# From sequence analysis of three novel ascorbate peroxidases from *Arabidopsis thaliana* to structure, function and evolution of seven types of ascorbate peroxidase

Hans M. JESPERSEN, Inger V. H. KJÆRSGÅRD, Lars ØSTERGAARD and Karen G. WELINDER[1]

Department of Protein Chemistry, Institute of Molecular Biology, University of Copenhagen, Øster Farimagsgade 2A, DK-1353 Copenhagen K, Denmark

Ascorbate peroxidases are haem proteins that efficiently scavenge $H_2O_2$ in the cytosol and chloroplasts of plants. Database analyses retrieved 52 expressed sequence tags coding for *Arabidopsis thaliana* ascorbate peroxidases. Complete sequencing of non-redundant clones revealed three novel types in addition to the two cytosol types described previously in *Arabidopsis*. Analysis of sequence data available for all plant ascorbate peroxidases resulted in the following classification: two types of cytosol soluble ascorbate peroxidase designated cs1 and cs2; three types of cytosol membrane-bound ascorbate peroxidase, namely cm1, bound to microbodies via a C-terminal membrane-spanning segment, and cm2 and cm3, both of unknown location; two types of chloroplast ascorbate peroxidase with N-terminal transit sequences, the stromal ascorbate peroxidase (chs), and the thylakoid-bound ascorbate peroxidase showing a C-terminal transmembrane segment and designated cht. Further comparison of the patterns of conserved residues and the crystal structure of pea ascorbate peroxidase showed that active site residues are conserved, and three peptide segments implicated in interaction with reducing substrate are similar, excepting cm2 and cm3 types. A change of Phe-175 in cytosol types to Trp-175 in chloroplast types might explain the greater ascorbate specificity of chloroplast compared with cytosol ascorbate peroxidases. Residues involved in homodimeric subunit interaction are conserved only in cs1, cs2 and cm1 types. The proximal cation ($K^+$)-binding site observed in pea ascorbate peroxidase seems to be conserved. In addition, cm1, cm2, cm3, chs and cht ascorbate peroxidases contain Asp-43, Asn-57 and Ser-59, indicative of a distal monovalent cation site. The data support the hypothesis that present-day peroxidases evolved by an early gene duplication event.

## INTRODUCTION

Publication of the first complete amino acid sequence of a plant ascorbate peroxidase [1] made it feasible to rationalize the evolution of the plant peroxidase superfamily by comparison of sequences [2]. It became evident that mitochondrial yeast cytochrome *c* peroxidase, intracellular (chloroplast or cytosol) plant ascorbate peroxidases and bacterial peroxidases belonged to the same evolutionary branch of prokaryotic origin (designated Class I), whereas extracellular fungal peroxidases such as lignin and Mn-peroxidases (Class II), and extracellular or vacuolar plant peroxidases such as the classical horseradish peroxidase HRP C (Class III) constituted separate lineages.

In the present study we have adopted a novel approach for the investigation of structure, function and evolution of ascorbate peroxidases (EC 1.11.1.11) by exploiting the biological sequence information deriving from partly sequenced, randomly picked cDNA clones (expressed sequence tags; ESTs) [3,4]. ESTs from several plant species are accumulating rapidly, with 29165 entries from *Arabidopsis thaliana* (thale cress) [5,6] and 12674 entries from *Oryza sativa* (rice) [7] present in the database of ESTs (dbEST) release of 20 February 1997. For *Arabidopsis*, once a sequence of interest has been identified, the corresponding cDNA clone can be obtained from the Arabidopsis Biological Resource Center, Ohio State University, for further study. A similar service is available for rice ESTs (Rice Genome Research Program, Ibaraki, Japan).

Ascorbate peroxidase activity has been demonstrated in higher plants, eukaryotic algae and certain cyanobacteria [8]. Plant ascorbate peroxidases are intracellular enzymes encoded in the nucleus [1]. They function as scavengers of $H_2O_2$, thereby protecting plant cells from the deleterious effects of $H_2O_2$ generated as a by-product during photosynthesis and respiration [9]. The enzyme catalyses the reduction of $H_2O_2$ to water by using ascorbate as an electron donor. In functional terms, ascorbate peroxidases differ from Class III peroxidases by exhibiting a higher affinity for ascorbate as reducing substrate [10]. Furthermore ascorbate peroxidases are inhibited by thiol reagents and salicylic acid [11,12], and chloroplast ascorbate peroxidase can lose activity in the absence of ascorbate [13].

The nucleotide sequences of two cytosol soluble ascorbate peroxidases from *Arabidopsis* (ATAs) have been described previously [14,15]. In this paper we analyse the cDNA sequences and predicted protein structures of three novel ATAs and compare them with all known plant ascorbate peroxidases. Our analysis reveals a total of seven types of plant ascorbate peroxidase that differ in their cellular locations and structural characteristics. The data provide essential new insights into catalytic site and substrate interaction site structure, predict the presence of a monovalent metal ion site in the protein domain distal to haem in five types of ascorbate peroxidase, distinguish monomeric and dimeric ascorbate peroxidases and support the argument that the distal and proximal domains in the superfamily of haem peroxidases originate from gene duplication.

