Phylogenetic analysis of the hybrid proline-rich protein families from seven plant species and expression profiles of *HyPRP* genes in potato

Lenka Dvořáková*, Zdeněk Opatrný and Lukáš Fischer**

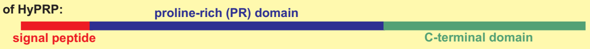
Department of Plant Physiology, Faculty of Science, Charles University in Prague, Viničná 5, 128 44 Prague 2, Czech Republic, *lenkadvorakova@centrum.cz, **lukasf@natur.cuni.cz

Introduction

Hybrid proline rich proteins (HyPRPs) create a subgroup of structural cell wall proteins rich in proline (PRPs). HyPRPs are composed of a hydrophobic **signal peptide** and two distinct domains: **proline-rich domain** and **C-terminal domain**. Repetitive proline-rich domains are very variable in respect of the length and amino acid composition. Hydrophobic C-terminal domains have similar length and are characterized by identical pattern of eigth cystein residues. Similar arrangement of cysteins is in sequences of structurally conserved non-specific lipid transfer proteins (ns-LTPs), whose structure has been described in detail: four disulphide bridges created between eigth conserved cystein residues stabilize four -helixes. A hydrophobic pocket occurs inside this structure (Kader 1997). Since disulphide bridges are important for stabilization of three-dimensional structure of proteins, the specific pattern of cysteins usually gives evidence of structural similarity (Josè-Estanyol and Puigdomènech 2000).

Here we present analysis of all available sequences of HyPRPs from seven plant species and characterization of expression profiles of *HyPRP* genes in potato.

Schema of HyPRP:



Aim

To find for possible functional specialization of HyPRPs

- **by phylogenetic analysis of C-terminal domains of HyPRPs from seven plant species**
- **by studying of expression profiles of *HyPRP* genes in potato organs**
- **by phylogenetic analysis of LTPs and C-terminal domains of HyPRPs from potato**

Results

Phylogenetic analysis of C-terminal domains of HyPRPs from seven plant species

HyPRPs create wide families in all studied plant species (approximately 10-30 members) with very variable proline-rich domains. Majority of studied proteins had short (or very short) proline-rich domain, few proteins had long proline-rich domain, low number of proteins had no proline-rich domain. In some proteins proline-rich domain was substituted by glycine-rich domain (short or long) and some proline-rich domains have also high content of glycine. With respect to proline-rich domains diversity, sequences of conserved C-terminal domains were used for phylogenetic analysis.

Legend:

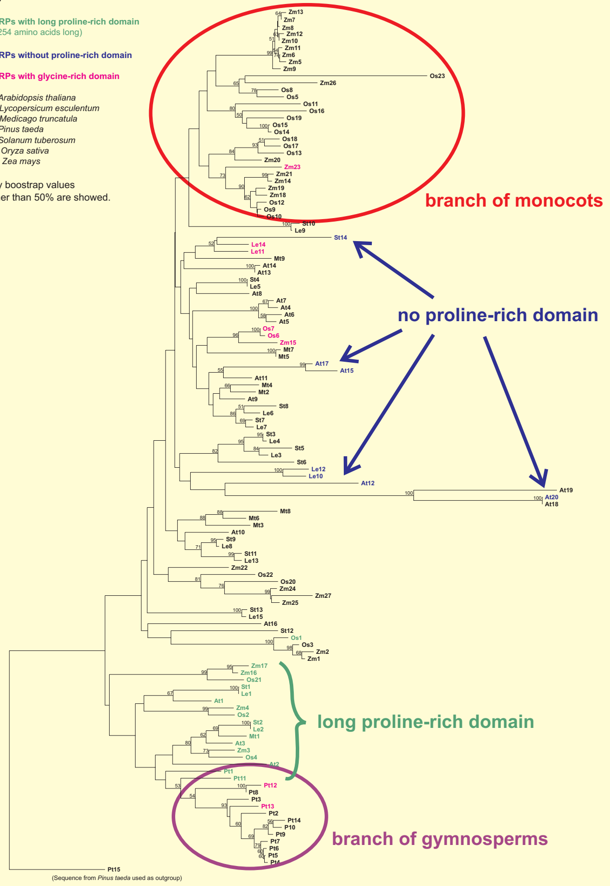
HyPRPs with long proline-rich domain
(86-254 amino acids long)

