

PeroxiBase: a database with new tools for peroxidase family classification

Dominique Koua^{1,2}, Lorenzo Cerutti¹, Laurent Falquet³, Christian J. A. Sigrist¹, Grégory Theiler², Nicolas Hulo¹ and Christophe Dunand^{2,*}

¹Swiss Institute of Bioinformatics, Swiss-Prot Group, CMU, 1 rue Michel Servet, ²Laboratory of Plant Physiology, University of Geneva, Quai Ernest-Ansermet 30, CH-1211 Geneva 4 and ³Swiss Institute of Bioinformatics, EMBnet Group, Quartier Sorge - Bâtiment Génopode, CH-1015 Lausanne, Switzerland

Received August 15, 2008; Revised September 19, 2008; Accepted September 23, 2008

ABSTRACT

Peroxidases (EC 1.11.1.x), which are encoded by small or large multigenic families, are involved in several important physiological and developmental processes. They use various peroxides as electron acceptors to catalyse a number of oxidative reactions and are present in almost all living organisms. We have created a peroxidase database (<http://peroxibase.isb-sib.ch>) that contains all identified peroxidase-encoding sequences (about 6000 sequences in 940 organisms). They are distributed between 11 superfamilies and about 60 subfamilies. All the sequences have been individually annotated and checked. PeroxiBase can be consulted using six major interlink sections 'Classes', 'Organisms', 'Cellular localisations', 'Inducers', 'Repressors' and 'Tissue types'. General documentation on peroxidases and PeroxiBase is accessible in the 'Documents' section containing 'Introduction', 'Class description', 'Publications' and 'Links'. In addition to the database, we have developed a tool to classify peroxidases based on the PROSITE profile methodology. To improve their specificity and to prevent overlaps between closely related subfamilies the profiles were built using a new strategy based on the silencing of residues. This new profile construction method and its discriminatory capacity have been tested and validated using the different peroxidase families and subfamilies present in the database. The peroxidase classification tool called PeroxiScan is accessible at the following address: <http://peroxibase.isb-sib.ch/peroxiscan.php>.

INTRODUCTION

Peroxidases are enzymes that use various peroxides (ROOH) as electron acceptors to catalyze a number of oxidative reactions. These peroxidases can be haem and non-haem proteins. They are extremely widespread and present in all living organisms. In mammals, they are implicated in biological processes as various as immune system or hormone regulation. In plants, they are involved in auxin metabolism, lignin and suberin formation, cross-linking of cell wall components, defense against pathogens or cell elongation. Humans contain more than 30 peroxidases whereas *Arabidopsis thaliana* has about 130 peroxidases that are grouped in 13 different families and nine subfamilies. There has been increased interest over the last few years in the role that mammalian haem peroxidase enzymes may play in both disease prevention and human pathologies. In general, haem peroxidases tend to promote rather than inhibit oxidative damage. Some mammalian haem peroxidases use H₂O₂ to generate more aggressive oxidants to fight intruding microorganisms (1). Peroxidase families from prokaryotic organisms, protists and fungi have been shown to promote virulence (2–5).

At the biochemical level, peroxidases can be found in the same enzyme sub-subclass E.C.1.11.1.x, donor:hydrogen-peroxide oxidoreductase (6). Currently, 15 different EC numbers have been ascribed to peroxidase: from EC 1.11.1.1 to EC 1.11.1.16 (EC 1.11.1.4 was removed) (7). Other peroxidase families with dual enzymatic domains were classified with the following numbers: EC 1.13.11.44, EC 1.14.99.1, EC 1.6.3.1 and EC 4.1.1.44 (7). The two independent EC numbers (1.11.1.9 and 1.11.1.12) both correspond to glutathione peroxidase and are based on the electron acceptor (hydrogen peroxide or lipid

*To whom correspondence should be addressed. Tel: 0033 562 193 557; Fax: 0033 562 193 502; Email: dunand@scsv.ups-tlse.fr
Present address:

Christophe Dunand, SCSV-UMR5546 CNRS/UPS, 24 Chemin de Borderouge, BP 42617 Auzeville, 31326 Castanet-Tolosan, France

peroxide, respectively). Two particular cases are also observed for numbers EC 1.11.1.2 (NADPH peroxidase) and 1.11.1.3 (fatty acid peroxidase) and no known peroxidase sequence has been assigned to NADPH peroxidase. Peroxidasins, peroxinectins, other non-animal peroxidases, Dyp-type peroxidases, hybrid ascorbate-cytochrome C peroxidases and other Class II peroxidases do not possess their own EC number and can only be classified in EC 1.11.1.7.

