Genome mining for the discovery of peptide halogenases and their biochemical characterization

Nirmal Saha[a] **and Vinayak Agarwal**[a,b,*]
[a]School of Chemistry & Biochemistry, Georgia Institute of Technology, Atlanta, GA, United States
[b]School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA, United States
*Corresponding author. e-mail address: vagarwal@gatech.edu

Contents

Abstract

While halogenation is one of the most versatile C–H functionalization strategy, regiospecific halogenation of peptides and proteins is outside the purview of traditional chemical catalysis. Enzymes that participate in the biosynthesis of ribosomally synthesized and post-translationally modified peptides and proteins can bridge this gap and offer a biocatalytic route for residue-specific incorporation of halogen handles onto amino acid side chains. Protocols described herein provide a guided approach for the discovery of peptide halogenases in the context of natural product biosynthetic gene clusters, and the preliminary reconstitution of their activity using a bacterial heterologous host. Also described are mass spectrometry-based analytical procedures and data analysis workflows that allow for deconvolution of halide specificity and preliminary insights into peptidic natural product biosynthetic schemes. As the available genomic data expands at a rapid rate, the methodology described here will enable the discovery and characterization of new halogenases that can be valuable partners in chemoenzymatic diversification of peptides and proteins.

Methods in Enzymology
ISSN 0076-6879, https://doi.org/10.1016/bs.mie.2025.02.006

1. Introduction

Halogenation is a versatile C–H activation strategy and is widely used in organic synthesis. Numerous reactions have been developed that leverage the halide activation of carbon centers. Among these are the transition metal-assisted C–C bond forming reactions (Rogge et al., 2021; Zhang, Vinogradova, Spokoyny, Buchwald, & Pentelute, 2019). A distinct advantage to these reactions is that they can be targeted to the C–X handle and are often orthogonal to other functional groups present on the organic chassis. While the utility of halogen activation is thusly undisputed, regiospecific halogen delivery to organic molecules needs to be carefully designed. In chemical synthesis, the regiospecificity of halogenation is *usually* determined by the reactivity of the organic substrate itself and not by the halogen-delivering co-substrate. Thus, halogenation in organic synthesis is usually an early-stage modification when the reactivity of individual building blocks is easier to manage. Sarlah's and Baran's syntheses of darobactin are illustrative examples wherein halogenated intermediates predominate and guide the cross-coupling reactions as is also represented in Boger's synthesis of vancomycin (Fig. 1) (Lin et al., 2022; Moore et al., 2020; Nesic et al., 2022).

Contrary to the rationale for the utilization of halogenated building blocks in chemical synthesis, the enzymological indole bromination during biosynthesis of bromodarobactin is a late stage modification, as is chlorination *en route* vancomycin biosynthesis (Böhringer et al., 2023; Schmartz, Zerbe, Abou-Hadeed, & Robinson, 2014). The challenge in controlling regiospecificity for late-stage halogenation reactions is magnified for peptides and proteins. None of the proteinogenic amino acids are halogenated and multiple electron-rich Tyr, Trp, His, and Phe side chains complicate regiospecific delivery of halogen atoms onto peptides and proteins. To circumvent this challenge, either the genetic code can be expanded such that halogenated amino acids can be incorporated, or amino acid side chains can be chemically modified after protein expression to bear halogenated organic handles. Both strategies have been successfully employed in chemical literature (Dumas et al., 2013; Gao, Gouverneur, & Davis, 2013; Spicer & Davis, 2011; Spicer, Triemer, & Davis, 2012).

Enzymes that participate in the biosynthesis of ribosomally synthesized and post-translationally modified peptides (RiPPs) offer an alternate route for halogen delivery onto peptides (Montalbán-López et al., 2021). The unique advantage of enzymatic halogenation of peptides and proteins is that