Abbreviations used: ATA, *Arabidopsis thaliana* ascorbate peroxidase; chs, chloroplast stromal; cht, chloroplast thylakoid-bound; cm, cytosol membrane-bound; cs, cytosol soluble; dbEST, database of ESTs; EST, expressed sequence tag; PAM, percentage accepted mutations.

[1] To whom correspondence should be addressed.

The nucleotide sequence results reported here will appear in DDBJ, EMBL and GenBank Nucleotide Sequence Databases under the accession numbers X98003, X98925, X98926 and X98927.

## MATERIALS AND METHODS

### Data analysis

Database searches were performed at the National Center of Biotechnology Information server (http://www.ncbi.nlm.nih.gov) with Entrez, BLAST [16] and keyword searches in dbEST. Sequence alignment was done with ClustalW [17] followed by minor manual adjustments. Phylogenetic tree construction was performed by the PHYLIP program package [18] with Fitch–Margoliash tree-fitting [19] of pairwise percentage accepted mutations (PAM) distances [20]. Hydrophobicity profiles and predictions of transmembrane segments [21] were performed with the program TopPred [22].

### DNA sequencing

EST clones 63G9T7, 109F7T7, 110L16T7 and 193D22T7 were obtained from the Arabidopsis Biological Resource Center in the form of DH10B(ZIP) cells containing ATP-encoding cDNA inserted in the pZL1 vector from Gibco BRL. Plasmid DNA was purified on Qiagen Tip-20 plasmid columns. Sequencing reactions were performed in accordance with the manufacturer's protocol with the *Taq* DyeDeoxy Terminator Cycle Sequencing Kit, and the samples were analysed on a Model 373A automated DNA Sequencer (Perkin Elmer). Clones were sequenced with T7 promoter, pucF and clone-specific primers (Pharmacia Biotech). Each base was unambiguously identified from both strands.

### Cloning and sequencing of a transit sequence

A clone encoding *A. thaliana* chloroplast stromal (chs) ascorbate peroxidase (ATA) (63G9T7) seemed to contain a partial transit sequence only and was therefore extended. An oligonucleotide primer, 5′-TGGCGGATCCTTAATGTTCTTGTTATACG-3′, corresponding to positions 337–309 in this clone, was designed including a *Bam*HI site (underlined) close to the 5′ end to facilitate cloning. The second primer corresponded to a sequence upstream of the polylinker sequence in the vector arm of the library. The standard PCR reaction was run on an aliquot of a cell suspension culture cDNA library, AC16H [23], kindly donated by Dr. Dominique Tremousaygue. Initially the phage DNA was unpacked by boiling for 10 min. The annealing temperature was set to 60 °C in accordance with the calculated melting temperature for the oligonucleotides. The largest product obtained by PCR was digested with *Eco*RI and *Bam*HI and purified by electrophoresis in a 1.5 % (w/v) agarose gel. A band corresponding to approx. 550 bp was excised from the gel, purified and ligated into the corresponding restriction sites of pUC18. An aliquot of the ligation mix was then transformed into the *Escherichia coli* strain DH5α.

## RESULTS AND DISCUSSION

### EST clones encoding ascorbate peroxidases

A search for *Arabidopsis* ascorbate peroxidases in dbEST revealed the presence of 52 entries. By nucleotide sequence comparisons these sequences could be divided into four groups. Sequences with more than 95 % pairwise sequence identity over a stretch of 100 or more bases were grouped together; 34 of the EST sequences were identical with the well-characterized soluble cytosol ATA [14,24] (ATAcs1) and therefore not studied further. Representative clones from each of the remaining three groups, which were subsequently assigned as ATAcm1 (ATA cytosol membrane-bound to microbodies), ATAchs (ATA chloroplast stromal or soluble) and ATAcht (ATA chloroplast, membrane-bound to thylakoid), were selected for complete DNA sequencing.

**Table 1 Sources of the ascorbate peroxidase sequences**

| Type | Plant | Sequence | Reference |
|---|---|---|---|
| cs1 | Pea | cDNA | [1] |
| | | Gene | [40] |
| cs1 | *Arabidopsis* | cDNA | [14] |
| | | Gene | [24] |
| cs1 | Pepper | cDNA | [41] |
| cs1 | Rice | cDNA | [42] |
| cs1 | Spinach | cDNA | [43] |
| cs1 | Tobacco | cDNA | [44] |
| cs1 | Maize | cDNA | [45] |
| cs1 | Radish | cDNA | [46] |
| cs1 | Soybean | cDNA | [47] |
| cs1 | Carrot | cDNA | GenBank accession no. Z17398 |
| cs2 | *Arabidopsis* | cDNA | GenBank accession no. X98275 |
| | | Gene | [15] |
| cm1 | Cotton | cDNA | [27] |
| cm1 | Arabidopsis | cDNA | This paper and [5] |
| cm2 | Spinach | cDNA | [32] |
| cm3 | *M. crystallinum* | cDNA | GenBank accession no. U43561 |
| cht | Pumpkin | cDNA | [33] |
| cht/chs | Spinach | cDNA | [28] |
| cht | *Arabidopsis* | cDNA | This work and [5] |
| chs | *Arabidopsis* | cDNA | This work and [5] |
| chs | Tea | Protein | [30] |

### Three novel ATAs

The sequence of clone 109F7T7 encoding ATAcm1 consists of 1103 nt excluding the 3′ poly(A) tail. The open reading frame codes for 287 amino acid residues, of which residues 260–280 were predicted by TopPred to be a membrane-spanning segment.

