ELSEVIER

Gene 256 (2000) 169–182

GENE
AN INTERNATIONAL JOURNAL ON GENES, GENOMES AND EVOLUTION

www.elsevier.com/locate/gene

Common phylogeny of catalase-peroxidases and ascorbate peroxidases

Marcel Zámocký [a,b,*], Štefan Janeček [b], Franz Koller [a]

[a] *Institute of Biochemistry and Molecular Cell Biology, University of Vienna, and Ludwig Boltzmann Forschungsstelle, Dr. Bohrgasse 9, A-1030 Vienna, Austria*

[b] *Institute of Microbiology, Slovak Academy of Sciences, Stefanikova 3, SK-81434 Bratislava, Slovakia*

Received 29 February 2000; received in revised form 19 July 2000; accepted 31 July 2000
Received by T. Gojobori

Abstract

Catalase-peroxidases belong to Class I of the plant, fungal, bacterial peroxidase superfamily, together with yeast cytochrome c peroxidase and ascorbate peroxidases. Obviously these bifunctional enzymes arose via gene duplication of an ancestral hydroperoxidase. A 230-residues long homologous region exists in all eukaryotic members of Class I, which is present twice in both prokaryotic and archaeal catalase-peroxidases. The overall structure of eukaryotic Class I peroxidases may be retained in both halves of catalase-peroxidases, with major insertions in several loops, some of which may participate in inter-domain or inter-subunit interactions.

Interspecies distances in unrooted phylogenetic trees, analysis of sequence similarities in distinct structural regions, as well as hydrophobic cluster analysis (HCA) suggest that one single tandem duplication had already occurred in the common ancestor prior to the segregation of the archaeal and eubacterial lines. The C-terminal halves of extant catalase-peroxidases clearly did not accumulate random changes, so prolonged periods of independent evolution of the duplicates can be ruled out. Fusion of both copies must have occurred still very early or even in the course of the duplication. We suggest that the sparse representatives of eukaryotic catalase-peroxidases go back to lateral gene transfer, and that, except for several fungi, only single copy hydroperoxidases occur in the eukaryotic lineage.

The N-terminal halves of catalase-peroxidases, which reveal higher homology with the single-copy members of the superfamily, obviously are catalytically active, whereas the C-terminal halves of the bifunctional enzymes presumably control the access to the haem pocket and facilitate stable folding. The bifunctional nature of catalase-peroxidases can be ascribed to several unique sequence peculiarities conserved among all N-terminal halves, which most likely will affect the properties of both haem ligands.
© 2000 Elsevier Science B.V. All rights reserved.

Keywords: Catalase; Cytochrome c peroxidase; Gene duplication; Hydrophobic cluster analysis; Unrooted phylogenetic tree

1. Introduction

Three groups of enzymes effectively catalyse the dismutation of hydrogen peroxide into oxygen and water: typical haem catalases ('true catalases'), manganese catalases ('pseudo-catalases'), and bifunctional haem proteins, generally addressed as catalase-peroxidases. The evolution of true catalases has been analysed in some detail (Klotz et al., 1997; Zámocký et al., 1997), whereas the phylogenetic trends within the two other groups have not been analysed so far. Catalase-peroxidases (currently 19 complete and two incomplete nucleotide sequences are known) constitute a fairly homogeneous group. Irrespective of their pronounced catalytic reactivity, they are not related to typical haem catalases. Sequence analysis revealed their close linkage to eukaryotic haem peroxidases of Class I of the plant, fungal, bacterial peroxidase superfamily (Welinder, 1992). This superfamily includes three independent evolutionary lineages: Class I (peroxidases of prokaryotic origin), which contains the large family of ascorbate peroxidases from higher plants and green algae, cytochrome c peroxidase (CCP) from *Saccharomyces cerevisiae*, and catalase-peroxidases. Class

Abbreviations: APX, ascorbate peroxidase; CCP, yeast cytochrome c peroxidase; CP, catalase-peroxidase; HPI, hydroperoxidase I; HCA, hydrophobic cluster analysis; ORF, open reading frame; ROS, reactive oxygen species.

* Corresponding author. Tel.: +43-1-4277-52809; fax: +43-1-4277-9528.
E-mail address: mz@abc.univie.ac.at (M. Zámocký)

0378-1119/00/$ - see front matter © 2000 Elsevier Science B.V. All rights reserved.
PII: S0378-1119(00)00358-9

II comprises extracellular fungal peroxidases (e.g. lignin peroxidase, manganese peroxidase), whereas the classical secretory plant peroxidases (e.g. horseradish peroxidase C) form Class III. With one exception, all catalase-peroxidases with known sequences are from prokaryotic sources (17 from bacteria, three from archaeons). Thus, Class I of the hydroperoxidase superfamily covers all three domains of life (Archaea, Bacteria, and Eukarya), but, owing to different physiological requirements, specialised subfamilies have evolved largely following this division of domains.

The crystal structures of two Class I peroxidases have been resolved to near atomic resolution: that of CCP from *S. cerevisiae* (Finzel et al., 1984) and that of the cytosolic ascorbate peroxidase (APX) from *Pisum sativum* (Patterson and Poulos, 1995). They show almost identical overall fold, which obviously is conserved within the entire superfamily. Both proteins reveal a high α-helical content (>46%, non-bundle) with no extended β-structures. Nine long and two short helices are numbered A to J according to the nomenclature of Finzel et al. (1984). Crystallisation of catalase-peroxidases has so far failed due to partial loss of prosthetic groups during purification, so no detailed structural information is yet available for them.

Ascorbate peroxidases play a key role in the removal of intracellular hydrogen peroxide, which they reduce with concomitant oxidation of ascorbate. Apparently, catalase-peroxidases and ascorbate peroxidases have similar functions in the defence of aerobically living cells against reactive oxygen species (ROS), and their reaction mechanisms presumably are closely related (Arnao et al., 1990). Functionally important sections, mainly involved in haem pocket formation, are highly conserved within this class, but catalase-peroxidases are characterised by several insertions in areas adjacent to the conserved regions.

The aim of this paper is a detailed phylogenetic analysis of all reported sequences of catalase-peroxidases. We conclude that a single tandem duplication event in a prokaryotic progenitor cell, accompanied or closely followed by concerted evolution of both duplicons, gave rise to all extant catalase-peroxidases. We try to attribute the bifunctional nature of this class of enzymes to several sequence peculiarities in both halves of all catalase-peroxidases.

2. Methods

2.1. Multiple sequence alignment

All amino acid sequences used in this study (Table 1) were extracted from the SWISS-PROT protein and GenBank DNA sequence databases. Multiple amino acid sequence alignment was performed on the prepared set of hydroperoxidases using CLUSTAL X (Jeanmougin et al., 1998). Slow/accurate alignments were performed with the following parameters: gap opening penalty 10.00, gap extension penalty 0.05, and the BLOSUM (Henikoff) protein weight matrices were used. The multiple alignment obtained was refined by eye. Evolutionary relatedness between individual enzymes is expressed as sequence similarity, i.e. the total percentage of identical residues and highly conservative replacements.

2.2. Hydrophobic cluster analysis

The sequences were visually inspected including the use of hydrophobic cluster analysis (HCA), which is a sensitive method at high sequence divergence (Callebaut et al., 1997). An HCA plot of the catalase-peroxidase from *M. tuberculosis* was prepared to demonstrate the eventuality of gene duplication in catalase-peroxidases (Fig. 2), and additional plots were prepared to allow a detailed analysis of sequence homologies between representative prokaryotic (bacterial and archaeal) and eukaryotic (plant cytosolic and thylakoid ascorbate peroxidases and yeast cytochrome c peroxidase) members of Class I of the plant peroxidase superfamily.

2.3. Construction of the phylogenetic trees

Unrooted distance trees were calculated by the PHYLIP (Phylogeny Inference Package) Version 3.5c (Felsenstein, 1989) using the Dayhoff PAM matrix. They were based on the alignment of the entire-length sequences of all enzymes included in this study. Before the analysis the bootstrap resampling method was applied with 100 replicates. The output was further subjected to the FITCH procedure (Fitch–Margoliash and least-squares distance methods) and the calculated trees were analysed by the CONSENSE method included in the PHYLIP package. The most probable tree was visualised by the program TreeView (Page, 1996; see Fig. 4).

