Plant Molecular Biology **18**: 617–619, 1992.

Update section

Sequence

The hydroxyproline-rich glycoprotein gene from *Oryza sativa*

Carme Caelles [1], Michel Delseny [2] and Pere Puigdomènech [1]*
[1] *Departamento de Genética Molecular, CID-CSIC, Jordi Girona 18, 08034 Barcelona, Spain (*author for correspondence);* [2] *Laboratoire de Physiologie et Biologie Moléculaire Végétale, URA 565 du CNRS, Université de Perpignan, 66860 Perpignan, France*

Received 4 October 1991; accepted 9 October 1991

Key words: rice, hydroxyproline-rich glycoprotein, Gramineae

Hydroxyproline-rich glycoproteins (HRGP) are among the major extractable components of plant cell walls ([1] for review). HRGPs from dicotyledonous species (also called extensins) are polypeptides formed by highly repetitive motifs including a typical Ser-$HPro_4$ repeat. Dicot extensins seem to be coded by multigene families expressing a variety of mRNAs [2]. It has been shown that these genes are developmentally regulated [3] and some of them can be induced by wounding and by fungal infection [4].

Most information on analogous proteins from monocotyledonous plants comes from maize and closely related species. The maize HRGP has been purified and partially sequenced [6]. The complete protein sequence has been obtained by cloning the corresponding gene from both maize [7] and sorghum [8]. Both proteins contain highly repeated motifs rich in proline but also in threonine. In contrast to dicotyledonous species, cereal HRGPs are coded by simple gene families, probably single genes [7, 8]. The maize HRGP mRNA accumulates in organs rich in dividing cells, it is induced by wounding [9] and it shows a specific pattern of expression in immature embryos [10].

In Southern hybridization of genomic rice DNA with either maize or sorghum HRGP probes a limited number of bands appeared [8]. Due to the interest of this species in terms of genomic simplicity, ability to be transformed, economic interest and taxonomic distance to the other cereals already studied, it was of interest to clone and characterize the corresponding HRGP gene. To this end, a genomic library from *Oryza sativa* (IR36 Indica variety) (see [11] for details of the library) was screened using a 842 bp restriction fragment from the 5′ and coding region of the maize HRGP gene [7] as a probe. A single positive clone was chosen for further analysis. A region of this genomic clone containing the complete transcription unit as well as 5′ and 3′ flanking regions was sequenced in both strands by the Exo III deletion method [12]. The sequence of the gene and that of the deduced protein sequence are shown in Fig. 1.

The general structure of the rice HRGP gene is similar to that of the maize and sorghum genes

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession number X61280

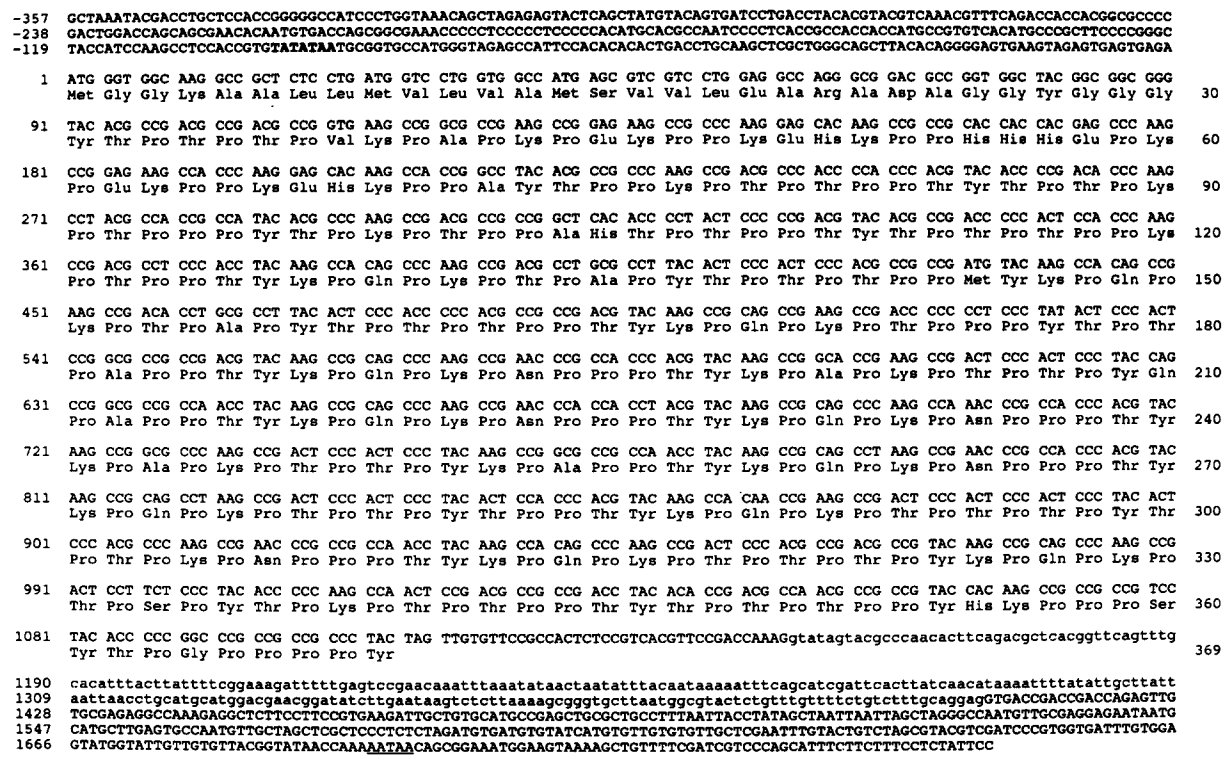

Fig. 1. Nucleotide and deduced protein sequence of the rice gene coding for a hydroxyproline-rich glycoprotein (HRGP). The putative TATA box and polyadenylation signal are underlined and a sequence having the features observed in other HRGP genes is indicated with lower case letters. Nucleotides are numbered from the A of the translation start site.

previously reported [7, 8]. In particular, it also includes a putative intron (shown in lower-case letters in the figure) in the 3′ untranslated region, a feature common to HRGP genes from both monocot and dicot species. In the figure the presence of putative TATA box and polyadenylation signals are indicated. Nevertheless, when the 5′ and 3′ flanking regions are compared to the other published sequences no obvious alignment can be achieved.