The cDNA sequence of EST clone 63G9T7 coding for ATAchs consisted of 1018 nt but contained no start codon. A sequence fragment extending the sequence was therefore isolated from a cDNA library by PCR as described in the Materials and Methods section. An agarose gel of the PCR fragments (results not shown) revealed three fragment sizes. The shortest and most abundant of these fragments started at the same site as EST clone 63G9T7, indicating that there might be DNA structural reasons for this preference of clone start. The longest fragment was sequenced and extended the sequence of 63G9T7 by 195 bases at the 5′ end, showing a start codon at position 95. The entire translated sequence, ATAchs, contains a segment 73 % identical with the 22 N-terminal residues of mature tea chloroplast ascorbate peroxidase (Figure 1 and see also Table 1). Assuming a similar cleavage site in pre-ATAchs, the mature ATAchs consists of 273 amino acid residues and is preceded by a 99-residue transit peptide as shown in Figure 2.

The complete sequence of EST clone 193D22T7 encoding ATAcht consists of 1442 nt excluding the 3′ poly(A) tail. The first AUG codon at position 47 initiates an open reading frame coding for 426 amino acid residues. Again, assuming a similar cleavage site in pre-ATAcht to that observed in the tea chloroplast ascorbate peroxidase [10], a transit peptide consisting of 48 residues is obtained, leaving, in this case, a mature ATAcht protein consisting of 348 residues (Figure 1). In addition, a membrane-spanning segment, residues 319–338, was predicted