At the sequence level, most haem peroxidases belong to two large families, one mainly found in plants and also in bacteria and fungi (7,8), and a second found mostly in animals (but also occasionally in some fungal and bacterial species) (9,10). These two independent groups, though possessing weak sequence homology, can still be identified with a common signature (see InterPro entry IPR010255). In addition to these two large superfamilies, four smaller protein families are indexed as capable of reducing peroxide molecules with the help of haem. Catalases (Kat), which can also oxidize hydrogen peroxide (unique feature); Di-haem cytochrome C peroxidases (DiHcCp); Dyp-type peroxidases (DypPrx); and haem Haloperoxidases (HalPrx). These families display no sequence homology between each other.

Non-haem peroxidases are not evolutionarily linked and form five independent families. The largest one is the thiol peroxidase, which currently contains more than 1000 members grouped in two different subfamilies (Glutathione peroxidases and Peroxiredoxines). Alkylhydroperoxidase, non-haem haloperoxidase, manganese catalase and NADH peroxidase are the remaining other four non haem peroxidase families.

According to the phylogenetic trees these 11 major groups can be subdivided in 60 subfamilies (Figure 1). These subdivisions based on evolution describe quite well the variety of peroxidase functions and can thus be used to predict the function of newly characterized proteins.

Due to the high diversity of peroxidase functions and increased interest of the medical research in pathologies related to the role of peroxidases there is an urgent need to

federate and organize data on peroxidases. The goal of our database is to centralize most sequences that belong to peroxidase superfamilies, to follow the evolution of peroxidase among living organism and to compile the information concerning putative functions and transcriptional regulation. Currently, PeroxiBase is a unique repository exclusively dedicated to peroxidase families and superfamilies from both Eukaryotes and Prokaryotes. It includes 6000 peroxidases encoding sequences from 940 organisms, and each sequence is individually annotated. We have also developed a new tool to facilitate the classification of new peroxidase members.

DATABASE INTERFACE ORGANIZATION

The PeroxiBase toolbar is divided into eight sections (Figure 2). The ‘Documents’ tab gives access to general information: ‘Introduction’, ‘Class description’, ‘Publications’ and ‘Links’. Several useful tools are available (‘Tools’) to classify and analyse peroxidases: ‘Search’ permits complex text queries on the database, ‘Blast’ allows a comparison between a query sequence and the peroxidases stored in PeroxiBase and, ‘FingerPrintscan’ and ‘PeroxiScan’ help classify a query sequence in the right group. The six following sections named ‘Classes’, ‘Organisms’, ‘Cellular localisations’, ‘Inducers’, ‘Repressors’ and ‘Tissue types’ permit the user to navigate within PeroxiBase using the specified criteria. Individual data sheets have been largely redesigned since the previous PeroxiBase publication (Figure 2). *Last sequence changes*, *Reviewer* and *Last annotation changes* fields exhibit the date of first entry (or of last sequence modification) with name of the contributor; the name of the curator who checked the entry, and the date of the last modification in any sections with name of the contributor, respectively. In an attempt to set up a unified nomenclature (*Name* field), we introduced a simple nomenclature based on species and class acronyms. The various original appellations have been conserved as synonyms in PeroxiBase. *Class* field refers to the class the peroxidase belongs based on the new PeroxiScan tool. *Cellular localisation*, *Tissue type*,

