

RESEARCH ARTICLE

# Independent Evolution of Six Families of Halogenating Enzymes

Gangming Xu\*, Bin-Gui Wang

Key Laboratory of Experimental Marine Biology, Institute of Oceanology, Chinese Academy of Sciences, Nanhai Road 7, Qingdao, 266071, People's Republic of China

\* [aericxu@gmail.com](mailto:aericxu@gmail.com)

## Abstract

Halogenated natural products are widespread in the environment, and the halogen atoms are typically vital to their bioactivities. Thus far, six families of halogenating enzymes have been identified: cofactor-free haloperoxidases (HPO), vanadium-dependent haloperoxidases (V-HPO), heme iron-dependent haloperoxidases (HI-HPO), non-heme iron-dependent halogenases (NI-HG), flavin-dependent halogenases (F-HG), and S-adenosyl-L-methionine (SAM)-dependent halogenases (S-HG). However, these halogenating enzymes with similar biological functions but distinct structures might have evolved independently. Phylogenetic and structural analyses suggest that the HPO, V-HPO, HI-HPO, NI-HG, F-HG, and S-HG enzyme families may have evolutionary relationships to the  $\alpha/\beta$  hydrolases, acid phosphatases, peroxidases, chemotaxis phosphatases, oxidoreductases, and SAM hydroxide adenosyltransferases, respectively. These halogenating enzymes have established sequence homology, structural conservation, and mechanistic features within each family. Understanding the distinct evolutionary history of these halogenating enzymes will provide further insights into the study of their catalytic mechanisms and halogenation specificity.



## OPEN ACCESS

**Citation:** Xu G, Wang B-G (2016) Independent Evolution of Six Families of Halogenating Enzymes. PLoS ONE 11(5): e0154619. doi:10.1371/journal.pone.0154619

**Editor:** Israel Silman, Weizmann Institute of Science, ISRAEL

**Received:** February 14, 2016

**Accepted:** April 17, 2016

**Published:** May 6, 2016

**Copyright:** © 2016 Xu, Wang. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** This work was supported by grants from the National Natural Science Foundation of China (31330009, 31570038 and 31200042).

**Competing Interests:** The authors have declared that no competing interests exist.

## Introduction

Halogenated compounds are widespread in nature, particularly in marine environment, and they have diverse structures and versatile bioactivities [1–3]. The halogen substituent is typically critical to the bioactivities of these compounds. There is great interest in the biosynthesis and mechanistic studies of halogenated compounds, including the antibiotics vancomycin and chloramphenicol, the antioxidants bromophenols, and the antitumor agents rebeccamycin and salinosporamides [4–6]. Previous studies have revealed that these halogenated natural products were biosynthesized by halogenating enzymes [5, 7]. A few examples of halogenated natural products catalyzed by different halogenating enzymes are shown in [S1 Fig](#). According to their cofactor dependence, six families of halogenating enzymes have been reported so far ([S1 Table](#)), including cofactor-free haloperoxidases (HPO), vanadium-dependent haloperoxidases (V-HPO), heme iron-dependent haloperoxidases (HI-HPO), non-heme iron-dependent

halogenases (NI-HG), flavin-dependent halogenases (F-HG), and S-adenosyl-L-methionine (SAM)-dependent halogenases (S-HG) [7–9].

In the cofactor-free HPO enzyme family, only five crystal structures have been determined, all from bacterial species [9]. These structures include the chloroperoxidases CPO-A1, CPO-A2, and CPO-T from *Streptomyces aureofaciens*, CPO-L from *S. lividans*, and CPO-F from *Pseudomonas fluorescens*. Although CPO-A2 has a catalytic triad, other halogenation mechanism was proposed. The cofactor-free HPO enzymes have an active site pocket, besides the specific halide-binding sites. These enzymes require organic acids, such as benzoate or propionate, as cosubstrate for their bioactivity [9, 10].

The V-HPO family has the most well-studied halogenating enzymes from algae, fungi, and bacteria [11, 12]. The V-HPO enzymes use vanadium (V) as a cofactor and catalyze the oxidation of halides (Cl, Br and I) using hydrogen peroxide ( $H_2O_2$ ). Several crystal structures of V-HPO enzymes have been reported, including V-bromoperoxidases (V-BPO), V-chloroperoxidases (V-CPO), and V-iodoperoxidases (V-IPO) [12]. The structure-function relationships of the V-HPO enzymes are relatively well understood and reviewed very recently [12, 13].

The HI-HPO family consists of heme-thiolate enzymes that catalyze halogenation reactions in the presence of  $H_2O_2$  [14]. Several crystal structures of HI-HPO enzymes with their respective substrates from fungi and mammals have been determined [14–16]. The fungal chloroperoxidase (CPO) from *Caldariomyces fumago* is involved in caldariomycin biosynthesis [14]. The HI-HPO family also includes mammalian enzymes, such as human myeloperoxidase (MPO) and animal lactoperoxidase (LPO) [15, 17]. The conserved halide-binding sites and bound substrates of some HI-HPO enzymes were also determined [14–16].

The NI-HG family is a class of highly homologous enzymes that halogenate amino acid methyl groups, using non-heme iron ( $Fe^{2+}$ ),  $O_2$  and  $\alpha$ -ketoglutarate ( $\alpha$ -KG) as cofactors [18]. Four enzymes of the NI-HG family have been well characterized [18–21]. CmaB (*P. syringae*) is involved in chlorinating the  $\gamma$ -position of L-*allo*-isoleucine during coronamic acid biosynthesis [18]. SyrB2 (*P. syringae*) is involved in the chlorination of threonine during the biosynthesis of syringomycin E [19]. CytC3 (*Streptomyces* sp.) catalyzes a double chlorination reaction during the  $\gamma,\gamma$ -dichloroaminobutyrate biosynthesis [20]. CurA-Hal (*Lyngbya majuscula*) catalyzes the cyclopropane ring formation of the curacin A [21]. After the crystal structures of SyrB2, CytC3 and CurA-Hal were determined, their halogenating mechanisms were also extensively studied [22–24].

All of the F-HG family enzymes contain a conserved flavin ( $FADH_2$ ) binding site, and six of them have crystal structures reported [8]. PyrH (*S. rugosporu*) is a regioselective tryptophan 5-halogenase during pyrroindomycin biosynthesis [25]. PrnA (*P. fluorescens*) is a tryptophan 7-halogenase with regioselective chlorination during pyrrolnitrin biosynthesis [26]. RebH (*Lechevalieria aerocolonigenes*) is also a tryptophan 7-halogenase, showing regioselective arene halogenation during rebeccamycin synthesis [27]. CmlS (*S. venezuelae*) has a covalent flavin-aspartate bond and is involved in chloramphenicol biosynthesis [28]. CndH (*Chondromyces crocatus*) is a myxobacterial chondrochloren halogenase of a new variant group [29]. PltA (*P. fluorescens*) is a  $FADH_2$ -dependent halogenase that catalyzes dichlorination during pyoluteorin biosynthesis [30].