The protein coded by the rice gene shares a number of features with those of the other monocots already reported [7, 8]. It begins with a typical signal peptide. The sequence Ala-Asp, where the peptide is cleaved in the maize protein, is also present between residues 19 and 20 in the rice one. The mature sequence of the protein begins with a stretch rich in glycine and tyrosine residues followed by sequences having a highly hydrophilic character and containing proline, lysine, glutamic acid and histidine residues. This region is formed by four hexapeptides repeated two by two and a pentapeptide that contains a His-His-His tripeptide. After this region the main repetitive proline-rich sequence begins. This part of the protein is formed by two alternating proline-rich motifs. One of these motifs (PPTYKP) is very similar to the main repetitive motif from maize (PPTYTP), which, in fact, is present three times in the rice protein. The other one (QPKPT/NP) is also similar to two repeated pentapeptides (SPKPP and TPKPT) forming the maize sequence.

Some interesting features make the rice HRGP sequence different from that of the other related species already known: (1) the almost complete absence of serine residues which are substituted by threonine or glutamine; (2) the disappearance of the single Ser-Pro_4 sequence, the essential el-

ement of dicot extensins, which is present in the other cereals analysed so far. These results indicate that neither serine residues nor the Ser-Pro_4 motif is required for the function of this protein and that threonine and glutamine may substitute serine which is one of the most abundant residues in HRGPs. It is also worth pointing out that, although the protein has a highly repetitive pattern, the motifs are not as perfectly conserved as in the maize protein.

A genomic Southern blot analysis of IR36 rice DNA was performed using a *Spe* I-*Pvu* I probe from the 3′-flanking region of the rice HRGP gene. The results indicate the presence of a single-copy gene which shows the same restriction sites as the isolated genomic clone. A transcript of about 1400 nucleotides is detected in northern blot analysis of young rice leaf RNA (not shown). These results indicate that the gene presented here has a single copy in the rice genome in accordance with the simplicity of this system also observed in other cereals. They also indicate that the gene is expressed as a mature mRNA in developing organs.

Acknowledgements

We thank Dr Steve Kay (Rockefeller University) for kindly providing the genomic rice library. The work was supported by grants from Plan Nacional de Investigación Científica y Técnica (grant BIO88-0242) to P.P. and from The Rockefeller Foundation Rice Biotechnology Network and from CNRS to M.D. C.C. was a recipient of a postdoctoral fellowship from CSIC.

References

1. Cassab GI, Varner JE: Cell wall proteins. Annu Rev Plant Physiol 39: 321–353 (1989).
2. Showalter AM, Zhou J, Rumeau D, Worst SG, Varner JE: Tomato extensin and extensin-like cDNAs: structure and expression in response to wounding. Plant Mol Biol 16: 547–565 (1991).
3. Memelink J, Hoge JHC, Schilperoort RA: Cytokinin stress changes the developmental regulation of several defense-related genes in tobacco. EMBO J 6: 3579–3583 (1987).
4. Showalter AM, Bell JN, Cramer CL, Bailey JA, Varner JE, Lamb CJ: Accumulation of hydroxyproline-rich glycoprotein mRNAs in response to fungal elicitor and infection. Proc Natl Acad Sci USA 82: 6551–6555 (1985).
5. Kieliszewski M, Leykam JF, Lamport DTA: Structure of the threonine-rich extensin from *Zea mays*. Plant Physiol 92: 316–326 (1990).
6. Ludevid MD, Ruiz-Avila L, Vallés MP, Stiefel V, Torrent M, Torné JM, Puigdomènech P: Expression of genes for cell-wall proteins in dividing and wounded tissues of *Zea mays*. Planta 180: 524–529 (1990).
7. Stiefel V, Ruiz-Avila L, Raz R, Vallés MP, Gómez J, Pagès M, Martínez-Izquierdo JA, Ludevid MD, Langdale JA, Nelson T, Puigdomènech P: Expression of a maize cell wall hydroxyproline-rich glycoprotein in early leaf and root vascular differentiation. Plant Cell 2: 785–793 (1990).
8. Raz R, Crétin C, Puigdomènech P, Martínez-Izquierdo JA: The sequence of a hydroxyproline-rich glycoprotein gene from *Sorghum vulgare*. Plant Mol Biol 16: 365–367 (1991).
9. Ludevid MD, Ruiz-Avila L, Vallés MP, Stiefel V, Torrent M, Torné JM, Puigdomènech P: Expression of genes for cell-wall proteins in dividing and wounded tissues of *Zea mays*. Plant 180: 524–529.
10. Ruiz-Avila L, Ludevid MD, Puigdomènech P: Differential expression of a hydroxyproline-rich cell-wall protein gene in embryonic tissues of *Zea mays* L. Planta 184: 130–136 (1991).
11. Kay SA, Keith B, Shinozaki K, Chye ML, Chua NH: The rice phytochrome gene: Structure, autoregulated expression and binding of GT-1 to a conserved site in the 5′ upstream region. Plant Cell 1: 351–360 (1989).
12. Henikoff S: Unidirectional digestion with exonuclease III in DNA sequence analysis. Meth Enzymol 155: 156–165 (1987).