

SCIENTIFIC REPORTS

OPEN

Fungal Hybrid B heme peroxidases – unique fusions of a heme peroxidase domain with a carbohydrate-binding domain

Marcel Zámocký^{1,2}, Štefan Janeček^{2,3} & Christian Obinger¹

Heme peroxidases, essential peroxide converting oxidoreductases are divided into four independently evolved superfamilies. Within the largest one – the peroxidase-catalase superfamily - two hybrid lineages were described recently. Whereas Hybrid A heme peroxidases represent intermediate enzymes between ascorbate peroxidases and cytochrome *c* peroxidases, Hybrid B heme peroxidases are unique fusion proteins comprised of a conserved N-terminal heme peroxidase domain and a C-terminal domain of various sugar binding motifs. So far these peculiar peroxidases are only found in the kingdom of Fungi. Here we present a phylogenetic reconstruction of the whole superfamily with focus on Hybrid B peroxidases. We analyse the domain assembly and putative structure and function of the newly discovered oligosaccharide binding domains. Two distinct carbohydrate binding modules (CBM21 and CBM34) are shown to occur in phytopathogenic ascomycetous orthologs of Hybrid B heme peroxidases only. Based on multiple sequence alignment and homology modeling the structure-function relationships are discussed with respect to physiological function. A concerted action of peroxide cleavage with specific cell-wall carbohydrate binding can support phytopathogens survival within the plant host.

Peroxidases (EC 1.11.1.1–1.11.1.19) are essential peroxide converting oxidoreductases present in all domains of life. Four heme peroxidase superfamilies (namely: peroxidase-catalase, peroxidase-cyclooxygenase, peroxidase-chlorite dismutase and peroxidase-peroxygenase) arose independently during evolution^{1,2}. They differ in overall fold, active site architecture and enzymatic activities, catalysing the hydrogen peroxide-mediated one- and two-electron oxidation of a myriad of cationic or anionic inorganic and organic molecules or even proteins (Reactions 1 and 2). Additionally, efficient dismutation of H₂O₂ can be performed by some representatives (Reaction 3).



The various physiological roles range from the degradation of hydrogen peroxide derived from aerobic life style or pathophysiological processes (Reactions 1 and 3) through H₂O₂-mediated formation of antimicrobial and halogenating oxidants (e.g. hypohalous acids, HOX, Reaction 2) to the production of radicals (HA[•], Reaction 1). Peroxidase-formed radicals are involved in either polymerization reactions^{3–5}, polymer modification⁶ or degradation reactions like plant cell wall degradation by white rot fungi⁷. The latter process recycles large amounts of carbon fixed by photosynthesis of land plants^{8,9}.

¹Department of Chemistry, Division of Biochemistry, University of Natural Resources and Life Sciences (BOKU), A-1190, Vienna, Austria. ²Institute of Molecular Biology, Slovak Academy of Sciences, SK-84551, Bratislava, Slovakia. ³Department of Biology, Faculty of Natural Sciences, University of SS Cyril and Methodius, SK-91701, Trnava, Slovakia. Correspondence and requests for materials should be addressed to M.Z. (email: marcel.zamocky@boku.ac.at)

Family #	Description new	Description old	Proteins matched	Architectures count
IPR024706	peroxiredoxins	AhpC type	37,193	13
IPR002016	peroxidase-catalase	“non-animal”	23,309	299
IPR018028	catalase typical	monofunctional	21,030	126
IPR000889	glutathione peroxidase	(the same)	17,317	72
IPR006314	peroxidase-dismutase	dyp-type	10,982	50
IPR019791	peroxidase-cyclooxygenase	“animal”	10,434	679
IPR004852	di-haem peroxidase	cyt. c peroxidase	8,522	104
IPR002065	thiol peroxidase	(the same)	7,367	12
IPR007760	manganese catalase	“pseudocatalase”	5,190	16
IPR000028	peroxidase-peroxygenase	chloroperoxidase	2,585	31

Table 1. Overview on all peroxidase and catalase families annotated in the InterPro database¹⁷ (<http://www.ebi.ac.uk/interpro/>) with numbers of matched proteins and count of various domain architectures (update June 2017).

The peroxidase-catalase superfamily is the most abundant heme peroxidase superfamily currently counting over 8,800 unique annotated members in PeroxiBase¹⁰ (<http://peroxibase.toulouse.inra.fr/>) and many more putative sequences in general databases (Table 1). This superfamily was originally named plant, fungal, and bacterial peroxidase superfamily and primarily divided in three structural classes according to a typical, rather conserved fold of their main catalytic domain¹¹. Since then, many attempts were made to analyse its phylogeny in detail^{12–15}. In 2015 we suggested to divide the superfamily in three families (instead of classes) thus providing the same systematic nomenclature as used in other superfamilies¹. Family I is comprised of (bifunctional) catalase-peroxidases, ascorbate peroxidases, cytochrome *c* peroxidases and all their evolutionary intermediates. In Family II fungal secretory peroxidases including all manganese and lignin peroxidases and their evolutionary intermediates like versatile peroxidases are found, but also numerous other peroxidases described yet as “generic” (expected to be nonlignolytic)⁹. Finally, Family III is represented by plant secretory peroxidases with hundreds of closely related genes in almost all sequenced genomes of the plant kingdom.

From an evolutionary point of view the phylogeny of the intermediates positioned between the three families is highly interesting. Important turning points of gene evolution are represented by (i) hybrid A or ascorbate-cytochrome *c* peroxidases¹⁶ and by (ii) Hybrid B heme peroxidases that were previously classified as Family I members¹³. However, recent analyses clearly demonstrated significant differences between Hybrid B peroxidases and Hybrid A or other Family I members^{2, 14, 15}. In the present study we demonstrate that Hybrid B heme peroxidases are found solely in the kingdom of Fungi and are comprised of two domains, i.e. a conserved N-terminal catalytic peroxidase domain and a C-terminal carbohydrate-binding domain with a high variability. We present the phylogeny of these fungal enzymes, discuss their domain assembly and carbohydrate sequence motifs (CBMs) as well as their putative tertiary structures derived from homology modelling. Additionally, the physiological role of these oxidoreductases is discussed.

Results and Discussion

Phylogeny of the peroxidase-catalase superfamily. The peroxidase-catalase superfamily annotated in databases as IPR002016 or PF00141 is currently represented by more than 23,300 protein sequences. As Table 1 demonstrates it represents the largest superfamily of heme peroxidases in InterPro database¹⁷ and the second largest (super)family of hydrogen peroxide reducing heme enzymes (including monofunctional catalases). Because there were recent attempts to group quite different peroxidase and catalase sequences together in a cladogram (e.g. a neighbour-joining reconstruction¹⁸) it is important to note that the heme peroxidase superfamilies summarized in Table 1 arose during genome evolution independently from each other and from non-heme peroxidases and all catalases² as explained also in PeroxiBase documentation at <http://peroxibase.toulouse.inra.fr/infos/documentation.php>.

