Analysis of the Molecular Evolutionary History of the Ascorbate Peroxidase Gene Family: Inferences from the Rice Genome

Felipe Karam Teixeira,[1] Larissa Menezes-Benavente,[1] Rogério Margis,[1,2] Márcia Margis-Pinheiro[1]

[1] Laboratório de Genética Molecular Vegetal, Departamento de Genética, UFRJ, 21944-970 Rio de Janeiro, Brasil
[2] Departamento de Bioquímica, Instituto de Química, UFRJ, 21944-970, Rio de Janeiro, Brasil

Abstract. Ascorbate peroxidase (APx) is a class I peroxidase that catalyzes the conversion of H_2O_2 to H_2O and O_2 using ascorbate as the specific electron donor. This enzyme has a key function in scavenging reactive oxygen species (ROS) and the protection against toxic effects of ROS in higher plants, algae, and *Euglena*. Here we report the identification of an APx multigene family in rice and propose a molecular evolutionary relationship between the diverse APx isoforms. In rice, the APx gene family has eight members, which encode two cytosolic, two putative peroxisomal, and four chloroplastic isoforms, respectively. Phylogenetic analyses were conducted using all APx protein sequences available in the NCBI databases. The results indicate that the different APx isoforms arose by a complex evolutionary process involving several gene duplications. The structural organization of APx genes also reflects this process and provides evidence for a close relationship among proteins located in the same subcellular compartment. A molecular evolutionary pathway, in which cytosolic and peroxisomal isoforms diverged early from chloroplastic ones, is proposed.

Key words: Ascorbate peroxidase — Antioxidant metabolism

Introduction

Aerobic life has evolved by exploiting the abundance of environmental O_2 in the atmosphere to oxidize organic compounds, thus obtaining chemical energy in a highly efficient manner. Paradoxically, the univalent reduction of molecular oxygen in metabolic reactions produces a plethora of partially reduced intermediates, commonly known as reactive oxygen species (ROS). If their levels are not tightly controlled, these chemical species can react with the majority of biological molecules and cause serious cellular damages (Fridovich 1975). However, ROS are also important in many physiological processes and their balance is of the utmost importance. As a result, a complex system, comprising enzymatic and nonenzymatic mechanisms, maintains the delicate balance between oxidant and antioxidant compounds in the cell (Scandalios 2002).

Peroxidases have a prominent position in antioxidant defenses, removing peroxides, including H_2O_2, generated by oxy-reduction reactions. Heme peroxidases are organized into three classes: (1) intracellular peroxidases, which correspond to peroxidases of prokaryotic origin (class I; e.g., yeast cytochrome *c* peroxidase); (2) secretory fungal peroxidases (class II; e.g., manganese peroxidase); (3) and plant peroxidases targeted to the secretory pathway (class III; e.g., horseradish peroxidase) (Zámocký et al. 2000; Duroux and Welinder 2003). Ascorbate peroxidase (APx, EC 1.11.1.11) belongs to the class I heme-containing peroxidases found in higher plants, chlorophytes (Takeda et al. 1998, 2000), red algae (Sano

Correspondence to: Márcia Margis-Pinheiro; *email:* margism@biologia.ufrj.br

et al. 2001), and members of the protist kingdom (Shigeoka et al. 1980; Wilkinson et al. 2002). It displays a high specificity toward ascorbate as the specific electron donor. H_2O_2 scavenging is accomplished through a series of reactions comprising the ascorbate–glutatione cycle (Noctor and Foyer 1998).

The catalytic mechanism, crystal structure, and ascorbate-binding of APx have been extensively studied (Patterson and Poulos 1995; Mandelman et al. 1998; Wada et al. 2002; Raven 2003; Sharp et al. 2003). These studies have shown the existence of a common APx catalytic core, harboring two typical domains of heme peroxidases. The active site contains a histidine residue, the distal histidine, which acts as an acid–base catalyst in the reaction between hydrogen peroxide (H_2O_2) and the enzyme. Another important histidine residue, the proximal histidine, is also present at the heme-binding site (Henrissat et al. 1990).

Ascorbate peroxidases in higher plants are encoded by small multigene families and the different isoforms are classified according to their subcellular localization. Soluble isoforms are found in cytosol and chloroplast stroma, while membrane-bound isoforms are found in peroxisomes and chloroplast thylakoids. The final subcellular localization of the isozyme is determined by the presence of organelle-specific targeting peptides and transmembrane domains that are found in the protein N-terminal and C-terminal regions. APx biochemical properties, such as molecular mass, substrate specificity, pH optimum, and stability in the absence of ascorbate, have also been correlated to their subcellular localization (Ishikawa et al. 1998). Although APx activity has been detected in mitochondria (Jiménez et al. 1997; De Leonardis et al. 2000), no corresponding gene, cDNA, or protein sequences have been isolated so far.

APx has a fundamental role in photosynthetic organisms. Chloroplasts are the major sources of superoxide and H_2O_2 as a consequence of the highly energetic reactions that take place in this cell compartment. Since this organelle does not possess catalase activity, superoxide is readily converted into hydrogen peroxide, which is then scavenged by APx. The importance of APx and the ascorbate–glutathione cycle is not restricted to the chloroplasts; it also plays a role in ROS scavenging in the cytosol and the peroxisomes (Noctor and Foyer 1998).

The study of ascorbate peroxidases is comparatively recent in the history of plant antioxidant metabolism. APx activity was first reported by Shigeoka et al. in 1980, but despite their importance, little is known about the evolutionary history or the gene structure of the different isoforms. Considering the different subcellular localization and roles of APx

isoforms, it seems likely that the structural diversity of APx genes resulted from a complex process of molecular evolution. Although some data are available in the literature, a thorough study of these aspects is missing. Furthermore, systematic study of the APx gene family in a single plant species is lacking. We aimed (i) to investigate these issues by taking advantage of the completion of the rice genome sequence to identify the APx gene family in rice and (ii), to investigate the phylogenetic evolution of APx through the comparison of genes, cDNAs, and protein sequences and structures.

Materials and Methods

Retrieval of Ascorbate Peroxidase Sequences

APx protein sequences were extracted from the National Center for Biotechnology Information (NCBI) database by keyword searches (September 2002). The redundant protein sequences were then removed and a final nonredundant list, comprising 74 protein sequences, was used in this work. Protein sequences are identified by their accession number at NCBI.

