PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

The peroxidase–cyclooxygenase superfamily: reconstructed evolution of critical enzymes of the innate immune system

Marcel Zamocky,[1,2*] Christa Jakopitsch,[1] Paul G. Furtmüller,[1] Christophe Dunand,[3] and Christian Obinger[1]

[1] Department of Chemistry, Division of Biochemistry, BOKU-University of Natural Resources and Applied Life Sciences, A-1190 Vienna, Austria
[2] Institute of Molecular Biology, Slovak Academy of Sciences, SK-84551 Bratislava, Slovakia
[3] Laboratoire de Physiologie Végétale, Université de Genève, Quai Ernest-Ansermet 30, CH-1211 Geneva 4, Switzerland

ABSTRACT

The authors have reconstructed the phylogenetic relationships of the main evolutionary lines of mammalian heme containing peroxidases. The sequences of intensively investigated human myeloperoxidase, eosinophil peroxidase, and lactoperoxidase, which participate in host defence against infections, were aligned together with newly found open reading frames coding for highly similar putative peroxidase domains in all kingdoms of life. The evolutionary relationships were reconstructed using neighbor-joining, maximum parsimony, and maximum likelihood methods. It is demonstrated that this enzyme superfamily obeys the rules of birth-and-death model of multigene family evolution and contains proteins with a variety of function that could be grouped in seven subfamilies. On the basis of occurrence and the fact that two main enzymatic activities are related with these metalloproteins, they propose the name peroxidase–cyclooxygenase superfamily for this widely spread group of heme-containing oxidoreductases. Well known structure–function relationships in mammalian peroxidases formed the basis for the critical inspection of all subfamilies. The presented data unequivocally suggest that predecessor genes of mammalian heme peroxidases have segregated very early in evolution. Before organisms developed an acquired immunity, their antimicrobial defence depended on enzymes that were recruited upon pathogen invasion and could produce antimicrobial reaction products. Thus, these peroxidatic heme proteins evolved to important components in the innate immune defence system. This work shows that even in certain prokaryotic organisms, genes encoding putative antimicrobial enzymes are found providing a group of bacteria with an evolutionary advantage over the others.

Proteins 2008; 72:589–605.

Key words: myeloperoxidase; lactoperoxidase; peroxidasin; peroxinectin; peroxicin; peroxidockerin; cyclooxygenase; prostaglandin synthase; dual oxidase; lateral gene transfer.

INTRODUCTION

Peroxidases are ubiquitous oxidoreductases capable of cleaving the peroxidic bond heterolytically. Most frequently, peroxidases reduce hydrogen peroxide to water with concomitant one- and/or two-electron oxidation of a great variety of both organic (AH_2) and inorganic (e.g., halides, X^-) substrates according to Reactions 1 and 2. Among others, reaction products can be organic radicals ($^\bullet AH$) or hypohalous acids (HOX).

$$H_2O_2 + 2\ AH_2 \rightarrow 2\ H_2O + 2\ {}^\bullet AH \qquad \text{Reaction 1}$$

$$H_2O_2 + X^- + H^+ \rightarrow H_2O + HOX \qquad \text{Reaction 2}$$

Although non-heme peroxidases, for example, thiol peroxidases represented by thioredoxins, peroxiredoxins, and glutathione peroxidases[1] are widespread in both prokaryotes and eukaryotes, the most abundant peroxidases in all living forms are probably those containing the prosthetic heme group.[2] The majority of currently known heme peroxidases are members of two known superfamilies that are ubiquitous in all kingdoms of life. Both superfamilies arose independently, thus their primary and tertiary structures and even the nature of the prosthetic group differ greatly. The superfamily of (archae) bacterial, fungal, and plant heme

The Supplementary Material referred to in this article can be found online at http://www.interscience.wiley.com/jpages/0887-3585/suppmat/

Abbreviations: CyOx, cyclooxygenase; DiOx, α-dioxygenase; DuOx, dual oxidase; EPO, eosinophil peroxidase; LGT, lateral gene transfer; LPO, lactoperoxidase; MPO, myeloperoxidase; ORF, open reading frame; PGHS, prostaglandin synthase; Pxc, peroxicin; Pxd, peroxidasin; PxDo, peroxidockerin; Pxt, peroxinectin; TPO, thyroid peroxidase.

Grant sponsor: Austrian Science Fund (FWF); Grant numbers: P19793, P14187.
*Correspondence to: Dr. Marcel Zamocky, Department of Chemistry, Division of Biochemistry, BOKU-University of Natural Resources and Applied Life Sciences, Muthgasse 18, A-1190 Vienna, Austria. E-mail: marcel.zamocky@boku.ac.at
Received 1 August 2007; Revised 29 November 2007; Accepted 5 December 2007
Published online 4 February 2008 in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/prot.21950

peroxidases (sometimes called "non-animal" superfamily) is represented by catalase-peroxidases, ascorbate peroxidases, cytochrome *c* peroxidases, manganese and lignin peroxidases, and plant secretoric peroxidases.[3,4] The other superfamily was originally called "mammalian peroxidase family"[5] since its members myeloperoxidase (MPO), eosinophil peroxidase (EPO), lactoperoxidase (LPO), and thyroid peroxidase (TPO) were originally identified in mammals. However, later on functional genes or predicted open reading frame (ORF) products showing significant similarity in amino acid sequence to mammalian peroxidases have also been identified in various invertebrates leading to the denomination "animal peroxidase family."[6] Finally, even plant, fungal, and bacterial representatives were found and so other metalloenzymes (e.g., peroxidasin, peroxinectin, and cyclooxygenase) were admitted to this multigene family.[7] A new synonym was suggested ("MPO family") that is misleading in the light of recent structural and functional findings (see below). Here, we focus on those members of this multigene family that—according to their sequence patterns—are expected to possess a peroxidase domain that could allow catalysis of Reactions 1 and 2.

The most obvious structural difference between the two heme peroxidase superfamilies is the nature of the prosthetic group. In contrast to the members of the (archae) bacterial, fungal, and plant heme peroxidase superfamily that have noncovalently bound heme *b* at the active site, the heme group of peroxidases from here described multigene family is post-translational modified and covalently linked with the protein via ester bond formation with highly conserved aspartate and glutamate residues. This has been proved by both X-ray crystallography and mass spectrometry for the mammalian enzymes MPO, EPO, LPO, and TPO.[8] MPO is singular in having additionally a sulfonium ion linkage between the 2-vinyl group and a conserved methionine. As a consequence, mammalian peroxidases contain an asymmetrical distorted heme with peculiar redox properties.[9,10] Together with conserved proximal and distal residues that allow the binding and oxidation of small anionic substrates like halides (chloride, bromide, iodide) and thiocyanate (Reaction 2), these structural features enable the mammalian peroxidases to fulfill their physiological role, that is, to produce hypohalous acids and hypothiocyanate that participate in unspecific immune defence (MPO, EPO, LPO) or hormone biosynthesis (TPO).

Regarding the other members of this growing superfamily, a critical systematic investigation that considers both phylogenetic and known structure–function relationships is missing. Peroxinectins (Pxt) have been described as peroxidatic enzymes with an integrin-binding motif that mediates cell adhesion and migration,[11,12] whereas peroxidasins (Pxd) have been demonstrated to be peroxidase-like enzymes, where the peroxidase domain is fused with motifs that occur within immunoglobulin loops and leucine-rich repeats.[13,14] However, valid data on structural features of peroxinectins and peroxidasins are insufficient and have never been analyzed against the background of knowledge about mammalian peroxidases. This will be done in the present comprehensive phylogenetic reconstruction that furthermore analysis the related group of cyclooxygenases[15] as well as the peroxidase domains in numerous other phylogenetically related multidomain and multifunctional enzymes—most interestingly among them the "dual oxidases."[16]

Here, we address the molecular evolution and surprising diversity of all the above-mentioned closely related peroxidase subfamilies. The number of gene and protein sequences has multiplied since the last report in the literature[7] aside from the appearance of new enzymatic subgroups. We demonstrate that the gene diversity in this superfamily evolved not only by conversion but also by various fusion events. Critical and comparative sequence analyses allow predictions about the functionality of the peroxidase domains, which will contribute to understand the physiological role(s) of these metalloproteins. Evidence is presented that the combination of adhesion to cells and/or pathogens with the enzymatic property to generate strong oxidants (Reaction 2) is a general strategy in unspecific innate immune defence that developed very early in evolution.

MATERIALS AND METHODS

Sequences of animal peroxidases

All the sequences analyzed here, of heme peroxidases and their phylogenetically related counterparts in all kingdoms, were collected from public databases (UniProt and NCBI—mainly the genomes databases). Blast searches were applied in order to collect a representative set of peroxidases from all possible (sub)groups of this superfamily. Selected representatives were grouped and abbreviated as MPO, EPO, LPO and salivary peroxidase (LPO), thyroid peroxidase (TPO), non-mammalian vertebrate peroxidases (POX), peroxidasin (Pxd), peroxinectin (Pxt), cyclooxygenase (CyOx) or prostaglandin H synthase (PGHS), peroxicins (Pxc), peroxidockerins (PxDo), α-dioxygenase (DiOX), and dual oxidase (DuOx). All sequences belonging to this superfamily are listed and annotated in PeroxiBase [17, http://peroxibase.isb-sib.ch/]. In this work, 134 selected members were used for sequence and phylogenetic analysis (Table I).

ORF data mining

ORFs in eukaryotic genomic sequences coding for putative heme containing peroxidases were searched with the program suite FGENESH (http://www.softberry.ru/berry.phtml) with the HMM-based model for the prediction of intron splicing. The closest organism profile available was selected in each case and the output was verified in the alignment with known peroxidases.