The three families and twelve subfamilies (catalase-peroxidases, ascorbate peroxidases, ascorbate-peroxidase-related, ascorbate-peroxidase-like, cytochrome *c* peroxidases, manganese peroxidases, lignin peroxidases, versatile peroxidases, “generic” peroxidases, plant secretory peroxidases and hybrid peroxidases of type A & B) contain sequences from all domains of life. Definitely, there are now numerous novel members stemming from taxonomic lineages beyond bacteria, fungi and plants thus it is more appropriate to denominate the whole superfamily as peroxidase-catalase superfamily reflecting the dominance of peroxidase (Reactions 1 and 2) and catalase (Reaction 3) reactivities^{1, 2}. The typical mainly α -helical fold of the catalytic domain including the architecture of the heme *b* cavity remained conserved during evolution of this superfamily. On the other hand, there is a rather high sequence variability in the heme periphery, around binding sites of electron donors¹⁹ and in non-essential regions.

Here we present a detailed molecular phylogeny of 500 full length protein sequences proportionally selected from all subfamilies mentioned above. For the phylogenetic reconstruction both MrBayes inference, version 3.2.6 with invariant gamma rates using the Whelan-Goldman model²⁰ (Fig. 1), and Maximum Likelihood method based on the same Whelan-Goldman model implemented in MEGA 7 suite (Supplementary material 1) were used. It was already suggested¹⁴ that the ancient representatives of this superfamily were bifunctional catalase-peroxidases (Fig. 1). Catalytic promiscuity is often observed in ancient enzymes²¹. In the later course

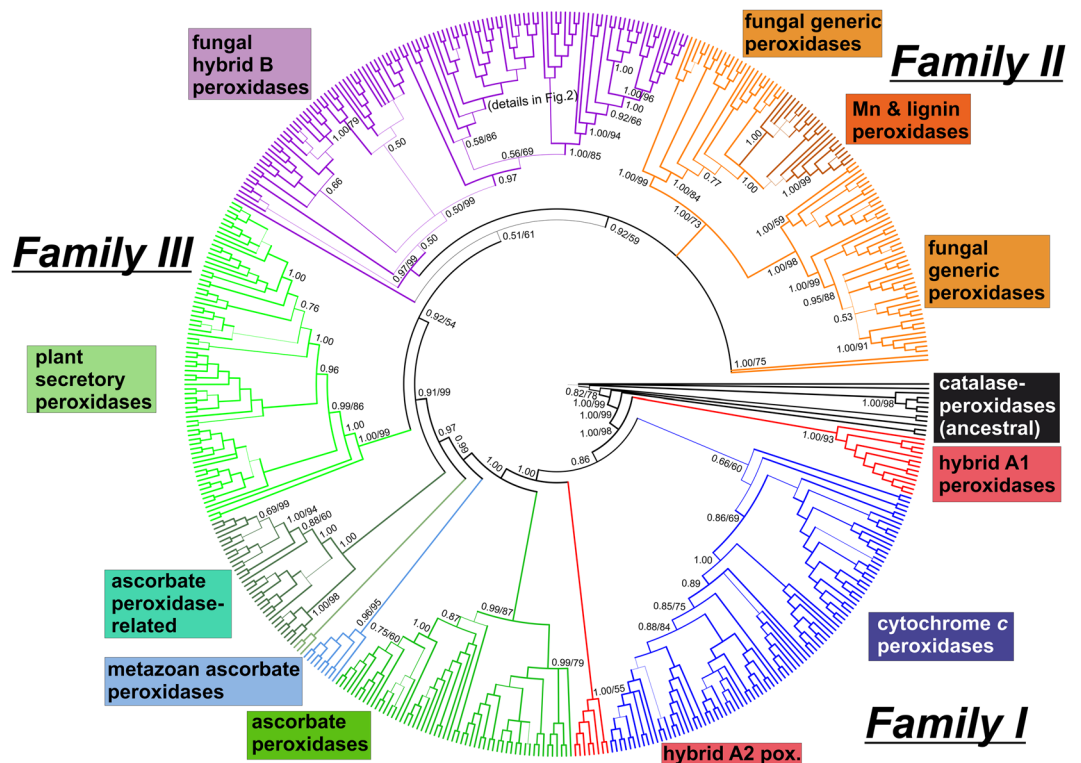


Figure 1. Global phylogenetic tree of the peroxidase-catalase superfamily. Presented is a circular tree for 500 full-length protein sequences obtained from Mr. Bayes (version 3.2) analysis with relative burn-in of 25%, WAG model of substitutions²⁰, 4 gamma categories and sampling over 3,000,000 generations. A very similar tree was obtained also with the maximum likelihood method of the MEGA suite (version 7) with 100 bootstrap replications and 4 gamma categories. Values in the (main) nodes represent posterior probabilities and bootstrap values (shown only values above 0.5/50), respectively. The thickness of the branches corresponds directly proportionally with obtained posterior probabilities. This evolutionary tree is deposited at <http://itol.embl.de/>.

of evolution bifunctional catalase-peroxidases diverged stepwise into monofunctional peroxidases with distinct substrate specificities.

Family I is currently comprised of ancient catalase-peroxidases, cytochrome *c* peroxidases, hybrid A peroxidases (abbreviated as APx-CcP), which segregated in two different main clades, and various clades of ascorbate peroxidases. A recent study focused mainly on these divergent clades of ascorbate peroxidases¹⁵. Besides “classical” Family I ascorbate peroxidases and ascorbate-peroxidase related (APx-R) genes, which segregated in two well-supported clades (Fig. 1), a new subfamily named “ascorbate-peroxidase like” (APx-L) situated on the evolutionary way from Family I towards ancestors of Families III and II was suggested¹⁵. The question remains whether APx-R and APx-L can still use ascorbate as (main) electron donor. With respect to the peroxidase domain APx-Ls might also represent pseudogenes¹⁵. Our present phylogenetic analysis shows that Metazoan (clearly non-plant) putative ascorbate peroxidases descended from the basal clades directly after the branches of “classical” intracellular ascorbate peroxidases from red & green alga and plants (Fig. 1). Thus their common ancestor was already present during the formation of primordial eukaryotic cells. It is also evident that “ascorbate peroxidase-related” proteins were segregated in distinct clades probably sooner than Family III plant secretory peroxidases (Fig. 1).

Recently, we have described the occurrence of Hybrid B peroxidases and started to analyse their phylogeny^{2, 14}. The present comprehensive phylogenetic reconstruction is mainly based on the Bayesian inference and a detailed comparative analysis of sequences. In contrast to Hybrid A peroxidases Hybrid B enzymes are strictly monophyletic (labelled violet in Fig. 1). With high statistical support the reconstruction reveals that there was a common ancestor for Family III enzymes (i.e. plant secretory peroxidases), Hybrid B peroxidases and all Family II descendants (i.e. manganese, lignin and all generic fungal secretory peroxidases) (Fig. 1). From a survey within PeroxiBase¹⁰ it can be expected that already the common ancestor of all known Hybrid B and Family II fungal peroxidases was a secretory protein. Genes for Hybrid B peroxidases can be found in the earliest diverging fungal lineage, in Chytridiomycetes (e.g. BdeHyBpox1 sequence from *Batrachomyces dendrobatis*). In contrast, there is no known sequence of a Family II representative found in a phylogenetically basal lineage of Fungi yet. Family II enzymes (generic, manganese & lignin peroxidases) occur in Dikarya (Ascomycetes & Basidiomycetes) only. Thus, Hybrid B peroxidases appear to have older roots than all Family II members but clearly more sequences from all basal fungal lineages are necessary to strengthen this hypothesis.

