

## Classification, expression pattern and comparative analysis of sugarcane expressed sequences tags (ESTs) encoding glycine-rich proteins (GRPs)

Adriana Fusaro<sup>1</sup>, Amanda Mangeon<sup>1</sup>, Ricardo Magrani Junqueira<sup>1</sup>, Carla Andréa Benício Rocha<sup>1</sup>,  
Tatiana Cardoso Coutinho<sup>1</sup>, Rogério Margis<sup>1,2</sup> and Gilberto Sachetto-Martins<sup>1\*</sup>

### Abstract

Since the isolation of the first glycine-rich proteins (GRPs) in plants a wealth of new GRPs have been identified. The highly specific but diverse expression pattern of *grp* genes, taken together with the distinct sub-cellular localization of some GRP groups, clearly indicate that these proteins are involved in several independent physiological processes. Notwithstanding the absence of a clear definition of the role of GRPs in plant cells, studies conducted with these proteins have provided new and interesting insights into the molecular biology and cell biology of plants. Complexly regulated promoters and distinct mechanisms for the regulation of gene expression have been demonstrated and new protein targeting pathways, as well as the exportation of GRPs from different cell types have been discovered. These data show that GRPs can be useful as markers and/or models to understand distinct aspects of plant biology. In this paper, the structural and functional features of these proteins in sugarcane (*Saccharum officinarum* L.) are summarized. Since this is the first description of GRPs in sugarcane, special emphasis has been given to the expression pattern of these GRP genes by studying their abundance and prevalence in the different cDNA-libraries of the Sugarcane Expressed Sequence Tag (SUCEST) project. The comparison of sugarcane GRPs with GRPs from other species is also discussed.

### INTRODUCTION

The occurrence of quasi-repetitive glycine-rich peptides has been reported in several different organisms. Glycine-rich regions are thought to be involved in protein-protein interactions in at least three families of mammalian proteins: keratins, lorincins and a series of single stranded RNA-binding proteins (Steinert *et al.*, 1991).

Glycine-rich proteins (GRPs) are distinguished by their characteristic primary structure, having a glycine content of 20 to 70% which forms glycine-rich domains predominantly arranged in quasi-repetitive (Gly-X)<sub>n</sub> motifs. In plants, the expression of genes encoding GRPs is developmentally regulated, and also induced by physical, chemical and biological factors such as wounding, virus infection, circadian rhythms, temperature, salinity, drought, flooding, light, salicylic acid, abscisic acid and ethylene. Besides having highly modulated expression, several GRPs also show tissue-specific expression patterns (reviewed in Sachetto-Martins *et al.*, 2000).

The first clue for the existence of GRPs came from the observation that, in certain plant tissues, glycine is the major fraction of total nitrogen. Examples are soybean and the seed-coat of gourds which contain 21% of glycine and the cell wall of milkweed stem (31% glycine) and oat coleoptile cells (27% glycine) (Varner and Cassab, 1986). Glycine-rich proteins have indeed been isolated from pumpkin seed-coat

(47% glycine) by Varner and Cassab (1986) and strawberry fruit (49% glycine) by Reedy and Poovaiah (1987). The first *grp* gene was isolated by Condit and Meagher (1986). The protein encoded by this gene contains 67% of glycine residues and has a predicted density very similar to the pumpkin protein isolated by Varner and Cassab (1986). These data, together with the identification of an eukaryotic signal peptide at the N-terminus of PtGRP-1, suggest that this protein is indeed a member of the GRP class of cell wall proteins.

Because the GRPs first described were cell wall located, a structural function has always been attributed to the protein encoded by every new *grp* gene described. However, the accumulated data clearly indicate that this concept is an oversimplification and GRPs may have very diverse locations and functions, with the only common feature among all the different GRPs being the presence of glycine-repeat domains (Steinert *et al.*, 1991). As discussed by Steinert *et al.* (1991) in mammalian keratins, the glycine-repeat domains are highly flexible and may act as a "velcro" zipper in protein-protein interactions. This high flexibility makes GRPs good candidates to work in conjunction with other proteins and other macromolecules. GRPs do not necessarily perform structural functions, but it is possible that their glycine-rich regions allow these proteins to assume a structure necessary for their correct conformation and/or interactions with other proteins (Sachetto-Martins *et al.*, 2000). It is quite possible that

<sup>1</sup>Laboratório de Genética Molecular Vegetal, Departamento de Genética, Instituto de Biologia, CCS, Bloco A - sala A2-076, Universidade Federal do Rio de Janeiro, Ilha do Fundão, 21944-970 Rio de Janeiro, RJ, Brazil.

<sup>2</sup>Departamento de Bioquímica, I.Q., Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brazil.

Send correspondence to Gilberto Sachetto-Martins. E-mail: sachetto@biologia.ufrj.br.

GRPs are components of different multi-molecular complexes, with the glycine-rich domain being required for the stabilization and flexibility of molecular interactions in such structures. Alberts (1998) introduced the idea of the cell as a collection of protein machines, and because multiple molecular systems play distinct roles in cell physiology they need to be differentially regulated, which could explain the highly complex gene expression pattern of GRPs.

The large spectrum of modulation and subcellular locations, together with the broad structural diversity of the GRPs (even in a plant with a compact genome such as *Arabidopsis thaliana*), indicates that GRPs do not represent a unique class of proteins. Instead, plant GRPs probably represent a diverse set of proteins which are not necessarily related (Sachetto-Martins *et al.*, 2000).

In this paper we describe the isolation of sugarcane GRPs and summarize the structural and functional features of these proteins. During our research, special emphasis was placed on the expression pattern of these genes by studying their abundance and prevalence in the different cDNA-libraries of the Sugarcane Expressed Sequence Tag (SUCEST) project. The comparison of GRPs from sugarcane with GRPs from other species will also be discussed.

## MATERIAL AND METHODS

### Sequence data, alignment and phylogenetic analysis

A basic local alignment sequence tool (BLAST) T-Blast-n search (Altschul *et al.*, 1997) was performed using GRP sequences listed in a recent review (Sachetto-Martins *et al.*, 2000) as baits against the full SUCEST expressed sequence tag (EST) data bank, using

the fragment assembly program Phrap ([www.phrap.com](http://www.phrap.com)) for clustering. The Clustal W program was used to align the different standards with the proteins deduced from the SUCEST clusters (Thompson *et al.*, 1994).

Phylogenetic analysis was performed using the Molecular Evolutionary Genetics Analysis (MEGA) software (Kumar *et al.*, 2000). The neighbor-joining distance method was used with the pairwise deletion option for the treatment of amino acid gaps during the multiple alignment of sugarcane GRPs. For construction of the phylogenetic tree the confidence levels assigned at various nodes were determined with 5000 replications using the Interior Branch test (Sitnikova *et al.*, 1995).

### SUCEST cDNA libraries

All sugarcane (*Saccharum officinarum* L.) sequences used during this work were obtained from the SUCEST project and derived from cDNA libraries specific to different sugarcane tissues, organs or conditions of growth (for detailed information see <http://sucest.lad.ic.unicamp.br/cgi-bin/prod/BD/webobjects/LibraryList.pl>).