The number of genes encoding APx in the rice genome was determined based on different databanks: NCBI (http://www.ncbi.nlm.nih.gov), The Institute of Genomic Research (TIGR; http://www.tigr.org/tdb/e2kl/osal/), Monsanto Company (http://www.rice-research.org), Syngenta Company (Goff et al. 2002) (http://portal.tmri.org/rice/RicePublicAccess.html), and Rice GD (Yu et al. 2002) (http://btn.genomics.org.cn:8080/rice/). Extensive searches using the BLAST program (Altschul et al. 1990) were conducted. First, cDNA sequences encoding ascorbate peroxidases from different organisms were used as queries in the BLASTN program. Then almost all protein sequences present on the nonredundant list were used as queries in the TBLASTN program. The resulting 110 genomic clusters were analyzed individually. The ProDom (http://prodes.toulouse.inra.fr/prodom/current/html/home.php) and PROSITE (http://www.expasy.org/prosite/) databases were used in the identification of the genes. The structural organization of APx genes was determined by aligning the genomic DNA and cDNAs/ESTs sequences. Genomic sequences were also analyzed in the FGENESH gene structure prediction program (http://www.softberry.com/) (Solovyev 2001) and GeneMark program (http://opal.biology.gatech.edu/GeneMark/). ESTs and cDNAs sequences were obtained from the NCBI, TIGR, and Rice GD databanks. The sequences of rice *OsAPx3* cDNA and protein appear in the GenBank database under accession numbers AY382617 and AAQ88105, respectively.

The genome databases of *A. thaliana* (http://www.ncbi.nlm.nih.gov/Genomes/index.html) and *C. reinhardtii* (http://genome.jgi-psf.org/chlrel/chlrel.home.html) were also analyzed using a methodology similar to that described for the rice sequences.

Sequence Alignments

Multiple sequence alignments were constructed using the program ClustalW 1.8 (Higgins and Sharp 1988) at the European Bioinformatics Institute server (http://www.ebi.ac.uk/clustalw/). Protein sequence alignments were performed with the following parameters: Gap Opening penalty = 10.0, Gap Extension penalty = 0.05, and BLOSUM (Henikoff) protein weight matrices. The multiple alignments were inspected by eye and edited using GeneDoc version 2.6.002 (Nicholas and Nicholas 1997).

Phylogenetic Constructions

The alignment used in the construction of trees contained the nonredundant proteins recovered from NCBI, the APx rice proteins in this work, and a yeast cytochrome *c* peroxidase protein (NP_012992). Sorting peptides and membrane-binding regions were excluded prior to tree construction. Three phylogenetic methods were used for tree construction: neighbor-joining (NJ), minimum evolution (ME), and maximum likelihood (ML). In the NJ and ME methods, the molecular evolutionary and phylogenetic analyses were conducted using Molecular Evolutionary Genetics Analysis (MEGA) version 2.1 (Kumar et al. 2001). The molecular distances of the aligned sequences were calculated according to the parameter *p*-distance. The parameter used to analyze gap/missing data was pairwise deletion. For both the NJ and the ME methods, two different tests of phylogeny were used: interior branch test and bootstrap test, with 1000 replications each. In the ML method, the rooted distance tree was calculated by PHYLIP (Phylogeny Inference Package) version 3.6a3 using the Dayhoff PAM model (PROTML) (Felsenstein 1989). All trees obtained exhibited the same topology. The tree presented in Fig. 3 was constructed with the following parameters: NJ, *p*-distance, pairwise deletion, and interior branch test (1000 replications).

Protein Sequence Analyses

Putative transmembrane domains in APx proteins were identified using SOSUI version 1.0 (Hirokawa et al. 1998) present in the BCM search launcher (http://searchlauncher.bcm.tmc.edu/). Molecular weight (MW) prediction of rice proteins was conducted by the ExPASy Molecular Biology Server (http://bo.expasy.org/tools/pi_tool.html).

Results

The Ascorbate Peroxidase Gene Family in Rice is Comprised of Eight Members

The APx gene family has been partially characterized in *Arabidopsis thaliana* (Jespersen et al. 1997) and in spinach *(Spinacia oleracea)* (Yoshimura et al. 2000). Recently, Agrawal et al. (2003) characterized two APx genes in rice, *OsAPX1* and *OsAPX2*, by Southern and northern analysis. However, no data concerning the structural organization of these genes were presented. We have identified the APx-encoding genes in the rice genome by extensive searches of different rice genomic sequence databases. A total of 36 genomic clones were identified and assembled into sequences corresponding to eight different APx loci. We named these genes *OsAPx1* to *OsAPx8*.

The identification and analyses of the genomic structure of these genes were performed using gene structure prediction programs and by comparison with the available ESTs clones. ESTs and cDNA sequences corresponding to each gene were obtained by searching the NCBI and Rice GD databanks as well as a recently described library of 28,000 full-length rice cDNAs (Kikuchi et al. 2003). The mRNA sequences and the predicted proteins corresponding to the genomic sequences were then determined. The predicted proteins were then analyzed for the presence of conserved APx domains and homology with typical APx sequences.

Among these sequences, *OsAPx1* and *OsAPx2* correspond to previously characterized cDNAs encoding cytosolic isoforms (Morita et al. 1997, 1999) and to the genes detected by Agrawal et al. (2003). The deduced protein sequences of two other genes, named *OsAPx3* and *OsAPx4*, harbor a putative hydrophobic transmembrane domain rich in valine and alanine residues followed by a short stretch of four or five positively charged residues, rich in serine and lysine. In OSAPX3, this domain is 70% similar to the cotton peroxisomal APx transit peptide. This similarity is higher in OSAPX4 (85%). OSAPX3 and OSAPX4 were designated putative peroxisomal membrane-bound proteins.

Four different gene products exhibited an N-terminal extension region characteristic of sorting peptide sequences typically found in chloroplastic APx isoforms. The genes encoding these proteins were named *OsAPx5*, *OsAPx6*, *OsAPx7*, and *OsAPx8*. Interestingly, *OsAPx5* and *OsAPx6* are located in tandem in the genome, separated by only 722 bp. OSAPX5 and OSAPX6 have a high level of similarity (>85%), suggesting that they have arisen by a recent gene duplication event. A third gene encoding a putative chloroplastic isoform, *OsAPx7*, has an intron interrupting the untranslated 3′ region that is not present in the other genes in this class. In addition to the N-terminal sorting peptide, OSAPX8 also possesses a C-terminal hydrophobic region characteristic of transmembrane domains found in the thylakoid-bound APx isoforms of spinach and pumpkin (Ishikawa et al. 1996). Therefore, while OSAPX5, OSAPX6, and OSAPX7 are likely to be soluble in the stroma, OSAPX8 is a putative thylakoid membrane-bound isoform.

