

RESEARCH ARTICLE

# A flavin-dependent halogenase from metagenomic analysis prefers bromination over chlorination

Pia R. Neubauer<sup>1</sup>, Christiane Widmann<sup>2</sup>, Daniel Wibberg<sup>3</sup>, Lea Schröder<sup>4</sup>, Marcel Frese<sup>1</sup>, Tilman Kottke<sup>4</sup>, Jörn Kalinowski<sup>3</sup>, Hartmut H. Niemann<sup>2</sup>, Norbert Sewald<sup>1\*</sup>

**1** Organic and Bioorganic Chemistry (OCIII), Bielefeld University, Bielefeld, Germany, **2** Structural Biochemistry (BCIV), Bielefeld University, Bielefeld, Germany, **3** Center for Biotechnology (CeBiTec), Bielefeld University, Bielefeld, Germany, **4** Physical Chemistry (PCIII), Bielefeld University, Bielefeld, Germany

\* [norbert.sewald@uni-bielefeld.de](mailto:norbert.sewald@uni-bielefeld.de)



**OPEN ACCESS**

**Citation:** Neubauer PR, Widmann C, Wibberg D, Schröder L, Frese M, Kottke T, et al. (2018) A flavin-dependent halogenase from metagenomic analysis prefers bromination over chlorination. PLoS ONE 13(5): e0196797. <https://doi.org/10.1371/journal.pone.0196797>

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

**Received:** December 13, 2017

**Accepted:** April 19, 2018

**Published:** May 10, 2018

**Copyright:** © 2018 Neubauer et al. 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:** The authors acknowledge support for the Article Processing Charge by the Deutsche Forschungsgemeinschaft and the Open Access Publication Fund of Bielefeld University. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Abstract

Flavin-dependent halogenases catalyse halogenation of aromatic compounds. In most cases, this reaction proceeds with high regioselectivity and requires only the presence of FADH<sub>2</sub>, oxygen, and halide salts. Since marine habitats contain high concentrations of halides, organisms populating the oceans might be valuable sources of yet undiscovered halogenases. A new Hidden-Markov-Model (HMM) based on the PFAM tryptophan halogenase model was used for the analysis of marine metagenomes. Eleven metagenomes were screened leading to the identification of 254 complete or partial putative flavin-dependent halogenase genes. One predicted halogenase gene (*brvH*) was selected, codon optimised for *E. coli*, and overexpressed. Substrate screening revealed that this enzyme represents an active flavin-dependent halogenase able to convert indole to 3-bromoindole. Remarkably, bromination prevails also in a large excess of chloride. The BrvH crystal structure is very similar to that of tryptophan halogenases but reveals a substrate binding site that is open to the solvent instead of being covered by a loop.

## Introduction

In many bioactive compounds, halogen substituents are an important factor for the biological activity [1]. Haloperoxidases and cofactor- and metal-free haloperoxidases (perhydrolases) were the first known halogenating enzymes. These require hydrogen peroxide for the halogenation reaction and lack substrate specificity and regioselectivity [2,3]. Flavin-dependent halogenases (FHals) play an essential role in the regioselective halogenation of natural products such as chloramphenicol [4,5], vancomycin [6,7], or cryptophycin [8,9]. Chemical halogenation of aromatic compounds often requires harsh reaction conditions using catalysts like Lewis acids and often lacks regioselectivity. FHals have the potential to overcome the drawbacks of conventional chemical approaches, as they work under more environmentally friendly conditions [10,11]. Only halide salts, oxygen, and the cofactor FADH<sub>2</sub> are required

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

for regioselective halogenation [10]. Moreover, it is possible to combine chemical synthesis with enzymatic halogenation in one-pot reactions. Frese et al. combined enzymatic halogenation of tryptophan with a subsequent Suzuki-Miyaura cross-coupling reaction leading selectively to C5-, C6- or C7-aryl-substituted tryptophan derivatives [12]. Preparative amounts of halogenated products can be obtained using immobilisation of the halogenases and their auxiliary enzymes as cross-linked enzyme aggregates (CLEAs) [13].

FHals can be divided into two major classes based on their substrate preference [14]. Halogenases belonging to variant A accept free soluble substrates. The tryptophan 7-halogenases RebH [15] and PrnA [16,17], the tryptophan 6-halogenase Thal [18] as well as the tryptophan 5-halogenase PyrH [19] belong to the most prominent members of this variant which can be subdivided into two subgroups with one group catalysing the halogenation of tryptophan and indole derivatives, whereas the other group only accepts phenol and pyrrole derivatives [20,21]. In contrast, variant B halogenases require substrates that are bound to carrier proteins. These halogenases catalyse a step in non-ribosomal peptide synthesis or polyketide synthesis. CndH and PltA are examples of enzymes that halogenate such carrier-bound tyrosyl- or pyrrolyl-residues during the biosynthesis of chondrochloren or pyoluteorin, respectively [14,22]. Most of the known FHals are of bacterial origin, but during the last years, fungal FHals have been identified as well. Rdc2, which is responsible for catalysing the halogenation of radicicol, was the first identified fungal FHal [23]. Recently, the fungal FHals, RadH and MalA, which are able to halogenate complex substrates, have been identified [24,25]. RadH is highly similar to Rdc2 on amino acid level and is able to halogenate monocillin II in radicicol biosynthesis [24]. MalA is involved in the synthesis of malbrancheamide and halogenates premalbrancheamide. It is proposed that MalA may be a part of a new class of FHals, which possess a Zn<sup>2+</sup>-binding C-terminus and a flexible active site for the halogenation of complex substrates [25].

Halogenases have the potential to provide a clean biocatalytic alternative to chemical halogenation. Yet, only a few FHals have been fully characterised. One of the major reasons for this is the cumbersome identification of the natural substrate that is being accepted by the enzyme, especially in case of variant B halogenases that act on carrier-bound intermediates of biosynthetic pathways. Hence, mainly FHals catalysing the regioselective halogenation of freely soluble tryptophan in different positions have been characterised until now. Based on the crystal structures of PrnA [26], RebH [27] and PyrH [21], a reaction mechanism has been suggested: FADH<sub>2</sub> reacts with molecular oxygen forming a flavin-hydroperoxide (FAD-OOH). This intermediate is attacked by a halide ion, leading to the formation of hypohalous acid (HOX) within the flavin binding site. The HOX is then transferred within the enzyme through a tunnel of 10 Å in length to the substrate binding site. A conserved lysine residue (K79 in PrnA [26]) is responsible for the electrophilic aromatic substitution of the substrate [21, 26, 27]. The putative influence of an *N*-haloamine of this particular lysine residue suggested on the basis of experimental evidence on the halogenation reaction is still under debate. Besides this, a conserved glutamic acid (E346 in PrnA) [26] is essential for appropriate orientation and further activation of HOX prior to halogenation [28]. Two additional regions are highly conserved in flavin-dependent halogenases: the FAD binding module (GxGxxG) [29,30] and a WxWxIP motif suggested to prevent the enzyme from functioning as a monooxygenase [26]. These conserved regions provide a promising basis for the identification of new FHals from other organisms based on a bioinformatic approach.

Considerable effort was put into the investigation of novel halogenases in the recent years. For example, directed evolution led to more thermostable halogenases [31]. The Lewis group used structure information for switching the regioselectivity of a halogenase [32]. In 2005, Zehner and coworkers positively identified the Trp 5-halogenase, PyrH in the model organism *Streptomyces rugosporus* LL-42D005 based on conserved regions by using degenerative

primers for PCR amplification [19]. Recently, Smith et al. identified a novel FHal from the metagenome of the marine sponge *Theonella swinhoei* WA that halogenates 5-hydroxytryptophan. Specific primers were used for identification of this novel halogenase [33]. Bioinformatic analyses for the identification of novel enzymes have so far been successfully carried out for different biotechnologically relevant enzymes, albeit not for FHals. For example, an unknown laccase-like enzyme [34] and an active chitinase [35] have been identified from metagenomic data sets by applying bioinformatics analyses. For the identification an adapted Hidden-Markov-model (HMM) was applied, which detected conserved regions of the particular enzymes [34,35].

Halides like chloride and bromide are present in high concentrations in marine habitats. Although the chloride concentration (19.345 ‰) is much higher than the bromide concentration (0.066 ‰) [36], many brominated natural products have been found in marine organisms [37]. It was reported that terrestrial halogenases mainly use chloride while marine halogenases tend to prefer bromide. Therefore, the ocean is a beneficial habitat for the evolution of brominating enzymes [20]. However, many of these potential enzymes are encoded by genes originating from microorganisms that are uncultivable under standard laboratory conditions. In this study, different marine metagenomes were screened for FHals by employing a HMM that is based on the PFAM tryptophan halogenase model (Trp\_halogenase, PF04820). We expected to find novel enzymes from cultivable as well as from uncultivable organisms using metagenomic data sets in order to increase the scope of possible halogenases.

## Materials and methods

### Materials

The chemicals and solvents were obtained, unless otherwise noted, from commercial suppliers in highest purity suitable for analytical applications (*p. a.*).

The plasmid vector pET-21\_ADH encoding for alcohol dehydrogenase was kindly donated by Prof. Dr. Werner Hummel and the plasmid pClBhis-PrnF encoding for the flavin reductase in *Pseudomonas fluorescens* was a gift from Prof. Dr. Karl-Heinz van Pée. The plasmid pGro7 for the chaperone GroEL-GroES was purchased from TaKaRa Bio Inc. The plasmid pETM-11 was obtained from Gunter Stier (EMBL). Competent cells *E. coli* DH5 $\alpha$  and *E. coli* BL21 (DE3) were obtained from Novagen. ThermoFisher GeneArt synthesised the flavin-dependent halogenase genes *brvH*.

### Analytical reversed-phase high performance liquid chromatography (RP-HPLC)

For analytical reversed-phase high performance liquid chromatography, three different methods were used. By using method A, reactions were monitored using a Thermo Scientific Accela 600 with Hypersil GOLD 3  $\mu$ m from Thermo Scientific (150  $\times$  2.1 mm, eluent A: H<sub>2</sub>O/CH<sub>3</sub>CN/TFA = 95:5:0.1, eluent B: H<sub>2</sub>O/CH<sub>3</sub>CN/TFA = 5:95:0.1, flow rate 700  $\mu$ l/min with a linear gradient from 0–100% of eluent B over 5 minutes, for 1 minute with eluent B, then in half a minute going back to 100% eluent A and staying with 100% A for 2 minutes).

In method B, reactions were monitored using a Thermo Scientific Accela 600 with NUCLEOSHELL<sup>®</sup> RP 18 column 18.5  $\mu$ m from Macherey-Nagel (150  $\times$  2.1 mm, eluent A: H<sub>2</sub>O/CH<sub>3</sub>CN/TFA = 95:5:0.1, eluent B: H<sub>2</sub>O/CH<sub>3</sub>CN/TFA = 5:95:0.1, flow rate 900  $\mu$ l/min, first minute 100% eluent A, then a linear gradient from 0–100% eluent B in 5 minutes, 1 minute at 100% B and in a half minute going back to 100% A. At the end staying for 2 minutes with eluent A).

