The peroxidase gene family of *Arabidopsis thaliana*

Simon, Patrice; Capelli, Nicolas; Flach, Jean; Overney, Serge; Tognolli, Michael; Penel, Claude; Greppin, Hubert

Plant Peroxidases: Biochemistry and Physiology,
C. Obinger, U. Burner, R. Ebermann, C. Penel, H. Greppin, eds.
University of Geneva, 1996

THE PEROXIDASE GENE FAMILY OF *ARABIDOPSIS THALIANA*

P. Simon, N. Capelli, J. Flach, S.Overney, M.Tognolli, C. Penel and H. Greppin

Laboratory of Plant Biochemistry and Physiology, University of Geneva,
place de l'Université 3, CH-1211 Genève 4, Switzerland

INTRODUCTION

The classical guaiacol peroxidase in plants is represented by numerous true or pseudo-isozymes, the expression of which obeys spatio-temporal regulation. There has been interest for a long time to purify each specific isoform in order to evaluate the intrinsic biochemical properties that would allow a better assessment of the putative physiological role of these diverse and omnipresent enzymes. Efforts to purify peroxidases are limited by the available amounts and can also be baffled by severe loss occurring during purification. These efforts have therefore often been concentrated on the highly expressed and also purification docile isoforms. Facing this kind of difficulties we have developped a strategy a few years ago to obtain sufficient amount of refractory and minoritary peroxidases. The strategy we follow is the isolation of molecular cDNA clones and the heterologous expression in an appropriate system, such as previously the *Xenopus* oocytes and now the baculovirus/insect cells. These eukaryotic expression systems are reputated to perform efficient expression and adequate posttranslational modifications. We were indeed able to produce catalytically active and properly maturated spinach isoforms in *Xenopus* oocytes (1,2). We have recently switched to the baculovirus system, which worked efficiently in the expression of the synthetic gene of HRP C (3). We have also chosen the model plant *Arabidopsis thaliana* in the frame of a joint EU-project and we have committed ourselves to the systematic screening for Arabidopsis cDNAs encoding isoperoxidases, which will be used in the next step for heterologous expression in the baculovirus systems. The results obtained from this screening are presented here.

MATERIAL AND METHODS

In a first step we have screened a cDNA library available at the ABRC (Arabidopsis Biological Resource Center at the Ohio State University), the cDNA library of T. Newman. The library was screened with the same probe derived from the conserved catalytical site that we used successfully with spinach (unpublished results). With the sudden explosion of EST clones (from the Expressed Sequence Tag sequencing projects) and their availability at the ABRC and also from individual French Labs, we also searched for peroxidase EST clones with available computer program facilities at the Center server (4). A virtual screening of the EST databases with the conserved catalytical site probe yielded partially the same results. The candidates were sequenced manually with Sequenase (Amersham). Sequence data were analyzed with the programs of the Wisconsin software package (5).