To demonstrate the phylogenetic relationships in the two copies of the ancestral gene in catalase-peroxidases, the N- and C-terminal halves (as defined by HCA) of each catalase-peroxidase were aligned separately, and unrooted distance trees were calculated by the neighbour-joining method (Saitou and Nei, 1987). The PHYLIP format tree output was applied in each case using the bootstrapping procedure; the number of bootstrap trials used here was 1000. The trees were drawn by the program TreeView (Fig. 5).

2.4. Blast similarity searches

In order to substantiate the hypothesis of ancestral gene duplication, gapped BLAST searches were per-

Table 1
Sources of enzymes. Abbreviations for catalase-peroxidases, ascorbate peroxidases and cytochrome c peroxidase included in this evolutionary analysis, with their accession numbers from GenBank (gb) or SWISS-PROT (sp) and organisms from which they originate

Abbreviation	Accession no.	Enzyme	Organism (strain)
Arcfu-CP	gbAE000951	catalase-peroxidase	*Archaeoglobus fulgidus*
Bacst-CI	gbM29876	catalase I	*Bacillus stearothermophilus*
Caucr-CP	gbAF027168	catalase-peroxidase	*Caulobacter crescentus*
Ecoli-HPI	spP13029	catalase HPI	*Escherichia coli*
EcoliP-CP	gbX89017	EHEC-catalase-peroxidase	*E. coli* (0157: H7)
Halma-CP	gbY16851	catalase-peroxidase	*Haloarcula marismortui*
Halsa-CP	gbAF069761	catalase-peroxidase	*Halobacterium salinarum*
Legpn-CP	gbAF078110	catalase-peroxidase	*Legionella pneumophila*
Mycbo-CP	spP46817	catalase-peroxidase	*Mycobacterium bovis*
Mycfo-CP	gbY07865	catalase-peroxidase	*Mycobacterium fortuitum*
Mycfo-CPII	gbY07866	catalase-peroxidase II	*M. fortuitum*
Mycin-CP	spQ04657	catalase-peroxidase	*Mycobacterium intracellulare*
Mycsm-CP	gbU46844	catalase-peroxidase	*Mycobacterium smegmatis*
Myctu-CP	spQ08129	catalase-peroxidase	*Mycobacterium tuberculosis*
Mysgr-CP	gbAW179968	catalase-peroxidase	*Mycosphaerella graminicola*
Rhoca-CP	gbX71420	catalase-peroxidase	*Rhodobacter capsulatus*
Salty-CP	spP17750	catalase HPI	*Salmonella typhimurium*
Scosp-CP	gbD61378	catalase-peroxidase	*Synechococcus*
Scysp-CP	gbD90910	catalase HPI	*Synechocystis* sp. (PCC6803)
Strre-CP	gbY14317	catalase-peroxidase	*Streptomyces reticuli*
Yerpe-CP	gbAF135170	catalase-peroxidase	*Yersinia pestis*
Arath-APX1	spQ05431	ascorbate peroxidase 1	*Arabidopsis thaliana*
Arath-APX2	gbX98275	ascorbate peroxidase 2	*A. thaliana*
Capan-APX	gbX81376	ascorbate peroxidase	*Capsicum annuum*
Chlre-APX	gbAJ223325	ascorbate peroxidase	*Chlamydomonas reinhardtii*
Cucsa-APX	gbD88649	ascorbate peroxidase	*Cucumis sativus*
Fraan-APX	gbAF022213	ascorbate peroxidase	*Fragaria* x *ananassa*
Glyma-APX1	gbL10292	ascorbate peroxidase 1	*Glycine max* (soybean)
Glyma-APX2	gbU56634	ascorbate peroxidase 2	*G. max* (soybean)
Goshi-APX	gbU37060	ascorbate peroxidase	*Gossypium hirsutum*
Mescr-APX	gbU43561	ascorbate peroxidase	*Mesembryanthemum crystallinum*
Nicta-APX1	gbU15933	ascorbate peroxidase 1	*Nicotiana tabacum*
Nicta-APX2	gbD85912	ascorbate peroxidase 2	*N. tabacum*
Orysa-APX	gbD45423	ascorbate peroxidase	*Oryza sativa*
Pissa-APX	spP48534	ascorbate peroxidase	*P. sativum*
Vigun-APX	gbU61379	ascorbate peroxidase	*Vigna unguiculata*
Zeama-APX	gbZ34934	ascorbate peroxidase	*Zea mays*
Arath-APXt	gbX98926	ascorbate peroxidase (thylakoid)	*A. thaliana*
Cucsp-APXt	gbD83656	ascorbate peroxidase (thylakoid)	*Cucurbita* sp.
Spiol-APXt	gbD77997	ascorbate peroxidase (thylakoid)	*Spinacia oleracea*
Sacce-CCP	spP00431	cytochrome c peroxidase	*S. cerevisiae*

formed following the recommendations of the authors (Altschul et al., 1997). The degree of homology between all investigated members of the plant peroxidase superfamily was determined with TBLASTN, which was also applied for screening of genomes for distantly related homologous areas.

2.5. Secondary structure prediction

Secondary structure prediction for catalase-peroxidases was accomplished by using the program PHDsec (Rost and Sander, 1993) and the method developed by Rychlewski and Godzik (1997). During the latter procedure a sequence similarity search based on 16-residues long overlapping segments was performed against the database of known structures. The most similar segments were identified and their secondary structures were averaged and used to predict the structure of the target protein. Only the method of Rychlewski and Godzik (1997) is included in the results, since the results obtained by PHD were essentially the same.

3. Results

3.1. Comparison of functionally important motifs in Class I peroxidases

We have analysed 17 sequences of catalase-peroxidases from Eubacteria, three members from the domain

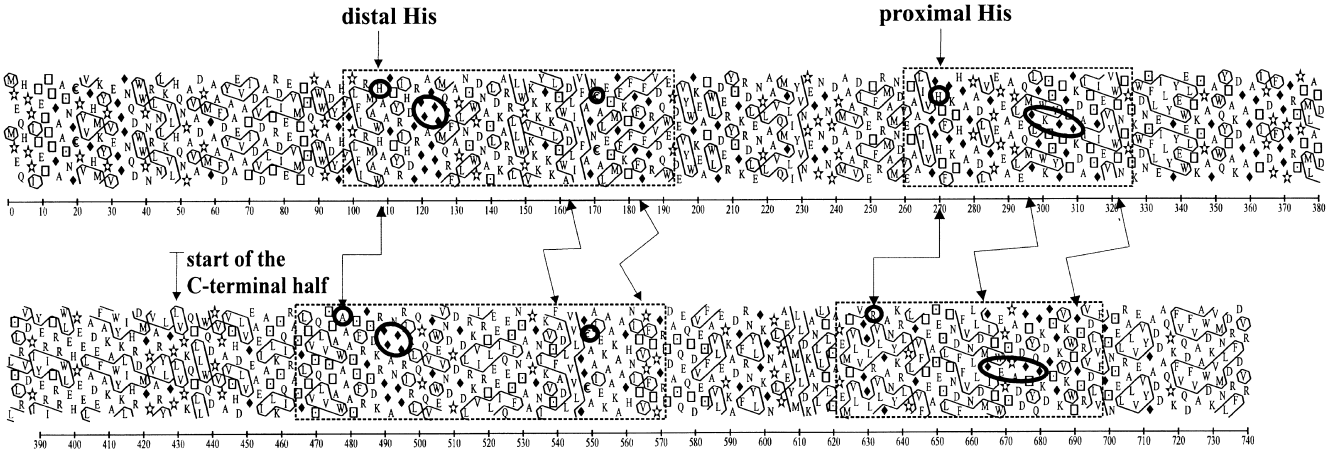

Fig. 1. HCA plot of the catalase-peroxidase amino acid sequence from *Mycobacterium tuberculosis* showing the correspondences between its N-terminal and C-terminal parts. In the HCA plot the amino acid sequence of a protein is drawn as an unrolled and duplicated longitudinal cut of a cylinder, where residues follow an α-helical pattern. Special symbols are used for glycine (♦), proline (★), cysteine (©), serine (□), and threonine (⊡). Clusters of hydrophobic residues are automatically contoured by the HCA-plot program. Analysis of the HCA plots was made following the published guidelines (Callebaut et al., 1997).