The accession numbers of proteins and cDNAs in the NCBI databank, chromosome allocation, number of exons and introns, cDNA and open reading frame (ORF) lengths, predicted number of amino acids, and molecular weight (MW), and putative subcellular localization of the proteins corresponding to the genes described here are summarized in Table 1.

Chloroplast-Specific Signatures in Ascorbate Peroxidase

A comparison of the rice APx isoforms with those of other species revealed a high degree of conservation among these proteins. All APx proteins share a central region, the catalytic core, harboring the active site and the heme-binding domain (Henrissat et al. 1990). The specific subcellular localization of each particular

Table 1. Gene structure and principal features of the rice ascorbate peroxidase isoforms

| | Chromosome | Number of | | cDNA/ORF (bp) | Protein (aa) | MW (kDa) | NCBI accession No. | | Putative subcellular localization |
		Exons	Intorns				cDNA	Protein	
OsAPx1	3	9	8	983/753	250	27.2	D45423	BAA08264	Cytosol
OsAPx2	7	9	8	1160/756	251	27.1	AB053297	BAA20889	Cytosol
OsAPx3	4	9	8	1060/876	291	32.0	AY382617	AAQ88105	Peroxisome
OsAPx4	8	9	8	1244/876	291	31.7	AK070842	—	Peroxisome
OsAPx5	12	11	10	1129/963	320	34.8	AK073910	—	Chloroplast
OsAPx6	12	11	10	1288/930	309	33.5	AK061107	—	Chloroplast
OsAPx7	4	12	11	1392/1080	359	38.3	AB1148551,AK103344	BAC79362	Chloroplast
OsAPx8	2	12	11	1698/1497	478	51.2	AK070842	BAC79363	Chloroplast
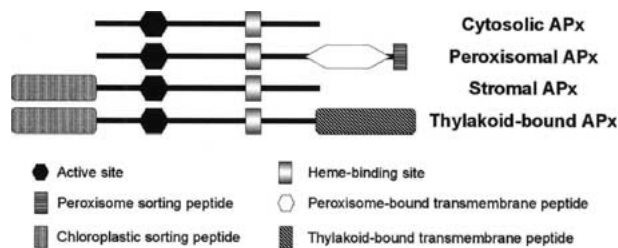
Fig. 1. Protein structure of APx isoforms. Comparison of the primary structure of APx proteins from distinct subcellular compartments. The different domains are represented by different symbols as indicated.

isoform is determined by the presence of additional sorting peptides (Shigeoka et al. 2002). Transmembrane domains also function to anchor peroxisomal and thylakoid-bound isoforms. A schematic representation of the structural organization of these proteins is given in Fig. 1. The APx isoforms identified in the rice genome display the combination of domains appropriate for the predicted subcellular location.

The alignment of all 74 APx protein sequences retrieved from the NCBI database (nonredundant list) revealed additional specific signatures for each particular isoform. Two signatures clearly identify higher plant chloroplastic isoforms (Fig. 2). The first signature is found near the active site and consists of seven residues (K-[ND]-I-[ETK]-E-W-P). The second signature has 16 residues and is found next to the heme-binding site (E-T-K-Y-T-[KE]-[DNTE]-G-P-G-[ANEK]-[PA]-G-G-Q-S). Additional differences among the isoforms, generally single—amino acid changes, are found throughout the alignment. Altogether, these differences clearly distinguish chloroplastic from nonchloroplastic isoforms (cytosolic and peroxisome membrane-bound). Although no particular function could be attributed to these signatures, they may affect the biochemical properties of the different APx isoforms. The proximity of these two new signatures to the active site and the heme-binding domain suggests a role for these domains in determining substrate specificity, which is higher in chloroplastic APx isoforms (Nakano and Asada 1987; Chen and Asada 1989; Miyake and Asada 1992,

1996). These two sequences could be useful to classify newly identified APx proteins in relation to their subcellular localization.

Phylogenetic Analyses

To investigate the molecular evolution and phylogenetic relationships among ascorbate peroxidases in plants, algae, and protists, APx protein sequences were aligned by CLUSTALW and analyzed using the MEGA 2.1 software (Kumar et al. 2001) (Fig. 3) and the PHYLIP package. To avoid a biased analysis, sorting peptides and membrane-binding domains were excluded. The closely related yeast cytochrome *c* peroxidase (CCP), a class I heme-binding peroxidase of the plant peroxidase superfamily, was included in the analyses as outgroup. Paterson and Poulos (1995) showed that the crystal structure of APx is nearly the same as that of CCP, despite their low amino acid sequence similarity (33%).

The phylogenetic tree structure reveals a clear divergence between chloroplastic and nonchloroplastic isoforms, represented in the first dichotomous branching. Two main groups are distinguished. One group includes all proteins known or predicted to be localized in the chloroplast. The second group comprises those proteins localized or predicted to be soluble in the cell cytosol or anchored in the peroxisomal membrane. This suggests that the nonchloroplastic isoforms (cytosolic and peroxisomal isoforms) were generated by duplication events of a single nonchloroplastic ancestral gene. Among them, a third small group was distinguished in the analyses, indicated in the phylogenetic tree as ''new'' APx isoforms. The sequences belonging to this group are so far restricted to dicotyledoneous plants and, to date, have been identified in only three species. It is important to emphasize that no isoform from this group was found in the rice genome. The new APx isoforms grouped closely to the peroxisomal isoforms, suggesting a common origin to these branches.

