Plant Molecular Biology **16**: 365–367, 1991.

Plant Molecular Biology Update

The sequence of a hydroxyproline-rich glycoprotein gene from *Sorghum vulgare*

Regina Raz[1], Claude Crétin[2], Pere Puigdomènech[1] and José A. Martínez-Izquierdo[1,*]
[1]*Departamento de Genética Molecular, CID-CSIC, Jordi Girona 18, 08034 Barcelona, Spain (*author for correspondence)*; [2]*Laboratoire de Physiologie Végétale Moléculaire, Université de Paris-XI, Orsay, France*

Received 16 October 1990; accepted 30 October 1990

Key words: hydroxyproline-rich glycoprotein, sorghum, Gramineae, 3′ intron

Cell wall hydroxyproline-rich glycoproteins (HRGPs) constitute a family of proteins found as structural components in the plant primary cell walls [1]. These proteins are characterized by the high content in hydroxyproline/proline, basic amino acids, serine and tyrosine [2]. The most studied HRGPs, so far, have been those from dicot species (also called extensins), although a maize HRGP has recently been characterized at the protein and genomic levels [3, 4]. This maize HRGP has the main features of dicot extensins, although its main repeated motif is different from the dicot ones and it is very rich in threonine [3]. In common with the carrot HRGP gene [5] the maize gene contains an intron in the 3′-non-coding region [4], but the GC content of the maize gene is significantly higher than for carrot and tobacco genes [4–6]. In this study we report on the sequence of a HRGP gene from sorghum which resembles maize HRGP gene (60% overall homology), indicating the existence of a Gramineae HRGP subfamily belonging to the plant cell wall HRGP family. Moreover, genomic Southern analysis of different graminaceous DNAs with two distinct DNA genomic probes from maize and sorghum reveal two different classes of similar sequences: the first one common to both probes and the other similar to the sorghum probe, suggesting the existence of another family of HRGP-related genes in Gramineae.

A genomic library of *Sorghum vulgare* made in λ EMBL4 was screened with a maize HRGP genomic probe 842 bp long, containing part of the 5′ and coding regions [4]. λ DNA from one positive clone was restricted with several enzymes and subjected to Southern/hybridization analysis. A 4.5 kbp *Eco*RI/*Sac*I fragment which showed strong hybridization with the maize probe was cloned in M13 vectors and sequenced by the dideoxy method on both strands. Nucleotide sequence analysis showed that such a fragment contains a single long open reading frame that codes for a protein of 283 amino acids. 2 kbp of nucleotide sequence of the gene and the derived amino acid sequence are presented in Fig. 1. There is a potential TATA box 77 bp upstream of the ATG start codon and a putative canonical polyadenylation signal 448 bp downstream the stop codon (underlined). There is a 122 bp long potential intron in the 3′-untranslated region (nucleotides in lower case) in the same position (ca. 25 bp after stop codon) as the maize HRGP one, flanked by AGgt/agGT intron consensus

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession number X56010.