The S-HG family contains SAM-dependent halogenating enzymes that use a nucleophilic substitution mechanism [31]. The 5'-fluoro-5'-deoxyadenosine synthase (FDAS) from *Streptomyces cattleya* is the first native fluorinating enzyme discovered that catalyzes the formation of a C-F bond using nucleophilic substitution [32, 33]. SalL (*Salinispora tropica*) is a marine bacterial SAM-dependent chlorinase that is involved in the biosynthesis of the anticancer agent salinosporamide A [34]. The fluorinase FDAS also has chlorinase activity; whereas the chlorinase SalL also utilizes bromide and iodide as substrates [34, 35].

These six families of halogenating enzymes have similar biological functions, but the various enzyme structures suggest that they might have divergent evolutionary processes. It was suggested that the V-HPO family have evolutionary relationship to the acid phosphatases [11, 12]. Based on the crystal structures and catalytic mechanisms, phylogenetic and structural analyses suggest that these halogenating enzymes may be evolutionarily related to other enzymes. The elucidation of their divergent evolutionary history will be helpful for the future investigation of their halogenating mechanisms and substrate selectivity.

## Materials and Methods

All the halogenating enzyme homologues with biological relevance were obtained using the position-specific iterated (PSI)-BLAST in the GenBank [36]. The conserved domain database (CDD) and the Pfam protein families database for putative conserved domains were also used [37, 38]. Distant homologues with similar structural macromolecular complexes were also tracked by the vector alignment search tool (VAST+) in the molecular modeling database (MMDB), and by the FUGUE 2.0 program using sequence-structure comparisons in the protein data bank (PDB) [39, 40]. The multiple sequence alignment (MSA) was aligned using the MUSCLE program in the MEGA 6.0 suite [41]. The structure-based MSA analyses were performed and visualized using the ESPript 3.0 program [42]. The predictions of protein quaternary structure were performed using the SWISS-MODEL homology-modeling server [43]. The 3D molecular structures of enzymes were visualized and prepared using the PyMOL molecular graphics system.

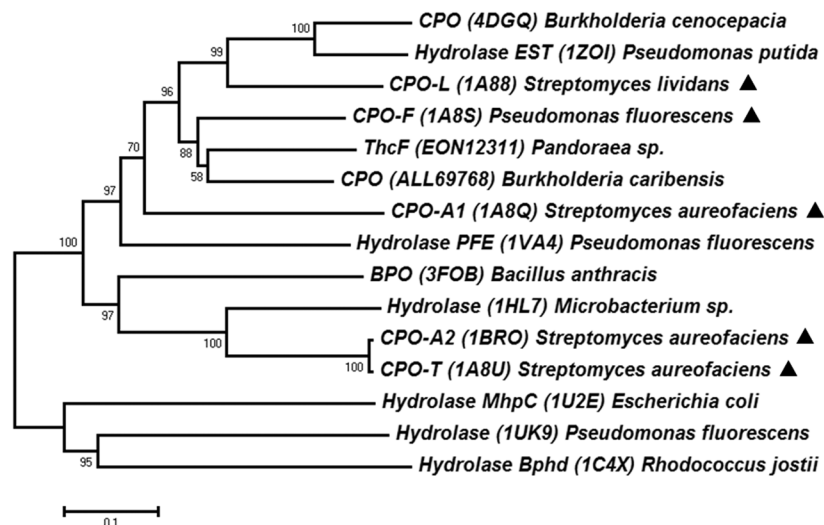
The reconstruction of phylogenetic trees was performed using neighbor-joining (NJ), minimum-evolution (ME), and maximum likelihood (ML) methods [44]. The phylogeny was tested by the bootstrap and interior-branch methods with 1000 replications. The substitution models were analyzed using the p-distance model for NJ trees and the Poisson model with gamma distributed correction for NJ, ME and ML methods in the MEGA 6.0 suite [45]. The gaps/missing data treatment is using partial deletion [41]. The ML trees were firstly to find best models and then reconstructed using the MEGA 6.0 software [45]. Similar topologies were obtained from different methods, and only the most robust phylogenetic trees from the NJ methods are shown.

## Results and Discussion

### The HPO family relationships to the $\alpha/\beta$ hydrolases

The crystal structures of five cofactor-free HPO family enzymes (CPO-A1, CPO-A2, CPO-T, CPO-L and CPO-F) from bacteria have been determined [9]. These HPO enzymes share more than 39% sequence identity. The HPO structures showed great similarity with the general topology of the  $\alpha/\beta$  hydrolase fold [9, 10]. In order to further understand their evolutionary relationships, these HPO enzyme homologues (particularly distant homologues) were obtained from the GenBank and the PDB databases. After MSA analysis, phylogenetic trees were reconstructed using various methods. As shown in Fig 1, these HPO family enzymes clustered separately, some of them with homologues of the  $\alpha/\beta$  hydrolase family. For example, CPO-A2 and CPO-T (*S. aureofaciens*) closely clustered with the hydrolase 1HL7 (*Microbacterium* sp.); and the CPO (*Burkholderia cenocepacia*) appears to be closely related to the hydrolase EST (*P. putida*) [46]. Some remote homologues of the  $\alpha/\beta$  hydrolase family clustered as a separate clade, such as the hydrolases MhpC (*Escherichia coli*) and Bphd (*Rhodococcus jostii*) [47].

Structure-based MSA analysis (S2 Fig) showed that the structures of the HPO enzymes also have the featured  $\alpha/\beta$  hydrolase fold with a catalytic triad (S98-D228-H257) in the active sites for halogenation [10]. Comparing the general scheme of the HPO enzymes and the  $\alpha/\beta$



**Fig 1. Evolutionary relationships between the cofactor-free HPO and the  $\alpha/\beta$  hydrolases.** The phylogenetic tree was reconstructed using the Neighbor-Joining method. The representative HPO enzymes are marked (▲).

doi:10.1371/journal.pone.0154619.g001

hydrolases, we found that the central  $\beta$ -sheet and six covering helices are mostly conserved. Additional conserved residues included the GYR and DRRG motifs, and also several other residues (HG, D-G-G-S and G-S-G-G) for structure arrangement and stabilization [9, 48]. The comparison between crystal structures of the HPO enzymes and the  $\alpha/\beta$  hydrolases suggest that they all have similar overall topology, which is featured with the  $\alpha/\beta$  hydrolase fold (S3 Fig). The  $\alpha/\beta$  hydrolase family also contains some esterases and lipases, which are structurally related but have diverse substrate specificity [49]. These enzymes have conserved arrangement of catalytic triad residues and structural features, and may share a common ancestor [50, 51]. The cofactor-free HPO enzymes may need organic acids as cosubstrates, such as CPO-T with benzoate and CPO-F with propionate [9]. The reaction mechanism may involve a conserved substrate-binding pocket for hydrophobic compounds halogenation at specific carbon center. These HPO enzymes may catalyze chlorination and bromination reactions, although no halogenated natural products have been proved yet [9]. Based on the phylogenetic clusters and structural similarities, the cofactor-free HPO family might have evolutionary relationships with the  $\alpha/\beta$  hydrolases, with similar substrates and bioactivities.