Although excluded in the generation of the phylogenetic tree, all proteins grouped in the chloroplastic
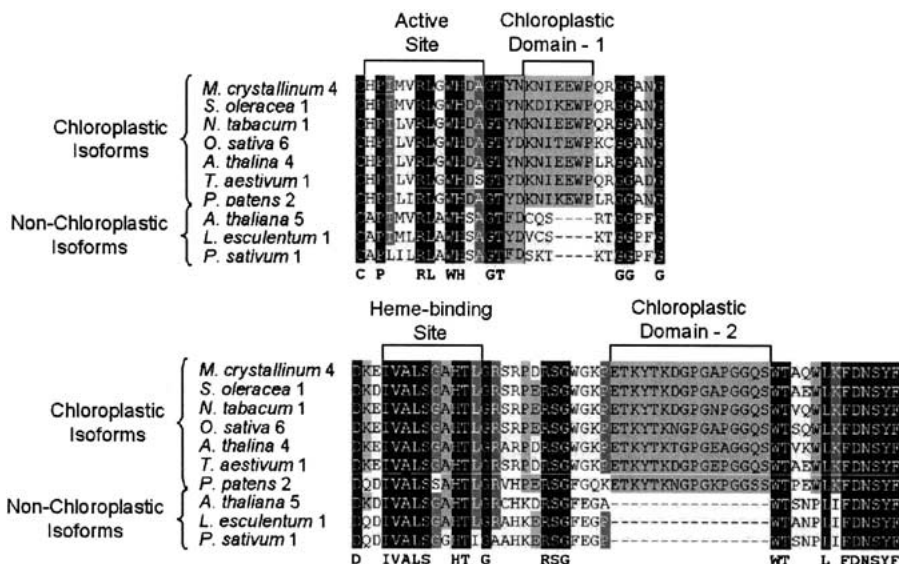
Fig. 2. Conserved sequences in chloroplastic APx. Representation of two noncontiguous conserved domains present in APx amino acid sequences. Semiconservative residues present in more than 50% of the aligned sequences are highlighted in gray. Residues conserved in all sequences are highlighted in black and are shown below the alignment. The active and heme-binding sites are indicated above. The new chloroplastic domains 1 and 2, found only in chloroplastic isoforms, are also indicated. Nonchloroplastic isoforms from *A. thaliana*, *L. esculentum*, and *P. sativum* were included in the analyses. Protein accession numbers, at NCBI: *M. crystallinum* 4 (AAC19394), *S. oleracea* 1 (BAA24610), *N. tabacum* 1 (BAA78553), *O. sativa* 6, *A. thaliana* 4 (CAA67427), *T. aestivum* 1 (AA77158), *P.patens* 2 (BQ042082), *A. thaliana* 5 (CAA42168), *L. esculentum* 1 (CAB58361), and *P. sativum* 1 (P48534).

branch have a chloroplast-targeting peptide at the N-terminal. In the nonchloroplastic group, all peroxisomal proteins have a hydrophobic transmembrane domain adjacent to a positively charged tail, while the cytosolic proteins are comprised of the main catalytic core and lack any additional sorting domains (Fig. 1).

The putative APx proteins described in the chloroplast thylakoid lumen of arabidopsis and its homologues in tomato (*Lycopersicon esculentum*) (Kieselbach et al. 2000) and rice lack both the active site and the heme-binding domain present in heme peroxidases. In addition, other regions that are relatively well conserved in all proteins with APx activity have diverged significantly in these sequences. These three proteins share a high degree of similarity ($\sim$70%) but have a low similarity to APx ($\sim$10%). Consistent with the alignment, these proteins formed a closed group in the phylogenetic tree, which is more distant from the APx protein group than is CCP, the outgroup used for the analyses. Taken together, these observations indicate that these proteins are not APx isozymes; their classification in the APx family was based on a weak similarity and there is no biochemical confirmation of their activity. Moreover, our results revealed no strong relationship among these proteins and the APx group.

Structural Organization of Ascorbate Peroxidase Genes

To further investigate the relationships among the distinct APx isoforms, we determined the structural organization of APx genes by comparing, where available, gene and cDNA sequences encoding the same isoform. All arabidopsis and rice APx genes were included in these analyses.

A detailed comparison of APx genes across plant species (Fig. 4) revealed a high degree of conservation in gene structure for APx isoforms localized in the same subcellular compartment. Moreover, the structure of the genes encoding cytosolic and peroxisome membrane-bound isoforms is very similar, differing markedly from that of the chloroplastic APx genes.

In higher plants, nonchloroplastic isoforms are encoded by genes that are generally comprised of nine exons and eight introns (Fig. 4 and Table 1). The one exception is the coding region of *A. thaliana* 5, encoding a cytosolic isoform, which has eight exons and seven introns. The genes *A. thaliana* 5, *Pisum sativum* 3, *Fragaria ananassa* 6, *F. ananassa* 7, and *F. ananassa* 8, all corresponding to cytosolic isoforms, have an intron in their 5′UTR. Comparison of the structural organizations further showed that genes encoding cytosolic isoforms lack the intron that separates exons 2 and 3 in the genes encoding peroxisome membrane-bound isoforms. In addition, the last two exons found in cytosolic isoforms were replaced in the peroxisome-bound isoforms by a large final exon that encodes the hydrophobic transmembrane domain and the conserved positively charged domain required for peroxisome targeting. A putative hydrophobic transmembrane domain was also found in the coding region of *A. thaliana* 2, a member of the New APx group. Among nonchloroplastic isoforms,
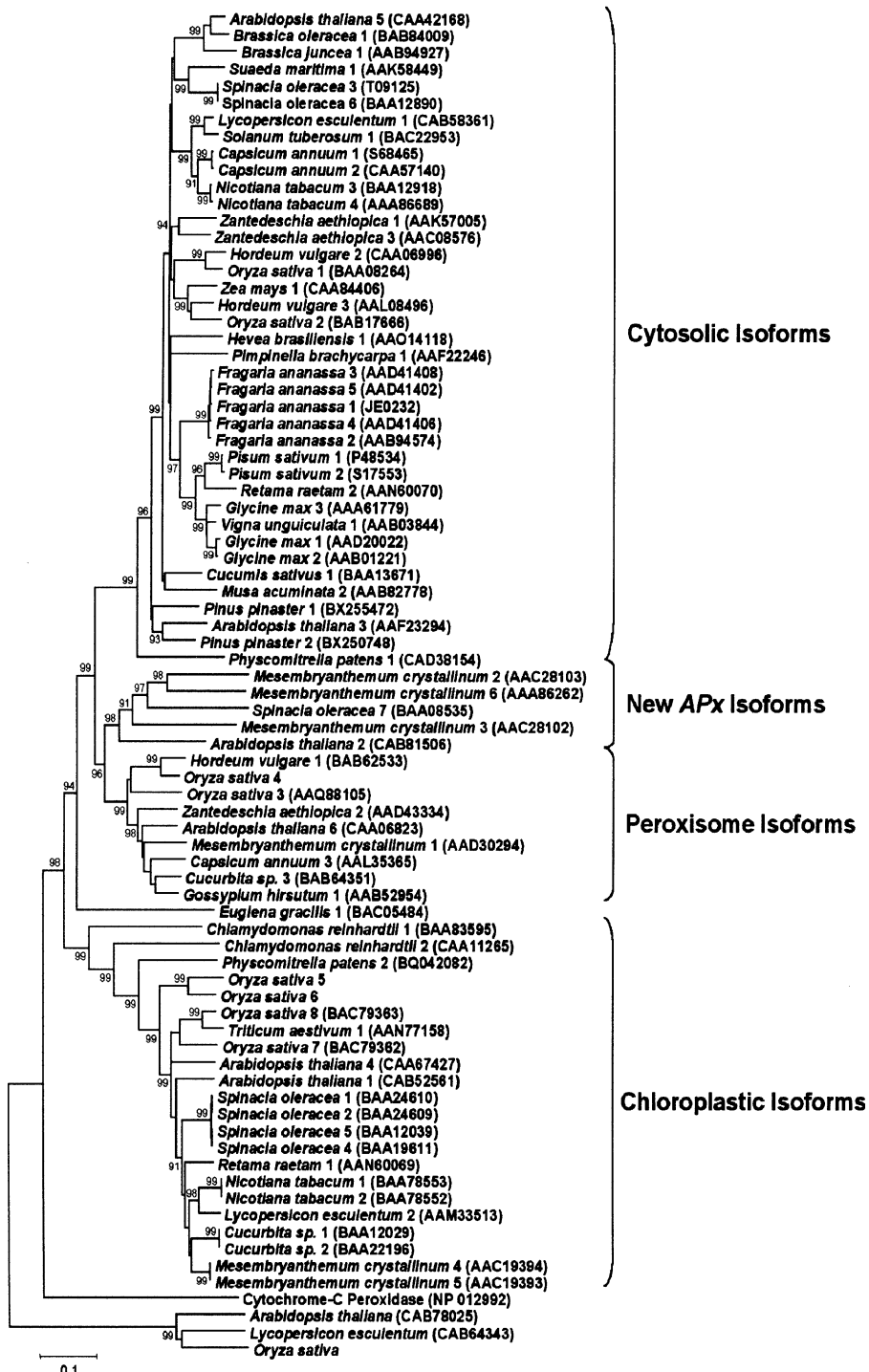
Fig. 3. Phylogenetic tree of APx proteins from different organisms. Phylogenetic analyses were conducted using the MEGA version 2.1 software. Molecular distances were calculated using the parameter of p-distance, and the trees were constructed using the neighbor-joining method with pairwise deletion. The test of phylogeny used was the interior branch test with 1000 replications. Confidence test values $> 90\%$ are shown. Sorting peptides and membrane-binding domains were excluded in these analyses. The same topology within each group was recovered by all methods described under Materials and Methods. Protein sequences are identified by their accession numbers in the NCBI database.