The monophyletic Hybrid B peroxidase subfamily with currently 114 full-ORF-length representatives can be subdivided in 9 main clades. Two of them are chytridiomycetous, three are basidiomycetous, and remaining four ascomycetous. In the well-resolved solely ascomycetous clade #7 formed by sequences from phytopathogenic

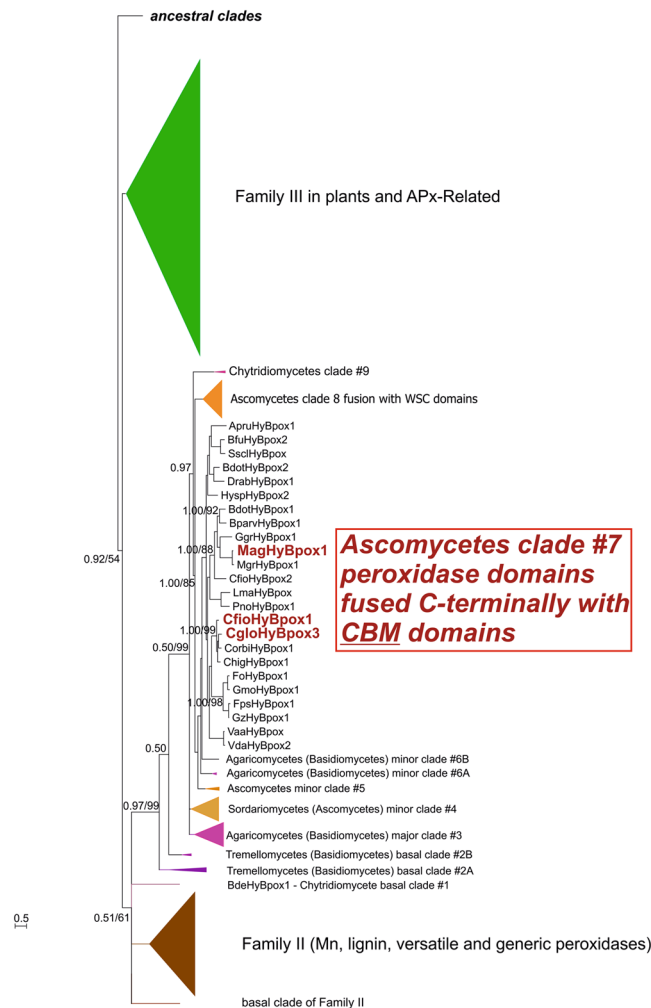


Figure 2. Detail of the evolutionary tree focused on Hybrid B peroxidases. Presented are those clades of the global peroxidase-catalase superfamily tree (Fig. 1) which show relationships among ascomycetous Hybrid B heme peroxidases possessing various carbohydrate binding domains. Values in the nodes represent posterior probability and bootstrap values (shown only above 0.5/50), respectively. Abbreviations of used peroxidase names are listed in Supplementary Table 1.

fungi (detail presented in Fig. 2) we have discovered a unique fusion of a N-terminal heme peroxidase domain with two different C-terminal carbohydrate binding motifs (CBMs) that are presented schematically in Fig. 3. For the domain architecture analysis we have selected one typical sequence of a Hybrid B peroxidase from a hemibiotrophic pathogen *Magnaporthe oryzae* causing rice and wheat blast (i.e. MagHyBpox1) and two other sequences from related hemibiotrophs (i.e. CfioHyBpox1 and CgloHyBpox3 – abbreviations explained in Supplementary Table 1). Observed two domain composition is quite different from the longer Hybrid B variants in clade #8 containing - besides a conserved peroxidase domain - at least two similar WSC domains (Fig. 3, lower part) as described previously¹⁴. The WSC domain with the InterPro accession IPR002889 (or PF01822) was formerly described also as a putative carbohydrate binding domain. Mostly, it contains up to eight conserved cysteine residues that may be involved in several disulfide bridges. However, there is currently no evidence on its ability to specifically bind carbohydrates similar to above mentioned CBMs. A detailed functional analysis revealed that WSC proteins are typically highly O-glycosylated²² and that they can serve as cell wall integrity stress sensors²³.

Domain architecture of Hybrid B peroxidases. From the multiple sequence alignment (Fig. 4 and Supplementary material 2) it is obvious that the N-terminal peroxidase domains of Hybrid B peroxidases have the same length, mainly α -helical overall fold and highly conserved heme cavity as all other members of the peroxidase-catalase superfamily. It is expected that the prosthetic group is non-covalently bound in this typical pocket that was preserved during the long evolution of this superfamily. Important invariantly conserved catalytic residues include the distal Arg/His pair, which supports the deprotonation of H_2O_2 and the heterolytic cleavage of the peroxide bond. Described Arg/His pair is part of the conserved triad Arg108 – Trp111 – His112 (BpKatG1 numbering in the upper sequence of Fig. 4A). The third amino acid in the distal triad is involved in the formation of covalent adduct only in ancestral catalase-peroxidases (KatGs) but during the evolution it was substituted mainly with phenylalanine. The latter event is reflected by the conversion of bifunctional KatGs to

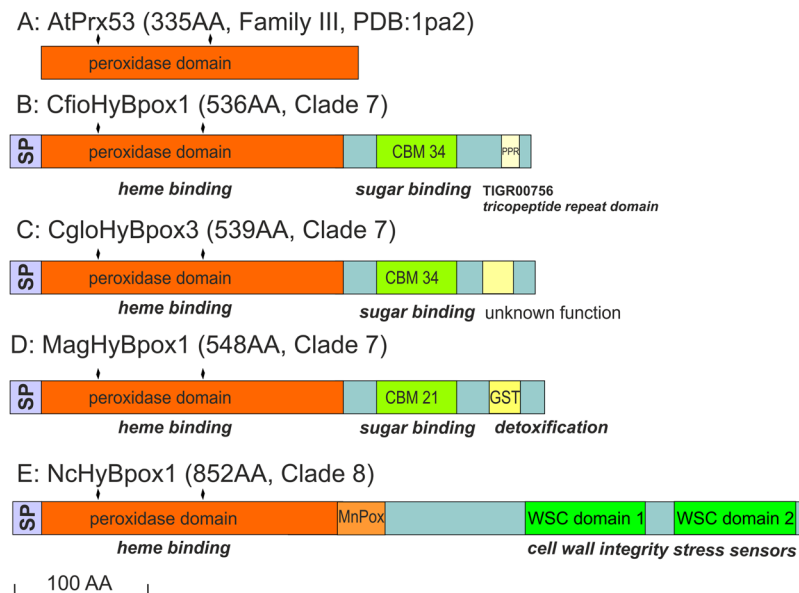


Figure 3. Schematic presentation of domain composition in selected Hybrid B heme peroxidase protein sequences. Four selected protein sequences of fungal Hybrid B peroxidases are compared with a sequence of a typical plant peroxidase. Abbreviations used: SP, signal peptide, MnPOX, motif known from a manganese peroxidase domain, CBM, carbohydrate binding module, WSC, cell wall stress-sensor component. Sequence motifs shown in yellow were identified with lower probability. Drawn to scale. Abbreviations of used peroxidase names are listed in Supplementary Table 1.

monofunctional peroxidases¹. Some rare and interesting variations within the whole superfamily are found only on the heme distal side in Hybrid B peroxidases, namely Arg69 – Tyr72 – His73 (e.g. BdotHyBpox2 numbering) or even Lys69 – Tyr72 – His73 (SsclHyBpox numbering). The latter unique variant opens the question about the role of lysine in heterolytic peroxide bond cleavage. In any case, the distal histidine is apparently invariantly conserved in all known sequences of the whole superfamily. Besides this conserved triad an invariant Asn142 (BpKatG1 numbering) occurs at the distal side, which is involved in H-bonding and modulation of the pK_a of the above mentioned catalytic His²⁴.

