0273-2289/05/120/0169/$30.00

Glycine-Rich Proteins

A Class of Novel Proteins

Amir Mousavi*,[1] and Yasuo Hotta[2]

*[1]Plant Molecular Biology Division,
National Research Center for Genetic Engineering and Biotechnology,
PO Box 14155-6343, Tehran, Iran, E-mail: m-amir@nrcgeb.ac.ir;
and [2]Niigata University of Health and Welfare, 1368 Shimami,
Niigata 950-3198, Japan*

**Received May 7, 2004; Revised October 20, 2004;
Accepted October 27, 2004**

Abstract

Glycine-rich proteins (GRPs) containing more than 60% glycine have been found in different tissues from many eukaryotic species. Despite the availability of literature on different groups of GRPs, there are few reports in which they are all considered and compared together. Some of these proteins are components of the cell walls of many higher plants. In most cases, it has been shown that they are accumulated in the vascular tissues and that their synthesis is part of the plant's defense mechanism. Other distinct types of GRPs are characterized by having structures and functions similar to animal cytokeratins or by a domain with typical RNA-binding motifs. The availability of cloned GRP genes facilitates the study of the function of this diverse class of proteins, which is expected to enhance the understanding of cell physiology.

Index Entries: Cell wall proteins; cytokeratin-like proteins; glycine-rich protein (GRP); stress response; RNA-binding proteins.

Introduction

Glycine-rich proteins (GRPs) are a group of proteins characterized by a high content and repetitive sequences of glycine residues based on (Gly-X)n motifs that are usually found in β-plated sheets with antiparallel strands or form flexible coiled structures. The GRPs isolated so far fall into three known classes: structural proteins in the cell wall *(1–3)*; RNA-binding proteins, or RNA-GRPs *(4–6)*; and cytokeratin-like proteins, or CL-GRPs *(7)*.

*Author to whom all correspondence and reprint requests should be addressed.

Plant Cell Wall GRPs

Most of the GRPs known to date have been found in the cell walls of many higher plants and form a third group of structural protein components of the wall, the other two being extensins and proline-rich proteins (for a review, see ref. *8*). Cell wall proteins are presumed to provide elasticity to the cell wall necessary for cell growth by extension.

The first report introducing cell wall GRPs was published in 1986 in *Nature* by Condit and Meagher *(9)*. In this article, they reported the isolation and sequencing of an expressed gene from petunia encoding a GRP (GRP1). At that time, extensins or hydroxyproline-rich glycoproteins (HRGPs) were the only well-characterized cell wall structural proteins of plants that had been isolated. However, it was clear that HRGPs could not be the only cell wall structural protein of plants, because some plant species and organs contain very little hydroxyproline in their cell walls but, instead, contain large amounts of glycine. This suggested that some plant cell walls may contain primarily glycine-rich rather than hydroxyproline-rich structural proteins.

Cell wall GRPs are often associated with the vascular system of plants, and their synthesis appears to be regulated during development *(10)*. There is evidence demonstrating the localization of the GRPs in French bean vascular tissue *(1)* and, more precisely, in all lignified cells and protoxylem of soybean tissues *(11,12)*. It is generally believed that these proteins are expressed in the vascular system and presumably located in the cell wall.

Structural GRPs in plant cell walls have been described in petunia *(9)*, Arabidopsis *(3)*, barley *(7)*, and rice *(13)*, in most of which their expression is regulated developmentally. Additionally, the expression of many of GRPs seems to be regulated by external stimuli ranging from pathogen infection to several forms of environmental stress such as light *(14)*, NaCl *(15)*, abscisic acid, and water *(16)*.

In loblolly pine (*Pinus taeda*), a water deficit-inducible GRP gene that is expressed almost exclusively in roots has been cloned *(17)*. The open reading frame in the cloned gene, named LP5, encodes a polypeptide of 194 amino acids that is also rich in serine (44% Gly and 20% Ser). The putative LP5 protein shows great similarity to silk fibroin *(18)* and the 14.7-kDa LIM14 protein of lily *(19,20)*. The striking similarities of LP5 to specialized animal structural proteins suggest that LP5 may play a similar role in mediating the elasticity and strength of the cells.