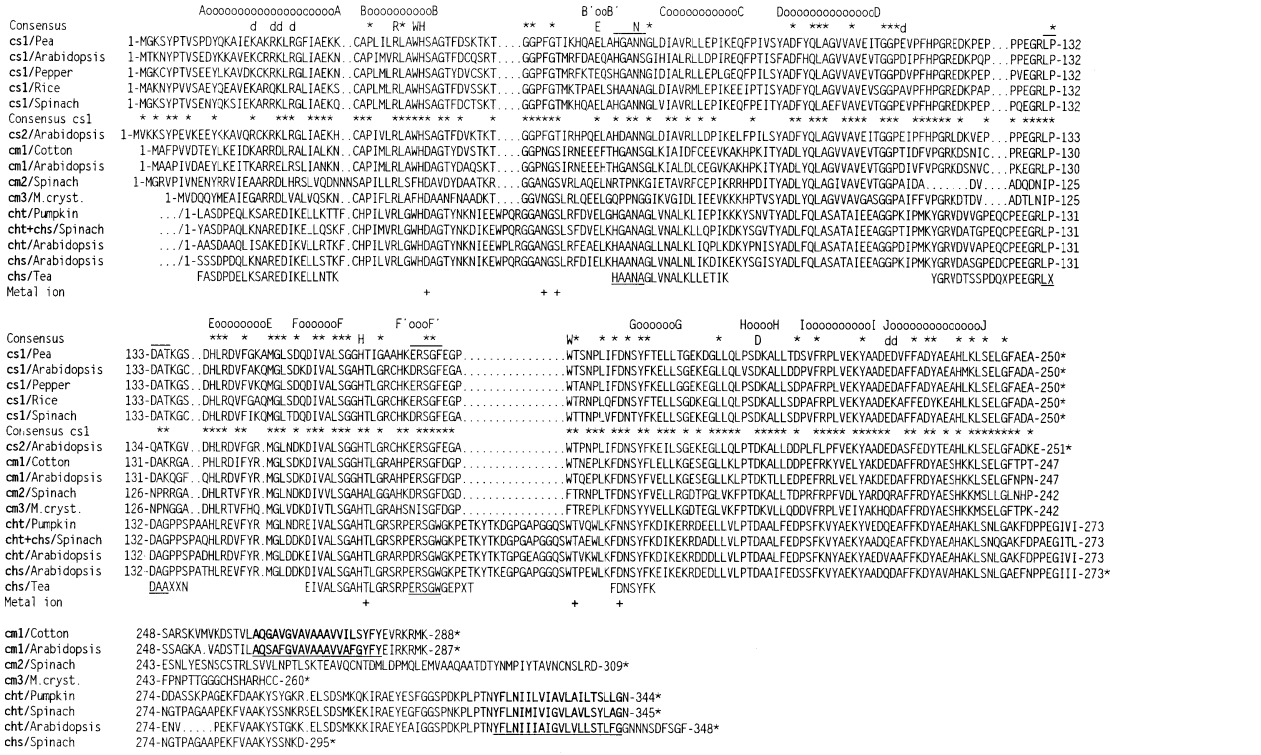

**Figure 1 Alignment of ascorbate peroxidase sequences**

References are as listed in Table 1. Sequences are categorized by experiment or homology as cs, cm, cht or chs. Only five out of ten known sequences of soluble cytosol cs1-type ascorbate peroxidases (Table 1) are shown. Peptide sequences from tea chloroplast ascorbate peroxidase [30] are shown in the last line. Putative signal peptides of the chloroplast sequences are indicated by . . . / and shown in Figure 2. A C-terminus defined by a stop codon is shown by a star. In the top line, helices A through J, observed in the crystal structure of pea cytosol ascorbate peroxidase [35] are shown. In the second line, residues conserved in all ascorbate peroxidases are indicated by stars or by a capital letter for active site residues, and by the letter d for residues involved in electrostatic interaction between dimers [35]. In the line below cytosol cs1 type peroxidases, residues conserved in 10 apparently orthologous cytosol ascorbate peroxidases are marked by stars. The last line indicates putative metal-ion ligands (+). Three conserved segments that form the surface of the distal haem entrance channel in pea ascorbate peroxidase are marked with a horizontal line above and below the aligned sequences. C-terminal transmembrane segments predicted by TopPred [22] are shown in bold and underlined.

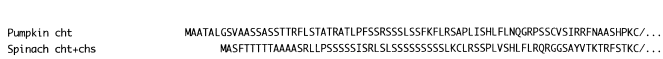

**Figure 2 Chloroplast transit sequences observed in ascorbate peroxidases**

A segment cloned by PCR extending the EST sequence of ATAchs is underlined. All sequences have a long segment with a high percentage of hydroxy amino acids as generally observed in transit peptides [26]. The last 20 residues seem to have a normal content of hydroxy amino acids and a relatively high content of basic amino acids.

by TopPred [22]. EST clone 110L16T7 was also sequenced. The coding sequence was identical with the C-terminal half of ATAcht, but the sites of polyadenylation were different.

Figure 2 shows chloroplast transit sequences for ascorbate peroxidases. It is seen that the general characteristics of a chloroplast transit sequence, such as a high content of hydroxy amino acids and the absence of tyrosine and carboxylic acid residues [25,26], are fulfilled in all cases, except for the presence of a glutamic residue near the first methionine of ATAchs. In all four sequences position −2 relative to the cleavage site was occupied by a lysine residue. The significance of this observation is unknown.

## Putative cellular locations of ascorbate peroxidases

Ascorbate peroxidases homologous with the three novel types from *Arabidopsis* have been found very recently also in other plants [27–29]. The data support our predicted cellular locations of these ascorbate peroxidases, on the basis of the finding of long chloroplast-type transit sequences in ATAchs and ATAcht, and predicted membrane-spanning segments in ATAcm1 and ATAcht, and as stromal ATAchs was similar to the stromal tea ascorbate peroxidase sequenced partly at the protein level [30]. In total, we currently know the complete sequence of 11 cytosol soluble ascorbate peroxidases, of four cytosol membrane-bound (putative) ascorbate peroxidases, and of two chloroplast soluble stromal ascorbate peroxidases and three ascorbate peroxidases predicted to be membrane-bound to the thylakoid. The structural divergence of the catalytic domain (i.e. excluding N-terminal signals and C-terminal membrane-spanning extensions) of all 20 ascorbate peroxidases is illustrated in Figure 3, whereas only the most diverse ascorbate peroxidases are aligned in Figure 1.

## Two types of cytosol soluble ascorbate peroxidase (cs1, cs2)

The right-hand side of the tree in Figure 3 shows a cluster of 10 soluble cytosol ascorbate peroxidases representing 10 plant species, and demonstrates the evolution of these plant species through the divergence of their presumed orthologous ascorbate

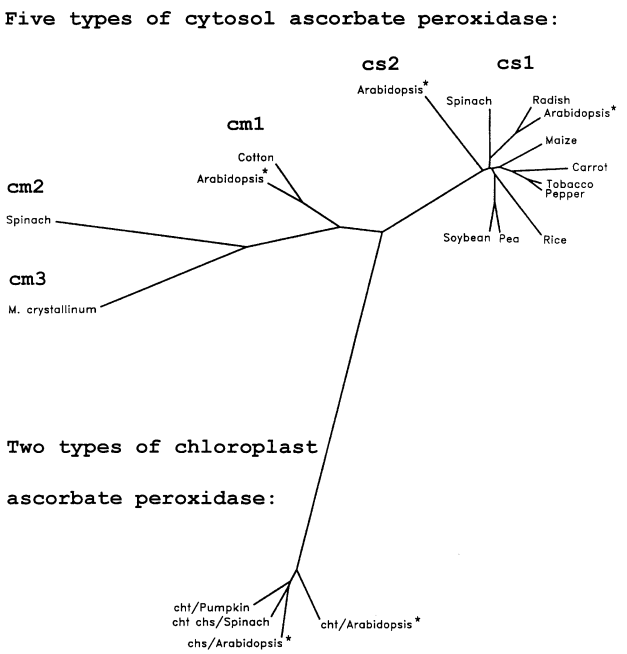


**Figure 3 Phylogenetic tree of seven types of ascorbate peroxidase**

References in Table 1. The tree was made with the programs DIST, FITCH and DRAWTREE of the PHYLIP program package [18]. The tree is fitted to the estimated pairwise PAM distances [20] between sequences by using the Fitch–Margoliash method [19]. The right branch of the tree, the cytosol-soluble (cs1) ascorbate peroxidases, represents orthologous ascorbate peroxidases only and is therefore indicative of the evolution of these 10 plant species, whereas the tree, including *Arabidopsis* ATAs only (marked with a star), is indicative of the paralogous evolution of the different types of ascorbate peroxidase.