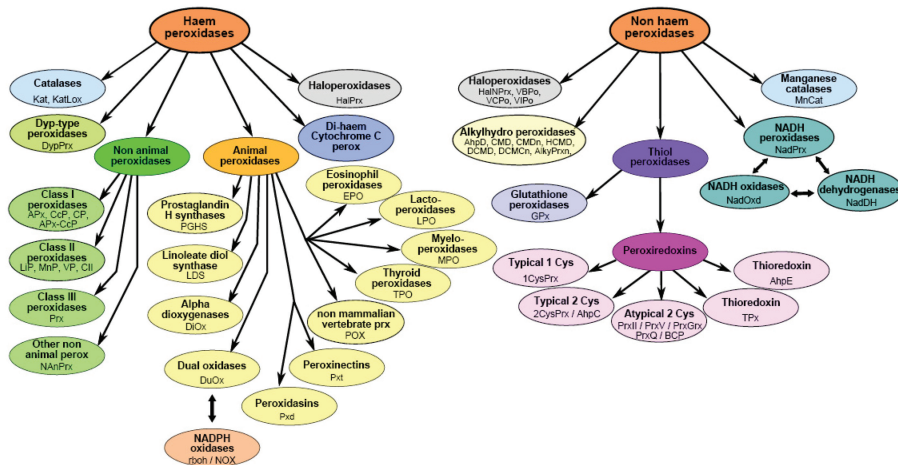


Figure 1. Schematic representation of the phylogenetic relationships between the different protein classes and families found in PeroxiBase.

Inducer and *Repressor* fields present data concerning the gene and protein expressions. These fields use fixed terms. *Best BLASTp hits* field reports the five closest hits to this entry obtained from daily updated BLAST searches. *Protein ref*, *DNA ref*, *mRNA ref* and *Cluster/prediction ref* fields refer to hyperlinks protein, DNA, mRNA sequences and cluster respectively. PeroxiBase entries are cross-referenced in UniProtKB (SwissProt/TREMBL). The data are stored in a MySQL relational database and the web interface is made of PHP and CGI/Perl scripts.

DATA ACQUISITION AND INTEGRATION

The automatic annotation of the complete genomes of numerous organisms and the automatic clustering and assembling of EST sequences led to the identification of numerous sequences coding for different peroxidase families and superfamilies. However, the automatic processing of the sequences is known to be of poor quality or not as specific as expected. Using the highly conserved motifs of each peroxidase class, manual annotation and editing can clearly identify the correct sequences even in low-quality sequences. In order to increase data reliability,

each new entry is individually controlled by a database curator. Each cross-reference is verified by the reviewer. The quality of the sequence is also examined by performing a sequence alignment with the other homologous sequences.

Thank to the continuous release of numerous genome sequencing projects (525 in March 2007 and 843 in August 2008 according to the Genomes OnLine Database (11)) and EST libraries, existing entries can be updated and, as more annotated sequences are integrated, the organism coverage is also increased. Existing entries are frequently verified and updated if any changes have occurred.

NEW CLASSIFICATION TOOLS FOR PEROXIDASES

To facilitate the classification of newly sequenced peroxidase proteins, we have developed a tool, based on PROSITE profile methodology that takes advantage of the manually curated hierarchical classification of PeroxiBase. One major problem with subfamily classification is the difficulty in separating proteins due to their high degree of similarity at the sequence level. The main principle of our new approach is to build a PROSITE profile on

PeroxiBase - The peroxidase database

Home | Documents... | Tools... | Classes | Organisms | Cellular localisations | Inducers | Repressors | Tissue types | Login

Introduction | Search | Class description | Blast | Publications | PeroxiScan | Links | FingerPrintscan

Search criteria

Organism: does contain human
 Class: does contain glutathione
 AllText: does contain
 AllText: does contain

Options

Combine criteria with: AND OR
 Include subcategories: Yes No (for Class or Taxonomic group only)

Entry info

Entry ID: 3600
 Last sequence changes: 2007-01-25 (Filippo Passardi)
 Sequence status: complete
 Reviewer: Christophe Dunand
 Last annotation changes: 2008-08-14 (Christophe Dunand)

Peroxidase information

Name (synonyms): HsGPx01-a (GSHPx-1, GPx-1)
 Class: Animal glutathione peroxidase
 Taxonomy: Fungi/Metazoa; Metazoa; Bilateria; Deuterostomia
 Organism: Homo sapiens (human) [Taxid: 9606]
 Cellular localisation: Cytosolic
 Tissue type: N/D
 Inducer: Constitutively induced
 Repressor: N/D

Best BLASTp hits

Perox	score	E-value
HsGPx01	409	1e-115
PpyGPx01	408	1e-115
MfGPx01	404	1e-114
CapGPx01	381	1e-107
SscGPx01	378	1e-106