Table I
Abbreviations of 134 Sequences Belonging to the Heme Peroxidase–Cyclooxygenase Superfamily Used in This Study with Corresponding Accession Numbers in UniProt and NCBI

Abbreviation	Type of peroxidase	Accession number	Organism	Remark
AaeDuOx01	Dual oxidase	Q171Q3	*Aegidius egypti* (yellowfewer mosquito)	●▼
AaePxt01	Peroxinectin	Q17CY5	*Aegidius egypti*	
AaePxt02	Peroxinectin	P82600	*Aegidius egypti*	
AalPxt	Peroxinectin (``salivary peroxidase'')	AF118391	*Anopheles albimanus*	
AgaPxd	Peroxidasin	Q7QJ29	*Anopheles gambiae*	▼
AgaPxt01	Peroxinectin	Q7PRL0	*Anopheles gambiae*	
AgaPxt02	Peroxinectin	Q7Q246	*Anopheles gambiae*	
AgaPxt03	Peroxinectin	Q7QH73	*Anopheles gambiae*	
AgaPxt04	Peroxinectin	Q5TUJ3	*Anopheles gambiae*	
AgaPxt05	Peroxinectin	Q7QH75	*Anopheles gambiae*	
AmelPxd01	Peroxidasin	XP_396476.2	*Apis melifera*	▼
AmelPxt01	Peroxinectin	XP_392481.3	*Apis melifera*	●
AmelPxt02	Peroxinectin	XP_623940.2	*Apis melifera*	
ArspPxt01	Peroxinectin	A0JUB7	*Arthrobacter sp.* FB24	
AtDiOx01	α-dioxygenase (``feebly like'' protein)	Q9SGH6	*Arabidopsis thaliana*	
AteCyOx1	PGHS-like protein	AAJN01000099	*Aspergillus tereus*	
AuspPxc01	Peroxicin with RTX toxin	Q1YMS2	*Aurantimonas sp.*	
BbePOX (BbeTPO)	Invertebrate peroxidase	Q9UAF8	*Branchiostoma belcheri*	
BmaPxDo01	Peroxidockerin	NZ_AANZ01000033	*Blastopirellula marina* (Planctomycetes)	
BtLPO	Lactoperoxidase	P80025	*Bos taurus*	
BtMPO	Myeloperoxidase	XP_588495.2	*Bos taurus*	
BtPxd	Peroxidasin	XP_593953	*Bos taurus*	▼
BtTPO	Thyroid peroxidase	XP_603356	*Bos taurus*	
CelDuOx01	Dual oxidase	O61213	*Caenorhabditis elegans*	▼
CelDuOX02	Dual oxidase	NP_490684	*Caenorhabditis elegans*	●▼
CelPxd01	Putative peroxidasin	AF022977	*Caenorhabditis elegans*	▼
CelPxd02	Peroxidasin	NM_077433	*Caenorhabditis elegans*	▼
CelPxd04	Peroxidasin	P91060	*Caenorhabditis elegans*	▼
CelPxt01	Peroxinectin	Q18647	*Caenorhabditis elegans*	
CelPxt03	Peroxinectin	O17241	*Caenorhabditis elegans*	
CelPxt04	Peroxinectin	O01892	*Caenorhabditis elegans*	
CelPxt05	Peroxinectin	P90820	*Caenorhabditis elegans*	
CfaEPO	Eosinophil peroxidase	XP_548229	*Canis familiaris* (dog)	●
CfaLPO	Lactoperoxidase	XP_548321.2	*Canis familiaris*	
CfaMPO	Myeloperoxidase	XP_852445.1	*Canis familiaris*	
CfaPxd	Peroxidasin	XP_544073	*Canis familiaris*	▼
CfaTPO	Thyroid peroxidase	Q8HYB7	*Canis familiaris*	
CgCyOx01	Hypothetical protein	Q2H8J5	*Chaetomium globosum*	●
CintPOX (CintTPO)	Related to thyroid peroxidase	Q9XXZ7	*Ciona intestinalis* (sea squirt)	
CwaPxDo01	Peroxidockerin	Q4C198	*Crocosphaera watsonii*	
DdPxDo01	Peroxidockerin	Q6TMK4	*Dictyostelium discoideum* (slime mold)	
DmPxd	Peroxidasin	Q23991	*Drosophila melanogaster*	▼
DmPxt01 (DmchorPOX)	Chorion peroxidase (peroxinectin)	Q01603	*Drosophila melanogaster*	
DmPxt01-b	Peroxinectin	Q9VEG6-2	*Drosophila melanogaster*	
DrPGHS01	Prostaglandin synthase	Q8JH44	*Danio rerio* (zebrafish)	
DrDuOx	Dual oxidase 1	Q08JS2	*Danio rerio*	▼
DrPOX	Similar to myelopoerixdase	Q90XS6	*Danio rerio*	
EscPxt01	Peroxidase-like protein	Q24926	*Euprimna scolopes* (squid)	
FchPxt01	Peroxinectin	Q0H836	*Fenneropenaeus chinensis* (shrimp)	
FhePGHS01	Cyclooxygenase 2	AY532639.2	*Fundulus heteroclitus* (killifish)	
FpePxc03	Peroxicin with RTX toxin	Q0G341	*Fulvimarina pelagi*	▼
FspCyOx01	Hypothetical protein	Q3WIM4	*Frankia* sp.	
GfrCyOX02	Cyclooxygenase B	Q6S375	*Gersemia fruticosa* (marine invertebrate)	
GgaPGHS02	Prostaglandin synthase 2	P27607	*Gallus gallus* (chicken)	
GgaPOX01	Hypothetical protein	XP_415715.1	*Gallus gallus*	
GgaPOX02 (GgaMPO)	Hypothetical fusion protein	XP_415716.1	*Gallus gallus*	▼
GgaPxd01	Peroxidasin	XP_419931	*Gallus gallus*	▼
GmoCyOx01	Hypothetical protein	AAIM02000117.1	*Gibberella moniliformis*	
GviPxc01	Peroxicin	Q7NFC7	*Gloeobacter violaceus*	
GzCyOx01	Hypothetical protein	XP_388327	*Gibberella zeae*	
HpuPxt01	Peroxinectin–ovoperoxidase	O02634	*Hemicentrotus pulcherrimus* (sea urchin)	
HrPOX (HrTPO)	Related to thyroid peroxidase	Q9XXZ8	*Halocynthia roretzi* (sea squirt)	

(*Continued*)

Table I
(Continued)

Abbreviation	Type of peroxidase	Accession number	Organism	Remark
HsDuOx01	Dual oxidase	Q9NRD9	*Homo sapiens*	▼
HsDuOx02	Dual oxidase	Q9NRD8	*Homo sapiens*	▼
HsEPO	Eosinophil peroxidase	P11678	*Homo sapiens*	
HsLPO	Lactoperoxidase	Q3KNQ2	*Homo sapiens*	
HsMPO	Myeloperoxidase	P05164	*Homo sapiens*	
HsPGHS01	Prostaglandin H synthase 1	P23219	*Homo sapiens*	
HsPGHS02	Prostaglandin H synthase 2	P35354	*Homo sapiens*	
HsPxd	Peroxidasin	Q92626	*Homo sapiens*	▼
HsTPO01	Thyroid peroxidase	P07202	*Homo sapiens*	
LspPxc01	Peroxicin	ZP_01620446	*Lyngbya* sp.	
LvaDuOx01	Dual oxidase	Q5XMJ0	*Lytechinus variegates*	●▼
LvaPxt01	Peroxinectin–ovoperoxidase	O44392	*Lytechinus variegatus* (green sea urchin)	
MauLPO	Lactoperoxidase	Q8R481	*Mesocricetus auratus* (golden hamster)	
MmEPO	Eosinophil peroxidase	P49290	*Mus musculus* (mouse)	
MmLPO	Lactoperoxidase	Q5SW46	*Mus musculus*	
MmMPO	Myeloperoxidase	Q5NCP1	*Mus musculus*	
MmPxd	Peroxidasin	Mm.251774	*Mus musculus*	▼
MmTPO	Thyroid peroxidase	P35419	*Mus musculus*	
MmulEPO	Eosinophil peroxidase	XP_001106246	*Macaca mulatta* (rhesus monkey)	
MmulMPO	Myeloperoxidase	XP_001103896	*Macaca mulatta*	
MmulTPO	Thyroid peroxidase	XP_001117795	*Macaca mulatta*	
MsmCyOx01	Hypothetical protein	YP_890542	*Mycobacterium smegmatis*	
MspPxc01	Peroxicin with calcium binding motif	Q11K84	*Mesorhizobium* sp.	▼
MvaCyOx01	Putative cyclooxygenase	ZP_01207097.1	*Mycobacterium vanbaalenii*	●
MxCyOx01	``Peroxidase'' family protein	Q1D1V4	*Myxococcus xanthus*	
NcCyOx01	Hypothetical protein	Q7RUD3	*Neurospora crassa*	
NeCyOx01	Putative cyclooxygenase	Q82V61	*Nitrosomonas europaea*	●
NePxc01	Putative peroxidase	Q82T80	*Nitrosomonas europaea*	
NmoPxDo01	Putative peroxidase	ZP_01126301	*Nitrococcus mobilis*	
NmuCyOx01	Heme peroxidase	Q2YBN0	*Nitrosospira multiformis*	●
NmuCyOx02	Heme peroxidase	Q2YBE9	*Nitrosospira multiformis*	●
NtaDiOx	Oxygenase	O82031	*Nicotiana tabacum*	
OarLPO	Lactoperoxidase	Q9MZY2	*Ovis aries* (sheep)	
OsDiOx01	Putative α-dioxygenase	Q2QRV3	*Oryza sativa* (japonica)	
OsiPxDo01	Peroxidockerin	CT836978	*Oryza sativa* (indica)	
OspCyOx01	Putative peroxidase	Q1N3B4	*Oceanobacter* sp.	
PalcCyOx01	Hypothetical protein	Q9ZFX7	*Pseudomonas alcaligenes*	
PhoCyOx01	Cyclooxygenase	Q962I8	*Plexaura homomalla* (carribean coral)	
PlenPxt01	Peroxinectin	Q26059	*Pacifastacus leniusculus* (crayfish)	
PmoPxt01	Peroxinectin	Q95X07	*Penaeus monodon* (black shrimp)	
PnoCyOx01	Hypothetical protein	Q0URU2	*Phaeosphaeria nodorum*	
PosPxc01	Heme peroxidase	Q12AX6	*Polaromonas* sp.	
PtroEPO	Eosinophil peroxidase	XP_523809.1	*Pan troglodytes* (chimpanzee)	
PtroMPO	Myeloperoxidase	XP_001162602	*Pan troglodytes*	
RbaPxDo01	Peroxidockerin	Q7UJQ5	*Rhodopirellula baltica*	
RbaPxDo02	Peroxidockerin	Q7UYG2	*Rhodopirellula baltica*	
RnoEPO	Eosinophil peroxidase	XP_220834	*Ratus norvegicus*	
RnoMPO	Myeloperoxidase	XP_220830.2	*Ratus norvegicus*	
RnoTPO	Thyroid peroxidase	P14650	*Ratus norvegicus*	
RosdCyOX	Putative cyclooxygenase	Q16BB2	*Roseobacter denitrificans*	●
RosPxc01	Peroxicin	A3XF15	*Roseobacter* sp. MED193	●
RpPxc01	Peroxicin	Q13AU2	*Rhodopseudomonas palustris*	●
RsphCyOx02	Putative cyclooxygenase-2	A3PQV0	*Rhodobacter sphaeroides*	●
SarPxDo01	Peroxidase precursor	A1G7A9	*Salinispora arenicola* CNS205	
SavCyOx01	Putative peroxidase and catalase	Q82M86	*Streptomyces avermitilis*	▼
SfoPGHS02	Prostaglandin synthase	Q9PW89	*Salvelinus fontinalis* (brook trout)	
SiniPOX	Putative peroxidase	Q1A3S0	*Siniperca chuatsi* (chinese perch)	
SmePxc01	Hypothetical protein	Q930F7	*Sinorhizobium melioti*	
SoffPxt01	Melanogenic peroxidase	O18504	*Sepia officinalis* (common cuttlefish)	
SpurPxd01	Similar to peroxidasin	XP_797821.2	*Strongylocentrotus purpuratus* (purple sea urchin)	▼
SpurPxt01	Ovoperoxidase	O44391	*Strongylocentrotus purpuratus*	
SscTPO	Thyroid peroxidase	P09933	*Sus scrofa*	
SuCyOx01	Putative cyclooxygenase	ZP_00519474	*Solibacter usitatus*	
TcasPxd01	Peroxidasin	XP_968570	*Tribolium castaneum* (red flour beetle)	▼

(*Continued*)

Table I
(Continued)

Abbreviation	Type of peroxidase	Accession number	Organism	Remark
TcasPxt01	Peroxinectin	XP_967241.1	*Tribolium castaneum*	
TcasPxt02	Peroxinectin	XP_969600.1	*Tribolium castaneum*	
TpsPxDo01	Peroxidockerin	AAFD01000071	*Thalassiosira pseudonana* (diatom)	
TrubPOX	Putative peroxidase	CAAB01001546	*Takifugu rubripes*	
XlPOX01	Polysomal ribonuclease 1	Q9YH34	*Xenopus laevis* (African clawed frog)	
XlPOX02	Peroxidase 2	Q8QFX3	*Xenopus laevis*	
XtPGHS01	Prostaglandin synthase	UniGene Str.14357	*Xenopus tropicalis*	
XtPxd	Peroxidasin	DAA05635	*Xenopus tropicalis*	▼

The extensions used in abbreviated names denote: myeloperoxidase (-MPO), eosinophil peroxidase (-EPO), lactoperoxidase (-LPO), thyroid peroxidase (-TPO), peroxidasin (-Pxd), peroxinectin (-Pxt), prostaglandin H synthase (-PGHS), cyclooxygenase (-CyOx), α-dioxygenase (-DiOx), dual oxidase (-DuOx), peroxicin (-Pxc), peroxidockerin (-PxDo), and -POX for any unspecified peroxidase of this superfamily. All abbreviations correlate with the nomenclature used in PeroxiBase (http://peroxibase.isb-sib.ch). Supposed pseudogenes within the peroxidase coding region are labeled with ● and fused proteins are labeled with ▼.