Ueki and Citovsky *(21)* identified a tobacco GRP, cdiGRP, specifically induced by a low concentration of cadmium and expressed in the cell walls of plant vascular tissues. Constitutive cdiGRP expression inhibits systemic transport of turnip vein-clearing tobamovirus (TVCV), whereas suppression of cdiGRP production allows TVCV movement in the presence of cadmium.

In general, the synthesis of cell wall proteins is stimulated under stress conditions. It seems that GRPs exert their inhibitory effect on stress elements by enhancing cell wall components and callus deposition. GRPs are proposed to be the part of a defense or repair system of the plants; however, their mechanism of action at the molecular level is still not clear.

More recently, plant GRP genes have been isolated from germ cells. Three stamen-specific genes of *Arabidopsis thaliana* are expressed predominantly in anthers at the later stage of flower development *(22)*. The primary structures of the encoded gene products have extensive hydrophobic N-terminal domains, suggesting that they are stored within the endomembrane system. Like other plant GRPs, these proteins are composed of tandemly repeated sequences. However, similar glycine-rich repeats were not found in a protein sequence data bank. Furthermore, instead of the characteristic antiparallel β-plated sheet predicted for the structure of cell wall GRPs, the glycine-rich repeats of the flower-specific Arabidopsis GRPs seem to form flexible coiled structures. These genes are currently being analyzed for tissue-specific expression, location of their encoded proteins, and their function in anther development.

Nadeau et al. *(23)* introduced an ovule-specific cDNA clone (O126) in phalaenopsis orchid that encodes a GRP distinct from previously reported sequences. O126 (PGRP-1) is expressed exclusively in ovule tissue at 11 wk after pollination. Because there is no vascularization in the phaleonopsis ovule and no expression of PGRP-1 in the vascularized tissues of the plant, it is unlikely that the O126 protein has tissue specificity similar to other GRPs. It is possible that O126 is a component of specialized cell walls in the ovule.

LIM14 was also identified as a glycine- and serine-rich protein with characteristic (S-G-X-G)n repeats expressed in anther wall and young microspores of *Lilium longiflorum (20)*. The putative LIM14 protein, however, is distinct from previously characterized GRPs and shows similarity to silk fibroin and cytokeratin in mammals, which are thought to fulfill structural roles *(18)*. Immunofluorescence microscopy and *in vitro* analysis of the reporter gene-tagged proteins identified a plastid-linked localization for the LIM14 protein. It is predicted that LIM14 is an anther-specific protein that may play a role in starch accumulation and amyloplast differentiation and integrity during anther development. The candidate homologous genes for *LIM14* have also been identified in Arabidopsis, which could lead to a greater understanding of the function of the protein during microsporogenesis (unpublished).

RNA–GRPs

Other plant GRPs may have different functions and localizations. In the past several years, a novel class of eukaryotic proteins has been identified whose members have been implicated in a number of diverse cellular processes requiring RNA recognition. A common structural feature of many of these proteins is an approx 90 amino acid domain surrounding a highly conserved octapeptide (RGFGFVXF) termed the ribonucleoprotein consensus sequence, or RNP-CS *(24)*. This putative RNA-binding domain (RBD) may be present in multiple copies in certain proteins and may be appended to regions highly enriched in acidic amino acids or glycine *(25)*. The RNP-CS has been proposed to be required for RNA

recognition whereas other domains, such as those rich in glycine, are thought to interact with other components required for RNA processing *(26)*. To date, several cDNAs encoding putative RNP-CS-type RNA-binding proteins have been isolated from plants, including two from *Sorghum vulgaris (4)*, an abscisic acid-induced transcript from *Zea mays (27)*, and a wound- and jasmonate-induced gene of *Picea glauca (28)*. All of the proteins encoded by these genes consist of a carboxyl-terminal glycine-rich domain appended directly to a single RNP-CS-type binding domain. These proteins are distinct from members of the cell wall group of GRPs in plants that lack the conventional RNP-CS domain. van Nocker and Vierstra *(5)* reported the isolation of two cDNAs from Arabidopsis encoding GRPs that showed homology to RNA-binding proteins containing the RNP-CS-type RNA-binding domain. They are distinct from the six GRPs (*AtGRP1–6*) previously identified in Arabidopsis *(3)* and, hence, were designated *AtGRP7* and *AtGRP8*. By analogy with other RNA-binding proteins, the bipartite structure of *AtGRP7* and *AtGRP8* suggests that the RNP-CS domain is required for RNA recognition whereas the glycine-rich domain may facilitate interprotein interactions.