At the proximal side the heme ligand His279 (Fig. 4B) and its H-bonding partner Asp389 (BpKatG1 numbering, see Supplementary material 2) are fully conserved in the whole superfamily and contribute to the stabilization of the ferric resting state. Together with a Trp or Phe they constitute the proximal triad. In almost all Hybrid B peroxidases a Phe is found (Phe216, BdotHyBpox2 numbering) whereas in Family I peroxidases a Trp is located at this position. Figure 5 demonstrates the high level of structural conservation within this superfamily by comparing the crystal structures of a fungal and plant peroxidase as well as two Phyre-predicted structural models of Hybrid B-peroxidases.

Almost all members of this superfamily are one domain proteins consisting of the peroxidase domain only. Exceptions are catalase-peroxidases and Hybrid B peroxidases (and few Hybrid A members). At the basis of evolution of the whole superfamily two-domain bifunctional catalase-peroxidases are found²⁵. They have a N-terminal catalytic heme domain and a shorter gen-duplicated homologous (heme-free) domain that supports the maintenance of the overall and heme cavity architecture²⁶. The phylogenetic origin of the C-terminal domain is still under discussion¹⁵. Anyway, in course of further evolution this C-terminal domain was lost. There are rare exceptions detected only among Hybrid A peroxidases where the second domain still remained preserved²⁷. However, the vast majority of members of the various subfamilies (except Hybrid B peroxidases) consist of only a single heme peroxidase domain (overview in PeroxiBase¹⁰ <http://peroxibase.toulouse.inra.fr/>).

Structural analysis of newly discovered CBMs present in phytopathogenic Ascomycetous Hybrid B heme peroxidases. All Hybrid heme B peroxidases are fused proteins consisting of the highly conserved heme peroxidase domain and at least one non-homologous and non-catalytic C-terminal domain (Figs 3 and 6). The C-terminal fusions are comprised of either multiple WSC domains (clade #8 in Figs 2 and 3) or a single carbohydrate binding motif with additional short variable motifs with mostly unknown function (clade #7). In contrast to our preliminary analysis of the C-terminal domains¹⁴ it is now obvious that not all Hybrid B peroxidases contain a WSC domain and the variability in this region of the fused peroxidases is much higher than expected before.

Our structural analysis (Figs 7 and 8) clearly demonstrates that these CBMs present in phytopathogenic Ascomycetous peroxidases belong to CBM21 and CBM34 families. We have selected CfiHyBpox1, CgloHyBpox3 and MagHyBpox1 that revealed in the first round of screening the highest probability for the presence of CBM domains by using the CDvist suite²⁸. It has to be emphasized that both CBM21 and CBM34 belong to the so-called raw starch-binding domains (SBD) found typically as modules of various microbial amylolytic

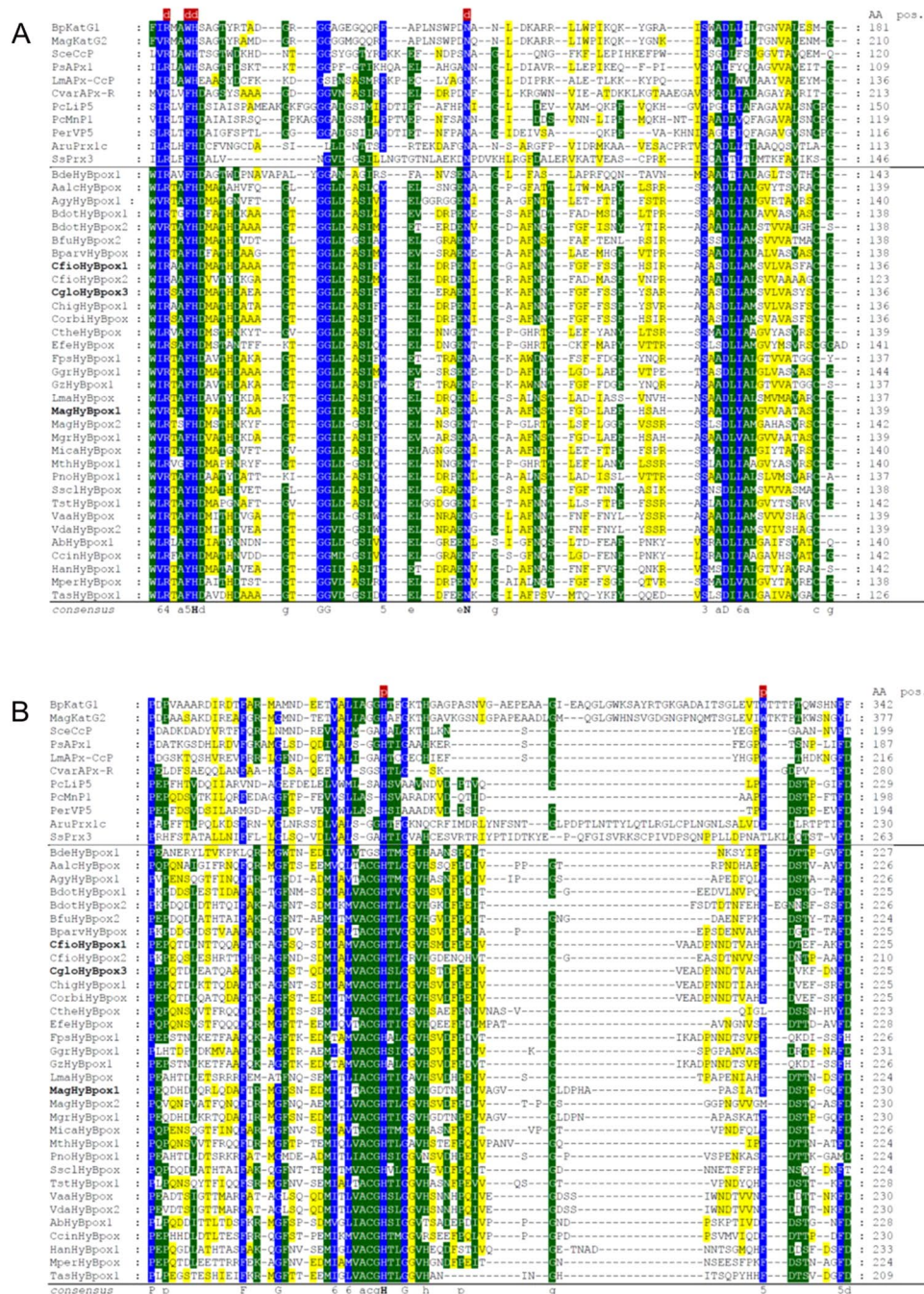


Figure 4. Multiple sequence alignment of 44 selected peroxidase-catalase superfamily members. Whole alignment in fasta format is deposited in Supplemental material 2. This part of the alignment shows highly conserved regions with essential residues of the heme peroxidase domain: (A) Region on the distal heme side and (B) region on the proximal side of the prosthetic heme group. Color scheme: blue > 90% conservation, green > 60% conservation, yellow > 25% conservation.

enzymes²⁹. Among 81 currently known CBM families there are at least 13 verified raw starch-binding domains as classified in the CAZY database (<http://www.cazy.org>)³⁰ and, indeed, some of them have already been identified in non-amyolytic enzymes²⁹. For example, CBM20 was found in the mammalian genethonin-1³¹ and laforin³² as well as in fungal lytic polysaccharide monoxygenases³³, whereas CBM48 was detected in the plant SEX4 protein³² and the β -subunit of AMP-activated protein kinase³⁴.