## The V-HPO family relationships to the acid phosphatases

Crystal structures of the V-BPO, V-CPO, and V-IPO enzymes (sequence identity 17~33%) have been determined from brown and red algae, fungi, and bacteria [13, 52–54]. The V-HPO enzymes can catalyze chlorination, bromination, and iodination reactions, with different efficiency [12]. Previous studies suggest that the V-HPO family may have relationships to the acid phosphatases [11, 12, 55]. Phylogenetic analysis of the V-HPO homologues showed that the V-HPO enzymes from different origins clustered closely together (S4 Fig). It suggests that these enzymes might be derived from a common ancestor. During the course of divergent evolution, the bacterial V-HPO enzymes may become clustered independently and have similarities to the type 2 phosphatidic acid phosphatase (PAP2) family [55]. The PAP2 superfamily includes the bacterial non-specific acid phosphatases and a variety of HPO enzymes, which may share a similar evolutionary history [12, 56].

Structure-based MSA analysis showed that the V-HPO enzymes contain many  $\alpha$ -helices, and the C-terminal has several conserved sites (S5 Fig). Despite a relatively low level of sequence identity, several short motifs in the C-terminal were still highly conserved between the V-HPO enzymes and acid phosphatases, including the cofactor vanadate coordination sites RP, Y-SGH, and R-G-H-D [12, 13]. The overall shape of the V-HPO family monomer enzyme looks like a cylinder, with a variable N-terminal helix-bundle and a conserved C-terminal helix-bundle (S6 Fig). The most conserved sites for  $\text{VO}_4$  coordination were located at the end of the C-terminal helix-bundle [12]. The non-specific acid phosphatase, such as EB-NSAP from *Escherichia blattae*, has only one helix-bundle motif, which contains 5 helices [55]. There is also great similarity in the structural fold between the V-HPO and the acid phosphatases [11, 12]. Phylogenetic analyses and structural similarities suggest that the V-HPO enzymes and the acid phosphatases are evolutionarily related, which may share a common ancestor. Gene duplication and fusion might have played a vital role during the divergent evolution of the V-HPO family [12, 13]. The significance of these events to function is that they may have similar bioactivity, due to the high similarity in structure and cofactor-binding manner [12]. For example, the V-CPO (*C. inaequalis*) was reported to exhibit phosphatase activity [57]. These all suggest that the V-HPO family may have evolutionary and functional relationships to the acid phosphatases [11, 12].

### The HI-HPO family relationships to the peroxidases

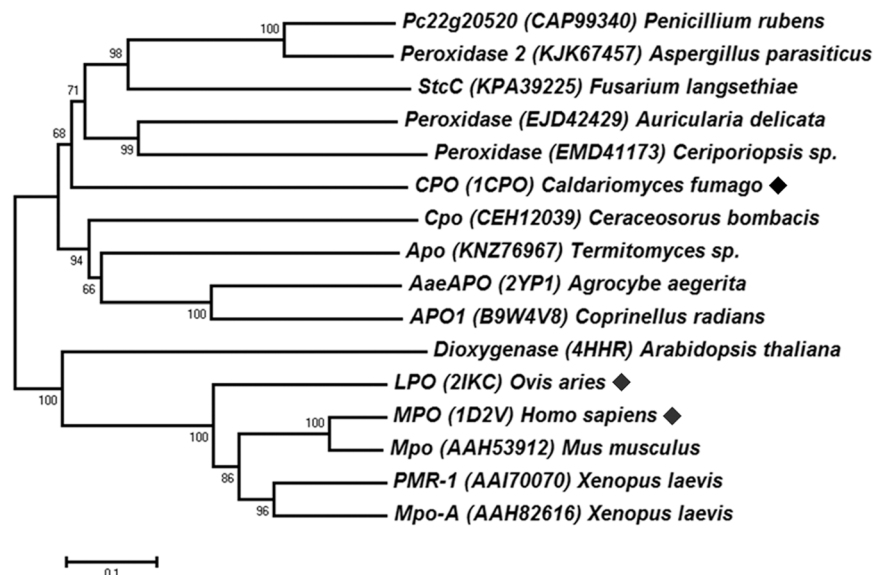
Only one enzyme crystal structure of the HI-HPO family has been determined from fungi, the fungal CPO from *C. fumago* [14, 58]. Phylogenetic trees were reconstructed after MSA analysis of CPO homologues, and the NJ tree is shown in Fig 2. The CPO (*C. fumago*) was closely clustered with other peroxidase family members. The peroxidase family 2 also includes cytochrome P450-like oxygenases; whereas the CPO was proved to be a functional hybrid of peroxidase-P450 [58]. The relatively close homologue (sequence identity 24%) is the fungal enzyme AaeAPO from *Agrocybe aegerita*. AaeAPO was the first reported aromatic peroxygenase, which also represents an evolutionary link between heme peroxidases and P450 [16]. Other distant homologues clustered separately, including the mammalian heme peroxidases MPO and LPO [15, 17]. The HI-HPO enzymes could oxidize the halides with a lower electronegativity. For example, the CPO may also accept bromide and iodide [7].

Structure-based MSA analysis of the HI-HPO and peroxidases (S7 Fig) suggest that the highly conserved regions include: the proximal heme-binding motif (D-R-PCP-N-LA-H) at the N-terminal, the E-D-S motif for heme propionates in the middle, and the acid-base catalyst E residue in the distal heme pocket [59, 60]. The specific residues in the substrate binding pocket of these enzymes might be variable to ensure their catalytic specificity [14, 16]. Comparison of the overall structures between the HI-HPO and the peroxidases also showed some structural similarities (S8 Fig). These enzymes are rich in  $\alpha$ -helices, particularly in the highly conserved core heme-binding site and the halide-binding pocket [16]. Moreover, their catalytic properties are also somewhat similar. For example, the typical CPO (*C. fumago*) has catalase and P450-like monooxygenase activity; whereas the predominant peroxygenase AaeAPO (*A. aegerita*) also displayed weak bromoperoxidase activity [14, 16, 61]. These indicate that there might be evolutionary and functional relationships between the HI-HPO enzymes and the peroxidases.