Multiple sequence alignment

Multiple sequence alignment of the 134 selected heme enzymes was conducted with Clustal X[18] with following optimized parameters: for the pairwise alignment: gap opening penalty 9, gap extension penalty 0.1 and for the multiple alignment gap opening penalty 8 and gap extension penalty 0.2. Gonnet protein weight matrix was used and the gap separation distance was set to 4. The output was depicted with Genedoc.[19]

Phylogeny reconstruction—distance method

The refined multiple sequence alignment of 134 peroxidase domains in the total length of 608 amino acid positions was subjected to the neighbor-joining (NJ) method of the MEGA package[20] with 1000 bootstrap replications, pairwise deletion of gaps, the Jones–Taylor–Thornton (JTT) model of amino acid substitution, and homogenous pattern among lineages. The optimized γ parameter for this set of sequences was estimated as $\gamma = 2.1$. The resulting tree was presented in the Tree Explorer within the MEGA package as a rectangular phylogram.

Phylogeny reconstruction—maximum likelihood method

The same multiple sequence alignment was subjected to the maximum likelihood (ML) method of the Phylip package.[21] After producing 100 bootstrap replicates, the data set was subjected to the ProML method. Global rearrangement of the sequence list and JTT model of amino acid substitution with the same γ parameter as for the distance method were used. The most likely unrooted tree was visualized with the program TreeView.[22]

Phylogeny reconstruction—maximum parsimony method

The phylogeny of this superfamily was also reconstructed using the maximum parsimony (MP) method of the Phylip package with following options: 100 bootstrap replicates, randomized input order of sequences, ordinary parsimony in all aligned sites. The most parsimonious unrooted trees were also visualized with the program TreeView.[22] Finally, the consensus bootstrap trees obtained by all three methods were compared, and the most likely reconstructed tree was visualized with TreeView.

Known protein structures and secondary structure prediction

PDB structure files that were used as reference structures in this work were 1CXP for human MPO,[23] 2IKC for a LPO from *Ovis aries*,[24] and 1DIY for human cyclooxygenase PGHS-1.[25] The prediction of secondary structure for selected enzymes was performed from the PSIPRED server.[26]

RESULTS

Phylogenetic reconstruction

Blast searches with the 608 amino acids long heme peroxidase domain as query in all public sequence databases, particularly in PeroxiBase, reveal the surprisingly broad abundance of this protein superfamily formerly known as "animal" or "MPO" family among all living forms (not just among animal genomes). The Interpro entry No. IPR002007 (also known as Pfam03098) currently counts 459 sequences (July 2007) with an ever-increasing portion being represented by putative ORFs. Thus, a critical sequence and phylogenetic analysis of these oxidoreductases is strongly needed as is the elucidation of their relation to the well-investigated mammalian peroxidases. This is important in order to detect potential pseudogenes among the new sequences as well as to obtain information about their putative enzymatic activity and physiological role(s).

The criteria for the definition of a new peroxidase superfamily are fulfilled for IPR002007 since (i) the

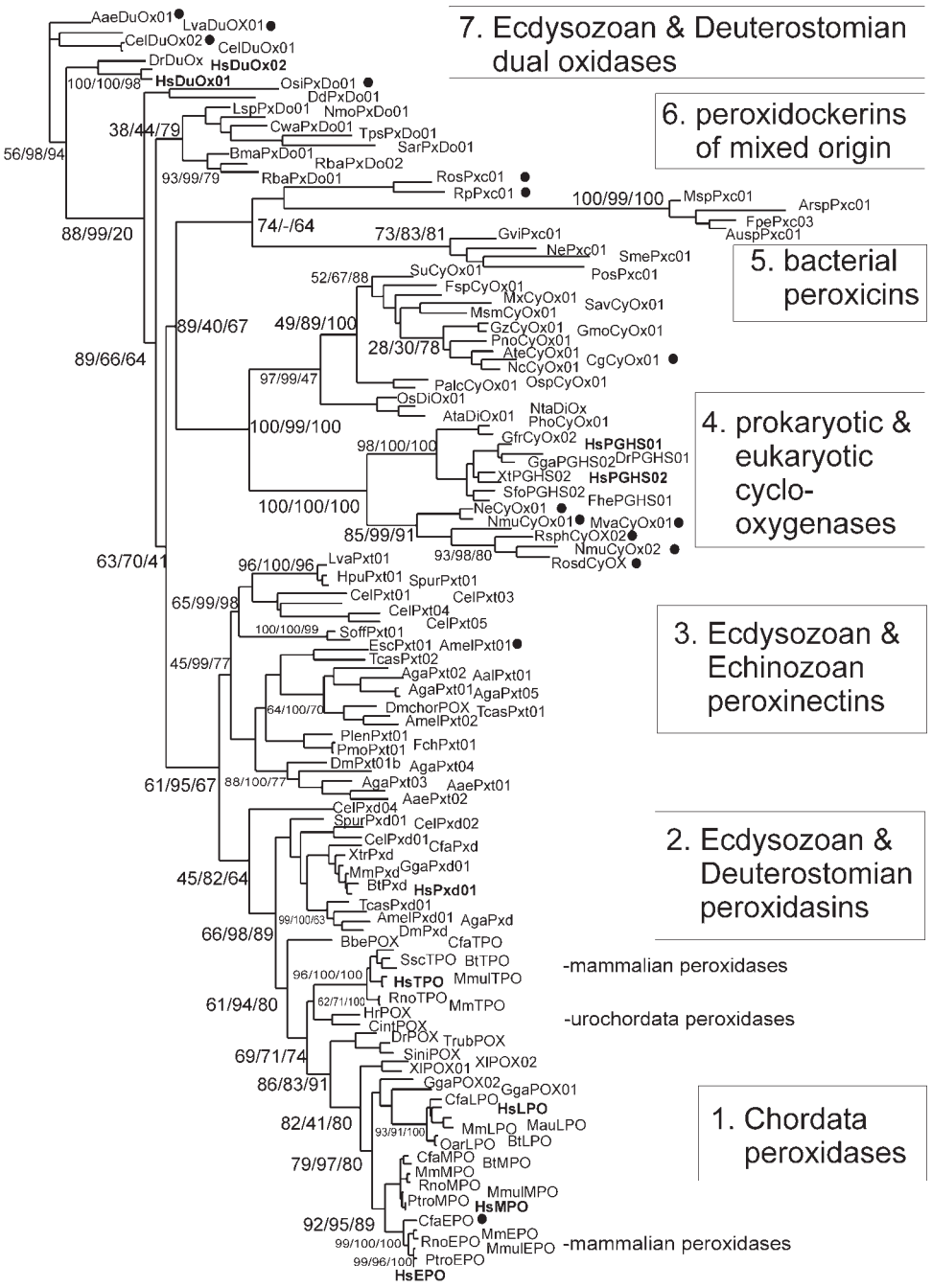

Figure 1

Reconstructed unrooted phylogenetic tree of 134 peroxidase domains. Bootstrap values in the nodes were obtained from ML/MP/NJ methods, respectively.

sequences of these heme proteins are homeomorphic, that is, their peroxidase domains can be completely aligned from end-to-end, and (ii) seven distinct subfamilies diverged from a common ancestor to perform different functions while retaining common structural elements.[27]

The phylogenetic relationships in the obtained multiple sequence alignment of 134 peroxidase domains (Supplemental Fig. A) were reconstructed by using three different methods. The reconstructed unrooted tree presented in Figure 1 is a consensus tree obtained by the ML method. Very similar consensus trees were also obtained

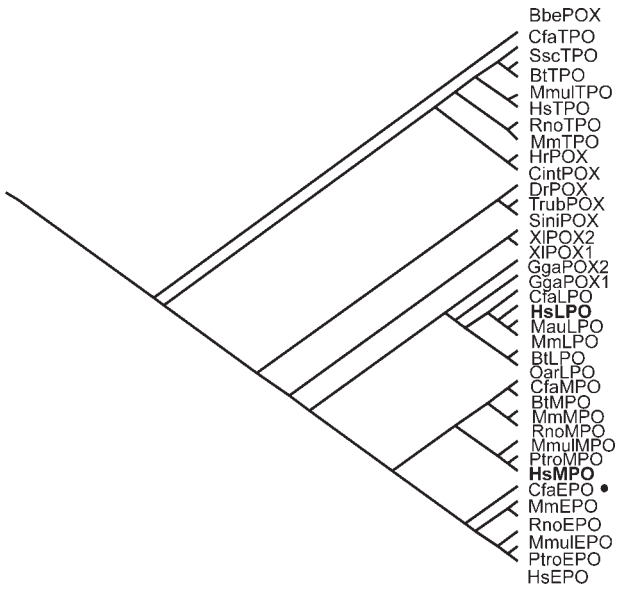

Figure 2

Detail of the reconstructed phylogenetic tree showing the subfamily of vertebrate peroxidases including the mammalian enzymes myeloperoxidase (MPO), eosinophil peroxidase (EPO), lactoperoxidase (LPO), and thyroid peroxidase (TPO).

by NJ and MP methods. The bootstrap values for all used methods are indicated on the branches of the ML tree.

The subfamily of chordata peroxidases (Fig. 2—a detail of Fig. 1) has a good bootstrap support in all branches. The main clades are represented by mammalian MPOs and EPOs that are well separated from the LPO branch (LPO). In the latter also related bird peroxidase genes, for example, GgaPOX01 and GgaPOX02 (the latter within a fusion protein), are found. The primary translation product of these enzymes contains a signal peptide, a propeptide, and the peroxidase domain [Fig. 3(A)]. From mammalian MPO, EPO, and LPO it is known that in mature soluble proteins both the signal and propeptide has been removed by proteolytic processing.[8]

Clearly segregated from the main clades are frog peroxidases, for example, XIPOX01 and XIPOX02, with XIPOX01 having been reported to possess a ribonuclease activity,[28] and the branch of fish peroxidases, for example, DrPOX, TrubPOX, and SiniPOX.