Literature and cross-references

REFERENCE 1 Sukenaga Y, Ishida K, Takeda T, Takagi K. cDNA sequence coding for human glutathione peroxidase. Nucleic Acids Res. 15:7178-7178(1987).
 REFERENCE 2 Ishida K, Morino T, Takagi K, Sukenaga Y. Nucleotide sequence of a human gene for glutathione peroxidase. Nucleic Acids Res. 15:10051-10051(1987).
 REFERENCE 3 Mullenbach G.T., Tabrizi A., Irvine B.D., Bell G.I., Halliwell R.A. Sequence of a cDNA coding for human glutathione peroxidase confirms TGA encodes active site selenocysteine. Nucleic Acids Res. 15:5484-5484(1987).
 REFERENCE 4 Chada S., le Beau M.M., Casey L., Newburger P.E. Isolation and chromosomal localization of the human glutathione peroxidase gene. Genomics 6:268-271(1990).

Protein ref. UniProtKB: P07203
DNA ref. GenBank: Y00483
Cluster/Prediction ref. UniGene: Hs.76686

Protein sequence

Sequence length: 201 amino acids
 Sequence: MCAKRLAA...
 Retrieve as FASTA

Remarks

Complete sequence from 1079 ESTs, 19 cDNA and DNA (transcript variants HsGPx01-a and HsGPx01-b cumulated). Two TREMBL accessions P07203 and Q6NSD4 (first 57 aa are missing; fragment or splicing variant?).

Glutathione peroxidase

Contact author: Marcia Pinheiro Margis
 Last update: 2007-11-28 (Christophe Dunand)

Glutathione peroxidase from Animals (Metazoa), Insect (Metazoa), Plants (Viridiplantae), Fungi/Bacteria and other (bacteria, fungus and protist)

Profiles

Glutathione peroxidase
 Pfam: PF00255, GSHPx_1
 Prosite: P500460, GLUTATHIONE_PEROXID_1 and P500763, GLUTATHIONE_PEROXID_2
 Interpro: IPR000889, Glut_peroxidase, IPR012316, Thordox-like_Id and IPR012315, Thoredoxin_fold

Description

Glutathione peroxidase are included in haem-free thiol peroxidase family as well as PeroxiRedoxins (Prx). Both families, although totally different from their primary sequence, have in common the formation of a sulfenic acid on a catalytic cysteine during the first step of peroxidase reduction.

Glutathione peroxidase (EC 1.11.1.9 for classical glutathione peroxidase and EC 1.11.1.12 phospholipid hydroperoxide glutathione peroxidase) encompasses a family of multiple isozymes, which catalyze the reduction of H₂O₂ or organic hydroperoxides to water or corresponding alcohols using reduced glutathione.

2 glutathione + H₂O₂ → glutathione disulfide + 2 H₂O
 2 glutathione + lipid hydroperoxide → glutathione disulfide + lipid + 2 H₂O

Some of these isozymes have selenium-dependent glutathione peroxidase activity, with selenocysteine being encoded by an opal TGA codon, while others do not contain any selenocysteine (Brown et al., 2000). The animal glutathione peroxidase family is characterized by the presence of a conserved motif (GPx signature 1) containing a totally conserved cysteine or selenocysteine (GLKRR(L)N(L)I(L)VE/T(N)A(L)S(L)T(A)E(Q)A(L)Y(L)C(A)G(L)T(T)T) it is almost totally dependent on GSH for its regeneration. Other non animal 'so called' glutathione peroxidases characterized so far possess at least two conserved cysteines which form a disulfide bridge reduced by thioredoxin but not GSH. Similar motif containing one of the cysteine is still present in plants: VNWS(R)K(Q)CC, bacteria: VNITV(V)S(L)T(R)A(R)K(E)Q(A)CC and fungi: VNETV(S)K(R)H(A)K(C)G(S)A(L). The usual definition based on homology to animal glutathione peroxidase is thus not adapted anymore as these enzymes are thioredoxin-dependent peroxidases.