peroxidases. All pairs in this group show between 78 and 95 % amino acid identity. This family of soluble cytosol ascorbate peroxidases represented by *Arabidopsis* ATAcs1 seems to fulfil an important function as a general scavenger of $H_2O_2$ and seems to be widely expressed as also indicated by the 34 EST clones found for ATAcs1. $H_2O_2$ has been proposed as a second messenger in plant cells [31], and because ascorbate peroxidase can efficiently remove low concentrations of $H_2O_2$, in contrast with catalase, it might have an important regulatory role.

The gene encoding a second family of soluble cytosol ascorbate peroxidases was recently cloned from *Arabidopsis* [15] (designated ATAcs2 in this paper), but is not represented in dbEST. ATAcs1 and ATAcs2 show 78 % amino acid sequence identity. Taking dbEST occurrence as a rough indication of expression, ATAcs2 seems to be rare, and no specific biological function has been proposed. However, the frequency of EST occurrence might be affected by cDNA library normalization procedures [5].

### Three types of putative membrane-bound cytosol ascorbate peroxidase (cm1, cm2, cm3)

The catalytic domain of ATAcm1 consisted of 247 residues but showed a C-terminal extension of 41 residues including a transmembrane segment. The catalytic domain of ATAcm1 shows only 65 % amino acid identity with ATAcs1 and ATAcs2. However, ATAcm1 matches a cotton ascorbate peroxidase bound to the membrane of glyoxysomes [27] by 85 % amino acid identity over their entire 287/288 residues (Figure 1). The cotton ascorbate peroxidase was shown to function on the matrix side of the glyoxysomal membrane, as trypsin treatment of intact and of perfused organelles released a 28 kDa ascorbate peroxidase in the latter case only [27]. This result is in accord with the 'positive-inside' rule of von Heijne [21], which predicts that the positively charged last five residues, RKRMK, of these proteins will remain on the cytosol side of the membrane. It was suggested, however [29], that the similar pumpkin ascorbate peroxidase was acting at the outside of microbodies. The biological function of this type of ascorbate peroxidase in protection of the glyoxysomal membrane against $H_2O_2$ in cotton has been discussed [27].

Figure 3 shows two additional branches of the cytosolic part of the tree: a spinach ascorbate peroxidase [32] represents the cm2 type ascorbate peroxidase, and an ascorbate peroxidase from the common ice plant *Mesembryanthemum crystallinum* (GenBank accession no. u43561) represents the cm3 type ascorbate peroxidase. No ATAs of these two types are known at present. The *M. crystallinum* ascorbate peroxidase shows a C-terminal extension of 18 residues largely hydrophobic in nature but ending with a basic segment, HARHCC. The spinach ascorbate peroxidase shows 67 additional residues at the C-terminus ending with a stretch of 18 hydrophobic residues and the sequence RD. These ascorbate peroxidases might be membrane-bound despite the fact that they do not attain a significant score in the TopPred program. The locations and functions of these two putative cytosol-type ascorbate peroxidases are unknown.

### Two types of chloroplast ascorbate peroxidase (chs, cht)

The catalytic domains, residues 1–273 (Figure 1), of ATAchs and ATAcht are 82 % identical, but only ATAcht shows a C-terminal extension with a predicted transmembrane segment. In addition the chloroplast transit sequences are different for the two (Figure 2). Hence *Arabidopsis* clearly has two different genes for chloroplast ATAs. This is in contrast with recent results obtained for stromal and thylakoid chloroplast ascorbate peroxidases from spinach [28]. In spinach the transit sequences and the catalytic domains are 100 % identical except that the stromal catalytic domain shows an additional C-terminal aspartic residue (Figure 1). This spinach enzyme also has a 22-residue C-terminal extension compared with ATAchs. Ishikawa et al. [28] suggested that the spinach stromal and thylakoid enzymes might arise by alternative splicing. The thylakoid type has also been cloned recently in pumpkin [33]. The pairwise sequence identities in the three cht-type ascorbate peroxidases are 82–84 %. The C-terminal extensions are very similar for all three, showing a high and similar content of basic and acidic residues and a hydrophobic segment (Figure 1). The cht type of ascorbate peroxidases is thought to be located on the stromal side of the thylakoid membrane [34].