Based on method C, reactions were monitored using a Shimadzu Nexera XR Luna<sup>®</sup> 3  $\mu\text{m}$  C18(2) 100  $\text{\AA}$ , LC Column from Phenomenex (100  $\times$  2 mm, eluent A: H<sub>2</sub>O/CH<sub>3</sub>CN/TFA = 95:5:0.1, eluent B: H<sub>2</sub>O/CH<sub>3</sub>CN/TFA = 5:95:0.1, flow rate 500  $\mu\text{L}/\text{min}$  isocratic with 35% of eluent B over 5 minutes).

For the reactions monitored by method D a column from Phenomenex (XR Luna<sup>®</sup> 3  $\mu\text{m}$  C18(2) 100  $\text{\AA}$ , LC column, 100  $\times$  2 mm) was used. Eluent A: H<sub>2</sub>O/CH<sub>3</sub>CN/TFA = 95:5:0.1, eluent B: H<sub>2</sub>O/CH<sub>3</sub>CN/TFA = 5:95:0.1, flow rate 650  $\mu\text{L}/\text{min}$  isocratic with 35% of eluent B over 7 minutes.

### Preparative reversed-phase high performance liquid chromatography (RP-HPLC)

For purification of products a preparative HPLC (Merck-Hitachi LaChrom) with Thermo Fisher Hypersil Gold 8  $\mu\text{m}$  column (250  $\times$  21.2 mm, eluent A: H<sub>2</sub>O/CH<sub>3</sub>CN/TFA = 95:5:0.1, eluent B: H<sub>2</sub>O/CH<sub>3</sub>CN/TFA = 5:95:0.1, flow rate 10 mL/min with a linear gradient from 0–100% B over 45 minutes) was used.

### Gas chromatography—Mass spectrometry (GC-MS)

For GC-MS analysis the gas chromatograph Trace GC Ultra (ThermoScientific) with a VF-5 column 0.25  $\mu\text{m}$  (30 m  $\times$  0.25 mm, 5% diphenylsiloxan, 95% dimethylsiloxan was employed; for mobile phase helium was used with temperature gradient of 5  $^{\circ}\text{C}/\text{min}$  from 80  $^{\circ}\text{C}$  to 325  $^{\circ}\text{C}$ ) and the mass spectrometer ITQ900 from ThermoFinnigan (20 measurements per minute, 50–750  $m/z$ ) were used.

### Nuclear magnetic resonance (NMR) spectroscopy

NMR spectra were recorded on a Bruker DRX-500 spectrometer (<sup>1</sup>H: 500 MHz, <sup>13</sup>C: 126 MHz). Chemical shifts are reported relative to residual solvent peaks (DMSO-*d*<sub>6</sub>: <sup>1</sup>H: 2.5 ppm; <sup>13</sup>C: 39.5 ppm)

## General methods

### Metagenomic analysis for the detection of flavin-dependent halogenases

All metagenomic data were obtained from the iMicrobe.us or gold.jgi.doe.gov databases. A two-step approach as recently described [34] with some modifications was used for the construction of a HMM for the detection of flavin-dependent halogenases. In the first step, the PFAM model for tryptophan halogenase (PF04820) was used as basis. To improve the initial model, conserved sequences were collected by applying BLAST [38] using known protein sequences of already characterized FHals with a threshold of 50% sequence identity and 90% query coverage as references. The obtained sequences were aligned using MUSCLE [39] and the alignment was processed manually. Finally, a halogenase HMM was generated based on the initial HMM for tryptophan halogenase and the obtained alignment using the HMMER3 package [40]. Assembly of metagenomic data was performed by applying MEGAHIT [41]. For gene prediction, the tool Prodigal [42] was used and the predicted genes were filtered based on their completeness and their match to the model. Completeness means that the gene contains a start and stop codon, conserved regions are present and the size of the genes fits the size of known halogenases. For example, the *rebH* gene contains 1593 base pairs, corresponding to 530 amino acids in length [15]. The predicted and translated genes were compared against the halogenase HMM using the HMMER3 software package. Only genes with an  $e\text{-value} < 1 \times 10^{-150}$  were used for further analyses.

## Phylogenetic analyses of the positive halogenases hits

For phylogenetic analyses MEGA7 [43] was used. First, a protein alignment (MUSCLE) was performed and then the phylogenetic tree was constructed based on neighbour joining (NJ) method and a bootstrap of 1000.

## Vector preparation, heterologous protein expression

BrvH has been identified in course of the bioinformatic approach explained above. The gene *brvH* was identified in Botany Bay metagenome (2016 iMicrobe, CAM\_SMPL\_001699). The gene of *brvH* was codon optimised for *E. coli* and extended with restriction sites (*NdeI* and *BamHI*) for further cloning. The synthetic gene was obtained from Invitrogen GeneArt. The gene was cloned into pET-28a vector and transformed to *E. coli* BL21 (DE3) pGro7. The pGro7 vector codes for the expression of a chaperone system GroEL-GroES. An overnight preculture (37 °C, 150 rpm) was used to inoculate a culture in LB medium containing the appropriate antibiotics (60 µg/mL kanamycin; 50 µg/mL chloramphenicol). This culture was grown (37 °C, 150 rpm) up to an OD<sub>600</sub> of 0.5. Protein expression was then induced with isopropyl-β-D-thiogalactopyranoside (IPTG) (0.1 mM) and L-arabinose (2 g/L) and the cells were cultivated at 25 °C, 150 rpm for 22 h. Afterwards, cells were harvested by centrifugation (3220 × g, 30 min), washed with 100 mM Na<sub>2</sub>HPO<sub>4</sub> buffer (pH7.4) and stored at -20 °C.

For protein crystallisation, the *brvH* gene was cloned into a pETM-11 vector. This vector includes a hexahistidine tag, followed by a tobacco etch virus (TEV) protease cleavage site. The plasmid was transformed into *E. coli* BL21 (DE3). An overnight preculture (30 °C, 100 rpm) was used to inoculate a culture in LB medium containing 30 µg/mL kanamycin. The culture was grown (37 °C, 100 rpm) to an OD<sub>600</sub> of 0.7 and protein expression was induced using 0.1 mM IPTG. After 18 h of expression at 20 °C, the cells were harvested by centrifugation (see above), and the cell pellet washed with PBS and frozen at -20 °C.

## Site-directed mutagenesis of *brvH*

A PCR was carried out to mutate the conserved lysine residue to an alanine (K83A). The following primers were designed:

BrvH\_K83A: GCAACCCAGGCAACCTGTGCGCTGGGTATTTCGTTTT

BrvH\_K83A-rev: AAAACGAATACCCAGCGCACAGGTTGCCTGGGTTGC

For the PCR, 70 ng of vector DNA (*brvH*\_pET-28a), 125 ng of each primer, 2.5 U/mL Pfu polymerase (Promega, Germany), 1 × reaction buffer and dNTP mix (200 µM of each nucleotide) were used. The PCR program is shown in Table 1.

Afterwards parental DNA was digested by incubating the PCR product with DpnI restriction enzyme (10 U/µl) at 37 °C. After 1 hour, the mutated DNA was transformed into *E. coli* DH5α (DE3). Protein expression and purification was identical to the non-mutated BrvH.

**Table 1. Program for the site-directed mutagenesis PCR.**

Cycles	Temperature [°C]	Minutes
1	95	0.5
16	95	0.5
	55	1
	68	7
1	68	5

<https://doi.org/10.1371/journal.pone.0196797.t001>

## Protein purification

For protein isolation, the cells from 1.5 L cultivation were thawed, suspended in 30 mL 100 mM Na<sub>2</sub>HPO<sub>4</sub> buffer (pH7.4) and lysed with the French Press (3 times 1000 psig). Soluble protein was separated from insoluble compounds by centrifugation (10000 × g, 30 min, 4 °C).

A HisTALON matrix was employed for purification. The elution of the bound enzymes was carried out with imidazole (300 mM imidazole, 100 mM Na<sub>2</sub>HPO<sub>4</sub>, 300 mM NaBr, pH 7.4) and resulting enzymes were collected in 0.5 mL fractions. Afterwards fractions with pure enzyme were pooled and a desalting column (HiTrap<sup>®</sup> Desalting (GE Healthcare)) was used to remove the imidazole and to change the buffer (50 mM Na<sub>2</sub>HPO<sub>4</sub>, 50 mM NaBr, pH 7.4).

For protein crystallisation, purification included Ni<sup>2+</sup>-affinity chromatography, ion exchange (IEX) and size exclusion chromatography (SEC). The cells were first resuspended in a lysis buffer (25 mL PBS containing cOmplete<sup>™</sup> protease inhibitor (EDTA-free; Merck, Germany) and 20 µg/mL DNase I) and lysed using a French Press (120 MPa). Soluble protein was obtained via centrifugation (16000 x g, 30 min, 4 °C) and captured via Ni<sup>2+</sup>-NTA affinity chromatography. Bound protein was eluted with imidazole (20 mM Tris 20 mM NaCl 200 mM imidazole pH 8) and an IEX of the protein-containing fractions was performed using an ÄKTA Purifier (self-packed Source 15 Q column (column volume (CV) = 7 mL; GE Healthcare, UK); 1 mL/min; gradient from 0 to 500 mM NaCl in 20 mM Tris pH 8 over 20 CV; fractions of 2 mL were collected). The FHal eluted in a single peak and was then subjected to an overnight TEV protease digest (1 mg TEV per 50 mg of BrvH) at 20 °C in order to remove the His<sub>6</sub>-tag. In a following Ni<sup>2+</sup>-NTA chromatography, the cleaved protein was obtained in the flowthrough and washing fractions and concentrated using a VivaSpin column (10000 MWCO; Sartorius, Germany). For SEC, a 16/60 Superdex 200 column (flow rate 1 mL/min; GE Healthcare, Germany) connected to an ÄKTA purifier at 4 °C was used (buffer 10 mM Tris 20 mM NaCl, pH8; fractions of 2 mL were collected). Eluted protein was concentrated to 22.3 mg/mL and stored at -80 °C prior to crystallisation.

## Cofactor regeneration system

The cofactor regeneration system consisting of flavin reductase (PrnF) and alcohol dehydrogenase (ADH) was generated as published previously [10].

## Crystallisation and structure determination of apo BrvH

BrvH was diluted to 10 mg/mL for crystallisation. Crystallisation conditions were screened using the JCSG Core IV Suite (Qiagen, Germany). Stacks of very thin plates were obtained after two weeks in 1 M Na/K tartrate, 0.1 M MES pH 6.0 with a drop size of 2 µL and a ratio of protein:reservoir solution of 1:1. Optimisation did not yield single crystals. However, using streak seeding, single plates could be obtained in 0.95 M Na/K tartrate 0.1 M MES pH 6.0 after one day. Prior to data collection, six-day-old crystals were transferred into the crystallisation solution supplemented with 20% glycerol and vitrified in liquid nitrogen. Data were collected at the ESRF, Grenoble, using the ID30 A-3 beamline (1800 frames; oscillation angle 0.15°; x-ray wavelength 0.9677 Å).