The distal and proximal haem ligands (His-108 and His-270 respectively) and the corresponding residues in the C-terminal half (Ala-478 and Arg-632), the corresponding glycines and the equivalent cysteine residues for both halves (Cys-171 and CYS-549) are highlighted in dark circles. Further sequence similarities are marked in contoured boxes with arrows showing the respective similarities between N- and C- terminal halves. The approximate start of the C-terminal part, in terms of the eventual duplication event (discussed in the text), is also indicated.

of Archaea, and one partial sequence from the fungus *M. graminicola* available in public databases (Table 1). The rather long coding sequence (average length 726 amino acids) can be divided into two halves, with the N-terminal halves typically being 32% longer than the C-terminal halves. Clear exceptions of this rule are the sequences of the putative catalase-peroxidase from *M. graminicola* and of the hydroperoxidase from *C. crescentus*, which are not yet completed, and that of *R. capsulatus*, possessing a 2.4-fold shorter C-terminal copy than its N-terminal counterpart.

The larger polypeptide size of catalase-peroxidases and distinct homologies between the two halves implicate a gene duplication event. This was first suggested by Welinder (1991) upon the analysis of repeated segments in hydroperoxidase I (HPI) from *E. coli*. The eventuality of this duplication could also be studied by classical sequence alignment; we preferred HCA analysis owing to the higher sensitivity of this method, which also provides us with some indisputable correspondences (see below). Duplication becomes immediately evident from inspection of Fig. 1, showing the HCA plot of the enzyme from *M. tuberculosis*, which is representative for the entire class. The longer N-terminal (residues 1–429) and the shorter C-terminal halves exhibit striking sequence similarities centred around the distal and proximal haem ligands (His108 and His270 respectively; unless otherwise specified, all amino acid numbering throughout the text corresponds to this representative), although only the N-terminal half contains these functionally essential residues. Additionally, three other regions appear well conserved between the two halves of *M. tuberculosis* catalase-peroxidase. (a) Residues 158–167: in the crystal structure of both APX and CCP the corresponding residues form the turn between helices C and D and the first part of helix D [throughout this paper the numbering of secondary structure elements refers to the crystal structure of *P. sativum* APX, Patterson and Poulos, 1995, a schematic presentation of which is shown in Fig. 2]. It also includes Asp163; an aspartate is invariantly found at the corresponding position in all plant peroxidases, which together with a glycine (186 in *M. tuberculosis*) and an arginine (187)

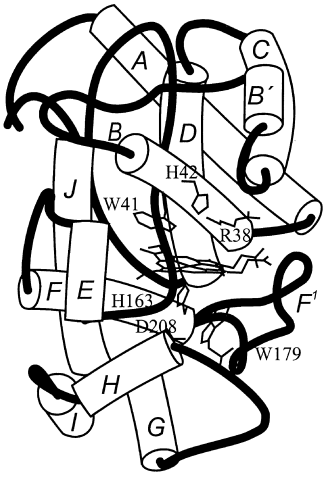

Fig. 2. Schematic presentation of the crystal structure of ascorbate peroxidase from *P. sativum*. The numbering of the secondary structural elements is according to Patterson and Poulos (1995). The prosthetic group and the side chains of functionally important residues on both haem sides are included.

forms a highly conserved buried salt bridge. (b) A short stretch between residues 179 and 189. In the two above mentioned Class I peroxidases of known structure this region reflects the first part of a long loop following helix D, connecting the two domains (Jespersen et al., 1997). It forms a sharp turn adjacent to helix D, and includes the residues Gly186 and Arg187, mentioned above. (c) The region between Gln295 and Trp321. In CCP and APX the tryptophan corresponding to Trp321 forms part of the catalytic triad His–Asp–Trp and therefore is of functional significance. A major part of region (c) belongs to a 35 residues long insertion at the position corresponding to helix F^1 in both peroxidases of known structure. This short helix forms part of a rather long loop forming the frontal edge of the haem cavity, controlling the lateral access to the prosthetic group. Thus, this loop is significantly longer in catalase-peroxidases than in the other members of Class I. Interestingly, the conservation of this area (c) among the two halves of catalase-peroxidases, though obvious from HCA-analysis, is not easily uncovered by more conventional multiple sequence alignments performed with CLUSTAL X. When only Class I peroxidases were analysed by this method, a unique insertion in the N-terminal halves of catalase-peroxidases was indicated in this region, and the situation did not change when Class II peroxidases were included. Only when members of Class III of the peroxidase superfamily were included, which also show an extended loop in this area, were the C-terminal halves of catalase-peroxidases correctly aligned. This clearly underlines the importance of applying two independent methods of sequence comparison.

The N-terminal halves of all catalase-peroxidases analysed exhibit high residue conservation along the whole sequence with 29% identity and 41% similarity, whereas the corresponding numbers are only 22% and 38% respectively for the C-terminal halves. Multiple sequence alignment (Fig. 3) reveals the main sequence features common to representatives of the different classes of this family. Only the N-terminal halves of the two catalase-peroxidases are included in Fig. 3. For the two enzymes of known structure the secondary structure elements are indicated in Fig. 3, as well as the corresponding data for the two catalase-peroxidases, pre-

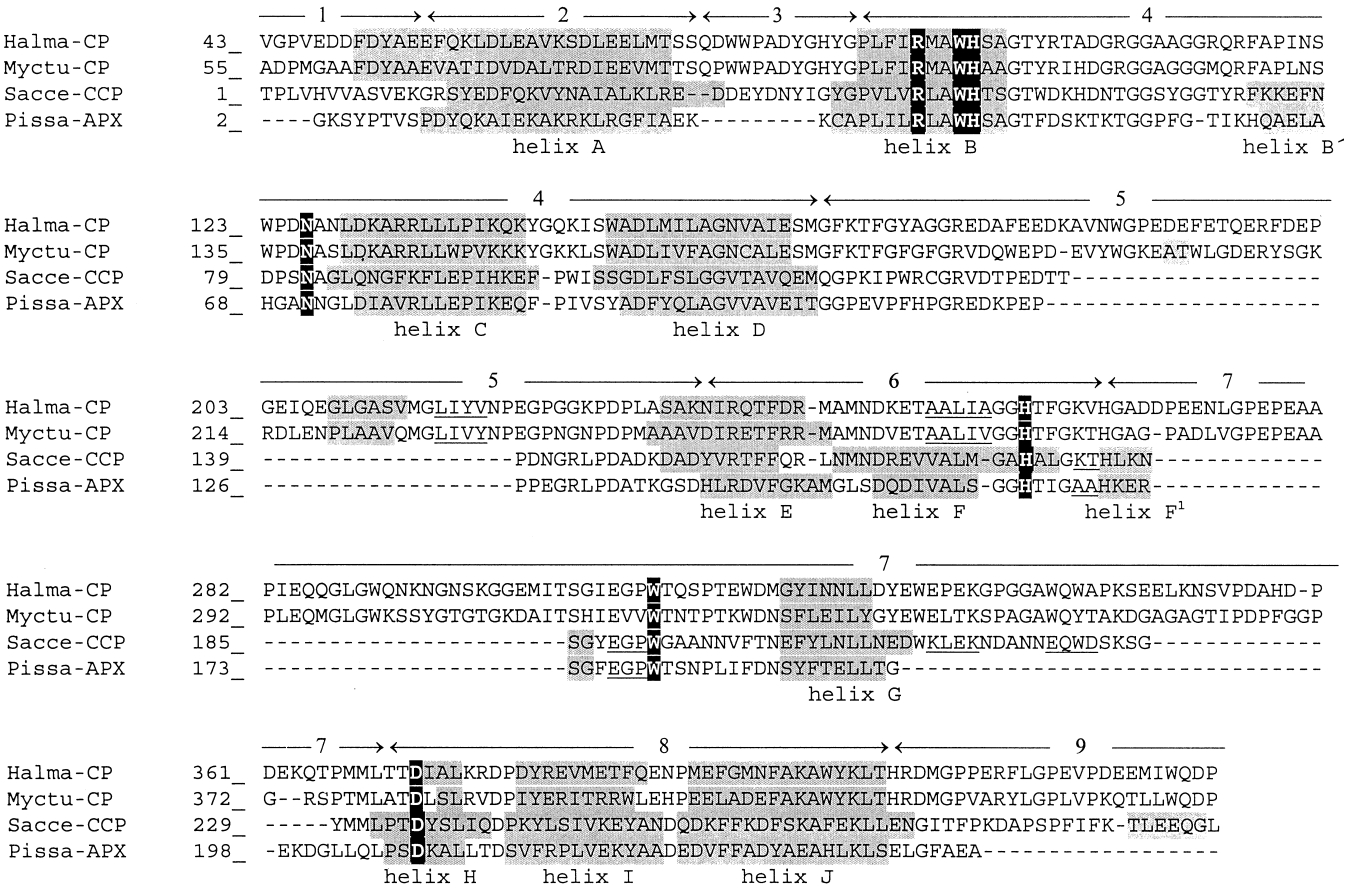

Fig. 3. Multiple sequence alignment of four representatives of Class I of the peroxidase superfamily. Abbreviations of enzyme sources are defined in Table 1. Numbers indicate the position of each segment within the corresponding sequence. Functionally important residues are in black boxes. Secondary structure elements (derived from crystal structure for Sacce-CCP and Pissa-APX, otherwise predicted as described in Section 2): α-helices are shown as grey boxes, β-strands are underlined. The borders of the segments that are presented in Table 2 are indicated with the respective numbering above the alignments.