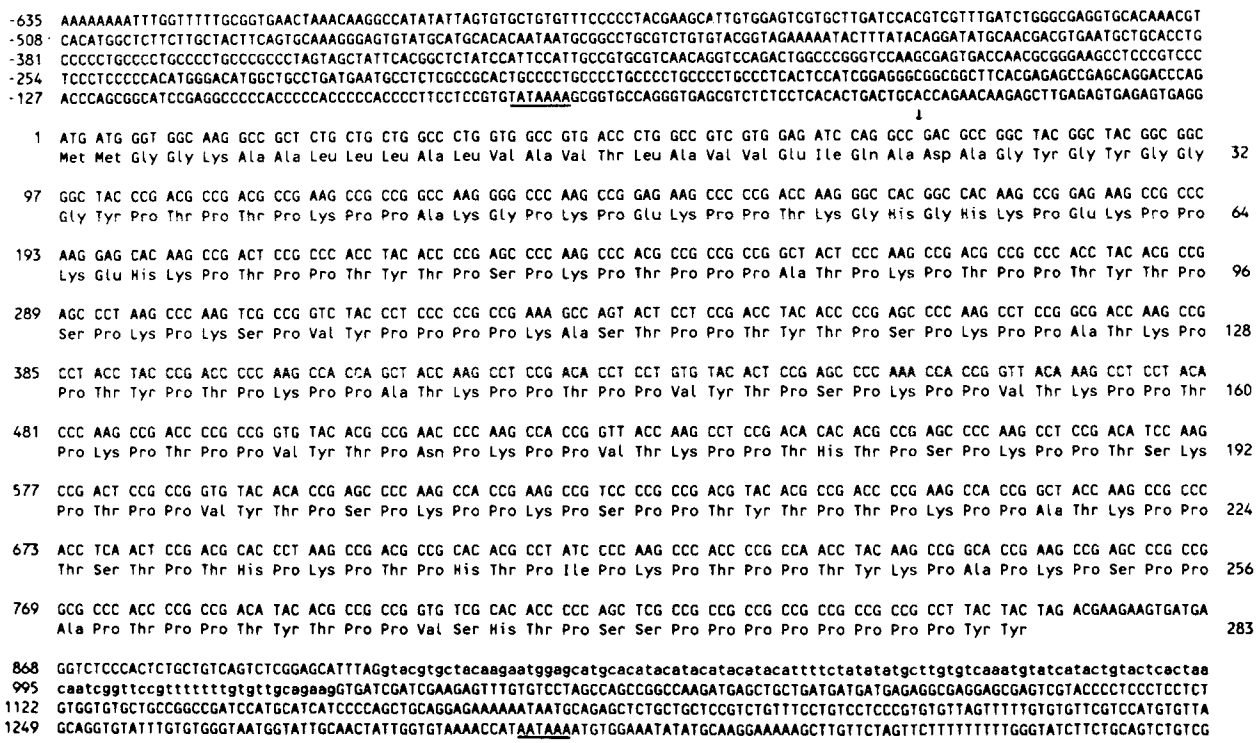

Fig. 1. Nucleotide and derived amino acid sequence of the sorghum HRGP gene. Nucleotides are numbered from the A of the start codon (+1) on the left and amino acids on the right. Putative TATA box and polyadenylation signal are underlined. The intron in the 3′-untranslated region is in lower case. The putative site of processing of the mature protein between Ala 24 and Asp 25 is indicated by a vertical arrow.

sequences. When the different functional regions of the gene, namely 5′, coding, intron and 3′ (intron) regions are compared separately with the corresponding regions of the maize HRGP gene, no significant variation was observed with the overall homology value (60%) except for the 3′ (intron) region with 66% homology. The higher homology observed in that region can be explained by the absence of long insertion/deletions shown in the other three regions.

The amino acid composition of the sorghum HRGP is similar to the maize one: rich in proline/hydroxyproline, lysine, tyrosine and threonine [3, 4], although the threonine content is reduced (16 versus 22%) and the valine content is doubled (3.9 versus 1.8%). In this respect the amino acid composition of sorghum HRGP is closer to carrot HRGP than the maize one [3]. However, the sorghum HRGP should be considered as belonging to the same HRGP family as the maize one, because the main repetitive motifs of the maize HRGP, one undecapeptide and two pentapeptides, are essentially the same in the sorghum HRGP. The loss of some of those motifs results in length reduction of the sorghum HRGP which is 45 amino acid shorter than the maize one [4]. The N-terminal part of the protein, rich in hydrophobic amino acids, has the characteristics of a signal peptide with a potential cleavage site (by comparing with the maize HRGP) between the residues Ala and Asp (marked by a vertical arrow in Fig. 1). Previous results have shown that maize HRGP gene is also present in species of the Panicoideae subfamily of Gramineae [4], as it is the case of the sorghum gene reported here. Figure 2A shows a genomic Southern hybridized with the maize probe described above. It can be seen that related sequences to maize HRGP gene are also present in species of Festucoideae subfamily, such as wheat, barley and rice, although