Raw data were processed with XDS and scaled with XSCALE [44]. The data were imported into CCP4 and merged with aimless [45] for structure solution via molecular replacement using Phaser (ensemble: PrnA, seq. identity 34%; PyrH, seq. identity 36%; RebH, seq. identity 33%; SttH, seq. identity 36%; [46]). An initial model was built by PHENIX AutoBuild [47] and further improved via manual building using COOT [48]. For crystallographic refinement of interim models, Refmac5 was used [49]. Validation with MOLPROBITY [50] was part of the iterative building process. The statistics of the final model are presented in S1 Table and the

crystal structure was deposited in the protein data bank (PDB ID 6FRL). Figures were generated in PyMOL [51].

### FAD reconstitution

For FAD reconstitution, the purified enzyme was incubated in FAD buffer (50 mM Na<sub>2</sub>HPO<sub>4</sub>, 150 mM NaBr, 1 mM FAD, pH 7.4) over night at 4 °C, 5 rpm. A desalting column (HiTrap<sup>®</sup> Desalting (GE Healthcare)) was used to change the buffer. After incubation overnight, BrvH was washed with buffer (50 mM Na<sub>2</sub>HPO<sub>4</sub>, 50 mM NaBr, pH 7.4) to remove free FAD using a filter device (Amicon, Ultra 4) with a 50 kDa cutoff. UV/vis spectra of the sample were measured in a quartz cuvette (Suprasil, Hellma) with a path length of 1 cm using a Shimadzu UV-2450 spectrometer.

### Assay conditions

The activity assays of PrnF and ADH were carried out as described by Frese et al. [10].

The activity assays with the novel flavin-dependent halogenases were carried out with cell lysate containing the enzymes (250 µl in 1 mL) with 1 µM FAD, 100 µM NAD, 100 mM NaBr/NaCl, PrnF (2.5 U/mL) and ADH (2 U/mL). Indole was dissolved in isopropanol to give a final substrate concentration of 1 mM and 5% solvent, respectively. When tryptophan was used as substrate, it was dissolved in water and isopropanol was added to a final concentration of 5% (v/v). Assays were incubated for 48 h at 25 °C, 300 rpm. For the activity assays using purified enzyme, a final enzyme concentration of 1.125 mg/mL was present in the reaction mixture, while the other conditions given above were unmodified. The activity assays were carried out in 1 mL or 0.5 mL volume, which was filled up with 100 mM Na<sub>2</sub>HPO<sub>4</sub> (pH 7.4). If appropriate, catalase (650 U/mL) was added to 500 µl reaction mixture of the activity assays to remove peroxide intermediates.

### Indole conversion on large scale by immobilised enzymes

Cross-linked enzyme aggregates (CLEAs) as described by Frese et al. [13] were used for conversion of higher amounts of indole for NMR analyses. First, the *E. coli* BL21\_pGro7 cells containing BrvH from a 1.5 L cultivation batch were disrupted using French Press and centrifuged as described above. 2.5 U/mL PrnF (flavin reductase) and 1 U/mL alcohol dehydrogenase (ADH) were added to the supernatant. This protein mixture was then separated into two equal parts because one cell pellet was used for two CLEA reactions. The proteins were precipitated by adding saturated ammonium sulfate solution in a tube rotator for 1 h at 4 °C. Glutaraldehyde was added to a final concentration of 0.5% and the mixture was further rotated for 2 h. Finally, the produced CLEAs were centrifuged and three times washed with 100 mM Na<sub>2</sub>HPO<sub>4</sub> buffer (pH7.4). For the biocatalysis with CLEAs 1.5 mM substrate, 1 µM FAD, 100 µM NAD, 15 mM Na<sub>2</sub>HPO<sub>4</sub>, 30 mM NaBr and 5% (v/v) isopropanol were employed. The pH was adjusted with phosphoric acid to pH 7.4. CLEAs were added to the reaction solution in a final volume of 500 mL and incubated for up to 10 days (depending on the conversion) at 25 °C, 150 rpm.

## Results and discussion

### Selection of metagenomic data sets, construction of the HMM, bioinformatic analyses and phylogenetic reconstruction of the identified FHal genes from Botany Bay to known FHal genes

For the identification of novel FHals, we analysed eleven metagenomes (S2 Table). For screening, we established a HMM based on a two-step approach for tryptophan halogenases like

recently described by Ausec et al. [34] with some modifications. After that, the halogenase HMM was applied to identify FHals within the metagenomic data sets. By applying the improved HMM model, several potential FHal genes were identified within the screened metagenomes. In total, 254 predicted FHal genes were identified, but only 77 were shown to be complete, meaning that the gene contains a start and stop codon, conserved regions are present and its size fits to the size of known halogenases. For example, the *rebH* gene contains 1593 base pairs, corresponding to 530 amino acids in length. The highest number of complete genes (42 in total) was identified from the Botany Bay metagenome (CAM\_SMPL\_001699). The two variants can be distinguished by multiple sequence alignment, which can be visualised by a phylogenetic tree. We built a phylogenetic tree with the predicted FHals from the Botany Bay metagenome in comparison to known FHals belonging to variant A and B on amino acid level. The tree was constructed via MUSCLE alignment and neighbour joining with a bootstrap of 1000 replicates using MEGA7 [43] (Fig 1). In this phylogenetic tree, the FHals of both classes cluster into two distinct clades consisting of variant A and variant B FHals. Remarkably, all predicted genes cluster in the clade of tryptophan halogenases, but only gene 38 clusters within the cluster of tryptophan halogenases. Some of the predicted FHal genes cluster more distantly from the tryptophan halogenases (5, 7, 8, 9, 16, 20, 21, 23, 24, 27, 30, 31, 32, 34, 35, 37, 40, 41). The predicted gene 17 seems to be an outgroup, because it is more distant from both the variant A and the variant B FHals. The fact that all predicted genes cluster closer to the tryptophan halogenases may be caused by the selection criteria since our HMM consists of the PFAM model of tryptophan halogenases. Moreover, the e-value and the gene size resembling the length of known tryptophan halogenases were used as criteria for selection. Weichold et al. postulated that variant B FHals seem to be around 70 amino acids shorter in comparison to variant A FHals [52]. Therefore, it can be assumed that our selection criteria rejected predicted genes belonging to variant B FHals and for this reason from 254 predicted genes only 77 were assigned as complete.

### Selection, expression and activity assay of a predicted FHal gene from Botany Bay

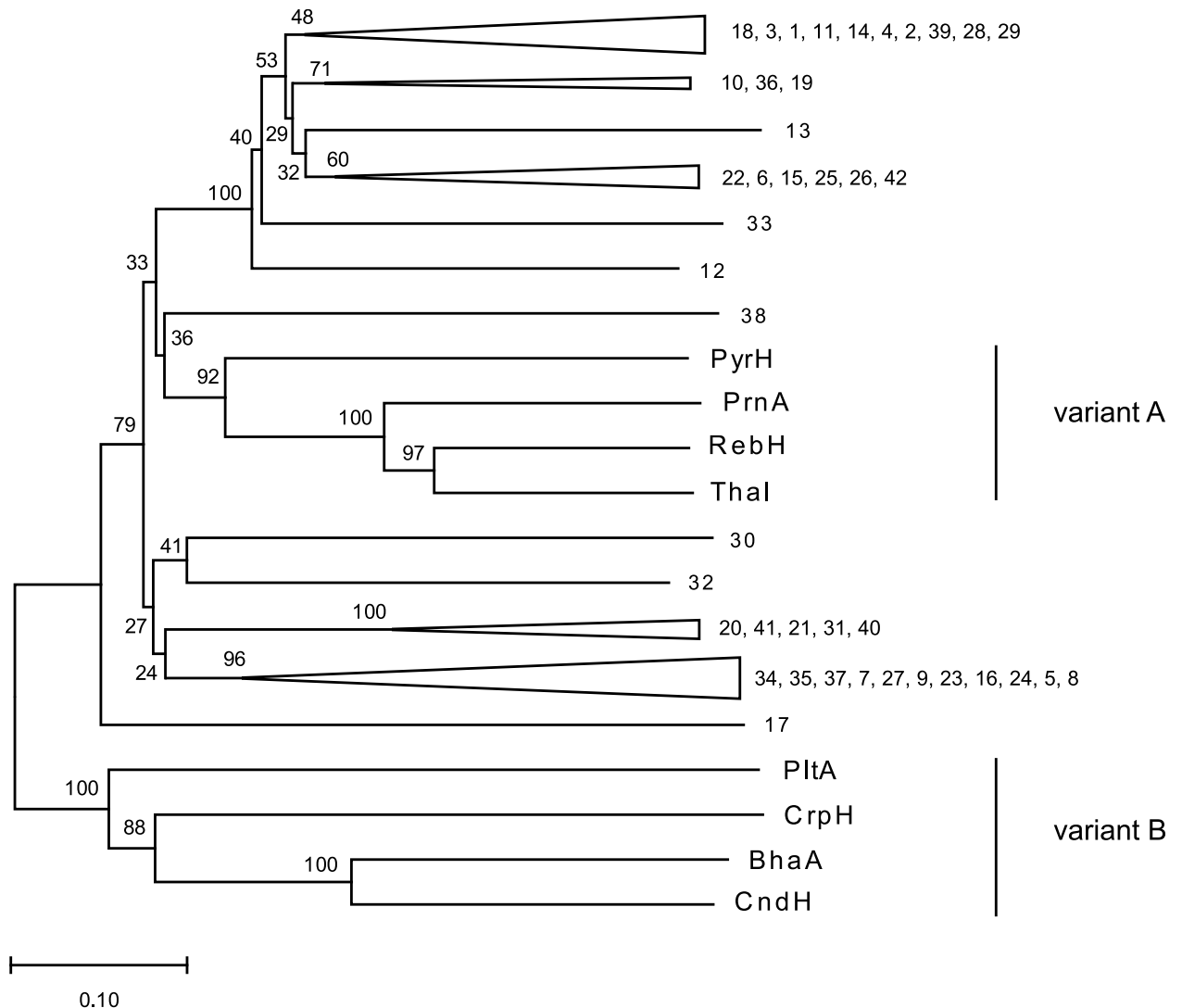
We decided to investigate one predicted FHal gene from the Botany Bay metagenome, which clusters close to tryptophan halogenases within the phylogenetic tree. We chose predicted gene 12 (1530 bp), which shows high similarities to tryptophan halogenases on amino acid level, but interestingly forms its own clade within the phylogenetic tree.