another important finding was that the structures of peroxisome and New APx encoding genes are more related to each other than to the structure of genes encoding cytosolic isoforms.

The structure of the genes encoding chloroplastic isoforms is distinct from that of the nonchloroplastic forms (Fig. 4). In rice and arabidopsis, the structure of the genes encoding thylakoid-bound isoforms
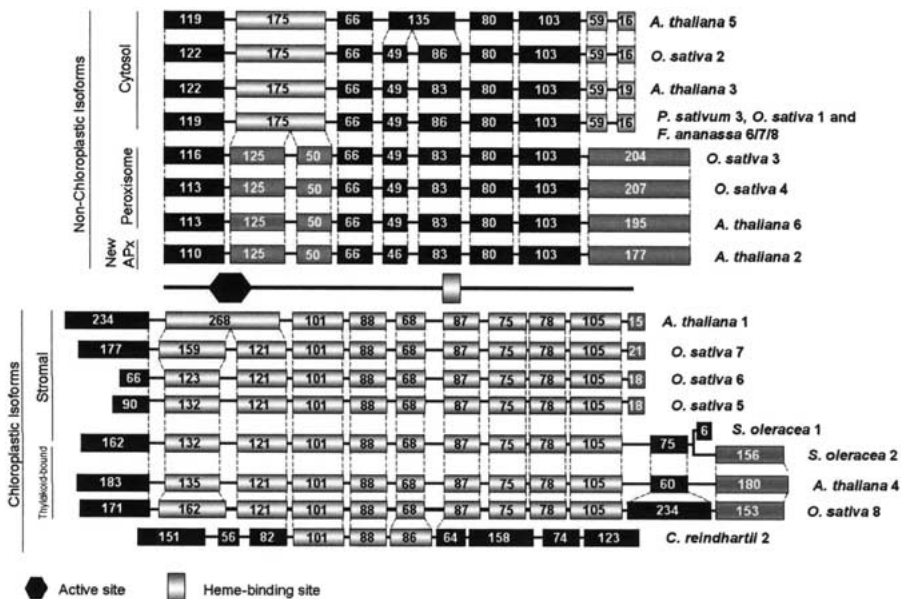
Fig. 4. Structural organization of plant APx genes. Exon sequences are represented as simplified boxes. The size of each exon (in bp) is given. Bars represent introns. *S. oleracea* 1 and *S oleracea* 2 are generated by alternative splicing. Relative positions of the active site and the heme-binding site are given. Gene accession numbers, at NCBI: *A. thaliana* 1 (At4g08390); *A. thaliana* 2 (At4g35970); *A. thaliana* 3 (At3g09640); *A. thaliana* 4 (Atlg77490); *A. thaliana* 5 (Atlg07890); *A. thaliana* 6 (At4g35000); *F. ananassa* 6 (AF158652); *F. ananassa* 7 (AF158653); *F. ananassa* 8 (AF158654); *P. sativum* 3 (M93051); *S. oleracea* 1/2 (AB002467). C. *reindhartii* and *0. sativa* gene sequences were identified by searches in different databanks as described under Materials and Methods.
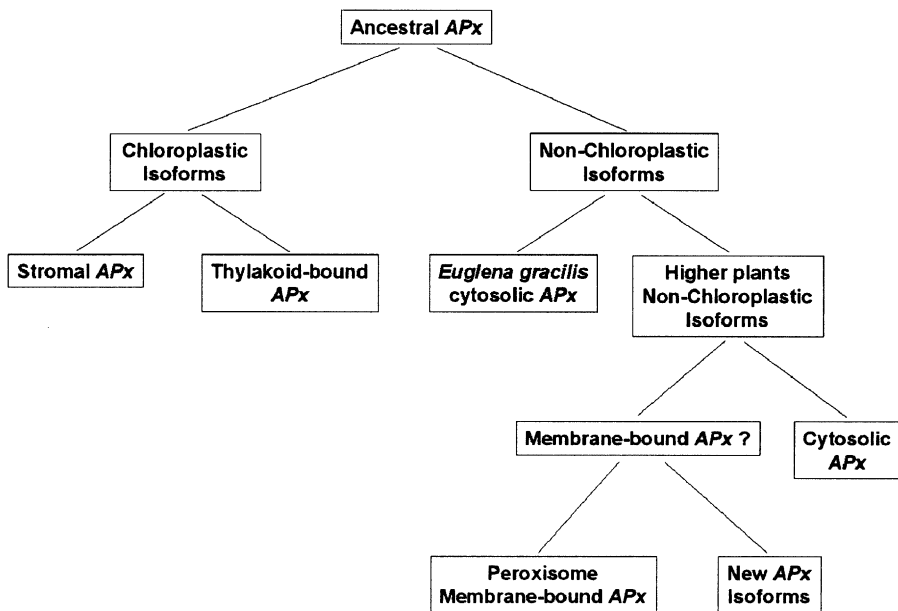