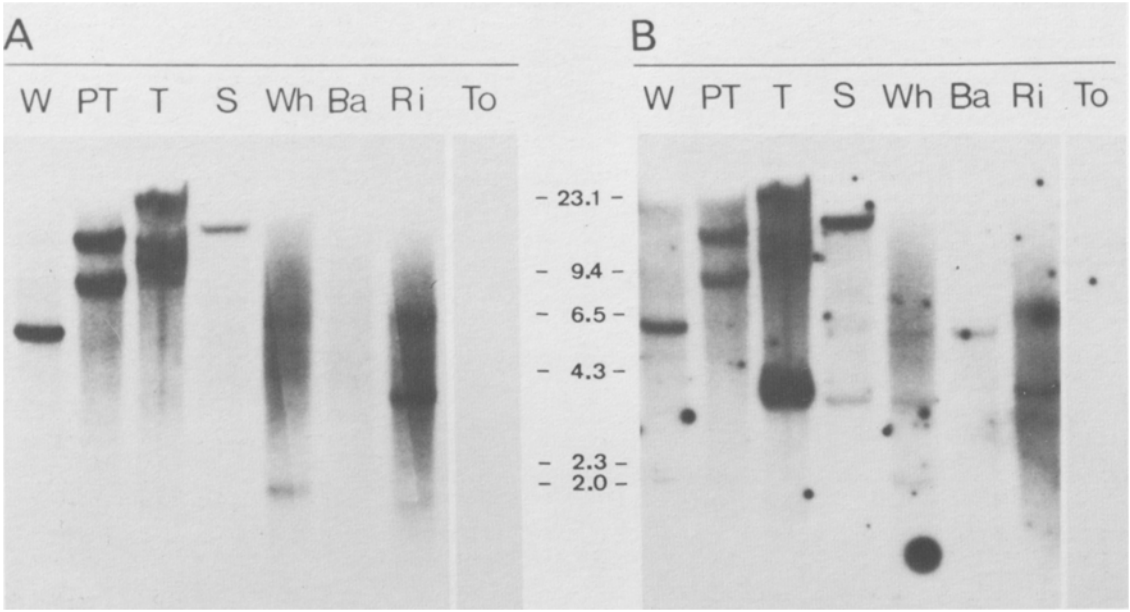

Fig. 2. Genomic Southern blots of several graminaceous DNAs hybridized with HRGP probes. Species analyzed are W64A maize inbred line (W), Palomero Toluqueño maize variety (PT), *Zea diploperennis* teosinte (T), sorghum (S), wheat (Wh), barley (Ba), rice (Ri) and tobacco (To), as external dicot control. DNAs (10 μg per lane) were digested with *Hind* III. The same filter was hybridized with two different probes: the *Hind* II-*Sna* BI 842 bp long fragment from maize inbred line AC1503 (panel A) and the *Alu* I-*Hinf* I 168 bp long fragment from sorghum. The position of size markers (kbp) is indicated.

they are not present in dicot species such as tobacco. There is a region within the sorghum gene which shows low homology to the maize one, compared to the overall homology between the two genes. When a 168 bp long probe (nucleotides 411–579 in Fig. 1) encompassing that region was used, the same bands which appeared with the maize genomic probe, lit up (Fig. 2B). In addition, some other new bands can be clearly seen for the cases of teosinte, sorghum and barley DNAs (Fig. 2B). These results suggest the existence of another HRGP-related gene family in Gramineae. Whether those genes code for proteins belonging to the same family as the dicot and monocot cell wall HRGPs, or to other hydroxyproline-rich glycoproteins [1] must be elucidated.

Acknowledgements

The authors are indebted to Mr. Patricio Gomez for technical assistance. This work has been supported by grants from PLANICYT (BIO 88/0242) and EECC (BAP 374). R.R. is recipient of a fellowship from PFPI.

References

1. Showalter AM, Varner JE: Plant hydroxyproline-rich glycoproteins. In: A. Marcus (ed), The Biochemistry of Plants, Vol. 15A, pp. 485–520. Academic Press, San Diego (1989).
2. Varner JE, Lin L-S: Plant cell wall architecture. Cell 56: 231–239 (1989).
3. Kieliszewski MJ, Leykam JF, Lamport DTA: Structure of the threonine-rich extensin from *Zea mays*. Plant Physiol 92: 231–239 (1990).
4. Stiefel V, Ruiz-Avila L, Raz R, Valles MP, Gomez J, Pages M, Martinez-Izquierdo JA, Ludevid MD, Langdale JA, Nelson T, Puigdomenech P: Expression of a maize cell wall hydroxyproline-rich glycoprotein gene in early leaf and root vascular differentiation. Plant Cell 2: 785–793 (1990).
5. Chen J, Varner JE: An extracellular matrix protein in plants: characterization of a genomic clone for carrot extensin. EMBO J 4: 2145–2151 (1985).
6. Keller B, Lamb CJ: Specific expression of a novel cell wall hydroxyproline-rich glycoprotein gene in lateral root initiation. Genes Dev 3: 1639–1646 (1989).