Sequence comparison of gene 12 using Blastp and the nr database resulted in one hit with 100% identity on amino acid level to an annotated putative tryptophan halogenase from *Brevundimonas* BAL3 (NCBI: Acc. # EDX81295.1), but no enzymatic activity of its gene product had been described. *Brevundimonas* BAL3 is a gram-negative alphaproteobacterium that was isolated from marine habitat (GOLD Project ID: Gp0006177). To the best of our knowledge, no halogenated product is known from this organism to date. AntiSMASH analysis of the genome of *Brevundimonas* BAL3 revealed one gene cluster leading to bacteriocin [53]. Since it is most likely that the predicted FHal gene is from *Brevundimonas* BAL3, we designated gene 12 as *brvH*.

Gene cluster analysis revealed another FHal gene 13 (1542 bp), located upstream of gene 12 (Fig 2). Gene 12 and 13 share 42.7% pairwise identity to each other.

A multiple sequence alignment based on amino acid sequences was carried out to compare BrvH with the tryptophan halogenases RebH, Thal and PyrH. The multiple alignment showed that BrvH contains all three conserved amino acid regions that were used for identification: (i) the FAD binding module, GxGxxG (amino acids (aa) 16–21), (ii) the conserved lysine residue



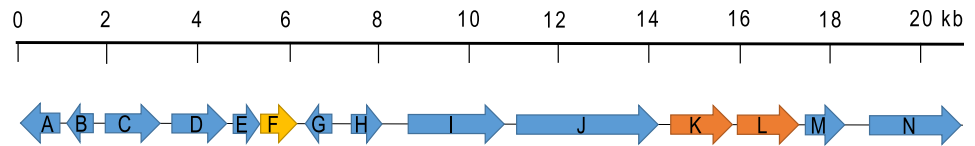


**Fig 1. Phylogenetic tree of positive detected FHals from the Botany Bay metagenome in comparison to known FHals on amino acid level.** The tree was constructed using neighbour joining method, bootstrap 1000, with the alignment based on amino acids by MEGA7 [43] and MUSCLE [39]. 1–42: positive FHal hits from the Botany Bay metagenome; PyrH, PrnA, RebH, Thal: Tryptophan halogenases belonging to variant A and accept free substrate; PltA, CrpH, BhaA, CndH: FHals belonging to variant B and require carrier bound substrate. The amino acid sequences are shown in S3 Table.

<https://doi.org/10.1371/journal.pone.0196797.g001>

(K83) and (iii) the conserved domain WxWxIP (aa 275–280). We decided to proceed with *in vitro* assays of BrvH to show halogenase activity and identify possible substrates. The synthetic gene *brvH* (codon optimised for *E. coli*) was cloned into the expression vector pET-28a. The gene was heterologously expressed in *E. coli* with coexpression of the chaperones GroEL-GroES. This resulted in adequate production of BrvH, which was obtained in soluble form and purified using an N-terminal His<sub>6</sub>-tag with immobilised metal ion chromatography (IMAC; S1 and S2 Figs).

Because our HMM is based on tryptophan halogenases and BrvH shows high similarities to these enzyme class, we suspected that tryptophan is halogenated by BrvH. However, substrate tests with cell lysate containing BrvH led to no halogenation of tryptophan, neither in presence of chloride nor in presence of bromide. Therefore, other substrates with an indole moiety, for



- A: Acyltransferase
- B: Oxidoreductase
- C: Cyclic diguanylate phosphodiesterase
- D: Hydrolase
- E: Acyl dehydratase
- F: FAD synthase
- G: Carbonic anhydrases/acetyltransferases
- H: PhoH family protein
- I: Glycosyl hydrolase
- J: Outer membrane receptor protein
- K: Tryptophan halogenase superfamily (12)
- L: Tryptophan halogenase superfamily (13)
- M: Regulatory protein
- N: Outer membrane receptor protein

**Fig 2. Gene cluster in the vicinity of gene 12 (K) in the Botany Bay metagenome.** Upstream to the gene 12 (K), another FHal, 13 (L) was identified.

<https://doi.org/10.1371/journal.pone.0196797.g002>

example 5-hydroxytryptohan, tryptophol, indole-3-propionic acid, indole-3-acetonitrile, and indole were tested. However, only indole was found to be accepted as substrate (Fig 3).

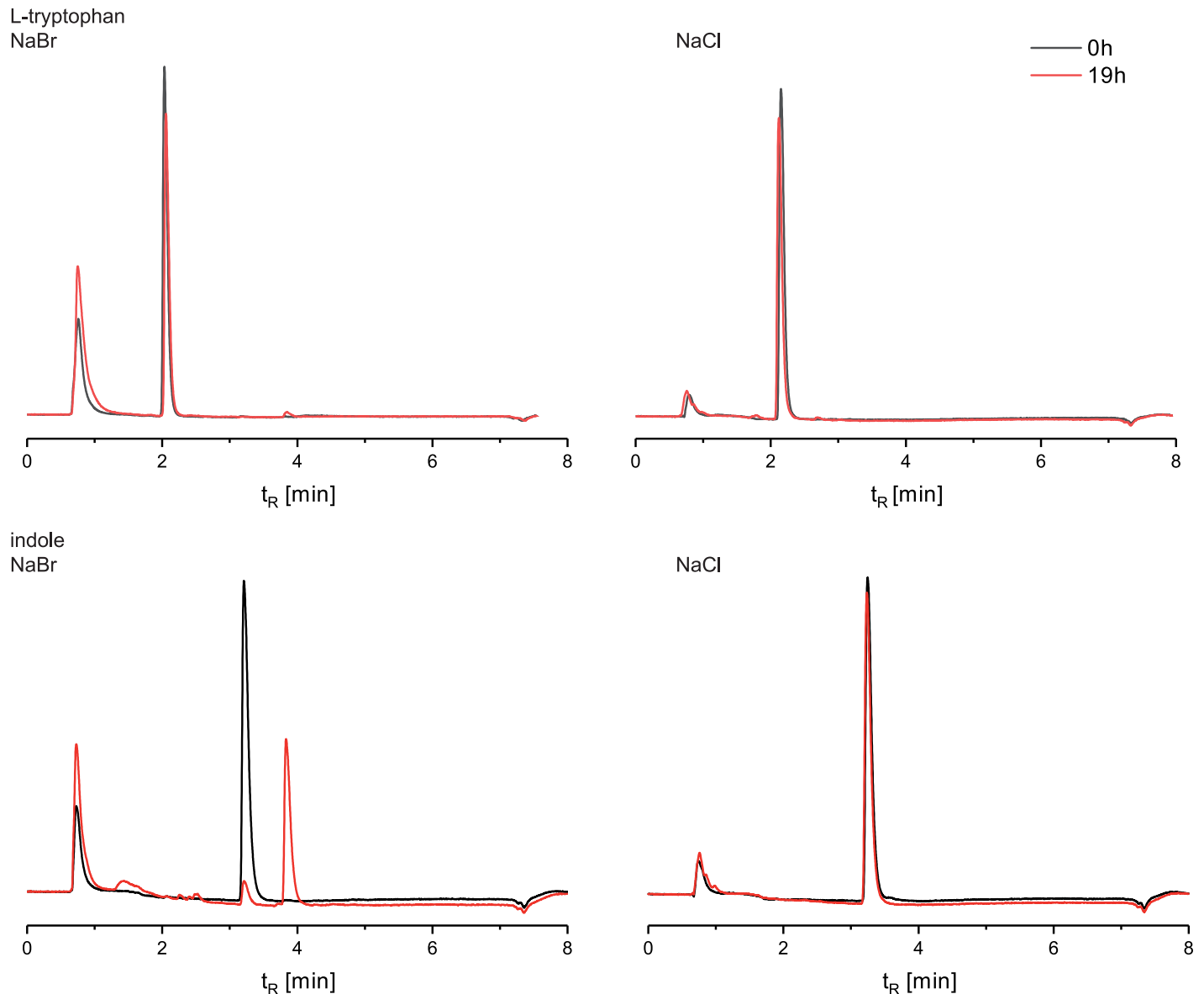
The bromoindole product was further investigated by GC-MS measurements (S3 Fig) as well as NMR analyses. COSY and HMQC experiments unambiguously proved that bromination took place in position C3 of the indole ring (S4–S7 Figs). The position of halogenation may be the reason for the non-acceptance of tryptophan. Indole and tryptophan possess an identical basic structure, but the C3 position of the indole ring is occupied in tryptophan.

The activity of BrvH underlines the feasibility of our method for the identification of novel FHals, showing that our HMM is able to identify active FHals only based on conserved regions. Low specific activity of 2.5 mU/mg and a turnover number of 39.6 lead to the conclusion that indole may not be the natural substrate of BrvH (Fig 4).

Nevertheless, it is challenging to find a suitable substrate for the identified halogenase if no halogenated natural product of the related organism is known.

Noteworthy, BrvH prefers bromide to chloride as halide source. In comparison, pure BrvH fully converts indole to bromoindole within 48 h while chlorination under identical conditions proceeds to 8.4% (S8 and S9 Figs). We need to state that we could not observe the mass of chloroindole in GC-MS measurements due to insufficient concentrations of produced chloroindole. Therefore, we obtained 3-chloroindole commercially and spiked it into the samples of the product obtained upon incubation of indole and NaCl with BrvH (Fig 5) giving the same retention times, while the halogenation product of indole with NaBr has a different retention time. This provides evidence that the halogenated product incubated with NaCl is chloroindole.

Most FHals catalyse chlorination as well as bromination, but chlorination activity usually is higher compared to bromination activity [52]. Bmp2 and Bmp5 are the only known FHals, which only catalyse bromination but not chlorination [54]. Interestingly, BrvH preferentially brominates its substrate but is also able to chlorinate in low amounts. BrvH prefers the bromination to chlorination even in excess of chloride as halide source. The HPLC chromatogram of the sample with tenfold excess of NaCl to NaBr (Fig 6) shows that small amounts of



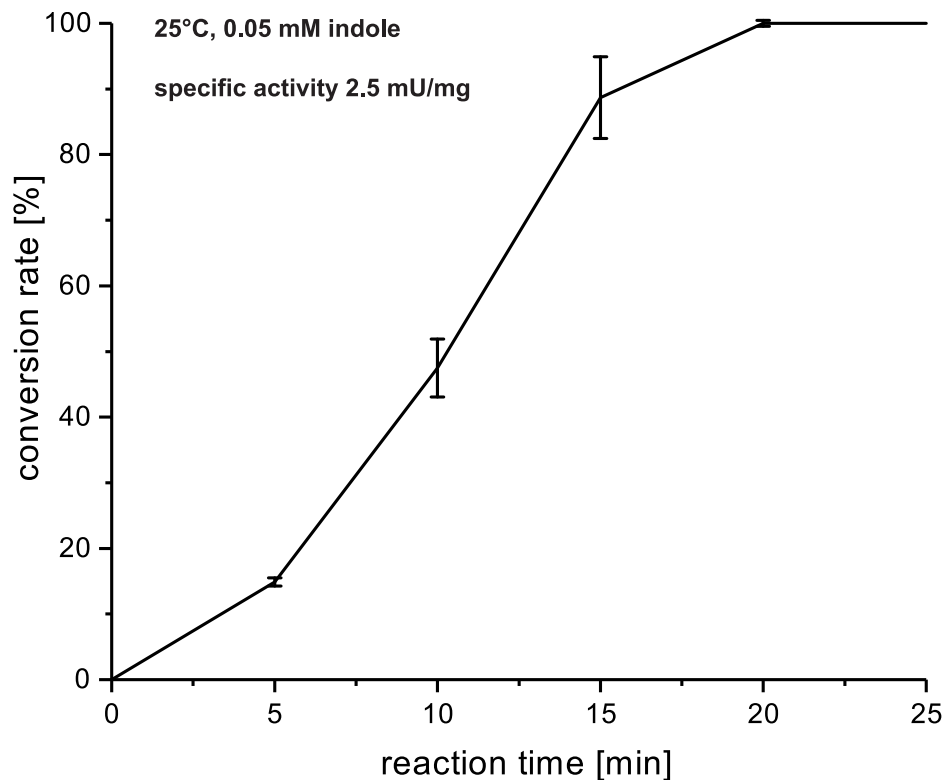
**Fig 3. HPLC traces of the enzymatic halogenation of L-tryptophan (1 mM) and indole (1 mM) by BrvH (in cell lysate for 19 h in presence of 100 mM NaCl or NaBr).** L-tryptophan ( $t_R$ : 2.2 minutes) is not being halogenated, while Indole ( $t_R$ : 3.2 minutes) is fully converted within 19 h to bromoindole ( $t_R$ : 3.8 minutes) (HPLC method A).