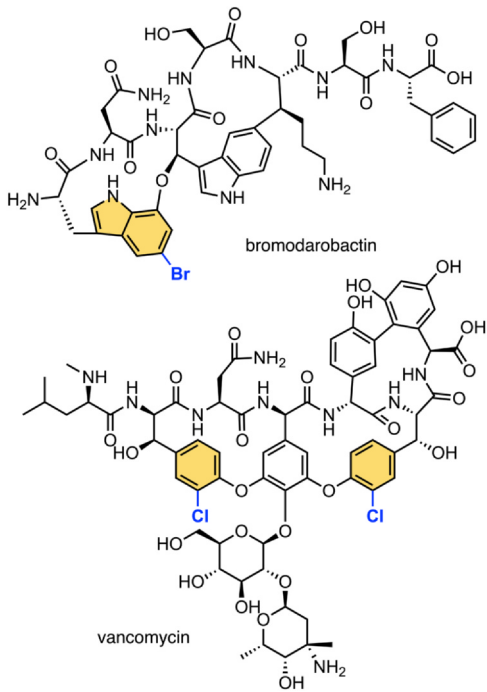

Fig. 1 Halogenated peptidic natural products vancomycin and bromodarobactin with favorable antimicrobial properties.

peptide halogenases are exquisitely selective for the amino acid side chain(s) that is(are) halogenated. Thus, late-stage halogenation of peptides and proteins can be achieved without recourse to genetic code expansion or chemical modification of reactive side chains. Typically, RiPP precursor peptides are divided into an N-terminal leader and a C-terminal core region. The RiPP biosynthetic enzymes, such as peptide halogenases, recognize and bind to the precursor peptide's leader region and modify the C-terminal core. The N-terminal leader is then proteolytically removed from the modified C-terminal core to furnish the halogenated peptide product. Thus, the enzymatic halogenation of RiPP precursor peptides suffers from poor atom economy as the unmodified leader is discarded and cannot be recycled. Efforts to shorten the leader peptides or discover halogenases that do not rely on the presence of the leader peptide are thus significant towards developing RiPP halogenases as viable biocatalysts (Montua & Sewald, 2024). In itself, halogenation can improve the bioactivities of RiPPs, in line with halogenation being critical for the antibiotic

activity of the non-ribosomal peptide synthetase-derived (NRPS-derived) peptidic antibiotic vancomycin (Fig. 1) (Pinchman & Boger, 2013a; 2013b). Bromination also improves the antimicrobial activity of the RiPP darobactin (Böhringer et al., 2023). Thus, the biocatalytic introduction of halogens into RiPPs can also be pharmacologically motivated.

Building on Bibb's description of the biosynthetic gene cluster (BGC), Nair and van der Donk reported the structure and activity of the RiPP halogenase MibH that participates in the biosynthesis of the antimicrobial RiPP NAI-107 by chlorinating the Trp side chain indole (Fig. 2) (Foulston & Bibb, 2010; Ortega et al., 2017). Mitchell has reported the discovery of the RiPP halogenase ChlH that participates in the biosynthesis of a halogenated lasso peptide natural product (Harris et al., 2024). By sequence similarity to MibH, we had identified the RiPP

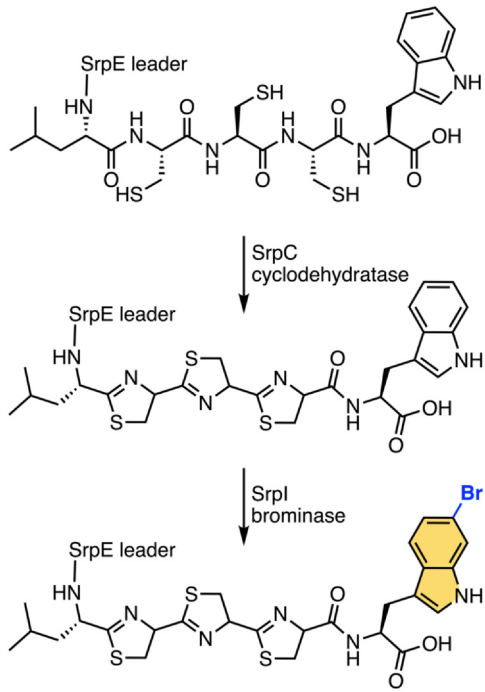

Fig. 2 Activity of the tryptophan-6-brominase SrpI. The cyclodehydratase SrpC dehydratases the three contiguous Cys residues in the SrpE pentapeptide core to install azoline heterocycles followed by bromination by SrpI. Note that SrpI was found to be a physiological brominase such that it did not characterize chlorination (Nguyen et al., 2021).