With regard to Hybrid B heme peroxidases the eventual presence of any CBM with assumed raw starch-binding capability is highly interesting. Despite the fact that MagHyBpox1 may contain CBM21, while both CfoHyBpox1 and CgloHyBpox3 possess CBM34, structural superimposition of their models with experimentally verified CBM21 and CBM34 templates clearly demonstrate that the overall respective folds, i.e. a typical

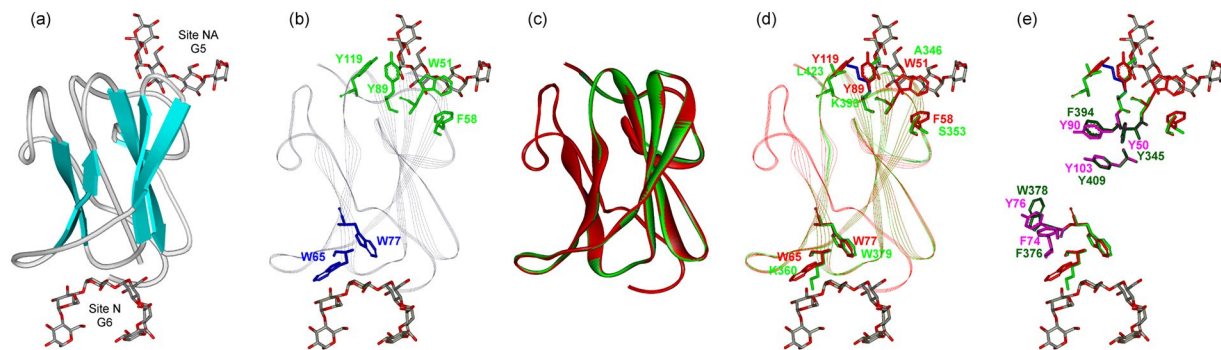


Figure 7. Structural comparison of second domain of ascomycetous Hybrid B peroxidases with starch-binding domain of the family CBM34. (a) Structure of the real family CBM34 starch-binding domain from the family GH13 α -amylase TVA-I from *Thermoactinomyces vulgaris* (PDB code: 1UH4³⁸) with two binding sites, the site N and the site NA with bound maltohexaose (G6) and maltopentaose (G5), respectively. (b) Aromatic residues responsible for saccharide binding acting in the sites N (blue) and NA (green). (c) Superimposed real CBM34 from *T. vulgaris* TVA-I (red) with CBM34 models from fungal Hybrid B heme peroxidases - CfloHyBpox1 (blue) and CgloHyBpox3 (green) - covering 87 C α -atoms with a 0.15 Å root-mean square deviation. (d) Binding residues in the two carbohydrate binding sites in the real CBM34 structure (red) and their counterparts in both fungal peroxidases CBM34-like models (green and blue); only the CgloHyBpox3 residues being labelled. (e) Emphasis on aromatic residues from the real CBM34 (magenta) that, although not involved in carbohydrate binding, have their aromatic counterparts in both CBM34-like structures from fungal Hybrid B heme peroxidases (cyan and dark green); only the CgloHyBpox3 residues being labelled.

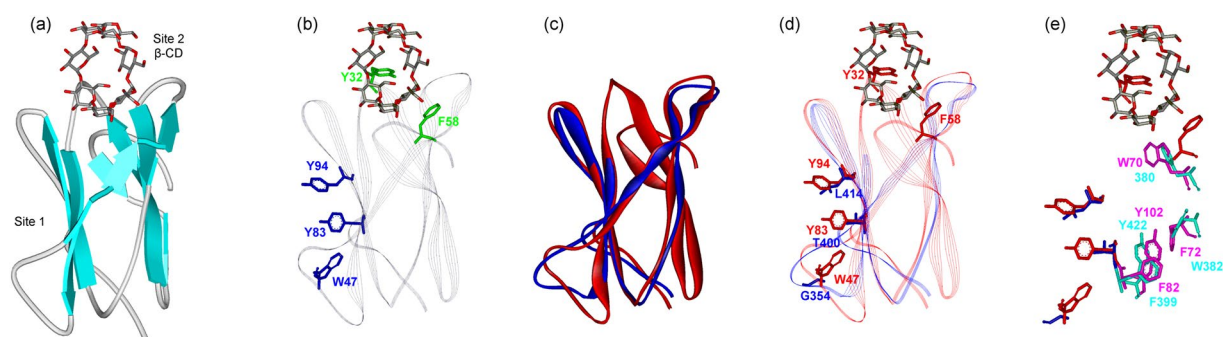


Figure 8. Structural comparison of second domain of ascomycetous Hybrid B peroxidases with starch-binding domain of the family CBM21. (a) Structure of the real family CBM21 starch-binding domain from the family GH15 glucoamylase from *Rhizopus oryzae* (PDB code: 2V8L⁴⁰) with two binding sites, the site 1 (with no saccharide bound) and the site 2 complexed with β -cyclodextrin (β -CD). (b) Aromatic residues responsible for saccharide binding acting in the sites 1 (blue) and 2 (green). (c) Superimposed real CBM21 from *R. oryzae* glucoamylase (red) with CBM21 model from fungal Hybrid B heme peroxidase MagHyBpox1 (blue) covering 64 C α -atoms with a 1.17 Å root-mean square deviation. (d) Binding residues in the two carbohydrate binding sites in the real CBM21 structure (red) and their counterparts in the site 1 of the fungal peroxidase CBM21-like model (blue); in the site 2, there were no corresponding residues in the overlap. (e) Emphasis on aromatic residues from the real CBM21 (magenta) that, although not involved in carbohydrate binding, have their aromatic counterparts in the CBM21-like structure from the fungal Hybrid B heme peroxidase (cyan).

(PDB code: 2DJM³⁵) for MagHyBpox1 (residues Asp345–Asp426) and CBM34 from *Thermoactinomyces vulgaris* α -amylase TVA-I (PDB code: 1J11³⁶) for both CfloHyBpox1 (residues Ile326–Gln426) and CgloHyBpox3 (residues Ile326–Glu426). In each case the models were selected in an effort to take into account the most appropriate combination of the three parameters confidence, sequence identity and alignment coverage.