<https://doi.org/10.1371/journal.pone.0196797.g003>

chloroindole ( $t_R$ : 3.75 min) as well as bromoindole ( $t_R$ : 3.8 min) were formed. Hence, chlorination is much less efficient than bromination and even in ten times excess of NaCl, NaBr is preferred.

In order to confirm that BrvH is a flavin-dependent halogenase we performed different experiments for evaluation. Besides FHals, haloperoxidases and perhydrolases are able to directly introduce halogen substituents into organic compounds in the presence of hydrogen peroxide [2,3], albeit with diminished regioselectivity.

Hydrogen peroxide was enzymatically removed *in situ* from the reaction mixture by adding catalase. Catalase enzymatically decomposes hydrogen peroxide into water and oxygen [55]. BrvH halogenates indole even in the presence of active catalase (S10 Fig). Hence, the



**Fig 4. Time course of the conversion of 0.05 mM indole by BrvH with NaBr over 25 minutes at 25 °C.** The conversion rate was identified via RP-HPLC by determining the ratio of the peak areas of indole and bromoindole. The specific activity was defined between 0 and 15 minutes reaction time.

<https://doi.org/10.1371/journal.pone.0196797.g004>

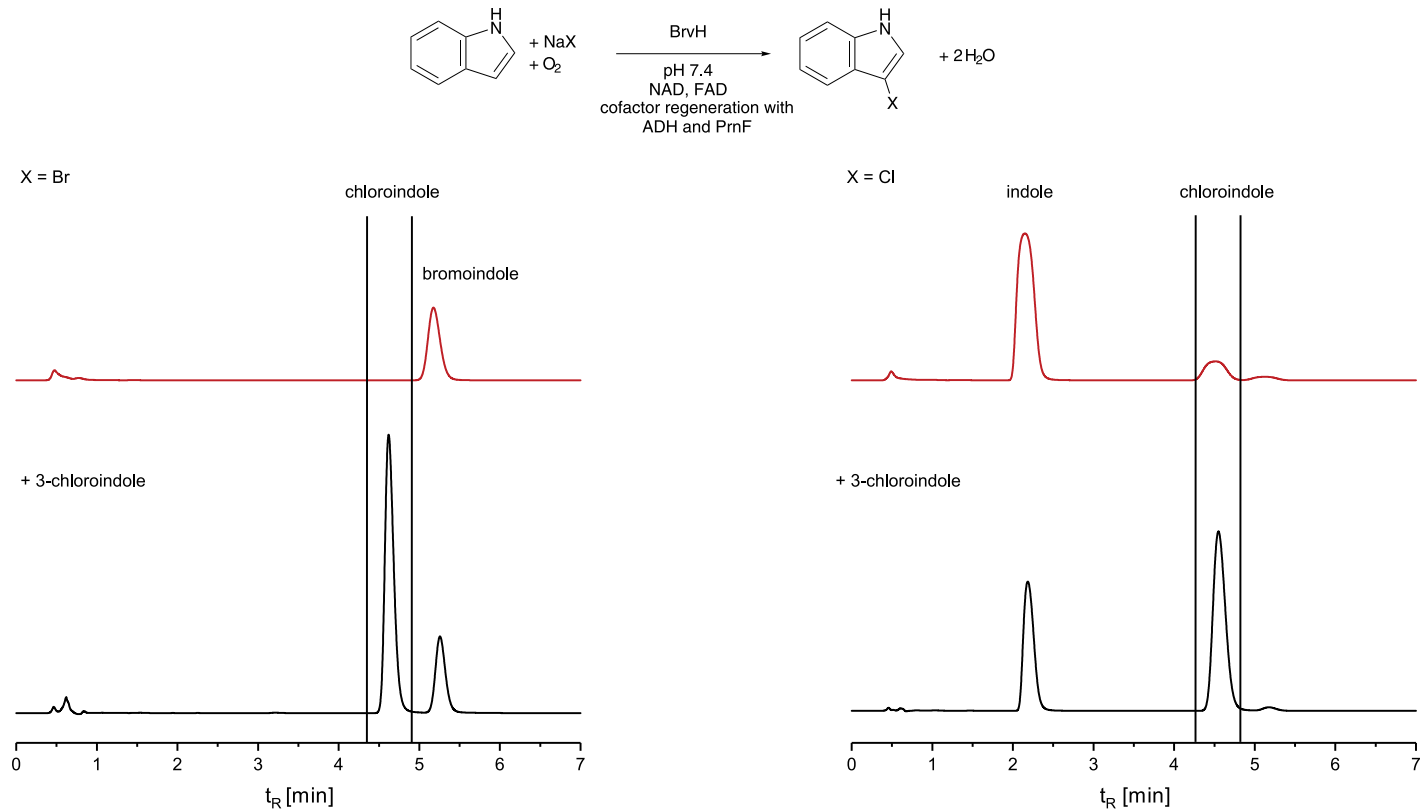
halogenation is not dependent on free hydrogen peroxide. Furthermore, a bromoperoxidase assay [56] gave evidence that BrvH is not a bromoperoxidase. Replacement of the lysine residue in position 83 that is conserved in FHals by alanine led to complete loss of enzyme activity (S11 Fig). Furthermore, we were able to show that FAD binds to the enzyme using UV/vis spectroscopy (S12 Fig). Upon binding of FAD to BrvH the band at 450 nm shows a slight shift to 448 nm while the band around 373 nm is shifted to around 360 nm. This latter shift is caused by the changed environment of FAD inside the protein compared to the polar and strongly hydrogen-bonded environment in solution [57,58]. All these experiments support BrvH being a FHal.

### Structure determination of BrvH

The observation that BrvH preferentially brominates in combination with the sequence similarity to tryptophan halogenases but the lack of halogenated tryptophan raised the question whether these characteristics could be explained by examining the 3D structure of the enzyme.

The structure of apo BrvH without any bound ligands was determined to a resolution of 2.5 Å (Fig 7A). There are two chains in the asymmetric unit with no major differences between the chains. BrvH eluted as a dimer from gel filtration. In the crystal, it forms a dimer very similar to that of tryptophan halogenases like PrnA or RebH. BrvH was modelled nearly completely, except for a few terminal residues and residues 46–51, a loop involved in FAD binding that can adopt various conformations in other FHals and is disordered in apo BrvH.

The substrate binding site is only partly conserved between BrvH and RebH. The region that is close to the active site lysine K83 (K79 in RebH) is highly similar in both proteins.



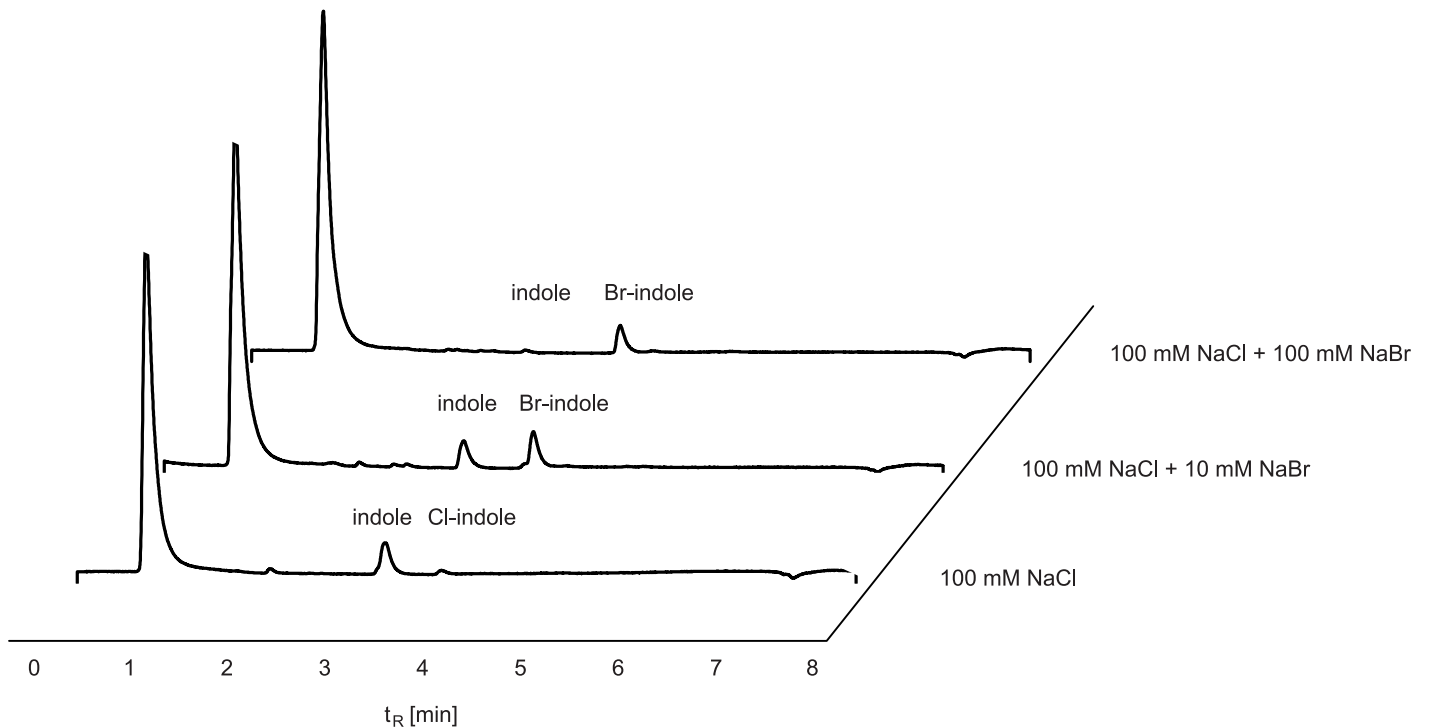
**Fig 5. HPLC traces of enzyme products of BrvH incubated with indole and NaBr/NaCl and the same products spiked with commercially available 3-chloroindole.** BrvH incubated with indole and NaBr leads to bromoindole ( $t_R$ : 5.2 min) and incubated with NaCl leads to chloroindole ( $t_R$ : 4.6 min). Both enzyme products spiked with 3-chloroindole ( $t_R$ : 4.6 min) show that the halogenated product of BrvH incubated with indole and NaCl has the same retention time with commercially 3-chloroindole (HPLC method D).