### Structure and function of ascorbate peroxidases

The catalytic domain of ascorbate peroxidases consists of 241–273 residues folded into characteristic $\alpha$-helices A to J [35], which are conserved in all members of the plant peroxidase superfamily [2]. In accord with the residue numbers given to pea ascorbate peroxidase in its crystal structure [35], we retain Met-1, despite the fact that it was removed from the native pea protein [36]. We further assume that Met-1 is absent from all cytosol ascorbate peroxidases found as in other proteins where methionine is followed by a small residue [37]. The mature chloroplast proteins are shorter by eight N-terminal residues relative to pea cytosol ascorbate peroxidase, as proved for the tea and spinach proteins [10,28]. Three larger loops between helices B–B′, D–E, F′–G and seven additional residues at the C-terminus of the catalytic

domain are seen in chloroplast ascorbate peroxidases compared with cytosol ascorbate peroxidases. Seven charged residues, which participate in electrostatic interactions in the dimer interface of pea cytosol ascorbate peroxidase [35] (indicated with a letter d in Figure 1) are highly conserved in all soluble cytosol ascorbate peroxidases and in the glyoxysomal membrane-bound ascorbate peroxidase, except for Gln-21 and Ala-112 in rice ascorbate peroxidase and Thr-110 in the glyoxysomal cotton ascorbate peroxidase. Therefore we conclude that the cytosol types of ascorbate peroxidase, cs1, cs2 and cm1, are all dimeric, in contrast with the putative cm2 type represented by spinach ascorbate peroxidase and the cm3 type represented by *M. crystallinum* ascorbate peroxidase. The charged residues are not conserved in the sequences of chloroplast ascorbate peroxidase (Figure 1), in agreement with experimental evidence that these enzymes are monomers [10]. This pattern of dimeric as opposed to monomeric ascorbate peroxidase suggests that the branch of the cm2 and cm3 types diverged from the other three cytosolic families at an early stage.

All fungal (Class II) and plant secretory (Class III) peroxidases bind a distal and a proximal $Ca^{2+}$ ion via the side chains marked in Figure 1 by + below the sequences. The crystal structure of pea cytosol ascorbate peroxidase [35] shows only a single monovalent ion (apparently $K^+$) in the proximal domain, in agreement with the presence of only one carboxylate ligand at this site. The $K^+$ ion is essential to the reaction mechanism of pea ascorbate peroxidase [38]. If we first consider the proximal metal-ion-binding site of pea ascorbate peroxidase (Thr-164, Thr-180, Asp-187), these three side chains are invariant in all types of ascorbate peroxidase shown in Figure 1, indicating that a monovalent metal ion is present in all (the pumpkin cht type might be an exception). Turning to the potential metal-binding site in the distal domain (positions 43, 57 and 59 in pea ascorbate peroxidase), which is filled by $Ca^{2+}$ in fungal and plant secretory peroxidases [39], all ascorbate peroxidases, except the soluble cytosol cs1 and cs2 types, have an Asp residue at position 43, Asn at position 57 and Ser at position 59. In fungal and plant peroxidases the equivalent residues are Asp, Asp and Ser. We therefore propose that ascorbate peroxidase types cm1, cm2, cm3, chs and cht also have their potential metal-binding site in the distal domain occupied by a monovalent metal ion.

## Putative active site residues

The plant peroxidase superfamily [2] is characterized by the presence of a histidine residue (His-42 of pea ascorbate peroxidase) and an arginine residue (Arg-38) in the distal cavity of the haem group. The distal histidine is hydrogen-bonded with the side chain of an asparagine residue (Asn-71), which in turn is hydrogen-bonded with the backbone carbonyl of a glutamic residue (Glu-65). At the closed proximal site of haem, a histidine residue (His-163) is bound to haem iron and forms a strong hydrogen bond with a buried aspartate residue (Asp-208). These six residues are conserved in all ascorbate peroxidases (Figure 1). So far, in the Class I peroxidases (of prokaryotic origin) a proximal tryptophan residue (Trp-179) located just below the haem is a third participant in a hydrogen-bonding network with the proximal histidine and aspartate residues. Also this tryptophan residue is conserved in most ascorbate peroxidases but interestingly is changed to phenylalanine in the spinach cm2 type and in *M. crystallinum* cm3-type ascorbate peroxidases. Phenylalanine is the predominant residue at this position in the fungal Class II and plant Class III peroxidases. Furthermore these same two ascorbate peroxidases have the conserved distal tryptophan residue (Trp-41) also changed to phenylalanine, the residue invariantly found at this position in Class II and III peroxidases.

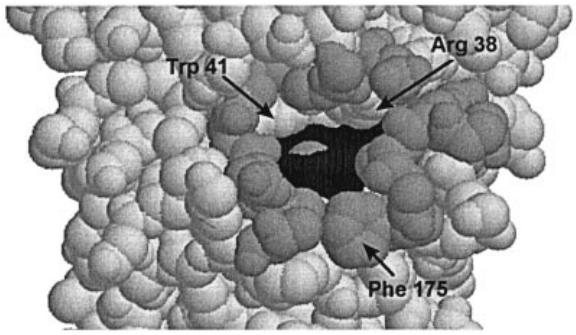


**Figure 4 Substrate channel of pea ascorbate peroxidase viewed from the molecular surface**

The three peptide segments 69–72, 131–135 and 171–175 are shown in grey. Phe-175 which is replaced by a Trp residue in the chloroplast ascorbate peroxidases, is labelled. A glimpse of the haem group is seen in black, the iron in grey. The active site Arg-38 and Trp-41 are also visible. The figure was made by using the atomic coordinates of pea ascorbate peroxidase (PDB ID Code: 1APX) and the RASMOL program [48].