dicted by the method of Rychlewski and Godzik (1997). An independent method, PHD, virtually revealed identical predictions (data not shown). Again, the region of highest homology centres around the essential residues in the distal haem pocket (Arg104, Trp107, and His108 in *M. tuberculosis* catalase-peroxidase), as well as the region adjacent to the proximal haem ligand (His270). APX and CCP are organised in two domains (distal and proximal with respect to the prosthetic group), connected by the long loop between helices D and E. Clearly, sequence homology between the different subgroups of Class I peroxidases is markedly higher in the distal domain. The most striking differences between catalase-peroxidases and the two other subclasses, however, are several long insertions. The loop connecting helices A and B is much longer in catalase-peroxidases than in the other members of Class I (see below). Two long insertions are located in stretches connecting helices D and E (38 to 40 residues), and G and H (17 to 33 residues). A 35 to 36 residues long insertion is found within the short α-helix F^1, which most likely does not exist in catalase-peroxidase. An insertion of about the same length is found in APX from *C. reinhardtii*, and also, though distinctly shorter, in another strain of *Chlamydomonas*, and in chloroplast-localised ascorbate peroxidases.

From a comparison of this alignment with a structure-based alignment including members of all three classes of the superfamily (Gajhede et al., 1997) the relative significance of these insertions can be evaluated. Similar, though shorter insertions in surface loops are found in Class II peroxidases (between helices G and H) and Class III (between helices F and G), whereas the long insertion between helices D and E is typical for catalase-peroxidases.

According to secondary structure prediction, most secondary structural elements defining the characteristic fold of the plant peroxidase superfamily (Patterson and Poulos, 1995) appear to be well conserved in both halves of catalase-peroxidases, including the essential α-helix on the distal haem side. Table 2 gives an overview of the reliability of the secondary structure elements predicted for representative catalase-peroxidases by the two methods. For most helices the score is very high. The short helix B′ may be absent, as well as the already mentioned helix F^1. The N-terminal halves of helices F and H appear less well defined in the bifunctional bacterial enzymes.

Sequence alignment is less convincing between the C-terminal halves of catalase-peroxidases and the corresponding regions of the N-terminal halves of catalase-peroxidases, ascorbate peroxidases, and yeast CCP (data not shown). The most striking peculiarity is the replacement of both catalytically important histidines in the C-terminal copies, by arginines at the position corresponding to the proximal haem ligand, and predominantly by alanines at the distal position. Both halves of catalase-peroxidases show a unique single residue-insertion close to the proximal haem ligand in the region corresponding to helix F in ascorbate peroxidases and yeast CCP. This probably causes a distortion or bending of this helix, a situation reminiscent of the distorted 'essential helix' in typical catalases. In both enzymes this unusual geometry may lead to an increased electronegativity of the proximal haem ligand due to the helical dipole oriented towards the catalytic residues on the haem proximal side (Mate et al., 1999).

Table 2
Reliability of secondary structure prediction of catalase-peroxidases. Two representative examples predicted with two different methods are given. In both methods reliability scores range from 0 to 9. Length of the predicted secondary structure element is given in parentheses

Secondary structure element (as of Pissa-APX)	Myctu-CP (PHDsec)	Halma-CP (R&G[a])	Length from X-ray data[b]
Helix A	8.11 (17)	7.00 (17)	17
Helix B	0	8.00 (9)	13
Helix B′	0	0	5
Helix C	8.23 (13)	8.15 (13)	13
Helix D	6.57 (7)	6.18 (11)	15
Helix E	9.00 (6)	9.00 (6)	6
Helix F	2.00 (3)	0	8
Helix F^1	0	0	4
Helix G	6.13 (8)	8.14 (7)	8
Helix H	0	6.33 (3)	6
Helix I	7.29 (7)	8.70 (10)	12
Helix J	8.93 (16)	8.88 (16)	16

[a] Method of Rychlewski and Godzik (1997).
[b] Length of overlapping helices in crystal structures of yeast CCP and Pissa-APX.

Table 3 summarises the respective sequence similarities for the various structural elements included in Fig. 3, calculated from all known sequences of catalase-peroxidases, and from a representative selection of ascorbate peroxidases. Several interesting features can be concluded from the presented data. First, for the entire distal domain (segments 1 to 4), significantly higher homology is seen within the N-terminal halves of catalase-peroxidases than in either ascorbate peroxidases, or the C-terminal halves of the bifunctional enzymes. The section connecting the presumptive helices A and B (segment 3) is of special interest. It is extremely well conserved, seven out of 12 residues are invariant. Furthermore, in ascorbate peroxidases there is only a two-residues long connection between these helices. In yeast CCP, a ten-residues long polar surface loop is found at the respective position. Secondly, in most parts of the proximal domain (segments 6 to 8) ascorbate peroxidases reveal the highest degree of sequence homology, the C-terminal halves of catalase-peroxidases being slightly better than the N-terminal halves. With few exceptions, pairwise alignment of the different families

Table 3
Similarity of amino acid sequences within the groups of Class I peroxidases and between members of these groups. N- and C-terminal halves of catalase-peroxidases are treated separately. Similarities (%) are presented in bold. The numbers of the aligned residues are given in parentheses[a]

Segment	CPn (20 species)	CPc (18 species)	APX (19 species)	CPn versus APX	CPc versus APX	CPn versus CPc	CPn versus CCP	CPc versus CCP
1 (preceding helix A)	**35** (23)	**0** (2)	**0** (3)	**0** (1)	**0** (1)	**0** (1)	**6** (16)	**0** (2)
2 (helix A)	**42** (19)	**29** (17)	**24** (17)	**12** (17)	**12** (17)	**5** (17)	**21** (19)	**0** (17)
3 (preceding helix B)	**75** (12)	**50** (4)	**33** (3)	**0** (3)	**0** (3)	**0** (3)	**30** (10)	**0** (4)
4 (helix B to D)	**70** (77)	**54** (74)	**53** (74)	**41** (74)	**38** (68)	**30** (69)	**27** (76)	**17** (74)
5 (preceding helix E)	**42** (65)	**35** (54)	**43** (30)	**13** (30)	**28** (25)	**12** (24)	**12** (33)	**15** (54)
6 (preceding helix F^1)	**62** (29)	**57** (28)	**66** (29)	**43** (28)	**29** (28)	**20** (28)	**41** (27)	**14** (28)
7 (preceding helix H)	**31** (95)	**40** (50)	**51** (37)	**18** (34)	**16** (19)	**11** (20)	**15** (52)	**26** (49)
8 (helix H to J)	**41** (37)	**52** (42)	**58** (38)	**24** (37)	**33** (30)	**18** (29)	**18** (37)	**18** (41)
9 (following helix J)	**57** (23)	**67** (9)	**29** (7)	**0** (6)	**17** (6)	**12** (6)	**0** (22)	**11** (9)
N-term. domain	**60.3**	**48.5**	**45.4**	**33.7**	**31.5**	**24.4**	**24.0**	**13.7**
domain connection	**41.5**	**35.2**	**43.3**	**13.3**	**28.0**	**12.5**	**12.1**	**14.8**
C-term. domain	**40.8**	**49.6**	**55.9**	**25.7**	**26.5**	**24.1**	**18.8**	**19.7**
entire sequence	**47.6**	**45.4**	**50.0**	**27.4**	**28.9**	**22.3**	**20.2**	**16.5**

[a] Abbreviations. CPn, CPc: N- and C-terminal halves of catalase-peroxidases; APX: ascorbate peroxidase; CCP: yeast cytochrome c peroxidase.

of Class I against each other reveals similarities which could be expected from the respective degree of conservation within the two families under investigation. This does not apply, however, to the segment including the proximal haem ligand in APX (segment 6). Although highly conserved within all three groups, the similarity of the C-terminal halves of catalase-peroxidases with all other members of Class I is unexpectedly low in this area.