## RESULTS AND DISCUSSION

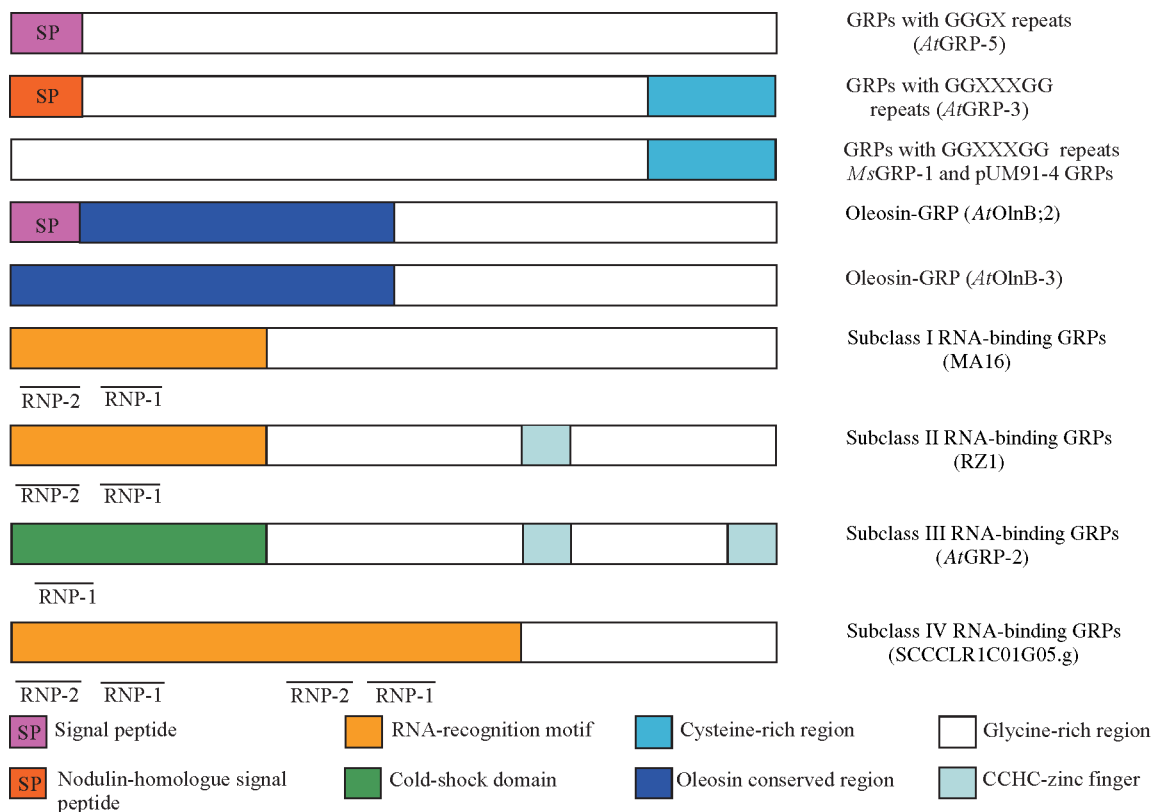
### Distribution of glycine-rich proteins genes on SUCEST data bank

As the name of this collection of proteins suggests, glycine repeats are the major structural feature of GRPs. The glycine-rich domains of plant GRPs consist of sequence repeats that can be summarized by the formula (Gly)n-X. However, several other distinct motifs can also be identified (Table I and Figure 1). Among these, two fre-

**Table I** - Glycine-rich proteins genes in the SUCEST databank.

GRP	General characteristics	Representative member	Number of genes published <sup>a</sup>	Number of SUCEST clusters	SUCEST clusters with preferential expression
GGGX repeats	Signal peptide, cell wall or membrane located	PvGRP1.8	20 (4)	37	28
GGXXXGG repeats	Signal peptide and C-terminal cysteine-rich domains homologues to nodulins	<i>At</i> GRP-3	11 (2)	8	7
GXGX repeats	Proteins with lower glycine content	-	19 (3)	20	11
Oleosin-GRP	Oleosin conserved sequenceTapetal-specific expression	<i>At</i> OlnB-2 and <i>At</i> OlnB-3	12 (0)	0	0
RNA-binding GRP (subclass I)	RRM and GGYGG repeats	MA16CCR1	20 (8)	62	14
RNA-binding GRP (subclass II)	RRM, CCHC zinc finger and GGYGG repeats	RZ-1	2 (1)	11	5
RNA-binding GRP (subclass III)	Cold-shock domain, CCHC zinc fingers and GGYGG repeats	<i>At</i> GRP-2	2 (0)	10	6
RNA-binding GRP(subclass IV)	Multiple RRM motifs	SCCCLR1C01G05.g	-	2	0
Total			86 (18)	150	70

<sup>a</sup>Numbers between parenthesis represents genes from monocotyledonous plants. In our work we considered as preferentially expressed clusters all those in which the reads came from two different libraries at the most.



**Figure 1** - Schematic representation of the primary structure of plant glycine-rich proteins.

quently observed consensus sequences split GRPs into two major groups: group 1, the consensus targeting the endoplasmic reticulum (signal peptide) which is present in most of the GRPs identified up to now, and group 2, the RNA-binding consensus sequences RNP-1 and RNP-2. As discussed by Sachetto-Martins *et al.* (2000), other motifs also exist, *e.g.* the RNA-recognition motif (RRM), the oleosin-conserved domain, the cold-shock domain, CCHC zinc-fingers, the C-rich carboxy-terminus, the amphiphilic-helix and H-rich, P-rich and T-rich sequences.

Five different classes of GRPs are shown in Table I, three groups based on the pattern of the glycine-rich repeats (class I, GGGX; class II, GGXXXGG; class III, GXGX) and two other groups based on the type of functional conserved motif (class IV, the oleosin glycine-rich proteins and class V, the RNA-binding GRPs).

In order to search the SUCEST database for genes encoding glycine-rich proteins, we used the typical GRP proteins sequences described for other species as well as the common motifs frequently found in these proteins. Different types of GRPs were found, a total of 150 different clusters being distributed among the general classes summarized in Table I, representing the first isolation of sugarcane GRP genes. These data show that the SUCEST sequences are equivalent to the large collection of GRPs from monocotyledonous plants already published. Even when compared in absolute terms, the number of GRP

genes listed in this paper surpasses the number of GRPs available before, suggesting that our inventory might contain almost all the GRPs expressed in monocotyledonous plants.

Oleosin GRPs are specifically expressed in the anther tapetum layer of *Brassica* and *Arabidopsis* (both dicotyledonous plants). The presence of oleosin-conserved sequences in GRPs is indicative of the involvement of these GRPs in lipid stabilization during tapetal development, as upon tapetal degradation lipid-containing complexes are sorted to the pollen coat to become part of tryphine (Ting *et al.*, 1998). Interestingly no member of the oleosin-GRP class was found in the SUCEST data bank. The absence of oleosin-GRPs in sugarcane and the fact that this kind of GRP has never been isolated from a monocotyledonous plant is evidence of the differences between the anthers and pollen grains of monocotyledonous and dicotyledonous plants.

The distribution of each EST sequence between the different SUCEST libraries was also analyzed. The SUCEST data bank comprises 291,689 EST sequences, arranged in 43,141 clusters, these reads coming from 37 different cDNA libraries constructed from different plant tissues under different culture conditions (see <http://sucest.lad.ic.unicamp.br/cgi-bin/prod/BD/webobjects/LibraryList.pl>). Since several GRP genes present tissue-specific expression in other plants, we analyzed the distribution of the

reads from each cluster per library, considering as preferentially expressed clusters all those in which the reads came from two different libraries at most. Several clusters presented a putative tissue-specific expression pattern.