Once the overall fold of CBM21 and CBM34 analogs in these phytopathogenic Ascomycetous peroxidases has been recognized, it was relevant to find out whether also the residues responsible for carbohydrate binding in the two CBMs have been conserved in peroxidases. In general, there is at least one, but usually two starch-binding sites in a CBM known as a starch-binding domain^{29, 37}. This is also the case of CBM34 (Fig. 7A,B) and CBM21 (Fig. 8A,B). Saccharide binding is provided mostly by aromatic residues involved in stacking interactions, but hydrogen bonds may also be involved^{29, 36–40}. Although no saccharide was seen complexed at binding site 1 (Fig. 8A,B) in the three-dimensional structure of CBM21 from the *Rhizopus oryzae* glucoamylase (PDB codes: 2DJM, 2V8L), the relevant aromatic residues are present at both binding sites⁴⁰. Comparison of saccharide binding residues from known CBM21s and CBM34s with putative CBMs from Hybrid B peroxidases (Figs 7D and

8D) shows that out of the six aromatic residues of characterized CBM34s (Fig. 7B), only Trp77 has the corresponding aromatic residue in the respective CBM34 models from CfloHyBpox1 (i.e. Trp379) and CgloHyBpox3 (Fig. 7D). The situation in putative CBM21 from MagHyBpox1 is even less convincing, i.e. out of the five aromatic residues of the two binding sites of characterized CBM21 (Fig. 8B), no corresponding aromatic amino acid was found. Moreover, the second binding site could not be identified due to incompleteness of model (Fig. 8C,D).

However, there are several other aromatic residues, although temporarily with no assigned functional role, positioned “inside” the CBM, which are found in real amylolytic starch-binding domains and the Hybrid B peroxidase models (Figs 7E and 8E). Five such residues can be seen in CBM34 of both CfloHyBpox1 and CgloHyBpox3 (Fig. 7E) and four in CBM21 of MagHyBpox1 (Fig. 8E). A similar observation has been reported for other starch-binding domains from the family CBM41^{29,41}, for which it has been hypothesised that these aromatic positions (neither totally conserved, nor functional role ascribed based on solved structures) may represent a relict from a primordial CBM ancestor before the current CBMs specialized during evolution.

A functional connection of a heme peroxidase with carbohydrate binding motifs thus observed among Hybrid B peroxidases from various important hemibiotrophic Ascomycetes might have significant impact for their phytopathogenicity. Transcripts of corresponding genes are currently detected in fungal families Magnaporthaceae & Glomerellaceae within mRNA libraries either non-induced or induced with some kind of oxidative stress (e.g. GenBank-EST database accession numbers [JZ969979.1](#), [JZ970399.1](#) or [DR621480.1](#)). The physiologically observed oxidative burst accomplished by a prompt accumulation of reactive oxygen species mainly from the action of plant host NADPH oxidases represents the main streamline of the apoplastic immunity⁴². A concerted action of peroxide bond cleavage with a specific binding on integral cell-wall carbohydrates can counteract the plant defence pathways and allow the fungal pathogen to survive within the host tissue. Concerning the taxonomy spectrum of organisms found currently in the families CBM21 or CBM34, the former can be considered a eukaryotic family with a majority of various amylases of yeast and fungal origin, whereas the latter is yet a solely prokaryotic family with an unambiguous dominance of bacterial amylolytic enzymes³⁰. To identify a homologue of CBM21, which is a typical fungal domain, among fungal Hybrid B peroxidases may thus not be so surprising, but to reveal a homologue of a typically bacterial CBM34 in a fungal hybrid B peroxidase should be of interest. Moreover, both CBM21 and CBM34 are best known as non-catalytic modules of amylolytic enzymes, which help their catalytic domains to bind and degrade raw starch or, in a wider sense, the α -glucans related to and/or derived from starch^{29,35–41}. Since, however, the residues responsible for binding the α -glucans in both CBM21 and CBM34 have not been found to be conserved in their counterparts from Hybrid B peroxidases, it is possible to expect also some changes in target bound carbohydrates, even in terms of their stereochemistry, i.e. a change to β -glucans. To determine the exact role these CBMs may play in the function of Hybrid B peroxidases represents therefore a relevant challenge for experiments on purified proteins that are already being undertaken.

Conclusion

The phylogenetic reconstruction of the peroxidase-catalase superfamily reveals three well resolved families and two distantly related polyphyletic Hybrid A (ascorbate-cytochrome *c* peroxidases) and monophyletic Hybrid B enzymes. The latter are unique fusion proteins containing a N-terminal highly conserved peroxidase domain and C-terminal domain comprised of variable carbohydrate binding motifs of two different types. The here observed unique domain fusion between a heme peroxidase and a CBM domain can open new horizons of future research exploring the physiological impact of the oligosaccharide binding domain(s) on the peroxidase function which might include hydrogen peroxide degradation during oxidative burst and/or site specific plant polymer degradation reactions in biotrophic and hemibiotrophic fungal pathogens.

Materials and Methods

Sequence data collection and multiple sequence alignment. All sequence data used for this analysis were collected from public databases. Protein sequences of herein analysed peroxidases were mainly from PeroxiBase¹⁰ at <http://peroxibase.toulouse.inra.fr>. Only in the case that a particular peroxidase sequence was not (yet) available in PeroxiBase corresponding Uniprot accession was used. All analysed peroxidase sequences are representative for the whole peroxidase-catalase superfamily divided in three main families and twelve subfamilies currently counting almost 8,800 manually annotated & curated sequences in PeroxiBase (in total already over 23,300 hits, provided mostly as automatic genomic annotation in InterPro database). Multiple sequence alignment of 500 selected full length protein sequences was performed with Muscle program⁴³ implemented in MEGA 7 package. Optimized alignment parameters were: gap open -0.8 gap extend -0.05 , hydrophobicity multiplier 0.9. Maximum of performed alignment iterations was set to 1,000. The used clustering method was UPGMB, for other interactions NJ and minimal diagonal length was set to 28. Alignment was inspected mainly for the presence of seven conserved catalytic residues on both distal and proximal sides involved in catalysis and binding of the heme prosthetic group^{19,44} and further refined in GeneDoc⁴⁵. Ambiguously aligned regions were excluded from further analysis. After inspection and refinements the final alignment used for molecular phylogeny contained 500 full length sequences from all subfamilies of the peroxidase-catalase superfamily. For bifunctional catalase-peroxidases analysed thoroughly in previous studies^{12,14} only the sequences of N-terminal domain known to bind the prosthetic heme group were used and not their gene-duplicated C-terminal (heme-free) counterpart.

Molecular phylogeny reconstruction. Molecular phylogeny was first reconstructed using the MEGA package, version 7⁴⁶. Muscle-aligned protein sequences including all sequences with currently known 3D structures were subjected to Maximum-Likelihood (ML) method of this package. Following optimised parameters were applied: 100 bootstraps, WAG model²⁰ of amino acid substitutions with four discrete gamma categories. The branch swap filter was set to very strong and the number of threads was set to 1. The branching patterns for particular subfamilies were presented with the Tree Explorer program of the MEGA 7 package in the rectangular

form. The same protein alignment of 500 peroxidase sequences was then subjected to phylogenetic reconstruction using MrBayes 3.2.6 suite⁴⁷. For calculating substitution rates the WAG model²⁰ applying invariant gamma option was used with 4 discrete gamma categories. For diagnostics a relative burn-in of 25.0% was applied. Majority consensus tree was obtained from all credible topologies sampled by MrBayes over 3,000,000 generations with finally achieved standard deviation of split frequencies below 0.09 (recommended limit 0.10). Resulting trees were displayed and annotated with Interactive tree of life (iTOL v.3⁴⁸) in a circular form with transformed branches.