In the eukaryotic domain, so far only two catalase-peroxidases have been reported, both from fungi. Both enzymes were purified and revealed enzymatic properties similar to those of typical prokaryotic catalase-peroxidases. The molecular mass of the *Penicillium simplicissimus* catalase-peroxidase ranges among those of bacterial homologues (Fraaije et al., 1996), whereas the subunit size of the tetrameric enzyme from the wheat pathogen *Septoria tritici* (*M. graminicola*) is significantly smaller (Levy et al., 1992). Very recently, the gene coding for the latter enzyme was partially sequenced. It comprised what appears to correspond to the major part of the N-terminal half, up to close to the proximal haem ligand. Almost the entire ORF reflects convincing homology with all other known catalase-peroxidases. However, the 20 C-terminal residues of the corresponding translated protein sequence are unique among the entire peroxidase superfamily. This area corresponds to a loop connecting the presumptive helices D and E, which is significantly longer in catalase-peroxidases than in all other members of the superfamily (see above). So eukaryotic and prokaryotic catalase-peroxidases may be quite distinct in this area, but at the moment one cannot rule out possible errors in the cDNA library. Evolutionary trees constructed from multiple alignments of Mysgr-CP with the corresponding regions of Class I peroxidases (not shown), therefore, depend on whether this region is included or not. In both cases, Mysgr-CP is located well within the catalase-peroxidase branch, closest to Bacst-CI.

The conservation of the proximal binding site for K^+ ions (important for the reaction mechanism of ascorbate peroxidase) can be easily followed by HCA (not shown). Three invariant residues (Thr164, Thr180 and Asp187 in *P. sativum* ascorbate peroxidase) are present in all ascorbate peroxidases, but only one of them is conserved in yeast CCP. This cation site is very similar to the proximal Ca^{2+} sites in fungal and plant secretory peroxidases. Formation of a tryptophan (Trp191) radical rather than a porphyrin π cation radical in compound I of CCP may be due to the absence of this cation site in this protein. These residues are conserved in the N-terminal halves of eubacterial catalase-peroxidases [Thr271, Thr322, and Asp329 in *M. tuberculosis* catalase-peroxidase)] but, interestingly, not in the N-terminal half of catalase-peroxidase from *A. fulgidus*, and they are not found in the C-terminal half of any catalase-peroxidase.

3.2. Phylogenetic relationships

The most likely tree (Fig. 4) was created by PROTDIST/FITCH/CONSENSE, all included in the PHYLIP package, to establish the degree of phylogenetic and evolutionary links among the various members of the catalase-peroxidase family and their integration in Class I of the peroxidase superfamily. This tree was based on the alignment of the entire amino acid sequences of the various proteins.

Two clusters are easily discernible on the constructed tree: prokaryotic catalase-peroxidases, and eukaryotic peroxidases (ascorbate peroxidases and yeast CCP). There is no mixture of members between these two

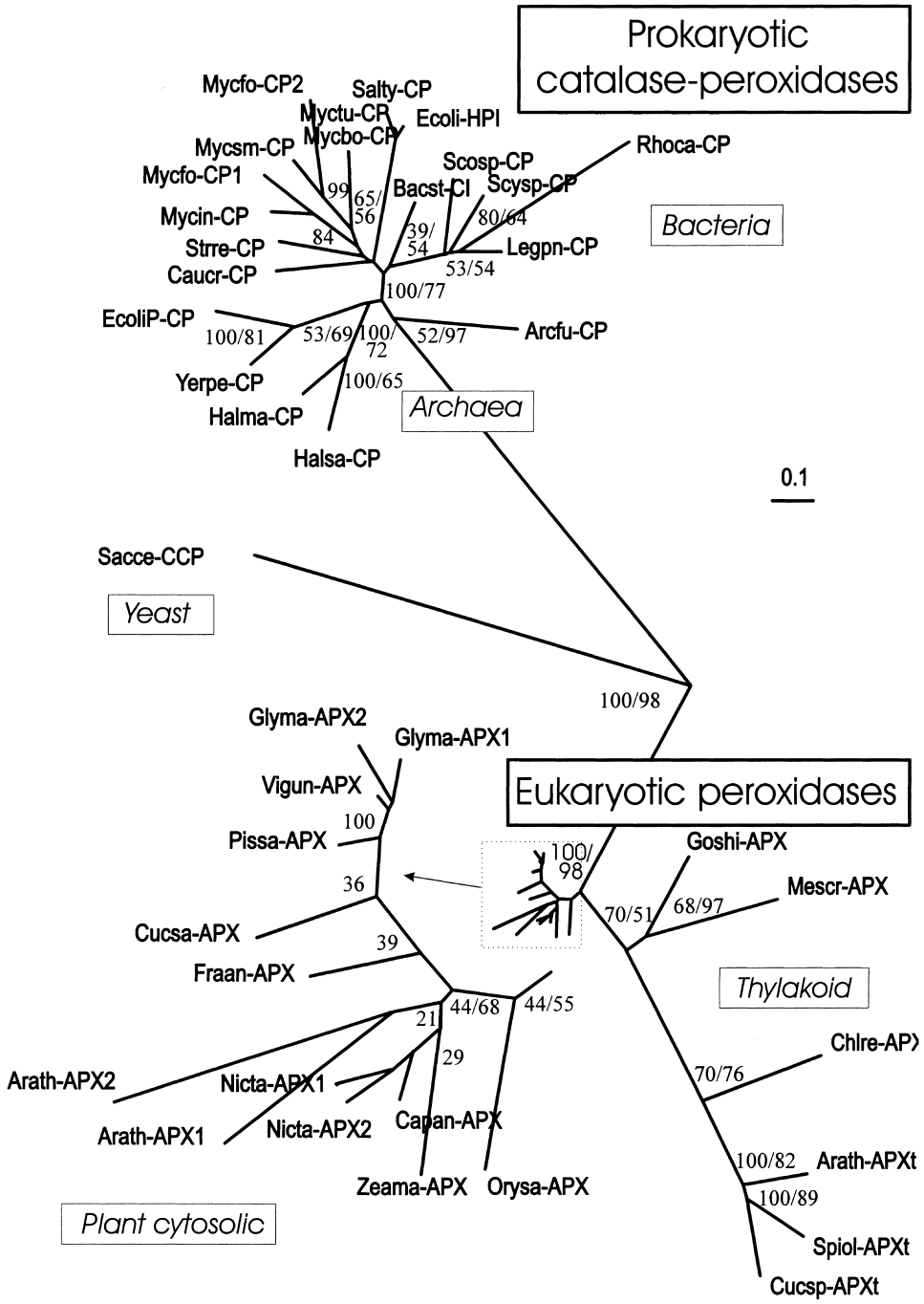

Fig. 4. Unrooted phylogenetic tree of eukaryotic ascorbate peroxidases, yeast cytochrome c peroxidase, and prokaryotic catalase-peroxidases. The abbreviations of the sources of the individual enzymes are given in Table 1. The tree is the most likely tree calculated with the programs PROTDIST, FITCH, CONSENSE of the PHYLIP program package. An essentially identical tree was obtained by the PUZZLE method (Strimmer and von Haeseler, 1996) with slightly different branch lengths. The tree is fitted to the estimated pairwise PAM distances between sequences by using the Fitch–Margoliash method. At important nodes numbers are given indicating the relative probability that the respective group occurred in all trees analysed (first numbers), together with maximum likelihood values as reported by PUZZLE. An enlarged view of the branch of cytosolic ascorbate peroxidases is shown as inset.

distinct lineages, suggesting that they segregated early in the phylogenetic history.

The eukaryotic half of the tree is organised in three branches: (i) cytosolic ascorbate peroxidases from higher plants; (ii) ascorbate peroxidases from green algae and higher plant thylakoids; and (iii) the unique CCP from bakers yeast. The arrangement of the large branch of cytosolic ascorbate peroxidases largely follows the taxonomic relationships. This does not hold for Zeama-APX, which does not segregate from the enzymes from

Eudicotyledons, whereas Orysa-APX shows some segregation. However, as indicated in Fig. 4, the respective nodes appear less reliable than other parts of this branch. The two membrane-bound, cytosolic ascorbate peroxidases from *G. hirsutum* and *M. crystallinum* appear more closely related to chloroplast-located ascorbate peroxidase than the other members of this group. Interestingly, the only known ascorbate peroxidase from a lower eukaryote (the green alga *C. reinhardtii*) goes well with the chloroplast ascorbate peroxidases, which may have survived as a relic of the former autonomy of this organelle.