#### Sugarcane ESTs encoding GRPs with GGGX repeats

The typical GGGX array is the most frequently observed array, and is generally present in GRPs displaying a signal peptide and high (40% to 70%) glycine content. However, it can also be observed in GRPs without a signal peptide but with RNA-binding sequences, in which case the repeats are usually GGGY (Sachetto-Martins *et al.*, 2000).

The existence of a signal peptide consensus sequence in the N-terminal end of most GRPs suggests that these proteins are located in the cell wall or cell membrane. Studies with PvGRP-1.8 have shown that this GRP is localized in the cell wall of proto-xylem elements and, surprisingly, is exported to these cells rather than synthesized there (Ryser and Keller, 1992). The same pattern of exportation to a non-producing cell has been observed for PtGRP-1 (Condit, 1993).

By analogy with HRGPs, Sachetto-Martins *et al.* (2000) have proposed that cell wall GRPs have a structural function, probably acting as a scaffold or agglutinating agent for deposition of cell wall constituents. The detection of PvGRP-1.8 in tissues undergoing lignification might indicate some interaction between these two cell wall components. The presence of tyrosine residues in many plant cell wall GRPs, the progressive reduction in solubility of PvGRP-1.8 and soybean GRP during development and the detection of a higher molecular weight isoform of ZmGRP-4 in root cap mucilage may indicate that the association of these proteins with themselves or with other components of the extracellular matrix might occur via isodityrosine cross-links, as described for the extensins (Sachetto-Martins *et al.*, 2000).

Thirty-seven clusters encoding GRPs with GGGX repeats were identified in the SUCEST database. These sequences were different to each other but related to different types of GRPs previously described from monocotyledonous and dicotyledonous plants (Table II). Among these 37 clusters, 28 were isolated from only one or two types of libraries, indicating that these genes could correspond to GRPs with a potential tissue-specific expression pattern. Four of these clusters (SCBFAD1094D08.g, SCEQHR1081H03.g, SCRLAD1102C04.g and SCSFHR1043H11.g) were specifically identified in the library prepared from sugarcane infected with the nitrogen-fixating endosymbiont bacteria *Herbaspirillum rubrisubalbicans* (HR) and *Gluconacetobacter diazotrophicans* (AD), suggesting a possible role for these genes in the mechanism of plant-microbe interaction. Four leaf-specific clusters (SCA CLB1046B05.g, SCCCLB1C04A01.g, SCJLLR1104 A01.b and SCRFLR2021C05.g) and four flower-specific clusters (SCQSFL1127H07.g, SCSBFL1046H04.g, SCVPFL1069E01.g and SCRLFL3006C09.g) were also observed.

To analyze the similarities between the 37 sugarcane GRPs and the 20 genes previously published by Sachetto-Martins *et al.* (2000) we aligned these sequences, but, as expected due to the high variation in the structure of this type of GRP, the alignments obtained presented many gaps and regions with no sequence overlapping, which made the construction of a dendrogram impossible.

#### Sugarcane ESTs encoding GRPs with N- and C-terminal domains homologous to nodulins

The GGXXXGG repeat occurs in a group of GRPs in which some members share similarities with soybean nodulin 24 (Sandal *et al.*, 1992), the tripeptide repetition interspersed between the glycine residues being generally composed of Y, H, R, N or Q. Some of these GRPs have been proposed to be located at the interface between the host plant membrane and the matrix surrounding the endosymbiont (Sandal *et al.*, 1992). Tripeptides with similar composition have been observed in non-GRP extracellular matrix proteins. We propose that it may be a ligand attachment region, in analogy to the RGD motif of the extracellular matrix adhesive proteins of mammals. The tripeptide repeats in plant GRPs may also represent interactive sites for association with other proteins or with other cellular structures (Sachetto-Martins *et al.*, 2000).

The binding function of this type of GRP has been recently proved by the interaction of AtGRP-3 with the extracellular domain of cell wall associated kinases (WAKs). It was shown that phosphorylation modulates this interaction and that the cysteine-rich C-terminal domain of AtGRP-3 is responsible for this interaction *in vitro*. AtGRP-3 regulates Wak1 function through binding to its cell wall domain and the interaction of Wak1 with AtGRP-3 occurs in a pathogenesis-related process *in planta* (Park *et al.*, 2001).

Searching the SUCEST data bank using the previously reported GRPs with N- and C-terminal domains homologous to nodulins, as well as the cysteine-rich domain, allowed us to identify 8 different GRPs (Table III). Although these 8 sequences are not very similar to each other, they appear to be related to the previously described GRP-1 and GRP-2 from barley. Four of these clusters were preferentially observed in one or two types of libraries. Alignments conducted with the sugarcane ESTs and other similar GRPs resulted in the unrooted tree shown in Figure 2. Two well-separated groups can be seen. All the sugarcane genes group together, being more related to the two barley GRPs and the several *Arabidopsis* AtGRP-3-like genes.

#### Sugarcane ESTs encoding GRPs with lower glycine content

This last pattern of glycine repeats, GXGX, is generally observed in GRPs with an average glycine content of 20%. Similar to the GGGX group (Table II) this GRP group shows a high degree of structural diversity and probably contains several different types of GRPs.