Two other conserved domains can be found in GPx: a glutathione peroxidase signature 2 (LAFPCNGC), and WNF(S)T(R)F that are critical sites for the catalytic activity of this enzyme (Cham et al., 1999). Photosynthetic organisms contain three conserved cysteine instead of two: the third Cys is outside the classical GPx catalytic domain (Iqbal et al., 2006). A third Cys, in similar position, can also be found in members from fungi. In mammalian tissues there are four major selenium dependent GPx isozymes: 1) classical GPx (GPx-1) which is found in red cells, liver, lung and kidney; 2) gastrointestinal GPx (GPx-2), 3) plasma GPx (GPx-3), which is present in different organs such as kidney, lung, endometrium, vas deferens, placenta, seminal vesicle, heart, and muscle, and 4) phospholipid GPx (PHGPx-4), which also present a broad distribution in different tissues. The divers GPx also have distinct subcellular locations: GPx-1 was identified in cytosol, nucleus and mitochondria; GPx-2 in cytosol and nucleus; GPx-3 is a secreted protein also found in cytosol; GPx-4 is present in nucleus, cytosol, mitochondria and bound to membranes (Herbette et al., 2007). Two other GPx isozymes, GPx-5 and 6, have been identified in mammalian tissues. (Thore et al., 2003) and are close related to GPx-3. GPx-7 compose the seventh and less characterizes group of mammalian GPxs.

In plant glutathione peroxidase protein or transcript increased in response to salt stress (Holland et al., 1993), mechanical stimulation (Degege et al., 1998), hydrogen peroxide treatment (Levine et al., 1994), and pathogen infections (Levine et al., 1994).

Literature

Brown KM, Pickard K, Nicol F, Beckett GJ, Duthie GG, Arthur JR. Effects of organic and inorganic selenium supplementation on selenoenzyme activity in blood lymphocytes, granulocytes, platelets and erythrocytes. Clin Sci (Lond). 2000 May;98(5):593-9. PMID: 10781391
 Churin Y, Schilling S, Bomer T. A gene family encoding glutathione peroxidase homologues in Hordeum vulgare (barley). A gene family encoding glutathione peroxidase homologues in Hordeum vulgare (barley). FEBS Lett. 1999 Oct 14;99(11):33-6. PMID: 10508912

Figure 2. Screenshot of one entry and description of the toolbar. The toolbar includes various sections. 'Documents' contains 'Introduction', 'Class description', 'Publications' (related to PeroxiBase) and 'Links' (specific and general databases). 'Tools' menu contains the following sections: 'Search' (multi criteria), 'Blast', 'PeroxiScan' and 'FingerPrintscan'.

Identify the class of your peroxidase sequence

ps_scan will be run with your sequence against profiles made by Dominique Koua (Master's student), Nicolas Hulo (Prosite team) and Christophe Dunand.

You may enter one raw sequence or up to 5 sequences in fasta format. Sequences longer than 2'000 amino acids will be truncated.

Query title (raw sequence only):

Paste your (fasta or raw) sequence here:

```
MCAARLAAAAQSVYAFSARPLAGGEPVSLGSLRGKVLLIENVASLCGTTVRDVTQMNELQRRLGPRGL
VVLGPRGLVVLGFPQIQFGHQENAKNEEILNSLKYVRPgggFEPNFMFLFEKCEVNGAGAHPLFAFLREALPAPSDDA
TALMTDPKLIWSPVCRNDVAWNFEKFLVGPDGVPLRRYSRRFQTDIEPDIEALLSQGPSCA
```

HsGPx01-a

10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200

PS52069: Glutathione Peroxidase Superfamily

PS52070: Animal glutathione peroxidase

PS52069: Glutathione Peroxidase Superfamily
Confidence level = '!'; Score = 31.395

```
13 SVYAFSARPLAGGEPVSLGSLRGKVLLIENVASLCGTTVRDVTQMNELQR
RLGPRGLVVLGFPQIQFGHQENAKNEEILNSLKYVRPgggFEPNFMFLFEK
CEVNGAGAHPLFAFLREALPAPSDDATAALMTDPKLIWSPVCRNDVAWNF
EKFLVGPDGVPLRRYSRR 180
```

PS52070: Animal glutathione peroxidase
Confidence level = '!'; Score = 26.405

```
13 SVYAFSARPLAGGEPVSLGSLRGKVLLIENVASLCGTTVRDVTQMNELQR
RLGPRGLVVLGFPQIQFGHQENAKNEEILNSLKYVRPgggFEPNFMFLFEK
CEVNGAGAHPLFAFLREALPAPSDDATAALMTDPKLIWSPVCRNDVAWNF
EKFLVGPDGVPLRRYSRR 180
```

[Top](#)

Documentation

Your sequence is represented by the graduated arrow, under which the significant matches between the sequence and the profiles are drawn. Two cutoffs were defined when the profile was built. '!' means the match is strong and the sequence is presumably part of the corresponding class. '?' defines a weaker match which requires further investigation.