The average interspecies distances are larger than those among eukaryotic peroxidases in the catalase-peroxidase part of the tree. The three archaeal enzymes form a loose cluster, which, however, is not clearly segregated from the eubacterial subbranch. Arcfu-APX appears more distantly related then the other two members of this group; in 52% of all trees it forms a branch of its own. Mycobacteria form a subbranch with comparatively small distances. Two catalase-peroxidase isoenzymes exist in *M. fortuitum*, which are apparently more distantly related to each other (70% similarity) than to katG-gene products from other mycobacteria. Based on the presence of a transposon sequence next to katGI, it is suggested that at least one of the two isoenzymes in *M. fortuitum* goes back to horizontal transmission from another bacterium (Menedez et al., 1997). This is also supported by the analysis of the evolutionary distance of C- and N-terminal copies of both katGI and katGII from this organism (see Fig. 5). It reveals that the C-terminal halves are 2.5-fold more distantly related from each other than their respective N-terminal segments. For a possible duplication of katG inside *M. fortuitum* approximately equal evolutionary distances would be expected.

The close relationship of the HPI enzymes from *E. coli* and *S. typhimurium* is in accordance with the common phylogeny of both organisms, as supported by numerous other loci. *E. coli* also expresses a plasmid-borne catalase-peroxidase. The closest relative of this plasmid-encoded catalase-peroxidase is the hydroperoxidase from *Y. pestis*, (88% similarity). It exhibits a 23-residues long signal sequence, unique among catalase-peroxidases. With significant reliability, both species appear rather closely related to archaeal catalase-peroxidases.

From a phylogenetic point of view catalase-peroxidases from phototropic bacteria probably are even more exciting. These organisms have to cope with higher levels of peroxides than other bacteria, since ROS inevitably evolve during photosynthesis. Catalytic activity was reported in all cyanobacterial species (Miyake et al., 1991), but some cyanobacteria may also possess ascorbate peroxidases (see Section 4). On the evolutionary tree the two catalase-peroxidases from phototropic bacteria and the hydroperoxidase from the purple non-sulphur photosynthetic bacterium *R. capsulatus* B10 form a branch of their own, clearly segregated form the eubacterial branch. Most surprisingly, the enzymes from *Legionella pneumoniae* and (though less closely and also less reliable) from *B. stearothermophilus* form part of this branch of phototrophic bacteria.

Evolutionary relationships within the subclass of catalase-peroxidases were studied separately. Trees were constructed for the N- and C-terminal halves (Fig. 5A and B respectively). For the N-terminal halves of catalase-peroxidases the archaebacterial representatives are slightly separate, but do not form a common lineage. Both halves of the enzyme from *A. fulgidus* are quite distinct from the two other archaeal members, but again the reliability of the position of this species is low in both trees. Both the plasmid-encoded *E. coli* enzyme and the catalase-peroxidase from *Y. pestis* are also found in this cluster. Among the other eubacterial catalase-peroxidases we can also distinguish the mycobacterial, *E. coli* HPI, and cyanobacterial sub-branches. The N-terminal half of the enzyme from *L. pneumoniae* again is found within the cyanobacterial branch, whereas *B. stearothermophilus* almost forms a branch of its own. Interestingly, the tree constructed only from the C-terminal halves of catalase-peroxidases generally reflects the same pattern, but some branches are more dispersed (mainly the mycobacterial, and to a lesser degree also the archaeal/*Y. pestis* subgroups). In this tree the *B. stearothermophilus* enzyme is linked more closely with the cyanobacterial subgroup. The C-terminal half of Strre-CP goes with the mycobacterial homologues, whereas its N-terminal half forms part of the *E. coli* HPI-subgroup. This pattern indicates some diversification of both copies of the ancestral hydroperoxidase gene after the duplication event.

When ascorbate peroxidases and yeast CCP are aligned with the N- and C-terminal halves of catalase-peroxidases separately (not shown), these four groups build three clearly separated clusters. Although the distances between these clusters are roughly of the same order, the N-terminal halves of catalase-peroxidases still appear closer to both groups of eukaryotic peroxidases than to the C-terminal halves of catalase-peroxidases. It is therefore supposed that the gene duplication eventually giving rise to catalase-peroxidases must have occurred at a stage still close to a common peroxidase ancestor.

4. Discussion

4.1. Gene duplication in the Class I peroxidase superfamily

The sequence analysis performed herein indicates that catalase-peroxidases evolved from a common ancestor

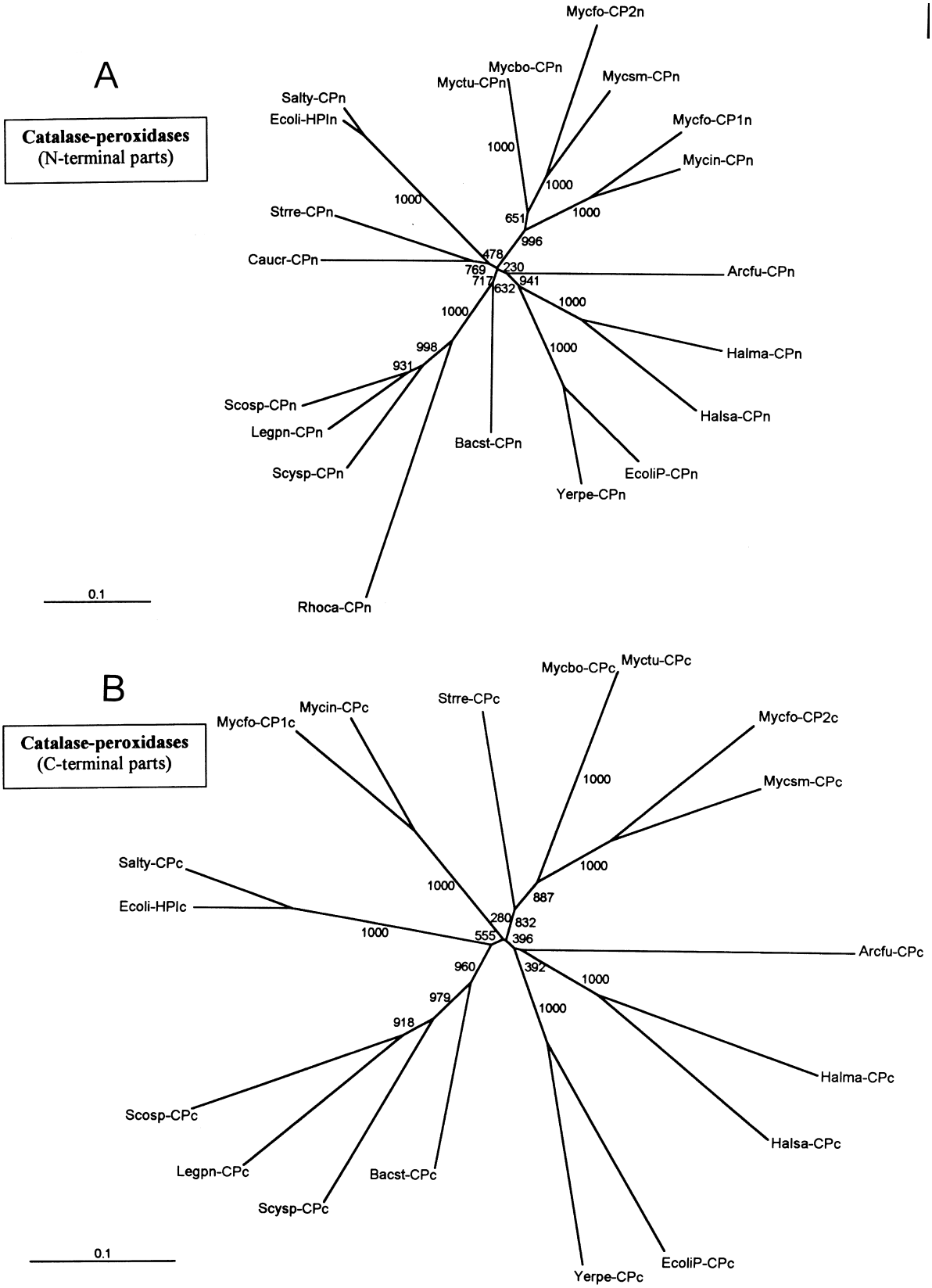

Fig. 5. Unrooted phylogenetic trees of catalase-peroxidases. The trees were based on the alignment of sequences of the N-terminal (tree A) and C-terminal (tree B) parts of catalase-peroxidases. For the N- and C-terminal halves of catalase-peroxidases see Fig. 1 (details in the Section 2). The abbreviations of the sources of the individual enzymes are given in Table 1. The branch lengths are proportional to the divergency of the individual amino acid sequences, the sum of the lengths of the branches linking any sources being a measure of the evolutionary distance between them. Bootstrap values are given for important branches and nodes.

which they share not only with yeast CCP and ascorbate peroxidases, but also with Class II and III. This species obviously was built from clearly separated distal and proximal domains, revealing a non-bundled α-helical fold consisting of ten helices with bound ferric protoporphyrin IX. As hypothesised by Welinder (1992), this ancestor itself could be the product of an earlier gene duplication event, giving rise to the above-mentioned domains. Today, catalase-peroxidases form the only subgroup that exists as duplicates of the ancestral peroxidase gene.