HyPRPs without proline-rich domain

HyPRPs with glycine-rich domain

At - *Arabidopsis thaliana*
Le - *Lycopersicum esculentum*
Mt - *Medicago truncatula*
Pt - *Pinus taeda*
St - *Solanum tuberosum*
Os - *Oryza sativa*
Zm - *Zea mays*

Only boostrap values higher than 50% are showed.



(Sequence from *Pinus taeda* used as outgroup)

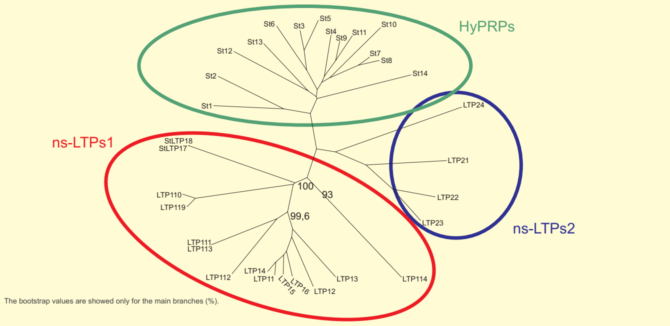
Main branches of the phylogenetic tree had very low bootstrap values, therefore the main dendrogram structure was not sufficiently robust. On the contrary, distal branches had often high bootstrap values even in cases of proteins with highly variable proline-rich domains.

Sequences from monocots and gymnosperms were partially separated probably due to independent evolution of *HyPRP* genes after diversification of these groups or different composition of the cell wall.

HyPRPs with long proline-rich domains were probably more ancestral than HyPRPs with short proline-rich domains (note the branch of long proline-rich domains and gymnosperms).

Different HyPRPs probably lost proline-rich or change it for glycine-rich domain independently.

Phylogenetic tree of LTPs and C-terminal domains of HyPRPs from potato



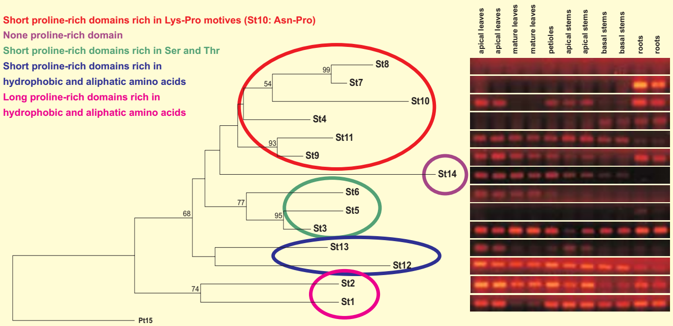
The bootstrap values are showed only for the main branches (%).

Though the structure of LTPs and C-terminal domains of HyPRPs is supposed to be similar, the branches of these families have high bootstrap values, making the dendrogram structure highly significantr.

St14 without proline-rich domain (thus resembling LTP) clustered to HyPRPs.

Expression profiles of *HyPRP* genes in potato organs

Short proline-rich domains rich in Lys-Pro motives (St10: Asn-Pro)
None proline-rich domain
Short proline-rich domains rich in Ser and Thr
Short proline-rich domains rich in hydrophobic and aliphatic amino acids
Long proline-rich domains rich in hydrophobic and aliphatic amino acids



(Sequence from *Pinus taeda* used as outgroup)

In the phylogenetic tree there are showed bootstrap values higher than 50%.

The expression profiles of *HyPRP* genes varied and were often overlapping. Some genes had complementary expression profiles.

Genes from all branches of phylogenetic tree (except the St14 branch) were expressed in every potato organ.