halogenase SrpI to be widely encoded in marine sponge microbiomes (Nguyen et al., 2021). The activity of SrpI was reconstituted to reveal a linear azoline containing peptide as the physiological substrate (Fig. 2). The promiscuous activity of SrpI allowed for the halogenation of short unmodified peptides as well, and their subsequent chemical modification by Suzuki-Miyaura cross-coupling and Sonogashira coupling reactions (Nguyen & Agarwal, 2023; Saha, Vidya, Xie, & Agarwal, 2024). As mentioned above, the atom economy of halogenated peptide product delivery by SrpI was compromised due to the requirement of the leader peptide. Intermolecular peptide-protein interactions were interrogated using solution NMR to reveal that the leader peptide could be drastically shortened while supporting SrpI activity (Nguyen et al., 2024). Here, in Section 2, we describe the procedures for bioinformatic discovery of peptide halogenases and RiPP-encoding BGCs using SrpI as a starting point. We also describe experiments and data analysis workflows to reconstitute activities of peptide halogenases *in vivo* using heterologous hosts (Section 4).

2. Genome mining for discovery of halogenase encoding BGCs

2.1 Discussion

Prior work from our laboratory has described the activity of the peptide halogenase SrpI (Nguyen & Agarwal, 2023; Nguyen et al., 2021; Nguyen et al., 2024). To discover enzymes with similar peptide halogenation activities, the online Enzyme Function Initiative (EFI) toolkit is used, accessible at https://efi.igb.illinois.edu/ (Gerlt et al., 2015; Zallot, Oberg, & Gerlt, 2019). This toolkit comprises the following components:

a. Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): Using an input query sequence, the EFI-EST tool performs a database search for similar sequences and arranges the output sequences in the form of sequence similarity networks (SSNs).

b. Enzyme Function Initiative-Genome Neighborhood Tool (EFI-GNT): This tool adds genomic context to the SSNs by generating gene neighborhood networks (GNNs). Specifically, this tool facilitates hypothesis formation by inferring activities of enzymes based on the genetic neighborhoods and BGCs that these enzymes are encoded within.

Below, we describe procedures for SSN generation and genome mining using the SrpI sequence as the input query. The user only requires basic computer hardware and infrastructure that is routinely available in research areas to access and use the EFI tools.

2.2 Procedure

The following SrpI sequence derived from marine sponge metagenomes is used as the query:

MIQPGSESLRKIAVIGRGTAGSLAAASVTRLHPDADHELHHI
YDSRIPVIGVGEGSWPSLVQEVQQLTGLPHETVQQRLKGTRKY
GVAFEGWGRRGRDFTHYFTPQQVSYAYHLSADLLADMLHESS
RARHIDAKVLDIARVDGGARVEFEGRAPERYDLVFDARGFPRE
LDTDEHIDISFIPTNTAVIRRCPAIVEEAAGPVLQHTYTRAVARP
HGWIFVIPLAVHTSYGYIFNRDVTGLDEVESDFDAFLETDGVPEF
EQRAVLRFPNFVHRRIYDGAVARIGNAAAFMEPLEATAIVSAQI
QIGMVLKTRLGRSVEHLDRDAPAVNRFLVKNVLRYGLFVGWH
YSCGSRYDSPFWRFARDRTWPRYRSAADPAAVDCNALGEFDE
MIRLLHQPVIDQGDWHRMCAVPLTSYAQMSQGLGC.

1. Using an internet browser, access the EFI tools webpage (https://efi.igb.illinois.edu). First-time users should create an account. Sign in to the user account.
2. Access the EFI-EST tool (https://efi.igb.illinois.edu/efi-est/). Under the 'Sequence BLAST' heading, paste the SrpI sequence above into the 'Query Sequence' text box.
3. Under the 'BLAST Retrieval Options' heading, the default values for the 'UniProt BLAST query e-value,' 'A maximum number of sequences retrieved,' and 'Sequence database' do not need to be changed. The 'UniProt BLAST query e-value' is a critical parameter; increasing this parameter can decrease the number of output sequences and bias the search for highly similar sequences only. Decreasing the parameter can make the output cumbersome to analyze. The user is encouraged to empirically decide the proper e-value parameter to be used by trial and error. Multiple jobs can be started in parallel.
4. Likewise, under the 'SSN Edge Calculation Option' heading, the default 'E-Value' does not need to be changed.
5. Enter a descriptive 'Job name'.
6. Enter an email address where the EFI toolkit can send updates regarding job progression.