The discussion of the evolution of Class I of the plant peroxidase superfamily, which essentially concerns the question when gene duplication occurred, and when fusion of the two copies occurred, has to take into account the following observations.

(1) Catalase-peroxidases presumably exist in all three major domains of living organisms. However, only preliminary data are available about eukaryotic catalase-peroxidases. The only known partial sequence (from *M. graminicola*) shows marked homology to eubacterial catalase-peroxidases, supporting the assumption that lateral gene transfer rather than an independent evolution within this eukaryotic branch lead to this gene.

BLAST searches were performed with several highly conserved regions of the C-terminal halves of various catalase-peroxidases showing no obvious sequence relationship to the corresponding N-terminal regions. These searches should unveil whether this second copy of the ancestral hydroperoxidase gene may have also survived in higher organisms, though independent of the peroxidase copy. Several of these regions showed distinct homology to non-peroxidase sequence patterns, mostly belonging to eukaryotic and prokaryotic transposons (data not shown). However, no convincing similarity was found when the entire C-terminal halves of catalase-peroxidases were applied, so most likely an independent second copy of the ancestral peroxidase gene does not exist in higher eukaryotes, or only survived as a pseudogene with insignificant residual sequence homology due to unconstrained mutation rate.

(2) 'Half-size'-peroxidases are not known from Archaea; ascorbate peroxidases are common in algae and higher plants, and may also occur in Euglenozoa (two short sequence stretches are known for the enzyme from *Euglena gracilis*); in prokaryotes ascorbate peroxidase activity was reported several times (Mittler and Tel-Or, 1991; Miyake et al., 1991), but some of these findings could not be confirmed by others (Obinger et al., 1997), and none of the putative enzymes were purified. Furthermore, no gene coding for the reported ascorbate peroxidase was found in the now completely sequenced genome of *Synechocystis* PCC6803. Taken together, the existence of ascorbate peroxidases in the prokaryotic lineage appears at least questionable.

(3) So far, single-copy- and double-copy-enzymes have never been reported to coexist in any organism.

(4) Evolutionary distances are slightly larger among C-terminal halves of catalase-peroxidases than among N-terminal halves. Nevertheless, the pattern of sequence similarity between ascorbate peroxidases and either the N-terminal halves or the C-terminal halves is remarkably similar (cf. Table 3). Obviously, all structurally essential areas were quite well conserved in both copies of the ancestral gene. Furthermore, the degree of overall similarity of residues is nearly identical in both halves. Taken together, natural selection must have largely protected the second copy of the ancestral peroxidase gene, which later developed into the C-terminal halves, from random mutagenesis.

(5) The evolutionary distance between N- and C-halves appears larger than between catalase-peroxidases and ascorbate-peroxidases. In some areas the similarity between C-and N-halves is of about the same order as that between the C-half and Class II or III peroxidases (data not shown).

(6) With the exception of Strre-CP, the trees of the N-terminal and the C-terminal halves of catalase-peroxidases show almost identical patterns of segregation. Again, an independent path of evolution of both halves, with pronounced heterogeneity of rate, or even random mutagenesis in either half, most likely can be ruled out.

Taken together, the most acceptable explanation for the evolution of catalase-peroxidases is a single gene duplication event in a common ancestor line. The alternative explanation stating that duplications occurred independently in each major branch, but still lead to very homologous products, appears very unlikely per se. Furthermore, it would then be difficult to explain why no single-copy protein survived in any archaeon or prokaryote. We also assume that a similar event did not occur or was not fixed in the eukaryotic domain.

Duplication of genes is the predominant and most important mechanism by which new genes can arise, giving rise to novel functional and structural properties of proteins. Since the most likely fate for deleterious or even neutral duplications is inactivation, the duplication event eventually leading to the formation of catalase-peroxidases must have been advantageous for the respective ancestor species.

Small gene families coding for Class III peroxidases are frequently found in higher plants. The members of these families generally are closely related, so obviously they arose by tandem gene duplications relatively recently. This mechanism clearly is ruled out in the case of catalase-peroxidase evolution. Both copies are significantly different to fulfil at least partially different roles, but on the other hand there is pronounced similarity between corresponding parts among all known species. All currently available data support the hypothesis that one single duplication event gave rise to the

evolution of extant catalase-peroxidases, and this gene duplication accordingly must have preceded speciation of all microorganisms involved. According to the currently accepted universal tree of life (bacterial rooting) this duplication then must have already occurred at the level of the last universal cellular ancestor (LUCA). However, this universal tree has recently been questioned, and alternative hypotheses were proposed. The tree proposed by Forterre and Philippe (1999) (root in the eukaryotic branch) allows a more plausible scenario of catalase-peroxidase evolution. The crucial gene duplication step would have taken place after the transition from complex to simple organisms, i.e. in an organism that evolved by simplification from a eukaryotic-like LUCA.

Whether direct tandem duplication occurred, or whether both copies remained separate for some time, and were fused after some further diversification, appears less obvious. Almost certainly, recent C-terminal halves of catalase-peroxidases are not functional as (haem) peroxidase. On the other hand, random exchanges do not appear to have occurred at any of the formerly essential positions, and more than half of them are even conserved in the C-terminal copies. So the corresponding changes must have either occurred (almost) simultaneously with the duplication, or some functional constraint must have virtually prevented further divergence.

The region linking both halves (segment 9 in Table 3, the area between the presumptive helix J of the N-terminal half and helix A of the C-terminal half) shows above-average degree of conservation (57% among all species) and only little variation with respect to its length. Separate evolution of both duplicons should have lead to a much higher divergence, especially at the ends of both genes and thus also in the region linking both halves. We therefore suggest that tandem duplication positioned both copies of the ancestor gene as close neighbours as they are found in recent catalase-peroxidases.

It cannot be ruled out that both copies of the ancestor gene originally retained their catalytic function, thus providing a selective advantage for the duplicated gene to be fixed in the population. Concerted evolution then could have lead to a modified gene product with two functionally non-identical halves. Again, the pattern of similarities found in both halves of catalase-peroxidase supports the view that the second gene copy had already lost its original function concomitant with tandem duplication. This single duplication/modification event could have been fixed in the population due to, for example, higher stability or increased resistance to oxidative damage of the corresponding gene product.

All other schemes would require convergent evolution in a rather large range of diversified organisms, and it would be more difficult to explain why Class I peroxidases are so unevenly distributed among the domains of life.

4.2. Predictions of structural and functional properties of catalase-peroxidases

In Class I of the peroxidase multigene family the two copies appear to have acquired non-equal functions from still homologous structures. There is compelling evidence that the original catalytic function is lost in the C-terminal copies of catalase-peroxidases: without exception the essential histidines are replaced by non-functional residues, and at least one representative with a truncated C-terminal copy exists still revealing typical catalase-peroxidase properties (*R. capsulatus*). A deletion of comparable length (150 residues) was never observed in the equivalent region of any N-terminal copy, clearly suggesting that the two halves are not functionally equivalent. The function of the second half of the molecule remains open. The very crude small-angle X-ray diffraction data for the katG-encoded enzyme from *M. tuberculosis* (Nagy, et al., 1997) do not allow any conclusions about the relative orientation of the two halves of each subunit. The C-terminal portion appears essential for proper functioning of the 'catalytic' half, as concluded from pronounced homology among these halves of the protein, as well as from the effects of random mutagenesis in mycobacterial catalase-peroxidases (Billman-Jacobe et al., 1996). The representative from *R. capsulatus* does not necessarily contradict this rule, since it also harbours major alterations in its N-terminal half. Owing to the long loops between helices D/E and F/G in catalase-peroxidases, the docking of the proximal domain to the distal domain may be less straightforward than in other Class I peroxidases, and thus may require further stabilisation by the C-terminal half of the polypeptide. Interestingly, Class II and III peroxidases, which also show some important insertions within these loops, clearly are stabilised by four disulphide bridges.