RNP-CP-type RNA-binding proteins from other eukaryotes have a complex structure, in many cases containing tandem repeats of the RBD appended on either the carboxyl- or amino-terminal side by long acidic tracts and/or tracts enriched in glycine *(25)*. Structurally, the plant glycine-rich RNP-CS-containing proteins appear to be the simplest of this class, having a single putative RBD appended directly to the glycine-rich tract. RNA-binding GRPs specifically bind poly(U) and poly(G) *(29)*, have high affinity for the RNA of the same cells in which they are present *(30)*, and are located in the nucleolar compartment *(31)*.

Given their possible role in RNA recognition and processing, plant GRPs with an RNA-binding capacity may have important roles in plant cell physiology. The availability of their cloned genes makes investigating their function possible, which, in turn, may facilitate the understanding of RNA biogenesis.

CL–GRPs

GRPs are not restricted to plants. In animal cells, quasirepetitive glycine-rich sequences are widespread in several distinct families of proteins. Such sequences have been proposed to comprise a new structural motif termed *glycine-loop*, which is expected to be highly flexible, without a unique defined conformation *(26)*. The flexibility may be the most important feature of the glycine-rich domains.

Interestingly, some of the plant GRPs, as shown for the barley protein HvGRP1 *(7)*, are highly related with respect to sequence (glycine stretches with interspersed tyrosine residues) and structural arrangement (β-conformation) to the glycine-rich repeat regions flanking the α-helical central core of epidermal keratins such as that of the 59-kDa mouse cytokeratin *(32)*.

The presence of a putative signal peptide suggests a cell wall location for HvGRP1 similar to that of some other GRPs. Epidermal keratins are a constituent of the 8- to 12-nm intermediate filaments of epithelial cells that form the protective layer (epidermis) of the animal body *(32)*. Expression of the keratins is regulated during epithelial differentiation, and the N- and C-terminal glycine-rich domains are not only necessary for filament formation *(33)*, but apparently also for the formation of an insoluble structure created by interaction with epidermal matrix proteins *(32)*. It is concluded that plant GRPs and animal cytokeratins are possibly related to each other with respect to both structure and function.

In the past two decades, a set of GRPs of plants was characterized and the new groups of GRPs were identified. These proteins, although all sharing the feature of having glycine-rich regions, are associated with various independent physiologic pathways. The diverse expression profile of these proteins in different cells and under different conditions identifies them as attractive model molecules or markers for molecular cell biology.

Acknowledgment

We thank Prof. Elahe Elahi for critically reading the manuscript.