A well resolved clade, distantly related with MPO, EPO, and LPO, contains enzymes homologous to mammalian TPOs. Here, interestingly related enzymes from invertebrates are found, for example, HroPOX and CintPOX, which are well separated from the mammalian membrane-anchored proteins that are known to be expressed and function in thyroid glands.

On the origin of the subfamily of vertebrate peroxidases, BbePOX from a marine ancestor of vertebrates is found. It is the closest phylogenetic neighbor of peroxidasins but does not contain the immunoglobulin motif of this subfamily [Fig. 3(B)].

The second subfamily is represented by peroxidasins (Pxd). Peroxidasins are found in invertebrates and vertebrates including mammals. The first peroxidasin was found in *Drosophila* and described as a protein combining peroxidase and extracellular matrix motifs.[13] The multidomain proteins contain a signal peptide, several (e.g., five in *Drosophila* or one in *Homo sapiens*) leucine-rich regions, followed by four immunoglobulin-like domains, a linking region, the peroxidase domain, and a carboxy terminal Willebrand type C protein–protein interaction domain [Fig. 3(B)].

The peroxidase domains of peroxidasins are well separated from other subfamilies and—with exception of

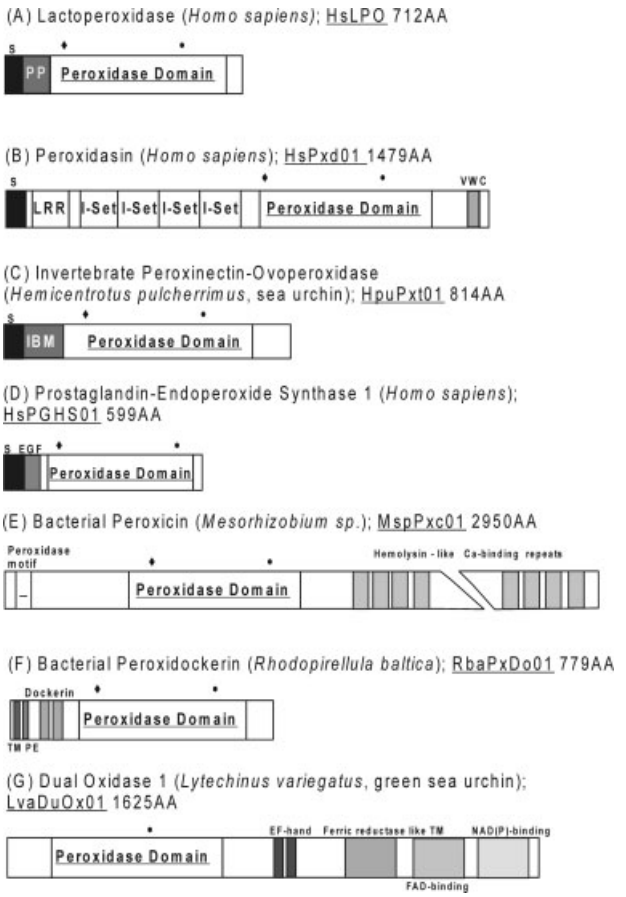

Figure 3

Schematic view of the domain structures of seven typical representatives of the peroxidase–cyclooxygenase superfamily. Abbrevations: S, signal peptide; PP, propeptide; LRR, leucin-rich repeats; I-Set, immunoglobulin I-set domain; VWC, von Willebrand factor type C domain; IBM, integrin-binding motif; EGF, epidermal growth factor like domain; TM, transmembrane helices; PE, planctomycete extracellular motif; EF-hand, EF-hand calcium binding domain. The domain architectures were adopted from the Pfam-Database at the Sanger-Institute (http://www.sanger.ac.uk/Software/Pfam/). The rhombus indicates position of distal and the circle of proximal histidine.

CelPxd04—closely related to each other. The profile of the phylogenetic tree suggests that vertebrate peroxidases probably evolved from a primitive peroxidasin ancestor by duplication of the peroxidase domain and loss of the residual protein part containing the leucine-rich and immunoglobulin-like motifs.

The third subfamily is formed by peroxinectins (Pxt) of predominantly Ecdyzoan origin. This abundant subfamily reveals the highest bootstrap support from the parsimony method. The first peroxinectin was detected in crayfish blood in 1988[29] and described as a cell adhesion molecule with a peroxidase domain and an integrin-binding motif [Fig. 3(C),[11,29]]. Two main peroxinectin clades are discernible, namely arthropod and worm peroxinectins. So far, no sequences from vertebrates peroxinectins are known suggesting that this evolutionary line represents an impasse of evolution (not spread among vertebrates). Among *Deuterostomia* only a minor subclade of *Echinozoa* peroxinectins exists (represented by, e.g., LvaPxt01), but this group is very distantly related with any known sequence of vertebrate peroxidases of this superfamily. In invertebrate genomes, duplicated or multiplicated variants of peroxinectin genes are found most probably as a result of repeated gene duplication.

The fourth subfamily is constituted by cyclooxygenases (PGHSs). This subfamily diverged together with the fifth subfamily very early in the history of the peroxidase–cyclooxygenase superfamily representing one of the major streams of evolution. The primary translation product of cyclooxygenases consists of a N-terminal signal peptide followed by an epidermal growth factor domain, a membrane-binding domain and the globular catalytic peroxidase domain [Fig. 3(D)]. The cyclooxygenase genes evolved partially as paralogs in vertebrate genomes that fulfill different physiological roles.

The subfamily of cyclooxygenases is divided into two main clades with a high bootstrap support. Whereas the clade containing both human PGHS paralogs is also directly connected with putative bacterial cyclooxygenases, the second clade is spread mainly among plant and fungal genomes. Very interestingly, the putative fungal cyclooxygenases are phylogenetically very closely related with bacterial ORFs where the physiological function of the corresponding genes is unknown at the moment. Also the native function of fungal cyclooxygenase genes has to be determined experimentally.

The fifth clade is represented by bacterial peroxicins that were formed early in the evolution as a distinct clade. A high portion of currently known members is formed by extremely long pseudogenes (up to 3300 amino acids), where the peroxidase domain or at least some motifs can be repeated several times—probably a relict of imperfect gene duplication(s) and intensive mutations. Moreover, these large fusion proteins contain hemolysin-type toxin and Ca^{2+}-binding motifs [Fig. 3(E)] that could indicate a role in defence of bacteria possessing this multidomain protein against other bacteria of the same environment that lack such a gene. Thus, in analogy with bacteriocins, that is, proteinaceous toxins produced by bacteria to inhibit the growth of similar or closely related bacterial strain(s), we suggest the denomination peroxicin. However, none of the fifth subfamily members has an experimentally verified physiological function. In certain bacteria (e.g., *Nitrosomnas europaea*) gene paralogs of the fourth and fifth subfamily represent divergent evolutionary lines in different genome locations.

Table II
GC Content of Selected Peroxidase Genes (From Subfamilies 4 and 6) and of the Corresponding Organisms

Peroxidase abbreviations (origin)	GC content of the whole gene (%)	GC content of whole organism (%)
DdPxDo01 (slime mold)	30.51	28.62
TpsPxDo01 (diatom)	44.55 ($\Delta = -5.1\%$)	49.64
OsiPxDo01 (from plant)	35.42 ($\Delta = -19.8\%$)	55.26
RbaPxDo02 (bacterial)	54.41	55.46
NmoPxDo01 (bacterial)	54.07	50.30
SuCyOx01 (bacterial)	62.11	62.43
FspCyOx01 (bacterial)	67.16	70.54
GzCyOx01 (fungal)	49.42	51.13
NcCyOx01 (fungal)	53.93	56.07

GC content of particular genes was calculated with Genedoc.[19] GC content of the whole organisms is according to the database maintained at http://www.kazusa.or.jp.

The sixth clade is represented by peroxidockerins of mixed origin. These multidomain proteins are composed of a transmembrane domain, a sequence motif characteristic for planctomycete extracellular proteins, two Dockerin type I repeats and the peroxidase domain [Fig. 3(F)]. Obvious gene transfers are revealed in the peculiar distribution of clades in this subfamily with closest neighborhood of prokaryotic and eukaryotic PxDo genes. This could indicate either a lateral gene transfer (LGT) or an endosymbiotic event. The comparison of GC content in PxDo genes analyzed here versus whole genomes revealed a significant difference only in the genome of rice and a slight difference in the genome of a diatom (Table II). Multiple DNA alignment showed significant similarity of corresponding prokaryotic and eukaryotic sequences only within a short region $\pm$500 bp up- and down-stream of the coding region (alignment not shown). The location of introns is very similar in DdPxDo01 and TpsPxDo01: in both cases short introns are located on the N-terminus of the region coding for the peroxidase domain and the major portion of this region is intronless (details not shown). All above-mentioned facts could indicate that only relative short stretches of genomic DNA were transferred between ancestral bacteria and protists, or that the ancestral eukaryotic genome was rearranged after the earliest LGT. In analogy with similar investigations on genes coding for the pentose phosphate pathway,[30] we cannot resolve

Table III
Differences Between Invertebrate Peroxinectins and Bacterial Peroxicins and Peroxidockerins Concerning Residues Involved in Heme to Protein Linkage, Distal Side Ca^{2+}-Binding, and H-Bonding to Proximal Histidine

HsMPO position aligned (processed)	Function in vertebrate peroxidases	Invertebrate peroxinectins	Peroxicins and peroxidockerins	Functional group involved in binding
D260 (D94)	Heme to protein linkage	32% D	96% D[a]	Carboxyl side chain
D262 (D96)	Ca^{2+}-binding	89% D (8% E)	70% D	Carboxyl side chain
T334 (T168)	Ca^{2+}-binding	95% T/S	87% T/S	Hydroxyl oxygen
F336 (F170)	Ca^{2+}-binding	100% F/Y/W	65% F/Y/W	Peptide carbonyl
D338 (D172)	Ca^{2+}-binding	100% D	83% D	Carboxyl side chain
S340 (S174)	Ca^{2+}-binding	97% S (3% N)	35% S	Hydroxyl oxygen
E408 (E242)	Heme to protein linkage	24% E (46% Q)	100% E	Carboxyl side chain
N587 (N421)	H-bonding to proximal histidine	68% N (35% D)	4% N (35% D)	Amide carbonyl group

Note that in all peroxinectins the distal triad His261-Arg405-Gln257 and the proximal His are present. Exception *Tribolium castaneum* peroxinectin 2 that contains Glu instead of Gln257. Numbering corresponds to human myeloperoxidase (HsMPO).
[a]One bacterial group has F preceding D260, whereas in majority of peroxinectins I/V/L precede D260.

whether the particular genes were acquired by endosymbiosis or via independent LGT. Clearly, further genome-wide analysis is necessary, which is outside the scope of this work. The physiological function of peroxidockerins that seem to have several common sequence motifs in the heme cavity with phylogenetically distantly related Ecdyzoan peroxinectins (Table III), remains to be elucidated.