Fig. 5. Reconstruction of the evolutionary pathway of APx genes. The proposed evolutionary relationships between APx family members are as follows: An ancient duplication event led to the divergence of chloroplastic from nonchloroplastic isoforms; chloroplastic isoforms diverged later, giving rise to stromal APx and thylakoid-bound APx; *E. gracilis* and nonchoroplastic plants isoforms diverged; a more recent duplication led to the divergence of cytosolic from a possible membrane-bound APx; peroxisomal APx diverged from new APx isoforms.

comprises 12 exons and 11 introns and the coding regions of the genes corresponding to stromal isoforms are usually comprised of 11 exons and 10 introns. The one exception is *A. thaliana 1*, with 10 exons separated by 9 introns. The *O. sativa 7* and *A. thaliana 1* genes have an intron inserted in their 3′UTR. In contrast, spinach stromal and thylakoid membrane-bound isoforms are generated by alternative splicing of a single gene comprised of 13 exons and 12 introns (Yoshimura et al. 2002). In higher

plants, the chloroplastic sorting signal spans exon 1 and the first half of exon 2, where the cleavage site of this peptide is found. In the thylakoid-bound isoforms, the hydrophobic transmembrane domain is encoded by the last exon of these genes. A common splicing site was found in genes encoding stromal and thylakoid-bound APx in higher plants and green algae (*Chlamydomonas reinhardtii* 2). There was no conservation of intron location between these two groups.

Discussion

Ascorbate peroxidases in higher plants are encoded by multigene families, which have been partially characterized in spinach (Yoshimura et al. 2000) and arabidopsis (Jespersen et al. 1997). The availability of the rice genome sequence and a large EST collection deposited in the NCBI GeneBank prompted us to investigate this class of proteins, a major enzyme in antioxidant defenses. The rice APx gene family is comprised of eight members. The rice isoforms harbor the typical heme-peroxidase domains and show a high degree of similarity to ascorbate peroxidases from other plant species. To our knowledge, this is the largest APx gene family described so far.

The eight genes identified encode two cytosolic isoforms, two putative peroxisomal proteins, and four putative chloroplastic ones. The subcellular location of these proteins was assigned based on the presence of targeting signals and transmembrane domains. Sorting of peroxisomal proteins is determined by a hydrophobic transmembrane domain, adjacent to a conserved positively charged five-residue domain at the C-terminal, and occurs in cotton via a reticular membranous network (Mullen and Trelease 2000). On the other hand, chloroplastic isoforms are preceded by a common chloroplast-targeting peptide present in the N-terminus, which is processed in mature proteins (Madhusudhan et al. 2003). Thylakoid isoforms differ from the stromal isoforms by the presence of a hydrophobic membrane-binding domain in the C-terminus (Ishikawa et al. 1996).

Chloroplastic isoforms (stromal and thylakoid-bound) are generated by alternative splicing of a single gene in spinach (Ishiwaka et al. 1997), tobacco (*Nicotiana tabacum*) (Yoshimura et al. 2000), pumpkin (*Curcubita* sp.) (Mano et al. 1997), and iceplant (*Mesembryanthemum crystallinum*). In these species, alternative splicing is determined by a conserved putative splicing regulatory *cis*-element (SRE), upstream of the acceptor site in intron 12 (Yoshimura et al. 2002). However, our results showed that rice chloroplastic isoforms are encoded by separate genes, as described in *A. thaliana* (Jespersen et al. 1997). Accordingly, these genes lack the SRE element.

The phylogenetic tree (Fig. 3) exhibits a strong dichotomous structure separating chloroplastic from nonchloroplastic isoforms, reflecting their subcellular localization. Thus, isoforms belonging to the same subcellular compartment are more related to each other than to isoforms from the same organism but with a different localization. This close relationship may indicate that proteins located at the same subcellular compartment have a common origin. This observation is supported by high interior branch test values. In other words, it is likely that all proteins located in chloroplasts, in such different organisms as green algae and higher plants, originated from a single common ancestral gene. This single origin event theory applies to all isoforms presented in Fig. 3.

The new APx isoform group was characterized only based on its common origin (Fig. 3). Unfortunately, no member of this group has been characterized with respect to subcellular localization or biochemical properties. We were not able to find any clear consensus pattern in this group, and their C-terminal regions show weak sequence similarity. Possibly, the characteristics of proteins from this group will be determined as more sequences belonging to the group are identified. However, our structure analyses showed that one member of this group, *A. thaliana* 2 (CAB81506), has a putative transmembrane domain in the C-terminus, very similar to those found in the peroxisomal isoforms. Furthermore, the last five residues correspond to a putative peroxisome sorting signal, suggesting that it may be a membrane-bound protein. On the other hand, other members of this group were predicted to be soluble proteins, due to the absence of hydrophobic transmembrane domains. The presence of the transmembrane domain in the *A. thaliana* 2 gene indicates that the ancestor of the peroxisome-bound and new APx isoforms should also encompass this domain. Thus, the membrane-binding domain may represent a gain of function in the lineage of the peroxisomal and new APx groups or a loss of function in the ancestral gene of the cytosolic isoforms from higher plants. Moreover, this suggests that the absence of a transmembrane domain in the other members of the new APx group results from a recent loss of function in these plant lineages.