One of the major challenges in investigations on catalase-peroxidases is to explain the bifunctional capability of these enzymes in molecular terms. Clearly, high-resolution crystal data are required to answer this question, but some conclusions can already be drawn from the available sequences, and their comparison with other peroxidases of known structural organisation. Obviously only one haem-containing active site exists in catalase-peroxidases, and it can be expected to be almost identical with that found in other members of Class I. Nearly all residues directly involved in the structural organisation of the active site in ascorbate peroxidases and yeast CCP are conserved in the N-terminal halves of the bifunctional enzymes. However, the hydrogen bonding network, which facilitates the function of the essential distal histidine as H^+-acceptor in eukaryotic peroxidases

(His52→Asn82→Glu76 in CCP) most likely is less extensive in catalase-peroxidases (non-polar residues at the position corresponding to Glu76). Consequently, the pK_a of the essential histidine may be substantially higher than in other members of this class, allowing a more flexible role of this residue comparable to the proposed function of the essential histidine in typical catalases (Mate et al., 1999). Additionally, the proposed partial reorientation of the helix incorporating the proximal haem ligand (see above) may increase its potential to 'push' electrons.

Almost certainly the opening of the substrate channel near the haem edge will be organised in a different way in catalase-peroxidases than with CCP or ascorbate peroxidases. The sequence of the short loop between helices B′ and C clearly differs from that in the other sub-families, and long insertions are found in catalase-peroxidases immediately adjacent to the other parts of this opening. One may speculate that the switching between the two modes of compound I reduction in catalase-peroxidases could be assisted or controlled by partial reorganisations within these long loops, in a similar way as leading to the opening of the hidden channel in yeast CCP (Fitzgerald et al., 1996). Major parts of the two insertions characteristic for all catalase-peroxidases — the extremely well-conserved connection between helices A and B, as well as the extra long loop between helices D and E – are hydrophobic and thus, unlike their shorter equivalents in all peroxidases of known structure, should not be surface exposed. Therefore, both loops are candidates for inter-domain or inter-subunit contacts. They could, however, also form a tightly packed, largely hydrophobic structure limiting access to the haem. The prosthetic group thus might be buried in a similar manner as in typical catalases, favouring the reduction of compound I by a second molecule of H_2O_2 rather than by bulky, typical peroxidase substrates.

Acknowledgements

This work was supported, in part by the grant P09968MOB from the Fonds zur Förderung der wissenschaftlichen Forschung (Austrian Foundation) to FK and MZ, grant 15 SR 13 from the Action Austria–Slovakia to all authors, and grant no. 2/6045/20 from the Slovak Grant Agency for Science VEGA to Š.J.

References

Altschul, S.F., Stephen, F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Arnao, M.B., Acosta, M., Rio, J.A., Garcia-Canovas, F., 1990. A kinetic study on the suicide inactivation of peroxidase by hydrogen peroxide. Biochim. Biophys. Acta 1038, 85–89.

Billman-Jacobe, H., Sloan, J., Coppel, R.L., 1996. Analysis of isoniazid-resistant transposon mutants of *Mycobacterium smegmatis*. FEMS Microbiol. Lett. 144, 47–52.

Callebaut, I., Labesse, G., Durand, P., Poupon, A., Canard, L., Chomilier, J., Henrissat, B., Mornon, J.P., 1997. Deciphering protein sequence information through hydrophobic cluster analysis (HCA): current status and perspectives. Cell. Mol. Life Sci. 53, 621–645.

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

Finzel, B.C., Poulos, T.L., Kraut, J., 1984. Crystal structure of yeast cytochrome c peroxidase refined at 1.7-A resolution. J. Biol. Chem. 259, 13 027–13 036.

Fitzgerald, M.M., Musah, R.A., McRee, D.E., Goodin, D.B., 1996. A ligand-gated, hinged loop rearrangement opens a channel to a buried artificial protein cavity. Nat. Struct. Biol. 3, 626–631.

Forterre, P., Philippe, H., 1999. Where is the root of the universal tree of life? BioEssays 21, 871–879.

Fraaije, M.W., Roubroeks, H.P., Hagen, W.R., Van-Berkel, W.J., 1996. Purification and characterization of an intracellular catalase-peroxidase from *Penicillium simplicissimum*. Eur. J. Biochem. 235, 192–198.

Gajhede, M., Schuller, D.J., Henriksen, A., Smith, A.T., Poulos, T.L., 1997. Crystal structure of horseradish peroxidase C at 2.15 Å resolution. Nat. Struct. Biol. 12, 1032–1038.

Jeanmougin, F., Thompson, J.D., Gouy, M., Higgins, M., Gibson, T.J., 1998. Multiple sequence alignment with ClustalX. Trends Biochem. Sci. 23, 403–405.

Jespersen, H.M., Kjærsgåard, I.V.H., Østergaard, L., Welinder, K.G., 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.

Klotz, M.G., Klassen, G.R., Loewen, P.C., 1997. Phylogenetic relationships among prokaryotic and eukaryotic catalases. Mol. Biol. Evol. 14, 951–958.

Levy, E., Eyal, Z., Hochman, A., 1992. Purification and characterization of a catalase-peroxidase from the fungus *Septoria tritici*. Arch. Biochem. Biophys. 296, 321–327.

Mate, M.J., Zamocky, M., Nykyri, L.M., Herzog, C., Alzari, P.M., Betzel, C., Koller, F., Fita, I., 1999. Structure of catalase-A from *Saccharomyces cerevisiae*. J. Mol. Biol. 286, 135–149.

Menedez, M.C., Ainsa, J.A., Martin, C., Garcia, M.J., 1997. katGI and katGII encode two different catalase-peroxidases in *Mycobacterium fortuitum*. J. Bacteriol. 179, 6880–6886.

Mittler, R., Tel-Or, E., 1991. Oxidative stress responses in the unicellular cyanobacterium *Synechococcus* PCC 7942. Free-Radic. Res. Commun. 12 (13), 845–850.

Miyake, C., Michihata, F., Asada, K., 1991. Scavenging of hydrogen peroxide in prokaryotic and eukaryotic algae: acqusition of ascorbate peroxidase during the evolution of cyanobacteria. Plant Cell Physiol. 32, 33–43.

Nagy, J.M., Svergun, D., Koch, M.H.J., Cass, A.E.G., Brown, K.A., 1997. Structural characterization of recombinant catalase-peroxidase from *Mycobacterium tuberculosis*. Biochem. Soc. Trans. 25, S617.

Obinger, C., Regelsberger, G., Strasser, G., Burner, U., Peschek, G.A., 1997. Purification and characterization of a homodimeric catalase-peroxidase from the cyanobacterium *Anacystis nidulans*. Biochem. Biophys. Res. Commun. 235, 545–552.

Page, R.D.M., 1996. TreeView: an application to display phylogenetic trees on personal computers. CABIOS 12, 357–358.

Patterson, W.R., Poulos, T.L., 1995. Crystal structure of recombinant pea cytosolic ascorbate peroxidase. Biochemistry 34, 4331–4341.

Rost, B., Sander, C., 1993. Prediction of protein structures of better than 70% accuracy. J. Mol. Biol. 232, 584–599.

Rychlewski, L., Godzik, A., 1997. Secondary structure prediction using segment similarity. Protein Eng. 10, 1143–1153.

Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.

Strimmer, K., von Haeseler, A., 1996. Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. Mol. Biol. Evol. 13, 964–969.

Welinder, K.G., 1991. Bacterial catalase-peroxidases are gene duplicated members of the plant peroxidase superfamily. Biochim. Biophys. Acta 1080, 215–220.

Welinder, K.G., 1992. Superfamily of plant, fungal and bacterial peroxidases. Curr. Opin. Struct. Biol. 2, 388–393.

Zámocký, M., Janeček, Š., Koller, F., 1997. The area of the main substrate channel is highly conserved among all true catalases. Biologia (Bratislava) 52, 723–730.