The last (seventh) subfamily is formed by eukaryotic dual oxidases. Dual oxidases are multidomain proteins with an N-terminal peroxidase domain linked with a C-terminal reductase domain by a cytoplasmic bridge with two EF hands. The peroxidase domains of dual oxidases were segregated early in the evolution of this superfamily and were further differentiated within this subfamily.

Further sequences can be awaited from future genome projects to elucidate the difference between the peroxidase domain of dual oxidases and peroxinectins.

By observing the main evolutionary streams and the abundance of the particular sequence, it is appropriate to name this superfamily peroxidase–cyclooxygenase superfamily. This new name is based on the properties of the main clades and not on the origin of particular genes. Numerous examples obvious in the reconstructed tree suggest frequent gene duplication events that in some cases led to a "death-end" product (i.e., observed pseudogenes within the peroxidase coding region in six of total seven subfamilies) but sometimes to proteins with a novel function (e.g., XlPOX01). Furthermore, the presence of multiple paralogs of a conserved peroxidase gene with different functions of the corresponding proteins, for example, MPO, EPO, LPO, TPO, Pxd01, Pxd02, PGHS01, PGHS02, DuOx01, and DuOx02 in human genome leads us to the conclusion that this superfamily obeys the model of birth-and-death process of multigene family evolution.[31] This observation is in accordance with the same conclusion for the evolution of the "non-animal" peroxidase superfamily.[32] It appears that heme peroxidase genes need the mechanism of birth-and-death evolution to maintain the fascinating diversity of heme peroxidases.

Conserved motifs in the multiple sequence alignment

As numerous members of IPR002007 are fusion proteins (see Fig. 3), we focused our current analysis on the peroxidase domain, which—although positioned in different locations of the fused proteins or standing alone—is highly conserved and essential for all members. In the optimized multiple sequence alignment, four highly conserved regions were obtained. They correspond to functionally and structurally essential motifs as known from mammalian peroxidases: two on the distal [Fig. 4(A,B)] and two on the proximal side of the prosthetic heme group [Fig. 4(C,D)].

The secondary structure of mammalian peroxidases is predominantly α-helical and each monomer has a central heme-containing core composed of five α-helices [Fig. 5(B)]. Both the distal and proximal histidine as well as its H-bonding partner are located within α-helices. Essential distal residues in mammalian peroxidases are Gln257, His261, and Arg405 [Fig. 5(A), if not specially indicated numbering corresponds to human MPO, HsMPO]. The peroxidase-typical distal pair His-Arg is found in all peroxidases from both superfamilies and has a function in the heterolytic cleavage of H_2O_2.[2] All the sequences analyzed here had the distal histidine and—with exception of the cyclooxygenase subfamily (and pseudogenes in all subfamilies)—also the distal arginine [Fig. 4(A,B)].

Gln257 is not found in the distal cavity of the (archae) bacterial, fungal, and plant heme peroxidase superfamily. In mammalian peroxidases, it is involved in maintenance of the distal hydrogen bond network and in halide binding (Fig. 5,[8]). Interestingly, this essential glutamine, which is part of a conserved motif "W/F-G/I/A-Q," is

A

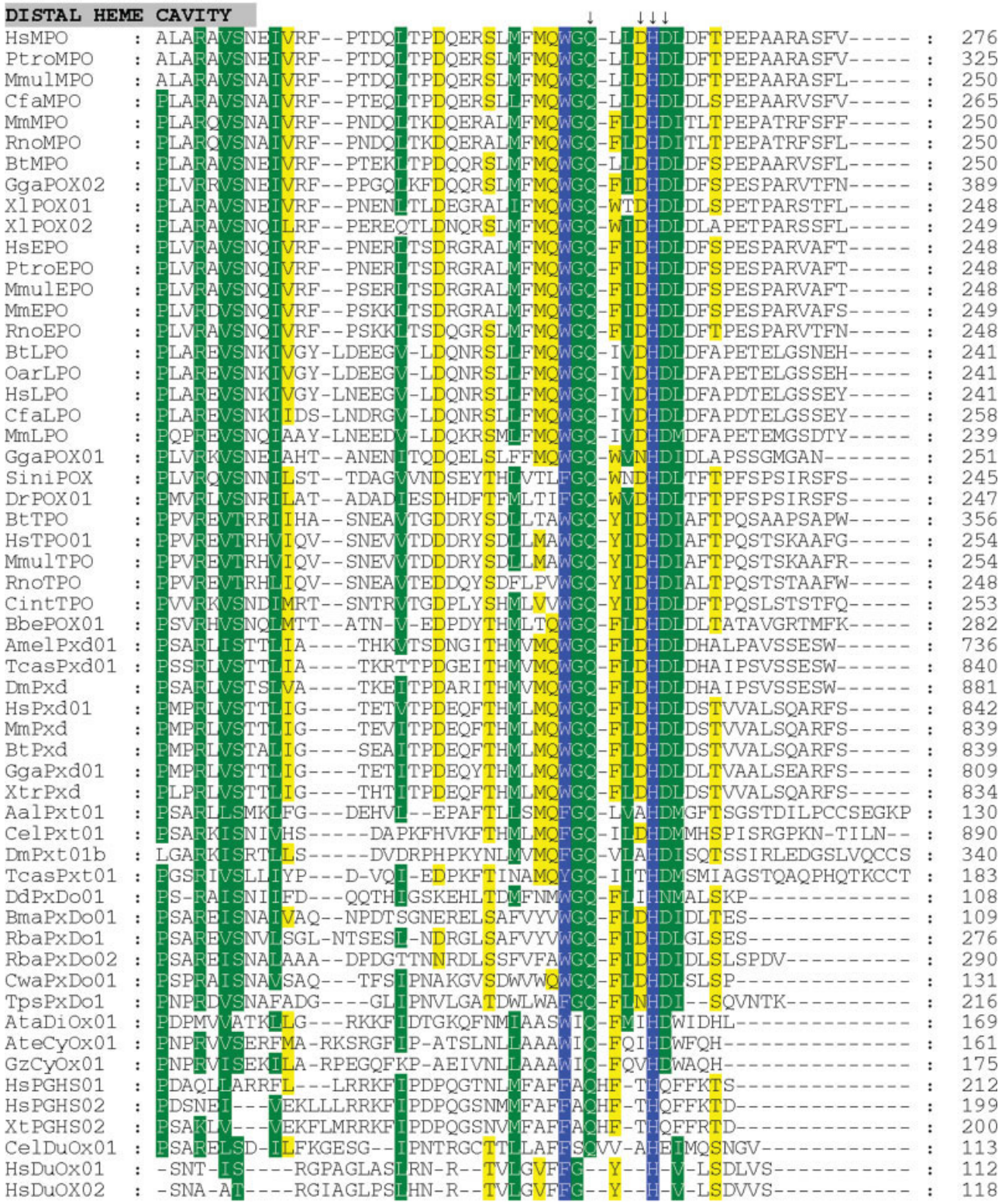

Figure 4

*Selected parts of the multiple sequence alignment of 56 typical members of the peroxidase–cyclooxygenase superfamily. Areas around the essential distal His and Gln (**A**), conserved distal Arg (**B**), essential proximal His (**C**), and conserved proximal Asn (**D**) are shown. Abbreviations of analyzed sequences are identical with those used in PeroxiBase (http://peroxibase.isb-sib.ch/). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]*

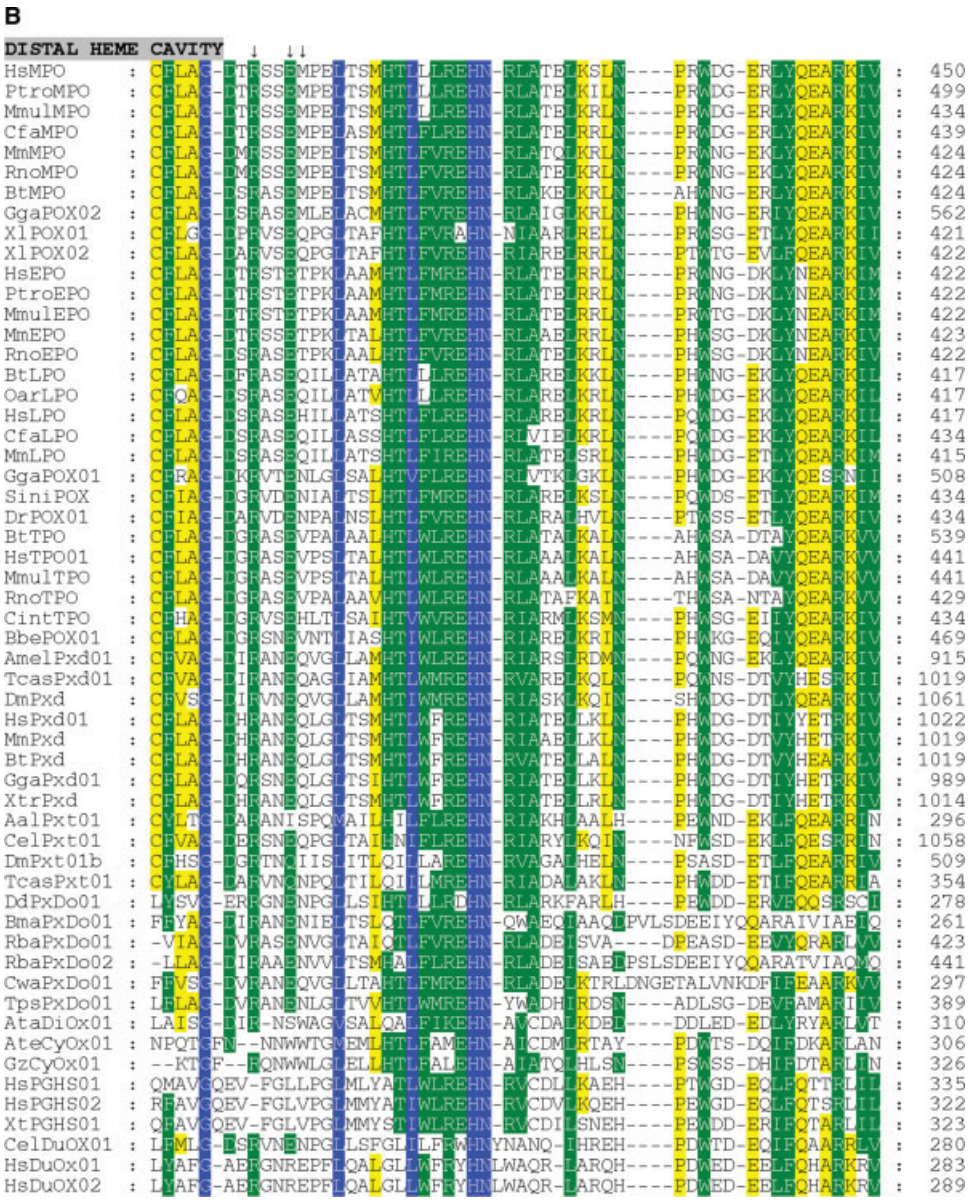

Figure 4
(Continued)

present in almost all analyzed sequences with the exception of both human dual oxidases and one peroxinectin from *Tribolium castaneum* [Fig. 4(A) and Supplemental Fig. A].

