Plant Molecular Biology **15**: 783–785, 1990.

Plant Molecular Biology Update

Glycine-rich RNA-binding proteins from *Sorghum vulgare*

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

Received 31 July 1990; accepted 13 August 1990

Key words: sorghum, maize, glycine-rich proteins, RNA-binding

Several sequences for plant glycine-rich proteins (GRPs) have been obtained by cDNA or genomic cloning. Some of these proteins are thought to be components of the cell wall [1, 5]. In at least one case it has been shown that they are accumulated in the vascular system [4, 5] and their synthesis is part of defense plant mechanisms [1, 4]. In cereals, the sequence of a glycine-rich protein with homology with vertebrate cytokeratins has recently been reported in barley [7].

A gene coding for a glycine-rich protein whose mRNA is present in dry embryos of maize has been described [3]. The gene shows a basal level of expression in many organs of the plant but it may be induced by ABA and desiccation in young embryos or leaves, respectively [3]. The existence of an RNA-binding consensus in its sequence has been observed [6]. Homologous proteins in another plant species have not yet been described. In *Arabidopsis thaliana* a number of glycine-rich protein cDNAs have been sequenced [2], one of them (atGRP-2) contains a site with an imperfect homology with the RNA-binding consensus. In the present report the sequence of two cDNAs from *Sorghum vulgare* coding for a protein with high similarity with the one described in maize is presented as well as evidence for heterogeneity in the cDNA population.

A cDNA library constructed in lambda gt11 from poly(A)$^+$ RNA extracted from 6-day-old *Sorghum vulgare* seedlings was screened and two positive clones were found. The two clones were sequenced by the dideoxy method in M13mp18 and 19 vectors. The sequence of the two cDNAs is presented in Fig. 1 aligned for maximum homology. The similarity in the 5′ half, corresponding to the coding sequence, is higher than in the 3′ non-translated region, which is very different in the two clones, indicating the existence of at least two different genes coding for this protein in sorghum. The poly(A) tail is added at different positions in the two cDNAs. No perfect consensus polyadenylation site is observed in the two cDNAs; however A-rich regions may be observed in the region − 30 bp from the poly(A) tail in the two cases.

The protein sequences deduced from the two cDNAs may be compared among themselves and to the maize protein [3]. The result of the alignment of the three proteins is shown in Fig. 2. In the figure the sequences are shown emphasizing the existence of the GGYGG element in the C-terminal half of the protein. A high degree of similarity between the three proteins is clearly oberved. The two sorghum proteins have the conserved RNA-binding domains (shown in boxes) and they have the glycine-rich domain formed by GGYGG elements as in the maize protein. The degree of similarity existing between the two sorghum proteins is comparable to that found

The nucleotide sequence data reported have been submitted to the EMBL, GenBank and DDBJ Nucleotide Sequence Databases.

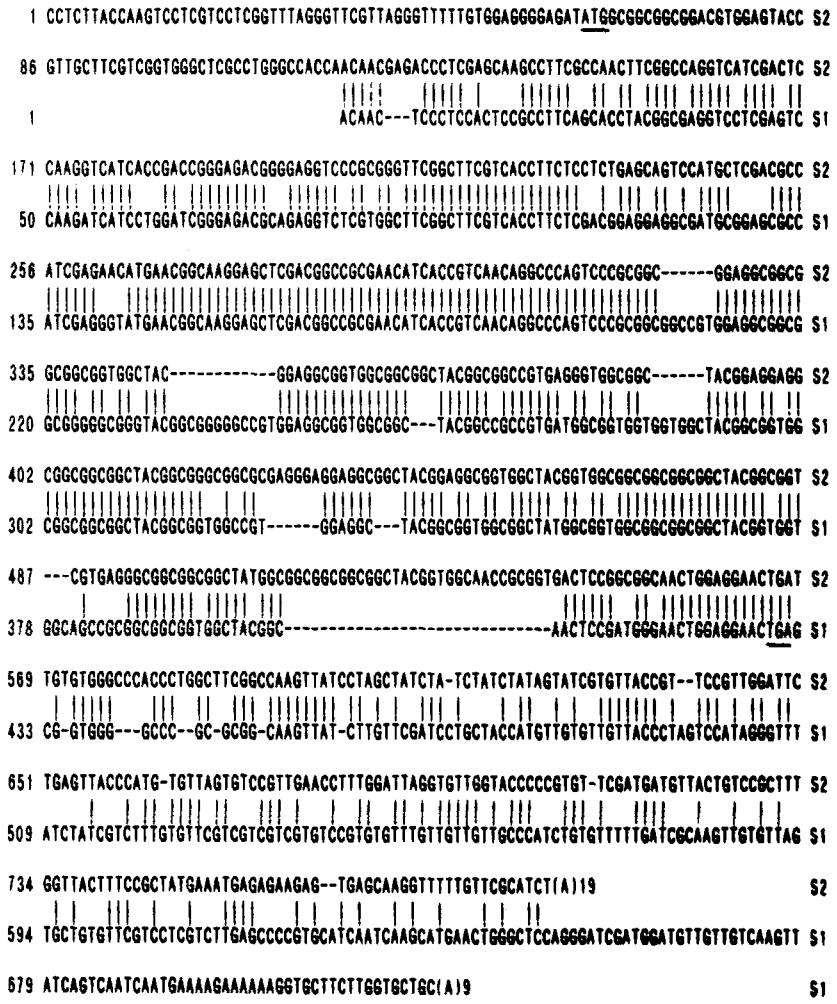

Fig. 1. Nucleotide sequence of two cDNAs coding for glycine-rich protein from *Sorghum vulgare*. The two sequences have been aligned for maximum homology. The ATG and stop signals are underlined.