7. Click 'Submit Analysis' to start the job. The EFI toolkit will send a confirmation to the email address provided above upon job initiation. The subject of the email will be 'EFI-EST - Initial submission received'.
8. Upon completing the job, the EFI toolkit will send a confirmation to the email address provided above. The subject of the email will be 'EFI-EST - Initial calculation complete'. The email will contain a hyperlink using which the results of the database search can be accessed.
9. Click on the abovementioned hyperlink and under the 'SSN Finalization' heading, enter a numerical value under 'Alignment Score Threshold'. These values typically range from 10 to 200, and multiple values must be empirically trialed for SSN generation.
10. Under the 'Sequence Length Restriction Options', the 'Minimum' sequence length should be set to '200', and the 'Maximum' sequence length should be set to '2000'. No other parameters should be changed from their default values.
11. Provide a unique 'Network name'. As multiple Alignment Score Threshold values (AST values) need to be empirically trialed for SSN generation, we recommend appending the 'Network name' with the AST values. For instance, if the AST value is specified to be '100', then the 'Network name' should be 'SrpI_AST100'. Section 2.3 describes the results for an AST value of 100.
12. Press the 'Create SSN' button. The EFI toolkit will send a confirmation to the email address provided above upon job initiation. The subject of the email will be 'EFI-EST - Your SSN is being finalized'.
13. Upon completing the job, the EFI toolkit will send a confirmation to the email address provided above. The subject of the email will be 'EFI-EST—Your SSN has now been completed and is available for download'. The email will contain a hyperlink to access the results of the SSN generation.
14. Click on the hyperlink mentioned above. The 'SSN Overview' tab provides the 'Analysis Summary,' which lists the input parameters, and the 'Database Summary,' which lists which databases were mined and the number of sequences identified.
15. The 'Network Files' tab lists various modalities for generating different SSNs. Typically, under the 'Representative Node Networks' table, a 95 % ID SSN should be transferred to the 'Genome Neighborhood Tool' that is accessed by hovering over the 'Transfer To:' button. This will transfer the job to the EFI-Genome Neighborhood Tool.

16. click on the 'Generate GNN' button without changing any default parameters. The EFI toolkit will send a confirmation to the email address provided above upon job initiation. The subject of the email will be 'EFI-GNT - SSN submission received'.
17. Upon completing the job, the EFI toolkit will send a confirmation to the email address provided above. The subject of the email will be 'EFI-GNT—GNN computation completed'. The email will contain a hyperlink to access the results of the GNN generation.
18. Click on the hyperlink mentioned above. The 'Submission Summary' tab lists the input parameters in the RESULTS section. The 'Networks and the GND' tab provides access to the Genome Neighborhood Diagrams (GNDs) critical for identifying enzymes similar in sequence to the query (SrpI) in the context of neighboring genes and BGCs.
19. Click the' View diagrams' button in the 'Genome Neighborhood Diagrams (GNDs)' section. This will open a new window.
20. In the 'SEARCH' text box, enter a numerical value starting from '1', and click 'Query'. The webpage will populate with gene neighborhood diagrams. Increasing the value from single-digit increments will populate the next cluster of gene neighborhoods.
21. Hovering over individual genes will deliver their functional annotation. Manually curate the gene neighborhoods to detect halogenase-encoding genes that are placed within BGCs harboring genes that encode RiPP precursor peptides and typical RiPP biosynthetic enzymes, such as lanthionine synthetases, YcaO cyclodehydratases, and peptidases.

2.3 Data analysis

1. Using the procedure delineated above, the EFI toolkit will search for sequences similar to the input query sequence within the bounds of the 'UniProt BLAST query e-value' and the 'Maximum number of sequences retrieved' parameters (see step in Section 2.2, above) and then arrange them into an SSN using the specified 'Alignment Score Threshold' parameter. For the SrpI input sequence delineated in Section 2.2, the SSN illustrated in Fig. 3 contained 789 nodes connected by 196,110 edges. Here, each node represents a protein sequence of a minimum of 200 and a maximum of 2000 amino acids in length (see step 10 in Section 2.2) that is similar to SrpI with a BLAST e-value less than or equal to the specified

'UniProt BLAST query e-value.' Sequences with 95 % or greater similarity (see step 15 in Section 2.2) are condensed into a single node. The SSN was visualized using Cytoscape.