The second adjacent conserved motif in mammalian peroxidases that includes the distal His261, is "D-H-D." Asp260 is known to be involved in ester bond formation with the 5-hydroxymethyl group on pyrrole ring C (see Fig. 5). It is found in all vertebrate peroxidases and peroxidasins. In many invertebrate and few peroxicins as well as all dioxygenases, cyclooxygenases, and dual oxidases this acidic residue is absent. Regarding Asp262, its role as a ligand of Ca^{2+} is well established and it is found in all vertebrate peroxidases as well as peroxidasins and almost all invertebrate peroxinectins and peroxidockerins (with the exception of a *Dictyostelium* peroxidockerin). In about 30% of peroxicins and in all cyclooxygenases and dual oxidases Asp262 is missing [Fig. 4(A)]. In cases where both essential distal motifs are not present (e.g., CfaEPO) or incomplete (e.g., AmelPxt01 or AuspPxc01, see Table I for details), we assume that these are pseudogenes that cannot be translated in functional peroxidases.

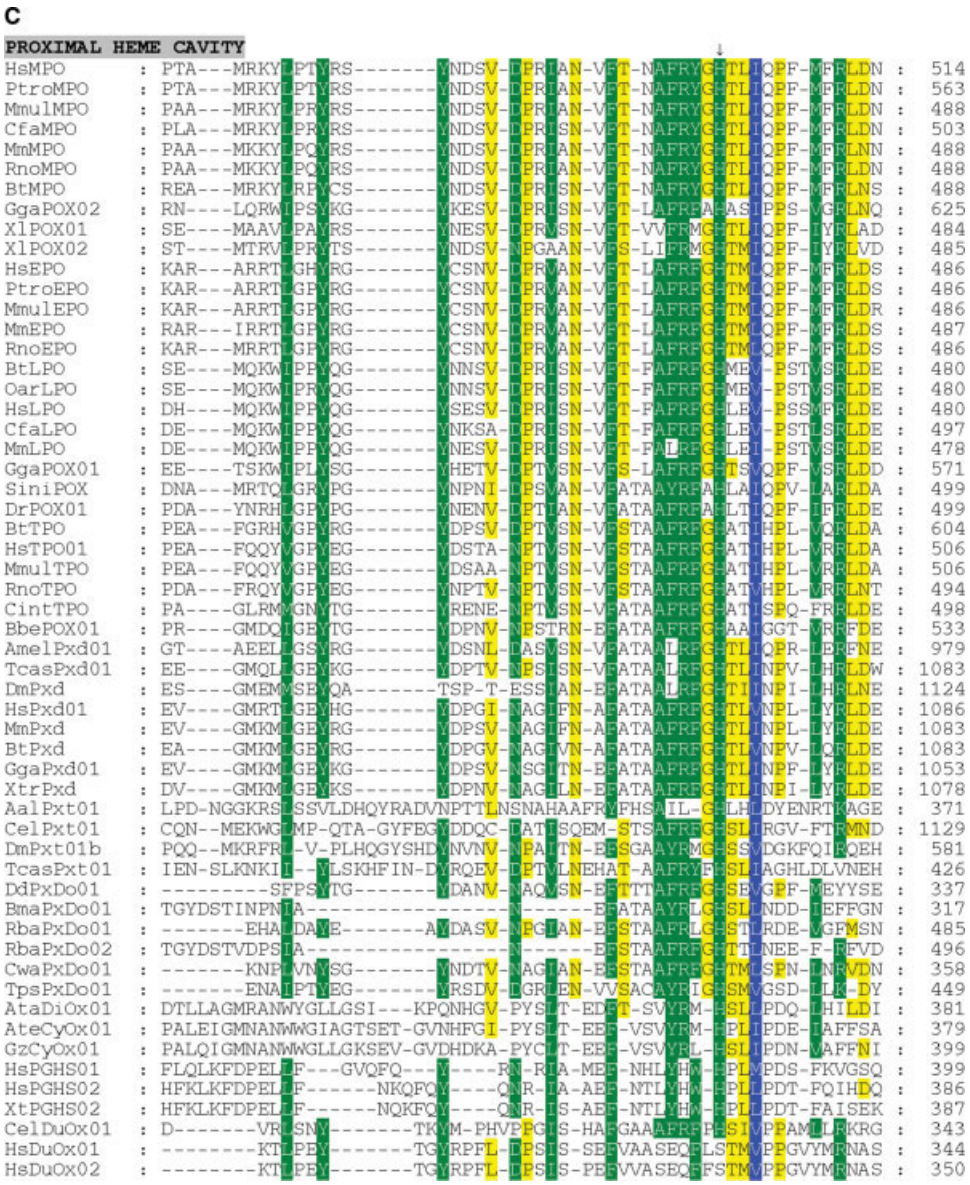

Figure 4
(Continued)

In MPO—besides Asp260—Glu408 and Met409 are involved in heme to protein linkages (see Fig. 5). Figure 4(B) clearly demonstrates that Glu408 is found in all vertebrate peroxidases, peroxidasins, peroxinectins, peroxicins, and peroxidockerins but is missing in dioxygenases, cyclooxygenases, and dual oxidases. By contrast, Met409 is conserved only in mammalian MPOs [Fig. 4(B)] underlining that the denomination "MPO superfamily" is inappropriate.

Structurally and functionally important motifs on the proximal side of mammalian peroxidases include the proximal histidine (His468) and its H-bonding partner Asn587 [Fig. 5(A)]. Both residues govern the heme iron reactivity by controlling the electron density at the metal.[2] By inspecting the alignment part in Figure 4(C), the most important fact is that in both human dual oxidases—but not in lower eukaryotic dual oxidases (not shown)—the His is exchanged by a serine, although the immediate vicinity is only moderately modified. Apparently this mutation occurred very late in the evolutionary history of these enzymes. All other subfamilies contain the essential proximal histidine. Figure 4(D) depicts that Asn587 is not present in dioxygenases, cyclooxygenases, and dual oxidases. In both invertebrate peroxi-

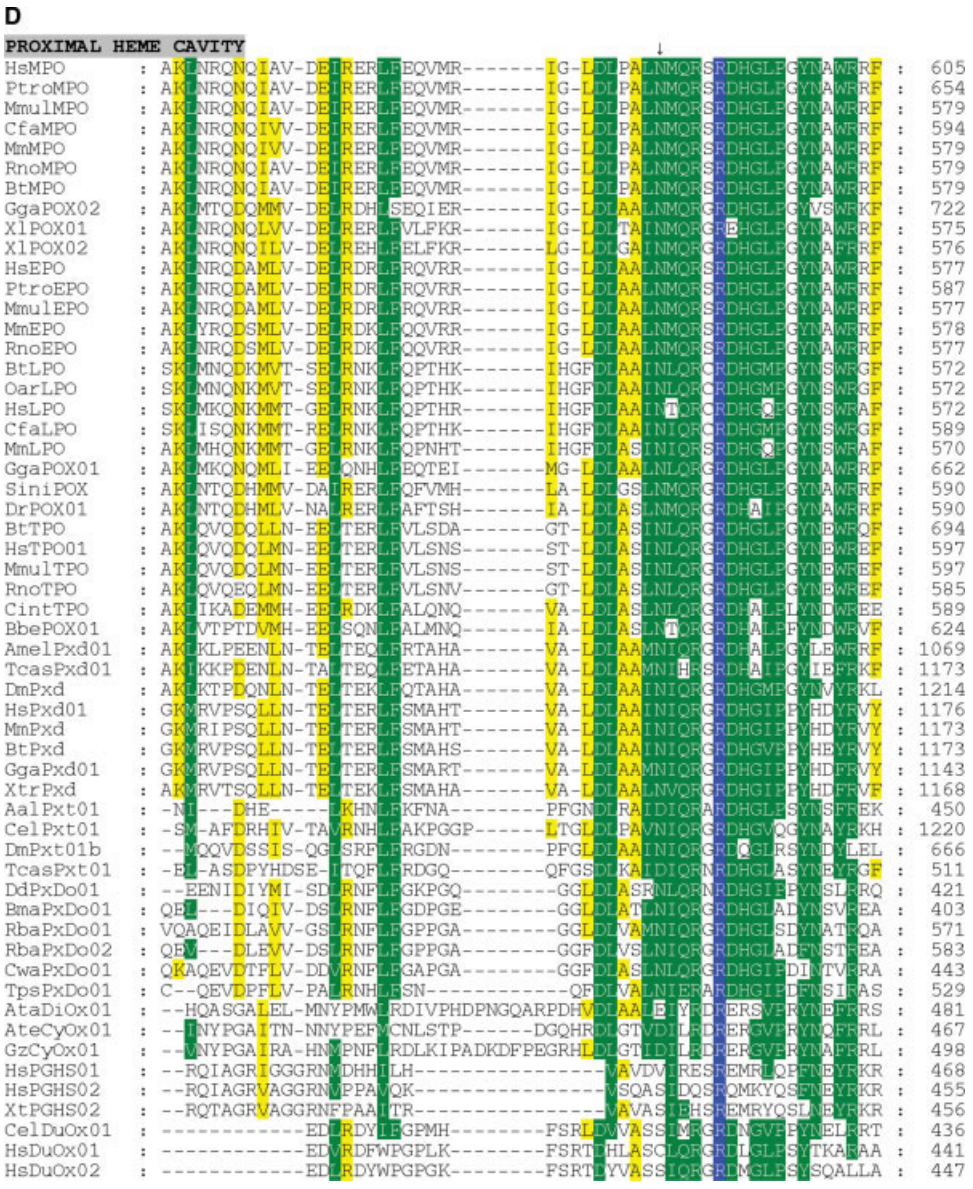

Figure 4
(Continued)

nectins and peroxicins, it can be substituted by aspartate (which renders these enzymes similar to peroxidases from the other superfamily) or even by residues that cannot function as H-bond acceptors of proximal histidine (Supplemental Fig. B and Table III).

The already-mentioned Ca^{2+}-binding site is a further peculiarity of mammalian peroxidases. Its role could be to stabilize both the distal heme cavity architecture as well as to mediate the assembly of the mature peroxidases.[8] In HsMPO, the Ca^{2+} binding site has a typical pentagonal bipyramidal coordination geometry and the cation is liganded by several highly conserved residues. One is the already-mentioned Asp260 neighbored to distal His. The other ligands are part of a loop that consists of eight conserved residues "Leu333-Thr-Ser-Phe-Val-Asp-Ala-Ser340" and is found in all vertebrate peroxidases, peroxidasins, most invertebrate peroxinectins, and peroxidockerins. It is not present in most peroxicins, cyclooxygenases, dioxygenases, and dual oxidases (data in Supplemental Fig. A).