References

1. Keller, B., Templeton, M. D., and Lamb, C. J. (1989), Proc. *Natl Acad. Sci.* USA **86,** 1529–1533.
2. Condit, C. M. and Meagher, R. B. (1990), *Plant Physiol.* **93,** 596–606.
3. de Oliveira, D. E., Seurinck, J., Inze, D., Van Montagu, M., and Botterman, J. (1990), *Plant Cell* **2,** 427–436.
4. Cretin, C. and Puigdomenech, P. (1990), *Plant Mol. Biol*. **15,** 783–785.
5. van Nocker, S. and Vierstra, R. D. (1993), *Plant Mol. Biol*. **21,** 695–699.
6. Aneeta, Sanan-Mishra, N., Tuteja, N., and Kumar Sopory, S. (2002), *Biochem. Biophys. Res. Commun*. **6,** 1063–1068.
7. Rohde, W., Rosch, K., Kroger, K., and Salamini, F. (1990), *Plant Mol. Biol*. **14,** 1057–1059.
8. Showalter, A. M., Kieliszewisky, M., Cheung, A., and Tierney, M. (1996), *Plant Mol. Biol. Rep*. **14,** 9, 10.
9. Condit, C. M. and Meagher, R. B. (1986), *Nature* **323,** 178–181.
10. Ringli, C., Keller, B., and Ryser, U. (2001), *Cell. Mol. Life Sci*. **58**, 1430–1441.
11. Ye, Z.-H. and Varner, J. E. (1991), *Plant Cell* **3,** 23–37.
12. Ryser, U. (2003), *Planta* **216,** 854–64.
13. Fang, R. X., Pang, Z., Gao, D. M., Mang, K. G., and Chua, N. H. (1991), *Plant Mol. Biol*. **17,** 1255–1257.
14. Kaldenhoff, R. and Richter, G. (1989), *Nucleic Acids Res*. **17,** 2853.
15. Tang, Y. X., Xia, G. X., and Liu, SG. (2002), *Sheng Wu Hua Xue Yu Sheng Wu Wu Li Xue Bao (Shanghai)* **34,** 737–742.
16. Didierjean, L., Frendo, P., and Burkard, G. (1992), *Plant Mol. Biol*. **18,** 847–849.
17. Chang, S., Puryear, J. D., Dias, M. A. D. L., Funkhouser, E. A., Newton, R. J., and Cairney, J. (1996), *Physiol. Plant* **97,** 139–148.
18. Mita, K., Ichimura, S., Zama, M., and James, T. C. (1988), *J. Mol. Biol*. **203,** 917–925.
19. Kobayashi, T., Kobayashi, E., Sato, S., Hotta, Y., Miyajima, N., Tanaka, A., and Tabata, S. (1994), *DNA Res*. **1,** 15–26.

20. Mousavi, A., Hiratsuka, R., Takase, H., Hiratsuka, K., and Hotta, Y. (1999), *Plant Cell Physiol*. **40,** 406–416.
21. Ueki, S. and Citovsky, V. (2002), *Nat. Cell. Biol*. **4,** 478–486.
22. de Oliveira, D. E., Franco, L. O., Simoens, C., Seurinck, J., Coppieters, J., Botterman, J., and Van Montagu, M. (1993), *Plant J*. **3,** 495–507.
23. Nadeau, J. A., Zhang, X. S., Li, J., and O'Neill, S. D. (1996), *Plant Cell* **8,** 213–239.
24. Dreyfuss, G., Swanson, M. S., and Pinol-Roma, S. (1988), *Trends Biochem. Sci*. **13,** 86–91.
25. Bandzilius, R. J., Swanson, M. S., and Dreyfuss, G. (1989), *Genes Dev*. **3,** 431–437.
26. Steinert, P. M., Mack, J. W., Korge, B. P., Gan, S. Q., Haynes, S. R., and Steven, A. C. (1991), *Int. J. Biol. Macromol*. **13,** 130–139.
27. Gomez, J., Sanchez-Martinez, D., Stiefel, V., Rigau, J., Puigdomenech, P., and Pages, M. (1988), *Nature* **334,** 262–264.
28. Richard, S., Drevet, C., Jouanin, L., and Seguin, A. (1999), *Gene* **240,** 379–388.
29. Ludevid, M. D., Freire, M. A., Gomez, J., Burd, C. G., Albericio, F., Giralt, E., Dreyfuss, G., and Pages, M. (1992), *Plant J*. **2,** 999–1003.
30. Hirose, T., Sugita, M., and Sugiura, M. (1994), *Mol. Gen. Genet*. **244,** 360–366.
31. Alba, M. M., Culianez-Macia, F. A., Goday, A., Freire, M. A., Nadal, B., and Pages, M. (1994), *Plant J*. **6,** 825–834.
32. Steinert, P. M., Jones, J. C., and Goldman, R. D. (1984), *J. Cell Biol*. **99,** 22s–27s.
33. Steinert, P. M., Steven, A. C., and Roop, D. R. (1985), *Cell* **42,** 411–420.