2. The subclusters identified in Cytoscape correlate to the 'SEARCH' parameter specified in step 20 in Section 2.2 above. The cluster with the maximum number of nodes is designated '1', and the next smaller cluster is designated as '2', and so on. Here, the node corresponding to the input SrpI sequence is placed in cluster '2' (yellow node, Fig. 3).
3. Cycling through the GNDs for the different SSN clusters and manual curation of the gene neighborhoods allows for the facile identification that halogenase nodes in cluster 2 correspond to the SrpI-like halogenases derived from the numerous *srp* BGCs that are widespread in marine sponge metagenomes (Fig. 4) (Nguyen et al., 2021).
4. The GNDs for cluster 4, illustrated in Fig. 5, demonstrated that the SrpI-like halogenases were encoded in a neighborhood of genes encoding RiPP biosynthetic enzymes such as ATP-dependent lanthionine synthetases that construct macrocyclizing (methyl)lanthionine rings (Repka, Chekan, Nair, & van der Donk, 2017). Hence, the gene neighborhoods represented in Fig. 5 likely represent BGCs encoding the production of halogenated RiPPs.
5. One of the BGCs identified in cluster 5 is derived from the marine cyanobacterium *Moorena producens,* which we have named the *Moorena producens*-derived RiPP BGC, henceforth referred to as the *mpp* BGC (Fig. 6). The organization of the *mpp* BGC, and the reconstitution of the activity of the SrpI-like peptide halogenase encoded in the *mpp* BGC is described in Section 4. While the RiPPs encoded by the *mpp* BGCs are cryptic, the *mpp* BGC presents an opportunity to explore the activity of a peptide halogenase in the context of RiPP biosynthesis (Saha, Vidya, Luo, van der Donk, & Agarwal, 2025).

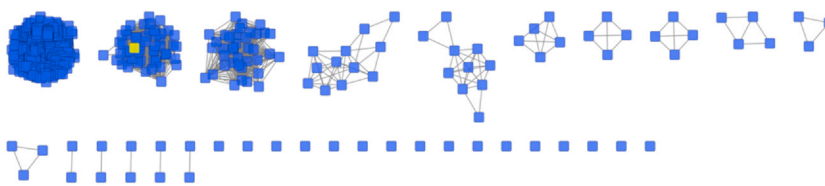

Fig. 3 EFI-generated SSN illustrating similarity-based clustering of protein sequences. The node corresponding to the input SrpI sequence is illustrated in yellow.

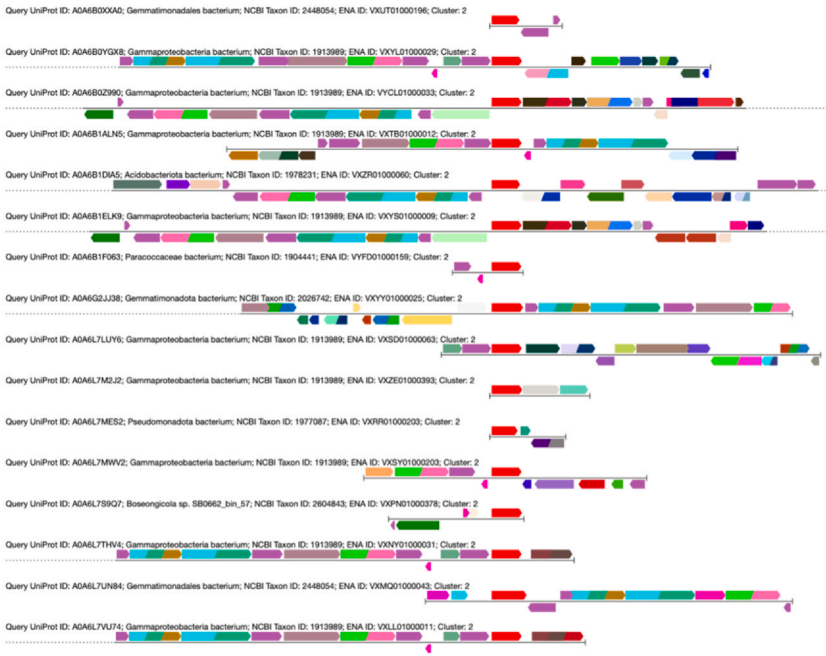

Fig. 4 The GND visualized using the online EFI toolkit corresponding to the SrpI cluster 2. The central gene colored red is the SrpI-like halogenase encoding gene. Hovering over the genes demonstrates their functional annotations. The BGCs illustrated here are derived from a marine sponge-associated Proteobacteria, as has been reported for *srp* BGCs previously.