The peptide segments residues 69–72, 131–135 and 171–175 form the surface of the active site entrance channel near the haem edge of pea ascorbate peroxidase, and are assumed to be essential to the interaction of ascorbate peroxidase with ascorbate (Figure 4). These segments are characterized by XANX, LPDAX and (E)RSGF/W respectively in all ascorbate peroxidases except for the *M. crystallinum* cm3 type and spinach cm2 type. It is seen that the spinach cm2 and *M. crystallinum* cm3 ascorbate peroxidases are distinct in many functionally important ways, and we speculate that they might have a preference for substrates other than ascorbate. The 171–175 segment has Phe-175 at the haem edge in all cytosol-type ascorbate peroxidases but Trp in chloroplast ascorbate peroxidases (Figure 1). We propose that a Trp residue at position 175 is the major determinant of the greater specificity of chloroplast ascorbate peroxidases towards ascorbate [10].

## Evolution

The structural characteristics of ascorbate peroxidases discussed above provide the basis for a discussion of the evolution of the seven ascorbate peroxidase types. The earliest event in ascorbate peroxidase evolution resulted in the rise of separate cytosol and chloroplast ascorbate peroxidases, clearly demonstrated by the sequence gap pattern in ascorbate peroxidases between these groups (Figure 1). The chloroplast branch has apparently diverged only recently into stromal and thylakoid-bound types in *Arabidopsis*, whereas results on the spinach ascorbate peroxidases [28] suggest that either differential splicing or a very recent gene duplication might provide the two forms. It seems that the chs–cht gene duplication is species-specific and not inherited from an ancestor plant.

The molecular properties of the cm2 (spinach) and cm3 (*M. crystallinum*) types are very different from those of the cytosol cs1, cs2 and cm1 types with respect to subunit composition (monomeric compared with dimeric), active site residues (two phenylalanines compared with two tryptophans), and site of reducing substrate interaction, indicating that these branches

diverged at a very early stage. However, ascorbate peroxidases from more plant species are needed to characterize the proposed cm2 and cm3 types more fully in respect of their cellular location, physiological and biochemical function. The same applies to the cs2 type, which might be species-specific and apparently separated from cs1 only recently in evolutionary terms. The intra-species PAM distances given for the cs1, cm1, chs and cht types in Figure 3 will be useful for the future classification of new ascorbate peroxidases belonging to one of the present seven types or constituting novel types.

Finally, the indication that all ascorbate peroxidase types, except cs1 and cs2, contain a conserved metal-ion-binding site in both the distal and proximal domains supports the hypothesis of Welinder and Gajhede [39] that present-day peroxidases arose by gene duplication of an ancestral gene that presumably encoded five helices, a His-to-haem ligand, an Asp or Asn residue hydrogen-bonded to this His, and a $Ca^{2+}$-binding site. On the present grounds a monovalent metal ion could equally well be the original ion. In the above-mentioned ascorbate peroxidases both ion-binding sites were retained, whereas in soluble cytosol ascorbate peroxidases one was lost. In yeast mitochondrial cytochrome *c* peroxidase both were lost [35,39].

## Conclusion

The present analysis of plant ascorbate peroxidases provides a rational basis for the selection of model peroxidases for the study of physiological roles, and for the study of structure–function relationships within the family of ascorbate peroxidases and within the superfamily of haem-containing peroxidases in general.

We thank the U. S. and European EST programs [5,6] and the *Arabidopsis* Biological Resource Center at Ohio State University for the provision of public services; Dr. S. K. Rasmussen (Risø National Laboratory, Denmark), for access to automated DNA sequencing and for assistance in interpretation of the data; Dr. D. Tremousaygue (UMR05 CNRS/INRA, Toulon, France) for generously donating the AC16H cDNA library; and Dr. N. C. Veitch (Royal Botanic Gardens, Kew, Surrey, U.K.) for a critical reading of the manuscript. This work was supported by the Danish Agricultural and Veterinary Research Council (5.23.26.10-1) and the Danish Natural Science Research Council (9502825).