**Table II** - Sugarcane ESTs encoding GRPs with GGGX repeats.

SUCEST Cluster	Related sequence	Accession number	e-value	Library expression pattern <sup>a</sup>
SCACLB1046B05.g	<i>At</i> GRP-1	S47405	3e-27	LB1
SCAGLB1072F05.g	<i>At</i> GRP-1	S47405	1e-72	AM2, LB1
SCBFAD1094D08.g	<i>Pv</i> GRP-1.8	S01820	5e-28	AD1 (2X)
SCCCCL3001D07.b	<i>At</i> GRP-1	S47405	5e-90	CL1, SD1, SD2 (3X)
SCCCCL4013A08.g	<i>At</i> GRP-1	S47405	8e-81	AD1, CL4
SCCCCL7C02E08.g	<i>Os</i> GRP-1	P25074	6e-49	CL7, LV1
SCCCCL7C05D02.g	<i>Pt</i> GRP-1	A26099	5e-46	CL7
SCCCCLB1003D10.g	<i>Pt</i> GRP-1	A26099	8e-93	HR1 (3X), LB1, LB2, FL1 (2X), SD2
SCCCCLB1C04A01.g	<i>At</i> GRP-1	S47405	2e-84	LB1(2X), LB2 (3X)
SCCCCLR1C10G03.g	<i>At</i> GRP-1	S47405	2e-92	AD1 (2X), SD2 (2X), LR1 (3X), RZ3
SCEQHR1081H03.g	<i>At</i> GRP-5	S47414	9e-30	HR1
SCEZAD1080B04.g	<i>At</i> GRP-1	S47405	3e-71	HR1 (2X), AD1
SCEZHR1049H05.g	<i>Pt</i> GRP-1	A26099	2e-65	HR1 (3X), SD2 (2X)
SCEZHR1054F02.g	<i>Os</i> GRP-1	P25074	1e-31	HR1, AD1, SD1, FL1, FL4 (2X)
SCEZHR1055D03.g	<i>At</i> GRP-1	S47405	5e-37	SD1 (2X), SD2, HR1
SCEZLB1005F03.g	<i>At</i> GRP-5	S47414	1e-36	LB1, RZ3
SCEZLB1010C10.g	<i>At</i> GRP-4	S47413	2e-08	SD2 (3X), LB1
SCEZRZ1017B01.g	<i>Pv</i> GRP1.8	S01820	7e-41	SD2 (3X), RZ1, RZ3 (3X)
SCJFRT2059G06.g	<i>Os</i> GRP-1	P25074	2e-35	HR1, CL7, AM2, RT2
SCJLHR1029D04.g	<i>At</i> GRP-1	S47405	3e-88	HR1, AD1, LB2
SCJLLR1104A01.b	<i>Pt</i> GRP-1	A26099	7e-37	LR1
SCQSFL1127H07.g	<i>At</i> GRP-6	P27483	6e-14	FL1
SCRFLR2021C05.g	<i>At</i> GRP-5	S47414	9e-42	LR2
SCRLAD1102C04.g	<i>At</i> GRP-1	S47405	7e-34	AD1
SCSBFL1046H04.g	<i>At</i> GRP-5	S47414	1e-51	FL1
SCSFHR1043H11.g	<i>Pt</i> GRP-1	A26099	5e-55	HR1
SCVPAM1059E11.g	<i>Pt</i> GRP-1	A26099	2e-34	AM1, ST3 (3X)
SCVPLB1015C05.g	<i>At</i> GRP-6	P27483	2e-12	HR1, LB1
SCVPLB1015G03.g	<i>At</i> GRP-5	S47414	5e-49	HR1, LB1
SCVPFL1069E01.g	<i>At</i> GRP-6	P27483	9e-15	FL1
SCSBRZ2021C09.g	<i>At</i> GRP-1	S47405	2e-30	RZ2, SD2 (3X), HR1
SCR FAM2071C04.g	root-specific protein <i>Oryza sativa</i>	AC037426	6e-55	AM2, HR1
SCJFRT2059D04.g	NT16	T03018	5e-54	AM1 (2X), AM2 (3X), LV1, LB2 (2X), ST1, RT1 (3X), RT2, FL1, FL4 (2X), FL8, SD1, SD2 (2X)
SCEZLR1031F06.g	NT16	T03018	3e-55	AM2 (3X), FL1, FL3, FL4 (3X), FL8 (2X), HR1 (2X), LR1, RT2 (2X), SD1 (6X), SD2 (5X), ST3 (3X)
SCEZRZ1015F10.g	NT16	T03018	2e-55	AD1 (3X), AM2, FL4 (2X), HR1, LB1, RT1 (7X), RT2 (5X), RZ1, RZ3, SB1, SD1 (4X), ST3 (3X)
SCRLFL3006C09.g	Keratin mouse	KRMS2	1e-25	FL3
SCR FAD1116B01.g	<i>Os</i> GRP-2	CAA38315	8e-41	SD2, RT2, AD1

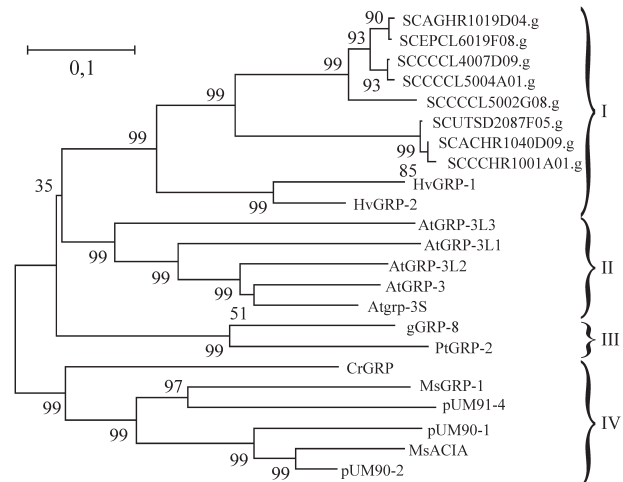
<sup>a</sup>SUCEST library codes: AM1 and AM2 (apical meristem from mature and immature plants); FL1,FL3,FL4,FL5 and FL8 (flowers at different stages of development); LB1 and LB2 (lateral buds); LR1 and LR2 (leaf-root transition zone); LV1 (etiolated leaves), NRn (non-redundant); RTn (root), SB1 (Stem-bark); SDn (seed); STn (stem); CL3, CL4, CL6 and CL7 (callus submitted to different light-dark cycles); AD1 (infected with *Gluconacetobacter diazotrophicans*) and HR1 (infected with *Herbaspirillum rubrisubalbicans*).

**Table III** - Sugarcane ESTs encoding GRPs with N- and C-terminal domains homologous to nodulins.

SUCEST Cluster	Related sequence <sup>a</sup>	e-value	Library expression pattern <sup>b</sup>
SCACHR1040D09.g	<i>HvGRP-2</i>	8e-20	HR1, SD2 (3X)
SCAGHR1019D04.g	<i>HvGRP-2</i>	5e-22	HR1, AM2
SCCCCL4007D09.g	<i>HvGRP-2</i>	2e-21	CL4, CL7, SD1, LV1
SCCCCL5002G08.g	<i>HvGRP-2</i>	1e-17	CL5
SCCCCL5004A01.g	<i>HvGRP-2</i>	2e-21	CL5, CL6, AD1 (2X), HR1,
SCCCHR1001A01.g	<i>HvGRP-2</i>	1e-15	HR1, SD2 (2X)
SCEPCL6019F08.g	<i>HvGRP-1</i>	1e-29	CL6 (2X)
SCUTSD2087F05.g	<i>HvGRP-2</i>	3e-18	SD2

a - Accession numbers: *HvGRP-1* - X52580; *HvGRP-2* - Z48625

b - SUCEST library codes: AM2 (apical meristem from immature plants); LV1 (etiolated leaves); SDn (seed); CL4, CL5, CL6 and CL7 (callus submitted to different light-to-dark cycles); AD1 (infected with *Gluconacetobacter diazotrophicans*) and HR1 (infected with *Herbaspirillum rubrisubalbicans*).



**Figure 2** - Unrooted dendrogram of glycine-rich proteins with N- and C-terminal sequences homologous to nodulins. The relationships were calculated using MEGA (p distance, neighbor-joining method and bootstrap test with 5000 replications, pairwise deletions). The analysis was based on Clustal W alignment of the sequences. Accession numbers: barley *HvGRP-1* (X52580) and *HvGRP-2* (Z48625); *Arabidopsis AtGRP-3* (S47409), *AtGRP-3s* (AAD11798), *AtGRP-3L1* (AAD24656), *AtGRP-3L2* (AAD24654) and *AtGRP-3L3* (AAD24653); *Nicotiana tabacum gGRP-8* (M37152); *Petunia PtGRP-2* (S11959); *Chenopodium rubrum CrGRP* (S04069); *Medicago sativa MsGRP-1* (X59930), pUM91-4 (AAA32652), pUM90-1 (AAA32653), pUM90-2 (AAA32651) and *MsACIA* (L03708). The sugarcane sequences use SUCEST nomenclature.