### The NI-HG family relationships to the chemotaxis phosphatases

Four chlorinases of the NI-HG family have been characterized, including CmaB (*P. syringae*), SyrB2 (*P. syringae*), CytC3 (*Streptomyces* sp.), and CurA-Hal (*Lyngbya majuscula*), which

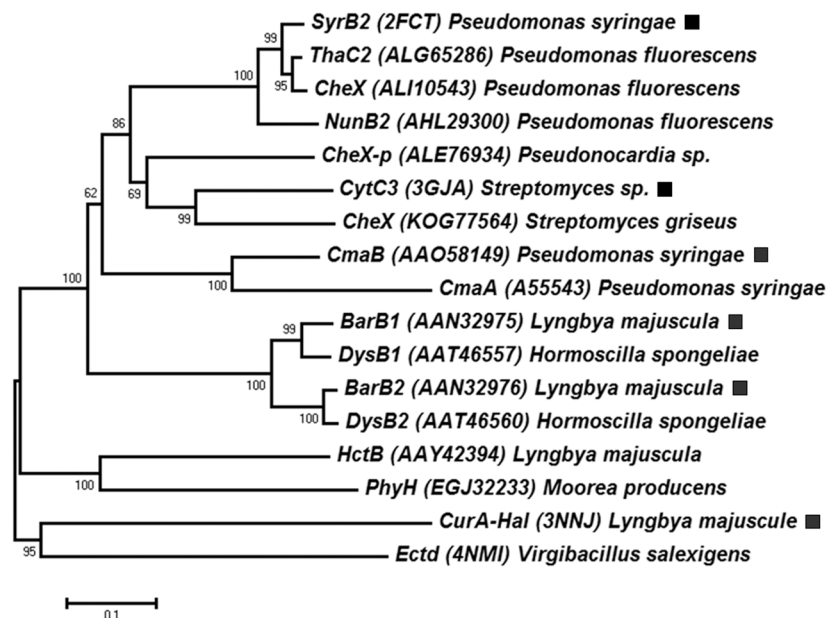




**Fig 2. Evolutionary relationships between the HI-HPO and the peroxidases.** The phylogenetic tree was reconstructed using the Neighbor-Joining method. The representative HI-HPO enzymes are marked (◆).

doi:10.1371/journal.pone.0154619.g002

could also catalyze bromination [18–20, 24]. These NI-HG enzymes share about 20~57% sequence identity. Phylogenetic trees were reconstructed using these protein homologues, and the NJ tree is shown in Fig 3. The NI-HG homologues also include BarB1 and BarB2, which is involved in a trichlorination during the biosynthesis of barbamide in the cyanobacterium



**Fig 3. Evolutionary relationships between the NI-HG and the chemotaxis phosphatases.** The phylogenetic tree was reconstructed using the Neighbor-Joining method. The representative NI-HG enzymes are marked (■).

doi:10.1371/journal.pone.0154619.g003

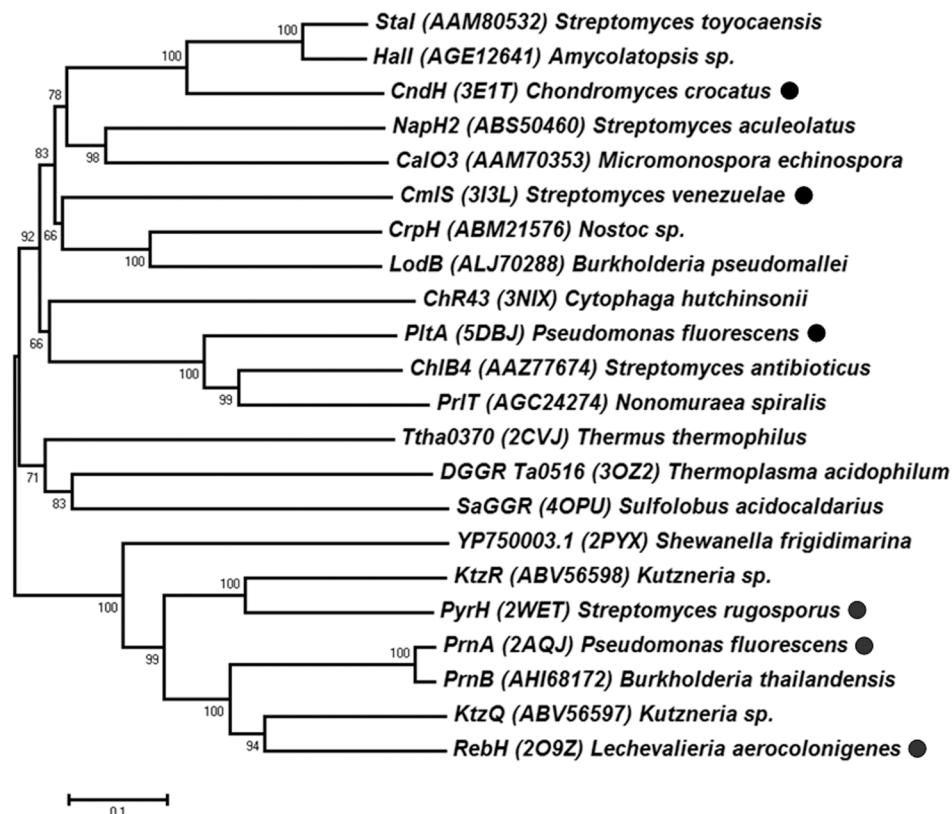
*Lyngbya majuscula* [62]. Surprisingly, the BlastP search of the SyrB2 and CytC3 homologues also revealed many chemotaxis proteins CheX in the GenBank database. The CheX protein is one of the phosphatases in the chemotaxis signal transduction system of bacteria [63, 64]. Some of the CheX coding genes (such as ALE76934) are located on plasmids, which suggest that horizontal gene transfer might play a role during the evolutionary process.

Structure-based MSA analysis (S9 Fig) showed that the NI-HG family has conserved motifs (HLD-H and HQA-H) for iron coordination, and also contains conserved active site residues (H, F, R and S) in a hydrophobic pocket for chloride binding [19]. The overall structures of the NI-HG enzymes are slightly different from the chemotaxis protein CheX; however, the conserved fold might be similar in these distant homologues (S10 Fig). Within the core structure of the NI-HG enzymes is a  $\beta$ -sandwich ‘jelly-roll’ motif (cupin fold) composed of eight anti-parallel strands [19, 20]. The chemotaxis proteins of the CheX family were reported to have a conserved (E-xx-N) motif; whereas only crystal structures of CheX in *Thermotoga maritima* and *Borrelia burgdorferi* have been determined [63, 64]. No other close homologue between the two families has been reported, and therefore more experimental data regarding the catalytic mechanisms and evolutionary relationships are needed.

## The F-HG family relationships to the oxidoreductases

The F-HG family enzymes all contain a conserved flavin-binding fold [8]. Thus far, six halogenase crystal structures have been determined, including PyrH (*S. rugosporu*), PrnA (*P. fluorescens*), RebH (*L. aerocolonigenes*), CmlS (*S. venezuelae*), CndH (*C. crocatus*), and PltA (*P. fluorescens*) [8, 25–29]. These F-HG enzymes share about 20–56% sequence identity. Phylogenetic trees were reconstructed using the close and distant homologues of the F-HG enzymes. As shown in Fig 4, the F-HG family enzymes form two main subgroups. The PyrH, PrnA, and RebH were categorized as tryptophan halogenases; the CmlS, CndH, and PltA were categorized as non-tryptophan halogenases. The F-HG enzymes were classified based on their halogenation substrates: variant A enzymes utilize free small molecules like tryptophan; variant B enzymes catalyze substrates that bind as a thioester [8, 28]. The tryptophan halogenases clustered closely together on a separate branch; whereas the non-tryptophan halogenases clustered on different sub-branches. These differences in clustering may be due to the substrate specificity of these enzymes. There might be no clear boundary between the two variants [8]. All of the F-HG enzymes reported are chlorinases, which can also function as brominases [7, 65, 66]. Moreover, distant homologues of the F-HG enzymes may also include some members of the glutathione reductase (GR) superfamily [8, 67].