Identification of introns and exons and prediction of signal sequences. Search for donor & acceptor splice sites in (mostly) putative fungal hybrid peroxidase genes was performed. For this purpose the program suite NetAspGene 1.0 of the CBS server was used (<http://www.cbs.dtu.dk/services/NetAspGene/>⁴⁹). GT-AG consensus sequence for the borders between exons and introns was present in most but not all hybrid peroxidase genes. Detailed output for each particular gene is presented in PeroxiBase¹⁰.

Putative signal sequences for protein secretion were revealed using the predictive algorithm of the program SignalP 4.1 (<http://www.cbs.dtu.dk/services/SignalP/>⁵⁰). The appropriate prediction database was chosen according to determined phylogenetic relationship of the analysed sequence. Those sequences that were found as intracellular with this approach were further subjected to subcellular localization analysis using TargetP 1.1 from the same online suite⁵⁰.

Analysis of domain assembly, sequences and tertiary structures of carbohydrate binding motifs (CBMs). CDvist²⁸ was used as a comprehensive visualization tool to delineate the presence of distinct domains in various fused proteins of the peroxidase-catalase superfamily. Following optimized parameters were used for screening: TMHMM for transmembrane prediction, HMMER3, domain split up to 5.0%, HH search 1 Pfam 75.0% cutoff, gap length 50aa, HH search 2 CDD 75.0% cutoff, gap length 50aa, HH search 3 PDB 75.0% cutoff, gap length 30aa, HH search 4 SCOP 75.0% cutoff, gap length 30aa, HH search 5 TIGR 75.0% cutoff, gap length 50aa and HHblits Uniprot with probability cutoff 60.0.

All used CBM sequences were retrieved from the UniProt knowledge database⁵¹; <http://www.uniprot.org/> and/or GenBank⁵²; <http://www.ncbi.nlm.nih.gov/genbank/>). The partial alignment, covering only the predicted CBM domain, was performed using the program Clustal-Omega available at the European Bioinformatics Institute's web-site (<http://www.ebi.ac.uk/>). In order to maximize similarities, the alignment was manually tuned taking into account previous bioinformatics studies^{37, 53–55}.

Three-dimensional structures were retrieved from the Protein Data Bank (PDB⁵⁶; <http://www.rcsb.org/pdb/>) for representatives of the individual CBM families, i.e.: CBM21 (PDB code: 2V8L⁴⁰) and CBM34 (PDB code: 1UH4³⁸). Three-dimensional models for domains without experimental 3D structure were created with the Phyre-2 server⁵⁷ (<http://www.sbg.bio.ic.ac.uk/phyre2/>) employing the “Normal” modelling mode. Obtained structures were superimposed using the program MultiProt⁵⁸ (<http://bioinfo3d.cs.tau.ac.il/MultiProt/>) and displayed with the WebLab Viewer Lite programme (Accelrys Inc.).

Accession codes. Of all peroxidases used in this work can be retrieved in PeroxiBase¹⁰ (<http://peroxibase.toulouse.inra.fr/>) and are listed in Supplementary Table 1.

References

- Zámocký, M. *et al.* Independent evolution of four heme peroxidase superfamilies. *Arch. Biochem. Biophys.* **574**, 108–119 (2015).
- Zámocký, M. *et al.* Genome sequence of the filamentous soil fungus *Chaetomium cochliodes* reveals abundance of genes for heme enzymes from all peroxidase and catalase superfamilies. *BMC Genomics* **17**, 763 (2016).
- Lee, I. K. & Yun, B. S. Peroxidase-mediated formation of the fungal polyphenol 3,14'-bihispidinyl. *J. Microbiol. Biotechnol.* **18**, 107–109 (2008).
- Mora-Pale, M. *et al.* Antimicrobial mechanism of resveratrol-trans-dihydrodimer produced from peroxidase-catalyzed oxidation of resveratrol. *Biotechnol. Bioeng.* **112**, 2417–2428 (2015).
- Ayala, M. & Torres, E. In *Heme Peroxidases*, RSC *Metallobiology Series No. 4* (eds E. Raven & B. Dunford) Ch. 13, 311–333 (The Royal Society of Chemistry, 2016).
- Hudson, C. M. *et al.* Lignin-modifying processes in the rhizosphere of arid land grasses. *Envir. Microbiol.* **17**, 4965–4978 (2015).
- Nagy, L. G. *et al.* Genetic bases of fungal white rot wood decay predicted by phylogenomic analysis of correlated gene-phenotype evolution. *Mol. Biol. Evol.* **34**, 35–44 (2017).
- Ruiz-Duenas, F. J. *et al.* Lignin-degrading peroxidases in Polyporales: an evolutionary survey based on 10 sequenced genomes. *Mycologia* **105**, 1428–1444 (2013).
- Floudas, D. *et al.* The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. *Science* **336**, 1715–1719 (2012).
- Fawall, N. *et al.* PeroxiBase: a database for large-scale evolutionary analysis of peroxidases. *Nucleic Acid Res.* **41**, D441–D444 (2013).
- Welinder, K. G. Superfamily of plant, fungal, and bacterial peroxidases. *Curr. Opin. Struct. Biology* **2**, 388–393 (1992).
- Zámocký, M. Phylogenetic relationships in class I of the superfamily of bacterial, fungal, and plant peroxidases. *Eur. J. Biochem.* **271**, 3297–3309 (2004).
- Passardi, F. *et al.* Prokaryotic origins of the non-animal peroxidase superfamily and organelle-mediated transmission to eukaryotes. *Genomics* **89**, 567–579 (2007).
- Zámocký, M., Gasselhuber, B., Furtmüller, P. G. & Obinger, C. Turning points in the evolution of peroxidase-catalase superfamily - molecular phylogeny of hybrid heme peroxidases. *Cell. Mol. Life Sci* **71**, 4681–4696 (2014).
- Lazzarotto, F., Turchetto-Zolet, A. & Margis-Pinheiro, M. Revisiting the non-animal peroxidase superfamily. *Trends in Plant Science* **20**, 807–813 (2015).
- Hugo, M. *et al.* Kinetics, subcellular localization and contribution to parasite virulence of a Trypanosoma cruzi hybrid type A heme peroxidase (TcAPx-CcP). *Proc. Natl. Acad. Sci. USA* **114**, E1326–E1335 (2017).
- Finn, R. D. *et al.* InterPro in 2017-beyond protein family and domain annotations. *Nucleic Acids Res.* **45**(D1), D190–D199 (2017).
- Mir, A. A. *et al.* Systematic characterization of the peroxidase gene family provides new insights into fungal pathogenicity in *Magnaporthe oryzae*. *Sci. Rep.* **5**, 11831 (2015).
- Gumiero, A., Murphy, E. J., Metcalfe, C. L., Moody, P. C. E. & Raven, E. L. An analysis of substrate binding interactions in the heme peroxidase enzymes: a structural perspective. *Arch. Biochem. Biophys.* **500**, 13–20 (2010).