Fig. 5 The GND visualized using the online EFI toolkit corresponding to the SrpI cluster 5. Hovering over one of the genes in the neighborhood of the SrpI-like halogenase encoding gene shows the annotation as a lanthionine synthetase.

3. Equipment and consumables

Equipment.

1. Temperature-controlled incubators and shakers for bacterial growth
2. Refrigerated centrifuges
3. French press or sonicator
4. Electrophoresis chamber and SDS-PAGE gel casting accessories
5. Transilluminator and gel imaging equipment
6. UV-Vis spectrophotometer
7. Äkta Go liquid chromatography system (Cytiva)
8. Agilent 1260 infinity quaternary liquid chromatography (LC) instrument coupled to an Agilent 6530 C time of flight (ToF) mass spectrometer (MS)

Consumables.

1. *Escherichia coli* BL21Gold(DE3) and DH5α competent cells (New England BioLabs).
2. Chromatography: His-Trap FF 5 mL (Cytiva); Hi-Trap Q 5 mL (Cytiva); Hi-Load 16/600 Superdex 75 pg (Cytiva).
3. Antibiotics: Kanamycin monosulfate, streptomycin, and chloramphenicol (Fisher).
4. Chemicals and reagents: All chemicals and reagents were purchased from Fisher Scientific unless noted otherwise. HEPES, Tris base, potassium chloride, sodium chloride, potassium bromide, magnesium chloride tetrahydrate, imidazole, glycerol, potassium acetate, sodium phosphate monobasic, sodium phosphate dibasic, isopropyl-β-D-1-thiogalactopyranoside (IPTG), L-arabinose (Sigma-Aldrich), FAD (Sigma-Aldrich), nicotinamide adenine dinucleotide (NAD^+) (Sigma-Aldrich), tris(2-carboxyethyl)phosphine hydrochloride (TCEP, Sigma-Aldrich), adenosine 5′-triphosphate disodium salt dihydrate (ATP) (Sigma-Aldrich), protein assay dye reagent (Bio-Rad), Coomassie Brilliant Blue R-250 dye (Bio-Rad), acrylamide and bis-acrylamide and buffers for denaturing sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).
5. Solvents: LC/MS grade acetonitrile, water, and formic acid (Fisher).
6. Media: Terrific broth (TB) (Fisher) and Luria-Bertani (LB) broth.
7. Centrifugal filters: Amicon® Ultra Centrifugal Filters – 30 kDa and 5 kDa MWCO (EMD Millipore).
8. 0.22 μM PVDF syringe filters (Millipore).
9. Dialysis membranes.

4. Heterologous expression of peptide halogenases with substrate peptides for reconstitution of activity

4.1 Discussion

The discovery of the *mpp* BGC using the EFI tools is described in Section 2, above. Harbored within the *mpp* BGC is the *mppE* gene that encodes for the RiPP precursor peptide (Fig. 6). The MppE precursor peptide is divided into the Nif11-like 77 amino acid long N-terminal leader peptide appended the C-terminal heptapeptide core (Haft, Basu, & Mitchell, 2010; Saha et al., 2025).

MppE sequence with the core sequence in boldface: MMSSEQIEQ FVEEIQRDPALKEQLQLQGSIDETIDKVIEIAKEKGYDFTATELKE YMENPSDEEEELSDSELEAVAGG**ACWRWSG**.

The SrpI-like peptide halogenase is encoded by the *mppI* gene, in addition to the macrocyclizing lanthionine synthetase encoded by the *mppM* gene and other tailoring enzymes as annotated in (Fig. 6). Given the organization of the mpp BGC, it is highly likely that the MppI halogenase would accept wild type MppE, or some form of post-translationally modified MppE as a substrate for halogenation. Experiments described in this section will be directed towards activity reconstitution of MppI.