**Table IV** - Sugarcane ESTs encoding GRPs with lower glycine content.

SUCEST Cluster	Related sequence	Accession number	e-value	Library expression pattern <sup>a</sup>
SCACHR1035G08.g	Root-specific protein RCc3 <i>Oryza sativa</i>	S53012	3e-19	HR1 (3X), RT3 (2X)
SCCCCL7002C06.g	<i>AtOlnB</i> ;3	S19933	6e-14	FL1 (2X), FL3, SD1, CL7, LR1
SCCCHR1001B01.g	PtGRP-1	A26099	1e-54	HR1 (6X), RT2 (2X), RT3 (2X)
SCCCLR2002A01.g	<i>OsGRP-2</i>	CAA38315	2e-41	FL4 (3X), LR1, LR2, RZ3
SCCART3001B10.g	Al-induced GRP <i>Zea mays</i>	AAB86493	e-101	RT3 (5X)
SCCART3008C02.g	Al-induced GRP <i>Zea mays</i>	AAB86493	7e-82	RT3 (2X)
SCCCST1004E07.g	<i>OsGRP-2</i>	CAA38315	3e-27	ST1, HR1
SCEPRZ1008B12.g	PvGRP1.8	S01820	5e-32	LB1 (3X), RZ1, RT2, RT3 (3X), FL1 (2X), CL6, LR2
SCQSFL1125A01.g	<i>AtGRP1</i>	S47405	3e-33	FL1 (2X), FL4
SCRURT2014C05.g	Al-induced GRP <i>Zea mays</i>	AAB86493	4e-83	RT2, RT3
SCSGHR1068E12.g	<i>OsGRP-2</i>	CAA38315	2e-32	HR1 (2X)
SCVPFL3040G09.g	Root-specific protein <i>Oryza sativa</i>	AC037426	1e-55	FL3, LV1
SCVPRT2080B02.g	Al-induced GRP <i>Zea mays</i>	AAB86493	1e-79	RT2, RT3 (4X)
SCRULR1020D10.g	<i>OsGRP-1</i>	P25074	6e-26	AM1 (2X), FL1 (2X), LR1, SD2
SCBFLR1046B02.g	p23 co-chaperone <i>Arabidopsis thaliana</i>	AJ297951	2e-69	AM1 (2X)
SCVPRT2083C11.g	Putative ubiquitin <i>Oryza sativa</i>	AC051633	2e-72	AM1, AM2, SB1, RT2, FL1, FL4, SD1
SCACSB1119B05.g	FGFR-4 gene <i>Homo sapiens</i>	Y13901	2e-15	SB1
SCRLFL4028F09.g	Putative ubiquitin <i>Oryza sativa</i>	AC051633	e-102	SB1, RT3 (2X), FL4 (2X)
SCCCLR2003B12.g	p23 co-chaperone <i>Arabidopsis thaliana</i>	AJ297951	6e-66	LR1, LR2, RT3, FL1 (2X), SD1
SCRULR1020D10.g	<i>OsGRP-1</i>	P25074	6e-26	AM1 (2X), FL1 (2X), LR1, SD2

a - SUCEST library codes: AM1 and AM2 (apical meristem from mature and immature plants); FL1,FL3,FL4,FL5 and FL8 (flowers at different developmental stages); LB1 and LB2 (lateral buds); LR1 and LR2 (leaf-root transition zone); LV1 (etiolated leaves), NRn (non-redundant); RTn (root), SB1 (stem-bark); SDn (seed); STn (stem); CL3, CL4, CL6 and CL7 (callus submitted to different light-dark cycles); AD1 (infected with *Gluconacetobacter diazotrophicans*) and HR1 (infected with *Herbaspirillum rubrisubalbicans*).

Twenty different clusters were identified, of which eleven were isolated from only one or two library types (Table IV). Interestingly six of these genes show preferential expression in the root system (SCACHR1035G08.g, SCCCHR1001B01.g, SCCCRT3001B10.g, SCCCRT3008C02.g, SCRURT2014C05.g and SCVPRT2080B02.g), and are also related to two previously identified root-specific GRPs of maize. As observed for the GGGX GRPs, the high variation in the structures of the GXGX GRPs did not allow the construction of a dendrogram.

#### Sugarcane ESTs encoding RNA-binding GRPs

Several GRPs with RNA-binding sequences have been identified in different plant species, most of them having an 80 to 100 amino acid N-terminus conserved region (the RNA-Recognition Motif (RRM) or Consensus Sequence-type RNA-Binding Domain (CS-RBD)) containing two conserved sequences (RNP-1 and RNP-2). Two plant GRPs (*At*GRP-2 and *Ns*GRP-2) have different RNA-binding sequences at their N-terminus. Instead of an RRM domain they have a cold-shock domain with only the RNP-1 sequence (Sachetto-Martins *et al.*, 2000).

Some of these GRPs are known to have *in vitro* RNA binding activity (Ludevid *et al.*, 1992; Hirose *et al.*, 1993; Freire *et al.*, 1995; Hanano *et al.*, 1996; Dunn *et al.*, 1996) and also bind single-stranded DNA (Hirose *et al.*, 1993; Dunn *et al.*, 1996) and are phosphorylated *in vitro* (Dunn *et al.*, 1996) and *in vivo* (Freire *et al.*, 1995). Deletion assays conducted with the RZ-1 protein have shown that the RNA-binding domain and the C-terminal glycine-rich region are essential for RNA-binding activity (Hanano *et al.*, 1996). Immuno-precipitation studies have shown that the MA16 protein interacts with RNAs through a complex association with several proteins (Freire *et al.*, 1995). These results, together with the immuno-localization of *Sa*GRP (Heintzen *et al.*, 1994), MA16 (Albà *et al.*, 1994) and RZ-1 proteins (Hanano *et al.*, 1996) in the nucleus have led to the hypothesis that these GRPs may be involved in RNA processing, maturation or the control of gene expression. The modulation of some RNA-binding GRPs by factors such as ABA (abscisic acid), salinity, wounding, cold-stress and circadian rhythm may reflect their involvement in the modulation of the pathways activated by these stimuli (Sachetto-Martins *et al.*, 2000).

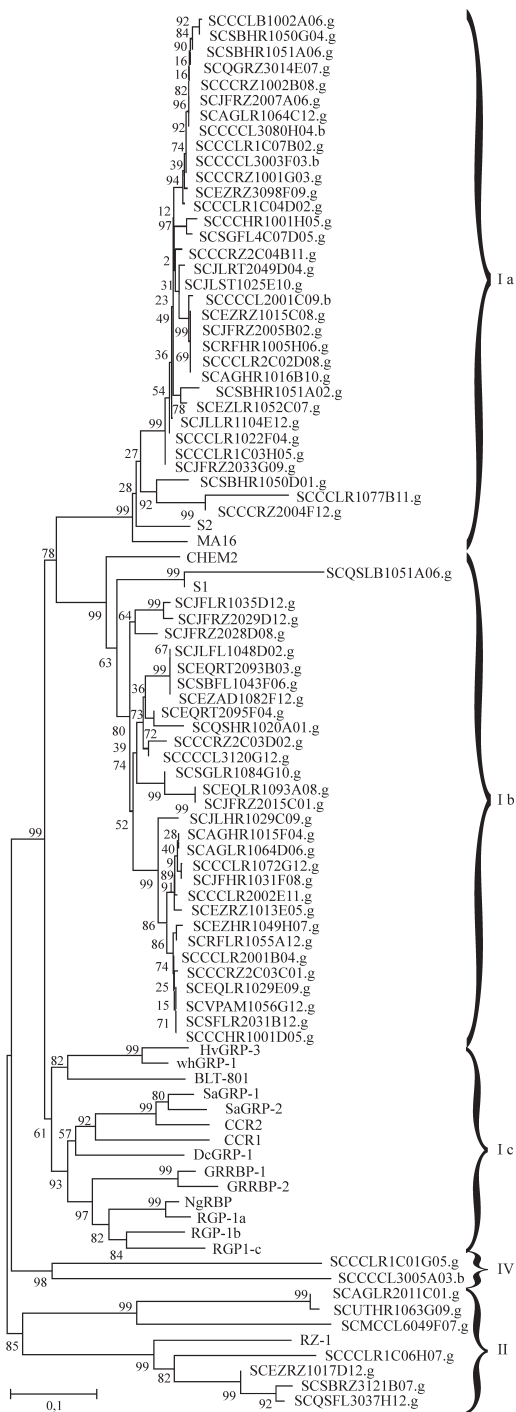
In addition to their glycine-rich and RNA-binding motifs, *At*GRP-2 (de Oliveira *et al.*, 1990), *Ns*GRP-2 (Obokata *et al.*, 1991) and RZ-1 (Hanano *et al.*, 1996) contain one or two CCHC zinc-fingers. Similar domains have been observed in yeast and mammals splicing factors, as well as in several retrovirus GAG proteins and in the human nucleic acid binding protein CNBP (Kingsley and Palis, 1994), and it is possible that these proteins represent components of the plant cell splicing machinery (Hanano *et al.*, 1996).

