




Genomic Determinants Encode the Reactivity and Regioselectivity of Flavin-Dependent Halogenases in Bacterial Genomes and Metagenomes

Jehyun Jeon,^a Jaehee Lee,^b Se-Min Jung,^b Jae Hong Shin,^a  Woon Ju Song,^b  Mina Rho^{a,c}

^aDepartment of Computer Science, Hanyang University, Seoul, Republic of Korea

^bDepartment of Chemistry, Seoul National University, Seoul, Republic of Korea

^cDepartment of Biomedical Informatics, Hanyang University, Seoul, Republic of Korea

Jehyun Jeon and Jaehee Lee contributed equally to this work. Author order was determined by mutual consent.

ABSTRACT Halogenases create diverse natural products by utilizing halide ions and are of great interest in the synthesis of potential pharmaceuticals and agrochemicals. An increasing number of halogenases discovered in microorganisms are annotated as flavin-dependent halogenases (FDHs), but their chemical reactivities are markedly different and the genomic contents associated with such functional distinction have not been revealed yet. Even though the reactivity and regioselectivity of FDHs are essential in the halogenation activity, these FDHs are annotated inaccurately in the protein sequence repositories without characterizing their functional activities. We carried out a comprehensive sequence analysis and biochemical characterization of FDHs. Using a probabilistic model that we built in this study, FDHs were discovered from 2,787 bacterial genomes and 17 sediment metagenomes. We analyzed the essential genomic determinants that are responsible for substrate binding and subsequent reactions: four flavin adenine dinucleotide-binding, one halide-binding, and four tryptophan-binding sites. Compared with previous studies, our study utilizes large-scale genomic information to propose a comprehensive set of sequence motifs that are related to the active sites and regioselectivity. We reveal that the genomic patterns and phylogenetic locations of the FDHs determine the enzymatic reactivities, which was experimentally validated in terms of the substrate scope and regioselectivity. A large portion of publicly available FDHs needs to be reevaluated to designate their correct functions. Our genomic models establish comprehensive links among genotypic information, reactivity, and regioselectivity of FDHs, thereby laying an important foundation for future discovery and classification of novel FDHs.

IMPORTANCE Halogenases are playing an important role as tailoring enzymes in biosynthetic pathways. Flavin-dependent tryptophan halogenases (Trp-FDHs) are among the enzymes that have broad substrate scope and high selectivity. From bacterial genomes and metagenomes, we found highly diverse halogenase sequences by using a well-trained profile hidden Markov model built from the experimentally validated halogenases. The characterization of genotype, steady-state activity, substrate scope, and regioselectivity has established comprehensive links between the information encoded in the genomic sequence and reactivity of FDHs reported here. By constructing models for accurate and detailed sequence markers, our work should guide future discovery and classification of novel FDHs.

KEYWORDS halogenase, active sites, regioselectivity, profile hidden Markov model, genomic patterns, phylogeny, mutagenesis

Citation Jeon J, Lee J, Jung S-M, Shin JH, Song WJ, Rho M. 2021. Genomic determinants encode the reactivity and regioselectivity of flavin-dependent halogenases in bacterial genomes and metagenomes. *mSystems* 6:e00053-21. <https://doi.org/10.1128/mSystems.00053-21>.

Editor Marnix Medema, Wageningen University

Copyright © 2021 Jeon et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Woon Ju Song, woonjusong@snu.ac.kr, or Mina Rho, minarho@hanyang.ac.kr.

Received 19 January 2021

Accepted 3 May 2021

Published 27 May 2021

Halogen-containing molecules are ubiquitous in the environment. An increasing number of organohalogen products have been isolated and characterized in recent decades (1). Organic halides produced as secondary metabolites are considered potential pharmaceuticals and agrochemicals, constituting an important class of antibiotics and antitumor drugs represented by rebeccamycin, chloramphenicol, and vancomycin derivatives (2). Because dehalogenation of these molecules often lowers binding affinity to the target (3) and biological activities (4), halogenation is one of the essential reactions in the functionalization of natural products. Halogenated molecules can also engage in specific noncovalent interactions with heteroatom-containing functional groups or aromatic rings by halogen bonds (5). The highly directional and specific halogen bonds have been exploited to enhance ligand binding affinities of drugs and inhibitors, with minimal disruption of other important intermolecular interactions (6).

Organohalogens are often generated by halogenases that belong to biosynthetic gene clusters. Depending on their native cofactors, cosubstrates, and reaction mechanisms, halogenases are categorized into four to five classes (2, 7–10). Among these, flavin-dependent halogenases (FDHs) have been extensively investigated due to their broad substrate scope and high selectivity. In particular, regioselective formation of the carbon-halogen bond represents an important class of reactions that produce aromatic compounds for feedstock. Consequently, artificial FDHs have emerged as viable surrogates to toxic halogenating chemical reagents (11–16).

Recently, numerous FDHs have been discovered and characterized in several directions. Sequence analysis of various FDHs were focused primarily on four strictly conserved motifs of the FAD-binding domain (17–19). Diverse halogenating activities of FDHs were investigated experimentally instead, resulting in FDHs that react with tryptophan (Trp-FDHs) and other aromatic functional groups, such as tyrosinyl (20), tryptophanyl (21, 22), and pyrrolyl moieties (23, 24). Phenolic halogenases, such as Rdc2 (25, 26), Gsfl (26, 27), and PltM (28), on the other hand, are reactive toward phenols and anilines. In addition, extensive activity profiling has been further carried out, expanding the substrate repertoire of FDHs (19, 29). Recent studies suggested that Trp-FDH enzymes have a subtle yet discrete substrate scope, revealing that not all react with tryptophan. MibH from *Microbispora corallina* (30) reacts with a lanthipeptide bearing a tryptophan moiety but not directly with tryptophan. BrvH from *Brevundimonas* sp. strain BAL3 (31) is inactive toward tryptophan but halogenates indole. Krml from *Theonella swinhoei* WA (22) reacts with tryptophan but not with indole.

In addition to the substrate scope, regioselectivity is another key element in the identification of Trp-FDH reactivities. Based on the site of halogenation on the indole ring, Trp-FDHs can be further categorized into 5-, 6-, and 7-tryptophan halogenases (5-, 6-, and 7-Trp-FDHs, respectively) (7, 32, 33). Previous studies demonstrated that the regioselectivity of carbon-halogen bond formation is determined by the relative positioning of the substrate with respect to the catalytically critical lysine residue at the active site in conjunction with a glutamate located between the flavin adenine dinucleotide (FAD)-binding and tryptophan-binding domains (34, 35). Additionally, site-directed mutagenesis and directed evolution studies revealed that tryptophan-binding sites are critical to determine the substrate scope and regioselectivity of selected enzymes (16, 32, 33, 36–39).

The reactivity and regioselectivity of FDH are essential in halogenation activity and determined by the active sites that recognize aromatic substrates. However, no definitive set of genomic markers was analyzed from the large-scale sequence analysis to distinguish their reactivities at atomic accuracy. This is presumably due to the high variability in the amino acid composition of the substrate-binding pocket (40), and FDH sequence motifs that underpin such functional diversity are yet to be elucidated (41).

Recent studies have shown that whole genomes as well as metagenomes can serve as an immense source of halogenase genes. Using PCR-based screening, Liao et al. found halogenase genes from *Streptomyces* and *Nocardiopsis* in Arctic marine actinomycete isolates (42). From forest soil microbiome and 11 Botany Bay metagenomes, Weigold et al. (43) and

Neubauer et al. (31) discovered putative FDHs. The increasing amount of metagenome data provides more opportunities to find novel halogenases with different substrate specificities.

Here, we report a comprehensive profiling of FDHs from bacterial genomes and metagenomes. A well-trained profile hidden Markov model (pHMM), built from experimentally validated halogenases, allowed us to find highly diverse halogenase genes from bacterial genomes and metagenomes. To discover remotely homologous genes, our sequence models include information for conserved class-specific cofactors and binding sites. Structural analysis on FDHs was integrated to identify key genomic determinants that are responsible for substrate binding and subsequent reactions. Using the isolated bacterial and metagenomic halogenases, we determined the steady-state activity, substrate scope, and regioselectivity of FDHs, thereby establishing comprehensive links between genotypic information and functional sites. Our work thus lays an important foundation for the future discovery and classification of novel FDHs by constructing accurate and detailed sequence models.

RESULTS

Genomic patterns of halogenases in bacterial genomes and metagenomes. We collected the experimentally validated FDH sequences to build a pHMM and searched homologous genes from 2,787 bacterial genomes and 17 sediment microbiomes (see Table S1 in the supplemental material). In total, 103 and 68 halogenases were identified using the pHMM search from the bacterial genomes and metagenomes, respectively (see Fig. S1). The 103 halogenases from bacterial genomes were further characterized into two groups: 20 located in the biosynthetic gene clusters (BGCs) and 83 that were not. Additionally, 109 halogenases were obtained from the NCBI protein sequence database by keyword and pHMM search.

To investigate the genomic patterns of halogenases, the 280 putative and 33 known halogenases were represented on embedding space (Fig. 1a). The amino acid conservation against the profiles of six different halogenase groups, 5-, 6-, and 7-Trp-FDHs, indole-Hals, phenolic-Hals, and pyrrole-Hals, was used to calculate the distance among halogenase genes using *t*-distributed stochastic neighbor embedding (t-SNE). Halogenases represented as the points were grouped according to their substrate types and scopes when the known halogenases were compared. Especially, Trp-FDHs were clustered into four groups, 7-Trp-FDH (A1), 6-Trp-FDH (A2), 6-Trp-FDH (B2), and 5-Trp-FDH (B1), implying that the regioselectivity is reflected in the embedding space. Interestingly, 6-Trp-FDHs in the B2 group, including KtzR (44), SttH (45), Tar14 (46), and thermophilic Hal (Th-Hal) (47), were relatively close to 5-Trp-FDHs, while 6-Trp-FDHs in the A2 group, including AORI_5336 (48), BorH (49), FmoD (50), and Thal (51), were close to 7-Trp-FDHs. Halogenases identified from marine sediments were clustered with most of the indole halogenases such as BrvH (31), Krml (22), MibH (30), VirX1 (52), Xcc-b100-1333, Xcc-b100-4156 (53), and Xcc-b100-4345 (54). Phenolic and pyrrole halogenases were clustered, which are apart from Trp-FDHs.

Similar patterns were also observed in the phylogenetic tree that was constructed with the same set of Trp-FDH and non-Trp-FDH sequences (Fig. 1b). The FDHs were grouped into three discrete groups, A, B, and C, in the tree. Consistently, the 17 known Trp-FDHs were clustered in groups A and B, whereas non-Trp-FDHs, such as MibH, BrvH, and Krml, were found in group C, showing that the substrate scope is associated with the branch grouping. Regioselectivity appeared to be represented phylogenetically: 7-Trp-FDHs and 5-Trp-FDHs were found in groups A and B, respectively. Interestingly, 6-Trp-FDHs were found in both groups A and B, which was also observed in the embedding space in Fig. 1a.