The phylogenetic tree of APx reveals ancient molecular divergence events but does not reflect more recent ones. For instance, in the branch of cytosolic isoforms an APx from a bryophyte (*Physcomitrella patens* 1; CAD38154) grouped outside of the Tracheophyta proteins, reflecting the divergence of these two taxons. Similarly, in the chloroplastic isoform branch, the bryophyte APx (*P. patens* 2; BQ042082) is grouped outside of the Tracheophyta chloroplastic isoforms. Moreover, the Chlorophyta (green algae)

chloroplastic isoforms (*C. reinhardtii* 1 and 2; BAA83595 and CAA11265, respectively) are grouped apart from those of higher plants (Embryophyta) chloroplastic isoforms, demonstrating a more ancient event.

As for nonchloroplastic isoforms, the speciation event that separated the Euglenazoa from the plants lineage (Viridiplantae) also separated the ancestor of the cytosolic *Euglena gracilis* gene from that of the plant nonchloroplastic isoforms. In other words, the divergence of Euglenazoa and Viridiplantae occurred before the duplication events that originated the diversity of nonchloroplastic isoforms (cytosolic, peroxisome, and new APx isoforms) in the plant lineages. This hypothesis is supported by the observation that the *E. gracilis* 1 (BAC05484) cytosolic isoform (Shigeoka et al. 1980) grouped together with the nonchloroplastic isoforms but separate from the isoform group of the plants lineage.

The comparisons of the structural organization of APx genes further corroborate the phylogenetic tree topology and reinforce the concept of a common origin with each isoform arising from a common ancestor for all APx genes. Isoforms located in the same cellular compartment have a similar gene structure, sharing most splicing sites used for determination of exon–intron borders. Furthermore, in genes encoding chloroplastic isoforms, the three widely conserved splicing sites among plants and green algae strongly support a common origin to this group.

An evolutionary pathway of the APx genes is proposed (Fig. 5), based on a sequence of gene duplications that led to the current diversity of APx isoforms. An initial duplication event generated the ancestral genes encoding the chloroplastic and the nonchloroplastic isoforms, from a common ancestral APx gene, encoding an isoform of unknown subcellular localization. It is likely that this first duplication event occurred before the divergence between Viridiplantae and Euglenazoa.

The two main APx groups, chloroplastic and nonchloroplastic, were thus formed, and each followed its own evolutionary path from there. In one branch, the ancestral gene for chloroplastic isoforms went through a second duplication event, generating the ancestor of the stromal and thylakoid-bound isoforms. It seems likely that the alternative splicing of the genes for the chloroplastic APx proteins has arisen very recently in the evolutionary history of a certain lineage of eudicots. This idea is supported by the absence of alternative splicing of the corresponding genes in rice, arabidopsis, and chlamydomonas.

The diversification within the higher plant nonchloroplastic APx branch occurred after the divergence between Euglenazoa and Viridiplantae, as shown in the phylogenetic tree. Because the non-chloroplastic APx isoform in *Euglena* is cytosolic, it is possible that the ancestor of the nonchloroplastic APx isoforms in plants was also cytosolic. The plant ancestral APx gene further duplicated, generating the ancestor of cytosolic isoforms in plants and a second gene, the common ancestor of peroxisomal and new APx isoforms. Moreover, this common ancestor of the peroxisomal and the new APx isoforms was probably a membrane-bound protein, since the peroxisomal isoforms are membrane-bound and at least one member of the new APx isoform group (*A. thaliana* 2) is also a putative membrane protein. This implies that this gene acquired the last exon that codes a transmembrane domain and the intron that separates exons 2 and 3. Another duplication event separated peroxisomal from new APx isoforms. Some members of the new APx group have subsequently lost their transmembrane domain.

The analyses presented here raise many questions about the role played by the different APx isoforms in antioxidant metabolism. Rice constitutes an excellent model to access the function of APx from different cell compartments. Thus, a comprehensive analysis of the expression profile of each of these genes should answer some of these questions. Analysis of the ability of rice mutants lacking different APx isoforms to scavenge ROS would also complement these studies. Understanding how cells activate their antioxidant mechanism in different subcellular compartments will enable us to manipulate these genes in order to improve crop tolerance against biotic and abiotic stress.

References

Agrawal GK, Jwa N, Iwahashi H, Rakwal R (2003) Importance of ascorbate peroxidases *OsAPx1* and *OsAPx2* in the rice pathogen response pathways and growth and reproduction revealed by their transcriptional profiling. Gene 322:93–103

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Chen G-X, Asada K (1989) Ascorbate peroxidase in tea leaves: Occurrence of two isoenzymes and the differences in their enzymatic and molecular properties. Plant Cell Physiol 30:987–998

De Leonardis S , Dipierro N, Dipierro S (2000) Purification and characterization of an ascorbate peroxidase from potato tuber mitochondria. Plant Physiol Biochem 38:773–779

Duroux L, Welinder KG (2003) The peroxidase gene family in plants: A phylogenetic overview. J Mol Evol 54:397–407

Felsenstein J (1989) PHYLIP—Phylogeny Inference Package (version 3.2). Cladistics 5:164–166

Fridovich I (1975) Superoxide dismutases. Annu Rev Biochem 44:147–159

Goff SA, Ricke D, Lan T et al. (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). Science 296:92–100

Henrissat B, Saloheimo M, Lavaitte S, Knowles JK (1990) Structural homology among the peroxidase enzyme family revealed by hydrophobic cluster analysis. Proteins 8:215–217

Higgins DG, Sharp PM (1988) CLUSTAL: A package for performing multiple sequence alignment on a microcomputer. Gene 73:237–244

Hirokawa T, Boon-Chieng S, Mitaku S (1998) SOSUI: Classification and secondary, structure prediction system for membrane proteins. Bioinformatics 14:378–379

Ishikawa T, Sakai K, Yoshimura K, Takeda T, Shigeoka S (1996) cDNAs encoding spinach stromal and thylakoid-bound ascorbate peroxidase, differing in the presence or absence of their 3′-coding regions. FEBS Lett 384:289–293

Ishikawa T, Yoshimura K, Tamoi M, Takeda T, Shigeoka S (1997) Alternative mRNA splicing of 3′-terminal exons generates ascorbate peroxidase isoenzyme in spinach (Spinacia oleracea) chloroplasts. Biochem J 328:795–800

Ishikawa T, Yoshimura K, Sakai K, Tamoi M, Takeda T, Shigeoka S (1998) Molecular characterization and physiological role of a glyoxysome-bound ascorbate peroxidase from spinach. Plant Cell Physiol 39:23–34