DISCUSSION

Chordata peroxidases represent the youngest subfamily of this superfamily. The phylogentic tree (see Fig. 1)

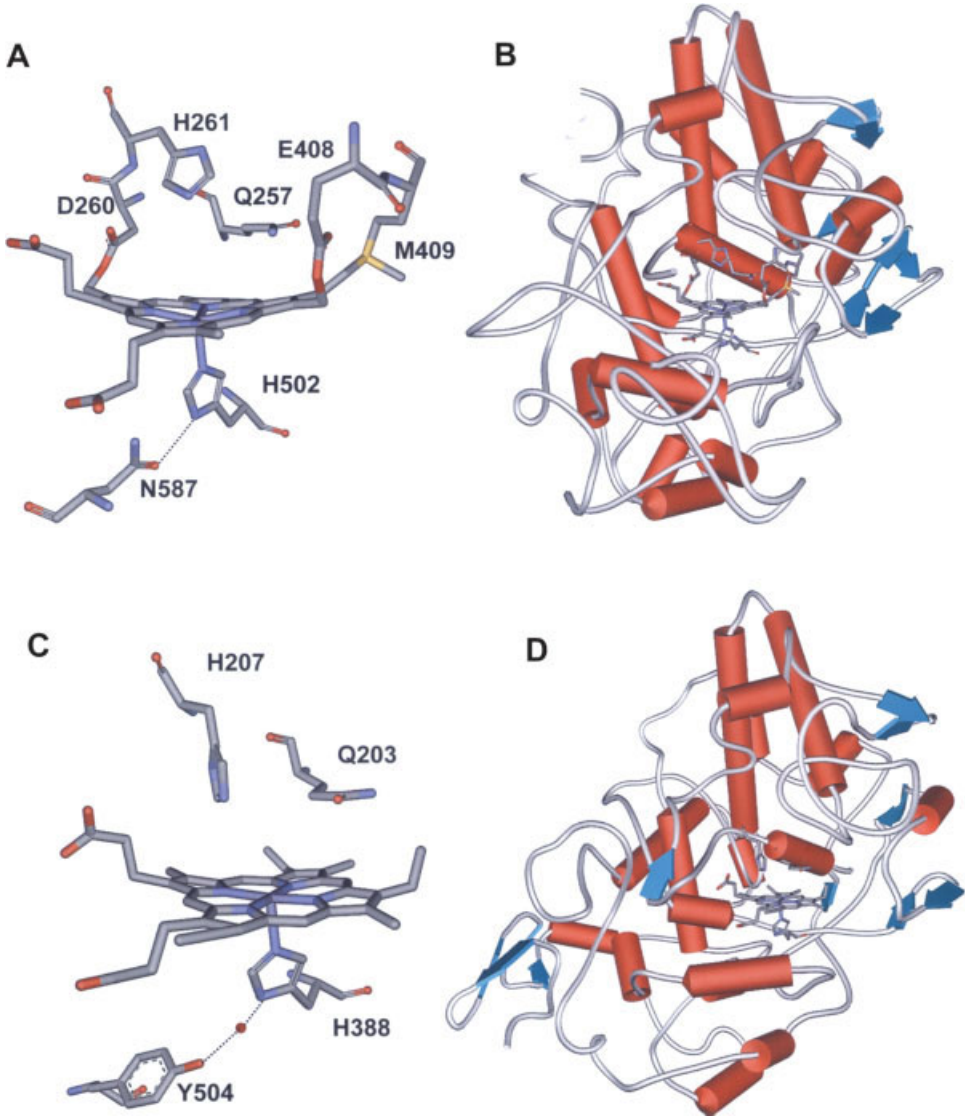

Figure 5

*Structural comparison of myeloperoxidase and cyclooxygenase. (**A**) The heme cavity architecture of human myeloperoxidase. Note the nonplanar porphyrin ring in MPO and its covalent attachments via two ester bonds (Asp260 and Glu408) and one sulfonium ion linkage (Met409). The latter is only found in the myeloperoxidase clade of mammalian peroxidases. In addition, the catalytic residues His261, Arg405, and Gln257 are shown, the latter being important in substrate (halide) binding. The proximal heme ligand is His502 that is hydrogen-bonded to Asn587. The figure was constructed using the coordinates deposited in the Protein Data Bank (accession code 1cxp). (**B**) Monomeric structure of human myeloperoxidase showing the assignment of secondary structure elements and the prosthetic group. (**C**) View of the active site of human cyclooxygenase (PGHS02) showing the conserved distal residues His207 and Gln203, as well as proximal His388 that is hydrogen bonded via a water molecule with Tyr504. The figure was constructed using the coordinates deposited in the Protein Data Bank (accession code 6COX). (**D**) Monomeric structure of human cyclooxygenase 2 (PGHS02) showing the assignment of secondary structure elements and the prosthetic group.*

very well reflects the well known physiological role of these metalloproteins. Both MPO and EPO are stored in granules of special leukocytes (neutrophils, monocytes, and eosinophils, respectively) and released when pathogens and parasites invade the human body as well as at sites of inflammation.[33,34] By contrast, LPO is not stored in special cell types but released from glands to mucosal surfaces and exocrine secretions such as milk, tears, and saliva.[35] All three peroxidases function in unspecific host defence. Closely related with this physiological role is the nature of the substrates oxidized by these oxidoreductases. They prefer small anionic molecules as electron donors such as halides (chloride, bromide, iodide), thiocyanate, and nitrite. The corresponding oxidation products, for example, hypohalous acids, are responsible for killing microorganisms. Oxidizing halides like chloride and bromide is thermodynamically challenging and needs a special design of the heme cav-

ity that guarantees reasonable oxidation rates as well as protects the enzyme from its own oxidizing and halogenating products. In this respect, the role of heme to protein linkages is definitively important.[10] This work clearly suggests that all vertebrate peroxidases have at least two conserved acidic amino acids (Asp, Glu) that allow the formation of ester bonds with modified heme *b*. Only one clade of mammalian enzymes is distinctive (MPOs) in that these enzymes form a third covalent sulfonium ion linkage, which has been demonstrated to be essential for chloride oxidation.[10]

Whereas MPOs and EPOs are found exclusively in mammals, enzymes homologous to mammalian lactperoxidase are also found in birds. Moreover, this article clearly demonstrates the occurrence of separate clades of fish and frog peroxidases suggesting that these organisms could also use these heme enzymes with antimicrobial activity in their innate unspecific immune system.

Regarding mammalian peroxidases, TPO is known to have the lowest homology to MPO.[8] This fits well with the fact that the TPO clade branches off very early in the evolution of vertebrate peroxidases and also correlates with its physiological role in thyroid gland function. Its preferred substrate is iodide and TPO is responsible for the biosynthesis of the hormones triiodotyronine and thyroxine. Interestingly, our phylogenetic analysis revealed that TPO-related enzymes are also found in basal chordates, the ascidians *Ciona intestinalis* and *Halocynthia roretzi*, and the cephalochordate *Brachiostoma belcheri* (see Fig. 2). This fits well with the general accepted fact that the endostyle of basal chordates, a pharyngeal organ involved in filter feeding, is a thyroid homolog and evolutionary predecessor of thyroid glands found in vertebrates.[36] Moreover, recent investigations revealed the presence of a TPO-related enzyme in the endostyle of the above-mentioned basal chordates.[36]

The peroxidase domains of members of the subfamily of peroxidasins show the closest relationship with vertebrate peroxidases. Peroxidasins are found in invertebrates and vertebrates including mammals. The present sequence (see Fig. 4) and secondary structure (Supplemental Fig. B) analysis clearly demonstrates that all catalytically relevant residues (with exception of the MPO-specific methionine that is responsible for the third covalent linkage between heme and protein) present in vertebrate peroxidases are also found in the peroxidase domain of peroxidasins. This suggests similar enzymatic features and fits with the findings that, for example, the *Drosophila* peroxidasin gene is associated with the function of hemocytes, which are migratory cells present in the hemolymph. Hemocytes are involved in both the intracellular destruction of phagocytosed apoptotic cells as well as of foreign materials[13] and in the deposition of extracellular matrix. *Drosophila* peroxidasin was shown to catalyze the hydrogen peroxide-mediated oxidation of halides as well as formation of dityrosine. This strongly resembles the enzymatic features and physiological roles of vertebrate peroxidases.

The physiological role of vertebrate peroxidasins is not very clear until now. In 2005, the first detailed expression analysis of a vertebrate peroxidasin from *Xenopus tropicalis* was published.[14] It is expressed in several distinct tissues during early development, thereby having a role in modifying extracellular matrix components necessary in morphogenesis. A role in scavenging reactive oxygen species was also discussed. The human peroxidasin orthologue has been identified in several contexts but the physiological role is even more speculative. It is expressed in adult tissues and cancer cell lines.[37]

So far, invertebrate peroxinectins have been described in *Drosophila melanogaster*,[38] the freshwater crayfish *Pacifastacus leniusculus*,[11] the black tiger shrimp *Penaeus monodon*[39] and in the giant freshwater prawn *Macrobrachium rosenbergii*.[40] Invertebrate peroxinectins are described to be exocytosed from blood cells. They gain cell adhesion activity by proteolytical processing after their release from the cell and concomitantly they possess the peroxidase activity. Both these activities again resemble the functions of mammalian leukocyte peroxidases. Indeed, invertebrates do not have acquired immunity; instead they have an innate immune system, which includes phagocytosis, encapsulation of foreign material, antimicrobial action, and cell agglutination.[41] Moreover, peroxinectins are synthesized and stored in semigranular and granular hemocytes and released in response to a stimulus like a foreign particle that enters the hemolymph of the host. Thus, in invertebrates peroxinectins seem to be essential in cellular defence reaction.[41] All peroxidase typical distal (including the Ca^{2+} ligands) and proximal residues are conserved in the peroxidase domains of invertebrate peroxinectins, but interestingly two-third of the members lost the ability to covalently bind the prosthetic group via ester bonds. Moreover, about 35% of invertebrate peroxinectins contain aspartate instead of asparagine as H-bonding partner of proximal histidine (similar to heme peroxidases from the other superfamily). The impact of this structural variation on substrate utilization and as a consequence on the physiological role is unknown.

The physiological role of bacterial peroxicins (subfamily 5) and of subfamily 6 (peroxidockerins of mixed origin) remains elusive. It is interesting to note that eubacteria like *Pirellula* that contain a gene for a peroxidockerin are known to be associated with digestive tracts and the haemolymph of marine invertebrates like *Penaeus monodon*, which are known to synthesize peroxinectins.[42] This fact as well as the presented phylogenetic tree (see Fig. 1) could suggest a LGT from an organism with an ancient member of subfamily 6 to primitive marine invertebrates.

In any case, a close inspection of sequences of subfamilies 5 and 6 clearly demonstrate the presence of all

essential distal (His-Arg-Gln) and proximal (His) residues. Variations are found in the H-bonding partner of proximal histidine as well as in the ligand pattern for a putative Ca^{2+}-binding site. Interestingly, the majority of peroxicins and peroxidockerins have both Asp and Glu for heme to protein linkage. This indicates that this structural feature, which has been shown to be essential for mammalian peroxidases to utilize halides as electron donor[10] has evolved very early in evolution and has been lost later on partially in subfamily 3 (peroxinectins) and totally in subfamily 4 (cyclooxygenases). These findings indicate that members of subfamilies 5 and 6 should have the ability to perform both one- and two-electron oxidation reactions (Reactions 1 and 2).