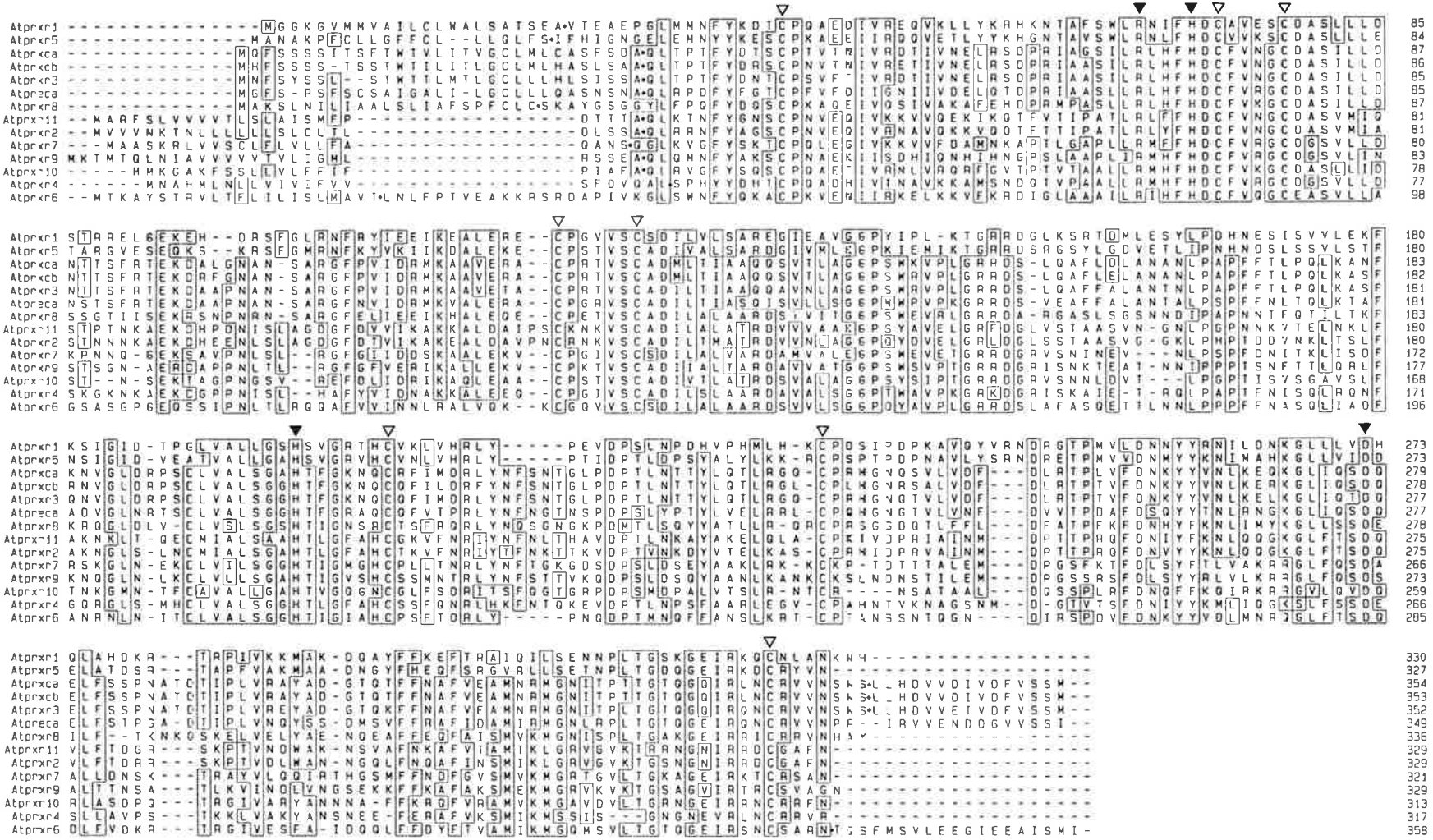

Figure 1. Alignment of the aminoacid encoded sequences of the 11 new peroxidases and of the previously known peroxidases in Arabidopsis. Explanations in the text.

RESULTS AND DISCUSSION

Arabidopsis sequences

Partial screening of the cDNA library yielded about 50 positive clones. Around thirty clones were clearly identified as peroxidases by partial sequencing at the 5' end and could be assigned to five different peroxidases. One of them revealed to be a cloning artefact and another one to an already known *Arabidopsis* gene. Three representative clones of the remaining types have been completely sequenced. More than 150 clones were found in thc databank non exhaustive

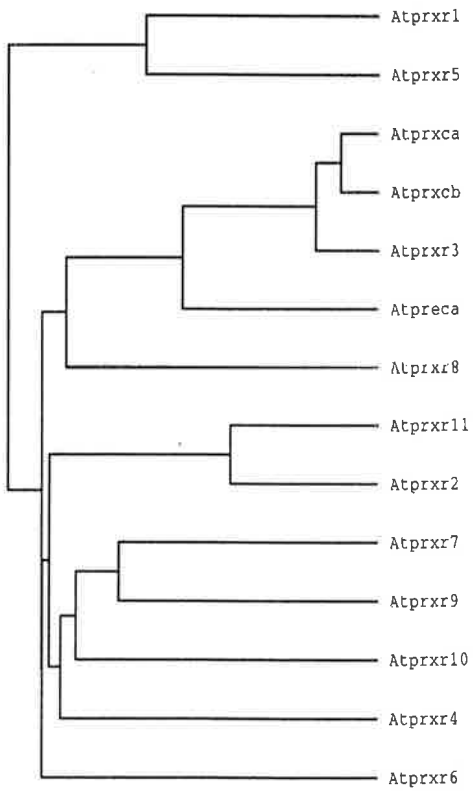

Figure 2. Level of similarity between the Arabidopsis peroxidases.