20. Whelan, S. & Goldman, N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* **18**, 691–699 (2001).
21. Nam, H. *et al.* Network context and selection in the evolution of enzyme specificity. *Science* **337**, 1101–1104 (2012).
22. Lommel, M., Bagnat, M. & Strahl, S. Aberrant processing of the WSC family and Mid2p cell surface sensors results in cell death of *Saccharomyces cerevisiae* O-mannosylation mutants. *Mol. Cell. Biology* **24**, 46–57 (2004).
23. Straede, A. & Heinisch, J. J. Functional analyses of the extra- and intracellular domains of the yeast cell wall integrity sensors Mid2 and Wsc1. *FEBS Letters* **581**, 4495–4500 (2007).
24. Jakopitsch, C. *et al.* The catalytic role of the distal site asparagine-histidine couple in catalase-peroxidases. *Eur. J. Biochem.* **270**, 1006–1013 (2003).
25. Zámocký, M., Furtmüller, P. G. & Obinger, C. Evolution of Catalases from Bacteria to Humans. *Antioxid. Redox Signal.* **10**, 1527–1547 (2008).
26. Wang, Y. & Goodwin, D. C. Integral role of the I'-helix in the function of the „inactive“ C-terminal domain of catalase-peroxidase (KatG). *Biochim. Biophys. Acta* **1834**, 362–371 (2013).
27. Ishikawa, T. *et al.* *Euglena gracilis* ascorbate peroxidase forms an intramolecular dimeric structure: its unique molecular characterization. *Biochem. J.* **426**, 125–134 (2010).
28. Adebali, O., Ortega, D. R. & Zhulin, I. B. CDvist: a webserver for identification and visualization of conserved domains in protein sequences. *Bioinformatics* **31**, 1475–1477 (2015).
29. Janecek, S., Svensson, B. & MacGregor, E. A. Structural and evolutionary aspects of two families of non-catalytic domains present in starch and glycogen binding proteins from microbes, plants and animals. *Enzyme Microb. Technol.* **49**, 429–440 (2011).
30. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The Carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490–D495 (2014).
31. Janecek, S. A motif of a microbial starch-binding domain found in human genethonin. *Bioinformatics* **18**, 1534–1537 (2002).
32. Emanuelle, S., Brewer, M. K., Meekins, D. A. & Gentry, M. S. Unique carbohydrate binding platforms employed by the glucan phosphatases. *Cell. Mol. Life Sci* **73**, 2765–2778 (2016).
33. Vu, V. V. & Marletta, M. A. Starch-degrading polysaccharide monooxygenases. *Cell. Mol. Life Sci* **73**, 2809–2819 (2016).
34. Polekhina, G. *et al.* Structural basis for glycogen recognition by AMP-activated protein kinase. *Structure* **13**, 1453–1462 (2005).
35. Liu, Y. N., Lai, Y. T., Chou, W. I., Chang, M. D. T. & Lyu, P. C. Solution structure of family 21 carbohydrate-binding module from *Rhizopus oryzae* glucoamylase. *Biochem. J.* **403**, 21–30 (2007).
36. Kamitori, S. *et al.* Crystal structures and structural comparison of *Thermoactinomyces vulgaris* R-47 α -amylase 1 (TVAI) at 1.6 Å resolution and α -amylase 2 (TVAII) at 2.3 Å resolution. *J. Mol. Biol.* **318**, 443–453 (2002).
37. Christiansen, C. *et al.* The carbohydrate-binding module family 20 – diversity, structure, and function. *FEBS J.* **276**, 5006–5029 (2009).
38. Abe, A., Tonzuka, T., Sakano, Y. & Kamitori, S. Complex structures of *Thermoactinomyces vulgaris* R-47 alpha-Amylase 1 with Malto-oligosaccharides demonstrate the role of domain N acting as a starch-binding domain. *J. Mol. Biol.* **335**, 811–822 (2004).
39. van Bueren, A. L. & Boraston, A. B. The structural basis of α -glucan recognition by a family 41 carbohydrate-binding module from *Thermotoga maritima*. *J. Mol. Biol.* **365**, 555–560 (2007).
40. Tung, J. Y. *et al.* Crystal structures of the starch-binding domain from *Rhizopus oryzae* glucoamylase reveal a polysaccharide-binding path. *Biochem. J.* **416**, 27–36 (2008).
41. Janecek, S., Majzlova, K., Svensson, B. & MacGregor, E. A. The starch-binding domain family CBM41 – an in silico analysis of evolutionary relationships. *Proteins* **85**, 1480–1492 (2017).
42. Doehlemann, G. & Hemetsberger, C. Apoplastic immunity and its suppression by filamentous plant pathogens. *New Phytologist* **198**, 1001–1016 (2013).
43. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
44. Kwon, H., Moody, P. C. E. & Raven, E. L. in *Heme Peroxidases*, *RSC Metallobiology Series No. 4* (eds E. Raven & B. Dunford) Ch. 3, 47–59 (The Royal Society of Chemistry, 2016).
45. Nicholas, K. B. & Nicholas, H. B. Genedoc: a tool for editing and annotating multiple sequence alignments. *Distributed by the authors.* (1997).
46. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).
47. Ronquist, F. *et al.* MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Syst. Biol.* **61**, 539–542 (2012).
48. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242–W245 (2016).
49. Wang, K., Wayne-Ussery, D. & Brunak, S. Analysis and prediction of gene splice sites in four *Aspergillus* genomes. *Fungal Genet. Biol.* **46**, 14–18 (2009).
50. Emanuelsson, O., Brunak, S., von Heijne, G. & Nielsen, H. Locating proteins in the cell using TargetP, Signal P, and related tools. *Nature Protocols* **2**, 953–971 (2007).
51. UniProt Consortium. Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res.* **42**, D191–D198 (2014).
52. Benson, D. A. *et al.* GenBank. *Nucleic Acids Res.* **42**, D32–D37 (2014).
53. Svensson, B., Jespersen, H., Sierks, M. R. & MacGregor, E. A. Sequence homology between putative raw-starch binding domains from different starch-degrading enzymes. *Biochem. J.* **264**, 309–311 (1989).
54. Park, K. H. *et al.* Structure, specificity and function of cyclomalto-dextrinase, a multispecific enzyme of the α -amylase family. *Biochim. Biophys. Acta* **1478**, 165–185 (2000).
55. Machovic, M. & Janecek, S. The evolution of putative starch-binding domains. *FEBS Letters* **580**, 6349–6356 (2006).
56. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
57. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* **10**, 845–858 (2015).
58. Shatsky, M., Nussinov, R. & Wolfson, H. J. A method for simultaneous alignment of multiple protein structures. *Proteins* **56**, 143–156 (2004).

Acknowledgements

This work was supported by the Austrian Science Fund (FWF project P27474-B22), by the Slovak Research and Development Agency (grant APVV-14-0375) and by the Slovak Grant Agency VEGA (grant 2/0021/14). SJ thanks for financial support the Slovak Grant Agency VEGA for the grant No. 2/0146/17.

Author Contributions

M.Z. collected and classified all protein sequences, performed the sequence and phylogenetic analyses and discovered novel domains in fused hybrid peroxidases, S.J. analysed the carbohydrate binding domains of hybrid

B peroxidases and classified them to particular CBM families, C.O. evaluated obtained results and contributed to the discussion on hybrid peroxidases. All authors reviewed the submitted manuscript.

Additional Information

Supplementary information accompanies this paper at doi:[10.1038/s41598-017-09581-8](https://doi.org/10.1038/s41598-017-09581-8)

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017