Group A was further divided into two subgroups, A1 and A2, in the phylogenetic tree (Fig. 1b), which correspond to two close groups in the embedding space (Fig. 1a, bottom). Here, 7-Trp-FDHs, including pyrrolnitrin halogenase (PrnA) from *Pseudomonas fluorescens* (55), were found in group A1, and 6-Trp-FDHs, including ThdH (Thal) from *Streptomyces albobriseolus* (51), ThnH from *Streptomyces* sp. FXJ1.172 (56), BorH (49),

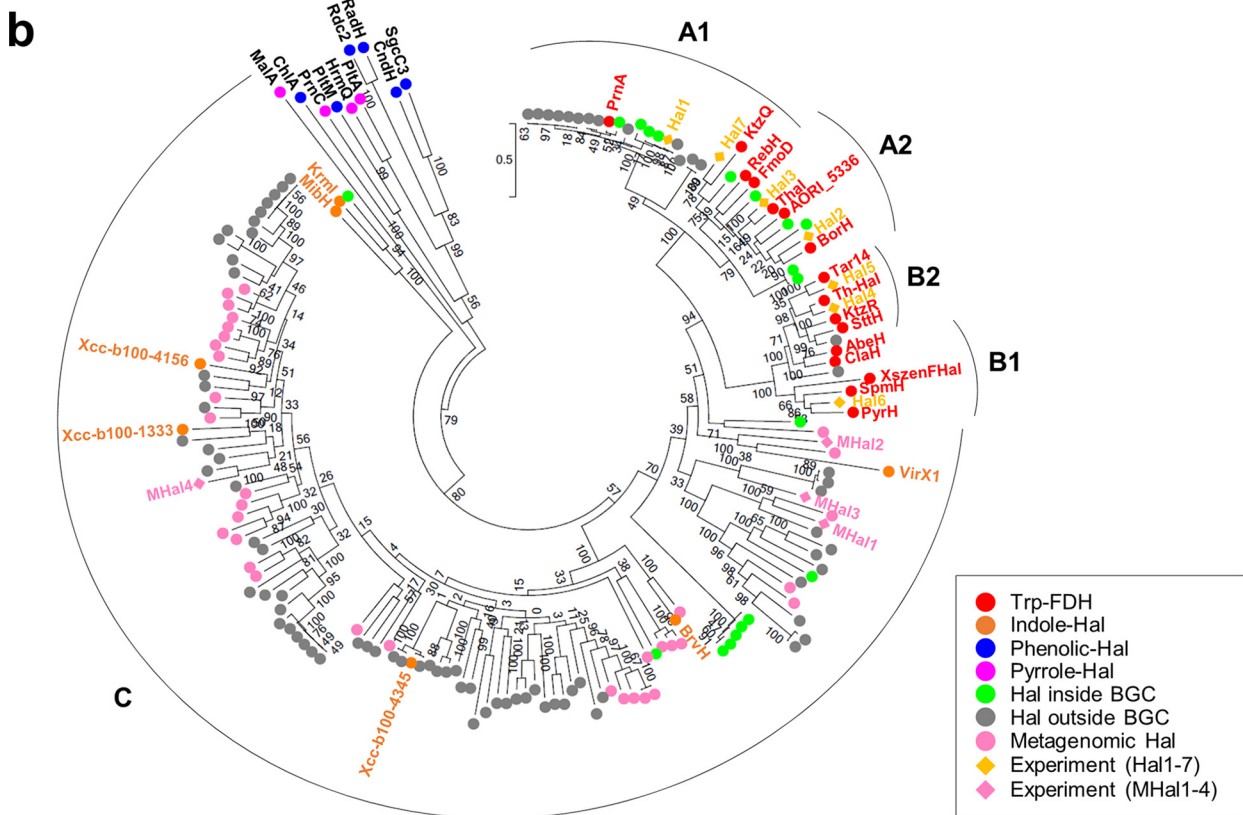
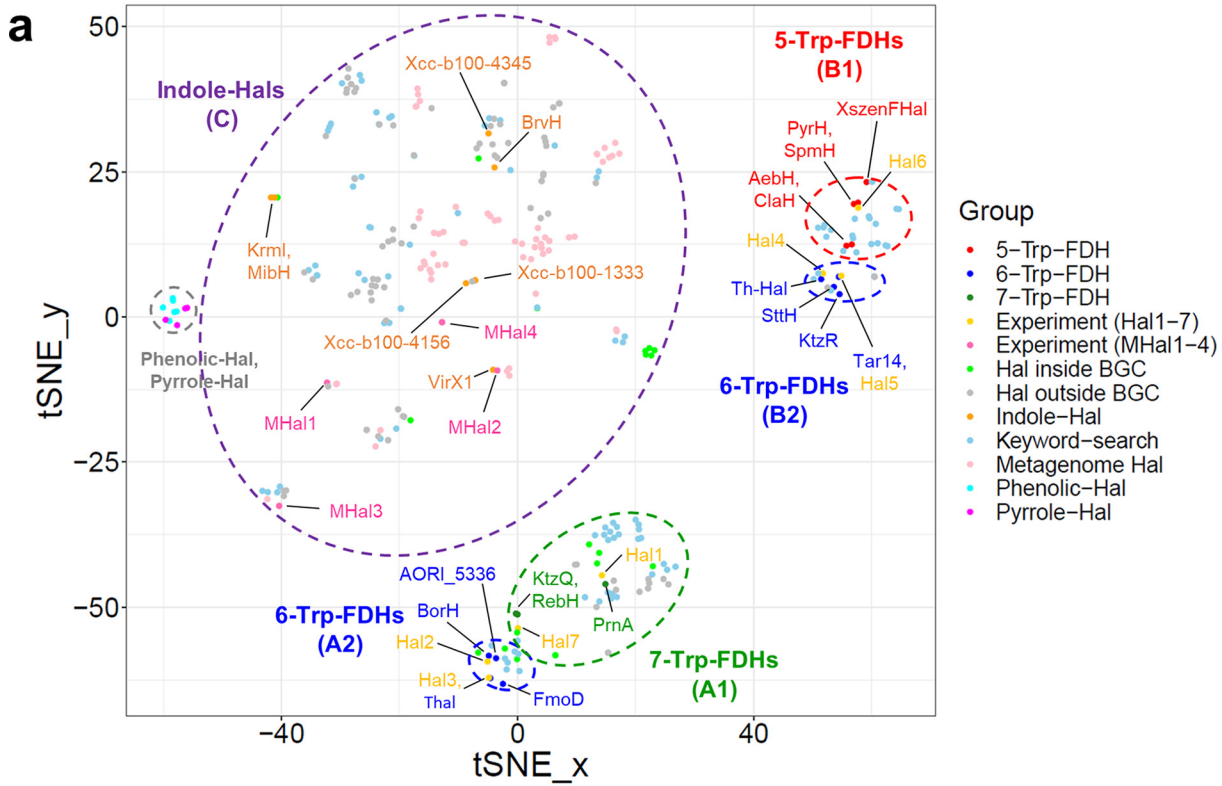


FIG 1 Known and putative halogenase sequences represented using t-SNE and a phylogenetic tree. (a) Embedding of the putative halogenases identified from the bacterial genomes and marine metagenomes, in conjunction with the known halogenases based on genomic patterns. (b) Phylogenetic tree of the halogenases.

and AORI_5336 (48), were found in group A2. Even though FmoD (50) was closely clustered with other 6-Trp-FDHs in the embedding space, it showed an ambiguous placement with KtzQ (44) and rebeccamycin halogenase (RebH) (57) of 7-Trp-FDHs in group A in the phylogenetic tree.

Group B was also divided into two subgroups, B1 and B2, in the phylogenetic tree. All 5-Trp-FDH genes, such as those encoding ClaH (58), AbeH (59), SpmH (60), XszenFHal (61), and PyrH (18), were included in group B1, whereas 6-Trp-FDH genes, such as those encoding Th-Hal (47), KtzR (44), Tar14 (46), and SttH (45), were in group B2. Therefore, Trp-FDHs showed distinctive sequence conservation patterns that impart regioselectivity to tryptophan. Even though ClaH (58) and AbeH (59) ambiguously exist between branch B1 and B2 in the phylogenetic tree, they were clustered more closely with 5-Trp-FDH genes in the embedding space according to the classification method.

Structure-guided analysis of FAD- and halide-binding regions in FDHs. To elucidate protein sequence-structure-activity relationships among FDHs, conserved regions were investigated. We screened 3-mers (i.e., sequences of three amino acids) using a multiple sequence alignment of 313 sequences (see Fig. S2A). Regions of low entropy were investigated using X-ray crystal structures of FDHs that are complexed with FAD, chloride, and tryptophan (Fig. 2a to f; Fig. S2B to H). We defined substrate-binding sites as the residues that make direct contacts with the molecule. As a result, 10 putative functional regions, FAD1 to -4, Lys (K), halide, and Trp1 to -4, were determined as well-conserved regions and subjected to further investigation (Fig. 2g).

We found four highly conserved regions related to FAD-binding (FAD1 to -4), as reported previously (18). We noted that FAD1 and FAD2 regions are consistently conserved in all identified FDHs in groups A to C. In contrast, FAD3 and -4 are less conserved than FAD1 and -2, exhibiting slight sequence alterations among groups. However, it is unclear whether the sequence variation in FAD3 and -4 is directly related to regioselectivity, as FAD3 and FAD4 are located remotely from the tryptophan-binding motifs.

The halide-binding site is also highly conserved in the sequences and structures of genomic and metagenomic FDHs. A large number of Trp-FDHs, including RebH (32, 57) and PyrH (18), react with both Cl^- and Br^- ; the yields are either comparable to or slightly higher with Cl^- . In contrast, the recently identified halogenase BrvH exhibits catalytic reactivity only with Br^- (31), indicating that the halogenation activity is less likely determined by the halide-binding motif alone. Because the activated halogenating species, such as hypochlorite (ClO^-) or hypobromite (BrO^-), are likely to transfer from the halide-binding sites to tryptophan-binding regions, unidentified residues that are associated with the dynamic process might be related to halogen-specific reactivity.

Identification of tryptophan-binding regions in Trp-FDHs. In addition to the FAD- and halide-binding sites, the interactions with tryptophan are essential for halogenation (Fig. 2a to f). We determined Trp1 to -4 regions with low entropy of amino acid sequences (Fig. 2g): Trp1 (amino acids [aa] 50 to 55 in PrnA), Trp2 (aa 101 to 105 in PrnA), Trp3 (aa 160 to 163 in PyrH), and Trp4 (aa 443 to 461 in PrnA). Notably, different groups showed different patterns of tryptophan-binding regions. Trp1 and Trp2 motifs are found in all genes of groups A and B, whereas Trp3 is present only in group B, and Trp4 is present differentially in groups A and B. The Trp4 motif was not observed in group C.

For halogenases that have different regioselectivity for tryptophan, the Trp1 to -4 regions show considerable sequence variations. In particular, the third and fourth positions in Trp1 have conservation with Ile (I), Val (V), and Phe (F) at 63%, 15%, and 14%, respectively, followed by Pro (P) and Ser (S) with 84% and 13%, respectively. They exhibit considerably high conservation within each subgroup, indicating that the regioselectivity could be related to the Trp1 motif. Structural analysis showed that the Trp1 motif undergoes considerable structural changes upon tryptophan binding (62). In particular, the third through fifth residues are involved in hydrophobic and hydrogen-bonding interactions with tryptophan (Fig. 2a to f), suggesting that Trp1 determines the regioselectivity (see the experiment results in the next section).