Each branch of the phylogenetic tree of C-terminal domains was characterized by similar proline-rich domains.

Similar results were obtained for *Arabidopsis thaliana* using Genevestigator.

Conclusions

Functional specialization of C-terminal domains was not confirmed

- **Sequences of C-terminal domains developed most likely by random changes of amino acid residues limited by requirement of structure conservation. Proline-rich domains developed independently.**
- **Concurrent expression of *HyPRP* genes with all different types of proline-rich domains might be essential for structure or function of every vegetative potato organ.**
- **Sequences of LTPs and C-terminal domains of HyPRPs belong to separate branches of the phylogenetic tree, documenting different function of these proteins families.**

Material and methods

Sequences used for analyses (genomic or ESTs: PUT - PlantGDB-assembled Unique Transcripts or unigenes) were searched by on-line application TBlastn with SbrPRP from *Solanum brevidens* (Fischer *et al.* 2002) as a query (http://arabidopsis.org, www.sgn.cornell.edu, www.plantgdb.org). Only sequences coding for whole potential protein with signal peptide, with or without proline-rich domain and with C-terminal domain and STOP codon were used for analysis. Phylogenetic trees were constructed from multiple alignments by Treecon (tree of C-terminal domains from seven plant species and from *Solanum tuberosum*; Van de Peer and De Wachter 1994) and ClustalX (tree of sequences of LTPs and HyPRPs from potato; Thompson *et al.*1997). Characteristic features of sequences (proline-rich domains without signal peptide) were counted by ProtParam (Gaisteiger *et al.* 2005, www.expasy.org) and signal peptides were predicted by Signal P (Bendtsen *et al.* 2004, www.expasy.org).

The expression of 14 *HyPRP* genes from potato was studied in 6 different organs via semikvantitative RT-PCR with primers specifically designed to every sequence. In all organs there was expression of *HyPRP* genes assessed in two parallels (independent isolation of RNA, reverse transcription and PCR). For reverse transcription was used primer oligo-T (23 nucleotides) and RevertAidTM M-MuLV Reverse Transcriptase (Fermentas) with appropriate buffer and RiboLockTM Ribonuclease inhibitor (Fermentas). For PCR was used 2 l of reverse transcription reaction as template and *Taq* DNA Polymerase (Fermentas). Primers for PCR were specific designed to every *HyPRP* sequence and had melting temperature 62°C computed with formula: (number of ATx2) + (number of GCx4). The primers allowed amplification of parts of C-terminal domains. Following program was used for PCR: starting denaturation 94°C for 3 minutes and 23 cycles with denaturation by 94°C for 30 seconds, annealing by 60°C for 45 seconds, polymeration by 72°C for 20 seconds. Elongation factor ef1 was used for standardization of reverse transcription reaction. The expression of two parallels was usually equal, but in some cases differed also when PCR was repeated and the signal of internal standard was identical. For electrophoresis was used 1,5% agarose gel in 0,5x TAE buffer and 10 l of each PCR reaction.

Acknowledgement

This work was supported by the grant from Grant Agency of Charles University, Czech Republic (149/2006/B-BIO/PrF).

References

Bendtsen, J. D., Nielsen, H., von Heijne, G., Brunak, S. Journal of Molecular Biology 340: 783-795, 2004.
Fischer, L., Lovas, A., Opatrný, Z., Bánfalvi, Z. Journal of Plant Physiology 159: 1271-1275, 2002.
Gaisteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., Bairoch, A. In: Walker, J. M. (ed.): The Proteomics Protocols Handbook. Humana Press, 571-607, 2005.
Josè-Estanyol, M., Puigdomènech, P. Plant Physiology and Biochemistry 38: 97-108, 2000.
Kader, J. C. Trends in Plant Science 2: 66-70, 1997.
Nicot, N., Hausman, J.-F., Hoffmann, L., Evers, D.
Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., Higgins, D. G. Nucleic Acids Research 24: 4876-4882, 1997.
Van de Peer, Y., De Wachter, R. Biochemistry 10: 569-570, 1994.