screening, from which it was possible to select 14 new peroxidases. Eight of these clones have been sequenced. In summary, 19 different peroxidase clones, clearly distinct from each other and representing not only allelic variants, were retrieved from this cloning abundancy, and 11 of them have been now fully sequenced. The alignment of the 11 encoded aminoacid sequences is shown in figure 1 together with the three already known Arabidopsis sequences (6). Region of similarity are boxed. The naming of the sequences is arbitrary, r stands for RNA and the numbering reflects the chronological order of determination. A peroxidaser expert eye will easily recognize the typically conserved residues, i.e. the 8 Cys involved in disulfide bridges, the proximal and distal His residues of the catalytical site, and the Arg and Asp residues also involved in catalysis, all marked with a triangle (7). The predicted signal peptide cleavage site is indicated by a dot. Two of the eleven peroxidases have a C-terminal extension and the cleavage site has been estimated by analogy to experimentally determined cases (see 7) and is also punctuated by a dot. The degree of homology between the various Arabidopsis peroxidases ranges from 31% to 89 % identity (52 to 92 % similarity), excluding the close relation between Atprxca and Atprxcb peroxidases. The level of similarity is illustrated in Figure 2. This relatively low degree of similarity between various peroxidases from the same plant has already been observed in spinach (1) and can be related to random cloning and analysis in opposition to