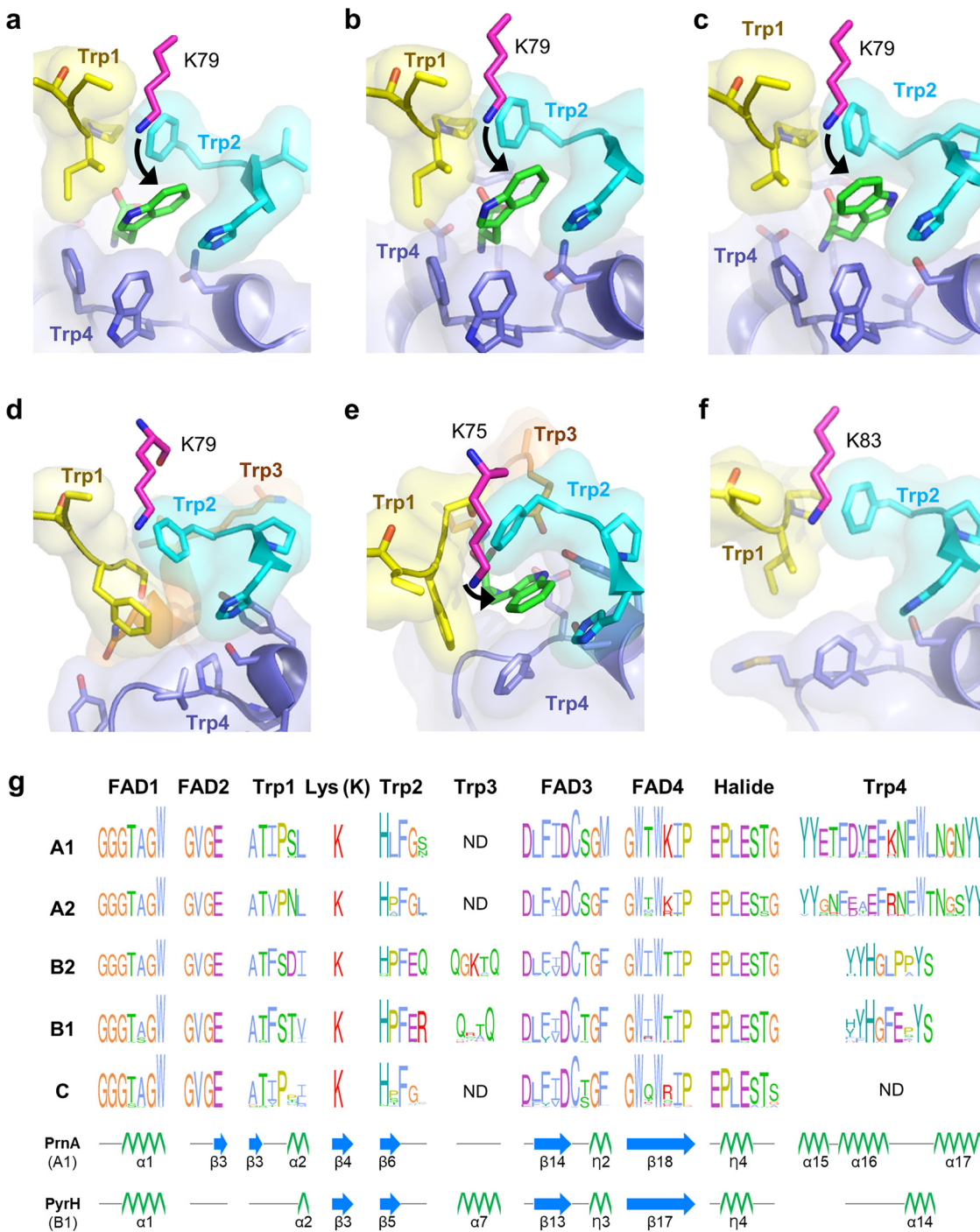


FIG 2 Integrated structure and sequence analysis of tryptophan and nontryptophan halogenases. Tryptophan-binding sites determined in this study are denoted for each group in Fig. 1. (a) PrnA in A1 (2AQJ); (b) RebH in A1 (2OA1); (c) Thal in A2 (6H44); (d) SttH in B2 (5HYS); (e) PyrH in B1 (2WET); (f) BrvH in C (6FRL). PDB codes are in parentheses. Residues comprising Trp1 to -4 are colored yellow, cyan, orange, and purple, respectively. Tryptophan molecules and catalytic lysine residues are shown in green and magenta bars, respectively. (g) Multiple-sequence alignment and conservation sites of FAD1 to -4, Lys (K), Trp1 to -4, and halide sites of FDHs. Secondary structure for each motif site is represented as α -helices, β -sheets, and loops using PrnA and PyrH.

The Trp2 region also exhibits high sequence conservation. Here, hydrophobic residues constitute the first and third positions with 95% and 98% conservation, respectively, and engage in π - π stacking interactions with the indole ring portion of tryptophan. Although the side chain of the second residue in the HxF motif is not directed toward the substrate, noticeable sequence conservation is observed depending on phylogeny: Leu (L) in branch

A1, and Pro (P) in branches A2, B1, and B2 (Fig. 2g). These observations imply that the least-conserved second residue in Trp2 might relate to altered orientations of the hydrophobic residues for π - π stacking interactions with tryptophan. Indeed, hydrogen-bonding interactions are observed between the indole N-H and the carbonyl group of Pro in the Trp2 motif in PyrH, suggesting its potential roles in regioselectivity (see the experiment results in the next section).

The Trp3 region occurs only in branch B, whereas variations are observed in a number of nonconserved residues between two Gln (Q)s as QxxQ and QxxxQ in groups B1 and B2, respectively (Fig. 2g). The enzyme PyrH belongs to group B1 and exhibits hydrogen bonds between tryptophan and the residues in Trp3, along with Trp1, -2, and Trp4 motifs (Fig. 2e). Notably, the residues positioned ahead of Trp3 (A153 to A158 in PyrH) become ordered upon complexation with tryptophan, indicating that these regions interact during substrate binding (62). Such a Trp3 region of PyrH (Q160 to Q163) exists in the alpha helix (α 7) (Fig. 2g), which follows a longer loop than that observed in PrnA of group A. The side chains of Q160 and Q163 in PyrH interact with the carboxylate oxygen atom of tryptophan, and they are 3.21 and 2.88 Å apart from tryptophan, respectively (62). While Trp3 regions were not observed in the halogenases in group A, a strong signal of Trp4 was observed in alpha-helix regions (e.g., α 15, α 16, and α 17 in PrnA; α 14, α 15, and α 16 in Thal), which plays an important role in substrate binding. To verify the Trp3 functional site, we performed experimental characterizations (see the experiment results in the next section).

The region Trp4 shows high sequence and structural variations relative to Trp1 and Trp2. The Trp4 motif is assigned as YYxxF(E|D)(A|Y)EF(R|K)NFW(L|T)N(G|S)(N|S)YY in group A and (Y|H)YHG(L|F)(P|E)PYS in group B, whereas no analogous motif is found in group C (Fig. 2g). The structures of Trp-FDHs indicate that Trp4 motifs form enzyme-substrate interactions with tryptophan, which suggests their critical roles for substrate binding (Fig. 2a to e). This observation is consistent with the altered substrate scope and regioselectivity upon mutation of Trp4 motifs (33, 37, 39, 63). The structure of BrvH (31) is in good agreement with the observation that halogenases lacking the Trp4 motifs do not interact with tryptophan. As such, it is likely that halogenases in group C might have a different substrate scope than those in groups A and B.

Characterization of putative Trp-FDHs. To validate the classification of FDHs, we isolated and characterized 11 putative halogenases: 7 bacterial genomic FDHs (Hal1 to -7) and 4 metagenomic FDHs (MHal1 to -4) (Fig. 1, Table 1, and Table S2). One or two sequences were randomly selected from each phylogenetic group (groups A to C). Both Hal1 and Hal7 are located close to 7-Trp-FDHs, such as PrnA and KtzQ, respectively, whereas Hal7 is located slightly apart from the majority of 7-Trp-FDHs. This pattern is also observed in the phylogenetic tree (Fig. 1b). In contrast, Hal2 and Hal3 are tightly clustered with multiple 6-Trp-FDHs in the embedding space and reside in group A2 with 6-Trp-FDHs such as ThdH (Thal), BorH, and AORI_5336 (Fig. 1). Notably, Hal3, which was previously named ThnH and reported as a part of the gene cluster for thienodolin synthesis (56), is clustered with a 6-Trp-FDH. Hal4 and Hal5 are tightly clustered with multiple 6-Trp-FDHs in the embedding space, which are close to 6-Trp-FDHs and belong to group B2 with 6-Trp-FDHs such as SttH, KtzR, and Th-Hal (Fig. 1). Lastly, Hal6 is close to 5-Trp-FDHs in the embedding space and is in group B1 with 5-Trp-FDHs such as PyrH, SpmH, and XszenFHal in the tree (Fig. 1). Notably, the selected enzymes possess strictly conserved residues involved in FAD and halide binding, whereas Hal1 to -7 genes exhibit the representative Trp1 to -4 motifs of each group shown in the conservation graph (Fig. 2g and Table 1). MHal1 to -4 exhibit an analogous Trp1 and -2 motif but no Trp4 motif.

The putative FDHs that we identified were heterologously expressed in *Escherichia coli* using the methods described in reference 32. We measured steady-state activities of putative Trp-FDHs from bacterial genomes (Hal1 to -7). The disappearance of the substrate, tryptophan, and the concurrent formation of new peaks were observed for all reactions (see Fig. S3). The products were analyzed by liquid chromatography mass

TABLE 1 Experimentally characterized FDHs^a

Group	Enzyme	Regioselectivity for Trp	Trp1	Trp2	Trp3	Trp4
A1	PrnA	7	ATIPSL	HLFGN		YYETFDFFKFNFWLNGNYY
	Hal1	7	ATIPSL	HLFGN		YYETFDFFKFNFWLNGNYY
	KtzQ	7	ATIPNL	HLFGQ		YYGNFEAEFRNFWTNSNYY
	RebH	7	ATIPNL	HSFGL		YYGNFEAEFRNFWNNSNYY
	Hal7	7	ATIPNL	HLFGL		YYGNFEAEFRNFWSNANYY
A2	ThdH (Thal)	6	ATVPNL	HPFGL		YYGNFEAEFRNFWTNGSYY
	FmoD	6	ATVPNL	HPFGL		YYGNFEAEFRNFWTNGSYY
	Hal3 (ThnH)	6	ATIPNL	HPFGL		YYGNFEAEFRNFWTNGSYY
	BorH	6	ATVPNL	HPFGL		YYGRFEAEFRNFWTNGSYY
	AORI_5336	6	ATVPNL	HPFGL		YYGNFDAEFRNFWTNGSYY
	Hal2	6	ATVPNL	HPFGL		YYDNFDVEFRNFWTNGSYY
B2	Thermophilic Hal	6	ATFSDI	HPFEQ	QGKTQ	YYHGLPPYS
	KtzR	6	ATFSDI	QPFEQ	QGKSQ	YYHGLPPYS
	SttH	6	ATFSDI	HPFEQ	QGKTQ	YYHGLPPYS
	Tar14	6	ATFSDI	HPFEQ	QGAQ	YYHGLPPYS
	Hal4	6	ATFSDI	HPFEQ	QGKTQ	YYHGLPAYS
	Hal5	6	ATFSDI	HPFEQ	QGTQ	HYHGLPPYS
B1	ClaH	5	ATFSTV	HPFER	QGTQ	HYHGFAYS
	AbeH	5	ATFSTV	HPFER	QTTQ	HYHGFESYS
	PyrH	5	ATFSTV	HPFER	QRAQ	YYHGFETYS
	SpmH	5	ATFSTV	HPFER	QRAQ	YYHGFEPYS
	XszenFHal	5	ATFSTV	HPFER	QRAQ	FYHGFEEYS
	Hal6	5	ATFSTV	HPFER	QRAQ	YYHGFESYS
C	Krml	6	ATTPSL	YPFGA		
	MibH	NA ^b	ATVSYM	APFDW		
	BrvH	NA	ATIPTI	HPFGL		
	VirX1	NA	STLGHF	YPFGP		
	Xcc-b100-1333	NA	STVPPI	HPFGS		
	Xcc-b100-4156	NA	ATIPSL	HSFGH		
	Xcc-b100-4345	NA	ATVPHI	HGFGT		
	MHal1	NA	ATMPNI	HSFDS		
	MHal2	NA	ATIIQM	NPFQP		
	MHal3	NA	ATVPGM	HPFNS		
MHal4	NA	ATIPPL	HSFGD			

^aResidues proposed to be important in the determination of reactivity and regioselectivity of tryptophan are shown in bold.

^bNA, not applicable.

spectrometry (LC-MS) and ¹H and ¹³C nuclear magnetic resonance (NMR) (Table S3; Fig. S3; Text S1). Hal1 and Hal7 primarily produced 7-chlorotryptophan (7-Cl-Trp), while Hal7 also produced 6,7-dichlorotryptophan (6,7-Di-Cl-Trp) as a minor product. Hal2 to -5 produced 6-chlorotryptophan (6-Cl-Trp), and Hal6 produced 5-chlorotryptophan (5-Cl-Trp), which is consistent with the genomic patterns and the position in the embedding space (Fig. 3a and b). Notably, these results are inconsistent with the annotations in NCBI gene database, where all selected genes, except for Hal3, are described as 7-Trp-FDHs. Such incorrect annotations might be due to the use of incorrect functional assignments based on homology transfer. Instead, the experimentally determined regioselectivity is consistent with the phylogenetic analysis (Fig. 1b), demonstrating that our bioinformatics approach correctly establishes the relationship between tryptophan-binding motifs and regioselectivity in halogenation.