Structure-based MSA analysis (S11 Fig) showed that the F-HG family enzymes have a conserved Rossmannoid-fold FAD-binding domain, consisting of a  $\beta$ -sheet flanked by helices. All the F-HG enzymes contain an N-terminal G-box (GxGxxG) motif, which is also conserved in the GR superfamily [8, 67]. There is also a conserved WxWxIP motif in the C-terminal, and several G residues are conserved for hydrogen bonding with FAD. The C-terminal domain is variable, and contains the active site residues responsible for specific substrate binding [8, 28]. Interestingly, the VAST+ search for structurally similar homologues of the F-HG family enzymes identified several oxidoreductases (sequence identity > 20%), including ChR43 (*Cytophaga hutchinsonii*), SaGGR (*Sulfolobus acidocaldarius*), Ta0516 (*Thermoplasma acidophilum*), and Ttha0370 (*Thermus thermophilus*) [8]. These enzymes have very similar overall structures, although the C-terminal domains are different (S12 Fig). Phylogenetic and structural analyses suggest that there might be evolutionary relationships between the F-HG family and the oxidoreductases.



**Fig 4. Evolutionary relationships between the F-HG and the oxidoreductases.** The phylogenetic tree was reconstructed using the Neighbor-Joining method. The representative F-HG enzymes are marked (●).

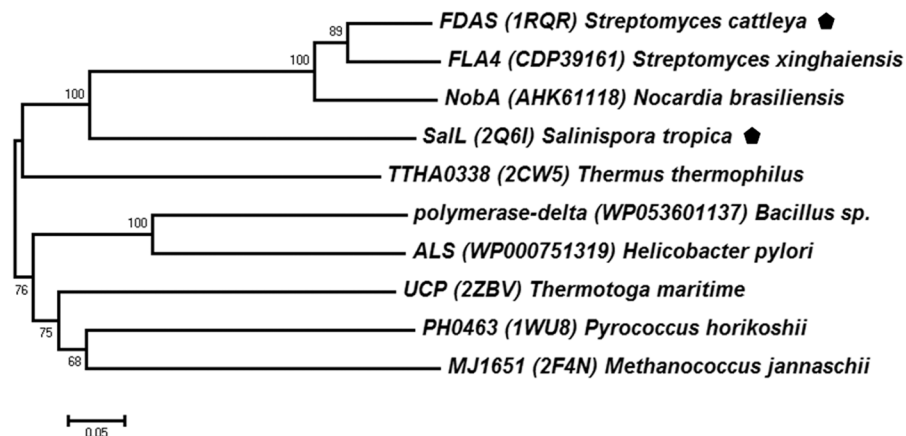
doi:10.1371/journal.pone.0154619.g004

## The S-HG family relationships to the SAM hydroxide adenosyltransferases

Only two crystal structures of the S-HG enzymes have been reported, the fluorinase FDAS (*Streptomyces cattleya*) and the chlorinase SalL (*Salinispora tropica*), which share 38% sequence identity [32, 34]. Homologues of these two S-HG enzymes were searched in the GenBank database, and only a few amino acid sequences were identified (Fig 5). One close homologue is NobA (*Nocardia brasiliensis*), which has 79% and 36% sequence identity to FDAS and SalL, respectively. Interestingly, several distant homologues of the SAM hydroxide adenosyltransferases (PF01887, previously known as duf-62 family) were also found, such as PH0463 (*Pyrococcus horikoshii*) and MJ1651 (*Methanococcus jannaschii*) [68, 69]. The SAM hydroxide adenosyltransferases catalyze a hydrolytic cleavage of SAM to generate adenosine and L-methionine. Similar  $S_N2$  nucleophilic substitution was employed, with the replacement of the halide ion by water or hydroxide [31].

Structure-based MSA analysis (S13 Fig) showed that two motifs (PxNGL and IDxxFGN), and several active site residues (such as D16 and N215) are conserved. The conserved residues in the binding pocket might form hydrogen bonds with substrate [31]. However, the key residues for the halide co-ordination are different. The S-HG family enzymes use T80-S158 in FDAS and Y70-G131 in SalL [32, 34]. The SAM hydroxide adenosyltransferases utilize a conserved amino acid triad (D68-R75-H127), which may activate water to hydroxide ions [68]. The variable organization of the active site may ensure their substrate specificity and reaction in different manners. These enzymes also have very similar tertiary structures, which are asymmetric homotrimer [31]. Each monomer unit has two domains: the N-terminal  $\alpha/\beta/\alpha$





**Fig 5. Evolutionary relationships between the S-HG and the SAM hydroxide adenosyltransferases.** The phylogenetic tree was reconstructed using the Neighbor-Joining method. The representative S-HG enzymes are marked (pentagon).

doi:10.1371/journal.pone.0154619.g005

sandwich domain and the C-terminal anti-parallel  $\beta$ -sheet domain (S14 Fig). However, these enzymes may have different biological functions. The fluorinase FADS could also act as a chlorinase [35]. The chlorinase SalL also accepts  $\text{Br}^-$  and  $\text{I}^-$ , but not  $\text{F}^-$  [34]. The SAM hydroxide adenosyltransferases utilize water/hydroxide, but no halogenation activity [68]. Therefore, phylogenetic and structural analyses suggest the S-HG enzymes might be evolutionarily related to the SAM hydroxide adenosyltransferases.

In summary, phylogenetic and structural analyses suggest that there might be evolutionary and functional relationships between the HPO and the  $\alpha/\beta$  hydrolases, the V-HPO and the acid phosphatases, the HI-HPO and the peroxidases, the NI-HG and the chemotaxis phosphatases, the F-HG and the oxidoreductases, and the S-HG and the SAM hydroxide adenosyltransferases, respectively. These enzymes have established conserved sequence, structural, and mechanistic features within each family. As summarized in the S1 Fig, by comparison of the halogenated natural products and the halogenating enzymes, and also considering the phylogenetic clusters of each family, it is possible to map different halogen specificities to different families. Understanding the distinct evolutionary process might be helpful for the study of their biological function and halogenation specificity.

## Supporting Information

**S1 Fig. Examples of halogenated natural products with their proposed specific halogenating enzymes.**

(PDF)

**S2 Fig. Multiple sequence alignment of the cofactor-free HPO and the  $\alpha/\beta$  hydrolases.**

(PDF)

**S3 Fig. Structure comparison of the cofactor-free HPO and the  $\alpha/\beta$  hydrolases.**

(PDF)

**S4 Fig. Evolutionary relationships between the V-HPO and the acid phosphatases.**

(PDF)

**S5 Fig. Multiple sequence alignment of the V-HPO and the acid phosphatases.**

(PDF)

**S6 Fig. Structure comparison of the V-HPO and the acid phosphatase.**  
(PDF)

**S7 Fig. Multiple sequence alignment of the HI-HPO and the peroxidases.**  
(PDF)

**S8 Fig. Structure comparison of the HI-HPO and the peroxidases.**  
(PDF)

**S9 Fig. Multiple sequence alignment of the NI-HG and the chemotaxis phosphatases.**  
(PDF)

**S10 Fig. Structure comparison of the NI-HG and the chemotaxis phosphatase.**  
(PDF)

**S11 Fig. Multiple sequence alignment of the F-HG and the oxidoreductases.**  
(PDF)

**S12 Fig. Structure comparison of the F-HG and the oxidoreductases.**  
(PDF)

**S13 Fig. Multiple sequence alignment of the S-HG and the SAM hydroxide adenosyltransferases.**  
(PDF)

**S14 Fig. Structure comparison of the S-HG and the SAM hydroxide adenosyltransferases.**  
(PDF)

**S1 Table. Six families of halogenating enzymes.**  
(PDF)

## Author Contributions

Conceived and designed the experiments: GX. Performed the experiments: GX. Analyzed the data: GX. Contributed reagents/materials/analysis tools: GX BW. Wrote the paper: GX BW.