The reconstructed phylogenetic tree of peroxidase domains demonstrates the occurrence of two further subfamilies (cyclooxygenases and dual oxidases), however, with totally different enzymatic features. This is well reflected by analyzing the peroxidase-typical structural features mentioned above as well as by inspection of known structures of prostanglandin synthases (see Fig. 5) or predicted secondary structure arrangements (Supplemental Fig. B). Cyclooxygenases have neither a distal side arginine nor covalently linked heme and a Ca^{2+}-binding site. In addition, the proximal histidine is H-bonded to a conserved tyrosine via a water molecule [Fig. 5(C)]. On the other hand, both His207 and Gln203 are important for the peroxidase activity of PGHS and the redox intermediates of PGHS are very similar to those of other (classical) peroxidases.[43] The physiological role in mammals is well known and includes both cyclooxygenase and peroxidase activity, that enable these enzymes to catalyze the committed step in prostanoid synthesis.[43] Related enzymes are found in bacteria, fungi, and plants, however their enzymatic features and physiological roles are unknown. These facts—the phylogenetic relationship to peroxidases of this superfamily by simultaneous different enzymatic activity—motivated us to denominate this superfamily as indicated in the title of this article.

In relation to mammalian peroxidases, the subfamily of dual oxidases shows the lowest sequence and structural homology. In contrast to the other subfamilies, dual oxidases contain a second catalytically active (flavo) domain that can bind NADPH. Analysis of the peroxidase domain suggests the absence of heme to protein linkages and of a calcium-binding site. In the human representatives even the proximal histidine is missing, which makes its role as a peroxidase questionable. The physiological role is not clear and is proposed to involve generation of hydrogen peroxide via the flavodomain (oxidase activity) eventually combined with a peroxidase function. An invertebrate dual oxidase was investigated to some detail[44] showing that the oxidase activity could have the same function as the respiratory burst in neutrophils and eosinophils in mammals (which is also mediated by an NADPH oxidase), namely, to generate superoxide by dioxgen reduction. The peroxidase domain was ascribed to have a hydrogen peroxide ("catalatic") degradation function. Clearly, more detailed enzymatic investigations are needed.

CONCLUSION

Summing up, we have reconstructed the phylogenetic relationships of enzymes that have peroxidase domains homologous to mammalian peroxidases. Enzymes from all living kingdoms were found and could be grouped in seven subfamilies. Newly constituted peroxidase–cyclooxygenase superfamily with a broad spectrum of enzymatic features and physiological roles unequivocally suggest that these enzymes have differentiated very early in the evolutionary history as cornerstone of the innate immune defence system. Before organisms developed an acquired immunity, their antimicrobial defence depended on enzymes that were recruited upon pathogen invasion and could produce antimicrobial reaction products. After the critical inspection of all subfamilies, we conclude that even certain prokaryotic organisms possess genes encoding putative antimicrobial enzymes possibly providing selective advantage.

REFERENCES

1. Rouhier N, Jacquot JP. The plant multigenic family of thiol peroxidases. Free Radic Biol Med 2005;38:1413–1421.
2. Dunford HB. Heme peroxidases. New York: Wiley-VCH; 1999.
3. Welinder KG. Superfamily of plant, fungal, and bacterial peroxidases. Curr Opin Struct Biol 1992;2:388–393.
4. Passardi F, Bakalovic N, Teixeira FK, Margis-Pinheiro M, Penel C, Dunand C. Prokaryotic origins of the non-animal peroxidase superfamily and organelle-mediated transmission to eukaryotes. Genomics 2007;89:567–579.
5. Kimura S, Ikeda-Saito M. Human myeloperoxidase and thyroid peroxidase, two enzymes with separate and distinct physiological functions, are evolutionarily related members of the same gene family. Proteins 1988;3:113–120.
6. Taurog A. Molecular evolution of thyroid peroxidase. Biochimie 1999;81:557–562.
7. Daiyasu H, Toh H. Molecular evolution of the myeloperoxidase family. J Mol Evol 2000;51:433–445.
8. Furtmüller PG, Zederbauer M, Jantschko W, Helm J, Bogner M, Jakopitsch CH, Obinger CH. Active site structure and catalytic mechanisms of human peroxidases. Arch Biochem Biophys 2006;445:199–213.
9. Arnhold J, Monzani E, Furtmüller PG, Zederbauer M, Casella L, Obinger C. Kinetics and thermodynamics of halide and nitrite oxidation by mammalian heme peroxidases. Eur J Inorg Chem 2006;19:3801–3811.
10. Zederbauer M, Furtmüller PG, Brogioni S, Jakopitsch C, Smulevich G, Obinger C. Heme to protein linkages in mammalian peroxidases: impact on spectroscopic, redox and catalytic properties. Nat Prod Rep 2007;24:571–584.
11. Johansson MW, Lind MI, Holmblad T, Thornqvist PO, Soderhall K. Peroxinectin, a novel cell adhesion protein from crayfish blood. Biochem Biophys Res Commun 1995;216:1079–1087.

12. Lin X, Cerenius L, Lee BL, Söderhäll K. Purification of properoxinectin, a myeloperoxidase homologue and its activation to a cell adhesion molecule. Biochim Biophys Acta 2007;1770:87–93.
13. Nelson RE, Fessler LI, Takagi Y, Blumberg B, Keene DR, Olson PF, Parker CG, Fessler JH. Peroxidasin: a novel enzyme-matrix protein of Drosophila development. EMBO J 1994;13:3438–3447.
14. Tindall AJ, Pownall ME, Morris ID, Isaacs HV. *Xenopus tropicalis* peroxidasin gene is expressed within the developing neural tube and pronephric kidney. Dev Dyn 2005;232:377–384.
15. Botting RM. Cyclooxygenase: past, present and future. A tribute to John R. Vane (1927–2004). J Therm Biol 2006;31:208–219.
16. Ris-Stalpers C. Physiology and pathophysiology of the DUOXes. Antioxid Redox Signal 2006;8:1563–1572.
17. Passardi F, Theiler G, Zamocky M, Cosio C, Rouhier N, Teixeira F, Margis-Pinheiro M, Ioannidis V, Penel C, Falquet L, Dunand C. PeroxiBase: the peroxidase database. Phytochemistry 2007;68:1605–1611.
18. Jeanmougin F, Thompson JD, Gouy M, Higgins M, Gibson TJ. Multiple sequence alignment with ClustalX. Trends Biochem Sci 1998;23:403–405.
19. Nicholas KB, Nicholas HB. Genedoc: a tool for editing and annotating multiple sequence alignments. Distributed by the author. 1997.
20. Nei M, Kumar S. Accuracies and statistical tests of phylogenetic trees. New York: Oxford University Press; 2000.
21. Felsenstein J. Inferring phylogenies from protein sequences by parsimony, distance, and likelihood method. Methods Enzymol 1996; 266:418–427.
22. Page RDM. TreeView: an application to display phylogenetic trees on personal computers. Comput Appl Biosci 1996;12:357–358.
23. Fiedler TJ, Davecy CA, Fenna RE. X-ray crystal structure and characterization of halide binding sites of human myeloperoxidase at 1.8A resolution. J Biol Chem 2000;275:11964–11971.
24. Singh AK, Singh N, Sharmas S, Singh SB, Kaur P, Bhushan A, Srinivasan A, Singh TP. Crystal structure of lactoperoxidase at 2.4A resolution. J Mol Biol 2007; doi: 10.1016/j.jmb.2007.12.012.
25. Malkowski MG, Ginell SL, Smith WL, Garavito RM. The productive conformation of arachidonic acid bound to prostaglandin synthase. Science 2000;289:1933–1937.
26. Mc Guffin L, Bryson K, Jones D. The PSIPRED protein structure prediction server. Bioinformatics 2000;16:404–405.
27. Meng EC, Polacco BJ, Babbitt PC. Superfamily active site templates. Proteins Struct Funct Bioinform 2004;55:962–976.
28. Chernokalskaya E, Dubell AN, Cunningham KS, Hanson MN, Dompenciel RE, Schoenberg DR. A polysomal ribonuclease involved in the destabilization of albumin mRNA is a novel member of the peroxidase gene family. RNA 1998;4:1537–1548.
29. Johansson MW, Söderhall K. Isolation and purification of a cell-adhesion factor from crayfish blood-cells. J Cell Biol 1988;106:1795–1803.
30. Anderson JO, Roger AJ. A cyanobacterial gene in nonphotosynthetic protests—an early chloroplast acquisition in Eukaryotes? Curr Biol 2002;12:115–119.
31. Nei M, Rooney AP. Concerted and birth-and-death evolution of multigene families. Annu Rev Genet 2005;39:121–152.
32. Zamocky M. Phylogenetic relationships in class I of the superfamily of bacterial, fungal, and plant peroxidases. Eur J Biochem 2004;271: 3297–3309.
33. Klebanoff JS. Myeloperoxidase: friend and foe. J Leukoc Biol 2005; 77:598–625.
34. Wang J, Slungaard A. Role of eosinophil peroxidase in host defense and disease pathology. Arch Biochem Biophys 2006;15: 256–260.
35. Ihalin R, Loimaranta V, Tenovuo J. Origin, structure, and biological activities of peroxidases in human saliva. Arch Biochem Biophys 2006;445:261–268.
36. Ogasawara M. Overlapping expression of amphioxus homologs of the thyroid transcription factor-1 gene and thyroid peroxidase gene in the endostyle: insight into evolution of the thyroid gland. Dev Genes Evol 2000;210:231–242.
37. Horikoshi N, Cong J, Kley N, Shenk T. Isolation of differentially expressed cDNAs from p53-dependent apoptotic cells: activation of the human homologue of the Drosophila peroxidasin gene. Biochem Biophys Res Commun 1999;261:864–869.
38. Vazquez M, Rodriguez R, Zurita M. A new peroxinectin-like gene preferentially expressed during oogenesis and early embryogenesis in *Drosophila melanogaster*. Dev Genes Evol 2002;212:526–529.
39. Sritunyalucksana K, Wongsuebsantati K, Johansson MW, Söderhäll K. Peroxinectin, a cell adhesive protein associated with the proPO system from the black tiger shrimp, *Penaeus monodon*. Dev Comp Immunol 2001;25:353–363.
40. Hsu P-I, Liu C-H, Tseng D-Y, Lee P-P, Cheng W. Molecular cloning and characterisation of peroxinectin, a cell adhesion molecule, from the giant freshwater prawn *Macrobrachium rosenbergii*. Fish Shellfish Immunol 2006;21:1–10.
41. Söderhall K. Invertebrate immunity. Dev Comp Immunol 1999;23: 263–266.
42. Fuerst JA, Sambhi SK, Paynter JL, Hawkins JA, Atherton JG. Isolation of a bacterium resembling Pirellula species from primary tissue culture of the giant prawn (*Penaeus monodon*). Appl Environ Microbiol 1991;57:3127–3134.
43. Smith WL, DeWitt DL, Garavito RM. Cycloxygenases: structural, cellular, and molecular biology Annu Rev Biochem 2000;145:145–182.
44. Wong JL, Creton R, Wessel GM. The oxidative burst at fertilization is dependent upon activation of the dual oxidase Udx1. Dev Cell 2004;7:801–814.