Substitution of Cl⁻ with Br⁻ yielded comparable conversion (%) of tryptophan (Fig. 3a and c). LC-MS analysis indicated that the major products are monobrominated. The regioselectivity in bromination was maintained except for Hal4 to -5 in group B2, which produced mixtures of 5-Br-Trp and 6-Br-Trp. A lower regioselectivity here might indicate that Trp motifs in group B2 less stringently constrain the positioning of tryptophan to the catalytic lysine residue. Additionally, the more electrophilic and polarized nature of BrO⁻ over that of ClO⁻ might be partially responsible for the more reactive and less selective halogenation.

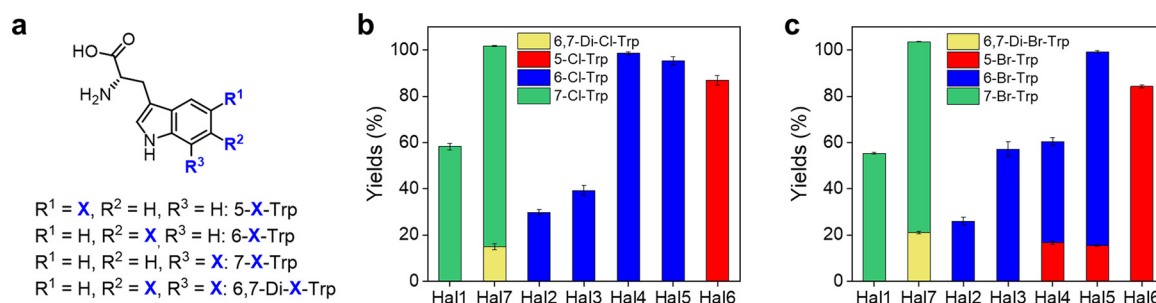


FIG 3 Halogenation of Trp-FDHs with tryptophan. (a) Chemical structures of the halogenated products. Product yields in chlorination (b) and bromination (c) reactions.

Altered regioselectivity of Trp-FDHs by the modification of Trp1 to -4 motifs.

Our genomic analysis integrated with structural information indicated that the genomic patterns and the Trp1 to -4 motifs might determine the regioselectivity of tryptophan halogenation. We thus proceeded to investigate whether the discrete non-covalent interactions could be changed by site-directed mutagenesis, thus altering the regioselectivity. Key amino acids in Trp1 to -4 motifs were replaced by comparing the motifs between two groups of halogenases that differ in the regioselectivity. The first mutagenesis study was performed between 6-Trp-FDH and 7-Trp-FDH in group A (Fig. 4a). The second study was performed between 5-Trp-FDH and 6-Trp-FDH in group B (Fig. 4b).

As shown in Fig. 4a, the first mutagenesis study was performed between 6-Trp-FDH and 7-Trp-FDH in group A. For the conversion of Hal2 (6-Trp-FDH in group A2) to 7-Trp-FDH, we carried out the following mutations: ATVP to ATIP (Trp1), HPF to HLF (Trp2), and YYX₁₃G₅YY to YYX₁₃G_NYY (Trp4) (Fig. 4a). Mutation of the Trp1 or Trp2 motif alone or together yielded no detectible changes in regioselectivity (Fig. 4a). Substitution of the Trp4 motif resulted in the 1:0.6 production of 6-Cl-Trp and 7-Cl-Trp. Inclusion of Trp1 or Trp2 mutation in the Trp4 variant inverted the product distribution to 1:4. Finally, when all three mutations were combined, the ratio became 1:21, which is consistent with the recent work on the quintuple mutations of Thal (40). These data indicate that seemingly subtle mutations, such as Val to Ile (Trp1), Pro to Leu (Trp2), and Ser to Asn (Trp4), are responsible for regioselectivity, and the orchestration of these motifs is necessary to alter the positioning of tryptophan exclusively. These results also indicate that Trp1, Trp2, and Trp4 motifs can be utilized as effective sequence markers to distinguish 6-Trp-FDH from 7-Trp-FDH proteins. Similar effects of these mutations were observed in the bromination with Hal2 (Fig. 4a). Again, the Trp4 motif makes the largest contribution, and the highest conversion to 7-Br-Trp is observed when Trp1, Trp2, and Trp4 mutations were combined, producing 6-Br-Trp and 7-Br-Trp in a 1:6 ratio. The distribution of the brominated products was not as prominent as that with Cl⁻, which is consistent with our earlier observation that BrO⁻ is less selective than ClO⁻.

The reverse mutagenesis was performed with Hal7 (7-Trp-FDH in group A1) toward 6-Trp-FDH. The effects of swapping the Trp motifs were similar to those with the A2 protein (Fig. 4a). While Trp1 and/or Trp2 mutations yielded no or slight perturbation, the addition of the Trp4 mutation drastically produced monohalogenated products at the 6 position in detectible amounts. When Trp1, Trp2, and Trp4 mutations were combined, the relative fraction of 6-Cl/Br-Trp was maximized, yielding 7-Br-Trp, 6,7-Di-Br-Trp, and 6-Br-Trp in a 1:0.46:1.23 ratio. We suspect that the 6,7-dihalogenated products were formed presumably due to the loosely regulated regioselectivity. Subsequent growth and decay upon the series of mutations might indicate that more selective interactions are made with tryptophan.

The second mutation study was carried out between 5-Trp-FDH and 6-Trp-FDH in group B (Fig. 4b). They possess identical Trp1 and Trp2 motifs as ATFS and HPF,

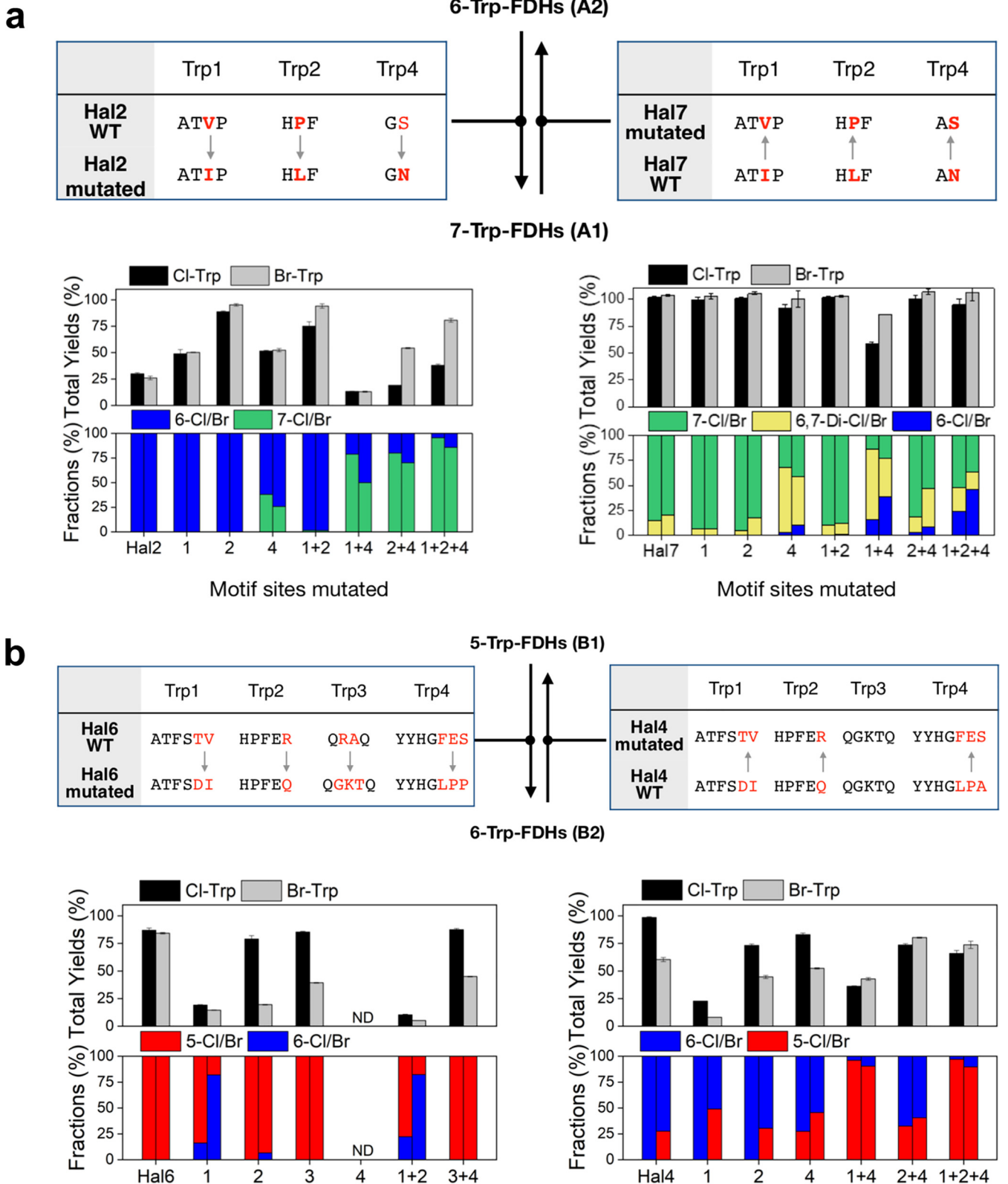


FIG 4 Altered regioselectivity of Trp-FDH by modification of Trp conservation sites. (a) Alteration between 6-Trp-FDH and 7-Trp-FDH. Total yields (top) and fractions (bottom) upon mutations of Hal2 to 7-Trp-FDH (left) and Hal7 to 6-Trp-FDH (right). (b) Alteration between 5-Trp-FDH and 6-Trp-FDH: Hal6 to 6-Trp-FDH (left) and Hal4 to 5-Trp-FDH (right). The fractions of the products in chlorination and bromination reactions are shown in left and right columns, respectively.

respectively. However, the tilted orientations of the conserved phenylalanines and a histidine were markedly different in the structures of PyrH (5-Trp-FDH in B1) and SttH (6-Trp-FDH in B2), thereby altering the directionality of the π - π stacking interactions. We speculated that the amino acids subsequent to the Trp1 and Trp2 motifs might be responsible for these distinct interactions. We thus extended the Trp1 and Trp2 motifs to include two additional amino acids: ATFSTV and HPFER in group B1 and ATFSDI and HPFEQ in group B2 (Fig. 2g and Table 1). Additionally, two glutamine residues in Trp3 (QxxQ in group B1 and QxxxQ in group B2) and Trp4 motifs (FE[T|S] in group B1 and LP [P|A] in group B2) were taken into consideration.

Since the role of Trp3 regions has not been investigated thus far, we first carried out experiments to confirm their involvement in the halogenation reaction. A double mutant was prepared with Hal6 by replacing both glutamines in Trp3 with alanines (from QxxQ to AxxA). Significantly lower conversion of tryptophan was observed in both chlorination and bromination reactions (see Table S3). The results suggest that two glutamine residues in Trp3 play an essential role in the reaction with tryptophan, possibly by the formation of hydrogen bonds with the amino and carboxyl moieties of the substrate.

To show the conversion from 5-Trp-FDH to 6-Trp-FDH in group B, the following mutations were made with Hal6 (5-Trp-FDH in group B1): ATFSTV to ATFSDI (Trp1), HPFER to HPFEQ (Trp2), QRAQ to QGKTQ (Trp3), or FES to LPP (Trp4) (Fig. 4b). Modification of Trp1 slightly altered the product distribution of 5-Cl-Trp and 6-Cl-Trp, whereas the ratio of brominated products was substantially changed to 1:4.6 (Fig. 4b). However, mutations of either the Trp2 or Trp3 motif reduced the overall yields without considerable changes in regioselectivity. Trp4 mutation alone or combined with a Trp1 or Trp2 mutation completely inactivated the enzyme, which is consistent with previous studies on PyrH (33). However, enzymatic activity was restored by the introduction of a Trp3 mutation without altering regioselectivity. When Trp1 and Trp2 mutations were combined, the ratios of 5-Cl-Trp to 6-Cl-Trp and of 5-Br-Trp to 6-Br-Trp were 1:0.3 and 1:4.7, respectively.