The coexpression of genes encoding the RiPP precursor peptide with genes encoding peptide modifying enzymes in *E. coli* using plasmid vectors is an extensively employed strategy for producing RiPPs (Shi, Yang, Garg, & van der Donk, 2011). Plasmid vectors with well-defined multiple cloning sites (MCSs), peptide fusion tags, inducible promoters, and

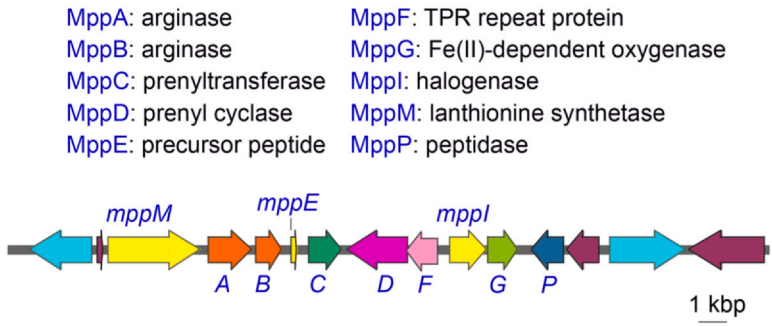

Fig. 6 The *mpp* BGC. The genes *mppM*, *mppE*, and *mppI* encode ATP-dependent lanthionine synthetase, the RiPP precursor peptide, and a flavin-dependent halogenase, respectively. The putative functional annotations of the *mpp* open reading frames are denoted. The genes colored cyan encode putative membrane transporters.

antibiotic selection markers provide the ideal vehicles for gene expression using standard techniques in molecular biology. In the procedures described below, we have employed commercially available plasmid vectors that possess two MCSs and allow for concomitant expression of two genes using a single plasmid backbone. Four vectors with non-overlapping origins of replication and non-overlapping antibiotic selection markers are available; thus, eight genes can be expressed in *E. coli* at a given time. Typically, the RiPP precursor peptide encoding gene is cloned into one of the two MCSs in a vector, and genes encoding the RiPP biosynthetic enzymes are cloned into MCSs of other compatible Duet vectors. Wild-type vectors with no gene inserts can facilitate negative control experiments (*vide infra*). The typical workflow involves the expression of the RiPP precursor peptide with and without concomitant expression of the peptide-modifying enzyme.

4.2 Procedure

4.2.1 Preparation of plasmid DNA vectors and E. coli strains for recombinant protein expression

1. Synthetic genes encoding MppE, MppI, and MppM were obtained from commercial vendors. For brevity, procedures describing introduction of these genes into plasmid vectors is not described. Instead, the reader is directed to excellent technical resources in cloning and standard molecular biology techniques (Lessard, 2013).
2. The *mppE* gene is introduced into the pET28a(+) vector such that a N-terminal hexahistidine tag (N-His_6) is appended to the recombinantly expressed protein. The pET28a(+) vector confers kanamycin resistance to the transformed *E. coli* strain. Plasmids are maintained in transformed *E. coli* DH5α strains under kanamycin selection and all plasmids are verified by Nanopore sequencing.
3. The *mppM* gene is introduced into the MCS-2 of the pCDF-Duet vector with, and without the introduction of the *mppI* gene into the MCS-1 of the same pCDF-Duet vector to yield pCDF-Duet/*mppI*, pCDF-Duet/*mppM*/*mppI* and pCDF-Duet/*mppM* plasmids. The pCDF-Duet vector confers streptomycin resistance to the transformed *E. coli* strain. Plasmids are maintained in transformed *E. coli* DH5α strains under streptomycin selection and all plasmids are verified by Nanopore sequencing.
4. The pET28a(+)/*mppE* vector is transformed into *E. coli* BL21Gold (DE3) cells together with the abovementioned pCDF-Duet vectors in

the following four combinations and transformants selected under kanamycin and streptomycin selection pressure:

a. pET28a(+)/*mppE* with pCDF-Duet
b. pET28a(+)/*mppE* with pCDF-Duet/*mppM*
c. pET28a(+)/*mppE* with pCDF-Duet/*mppI*
d. pET28a(+)/*mppE* with pCDF-Duet/*mppI*/*mppM*

5. To assist in the folding of recombinant MppM and MppI, additionally transform the abovementioned BL21(Gold)DE3 strains with the pGro7 chaperone-encoding plasmid and select the transformants under a combined kanamycin, streptomycin, and chloramphenicol selection pressure in solid and liquid media.