<https://doi.org/10.1371/journal.pone.0196797.g005>

Among the residues contacting the six-membered ring of the indole, there is only a single substitution (BrvH: S447 vs. RebH: N470). More variation is present among residues contacting the five-membered indole ring and the C $\beta$  atom (BrvH: M442, F443 vs. RebH: F465, W466). The biggest difference maps to the loop that covers the tryptophan binding site in RebH and contains the amino acids (Y454, Y455, E461, F465) that form hydrogen bonds with the amino and carboxy group of the substrate (Fig 7B). There are no equivalent residues in BrvH, as a short loop (residues 432–444) replaces its much longer counterpart in RebH (residues 440–467). As a consequence, the substrate binding site of BrvH is noticeably more open than that of tryptophan halogenases and the  $\epsilon$ -amino group of the active site lysine K83 is directly accessible to the solvent (S13 Fig).

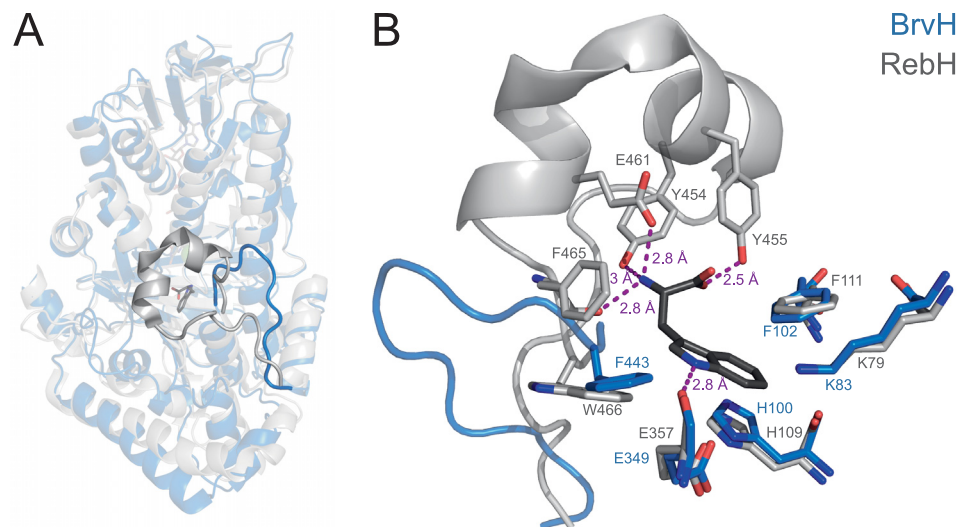
The lack of side chains to keep the amino and carboxy group of tryptophan in place could potentially explain why BrvH does not accept tryptophan as substrate. Instead, BrvH may be able to convert larger substrates. The openness of the substrate binding site and the absence of large non-polar patches—which would point to a variant B halogenase [14]—resemble the structure of MibH [59], a tryptophan halogenase that chlorinates tryptophan only when the latter is part of the enzyme’s cognate substrate peptide (deschloro NAI-107).

As for the halide specificity, no definite conclusions can be drawn. The structure does not contain FAD and there is no halide bound at the assumed halide binding site, the amide nitrogens of T351 and S352 (T359 and G360 in RebH). The loop containing the halide binding motif of BrvH structurally resembles that of other FHals and its sequence is conserved between



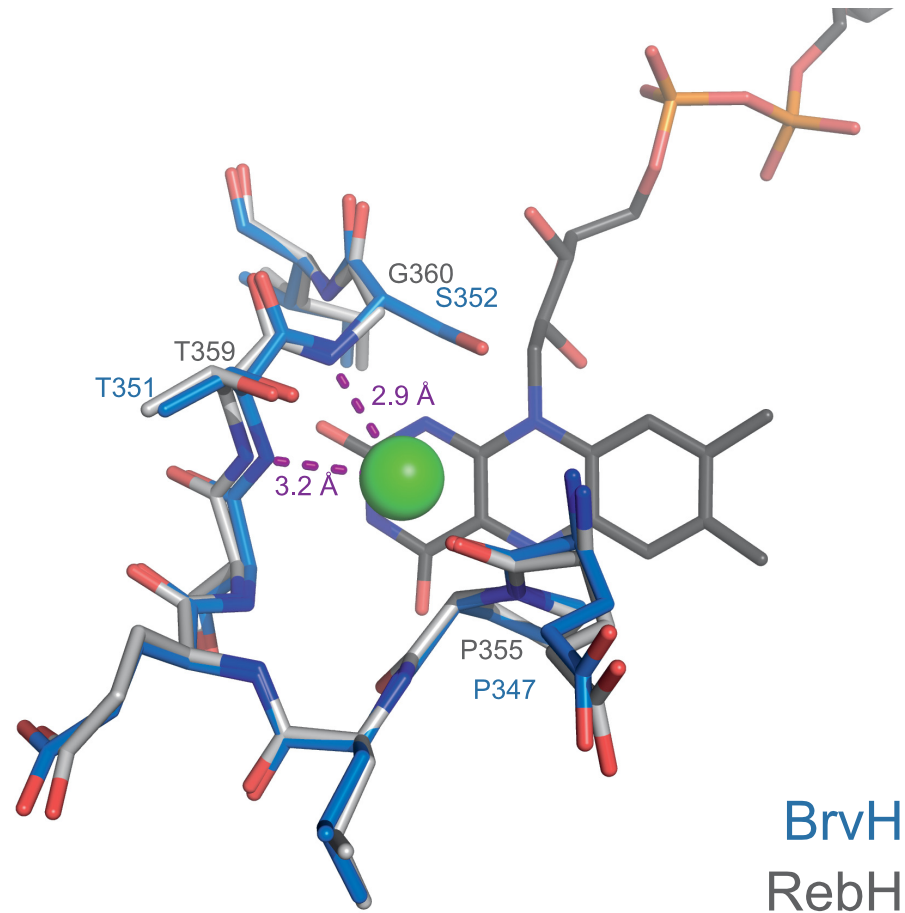
**Fig 6. HPLC traces of the halogenation of indole with BalH in excess of NaCl (100 mM) and three different concentrations of NaBr (0 mM, 10 mM, 100 mM).** Even in 10 times higher concentrations of chloride over bromide, bromide is the preferred halide source (HPLC method A).

<https://doi.org/10.1371/journal.pone.0196797.g006>



**Fig 7. Structure of apo BrvH compared to RebH bound to FAD, Trp and Cl<sup>-</sup>.** **A:** The structural alignment was made in PyMOL based on the conserved “box” domain of the two halogenases (BrvH: 6–99, 162–416, 494–502 vs. RebH: 2–98, 167–426, 519–528; PDB ID 2oa1). FAD, Trp (both dark grey) and Cl<sup>-</sup> (green) are present only in the structure of RebH. BrvH (blue) and RebH (grey) are structurally very similar. A major difference is found in the substrate binding site, which is accessible to the solvent in BrvH, while it is covered by a longer loop in RebH. **B:** Empty substrate binding site of BrvH overlaid with the tryptophan binding site of RebH containing Trp (dark grey). Several of the RebH residues directly contacting the bound Trp are shown as grey sticks along with the corresponding residues of BrvH shown in blue. The loops that are shown correspond to those in A. While the residues contacting the indole ring are conserved, those residues that form hydrogen bonds to the carboxylate and the  $\alpha$ -amino group of Trp in RebH are not present in BrvH. K83 and E349 of BrvH correspond to the amino acids shown to be important for the catalytic mechanism in other halogenases (K79 and E357 in RebH).

<https://doi.org/10.1371/journal.pone.0196797.g007>



**Fig 8. Comparison of the halide binding site in RebH and BrvH.** FAD (dark grey) and Cl<sup>-</sup> (green) are present only in the structure of RebH. Coordination of the halide takes place via the amide nitrogens of the backbone. Whilst the halide binding motif in RebH (grey) is T359, G360 in BrvH (blue) we see a T351, S352 at the same site. However, the amino acid exchange does not seem to alter the overall structure of the halide binding site, thus not giving a definite hint as to why BrvH would preferentially brominate.

<https://doi.org/10.1371/journal.pone.0196797.g008>

BrvH and RebH except for a single substitution (BrvH: S352 vs. RebH: G360; Fig 8). Thus, the halide specificity cannot readily be explained by the structure of the empty halide binding site.

## Conclusion

It is possible to identify FHals on genome level using bioinformatic analyses by employing a Hidden-Markov-model (HMM) based on the PFAM model for tryptophan halogenases. Several novel putative FHal genes were successfully identified from eleven metagenomes. From the 77 identified complete genes 42 originated from a metagenome from Botany Bay/Australia. Phylogenetic analysis based on FHals belonging to variant A and B halogenases in comparison to the identified genes from the Botany Bay metagenome revealed that the phylogenetic tree clusters in two distinct clades. One consists of the variant A halogenases including tryptophan halogenases and one of the variant B FHals. Interestingly, all predicted genes cluster within the tryptophan halogenases and seem to be phylogenetically closer to them than to variant B FHals. It is assumed that the predicted halogenase genes belong to variant A FHals and catalyse the halogenation of free substrates. Only the putative halogenase gene 38 clusters within the clade of tryptophan halogenases and might be a tryptophan halogenase. The other predicted

genes are similar to tryptophan halogenase genes, but the corresponding enzymes may accept different substrates like BrvH (gene 12). The putative halogenase gene *brvH* from *Brevundimonas* BAL3 was further investigated *in vitro*. We were able to show that the gene product forms an active enzyme and catalyses the halogenation of free indole in C3 position. Tryptophan, however, was not halogenated. This behaviour might be explained by the lack of residues that could form hydrogen bonds to the amino and carboxy group of tryptophan in the substrate binding site of BrvH. The low specific activity towards indole suggests that indole may not be the natural substrate of BrvH. In addition, the structure of the active site could allow larger substrates to enter and possibly be halogenated. Interestingly, BrvH possesses an intriguing selectivity for bromide over chloride in halogenation that, however, cannot be directly explained from the crystal structure. Bromide is the preferred halogen source even in the presence of tenfold excess of chloride to bromide. This distinguishes BrvH from other known FHals, as these enzymes show higher activities in chlorination compared to bromination with Bmp2 and Bmp5 being the only FHals that are known to brominate exclusively. Finally, we were able to show that metagenomic data analyses constitutes a powerful tool to find new biocatalysts, for example for enzymatic halogenation. Metagenomic data has the great advantage that it not only provides genomic data of cultivable, but also of non-cultivable organisms. This expands the search field for novel enzymes. At the same time, the difficulty in determining the natural substrate of an enzyme from an organism that is not well-characterised remains a challenge.