In the reverse mutagenesis performed with Hal4 (6-Trp-FDH in group B2) toward 5-Trp-FDH in group B1, the regiochemical preference was improved. While Trp1 or Trp2 mutations of Hal4 (6-Trp-FDH in group B2) alone yielded no alternation in the distribution of chlorinated products, the Trp4 mutation by itself altered the ratio of 6-Cl-Trp and 5-Cl-Trp to 1:0.4 (Fig. 4b), which is consistent with a previous report on the triple mutant of SttH (33). Substitutions on both Trp2 and Trp4 motifs resulted in negligible changes, whereas substitutions on Trp1 and Trp4 enhanced the ratio to 1:25. When Trp1, Trp2, and Trp4 motifs were altered simultaneously, the ratio was increased to 1:34 in chlorination, indicating that the combined interactions of these three motifs with tryptophan dictate regioselectivity in a nearly exclusive manner. The effects of this mutation were conserved in the bromination reaction except that Trp1 mutation alone induced considerable changes in the regioselectivity. Trp4 mutation produced mixed products at the ratio of 1:1; the combination of Trp1, Trp2, and Trp4 mutations maximized the ratio to 1:9.

Biochemical characterization of metagenomic FDHs. Steady-state activity assays were carried out on MHal1 to -4. MHal2 to -4 actively halogenated indole, but no conversion was observed for tryptophan (Tables S3 and S4; Fig. S3; Text S1). The discrete reactivity indicates that Trp-binding site Trp4, which is absent in MHal1 to -4, is critical for the halogenation activity toward tryptophan. Notably, MHal2 to -4 were considerably more reactive in bromination than in chlorination, although there is no significant difference in the sequences and structures of FAD- and halide-binding sites in genomic and metagenomic FDHs. Similarly to that for genomic FDHs, the molecular basis for halide selectivity observed in MHals needs further investigation. The discrete chemical nature of BrO^- and ClO^- and/or kinetic and/or thermodynamic parameters in transfer reactions of hypohalite to tryptophan or the indole-binding region via catalytic lysine might be in part responsible for the determination of the reactivities. The latter is

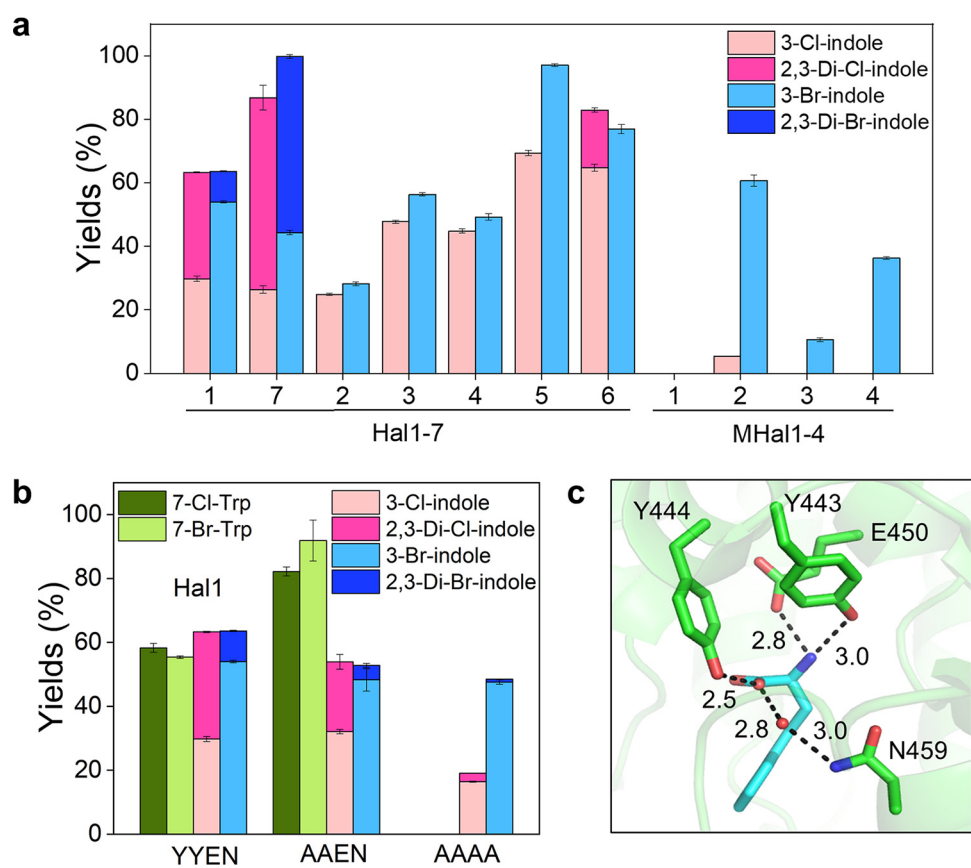


FIG 5 Halogenation of bacterial and metagenomic FDHs with indole. (a) Yields (%) of chlorinated and brominated indoles. (b) Reactivities of Hal1 variants at the Trp4 motif (YYEN) with tryptophan and indole. (c) Hydrogen-bonding interactions between Trp4 residues (YYEN) and a tryptophan shown in the structure of PrnA in A1 (2AQJ).

consistent with the molecular dynamics (MD) simulation studies of PrnA and PyrH (64), where significant conformational changes occur in the transfer of hypohalite. MHal1 was inactive even with indole, which is presumably related to the substantially altered sequence nearby the Trp-binding region (see Fig. S4).

For the reactions of MHal2 to -4 with indole, they exhibited identical regioselectivity, exclusively producing 3-Cl and 3-Br-indoles. We also measured the reactivities of Hal1 to -7 with indole. They also produced 3-Cl and 3-Br-indoles, displaying the regioselectivity identical to that of MHal2 to -4 (Fig. 5a). Additionally, Hal1, Hal6, and Hal7 yielded doubly halogenated products from indole, i.e., 2,3-dichloroindole or 2,3-dibromoindole. The formation of doubly halogenated products indicates that the binding affinity and additional halogenation reactivity of 3-Cl-indole or 3-Br-indole might be at least comparable to those of indole.

As the distinction between Hal1 to -7 and MHal2 to -4 is related to the Trp4 motif, we also explored whether the sequences in Trp4 may determine the reactivity with tryptophan versus that with indole. When two tyrosine residues (YY) at the beginning of Trp4 motif in Hal1 were mutated to alanines (Fig. 5b and c), the mutant yielded even higher conversion of tryptophan. When the other conserved residues in Trp4, glutamate and asparagine (EN), were mutated to alanines, the quartet mutant became completely inactive to tryptophan. Double or quartet mutants produced halogenated indoles with comparable or lower yields than Hal1, indicating that the absence of these four residues might severely weaken substrate-binding affinity, leading to lower catalytic activities, particularly for tryptophan but not as much as with indole. Therefore,

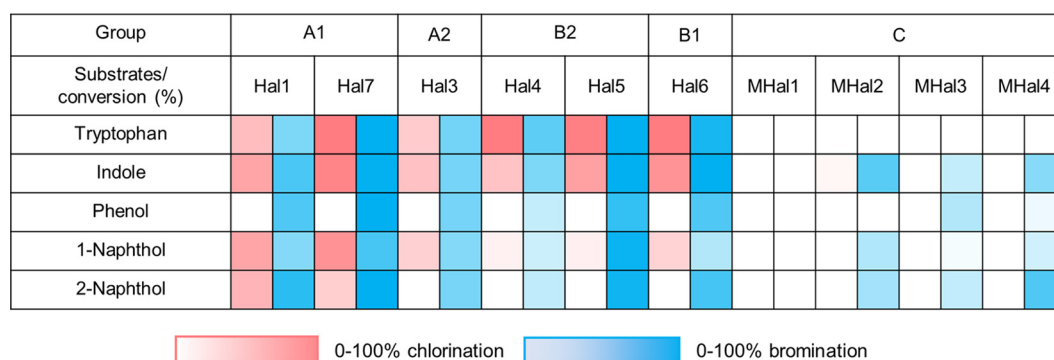


FIG 6 Catalytic activities of Trp-FDH identified from the bacterial genomes and metagenomes on diverse aromatic substrates.

the presence of the Trp4 motif might be critical to determining the substrate scopes, such as tryptophan versus indole, and the product yields.

We further investigated the substrate scopes of Hal1, Hal3 to -7, and MHal1 to -4 beyond indole (Fig. 6; Table S4; Fig. S3; Text S1). Hal2 was excluded due to protein instability. First, we observed whether other aromatic amino acids, such as tyrosine and phenylalanine, can be halogenated. No reactivity was observed, implying that the sp^2 carbon of aromatic rings might not be closely placed to the catalytic lysine. It is possible that the relatively smaller size of aromatic rings might have altered the mode of binding and interactions with halogenases. When indole derivatives, such as benzimidazole and benzothiazole having a heteroatom such as S and N at the C-3 position of indole, respectively, were added, no halogenated product was again observed, indicating that an alternative mode of binding of the indole moiety is disallowed, which contrasted with the discrete regioselectivity of tryptophan. Phenol was brominated by most of the proteins that exhibited activity with indole (Hal1, Hal3 to -7, and MHal3 and -4). The genomic and metagenomic FDHs were also active with 1-naphthol and 2-naphthol. Hal1 and Hal3 to -7 yielded both chlorinated and brominated products (11% to 83% and 21% to 100% conversion, respectively), whereas MHal2 to -4 reacted with 1- and 2-naphthols only in the presence of Br^- (4% to 69% conversion). Notably, irrespective of the phylogenetic positions of the genes, only the *ortho* and *para* positions of phenol and naphthols were halogenated, which reflects a strong intrinsic electronic bias for regioselectivity not overridden by substrate-protein interactions (Fig. 1b). Our results reveal that FDHs can act on diverse substrates, including a 6-membered benzene ring (phenol), 5-membered ring fused with 6-membered ring (indole), and two fused 6-membered rings (naphthol), with substantial variations in conversion and product distribution.

DISCUSSION

Many studies have performed computational analyses to estimate the functions of the genes in newly assembled genomes. For this purpose, sequence searches have been extensively used to find homologous genes of the same function. However, there exists limitations in finding subtle differences in biochemical reactivities such as regioselectivity in functional group transformation and selectivity of substrates. Since halogenation is an important chemical reaction in pharmaceutical or agrochemical applications, we have carried out an integrated genomic and protein structural analysis to identify the genotypes that are critical to the biochemical functions of FDHs. From both genome and metagenome mining, halogenases with similar genomic patterns of active sites are clustered in the phylogenetic tree as well as embedding space.

In this study, we found primarily genomic patterns that determine the enzymatic activity toward tryptophan (Trp-FDH versus non-Trp-FDH) and the site of halogenation for

tryptophan (5-, 6-, and 7-Cl/Br-Trp-FDH). Recent studies found that the sequences with different substrate scope were clustered based on the sequence similarity in the network analysis (19) and screened the halogenation with various substrates (65). Structure analysis of the known proteins identified the residues that bind to FAD and substrates (66–69). Compared with such studies, our work utilizes large-scale genomic information to propose a comprehensive set of active sites. Such sequence motifs for the active sites showed different patterns for the discrete regioselectivity of Trp-FDHs. Notably, the sequence motifs of Trp-FDHs that belong to A1 and A2 differ only moderately while they exhibit discrete regioselectivity. The results suggest that the halogenating reactivities are determined by the actions of several nonsequential and subtly different sequence motifs in concerted and complicated manners. In conjunction with biochemical assays, mutagenesis experiments suggest that the sequence motifs can be exchanged to create altered product distributions, implicating potential biotechnological applications in protein engineering and synthetic biology for the synthesis of novel natural products.

The correlation of the substrate specificity and regioselectivity with source organisms and their habitats is also an interesting and challenging question to characterize FDHs. In terms of taxonomic distribution, the FDHs were found in *Actinobacteria*, *Cyanobacteria*, and *Proteobacteria* (see Table S1 in the supplemental material). Since we do not have enough FDHs identified, it is hard to determine the correlation between substrate specificity/regioselectivity and taxonomic distribution. In this study, we used seawater microbiome data to find more diverse FDHs and found that most of the sequences were indole-FDHs. Screening of FDHs in more diverse environments could provide a better chance to characterize FDHs. The demonstration of discrete and diverse substrate scopes of genomic and metagenomic FDHs suggests that the FDHs annotated in the last decade are only a small portion of the genes, and more FDH genes could be discovered.