You may analyze your sequence on [MyHits](#). It provides a better but more complex graphical representation of the match. Just press the button below. Be sure to read the documentation about [match interpretation](#).

Figure 3. The new PeroxiScan interface and result. PeroxiScan tool enables the identification of a given peroxidase sequence. PeroxiScan can be performed directly from one entry or independently from the 'Tools' section for an unknown sequence. Fine descriptions of the matching scores are available from a direct submission through MyHits web site.

the whole conserved region of each subfamily, but to make the profile more specific, residues that are conserved in the whole family are lightened and residues specific to each subfamily are emphasized. We started by merging all families that overlap to construct general alignments. In these alignments, we specifically tag well-conserved residues. The family alignments are then simply split (without modifying the alignment of residues) in several

sub-alignments according to our subfamily classification. Each subfamily alignment now contains an annotation line where residues conserved in the whole family and residues specific to the subfamily are tagged. This annotation line is then used by our profile construction program to down weigh family-conserved columns and over weight subfamily-specific ones (see http://www.expasy.org/tools/subprofiler/subprofiler_help.html for more details).

We first built profiles or used existing PROSITE profiles for the 11 major families that do not overlap. We used these profiles to build multiple sequence alignments (MSA) and integrate the PeroxiBase classification into the MSA. These 11 families were then split into 60 sub-families according to the PeroxiBase classification. For each of the subfamilies the MSA contains annotation of residues conserved in the whole family and residues specific to the subfamily. This information was used to build 60 sub-profiles specific to each subfamily, which cover all the diversity of peroxidases. During the scanning process the various sub-profiles are in competition and only the best score is reported as is done for overlapping profiles in the PROSITE database (12). This sub-profile classification allows the identification of wrongly annotated sequences in PeroxiBase and reassignment of them to their correct sub-families. It has also improved the classification of some classes of peroxidases that were difficult to distinguish with classical tools. For example the classification of the Vanadium peroxidase has been separated into three subcategories (bromoperoxidase, chloroperoxidase and iodoperoxidase). Each profile is associated to a specific function or to a biological process in order to facilitate functional classification of newly discovered proteins. New sequences can be scanned against the subfamily profile peroxidases at the following address: <http://peroxibase.isb-sib.ch/peroxiscan.php> (Figure 3). Fine descriptions of the matching residues as well as matching scores can be obtained from a direct submission through MyHits web site (<http://myhits.isb-sib.ch>) (13).

FUTURE DEVELOPMENTS

The PeroxiBase is a unique, powerful and reliable database dedicated to a large superfamily composed of several families (multigenic or not) and present in all kingdoms. The database currently contains over 6000 complete or partial peroxidase-encoding sequences distributed among 60 different protein classes. The number of peroxidase families should not undergo major changes in the future. We expect only minor modifications in the sub-classification of a few classes due to better coverage and to the biochemical characterization of the enzymes. Profiles will be updated continuously to account for such modifications, thus maintaining high quality discriminators to pursue our effort in data mining of non-annotated sequences.

Even with the large extension of the database (from 4700 in March 2007 (14) to 6026 in August 2008), it is still mainly composed of sequences originated from Viridiplantae (68%). The next step forward is to extend the coverage and to increase the number of sequences from exotic and poorly represented organisms. As the number of new sequences increases rapidly, the subsequent expansion of PeroxiBase will facilitate peroxidase gene-family studies.

Even if the manual integration of sequences is a guarantee of quality we need automatic methods to speed up the annotation of new sequences. Our classification

method will help curators to rapidly integrate new peroxidases and assign them to the correct sub-families.

To make the PeroxiBase more user-friendly to anyone who would like to add new entries or to modify present entries, a Wiki page is in development. It will surely create more collaborative interactions for the peroxidase scientific community.