Based on their structural features, the RNA-binding GRPs can be classified into three different sub-classes. Proteins from the first class show a RRM conserved motif at the N-terminal end followed by a glycine-rich region with GGYGG repeats. GRPs from the second class show similar organization, but present a CCHC zinc-finger inside their glycine-rich region. Proteins from the third subclass are organized with a cold-shock domain at the N-terminus and two copies of the CCHC zinc-finger in their glycine-rich region (Table I and Figure 1). Our analysis of the SUCEST database identified a novel group of RNA-binding GRPs (subclass IV, clusters SCCCCL3005A03.b and SCCCLR1C01G05.g), this new subclass having two copies of the RRM motif followed by a C-terminal glycine-rich region distinct from previous GRPs.

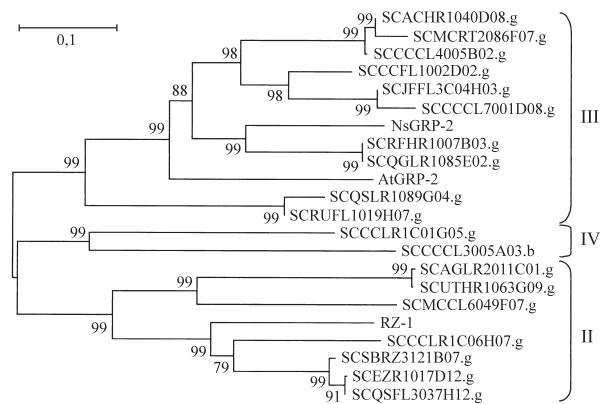
The RNA-binding GRPs from subclasses I and III show a high degree of divergence, and these differences introduced a high number of gaps during the alignment of all the RNA-binding sequences and did not allow correlation between the sequences or construction of a dendrogram. To analyze the correlation between the four RNA-binding GRP subclasses, the alignments were constructed mutually excluding GRPs from subclasses I and III (Figures 3 and 4). The observation that class IV RNA-binding GRPs remain as a separated group in both analysis justifies this treatment and shows the existence of a fourth subclass of RNA-binding GRPs.

Subclass I RNA-binding GRPs were subdivided into at least three subgroups, subgroups Ia and Ib being composed of sequences from monocotyledonous plants and subgroup Ic containing all the sequences from dicotyledonous plants (Figure 3). We found 62 SUCEST clusters encoding subclass I RNA-binding GRPs (Table I). These clusters are preferentially related to the CHEM2, MA16, S1 and S2 GRPs and present e-values ranging from 7e-39 to 9e-81. In general, these clusters were expressed in a large spectrum of libraries, being represented in virtually all the libraries analyzed. These results indicate that the transcripts of these genes accumulated at high levels in most of the organs investigated, suggesting that they may be constitutively synthesized and involved in fundamental cell processes. However, four clusters presented a more restricted expression pattern and seven (SCAGHR1015F04.g, SCCCHR1001D05.g, SCCCHR1001H05.g, SCEZAD1082F12.g, SCJLFL1048D02.g, SCJLHR1029C09.g, and SCSGFL4C07D05.g) were preferentially expressed in flower-libraries. Some clusters were expressed during infection with nitrogen-fixating endosymbionts (Table V).

Eleven subclass II RNA-binding GRPs were detected in the SUCEST database, and five of them presented a preferential expression pattern (Table V). GRPs of this subclass have only recently been isolated from monocotyledonous plants (Ni *et al.*, 2000), and the detection of sugarcane



**Figure 3** - Unrooted dendrogram of RNA-binding glycine-rich protein sequences. The relationships were calculated using MEGA (p distance, neighbor-joining method and bootstrap test with 5000 replications, pairwise deletions). The analysis was based on the Clustal W alignment of sequences. Accession numbers: maize MA16 (P10979) and CHEM2 (CAA43431); Sorghum S1 (S12311) and S2 (S12312); barley *HvGRP-3* (Z48624) and BLT-801 (S71453); wheat *WhGRP-1* (U32310); *Sinapis alba* *SaGRP-1* (L31374) and *SaGRP-2* (L31377); *Arabidopsis* CCR1 (Q03251) and CCR2 (Z14987); *Daucus carota* *DeGRP-1* (X58146); *Euphorbia esula* *GRRBP-1* (AAC61786) and *GRRBP-2* (AAC61787); *Nicotiana glutinosa* *NgRBP* (AF005359); *Nicotiana sylvestris* *RGP-1a* (D16204), *RGP-1b* (D16205), *RGP-1c* (D16206) and *RZ-1* (D28861). The sugarcane sequences use SUCEST nomenclature. Roman numerals represent the different subclasses of RNA-binding GRPs.



**Figure 4** - Unrooted dendrogram of RNA-binding glycine-rich protein sequences. The relationships were calculated using MEGA (p distance, neighbor-joining method and bootstrap test with 5000 replications, pairwise deletions). The analysis was based on the Clustal W alignment of sequences. Accession numbers: *Nicotiana sylvestris* *NsGRP-2* (CAA42622) and *RZ-1* (D28861); *Arabidopsis* *AtGRP-2* (S47408). The sugarcane sequences use SUCEST nomenclature. Roman numerals represent the different subclasses of RNA-binding GRPs.

GRPs from this class may help to elucidate their biological function.

Subclass III GRPs have never before been isolated from monocotyledonous plants, but we identified ten clusters in the SUCEST data bank, six of them representing putative tissue-specific genes (Table V). The principal characteristics of these GRPs are the combination of two well-defined nucleic acid-binding motifs, a cold-shock domain and CCHC zinc-fingers.

The diversity of the glycine-repeats in GRPs suggests that some of these genes may have originated from an ancestral gene by recombination events. Genes of GRPs organized *in tandem* on the genome have been isolated in *Arabidopsis thaliana* and bean. The high GC content of the *grp* genes makes them hot spots for recombination, and this mechanism is responsible for the genetic diversity of the mammalian keratin gene family in which evolutionary divergence is due to recombination at glycine-rich regions (Steinert *et al.*, 1991). Because of its multi-species origin, sugarcane is thought to have one of the most complex plant genomes, with a variable number of chromosomes (generally  $2n = 70-120$ ) and a large DNA content. Hot-spot recombination could explain the high variability in the different classes of sugarcane GRPs.