## REFERENCES

1 Mittler, R. and Zilinskas, B. A. (1991) FEBS Lett. **289**, 257–259
2 Welinder, K. G. (1992) Curr. Opin. Struct. Biol. **2**, 388–393
3 Boguski, M. S., Lowe, T. M. J. and Tolstoshev, C. M. (1993) Nature Genet. **4**, 332–333
4 Boguski, M. S. (1995) Trends Biochem. Sci. **20**, 295–296
5 Newman, T., De Bruijn, F. J., Green, P., Keegstra, K., Kende, H., McIntosh, L., Ohlrogge, J., Raikhel, N., Somerville, S., Thomashow, M. et al. (1994) Plant Physiol. **106**, 1241–1255
6 Cooke, R., Raynal, M., Laudié, M., Grellet, F., Delseny, M., Morris, P. C., Guerrier, D., Giraudat, J., Quigley, F., Clabault, G. et al. (1996) Plant J. **9**, 101–124
7 Sasaki, T., Song, J., Koga-Ban, Y., Matsui, E., Fang, F., Higo, H., Nagasaki, H., Hori, M., Miya, M., Murayama-Kayano, E. et al. (1994) Plant J. **6**, 615–624
8 Miyake, C., Michihata, F. and Asada, K. (1991) Plant Cell Physiol. **32**, 33–44
9 Dalton, D. A. (1991) in Peroxidases in Chemistry and Biology (Everse, J., Everse, K. E. and Grisham, M. B., eds.), pp. 139–153, CRC Press, Boca Raton, FL
10 Chen, G. X. and Asada, K. (1989) Plant Cell Physiol. **30**, 987–998
11 Chen, G. X. and Asada, K. (1992) Plant Cell Physiol. **33**, 117–123
12 Durner, J. and Klessig, D. F. (1995) Proc. Natl. Acad. Sci. U.S.A. **92**, 11312–11316
13 Nakano, Y. and Asada, K. (1987) Plant Cell Physiol. **28**, 131–140
14 Kubo, A., Saji, H., Tanaka, K. and Kondo, N. (1992) Plant Mol. Biol. **18**, 691–701
15 Santos, M., Gousseau, H., Lister, C., Foyer, C., Creissen, G. and Mullineaux, P. (1996) Planta **198**, 64–69
16 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) J. Mol. Biol. **215**, 403–410
17 Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994) Nucleic Acids Res. **22**, 4673–4680
18 Felsenstein, J. (1989) Cladistics **5**, 164–166
19 Fitch, W. M. and Margoliash, E. (1967) Science **155**, 279–284
20 Dayhoff, M. O. (1978) Atlas of Protein Sequence and Structure 5, suppl. 3, National Biomedical Research Foundation, Washington, DC
21 von Heijne, G. (1992) J. Mol. Biol. **225**, 487–494
22 Claros, M. G. and von Heijne, G. (1994) Comp. Appl. Biosci. **10**, 685–686
23 Regad, F., Bardet, C., Tremousaygue, D., Moisan, A., Lescure, B. and Axelos, M. (1993) FEBS Lett. **316**, 133–136
24 Kubo, A., Saji, H., Tanaka, K. and Kondo, N. (1993) FEBS Lett. **315**, 313–317
25 Schatz, G. and Dobberstein, B. (1996) Science **271**, 1519–1526
26 Douwe de Boer, A. and Weisbeek, P. J. (1991) Biochim. Biophys. Acta **1071**, 221–253
27 Bunkelmann, J. and Trelease, R. N. (1996) Plant Physiol. **110**, 589–598
28 Ishikawa, T., Sakai, K., Yoshimura, K., Takeda, T. and Shigeoka, S. (1996) FEBS Lett. **384**, 289–293
29 Yamaguchi, K., Mori, H. and Nishimura, M. (1995) Plant Cell Physiol. **36**, 1157–1162
30 Chen, G. X., Sano, S. and Asada, K. (1992) Plant Cell Physiol. **33**, 109–116
31 Chen, Z., Silva, H. and Klessig, D. F. (1993) Science **262**, 1883–1886
32 Ishikawa, T., Sakai, K., Takeda, T. and Shigeoka, S. (1995) FEBS Lett. **367**, 28–32
33 Yamaguchi, K., Hayashi, M. and Nishimura, M. (1996) Plant Cell Physiol. **37**, 405–409
34 Miyake, C. and Asada, K. (1992) Plant Cell Physiol. **33**, 541–553
35 Patterson, W. R. and Poulos, T. L. (1995) Biochemistry **34**, 4331–4341
36 Mittler, R. and Zilinskas, B. A. (1991) Plant Physiol. **97**, 962–968
37 Hirel, P. H., Schmitter, J. M., Dessen, P., Fayat, G. and Blanquet, S. (1989) Proc. Natl. Acad. Sci. U.S.A. **86**, 8247–8251
38 Poulos, T. L., Patterson, W. R. and Sundaramoorthy, M. (1995) Biochem. Soc. Trans. **23**, 228–232
39 Welinder, K. G. and Gajhede, M. (1993) in Plant Peroxidases: Biochemistry and Physiology (Welinder, K. G., Rasmussen, S. K., Penel, C. and Greppin, H., eds.), pp. 35–42, University of Geneva, Geneva
40 Mittler, R. and Zilinskas, B. A. (1992) J. Biol. Chem. **267**, 21802–21807
41 Schantz, M. L., Schreiber, H., Guillemaut, P. and Schantz, R. (1995) FEBS Lett. **358**, 149–152
42 Morita, S., Kaminaka, H., Yokoi, H., Masumura, T. and Tanaka, K. (1997) Plant Physiol. **113**, 306
43 Webb, R. P. and Allen, R. D. (1995) Plant Physiol. **108**, 1325
44 Örvar, B. L. and Ellis, B. E. (1995) Plant Physiol. **108**, 839–840
45 Van Breusegem, F., Villarroel, R., Van Montagu, M. and Inze, D. (1995) Plant Physiol. **107**, 649–650
46 Lopez, F., Vansuyt, G., Derancourt, J., Fourcroy, P. and Casse-Delbart, F. (1994) Cell Mol. Biol. **40**, 85–90
47 Chatfield, M. and Dalton, D. A. (1993) Plant Physiol. **103**, 661–662
48 Sayle, R. A. and Milner-White, E. J. (1995) Trends Biochem. Sci. **20**, 374–376

Received 10 March 1997/6 May 1997; accepted 14 May 1997