Jespersen HM, Kjaersgard IV, Ostergaard L, Welinder KG (1997) From sequence analysis of three novel ascorbate peroxidases from Arabidopsis thaliana to structure, function and evolution of seven types of ascorbate peroxidase. Biochem J 326:305–310

Jiménez A, Hernández JA, del Rio LA, Sevilla F (1997) Evidence for the presence of the ascorbate-glutathione cycle in mitochondria and peroxisomes of pea leaves. Plant Physiol 114:275–284

Kieselbach T, Bystedt M, Hynds P, Robinson C, Schröder WP (2000) A peroxidase homologue and novel plastocyanin located by proteomics to the Arabidopsis chloroplast thylakoid lumen. FEBS Lett 480:271–276

Kikuchi S, Satoh K, Nagata T et al. (2003) Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice. Science 301:376–379

Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: Molecular Evolutionary Genetics Analysis software. Bioinformatics 17:1244–1245

Madhusudhan R, Ishikawa T, Sawa Y, Shigeoka S, Shibata H (2003) Characterization of an ascorbate peroxidase in plastids of tobacco BY-2 cells. Physiol Plantarum 117:550–557

Mandelman D, Schwarz FP, Li H, Poulos T (1998) The role of quaternary interactions on the stability and activity of ascorbate peroxidase. Protein Sci 7:2089–2098

Mano S, Yamaguchi K, Hayashi M, Nishimura M (1997) Stromal and thylakoid-bound ascorbate peroxidase are produced by alternative splicing in pumpkin. FEBS Lett 413:21–26

Miyake C, Asada K (1992) Thylakoid-bound ascorbate peroxidase in spinach chloroplasts and photoreduction of its primary oxidation product monodehydroascorbate radicals in thylakoids. Plant Cell Physiol 33:54l–553

Miyake C, Asada K (1996) Inactivation mechanism of ascorbate peroxidase at low concentrations of ascorbate: Hydrogen peroxide decomposes compound I of ascorbate peroxidase. Plant Cell Physiol 37:423–553

Morita S, Kaminaka H, Yokoi H, Masumura T, Tanaka K (1997) Cloning and characterization of cytosolic ascorbate peroxidase cDNA from rice (PGR97-012). Plant Physiol 113:306

Morita S, Kaminaka H, Masumura T, Tanaka K (1999) Induction of rice cytosolic ascorbate peroxidase mRNA by oxidative stress: The involvement of hydrogen peroxide in oxidative stress signalling. Plant Cell Physiol 40:417–422

Mullen RT, Trelease RN (2000) The sorting signal for peroxisomal membrane-bound ascorbate peroxidase are within its C-terminal tail. J Biol Chem 275:16337–16344

Nakano Y, Asada K (1987) Purification of ascorbate peroxidase in chloroplasts: its inactivation in ascorbate-depleted medium and reaction by monodehydroascorbate radicals. Plant Cell Physiol 28:131–140

Nicholas KB, Nicholas HB, Deerfield DW (1997) GeneDoc: Analysis and visualization of genetic variation. EMBNEW.NEWS 4:14

Noctor G, Foyer CH (1998) Ascorbate and glutathione: Keeping active oxygen under control. Annu Rev Plant Physiol Plant Mol Biol 49:249–279

Patterson WR, Poulos TL (1995) Crystal structure of recombinant pea cytosolic ascorbate peroxidase. Biochemistry 34:4331–4341

Raven EL (2003) Understanding functional diversity and substrate specificity in haem peroxidases: What can we learn from ascorbate peroxidase? Nat Prod Rep 20:367–382

Sano S, Ueda M, Kitajima S, Takeda T, Shigeoka S, Kurano N, Miyachi S, Miyake C, Yokota A (2001) Characterization of ascorbate peroxidases from unicellular red alga Galdieria partita. Plant Cell Physiol 42:433–440

Scandalios JG (2002) The rise of ROS. Trends Biochem Sci 27:483–486

Sharp KH, Mewies M, Moody PC, Raven EL (2003) Crystal structure of the ascorbate peroxidase-ascorbate complex. Nat Struct Biol 10:303–307

Shigeoka S, Nakano Y, Kitaoka S (1980) Metabolism of hydrogen peroxide in Euglena gracilis Z by L-ascorbic acid peroxidase. Biochem J 186:377–380

Shigeoka S, Ishikawa T, Tamoi M, Miyagawa Y, Takeda T, Yabuta Y, Yoshimura K (2002) Regulation and function of ascorbate peroxidase isozymes. J Exp Bot 53:1305–1319

Solovyev VV (2001) Statistical approaches in eukaryotic gene prediction. In: Balding D (eds) Handbook of statistical genetics. John Wiley & Sons, New York, pp 83–127

Takeda T, Yoshimura K, Ishikawa T, Shigeoka S (1998) Purification and characterization of ascorbate peroxidase in Chlorella vulgaris. Biochimie 80:295–301

Takeda T, Yoshimura K, Yoshii M, Kanahoshi H, Miyasaka H, Shigeoka S (2000) Molecular Characterization and physiological role of ascorbate peroxidase from halotolerant Chlamydomonas sp. W80 strain. Arch Biochem Biophys 376:82–90

Yoshimura K, Yabuta Y, Ishikawa T, Shigeoka S (2000) Expression of spinach ascorbate peroxidase isoenzymes in response to oxidative stresses. Plant Physiol 123:223–234

Yoshimura K, Yabuta Y, Ishikawa T, Shigeoka S (2002) Identification of a cis element for tissue-specific alternative splicing of chloroplast ascorbate peroxidase pre-mRNA in higher plants. J Biol Chem 277:40623–40632

Wada K, Tada T, Nakamura Y, Yabuta Y, Yoshimura K, Takeda T, Shigeoka S, Nishimura K (2002) Crystallization and preliminary X-ray diffraction analysis of chloroplastic ascorbate peroxidase of tobacco plants. Acta Crystallogr D Biol Crystallogr 58:559–561

Wikinson SR, Obado SO, Mauricio IL, Kelly JM (2002) Trypanosoma cruzi expresses a plant-like ascorbate-dependent hemoperoxidase localized to the endoplasmic reticulum. Proc Natl Acad Sci USA 99:13453–13458

Yu J, Hu S, Wang J et al. (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. indica). Science 296:79–92

Záimocký M, Janecek S, Koller F (2000) Common phylogeny of catalase-peroxidases and ascorbate peroxidases. Gene 256:169–182