## References

1. Gribble GW. The diversity of naturally produced organohalogens. *Chemosphere*. 2003; 52(2):289–97. PMID: [12738253](#)
2. Cabrita MT, Vale C, Rauter AP. Halogenated compounds from marine algae. *Mar Drugs*. 2010; 8(8):2301–17. doi: [10.3390/md8082301](#) PMID: [20948909](#)
3. Wang BG, Gloer JB, Ji NY, Zhao JC. Halogenated organic molecules of Rhodmelaceae origin: chemistry and biology. *Chem Rev*. 2013; 113(5):3632–85. doi: [10.1021/cr9002215](#) PMID: [23448097](#)
4. Vaillancourt FH, Yeh E, Vosburg DA, Garneau-Tsodikova S, Walsh CT. Nature's inventory of halogenation catalysts: oxidative strategies predominate. *Chem Rev*. 2006; 106(8):3364–78. PMID: [16895332](#)
5. Butler A, Sandy M. Mechanistic considerations of halogenating enzymes. *Nature*. 2009; 460(7257):848–54. doi: [10.1038/nature08303](#) PMID: [19675645](#)
6. van Pee KH. Microbial biosynthesis of halometabolites. *Arch Microbiol*. 2001; 175(4):250–8. PMID: [11382220](#)
7. Blasiak LC, Drennan CL. Structural perspective on enzymatic halogenation. *Acc Chem Res*. 2009; 42(1):147–55. doi: [10.1021/ar800088r](#) PMID: [18774824](#)
8. Pang AH, Garneau-Tsodikova S, Tsodikov OV. Crystal structure of halogenase PltA from the pyoluteorin biosynthetic pathway. *J Struct Biol*. 2015; 192(3):349–57. doi: [10.1016/j.jsb.2015.09.013](#) PMID: [26416533](#)
9. Hofmann B, Tolzer S, Pelletier I, Altenbuchner J, van Pee KH, Hecht HJ. Structural investigation of the cofactor-free chloroperoxidases. *J Mol Biol*. 1998; 279(4):889–900. PMID: [9642069](#)

10. Hecht HJ, Sobek H, Haag T, Pfeifer O, Vanpee KH. The metal-ion-free oxidoreductase from *Streptomyces aureofaciens* has an  $\alpha/\beta$  hydrolase fold. *Nat Struct Biol*. 1994; 1(8):532–7. PMID: [7664081](#)
11. Winter JM, Moore BS. Exploring the chemistry and biology of vanadium-dependent haloperoxidases. *J Biol Chem*. 2009; 284(28):18577–81. doi: [10.1074/jbc.R109.001602](#) PMID: [19363038](#)
12. Leblanc C, Vilter H, Fournier JB, Delage L, Potin P, Rebuffet E, et al. Vanadium haloperoxidases: From the discovery 30 years ago to X-ray crystallographic and V K-edge absorption spectroscopic studies. *Coord Chem Rev*. 2015; 301:134–46.
13. Fournier JB, Rebuffet E, Delage L, Grijol R, Meslet-Cladiere L, Rzonca J, et al. The vanadium iodoperoxidase from the marine *Flavobacteriaceae* species *Zobellia galactanivorans* reveals novel molecular and evolutionary features of halide specificity in the vanadium haloperoxidase enzyme family. *Appl Environ Microbiol*. 2014; 80(24):7561–73. doi: [10.1128/AEM.02430-14](#) PMID: [25261522](#)
14. Kuhnel K, Blankenfeldt W, Turner J, Schlichting I. Crystal structures of chloroperoxidase with its bound substrates and complexed with formate, acetate, and nitrate. *J Biol Chem*. 2006; 281(33):23990–8. PMID: [16790441](#)
15. Fiedler TJ, Davey CA, Fenna RE. X-ray crystal structure and characterization of halide-binding sites of human myeloperoxidase at 1.8 Å resolution. *J Biol Chem*. 2000; 275(16):11964–71. PMID: [10766826](#)
16. Piontek K, Strittmatter E, Ullrich R, Grobe G, Pecyna MJ, Kluge M, et al. Structural basis of substrate conversion in a new aromatic peroxigenase: cytochrome P450 functionality with benefits. *J Biol Chem*. 2013; 288(48):34767–76. doi: [10.1074/jbc.M113.514521](#) PMID: [24126915](#)
17. Loughran NB, O'Connor B, O'Fagain C, O'Connell MJ. The phylogeny of the mammalian heme peroxidases and the evolution of their diverse functions. *BMC Evol Biol*. 2008; 8:101. doi: [10.1186/1471-2148-8-101](#) PMID: [18371223](#)
18. Vaillancourt FH, Yeh E, Vosburg DA, O'Connor SE, Walsh CT. Cryptic chlorination by a non-haem iron enzyme during cyclopropyl amino acid biosynthesis. *Nature*. 2005; 436(7054):1191–4. PMID: [16121186](#)
19. Blasiak LC, Vaillancourt FH, Walsh CT, Drennan CL. Crystal structure of the non-haem iron halogenase SyrB2 in syringomycin biosynthesis. *Nature*. 2006; 440(7082):368–71. PMID: [16541079](#)
20. Wong C, Fujimori DG, Walsh CT, Drennan CL. Structural analysis of an open active site conformation of nonheme iron halogenase CytC3. *J Am Chem Soc*. 2009; 131(13):4872–9. doi: [10.1021/ja8097355](#) PMID: [19281171](#)
21. Gu L, Wang B, Kulkarni A, Geders TW, Grindberg RV, Gerwick L, et al. Metamorphic enzyme assembly in polyketide diversification. *Nature*. 2009; 459(7247):731–5. doi: [10.1038/nature07870](#) PMID: [19494914](#)
22. Wong SD, Srncic M, Matthews ML, Liu LV, Kwak Y, Park K, et al. Elucidation of the Fe(IV) = O intermediate in the catalytic cycle of the halogenase SyrB2. *Nature*. 2013; 499(7458):320–3. doi: [10.1038/nature12304](#) PMID: [23868262](#)
23. Matthews ML, Neumann CS, Miles LA, Grove TL, Booker SJ, Krebs C, et al. Substrate positioning controls the partition between halogenation and hydroxylation in the aliphatic halogenase, SyrB2. *Proc Natl Acad Sci U S A*. 2009; 106(42):17723–8. doi: [10.1073/pnas.0909649106](#) PMID: [19815524](#)
24. Khare D, Wang B, Gu L, Razelun J, Sherman DH, Gerwick WH, et al. Conformational switch triggered by  $\alpha$ -ketoglutarate in a halogenase of curacin A biosynthesis. *Proc Natl Acad Sci U S A*. 2010; 107(32):14099–104. doi: [10.1073/pnas.1006738107](#) PMID: [20660778](#)
25. Zehner S, Kotzsch A, Bister B, Sussmuth RD, Mendez C, Salas JA, et al. A regioselective tryptophan 5-halogenase is involved in pyrroindomycin biosynthesis in *Streptomyces rugosporus* LL-42D005. *Chem Biol*. 2005; 12(4):445–52. PMID: [15850981](#)
26. Dong C, Flecks S, Unversucht S, Haupt C, van Pee KH, Naismith JH. Tryptophan 7-halogenase (PrnA) structure suggests a mechanism for regioselective chlorination. *Science*. 2005; 309(5744):2216–9. PMID: [16195462](#)
27. Bitto E, Huang Y, Bingman CA, Singh S, Thorson JS, Phillips GN Jr. The structure of flavin-dependent tryptophan 7-halogenase RebH. *Proteins*. 2008; 70(1):289–93. PMID: [17876823](#)
28. Podzelinska K, Latimer R, Bhattacharya A, Vining LC, Zechel DL, Jia Z. Chloramphenicol biosynthesis: the structure of CmlS, a flavin-dependent halogenase showing a covalent flavin-aspartate bond. *J Mol Biol*. 2010; 397(1):316–31. doi: [10.1016/j.jmb.2010.01.020](#) PMID: [20080101](#)
29. Buedenbender S, Rachid S, Muller R, Schulz GE. Structure and action of the myxobacterial chondrochloren halogenase CndH: a new variant of FAD-dependent halogenases. *J Mol Biol*. 2009; 385(2):520–30. doi: [10.1016/j.jmb.2008.10.057](#) PMID: [19000696](#)
30. Dorrestein PC, Yeh E, Garneau-Tsodikova S, Kelleher NL, Walsh CT. Dichlorination of a pyrrolyl-S-carrier protein by FADH<sub>2</sub>-dependent halogenase PltA during pyoluteorin biosynthesis. *Proc Natl Acad Sci U S A*. 2005; 102(39):13843–8. PMID: [16162666](#)