## Supporting information

**S1 Fig. SDS-PAGE (12%) analysis performed after cell disruption of *E. coli* BL21\_pGro7 with the overexpressed BrvH protein.** M: prestained proteinladder, NEB (11–245 kDa); P: pellet with insoluble proteins; S: supernatant with soluble protein fraction; before induction: sample taken from *E. coli* BL21\_pGro7 without induction with IPTG and L-arabinose. BrvH possess a mass of 56 kDa and chaperone GroEL of 60 kDa.  
(TIF)

**S2 Fig. SDS-PAGE (12%) analysis performed after cell disruption and Co-IMAC purification by His-Tag of *E. coli* BL21\_pGro7 with the overexpressed BrvH protein.** 1–8: collected fractions after Co-TALON purification and elution with 300 mM imidazole. BrvH possess a mass of 56 kDa and chaperone GroEL of 60 kDa.  
(TIF)

**S3 Fig. GC-MS analysis of the brominated indole product after 48 h of enzyme incubation.** A: gas chromatogram; B: mass spectrum of the product from  $t_R = 20.79$  min ( $[M+H]^+$  obs. 195.138 ( $^{79}\text{Br}$ ); 197.119 ( $^{81}\text{Br}$ ), calc. 194.968 ( $^{79}\text{Br}$ ); 196.968 ( $^{81}\text{Br}$ ).  
(EPS)

**S4 Fig.  $^{13}\text{C}$ -NMR (500 MHz, DMSO- $d_6$ ) spectrum of the brominated indole by BrvH.**  
(EPS)

**S5 Fig.  $^1\text{H}$ -NMR (500 MHz, DMSO- $d_6$ ) spectrum of the brominated indole by BrvH.**  
(EPS)

**S6 Fig.  $^1\text{H}$ -NMR COSY (500 MHz, DMSO- $d_6$ ) spectrum of the brominated indole by BrvH.**  
(EPS)



**S7 Fig. HMQC (500 MHz, DMSO- $d_6$ ) spectrum of the brominated indole by BrvH.**  
(EPS)

**S8 Fig. RP-HPLC analysis of BrvH catalysed bromination of indole after 0 h, 24 h and 48 h at 280 nm.** After 24 h of incubation the substrate, indole ( $t_R = 5.5$  min), was fully converted to the halogenated indole ( $t_R = 6.1$  min). Under these conditions a by-product is formed ( $t_R = 6.5$  min, HPLC method B).  
(EPS)

**S9 Fig. RP-HPLC analysis of BrvH catalysed chlorination of indole after 0 h, 24 h and 48 h at 280 nm.** After 24 h of incubation the substrate, indole ( $t_R = 5.5$  min), was converted in traces to the chlorinated indole ( $t_R = 6$  min; HPLC method B).  
(EPS)

**S10 Fig. RP-HPLC analysis of BrvH catalysed bromination of indole in presence of catalase after 0 h, 24 h and 48 h at 280 nm. Catalase was added to this enzyme assay for decomposition of hydrogen peroxide.** After 24 h of incubation the substrate, indole (5.5 min), was fully converted to the halogenated indole ( $t_R = 6.1$  min). A second by-product is catalysed ( $t_R = 6.5$  min). These results show that free hydrogen peroxide is not responsible for the halogenation (HPLC method B).  
(EPS)

**S11 Fig. RP-HPLC traces of indole incubated with BrvH\_K83A and NaBr after 0 h, 24 h and 48 h at 280 nm.** A point mutation at position 83 on amino acid level in mutant BrvH\_k83A leads to an exchange of the conserved lysine residue to alanine. This experiment revealed that the conserved lysine residue is essential for the activity of the FHal BrvH. The mutated enzyme did not halogenate indole (1.2 min; HPLC method C).  
(EPS)

**S12 Fig. UV/vis spectra of the halogenase BrvH after successful reconstitution with FAD in comparison to FAD in solution.** Upon binding of FAD to BrvH, the band at 450 nm shows a slight shift to 448 nm and some fine structure while the band around 373 nm is shifted to around 360 nm. This latter shift is caused by the changed environment of FAD inside the protein compared to the polar and strongly hydrogen-bonded environment in solution, proofing that BrvH binds FAD.  
(EPS)

**S13 Fig. Difference in the openness of the active site between BrvH und RebH. A: Surface representation of BrvH.** The structurally conserved part between different halogenases („box“) is shown in dark blue; the structurally variable „pyramid“ is shown in turquoise. B: Zoomed into the active site. The  $\epsilon$ -amino group of lysin is shown in orange (centre of the image), showing its accessibility to the solvent in the apo form with no substrate bound. C: Trp (from RebH; shown as dark grey space-filling spheres) and RebH (dark grey, cartoon representation) are superimposed on the structure of BrvH to highlight the loop covering the active site in RebH. The RebH loop shown is the same as that shown in Fig 7A. D: Surface representation of RebH. The substrate binding site is not accessible to the solvent.  
(EPS)

**S14 Fig. RP-HPLC analysis of the enzyme assay with all cofactors, compounds and indole, but without the halogenase BrvH for zero sample after 0 h, 24 h and 48 h at 280 nm.** The substrate, indole ( $t_R = 5.5$  min), is not converted after 48 h. This assay proves that the conversion of indole to bromoindole is BrvH (HPLC method B).  
(EPS)

**S1 Table. Data collection and refinement statistics of the crystallisation of BrvH.**

(PDF)

**S2 Table. Metagenomic data, which were screened with the created HMM algorithm for FHals.**

(PDF)

**S3 Table. The 42 identified putative FHal genes from Botany Bay metagenome.** The genes were identified by using our created HMM in a two-step approach based on the conserved regions of tryptophan halogenases.

(PDF)

## Acknowledgments

We thank Prof. Dr. Karl-Heinz van Pée for donating the plasmid encoding for the flavin reductase PrnF, as well as Prof. Dr. Werner Hummel for donating the plasmid encoding for the alcohol dehydrogenase. Bielefeld University is acknowledged for a PhD fellowship to P.R. N. We acknowledge support for the Article Processing Charge by the Deutsche Forschungsgemeinschaft and the Open Access Publication Fund of Bielefeld University. The experiments for structure determination were performed on beamline ID30 A-3 at the European Synchrotron Radiation Facility (ESRF), Grenoble, France. We are grateful to Cristiano de Sousa Mota at the ESRF for providing assistance in using that beamline.

## Author Contributions

**Conceptualization:** Pia R. Neubauer, Daniel Wibberg, Marcel Frese.

**Data curation:** Pia R. Neubauer.

**Funding acquisition:** Norbert Sewald.

**Investigation:** Pia R. Neubauer, Christiane Widmann, Daniel Wibberg, Lea Schröder, Hartmut H. Niemann.

**Project administration:** Norbert Sewald.

**Software:** Daniel Wibberg.

**Supervision:** Marcel Frese, Tilman Kottke, Jörn Kalinowski, Hartmut H. Niemann, Norbert Sewald.

**Validation:** Pia R. Neubauer, Christiane Widmann.

**Visualization:** Pia R. Neubauer.

**Writing – original draft:** Pia R. Neubauer, Christiane Widmann, Tilman Kottke, Hartmut H. Niemann.

**Writing – review & editing:** Daniel Wibberg, Norbert Sewald.

## References

1. Rodrigues Pereira E, Belin L, Sancelme M, Prudhomme M, Ollier M, Rapp M et al. Structure–Activity Relationships in a Series of Substituted Indolocarbazoles: Topoisomerase I and Protein Kinase C Inhibition and Antitumoral and Antimicrobial Properties. *J. Med. Chem.* 1996; 39(22):4471–4477. <https://doi.org/10.1021/jm9603779> PMID: 8893841
2. van Pée K H, Keller S, Wage T, Wynands I, Schnerr H, Zehner S. Enzymatic halogenation catalyzed via a catalytic triad and by oxidoreductases. *Biol. Chem.* 2000; 381(1):1–5. <https://doi.org/10.1515/BC.2000.001> PMID: 10722044