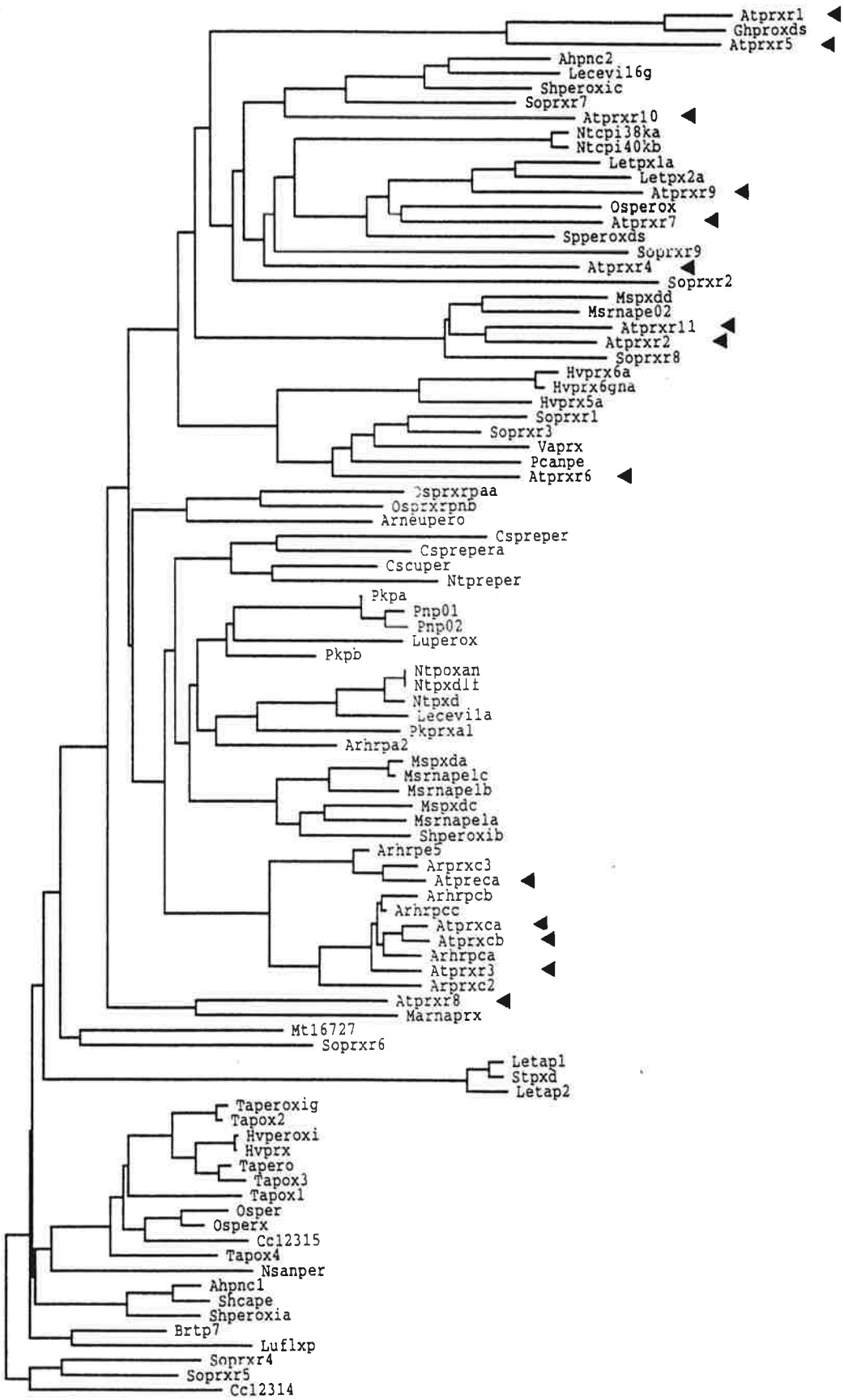

Figure 3. Evolutionary tree of plant peroxidases. The names are the entry names at the EMBL databank, with the exception of Arhrpa2 and Arhrpe5 which are not yet available in databanks and Brtp7 which corresponds to Per_brara in Swissprot. Arabidopsis peroxidases are designated by a triangle.

selective cloning, that uses a probe that preferentially detects closely related sequences (8). The calculated pIs are all basic, and reach a value of 10.7 for Atprx10. One to 7 potential glycosylation sites are found.

Phylogeny

The evolutionary relationships with available peroxidases from other plants is shown in a phylogenic tree (Figure 3). The tree is only approximate since not all sequences are complete and therefore biased the calculations. Nine sequences were nevertheless omitted because of the excessive size of the missing fragments. It is apparent from this tree that Arabidopsis peroxidases, besides a Brassicaceae cluster, are disseminated among the various branches with other plant peroxidases, independently of the taxonomic relation. This is best illustrated by Atprxr1 which shows a clear cut high degree of similarity to a cotton peroxidase, isolated in the top of the tree. This situation with plant peroxidases was already depicted in previous works (1,9). The grouping of various peroxidases irrespective of the taxonomic relations derives most likely from some functional specificities rather than random molecular variation on a theme. More refined phylogenic analysis is needed to strengthen this hypothesis.

CONCLUSIONS AND PERSPECTIVES

Molecular analysis of the Arabidopsis cDNA clones has shown a great variety of peroxidases isozymes in this model plant. It is questionable whether each single isozyme is linked to a specific peroxidative function. We rather think that the situation we observe is the result of both intertwined redondant diversification and functional speciation. With all the available clones, we are now headed towards the heterologous expression of the recombinant proteins, the biochemical analysis of the active enzymes and the spatio-temporal expression analysis of the genes. A better understanding of peroxidase diversity should soon emanate from this work.

Acknowlegments. We are thankful to the ABRC at the Ohio State University and to Dr T. Newman at the MSU for providing the Est clones. This research is part of a joint EU-project (AIR2-CT93-1661) and is supported by the Swiss Federal Office of Science and Education (grant OFES 93.0090).

Note added in proof. The Atprxr1 to Atprxr11 sequences are deposited in the EMBL database under the accession numbers X98313 to X98323, respectively.

REFERENCES

1. **Simon P** (1993) Plant Peroxidase Newsletter, C Penel, Th Gaspar, H Greppin eds, University of Geneva, 1, pp 4-7.
2. **Simon P, Tacchini P, Penel C and H Greppin** (1993) *In:* Plant Peroxidases: Biochemistry and Physiology, KG Welinder, SK Rasmussen, C Penel, H Greppin, eds, University of Geneva, pp 465-474.
3. **Hartmann C and Ortiz de Montellano PR** (1992) Arch Biochem Biophys 297: 61-72.
4. **Newman T, de Bruijn FJ, Green P, Keegstra K, Kende H, Mcintosh L, Ohlrogge J, Raikhel N, Somerville S, Thomashow M, Retzel E and Somerville C** (1994Plant Physiol 106:1241-1255.
5. **Devereux J, Haeberli P and Smithies OA** (1984) Nucl Acid Res 12: 387-395.
6. **Intapruk C, Higashimura N, Yamamoto K, Okada N, Shinmyo A and Takano M** (1991) Gene 98: 237-241.
7. **Welinder KG** (1992) *In:* Plant Peroxidases 1980-1990, Topics and Detailed Literature on Molecular, Biochemical, and Physiological Aspects, C Penel, Th Gaspar, H Greppin, eds University of Geneva, pp 1-24.
8. **Simon P** (1992) *In:* Plant Peroxidases 1980-1990, Topics and Detailed Literature on Molecular, Biochemical, and Physiological Aspects, C Penel, Th Gaspar, H Greppin, eds University of Geneva, pp 47-58.
9. **Esnault R and van Huystee RB** (1994) Plant Peroxidase Newsletter, C Penel, Th Gaspar, H Greppin eds, University of Geneva, 1, pp 7-11.