Several GRPs were expressed in a tissue-specific manner, and the identification of the cell types in which GRPs are expressed opens the possibility of addressing questions on the function of these genes. In other plant species GRPs have been described that are preferentially expressed in xylem (Keller *et al.*, 1989), phloem (Condit, 1993), epidermis (Sachetto-Martins *et al.*, 1995), tapetum (Ferreira *et al.*, 1997), embryo (Gómez *et al.*, 1988) and roots (Goddemeier *et al.*, 1998). Tissue-specific genes have



**Table V** - Sugarcane ESTs encoding RNA-binding GRPs.

SUCEST Cluster	Related sequence <sup>a</sup>	e-value	Library expression pattern <sup>b</sup>
Subclass I RNA-binding GRPs with library-specific expression pattern			
SCAGHR1015F04.g	CHEM2	9e-77	HR1, FL3
SCCCHR1001D05.g	MA16	2e-62	FL3 (2X), FL4 (2X), SD1, HR1 (7X)
SCCCHR1001H05.g	MA16	4e-71	FL1, FL3 FL4 (2X), HR1
SCEZAD1082F12.g	CHEM2	1e-70	FL1, FL3 (2X), AD1
SCEZHR1049H07.g	CHEM2	3e-74	SD2 (3X), FL3, HR1 (4X), AD1
SCJFHR1031F08.g	CHEM2	1e-76	LV1, SD1, HR1 (5X)
SCJFRZ2028D08.g	CHEM2	7e-39	AM2, RT2, RZ2
SCJFRZ2033G09.g	MA16	3e-41	AM2, RT2, RZ2 (2X)
SCJLFL1048D02.g	CHEM2	3e-76	FL1 (3X), FL3, FL4
SCJLHR1029C09.g	CHEM2	3e-72	FL5 (7X), HR1 (2X), AD1
SCAGHR1016B10.g	MA16	4e-77	HR1 (7X) RZ2, NR1
SCSBHR1050D01.g	MA16	6e-35	HR1
SCSBHR1051A02.g	MA16	4e-67	HR1
SCSGFL4C07D05.g	MA16	2e-72	FL1 (2X), FL3 (3X), FL4 (4X)
Subclass II RNA-binding GRPs			
SCAGLR2011C01.g	RZ1	6e-36	LR2, FL3 (2X), FL4, RZ3, AM1 (2X)
SCCCLR1C06H07.g	RZ1	9e-75	LR1 (2X), RT1, LB1 (2X), FL3
SCEZRZ1017D12.g	RZ1	6e-64	RZ1, RZ2, AM1
SCMCCL6049F07.g	RZ1	3e-40	CL6 (2X), SB1 (3X), RT3, RZ3, FL3, FL4, LB1
SCUTHR1063G09.g	RZ1	1e-35	FL1 (3X), FL4 (2X), HR1 (2X), LB2, AM1, RZ3
SCSBRZ3121B07.g	RZ1	5e-74	RZ3 (2X), AD1
SCQSFL3037H12.g	RZ1	3e-64	FL1, FL3 (2X), SD2
SCQSLR1090G01.g	RZ1	3e-63	AM2, LB2, ST3, LR1, RZ2, FL1, FL4 (3X)
SCEZLR1031E02.g	RZ1	2e-77	LB1, RT3, RZ2 (2X), RZ3 (2X), LR1 (2X), FL3, FL5 (2X), HR1
SCEZLB2066F06.g	RZ1	3e-45	LB2
SCVPRT2082E08.g	RZ1	1e-56	RT2, RZ3, CL2, HR1 (2X)
Subclass III RNA-binding GRPs			
<i>SCRFHR1007B03.g</i>	<i>AtGRP-2</i>	3e-31	HR1 (3X)
SCACHR1040D08.g	<i>AtGRP-2</i>	5e-46	HR1 (2X)
SCJFFL3C04H03.g	<i>NsGRP-2</i>	2e-10	FL3 (2X)
SCCCFL1002D02.g	<i>AtGRP-2</i>	6e-38	FL1, FL3, FL5 (2X), AM1, AD1
SCCCCL4005B02.g	<i>AtGRP-2</i>	6e-47	FL1
SCCCCL7001D08.g	<i>AtGRP-2</i>	5e-65	CL7 (2X), RT3, AM2, SB1, RT2
SCMCRT2086F07.g	<i>AtGRP-2</i>	9e-59	RT2, RT3, ST1, AM2, FL1, FL3
SCQGLR1085E02.g	<i>AtGRP-2</i>	7e-29	LR1, SD2
SCQSLR1089G04.g	<i>AtGRP-2</i>	2e-34	LR1, RT1, FL1 (2X), ST3
SCRUFL1019H07.g	<i>AtGRP-2</i>	7e-37	FL1, AD1

a - Accession numbers MA16 - P10979; CHEM2 - CAA43431; RZ1 - D28861; *AtGRP-2* - S47408; *NsGRP-2* - CAA42622.

b - SUCEST library codes: AM1 and AM2 (apical meristem from mature and immature plants); FL1, FL3, FL4, FL5 and FL8 (flowers at different stages of development); LB1 and LB2 (lateral buds); LR1 and LR2 (leaf-root transition zone); LV1 (etiolated leaves), NRn (non-redundant); RTn (root), SB1 (stem-bark); SDn (seed); STn (stem); CL3, CL4, CL6 and CL7 (callus submitted to different light-dark cycles); AD1 (infected with *Gluconacetobacter diazotrophicans*) and HR1 (infected with *Herbaspirillum rubrisubalbicans*).

received a great deal of attention because of their biotechnological potential. Using the promoter of the tape-tum-specific *ta29* GRP male-sterile tobacco and *Brassica* plants were obtained for hybrid production in crop improvement programs (Mariane *et al.*, 1990). The identification of sugarcane GRP genes preferentially expressed in specific organs opens up the possibility of isolating tissue-specific promoters that can be used to drive the expression of exogenous genes in sugarcane as well as in other monocotyledonous plants.

## RESUMO

Desde o isolamento da primeira proteína rica em glicina (GRP) em plantas, um grande número de novas GRPs vem sendo identificado. Seu padrão de expressão altamente específico, embora variado, em conjunto com as diferentes localizações sub-celulares de alguns dos tipos de GRPs, claramente indica que estas proteínas encontram-se envolvidas em diversos processos fisiológicos independentes. Embora ainda sem uma clara definição do papel de GRPs na célula vegetal, estudos realizados com estas proteínas têm resultado em novos e interessantes esclarecimentos da biologia celular e molecular de plantas. Promotores com regulação complexa, assim como distintos mecanismos de regulação da expressão gênica tem sido demonstrados. Novas vias de endereçamento de proteínas tais como a exportação de GRPs para diferentes tipos celulares, tem sido observados. Estes dados mostram que as GRPs podem constituir marcadores e/ou modelos interessantes para a compreensão de distintos aspectos da biologia vegetal. Neste trabalho, características estruturais e funcionais deste tipo de proteínas em cana-de-açúcar (*Saccharum officinarum* L.) foram analisadas. Uma vez que esta é a primeira descrição deste tipo de proteínas em cana-de-açúcar, especial atenção foi dada para o padrão de expressão destes genes, analisando-se a abundância e prevalência de cada um dos genes nas diferentes bibliotecas de cDNA do projeto Sequenciamento de ESTs de cana-de-açúcar (SUCEST). A comparação das GRPs de cana-de-açúcar com as GRPs descritas em outras espécies também será discutida.