3. Puk O, Huber P, Bischoff D, Recktenwald J, Jung G, Süßmuth RD et al. Glycopeptide Biosynthesis in *Amycolatopsis mediterranei* DSM5908: Function of a Halogenase and a Haloperoxidase/Perhydrolase. *Chem. Biol.* 2002; 9(2):225–35. PMID: [11880037](#)
4. Ehrlich J, Bartz QR, Smith RM, Joselyn DA, Burkholder PR. Chloromycetin, a new antibiotic from a soil actinomycete. *Sci.* 1947; 106(2757):417
5. 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):316–331. <https://doi.org/10.1016/j.jmb.2010.01.020> PMID: [20080101](#)
6. Williams DH, Bardsley B. The vancomycin group of antibiotics and the fight against resistant bacteria. *Angew. Chem. Int. Ed.* 1999; 38(9):1173–1193.
7. Bischoff D, Pelzer S, Bister B, Nicholson GJ, Stockert S, Schirle M et al. The biosynthesis of vancomycin-type glycopeptide antibiotics—The order of the cyclization steps. *Angew. Chem. Int. Ed.* 2001; 40(9):1693–1696.
8. Weiss C, Sammet B, Sewald N. Recent approaches for the synthesis of modified cryptophycins. *Nat. Prod. Rep.* 2013; 30(7):924–940. <https://doi.org/10.1039/c3np70022d> PMID: [23732943](#)
9. Magarvey NA, Beck ZQ, Golakoti T, Ding Y, Huber U, Hemscheidt TK et al. Biosynthetic Characterization and Chemoenzymatic Assembly of the Cryptophycins. Potent Anticancer Agents from *Nostoc Cyanobionts*. *ACS Chem. Biol.* 2006; 1(12):766–779. <https://doi.org/10.1021/cb6004307> PMID: [17240975](#)
10. Frese M, Guzowska PH, Voss H, Sewald N. Regioselective Enzymatic Halogenation of Substituted Tryptophan Derivatives using the FAD-Dependent Halogenase RebH. *ChemCatChem* 2014; 6(5):1270–1276.
11. Schnepel C, Sewald N. Enzymatic Halogenation: A Timely Strategy for Regioselective C-H Activation. *Chem Eur J.* 2017; 23(50):12064–86. <https://doi.org/10.1002/chem.201701209> PMID: [28464370](#)
12. Frese M, Schnepel C, Minges H, Voss H, Feiner R, Sewald N. Modular Combination of Enzymatic Halogenation of Tryptophan with Suzuki-Miyaura Cross-Coupling Reactions. *ChemCatChem* 2016; 8(10):1799–1803.
13. Frese M, Sewald N. Enzymatic halogenation of tryptophan on a gram scale. *Angew. Chem. Int. Ed.* 2015; 54(1):298–301.
14. Buedenbender S, Rachid S, Müller 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. <https://doi.org/10.1016/j.jmb.2008.10.057> PMID: [19000696](#)
15. Sanchez C, Butovich IA, Brana AF, Rohr J, Mendez C, Salas JA. The biosynthetic gene cluster for the antitumor rebeccamycin: Characterization and generation of indolocarbazole derivatives. *Chem. Biol.* 2002; 9(4):519–531. PMID: [11983340](#)
16. Kirner S, Hammer PE, Hill DS, Altmann A, Fischer I, Weislo LJ et al. Functions encoded by pyrrolnitrin biosynthetic genes from *Pseudomonas fluorescens*. *J. Bacteriol.* 1998; 180(7):1939–1943. PMID: [9537395](#)
17. Keller S, Wage T, Hohaus K, Hölzer M, Eichhorn E, van Pée KH. Purification and Partial Characterization of Tryptophan 7-Halogenase (PrnA) from *Pseudomonas fluorescens*. *Angew. Chem. Int. Ed.* 2000; 39(13):2300–2302.
18. Seibold C, Schnerr H, Rumpf J, Kunzendorf A, Hatscher C, Wage T et al. A flavin-dependent tryptophan 6-halogenase and its use in modification of pyrrolnitrin biosynthesis. *Biocatal. Biotransformation* 2006; 24(6):401–408.
19. Zehner S, Kotsch A, Bister B, Süßmuth RD, Méndez 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):445–452. <https://doi.org/10.1016/j.chembiol.2005.02.005> PMID: [15850981](#)
20. van Pée K. Microbial biosynthesis of halometabolites. *Archives of Microbiology* 2001; 175(4):250–258. <https://doi.org/10.1007/s002030100263> PMID: [11382220](#)
21. Zhu X, de Laurentis W, Leang K, Herrmann J, Ihlefeld K, van Pée K et al. Structural insights into regioselectivity in the enzymatic chlorination of tryptophan. *J. Mol. Biol.* 2009; 391(1):74–85. <https://doi.org/10.1016/j.jmb.2009.06.008> PMID: [19501593](#)
22. 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):13843–13848. <https://doi.org/10.1073/pnas.0506964102> PMID: [16162666](#)
23. Zeng J, Zhan J. A Novel Fungal Flavin-Dependent Halogenase for Natural Product Biosynthesis. *ChemBioChem* 2010; 11(15):2119–23. <https://doi.org/10.1002/cbic.201000439> PMID: [20827793](#)
24. Menon Binuraj R K, Brandenburger E, Sharif HH, Klemstein U, Shepherd SA, Greaney MF et al. RadH: A Versatile Halogenase for Integration into Synthetic Pathways. *Angew. Chem. Int. Ed.* 2017; 56(39):11841–5.

25. Fraley AE, Garcia-Borràs M, Tripathi A, Khare D, Mercado-Marin EV, Tran H et al. Function and Structure of MalA/MalA', Iterative Halogenases for Late-Stage C-H Functionalization of Indole Alkaloids. *J. Am. Chem. Soc.* 2017; 139(34):12060–8. <https://doi.org/10.1021/jacs.7b06773> PMID: 28777910
26. Dong C, Flecks S, Unversucht S, Haupt C, van Pée K, Naismith JH. Tryptophan 7-Halogenase (PrnA) Structure Suggests a Mechanism for Regioselective Chlorination. *Sci* 2005; 309(5744):2216–2219.
27. Yeh E, Blasiak LC, Koglin A, Drennan CL, Walsh CT. Chlorination by a Long-Lived Intermediate in the Mechanism of Flavin-Dependent Halogenases. *Biochemistry* 2007; 46(5):1284–92. <https://doi.org/10.1021/bi0621213> PMID: 17260957
28. Flecks S, Patallo EP, Zhu X, Ernyei AJ, Seifert G, Schneider A et al. New insights into the mechanism of enzymatic chlorination of tryptophan. *Angew. Chem. Int. Ed.* 2008; 47(49):9533–9536.
29. Wierenga RK, Drenth J, Schulz GE, Huber R. Comparison of the three-dimensional protein and nucleotide structure of the FAD-binding domain of p-hydroxybenzoate hydroxylase with the FAD- as well as NADPH-binding domains of glutathione reductase. *J. Mol. Biol.* 1983; 167(3):725–39. PMID: 6876163
30. Dym O, Eisenberg D. Sequence-structure analysis of FAD-containing proteins. *Protein Science* 2001; 10(9):1712–28. <https://doi.org/10.1110/ps.12801> PMID: 11514662
31. Schnepel C, Minges H, Frese M, Sewald N. A High-Throughput Fluorescence Assay to Determine the Activity of Tryptophan Halogenases. *Angew. Chem. Int. Ed.* 2016; 55(45): 14159–14163.
32. Shepherd SA, Menon Binuraj R. K., Fisk H, Struck A, Levy C, Leys D et al. A Structure-Guided Switch in the Regioselectivity of a Tryptophan Halogenase. *ChemBioChem* 2016; 17(9):821–824. <https://doi.org/10.1002/cbic.201600051> PMID: 26840773
33. Smith DRM, Uria AR, Helfrich EJM, Milbredt D, van Pée KH, Piel J et al. An Unusual Flavin-Dependent Halogenase from the Metagenome of the Marine Sponge *Theonella swinhoei* WA. *ACS Chem. Biol* 2017 12 (5): 1281–1287. <https://doi.org/10.1021/acscchembio.6b01115> PMID: 28198609
34. Ausec L, Zakrzewski M, Goesmann A, Schlüter A, Mandic-Mulec I. Bioinformatic analysis reveals high diversity of bacterial genes for laccase-like enzymes. *PloS ONE* 2011; 6(10):e25724. <https://doi.org/10.1371/journal.pone.0025724> PMID: 22022440
35. Stoeveken J, Singh R, Kolkenbrock S, Zakrzewski M, Wibberg D, Eikmeyer FG et al. Successful heterologous expression of a novel chitinase identified by sequence analyses of the metagenome from a chitin-enriched soil sample. *J. Biotechnol* 2015; 201:60–68. <https://doi.org/10.1016/j.jbiotec.2014.09.010> PMID: 25240439
36. Castro P, Huber ME. *Marine biology*. 9th ed. New York, NY: McGraw Hill; 2013.
37. Tasdemir D, Mallon R, Greenstein M, Feldberg LR, Kim SC, Collins K et al. Aldisine Alkaloids from the Philippine Sponge *Stylissa massa* Are Potent Inhibitors of Mitogen-Activated Protein Kinase Kinase-1 (MEK-1). *J. Med. Chem.* 2002; 45(2):529–532. PMID: 11784156
38. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J. Mol. Biol.* 1990; 215(3):403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2) PMID: 2231712
39. Robert CE. MUSCLE: multiple sequence alignment with high accuracy and high throughput: Oxford University Press. *Nucleic Acids Res.* 2004 32(5):1792–1797. <https://doi.org/10.1093/nar/gkh340> PMID: 15034147
40. Mistry J, Finn RD, Eddy SR, Bateman A, Punta M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.* 2013; 41(12).
41. Li D, Liu C, Luo R, Sadakane K, Lam T. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 2015; 31(10):1674–1676. <https://doi.org/10.1093/bioinformatics/btv033> PMID: 25609793
42. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *Bioinformatics* 2010; 11: 119. <https://doi.org/10.1186/1471-2105-11-119> PMID: 20211023
43. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* 2016; 33(7):1870–1874. <https://doi.org/10.1093/molbev/msw054> PMID: 27004904
44. Kabsch W. Integration, scaling, space-group assignment and post-refinement. *Acta Cryst. D* 2010; 66(2):133–44.
45. Evans PR, Murshudov GN. How good are my data and what is the resolution? *Acta Cryst. D* 2013; 69(7):1204–14.
46. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. *J. Appl. Cryst.* 2007; 40(4):658–74.

47. Terwilliger TC, Grosse-Kunstleve RW, Afonine PV, Moriarty NW, Zwart PH, Hung L et al. Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Cryst. D* 2008; 64(1):61–9.
48. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Cryst. D* 2010; 66(4):486–501.
49. Murshudov GN, Vagin AA, Dodson EJ. Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Cryst. D* 1997; 53(3):240–55.
50. Chen VB, Arendall W. Bryan III, Headd JJ, Keedy DA, Immormino RM, Kapral GJ et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Cryst. D* 2010; 66(1):12–21.
51. Schrödinger L. PyMOL The PyMOL Molecular Graphics System: Version; 2010.
52. Weichold V, Milbredt D, van Pée K. Specific Enzymatic Halogenation-From the Discovery of Halogenated Enzymes to Their Applications In Vitro and In Vivo. *Angew. Chem. Int. Ed.* 2016; 55(22):6374–89.
53. Weber T, Blin K, Duddela S, Krug D, Kim HU, Brucoleri R et al. antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res.* 2015; 43(W1): W237–43. <https://doi.org/10.1093/nar/gkv437> PMID: 25948579
54. Agarwal V, El Gamal Abraham A, Yamanaka K, Poth D, Kersten RD, Schorn M et al. Biosynthesis of polybrominated aromatic organic compounds by marine bacteria. *Nat. Chem. Biol* 2014; 10(8):640–7. <https://doi.org/10.1038/nchembio.1564> PMID: 24974229
55. Wiesner W, van Pée K H, Lingens F. Purification and characterization of a novel bacterial non-heme chloroperoxidase from *Pseudomonas pyrrocinia*. *J Biol Chem* 1988; 263(27):13725–13732. PMID: 3417677
56. Beers RF., Sizer IW. A spectrophotometric method for measuring the breakdown of hydrogen peroxide by catalase. *J Biol Chem* 1952; 195(1):133–140. PMID: 14938361
57. Kotaki A, Naoi M, Yagi K. Effect of Proton Donors on the Absorption Spectrum of Flavin Compounds in Apolar Media. *J. Biochem.* 1970; 68(3):287–92. PMID: 5472345
58. Raffelberg S, Gutt A, Gärtner W, Mandalari C, Abbruzzetti S, Viappiani C et al. The amino acids surrounding the flavin 7a-methyl group determine the UVA spectral features of a LOV protein. *Biol. Chem.* 2013; 394(11):1517–28. <https://doi.org/10.1515/hsz-2013-0163> PMID: 23828427
59. Ortega MA, Cogan DP, Mukherjee S, Garg N, Li B, Thibodeaux GN et al. Two Flavoenzymes Catalyze the Post-Translational Generation of 5-Chlorotryptophan and 2-Aminovinyl-Cysteine during NAI-107 Biosynthesis. *ACS Chem. Biol.* 2017; 12(2):548–57. <https://doi.org/10.1021/acscchembio.6b01031> PMID: 28032983