG1, λOG11, λOG15, and λOG17, are shown with the symbols: K, KpnI; S, StuI; A, Accl; P, PstI; E, EcoRI; EV, EcoRV; Sm, SmaI; H, HindIII; X, XbaI; and B, BamHI. Synthesized oligonucleotides (black bars) were used as primers for sequencing by the dideoxy method (see Methods). The direction and extent of sequencing are shown by arrows. The black boxes each show an exon of glutelin.

**Fig. 2.** Nucleotide and Deduced Amino Acid Sequences for Glutelin Gene.

Column A, B, C, and D showed the sequences of λOG15, λOG1, λOG11, and λOG17, respectively. Nucleotides in λOG15 are numbered to start at the A nucleotide of the presumptive translation initiation codon. The amino acid sequences are shown above the nucleotide sequences. The sequences of the introns are represented by small letters. Consensus sequences, such as the TATA box, CAAT box, poly A addition signal sequence (AATAAA), and poly A site consensus sequences (TGTGTTC and AATTG) common in higher plant genes12,17 are underlined. The stop codon, TAA, is also underlined. The repeated sequences discussed in the text are indicated: CTCTGTTTTG by thick lines, CATAATGCAAA by double lines, TTGGAAAG by horizontal arrows, AACCTA by boxes, CAAAAAGAGGAG by dashed lines, and CCTTTCGTGT by wavy lines. In the sequences of λOG1, λOG11, and λOG17, only different nucleotides are shown, dots standing for nucleotides that are same as in λOG15. Changed amino acids are shown in brackets below the sequences. The asterisks indicate the different sequences found when λOG15 is compared with the gene isolated by Takaiwa et al8: *1, T→C at -290; *2, T→TAA at -136; *3, T→C at -124; *4 T→C at 722; *5, C→G at 1217 (T→S142); *6, G→A at 2147.
detailed analysis. The 5.1-kb sequencing done by the strategy shown in Fig. 1. Covers the entire transcribed region plus the 5'- and 3'-untranscribed regions. By comparing the nucleotide sequence of λOOG15 with that of the cDNA, three exons were identified (Fig. 1), and λOOG15 was believed to encode a type II mRNA species of glutelin.

Comparison of the results (Fig. 1) with the consensus sequences of the 5'-upstream regions of plant genes showed that the TATA box and the CAAT sequences in λOOG15 are at positions 67 bases upstream and 202 bases upstream from the initiation methionine (ATG), respectively. Interestingly, this 5'-upstream region is also characterized by the presence of a variety of repeated sequences, such as CTCTGTTTTG (−2171 to −2162), CATAATGCAAA (−2249 to −2239, −570 to −560, −344 to −334 and −245 to −235), CAAAAAGAGGAG (−337 to −326 and −276 to −265) and CTTTTCGTGT (−286 to −277 and −164 to −155). In particular, the sequence AACCTA is repeated five times between positions −1037 and −446. It should be noted that a similar sequence (AGCCCA) occurs in the 5'-upstream region of the soybean β-conglycinin α'-subunit gene and is considered to be important in the regulation of tissue-specific expression of the gene. A similar sequence resides in the 5'-upstream region of the phaseolin β-subunit gene. Thus, the sequence AACCCA or its analogue may be a cis-acting sequence targeted by a tissue-specific transcription-activating factor in plants.

Furthermore, the sequence TTGGAAAG (−1593, −978 and −911) is homologous not only to the animal core enhancer sequence GTGGAAGA, but also to a higher plant gene promoter sequence which may be involved in light regulation during photosynthesis. These features in the 5'-upstream region of the rice glutelin gene might contribute to stage- and tissue-specific expression of the mRNA, the amount of which reaches a maximum three weeks after flowering in the rice plant we investigated.

Next, we sequenced the other three clones using cDNA-specific primers (see Materials and Methods, and Fig. 1). The sequencing analyses showed that all three DNA sequences are very similar to that of λOOG15 and encode type II mRNA species. For these clones, however, the following nucleotide and amino acid changes were observed: G→T at 629, A→T at 1151 (N320→I320), and A→T at 1263 for λOOG1 (symbol B in Fig. 2); T→C at −290, A→T, at 1083 and C→G at 1748 for λOOG17 (symbol D in Fig. 2); and A→T at 1151 (N320→I320) and A→C at 1636 (D354→A354) for λOOG11 (symbol C in Fig. 2). λOOG11 was totally different from the others in the nucleotide sequence of the region upstream from position −2009 (Fig. 2).

Takaiwa et al. isolated and sequenced a rice glutelin gene encoding type II mRNA. However, all of our four clones differ from their gene in several respects. Compared to their gene, λOOG15 has three different nucleotides in the 5'-upstream region, two different nucleotides in the coding region, and one different nucleotide in the 3'-downstream region (Fig. 2). The differences in the coding region result in the altered amino acid sequence of the encoded peptide (Thr342→Ser342).

On the basis of these data, we conclude that the five rice glutelin genes isolated by us and by Takaiwa et al. are distinct from each other and must be at different chromosomal positions because of the differences in their nucleotide sequences, as described above, although the possibility remains that some of these differences can be ascribed to allelic polymorphism. However, only type II clones were obtained from the library by using type I cDNA probe, and type I and type III genomic clones were not isolated as yet, although they are expressed to be found in the cDNA library. This suggests that the copy numbers of type II genes are larger than those of...
the others and that type II genomic fragments were easier to isolate than type I or III. Type II mRNAs encoding the rice prepro-glutelin may thus be produced by such multiple genes of closely related members.

References