31. Deng H, O'Hagan D. The fluorinase, the chlorinase and the duf-62 enzymes. *Curr Opin Chem Biol*. 2008; 12(5):582–92. doi: [10.1016/j.cbpa.2008.06.036](https://doi.org/10.1016/j.cbpa.2008.06.036) PMID: [18675376](https://pubmed.ncbi.nlm.nih.gov/18675376/)
32. Dong C, Huang F, Deng H, Schaffrath C, Spencer JB, O'Hagan D, et al. Crystal structure and mechanism of a bacterial fluorinating enzyme. *Nature*. 2004; 427(6974):561–5. PMID: [14765200](https://pubmed.ncbi.nlm.nih.gov/14765200/)
33. Zhu X, Robinson DA, McEwan AR, O'Hagan D, Naismith JH. Mechanism of enzymatic fluorination in *Streptomyces cattleya*. *J Am Chem Soc*. 2007; 129(47):14597–604. PMID: [17985882](https://pubmed.ncbi.nlm.nih.gov/17985882/)
34. Eustaquio AS, Pojer F, Noel JP, Moore BS. Discovery and characterization of a marine bacterial SAM-dependent chlorinase. *Nat Chem Biol*. 2008; 4(1):69–74. PMID: [18059261](https://pubmed.ncbi.nlm.nih.gov/18059261/)
35. Deng H, Cobb SL, McEwan AR, McGlinchey RP, Naismith JH, O'Hagan D, et al. The fluorinase from *Streptomyces cattleya* is also a chlorinase. *Angew Chem Int Edit*. 2006; 45(5):759–62.
36. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997; 25(17):3389–402. PMID: [9254694](https://pubmed.ncbi.nlm.nih.gov/9254694/)
37. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, et al. CDD: NCBI's conserved domain database. *Nucleic Acids Res*. 2015; 43(Database issue):D222–6. doi: [10.1093/nar/gku1221](https://doi.org/10.1093/nar/gku1221) PMID: [25414356](https://pubmed.ncbi.nlm.nih.gov/25414356/)
38. Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res*. 2016; 44(D1):D279–85. doi: [10.1093/nar/gkv1344](https://doi.org/10.1093/nar/gkv1344) PMID: [26673716](https://pubmed.ncbi.nlm.nih.gov/26673716/)
39. Shi J, Blundell TL, Mizuguchi K. FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *J Mol Biol*. 2001; 310(1):243–57. PMID: [11419950](https://pubmed.ncbi.nlm.nih.gov/11419950/)
40. Madej T, Lanczycki CJ, Zhang D, Thiessen PA, Geer RC, Marchler-Bauer A, et al. MMDB and VAST+: tracking structural similarities between macromolecular complexes. *Nucleic Acids Res*. 2014; 42(Database issue):D297–303. doi: [10.1093/nar/gkt1208](https://doi.org/10.1093/nar/gkt1208) PMID: [24319143](https://pubmed.ncbi.nlm.nih.gov/24319143/)
41. Hall BG. Building phylogenetic trees from molecular data with MEGA. *Mol Biol Evol*. 2013; 30(5):1229–35. doi: [10.1093/molbev/mst012](https://doi.org/10.1093/molbev/mst012) PMID: [23486614](https://pubmed.ncbi.nlm.nih.gov/23486614/)
42. Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res*. 2014; 42(Web Server issue):W320–4. doi: [10.1093/nar/gku316](https://doi.org/10.1093/nar/gku316) PMID: [24753421](https://pubmed.ncbi.nlm.nih.gov/24753421/)
43. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res*. 2014; 42(W1):W252–W8.
44. Hall BG. *Phylogenetic trees made easy: a how-to manual*. 4th ed. Sunderland, MA, USA: Sinauer Associates, Inc.; 2011.
45. Tamura K, Stecher G, Peterson D, Filipinski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol*. 2013; 30(12):2725–9. doi: [10.1093/molbev/mst197](https://doi.org/10.1093/molbev/mst197) PMID: [24132122](https://pubmed.ncbi.nlm.nih.gov/24132122/)
46. Elmi F, Lee HT, Huang JY, Hsieh YC, Wang YL, Chen YJ, et al. Stereoselective esterase from *Pseudomonas putida* IFO12996 reveals  $\alpha/\beta$  hydrolase folds for D- $\beta$ -acetylthioisobutyric acid synthesis. *J Bacteriol*. 2005; 187(24):8470–6. PMID: [16321951](https://pubmed.ncbi.nlm.nih.gov/16321951/)
47. Dunn G, Montgomery MG, Mohammed F, Coker A, Cooper JB, Robertson T, et al. The structure of the C-C bond hydrolase MhpC provides insights into its catalytic mechanism. *J Mol Biol*. 2005; 346(1):253–65. PMID: [15663942](https://pubmed.ncbi.nlm.nih.gov/15663942/)
48. Shaw E, McCue LA, Lawrence CE, Dordick JS. Identification of a novel class in the  $\alpha/\beta$  hydrolase fold superfamily: the N-myc differentiation-related proteins. *Proteins*. 2002; 47(2):163–8. PMID: [11933063](https://pubmed.ncbi.nlm.nih.gov/11933063/)
49. Lenfant N, Hotelier T, Velluet E, Bourme Y, Marchot P, Chatonnet A. ESTHER, the database of the  $\alpha/\beta$ -hydrolase fold superfamily of proteins: tools to explore diversity of functions. *Nucleic Acids Res*. 2013; 41(Database issue):D423–9. doi: [10.1093/nar/gks1154](https://doi.org/10.1093/nar/gks1154) PMID: [23193256](https://pubmed.ncbi.nlm.nih.gov/23193256/)
50. Ollis DL, Cheah E, Cygler M, Dijkstra B, Frolow F, Franken SM, et al. The alpha/beta hydrolase fold. *Protein Eng*. 1992; 5(3):197–211. PMID: [1409539](https://pubmed.ncbi.nlm.nih.gov/1409539/)
51. Nardini M, Dijkstra BW. Alpha/beta hydrolase fold enzymes: the family keeps growing. *Curr Opin Struct Biol*. 1999; 9(6):732–7. PMID: [10607665](https://pubmed.ncbi.nlm.nih.gov/10607665/)
52. Weyand M, Hecht H, Kiess M, Liaud M, Vilter H, Schomburg D. X-ray structure determination of a vanadium-dependent haloperoxidase from *Ascophyllum nodosum* at 2.0 Å resolution. *J Mol Biol*. 1999; 293(3):595–611. PMID: [10543953](https://pubmed.ncbi.nlm.nih.gov/10543953/)
53. Isupov MN, Dalby AR, Brindley AA, Izumi Y, Tanabe T, Murshudov GN, et al. Crystal structure of dodecameric vanadium-dependent bromoperoxidase from the red algae *Corallina officinalis*. *J Mol Biol*. 2000; 299(4):1035–49. PMID: [10843856](https://pubmed.ncbi.nlm.nih.gov/10843856/)