As discussed above, many gene annotations are based primarily on the homology search, which is fast but less accurate to predict subtle functional differences. While the NCBI annotation reported six FDHs (Hal1 and -2 and Hal4 to -7) as 7-Trp-FDHs, we found that four genes are in fact 5- or 6-Trp-FDH. To assign subtle functions such as regioselectivity, the genomic sequences of halogenases in the public repository need additional confirmation procedures using more informative models such as motif patterns or phylogenetic characteristics.

MATERIALS AND METHODS

Construction of Trp-FDH pHMM and halogenase identification. To search halogenase genes in the bacterial genomes and metagenomes, the pHMM was constructed using 13 Trp-FDHs experimentally validated in previous studies (see Fig. S1 in the supplemental material): AbeH (AEF32095.1), ClaH (AEO12707.1), PyrH (AAU95674.1), KtzR (ABV56598.1), SttH (ADW94630.1), Th-Hal (WP_023586065.1), ThdH (ABK79936.1), BorH (AGI62217.1), FmoD (BAP16687.1), AORI_5336 (AGM07919.1), KtzQ (ABV56597.1), PrnA (AAB97504.1), and RebH (CAC93722.1). A multiple-sequence alignment was conducted using Clustal Omega (v1.2.4) (70). The pHMM of halogenase genes was constructed using hmmbuild in HMMER3.0 (ver. 3.1b2) (71). To identify the genes that are homologous to the known Trp-FDHs in bacterial genomes and metagenomes, hmscan in HMMER3.0 (71) was used with the `-domtblout` option. Halogenases identified in the BGC cluster were included regardless of the E value. Using an in-house Python script, homology search results were filtered. The models and script are available at GitHub site (<https://github.com/DMnBI/Halogenase>). The E value threshold was set to $<10^{-130}$, and the model length covered $>50\%$. A total of 351 sequences were collected from the NCBI repository by searching with the keyword “tryptophan n-halogenase [All Fields] AND bacteria [filter],” where $n=5, 6, \text{ and } 7$. Partial and redundant sequences were discarded, resulting in 333 sequences obtained (Fig. S1). In addition, we found three additional Trp-FDHs from the literature survey, which were also included in the motif analysis.

Detection of halogenase genes in the BGCs. To identify BGCs in 2,787 bacterial complete genomes obtained from the NCBI repository, antiSMASH3.0 (72) was used. In total, 10,166 BGCs were predicted; halogenases inside a BGC were characterized subsequently. In total, 20 putative halogenases inside BGCs and 83 outside BGCs were classified.

Sample collection, preparation, and metagenome shotgun sequencing. A total of 17 metagenomic samples were obtained from different marine environments, such as coastal and deep-sea sediments (Table S1). The metagenome data sets analyzed in this study were collected in our previous studies and downloaded from the European Nucleotide Archive with the accession number ERP107268 (Table S1). All reads were 151-bp paired-end reads. Low-quality reads, which have the ambiguous sequence “N,” were trimmed and filtered using Sickle (downloaded from <https://github.com/najoshi/sickle>) with the option of Phred quality score of >20 . The

metagenomic reads were assembled using MEGAHIT (v1.0.3) (73) with default k-mer options. The assembled contigs longer than 1,000 bp were translated into protein sequences using FragGeneScan (v1.19) (74).

Embedding of halogenase genes and classifying halogenases. To investigate the similarity of genomic patterns among FDHs, *t*-distributed stochastic neighbor embedding (t-SNE) was carried out using "Rtsne" in the R package (75) with the parameters of perplexity of 10 and iterations of 1,000. From multiple-sequence alignment of the known halogenase sequences, the profiles were constructed for six groups (5-, 6-, and 7-Trp-FDHs, indole-Hals, phenolic-Hals, and pyrrole-Hals). Using these profiles, the probability of the amino acid at each position of the halogenase genes was calculated as a feature value to learn their embedding. After obtaining embedding vectors for each halogenase, the class (5-Trp-FDHs in B1, 6-Trp-FDHs in B2 and A2, 7-Trp-FDHs in A1, and indole-Hals in C) was assigned using the k-nearest neighbor method with a k of 1.

Phylogenetic analysis. Multiple-sequence alignments were conducted using MUSCLE (v3.8.31) (76), and a phylogenetic tree was constructed using MEGA6.0 (77). The evolutionary distance was calculated using the maximum likelihood method based on a Jones-Taylor-Thornton (JTT) matrix-based model. In the phylogenetic analysis, a total of 180 sequences (of 313) were used for better representation of the tree (Fig. S1). Among the 109 sequences from the keyword search, 6 were included. Among the 68 sequences obtained from metagenome data, 38 sequences were included by random sampling.

Examination of FDH structures. Three-dimensional structures of FDHs cocrystallized with native substrates such as FAD, halide, and/or tryptophan were obtained from the RCSB database and inspected for structural analysis. The figures were created using PyMOL software (78). The residues involved in direct enzyme-substrate interactions with the native substrates were manually examined to correlate the structural data with the sequence analysis (Table 1). Residues that form enzyme-substrate interactions were identified as FAD1 to -4, halide, and Trp1 to -4 (Fig. S2B to H). To represent the conservation of functional regions, the sequence logo was constructed using WebLogo (v3.6.0) (79) with default options.

Biochemical characterization of putative halogenases from genomic and metagenomic sequences. From the sequence analysis data, we selected 11 putative FDH genes, 7 from genome data (Hal1 to -7) and 4 from metagenome data (MHal1 to -4). For steady-state coupled assays, an NADH-dependent reductase (MR) was selected from the metagenome with MHal4. The MR is functionally analogous to NAD(P)H-dependent flavin oxidoreductase (Fre) found in *Escherichia coli* (80).

The 12 genes were synthesized after codon optimization for *E. coli* heterologous expression (General Biosystems, NC Morrisville). MHal1 to -3 were inserted into the pET-22b(+) vector using NdeI and XhoI as the cut sites. MR, Hal1 to -7, and MHal4 were inserted into the pET-28a(+) vector using NdeI and BamHI as the cut sites. A stop codon was inserted at the end of the sequence due to its native C-terminal His tag sequence in the pET-28a(+) vector. The plasmids were transformed to DH5 α or BL21(DE3) strains of *E. coli* competent cells by heat shock method for sequencing or protein expression, respectively. All putative FDHs were coexpressed with GroES and GroEL chaperones using pGro7 plasmid (TaKaRa Bio) as reported previously (32). For site-directed mutagenesis of FDHs, custom-made primers were used for PCR (Table S2). Detailed experimental procedures for cell growth, protein isolation, purification, and characterization are available in Text S1 and Fig. S5.

Characterization of an NADH-dependent reductase from arctic environments. The catalytic activity of a newly discovered reductase (MR) was determined by monitoring the consumption rates of NADH by the addition of 5 mM NADH, 10 μ M FAD, 5 μ M reductase, and 65 mM halide source, such as NaCl and NaBr. The conversions (%) of indole were measured using high-pressure liquid chromatography (HPLC) by calculating the residual indole in the presence of 5 mM NADH, 10 μ M FAD, 5 μ M reductase, 65 mM NaBr, and 25 μ M MHal2 in 25 mM HEPES (pH 7.4) buffer at room temperature. For preliminary studies, we measured the activity of another reductase, Fre (Fig. S5) (81).

Activity assays of putative FDHs. For steady-state coupled activity assays, NADH (5 mM, 20 eq to the aromatic substrate), FAD (10 μ M, 0.04 eq), and aromatic substrates (250 μ M, 1 eq), NaCl or NaBr (65 mM, 260 eq) in 25 mM HEPES buffer, pH 7.4, were prepared in a 1.5-ml microcentrifuge tube at final concentrations. The reaction was initiated by adding putative FDH (25 μ M, 0.1 eq) and MR (5 μ M, 0.02 eq) to the premixed solution to make the final volume 100 μ l. The assays were carried out at room temperature in darkness. After 5 h, 100 μ l of methanol was added, and the resulting precipitates were removed by centrifugation at $15,928 \times g$ for 10 min at 4°C. After syringe filtration, the supernatant was analyzed by reverse-phase HPLC (Agilent 1260 Infinity II) with InfinityLab Poroshell column (120 EC-C₁₈, 4.6 by 100 mm, 2.7 μ m, or 120 EC-C₁₈, 4.6 by 150 mm, 2.7 μ m) with a linear gradient of H₂O containing 0.05% trifluoroacetic acid (TFA) and acetonitrile (ACN) containing 0.1% TFA. A representative HPLC trace and the retention time are shown in Fig. S3. Averages and standard deviations of activity assays were obtained from triple replications. The conversions and yields are listed in Tables S3 and S4.

Characterization of the halogenated products. For NMR characterizations, we scaled up the reaction volumes to 90 ml without altering the concentrations or the ratios of the reaction components. We observed that increasing the volume of the reactions considerably increased the total yields of the products, presumably because the activities of coupled enzymatic reactions are dependent on the concentration of dissolved dioxygen in the buffer solution. To maximize the yields, the reaction mixtures were incubated overnight. Depending on the substrates, the workup procedure was slightly modified. For the reactions with tryptophan, 90 ml of methanol (MeOH) containing 0.1% TFA was added for quenching. After centrifugation at $2,935 \times g$ for 10 min at 4°C and syringe filtration, the solution was directly injected to the semiprep HPLC (ZORBAX SB-C₁₈ semipreparative 9.4- by 250-mm 5- μ m column). For the reactions with substrates other than tryptophan, ca. 5 drops of concentrated HCl were added for quenching. Precipitates were removed after centrifugation at $2,935 \times g$ for 10 min at 4°C, and the supernatants were extracted with dichloromethane (DCM). After vacuum evaporation, ACN-H₂O (1:9 [vol/vol])

was added to the extracted fraction. The solution was filtered by a 0.25- μm syringe filter and purified by HPLC (ZORBAX SB-C₁₈ semipreparative 9.4- by 250-mm 5- μm column).

The reaction products of non-amino acidic substrates (indole, phenol, 1-naphthol, and 2-naphthol) were characterized by gas chromatography-mass spectrometry (GC-MS) (Hewlett Packard HP 6890 series GC system equipped with Agilent 5973 Network mass selective detector or Agilent Technologies 7820A GC system equipped with Agilent Technologies 5977E MSD). The reaction products from tryptophan were analyzed by LC-MS (Agilent Technologies 1200 Infinity series equipped with Agilent Technologies 6120 quadrupole LC-MS). The incorporation of chlorine or bromine was determined by the isotope ratios of ³⁵Cl to ³⁷Cl of 3:1 or ⁷⁹Br to ⁸¹Br of 1:1, respectively. The products were analyzed by ¹H and ¹³C NMR spectroscopy (Fig. S3; Text S1).

Synthesis of 3-chloroindole and 3-bromoindole. To isolate large quantities of 2,3-dichloroindole and 2,3-dibromoindole for ¹H and ¹³C NMR experiments, we synthesized 3-chloroindole or 3-bromoindole and used as the substrate for the reactions with Hal1. To a dried 50-ml round-bottom flask, 588 mg of indole (1 eq) and 377 μl of 1,4-dimethylpiperazine (0.558 eq) were dissolved into 5 ml DCM. The temperature was lowered by placing the flask in iced water. After 10 min, 800 mg of *N*-chlorosuccinimide or *N*-bromosuccinimide (0.9 eq) was added slowly and stirred for 15 min to yield 3-chloroindole or 3-bromoindole, respectively. After 15 min, the temperature was raised to room temperature, and the reaction mixture was stirred for 3 h. Volatile fractions were removed by rotary evaporation, and the product was isolated by silica chromatography (ethyl acetate-hexane, 5:95). The isolated yields of 3-chloroindole and 3-bromoindole were 55.5% and 57%, respectively.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

TEXT S1, DOCX file, 0.1 MB.