between each one of the two proteins and that of maize. In fact, the S2 sequence is more homologous to the maize protein than to the S1 sequence in its N-terminal half while the two sorghum proteins have a better homology between themselves in their C-terminal half. This may be an indication of two distinct genes with specific functions for the two glycine-rich proteins. In conclusion, the presence of glycine-rich proteins having complete consensus RNA-binding domains is shown in a species different from maize. There are at least two different species of mRNA in sorghum but the proteins keep their main structural features even in their more variable glycine-rich part, such as the GGYGG pattern. This family of glycine-rich proteins has to be distinguished from the putative cell wall proteins rich in the same amino acid but having significant differences in their structure and in their pattern of mRNA accumulation. With the exception of some of the *Arabidopsis* proteins [2] all of these glycine-rich proteins have tyrosine residues interspersed with the glycine residues and with amino acids having an acid or basic character. The cytokeratin-like protein from barley [7] (as well as the atGRP-3 protein from *Arabidopsis*, see [2] has a C-terminus rich in cysteine, while it shares with dicot GRPs the

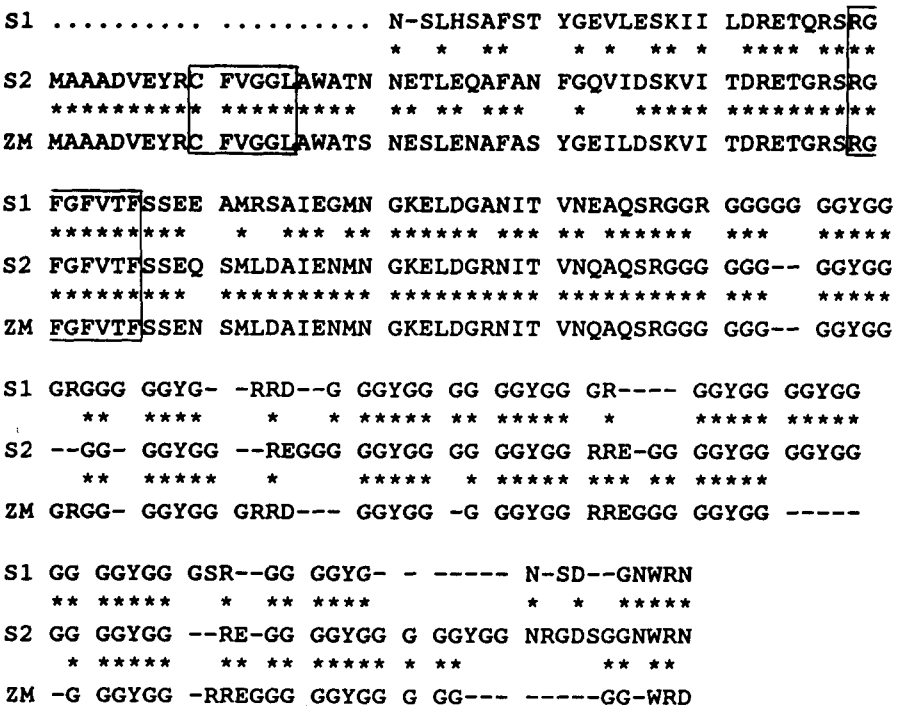

Fig. 2. Comparison of the sequence of the two sorghum glycine-rich proteins with that of the maize protein. The two RNA-binding consensus sequences are boxed. The C-terminal sequence has been aligned to emphasize the conservation of the GGYGG elements.

presence of a signal peptide which is absent in the monocot RNA-binding proteins. Therefore, at least three families of plant GRPs may be distinguished at present: (1) the putative components of cell wall described in some dicotyledonous species [1, 4], (2) the barley cytokeratin-like [7] and (3) the RNA-binding proteins ([3] and this paper). All these proteins, while sharing the glycine-rich domain in part of their sequence that may serve as a point of interaction with other cell components, probably have very distinct functions in the cell.

Acknowledgements

The authors are indebted to Miss Regina Taz and Dr. José Antonio Martínez Izquierdo for assistance and helpful discussions and to Drs. Montserrat Pagès and M. Dolors Ludevid for critical reading of the manuscript. The work has partially been financed by Plan Nacional de Investigación Científica y Técnica (grant BIO99-0242).

References

1. Condit CM, Meagher RB: A gene encoding a novel glycine-rich structural protein of petunia. Nature 232: 178–181 (1986).
2. de Oliveira DE, Seurinck J, Inzé D, Van Montagu M, Botterman J: Differential expression of five *Arabidopsis* genes encoding glycine-rich proteins. Plant Cell 2: 427–436.
3. Gómez J, Sánchez-Martínez D, Stiefel V, Rigau J, Puigdomènech P, Pagès M: A gene induced by the plant hormone abscisic acid in response to water stress encodes a glycine-rich protein. Nature 334: 262–264.
4. Keller B, Sauer N, Lamb CJ: Glycine-rich cell wall proteins in bean: gene structure and association of the protein with the vascular system. EMBO J 7: 3625–3633 (1988).
5. Keller B, Templeton MD, Lamb CJ: Specific localization of a plant cell wall glycine-rich protein in protoxylem cells of the vascular system. Proc Natl Acad Sci USA: 86: 1529–1533 (1989).
6. Mortenson E, Dreyfuss G: RNP in maize protein. Nature 337: 425 (1989).
7. Rohde W, Rosch K, Kröger K, Salamini F: Nucleotide sequence of a *Hordeum vulgare* gene encoding a glycine-rich protein with homology to vertebrate cytokeratins. Plant Mol Biol 14: 1057–1059 (1990).