ACKNOWLEDGEMENTS

We thank Filippo Passardi, Nenad Bakalovic and Vassilios Ioannidis for their efforts in the development of the PeroxiBase database, as well as the Swiss Institute of Bioinformatics for web hosting. We are also indebted to Amos Bairoch and his team for cross-referencing PeroxiBase entries in UniProt Knowledgebase, and Tania Lima for critical reading of the article.

FUNDING

Swiss National Science Foundation (31-068003.02 to C.D., 315200-116864 to L.C. and N.H.).

Conflict of interest statement. None declared.

REFERENCES

- Flohe, L. and Ursini, F. (2008) Peroxidase: a term of many meanings. *Antioxid. Redox. Signal.*, **10**, 1485–1490.
- Brenot, A., King, K.Y., Janowiak, B., Griffith, O. and Caparon, M.G. (2004) Contribution of glutathione peroxidase to the virulence of *Streptococcus pyogenes*. *Infect. Immun.*, **72**, 408–413.
- Heym, B., Stavropoulos, E., Honore, N., Domenech, P., Saint-Joanis, B., Wilson, T.M., Collins, D.M., Colston, M.J. and Cole, S.T. (1997) Effects of overexpression of the alkyl hydroperoxide reductase AhpC on the virulence and isoniazid resistance of *Mycobacterium tuberculosis*. *Infect. Immun.*, **65**, 1395–1401.
- Missall, T.A., Cherry-Harris, J.F. and Lodge, J.K. (2005) Two glutathione peroxidases in the fungal pathogen *Cryptococcus neoformans* are expressed in the presence of specific substrates. *Microbiology*, **151**, 2573–2581.
- Pineyro, M.D., Parodi-Talice, A., Arcari, T. and Robello, C. (2008) Peroxiredoxins from *Trypanosoma cruzi*: virulence factors and drug targets for treatment of Chagas disease? *Gene*, **408**, 45–50.
- Fleischmann, A., Darsow, M., Degtyarenko, K., Fleischmann, W., Boyce, S., Axelsen, K.B., Bairoch, A., Schomburg, D., Tipton, K.F. and Apweiler, R. (2004) IntEnz, the integrated relational enzyme database. *Nucleic Acids Res.*, **32**, D434–D437.
- Passardi, F., Bakalovic, N., Teixeira, F.K., Pinheiro-Margis, M., Penel, C. and Dunand, C. (2007) Prokaryotic origins of the peroxidase superfamily and organellar-mediated transmission to eukaryotes. *Genomic*, **89**, 567–579.
- Welinder, K.G. (1992) Plant peroxidases: structure-function relationships. In Penel, C., Gaspar, T. and Greppin, H. (eds), *Plant Peroxidases*, University of Geneva, Switzerland, pp. 1–24.
- Daiyasu, H. and Toh, H. (2000) Molecular evolution of the myeloperoxidase family. *J. Mol. Evol.*, **51**, 433–445.
- Furtmuller, P.G., Zederbauer, M., Jantschko, W., Helm, J., Bogner, M., Jakopitsch, C. and Obinger, C. (2006) Active site structure and catalytic mechanisms of human peroxidases. *Arch. Biochem. Biophys.*, **445**, 199–213.
- Liolios, K., Mavromatis, K., Tavernarakis, N. and Kyrpides, N.C. (2008) The Genomes On Line Database (GOLD) in 2007: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res.*, **36**, D475–D479.
- Hulo, N., Bairoch, A., Bulliard, V., Cerutti, L., Cuče, B.A., Castro, E.D., Lachaize, C., Langendijk-Genevaux, P.S. and Sigrist, C.J.A. (2008) The 20 years of PROSITE. *Nucleic Acids Res.*, **36**, D245–D249.

13. Pagni,M., Ioannidis,V., Cerutti,L., Zahn-Zabal,M., Jongeneel,C.V., Hau,J., Martin,O., Kuznetsov,D. and Falquet,L. (2007) MyHits: improvements to an interactive resource for analyzing protein sequences. *Nucleic Acids Res.*, **35**, W433–W437.
14. Passardi,F., Theiler,G., Zamocky,M., Cosio,C., Rouhier,N., Teixeira,F., Margis-Pinheiro,M., Ioannidis,V., Penel,C., Falquet,L. *et al.* (2007) PeroxiBase: the peroxidase database. *Phytochemistry*, **68**, 1605–1611.