FIG S1, TIF file, 0.6 MB.

FIG S2, TIF file, 1.1 MB.

FIG S3, PDF file, 2.9 MB.

FIG S4, TIF file, 2.2 MB.

FIG S5, TIF file, 0.8 MB.

TABLE S1, DOCX file, 0.1 MB.

TABLE S2, DOCX file, 0.1 MB.

TABLE S3, DOCX file, 0.1 MB.

TABLE S4, DOCX file, 0.1 MB.

ACKNOWLEDGMENTS

This work was supported by the Collaborative Genome Program of the Korea Institute of Marine Science and Technology Promotion (KIMST), funded by the Ministry of Oceans and Fisheries (MOF) (no. 20180430) and the National Research Foundation (NRF) of the Korean government (NRF-2019R1C1C1003863).

The Fre plasmid used in the initial activity assays was a kind gift from Marc Foncave at Laboratoire de Chimie des Processus Biologiques, Collège de France–CNRS–Sorbonne Université. We thank Dongwhan Lee for helpful discussions.

J.J., J.L., J.H.S., W.J.S., and M.R. designed the project. J.J. and J.H.S. performed the bioinformatic analysis. J.L. and S.-M.J. carried out biochemical experiments. J.J., J.L., J.H.S., W.J.S., and M.R. wrote the paper.

REFERENCES

- Gribble GW. 2010. Naturally occurring organohalogen compounds: a comprehensive update. Springer, Verlag, Germany.
- Latham J, Brandenburger E, Shepherd SA, Menon BRK, Micklefield J. 2018. Development of halogenase enzymes for use in synthesis. *Chem Rev* 118:232–269. <https://doi.org/10.1021/acs.chemrev.7b00032>.
- Harris CM, Kannan R, Kopecka H, Harris TM. 1985. The role of the chlorine substituents in the antibiotic vancomycin: preparation and characterization of mono- and didechlorovancomycin. *J Am Chem Soc* 107:6652–6658. <https://doi.org/10.1021/ja00309a038>.
- Rodrigues Pereira E, Belin L, Sancelme M, Prudhomme M, Ollier M, Rapp M, Sevère D, Riou J-F, Fabbro D, Meyer T. 1996. Structure–activity relationships in a series of substituted indolocarbazoles: topoisomerase I and protein kinase C inhibition and antitumoral and antimicrobial properties. *J Med Chem* 39:4471–4477. <https://doi.org/10.1021/jm9603779>.
- Parisini E, Metrangolo P, Pilati T, Resnati G, Terraneo G. 2011. Halogen bonding in halocarbon-protein complexes: a structural survey. *Chem Soc Rev* 40:2267–2278. <https://doi.org/10.1039/c0cs00177e>.
- Voth AR, Hays FA, Ho PS. 2007. Directing macromolecular conformation through halogen bonds. *Proc Natl Acad Sci U S A* 104:6188–6193. <https://doi.org/10.1073/pnas.0610531104>.
- Agarwal V, Miles ZD, Winter JM, Eustaquio AS, El Gamal AA, Moore BS. 2017. Enzymatic halogenation and dehalogenation reactions: pervasive and mechanistically diverse. *Chem Rev* 117:5619–5674. <https://doi.org/10.1021/acs.chemrev.6b00571>.
- Weichold V, Milbredt D, van Pée K-H. 2016. Specific enzymatic halogenation—from the discovery of halogenated enzymes to their applications *in vitro* and *in vivo*. *Angew Chem Int Ed Engl* 55:6374–6389. <https://doi.org/10.1002/anie.201509573>.

9. Butler A, Sandy M. 2009. Mechanistic considerations of halogenating enzymes. *Nature* 460:848–854. <https://doi.org/10.1038/nature08303>.
10. Neumann CS, Fujimori DG, Walsh CT. 2008. Halogenation strategies in natural product biosynthesis. *Chem Biol* 15:99–109. <https://doi.org/10.1016/j.chembiol.2008.01.006>.
11. Frese M, Sewald N. 2015. Enzymatic halogenation of tryptophan on a gram scale. *Angew Chem Int Ed Engl* 54:298–301. <https://doi.org/10.1002/anie.201408561>.
12. Schroeder L, Frese M, Müller C, Sewald N, Kottke T. 2018. Photochemically driven biocatalysis of halogenases for the green production of chlorinated compounds. *ChemCatChem* 10:3336–3341. <https://doi.org/10.1002/cctc.201800280>.
13. Durak LJ, Payne JT, Lewis JC. 2016. Late-stage diversification of biologically active molecules via chemoenzymatic C–H functionalization. *ACS Catal* 6:1451–1454. <https://doi.org/10.1021/acscatal.5b02558>.
14. Ismail M, Schroeder L, Frese M, Kottke T, Hollmann F, Paul CE, Sewald N. 2019. Straightforward regeneration of reduced flavin adenine dinucleotide required for enzymatic tryptophan halogenation. *ACS Catal* 9:1389–1395. <https://doi.org/10.1021/acscatal.8b04500>.
15. Runguphan W, Qu X, O'Connor SE. 2010. Integrating carbon–halogen bond formation into medicinal plant metabolism. *Nature* 468:461–464. <https://doi.org/10.1038/nature09524>.
16. Latham J, Henry J-M, Sharif HH, Menon BRK, Shepherd SA, Greaney MF, Micklefield J. 2016. Integrated catalysis opens new arylation pathways via regiodivergent enzymatic C–H activation. *Nat Commun* 7:11873. <https://doi.org/10.1038/ncomms11873>.
17. Kong L, Wang Q, Deng Z, You D. 2020. Flavin adenine dinucleotide-dependent halogenase XanH and engineering of multifunctional fusion halogenases. *Appl Environ Microbiol* 86:e01225–20. <https://doi.org/10.1128/AEM.01225-20>.
18. Zehner S, Kotszsch A, Bister B, Sussmuth RD, Mendez C, Salas JA, van Pee KH. 2005. A regioselective tryptophan 5-halogenase is involved in pyroindomycin biosynthesis in *Streptomyces rugosporus* LL-42D005. *Chem Biol* 12:445–452. <https://doi.org/10.1016/j.chembiol.2005.02.005>.
19. Fisher BF, Snodgrass HM, Jones KA, Andorfer MC, Lewis JC. 2019. Site-selective C–H halogenation using flavin-dependent halogenases identified via family-wide activity profiling. *ACS Cent Sci* 5:1844–1856. <https://doi.org/10.1021/acscentsci.9b00835>.
20. Buedenbender S, Rachid S, Müller R, Schulz GE. 2009. Structure and action of the myxobacterial chondrochloren halogenase CndH: a new variant of FAD-dependent halogenases. *J Mol Biol* 385:520–530. <https://doi.org/10.1016/j.jmb.2008.10.057>.
21. Rachid S, Krug D, Kunze B, Kochems I, Scharfe M, Zabriskie TM, Blöcker H, Müller R. 2006. Molecular and biochemical studies of chondramide formation—highly cytotoxic natural products from *Chondromyces crocatus* Cm c5. *Chem Biol* 13:667–681. <https://doi.org/10.1016/j.chembiol.2006.06.002>.
22. Smith DRM, Uria AR, Helfrich EJN, Milbredt D, van Pée K-H, Piel J, Goss RJM. 2017. An unusual flavin-dependent halogenase from the metagenome of the marine sponge *Theonella swinhoei* WA. *ACS Chem Biol* 12:1281–1287. <https://doi.org/10.1021/acscchembio.6b01115>.
23. Dorrestein PC, Yeh E, Garneau-Tsodikova S, Kelleher NL, Walsh CT. 2005. Dichlorination of a pyrrolyl-S-carrier protein by FADH₂-dependent halogenase PltA during pyoluteorin biosynthesis. *Proc Natl Acad Sci U S A* 102:13843–13848. <https://doi.org/10.1073/pnas.0506964102>.
24. El Gamal A, Agarwal V, Diethelm S, Rahman I, Schorn MA, Sneed JM, Louie GV, Whalen KE, Mincer TJ, Noel JP, Paul VJ, Moore BS. 2016. Biosynthesis of coral settlement cue tetrabromopyrrole in marine bacteria by a uniquely adapted brominase–thioesterase enzyme pair. *Proc Natl Acad Sci U S A* 113:3797–3802. <https://doi.org/10.1073/pnas.1519695113>.
25. Zeng J, Zhan J. 2010. A novel fungal flavin-dependent halogenase for natural product biosynthesis. *Chembiochem* 11:2119–2123. <https://doi.org/10.1002/cbic.201000439>.
26. Andorfer MC, Grob JE, Hajdin CE, Chael JR, Siuti P, Lilly J, Tan KL, Lewis JC. 2017. Understanding flavin-dependent halogenase reactivity via substrate activity profiling. *ACS Catal* 7:1897–1904. <https://doi.org/10.1021/acscatal.6b02707>.
27. Chooi Y-H, Cacho R, Tang Y. 2010. Identification of the viridicatumtoxin and griseofulvin gene clusters from *Penicillium aethiopicum*. *Cell Chem Biol* 17:483–494. <https://doi.org/10.1016/j.chembiol.2010.03.015>.
28. Mori S, Pang AH, Thamban Chandrika N, Garneau-Tsodikova S, Tsodikov OV. 2019. Unusual substrate and halide versatility of phenolic halogenase PltM. *Nat Commun* 10:1255. <https://doi.org/10.1038/s41467-019-09215-9>.
29. Andorfer MC, Lewis JC. 2018. Understanding and improving the activity of flavin-dependent halogenases via random and targeted mutagenesis. *Annu Rev Biochem* 87:159–185. <https://doi.org/10.1146/annurev-biochem-062917-012042>.
30. Ortega MA, Cogan DP, Mukherjee S, Garg N, Li B, Thibodeaux GN, Maffioli SI, Donadio S, Sosio M, Escano J, Smith L, Nair SK, van der Donk WA. 2017. Two flavoenzymes catalyze the post-translational generation of 5-chloro-tryptophan and 2-aminovinyl-cysteine during NAI-107 biosynthesis. *ACS Chem Biol* 12:548–557. <https://doi.org/10.1021/acscchembio.6b01031>.
31. Neubauer PR, Widmann C, Wibberg D, Schröder L, Frese M, Kottke T, Kalinowski J, Niemann HH, Sewald N. 2018. A flavin-dependent halogenase from metagenomic analysis prefers bromination over chlorination. *PLoS One* 13:e0196797. <https://doi.org/10.1371/journal.pone.0196797>.
32. Payne JT, Andorfer MC, Lewis JC. 2013. Regioselective arene halogenation using the FAD-dependent halogenase RebH. *Angew Chem Int Ed Engl* 52:5271–5274. <https://doi.org/10.1002/anie.201300762>.
33. Shepherd SA, Menon BR, Fisk H, Struck AW, Levy C, Leys D, Micklefield J. 2016. A structure-guided switch in the regioselectivity of a tryptophan halogenase. *Chembiochem* 17:821–824. <https://doi.org/10.1002/cbic.201600051>.
34. Dong C, Flecks S, Unversucht S, Haupt C, van Pee KH, Naismith JH. 2005. Tryptophan 7-halogenase (PrnA) structure suggests a mechanism for regioselective chlorination. *Science* 309:2216–2219. <https://doi.org/10.1126/science.1116510>.
35. Flecks S, Patallo EP, Zhu X, Ernyei AJ, Seifert G, Schneider A, Dong C, Naismith JH, van Pee KH. 2008. New insights into the mechanism of enzymatic chlorination of tryptophan. *Angew Chem Int Ed Engl* 47:9533–9536. <https://doi.org/10.1002/anie.200802466>.
36. Lang A, Polnick S, Nicke T, William P, Patallo EP, Naismith JH, van Pee KH. 2011. Changing the regioselectivity of the tryptophan 7-halogenase PrnA by site-directed mutagenesis. *Angew Chem Int Ed Engl* 50:2951–2953. <https://doi.org/10.1002/anie.201007896>.
37. Glenn WS, Nims E, O'Connor SE. 2011. Reengineering a tryptophan halogenase to preferentially chlorinate a direct alkaloid precursor. *J Am Chem Soc* 133:19346–19349. <https://doi.org/10.1021/ja2089348>.
38. Payne JT, Poor CB, Lewis JC. 2015. Directed evolution of RebH for site-selective halogenation of large biologically active molecules. *Angew Chem Int Ed Engl* 54:4226–4230. <https://doi.org/10.1002/anie.201411901>.
39. Andorfer MC, Park HJ, Vergara-Coll J, Lewis JC. 2016. Directed evolution of RebH for catalyst-controlled halogenation of indole C–H bonds. *Chem Sci* 7:3720–3729. <https://doi.org/10.1039/C5SC04680G>.
40. Moritzer A-C, Minges H, Prior T, Frese M, Sewald N, Niemann HH. 2019. Structure-based switch of regioselectivity in the flavin-dependent tryptophan 6-halogenase Thal. *J Biol Chem* 294:2529–2542. <https://doi.org/10.1074/jbc.RA118.005393>.
41. Rho M, Song WJ. 2020. Discovery of novel gene functions by chemistry-guided targeted sequence analysis. *Biochemistry* 59:10–11. <https://doi.org/10.1021/acs.biochem.9b00790>.
42. Liao L, Chen R, Jiang M, Tian X, Liu H, Yu Y, Fan C, Chen B. 2016. Bioprospecting potential of halogenases from Arctic marine actinomycetes. *BMC Microbiol* 16:34. <https://doi.org/10.1186/s12866-016-0662-2>.
43. Weigold P, El-Hadidi M, Ruecker A, Huson DH, Scholten T, Jochmann M, Kappler A, Behrens S. 2016. A metagenomic-based survey of microbial (de)halogenation potential in a German forest soil. *Sci Rep* 6:28958. <https://doi.org/10.1038/srep28958>.
44. Heemstra JR, Walsh CT. 2008. Tandem action of the O₂⁻ and FADH₂-dependent halogenases KtzQ and KtzR produce 6,7-dichlorotryptophan for kutzneride assembly. *J Am Chem Soc* 130:14024–14025. <https://doi.org/10.1021/ja806467a>.
45. Zeng J, Zhan J. 2011. Characterization of a tryptophan 6-halogenase from *Streptomyces toxytricini*. *Biotechnol Lett* 33:1607–1613. <https://doi.org/10.1007/s10529-011-0595-7>.
46. Luhavaya H, Sigrist R, Chekan JR, McKinnie SMK, Moore BS. 2019. Biosynthesis of l-4-chlorokynurenine, an antidepressant prodrug and a non-proteinogenic amino acid found in lipopeptide antibiotics. *Angew Chem Int Ed Engl* 58:8394–8399. <https://doi.org/10.1002/anie.201901571>.
47. Menon BRK, Latham J, Dunstan MS, Brandenburger E, Klemstein U, Leys D, Karthikeyan C, Greaney MF, Shepherd SA, Micklefield J. 2016. Structure and biocatalytic scope of thermophilic flavin-dependent halogenase and flavin reductase enzymes. *Org Biomol Chem* 14:9354–9361. <https://doi.org/10.1039/c6ob01861k>.
48. Xu L, Han T, Ge M, Zhu L, Qian X. 2016. Discovery of the new plant growth-regulating compound LYXLF2 based on manipulating the halogenase in *Amycolatopsis orientalis*. *Curr Microbiol* 73:335–340. <https://doi.org/10.1007/s00284-016-1052-6>.