54. Messerschmidt A, Wever R. X-ray structure of a vanadium-containing enzyme: chloroperoxidase from the fungus *Curvularia inaequalis*. *Proc Natl Acad Sci U S A*. 1996; 93(1):392–6. PMID: [8552646](#)
55. Ishikawa K, Mihara Y, Gondoh K, Suzuki E, Asano Y. X-ray structures of a novel acid phosphatase from *Escherichia blattae* and its complex with the transition-state analog molybdate. *EMBO J*. 2000; 19(11):2412–23. PMID: [10835340](#)
56. Littlechild J, Garcia-Rodriguez E, Dalby A, Isupov M. Structural and functional comparisons between vanadium haloperoxidase and acid phosphatase enzymes. *J Mol Recognit*. 2002; 15(5):291–6. PMID: [12447906](#)
57. de Macedo-Ribeiro S, Renirie R, Wever R, Messerschmidt A. Crystal structure of a trapped phosphate intermediate in vanadium apochloroperoxidase catalyzing a dephosphorylation reaction. *Biochemistry*. 2008; 47(3):929–34. doi: [10.1021/bi7018628](#) PMID: [18163651](#)
58. Sundaramoorthy M, Turner J, Poulos TL. The crystal structure of chloroperoxidase: a heme peroxidase-cytochrome P450 functional hybrid. *Structure*. 1995; 3(12):1367–77. PMID: [8747463](#)
59. Pecyna MJ, Ullrich R, Bittner B, Clemens A, Scheibner K, Schubert R, et al. Molecular characterization of aromatic peroxxygenase from *Agroclybe aegerita*. *Appl Microbiol Biotechnol*. 2009; 84(5):885–97. doi: [10.1007/s00253-009-2000-1](#) PMID: [19434406](#)
60. Hofrichter M, Ullrich R, Pecyna MJ, Liers C, Lundell T. New and classic families of secreted fungal heme peroxidases. *Appl Microbiol Biotechnol*. 2010; 87(3):871–97. doi: [10.1007/s00253-010-2633-0](#) PMID: [20495915](#)
61. Hofrichter M, Kellner H, Pecyna MJ, Ullrich R. Fungal unspecific peroxxygenases: heme-thiolate proteins that combine peroxidase and cytochrome p450 properties. *Adv Exp Med Biol*. 2015; 851:341–68. doi: [10.1007/978-3-319-16009-2\\_13](#) PMID: [26002742](#)
62. Chang Z, Platt P, Gerwick WH, Nguyen VA, Willis CL, Sherman DH. The barbamide biosynthetic gene cluster: a novel marine cyanobacterial system of mixed polyketide synthase (PKS)-non-ribosomal peptide synthetase (NRPS) origin involving an unusual trichloroleucyl starter unit. *Gene*. 2002; 296(1–2):235–47. PMID: [12383521](#)
63. Pazy Y, Motaleb MA, Guarneri MT, Charon NW, Zhao R, Silversmith RE. Identical phosphatase mechanisms achieved through distinct modes of binding phosphoprotein substrate. *Proc Natl Acad Sci U S A*. 2010; 107(5):1924–9. doi: [10.1073/pnas.0911185107](#) PMID: [20080618](#)
64. Park SY, Chao X, Gonzalez-Bonet G, Beel BD, Bilwes AM, Crane BR. Structure and function of an unusual family of protein phosphatases: the bacterial chemotaxis proteins CheC and CheX. *Mol Cell*. 2004; 16(4):563–74. PMID: [15546616](#)
65. Yeh E, Garneau S, Walsh CT. Robust *in vitro* activity of RebF and RebH, a two-component reductase/halogenase, generating 7-chlorotryptophan during rebeccamycin biosynthesis. *Proc Natl Acad Sci U S A*. 2005; 102(11):3960–5. PMID: [15743914](#)
66. Lin S, Van Lanen SG, Shen B. Regiospecific chlorination of (S)- $\beta$ -tyrosyl-S-carrier protein catalyzed by SgcC3 in the biosynthesis of the enediyne antitumor antibiotic C-1027. *J Am Chem Soc*. 2007; 129(41):12432–8. PMID: [17887753](#)
67. Dym O, Eisenberg D. Sequence-structure analysis of FAD-containing proteins. *Protein Sci*. 2001; 10(9):1712–28. PMID: [11514662](#)
68. Deng H, Botting CH, Hamilton JT, Russell RJ, O'Hagan D. S-adenosyl-L-methionine:hydroxide adenosyltransferase: a SAM enzyme. *Angew Chem Int Ed*. 2008; 47(29):5357–61.
69. Rao KN, Burley SK, Swaminathan S. Crystal structure of a conserved protein of unknown function (MJ1651) from *Methanococcus jannaschii*. *Proteins*. 2008; 70(2):572–7. PMID: [17910070](#)