## ACKNOWLEDGMENTS

A. Fusaro is recipient of a doctoral fellowship from the Brazilian agency CAPES. A. Mangeon, C.A.B. Rocha, R.M. Junqueira and T.C. Coutinho were supported by fellowships from the Brazilian agencies CNPq (CNPq-PIBIC). The Authors thank to Dr Maite F.S. Vaslin for the critical reading of the manuscript and to FAPESP for the support and development of the SUCEST project. This work was supported by a grant from the 'young state scientist' (*Cientista Jovem do Estado*) program from the Rio de Janeiro state agency FAPERJ.

## REFERENCES

- Albà, M.M., Culiñez-Macià, F.A., Goday, A., Freire, M.A., Nadal, B. and Pagès, M. (1994). The maize RNA-binding protein, MA16, is a nucleolar protein located in the dense fibrillar component. *Plant J.* 6: 825-834.
- Alberts, B. (1998). The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell* 92: 291-294.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25: 3389-3402.
- Condit, C.M. (1993). Developmental expression and localization of petunia glycine-rich protein 1. *Plant Cell* 5: 277-288.
- Condit, C.M. and Meagher, R.B. (1986). A gene encoding a novel glycine-rich structural protein of petunia. *Nature* 323: 178-181.
- de Oliveira, D.E., Franco, L.O., Simoens, C., Seurink, J., Coppieters, J., Botterman, J. and van Montagu, M. (1993). Inflorescence-specific genes from *Arabidopsis thaliana* encoding glycine-rich proteins. *Plant J.* 3: 495-507.
- Dunn, M.A., Brown, K., Lightowers, R. and Hughes, M.A. (1996). A low temperature responsive gene from barley encodes a protein with single stranded nucleic acid binding activity which is phosphorylated in vitro. *Plant Mol. Biol.* 30: 947-959.
- Ferreira, M.A., Almeida-Engler, J., Miguens, F.C., van Montagu, M., Engler, G. and de Oliveira, D.E. (1997). Oleosin gene expression in *Arabidopsis thaliana* coincides with accumulation of lipids in plastids and cytoplasmic bodies. *Plant Physiol. Biochem.* 35: 729-739.
- Freire, M.A. and Pages, M. (1995). Functional characterization of the maize RNA binding protein MA16. *Plant Mol. Biol.* 29: 797-807.
- Goddemeier ML, Wulff D and Feix G (1998). Root-specific expression of a *Zea mays* gene encoding a novel glycine-rich protein, zmGRP3. *Plant Mol Biol* 36: 799-802.
- Gómez, J., Sánchez-Martínez, D., Stiefel, V., Rigau, J., Puigdomènech, P. and Pagès, M. (1988). A gene induced by the plant hormone abscisic acid in response to water stress encodes a glycine-rich protein. *Nature* 334: 262-264.
- Hanano, S., Sugita, M. and Sugiura, M. (1996). Isolation of a novel RNA-binding protein and its association with a large ribonucleoprotein particle present in the nucleoplasm of tobacco cells. *Plant Mol. Biol.* 31: 57-68.
- Heintzen, C., Melzer, S., Fischer, R., Kappeler, S., Apel, K. and Staiger, D. (1994). A light- and temperature-entrained circadian clock controls expression of transcripts encoding nuclear proteins with homology to RNA-binding proteins in meristematic tissue. *Plant J.* 5: 799-813.
- Hirose, T., Sugita, M. and Sugiura, M. (1993). cDNA structure, expression and nucleic acid-binding properties of three RNA-binding proteins in tobacco: occurrence of tissue-specific alternative splicing. *Nucleic Acids Res.* 21: 3981-3987.
- Keller, B., Schmid, J. and Lamb, C.J. (1989). Vascular expression of a bean cell wall glycine-rich protein -  $\beta$ -glucuronidase gene fusion in transgenic tobacco. *Embo J.* 8: 1309-1314.
- Kumar, S., Tamura, K., Jacobsen, I. and Nei, M. (2000). MEGA2: Molecular Evolutionary Genetics Analysis, ver-

- sion 2.0. Pennsylvania and Arizona State Universities, University Park, Pennsylvania and Tempe, Arizona.
- Ludevid, M.D., Freire, M.A., Gómez, J., Burd, C.G., Albericio, F., Giral, E., Dreyfuss, G. and Pagès, M.** (1992). RNA binding characteristics of a 16kDa glycine-rich protein from maize. *Plant J. 2*: 999-1003.
- Mariane, C., de Beuckeleer, M., Truettner, J., Leemans, J. and Goldberg, R.B.** (1990). Induction of male sterility in plants by a chimeric ribonuclease gene. *Nature 347*: 737-741.
- Ni, Z., Sun, Q., Liu, Z., Wu, J.L. and Wang, X.** (2000). Identification of a hybrid-specific expressed gene encoding novel RNA-binding protein in wheat seedling leaves using differential display of mRNA. *Mol. Gen. Genet. 263*: 934-938.
- Obokata, J., Ohme, M. and Hayashida, N.** (1991). Nucleotide sequence of a cDNA clone encoding a putative glycine-rich protein of 19.7 kDa in *Nicotiana sylvestris*. *Plant Mol. Biol. 17*: 953-955.
- Park, A.R., Cho, S.K., Yun, U.J., Jin, M.Y., Lee, S.H., Sachetto-Martins G. and Park, O.K.** (2001). Interaction of the *Arabidopsis* Receptor Protein Kinase Wak1 with a Glycine-rich Protein, AtGRP-3. *J. Biol. Chem. 276*: 26688-26693.
- Reddy, A.S.N. and Poovaiah, B.W.** (1987). Accumulation of a glycine rich protein in auxin- deprived strawberry fruits. *Biochem. Biophys. Res. Commun. 147*: 885-891.
- Ryser, U. and Keller, B.** (1992). Ultrastructural localization of bean glycine-rich protein in un lignified primary walls of protoxylem cells. *Plant Cell 4*: 773-783.
- Sachetto-Martins, G., Fernandes, L.D., Felix, D. and de Oliveira, D.E.** (1995). Preferential Transcriptional activity of a glycine-rich protein gene from *Arabidopsis thaliana* in pro-toderm derived cells. *Int. J. Plant Science. 156*(4): 460-470.
- Sachetto-Martins, G., Franco, L.O. and de Oliveira, D.E.** (2000). Plant glycine-rich proteins: a family or just proteins with a common motif? *BBA 1492*: 1-14.
- Sandal, N.N., Bojsen, K., Richter, H., Sengupta-Gopalan, C and Marcker, K.A.** (1992). The nodulin 24 protein family shows similarity to a family of glycine-rich plant proteins. *Plant Mol. Biol. 18*: 607-610.
- Sitnikova, T., A. Rzhetsky and M. Nei.** (1995). Interior-branch and bootstrap tests of phylogenetic trees. *Molecular Biology and Evolution 12*: 319-333.
- Steinert, P.M., Mack, J.W., Korge, B.P., Gan, S-Q., Haynes, S.R. and Steven, A.C.** (1991). Glycine loops in proteins: their occurrence in certain intermediate filament chains, loricrins and single-stranded RNA-binding proteins. *Int. J. Biol. Macromol. 13*: 130-139.
- Ting, J.T.L., Wu, S.S.H., Ratnayake, C. and Huang, A.H.C.** (1998). Constituents of the tapetosomes and elaioplastids in *Brassica campestris* tapetum and their degradation and retention during microsporogenesis. *Plant J. 16*: 541-551.
- Thompson, J.D., Higgins, D.G. and Gibson, T.J.** (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Research, 22*: 4673-4680.
- Varner, J.E. and Cassab, G.I.** (1986). A new protein in petunia. *Nature 323*: 110.