49. Chang F-Y, Brady SF. 2013. Discovery of indolotryptoline antiproliferative agents by homology-guided metagenomic screening. *Proc Natl Acad Sci U S A* 110:2478–2483. <https://doi.org/10.1073/pnas.1218073110>.
50. Muliandi A, Katsuyama Y, Sone K, Izumikawa M, Moriya T, Hashimoto J, Kozono I, Takagi M, Shin-Ya K, Ohnishi Y. 2014. Biosynthesis of the 4-methylloxazoline-containing nonribosomal peptides, JBIR-34 and -35, in *Streptomyces* sp. Sp080513GE-23. *Chem Biol* 21:923–934. <https://doi.org/10.1016/j.chembiol.2014.06.004>.
51. Milbredt D, Patallo EP, van Pee KH. 2014. A tryptophan 6-halogenase and an amidotransferase are involved in thienodolin biosynthesis. *Chembiochem* 15:1011–1020. <https://doi.org/10.1002/cbic.201400016>.
52. Gkotsi DS, Ludewig H, Sharma SV, Connolly JA, Dhaliwal J, Wang Y, Unsworth WP, Taylor RJK, McLachlan MMW, Shanahan S, Naismith JH, Goss RJM. 2019. A marine viral halogenase that iodinated diverse substrates. *Nat Chem* 11:1091–1097. <https://doi.org/10.1038/s41557-019-0349-z>.
53. Widmann C, Ismail M, Sewald N, Niemann HH. 2020. Structure of apo flavin-dependent halogenase Xcc4156 hints at a reason for cofactor-soaking difficulties. *Acta Crystallogr D Struct Biol* 76:687–697. <https://doi.org/10.1107/S2059798320007731>.
54. Ismail M, Frese M, Patschkowski T, Ortseifen V, Niehaus K, Sewald N. 2019. Flavin-dependent halogenases from *Xanthomonas campestris* pv. *campestris* B100 prefer bromination over chlorination. *Adv Synth Catal* 361:2475–2486. <https://doi.org/10.1002/adsc.201801591>.
55. Keller S, Frese M, Hohauser K, Holzer M, Eichhorn E, van Pee KH. 2000. Purification and partial characterization of tryptophan 7-halogenase (PmA) from *Pseudomonas fluorescens*. *Angew Chem Int Ed* 39:2300–2302. [https://doi.org/10.1002/1521-3773\(20000703\)39:13<2300::AID-ANIE2300>3.0.CO;2-I](https://doi.org/10.1002/1521-3773(20000703)39:13<2300::AID-ANIE2300>3.0.CO;2-I).
56. Wang Y, Wang J, Yu S, Wang F, Ma H, Yue C, Liu M, Deng Z, Huang Y, Qu X. 2016. Identifying the minimal enzymes for unusual carbon–sulfur bond formation in thienodolin biosynthesis. *Chembiochem* 17:799–803. <https://doi.org/10.1002/cbic.201500670>.
57. Yeh E, Garneau S, Walsh CT. 2005. 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* 102:3960–3965. <https://doi.org/10.1073/pnas.0500755102>.
58. Ryan KS. 2011. Biosynthetic gene cluster for the cladoniamides, bis-indoles with a rearranged scaffold. *PLoS One* 6:e23694. <https://doi.org/10.1371/journal.pone.0023694>.
59. Chang F-Y, Brady SF. 2011. Cloning and characterization of an environmental DNA-Derived gene cluster that encodes the biosynthesis of the antitumor substance BE-54017. *J Am Chem Soc* 133:9996–9999. <https://doi.org/10.1021/ja2022653>.
60. Liu Z, Ma L, Zhang L, Zhang W, Zhu Y, Chen Y, Zhang W, Zhang C. 2019. Functional characterization of the halogenase SpmH and discovery of new deschloro-tryptophan dimers. *Org Biomol Chem* 17:1053–1057. <https://doi.org/10.1039/c8ob02775g>.
61. Domergue J, Erdmann D, Fossey-Jouenne A, Petit JL, Debard A, de Berardinis V, Vergne-Vaxelaire C, Zaparucha A. 2019. XszenFHal, a novel tryptophan 5-halogenase from *Xenorhabdus szentirmaii*. *AMB Express* 9:175. <https://doi.org/10.1186/s13568-019-0898-y>.
62. Zhu X, De Laurentis W, Leang K, Herrmann J, Ihlefeld K, van Pee KH, Naismith JH. 2009. Structural insights into regioselectivity in the enzymatic chlorination of tryptophan. *J Mol Biol* 391:74–85. <https://doi.org/10.1016/j.jmb.2009.06.008>.
63. Shepherd SA, Karthikeyan C, Latham J, Struck A-W, Thompson ML, Menon BRK, Styles MQ, Levy C, Leys D, Micklefield J. 2015. Extending the biocatalytic scope of regiocomplementary flavin-dependent halogenase enzymes. *Chem Sci* 6:3454–3460. <https://doi.org/10.1039/c5sc00913h>.
64. Ainsley J, Mulholland AJ, Black GW, Sparagano O, Christov CZ, Karabencheva-Christova TG. 2018. Structural insights from molecular dynamics simulations of tryptophan 7-halogenase and tryptophan 5-halogenase. *ACS Omega* 3:4847–4859. <https://doi.org/10.1021/acsomega.8b00385>.
65. Neubauer PR, Pienkny S, Wessjohann L, Brandt W, Sewald N. 2020. Predicting the substrate scope of the flavin-dependent halogenase BrvH. *Chembiochem* 21:3282–3288. <https://doi.org/10.1002/cbic.202000444>.
66. Moritzer AC, Niemann HH. 2019. Binding of FAD and tryptophan to the tryptophan 6-halogenase Thal is negatively coupled. *Protein Sci* 28:2112–2118. <https://doi.org/10.1002/pro.3739>.
67. Lee J, Kim J, Kim H, Kim EJ, Jeong HJ, Choi KY, Kim BG. 2020. Characterization of a tryptophan 6-halogenase from *Streptomyces albus* and its regioselectivity determinants. *Chembiochem* 21:1446–1452. <https://doi.org/10.1002/cbic.201900723>.
68. Podzelinska K, Latimer R, Bhattacharya A, Vining LC, Zechel DL, Jia Z. 2010. Chloramphenicol biosynthesis: the structure of CmlS, a flavin-dependent halogenase showing a covalent flavin-aspartate bond. *J Mol Biol* 397:316–331. <https://doi.org/10.1016/j.jmb.2010.01.020>.
69. Frese M, Guzowska PH, Voß H, Sewald N. 2014. Regioselective enzymatic halogenation of substituted tryptophan derivatives using the FAD-dependent halogenase RebH. *ChemCatChem* 6:1270–1276. <https://doi.org/10.1002/cctc.201301090>.
70. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7:539. <https://doi.org/10.1038/msb.2011.75>.
71. Mistry J, Finn RD, Eddy SR, Bateman A, Punta M. 2013. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res* 41:e121. <https://doi.org/10.1093/nar/gkt263>.
72. Weber T, Blin K, Duddela S, Krug D, Kim HU, Bruccoleri R, Lee SY, Fischbach MA, Müller R, Wohleben W, Breitling R, Takano E, Medema MH. 2015. antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res* 43:W237–W243. <https://doi.org/10.1093/nar/gkv437>.
73. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31:1674–1676. <https://doi.org/10.1093/bioinformatics/btv033>.
74. Rho M, Tang H, Ye Y. 2010. FragGeneScan: predicting genes in short and error-prone reads. *Nucleic Acids Res* 38:e191. <https://doi.org/10.1093/nar/gkq747>.
75. van der Maaten L, Hinton G. 2008. Visualizing data using t-SNE. *J Mach Learn Res* 9:2579–2605.
76. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797. <https://doi.org/10.1093/nar/gkh340>.
77. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* 30:2725–2729. <https://doi.org/10.1093/molbev/mst197>.
78. Schrödinger LLC. 2015. The PyMOL molecular graphics system, version 1.8. Schrödinger LLC, New York, NY.
79. Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. *Genome Res* 14:1188–1190. <https://doi.org/10.1101/gr.849004>.
80. Fontecave M, Eliasson R, Reichard P. 1987. NAD(P)H:flavin oxidoreductase of *Escherichia coli*. A ferric iron reductase participating in the generation of the free radical of ribonucleotide reductase. *J Biol Chem* 262:12325–12331. [https://doi.org/10.1016/S0021-9258\(18\)45356-2](https://doi.org/10.1016/S0021-9258(18)45356-2).
81. Fieschi F, Nivière V, Frier C, Décout J-L, Fontecave M. 1995. The mechanism and substrate specificity of the NADPH:Flavin oxidoreductase from *Escherichia coli*. *J Biological Chemistry* 270:30392–30400. <https://doi.org/10.1074/jbc.270.51.30392>.