4.2.2 Coexpression of genes encoding the RiPP precursor peptide with genes encoding the RiPP biosynthetic enzyme(s)

1. Plate transformed *E. coli* BL21Gold(DE3) cells containing the combinations of plasmids described above on LB-agar with appropriate antibiotics and incubate at 37 °C for 16 h.
2. Pick a single bacterial colony and inoculate into 10 mL TB media supplemented with appropriate antibiotics.
3. Incubate at 37 °C for 16 h with shaking.
4. Inoculate 2 L TB media with the overnight culture supplemented with appropriate antibiotics (final antibiotic concentrations: streptomycin 50 μg/mL; kanamycin 50 μg/mL; chloramphenicol 34 μg/mL), 1 g/L KBr, and 10 mg/L riboflavin. Incubate the culture flasks at 37 °C with shaking until optical density measured by absorbance at 600 nm wavelength (OD_{600}) reaches 0.4.
5. To induce the expression of chaperones encoded by the pGro7 plamsids, add 0.5 mg/mL (final concentration) L-arabinose and continue incubation at 37 °C with shaking until the OD_{600} reaches 0.6.
6. Reduce the incubation temperature to 18 °C. Allow the culture flasks to incubate for further 1 h.
7. Induce protein expression by adding 0.5 mM IPTG. Incubate at 18 °C for 12–16 h with vigorous shaking.
8. Centrifuge cultures at 4,000g for 30 min at 4 °C and collect cell pellets after decanting off the supernatant. Steps 9–16, below, are conducted on ice or at 4 °C.
9. Resuspend the cell pellet in 50 mL 'cell lysis buffer' with the following composition: 20 mM Tris-Cl (pH 7.9), 500 mM NaCl, 10 % (v/v) glycerol.

10. Lyse the *E. coli* cells by sonication using the manufacturer's recommendations for the sonicator probe, or by passage through a homogenizer.
11. Clarify the cell lysate by centrifugation at at 30,000g for 45 min at 4 °C. Collect the supernatant by decantation.
12. Apply the supernatant to a 5 mL His-Trap FF column that is equilibrated in the abovementioned cell lysis buffer using a peristaltic pump of the FPLC instrument at a flow rate of 1 mL/min.
13. Wash the His-Trap FF column with 60 mL of cell lysis buffer supplemented with 30 mM imidazole using a peristaltic pump of the FPLC instrument at a flow rate of 2 mL/min.
14. Using a FPLC instrument, elute the bound proteins by a linear gradient across 50 mL to the cell lysis buffer supplemented with 200 mM imidazole. Collect 5 mL fractions.
15. Check for the presence of the MppE peptide in the elution fractions using SDS-PAGE and comparison to a molecular weight marker.
16. For the fraction(s) that possess the MppE peptide in high purity, desalt into the cell lysis buffer using PD-10 desalting columns operating under gravity using the manufacturer's recommended procedures.

4.2.3 LC/MS characterization of post-translational tailoring of the precursor peptide

1. The post-translational modifications are affected upon the C-terminal core region of the RiPP precursor peptide. Proteolytic digestion and removal of the N-terminal leader region facilitates MS-based characterization of the core peptide. Incubate 20 μM desalted MppE peptide obtained in Section 4.2.2 above for 2 h at 30 °C with 1 mM DTT and 5 μM of LahT150 peptidase (Bobeica et al., 2019).
2. Add equal volume of 'quenching solution' (MeOH with 2 % (v/v) trifluoroacetic acid (TFA)). Centrifuge at 14,000g at room temperature for 30 min. Carefully withdraw the supernatant into an appropriate vial for LC/MS analysis.
3. Inject 10 μL of this sample onto an LC/MS system. The injection volume may need to be adjusted depending upon the sensitivity of the mass spectrometer. Collect MS data in the positive ionization mode in the MS^1 and MS^2 mass range m/z 50–3000 Da. These mass rages can be empirically adjusted to improve resolution and sensitivity.
4. Small core peptides are typically well retained on an Agilent Poroshell 120 2.7 μm particle size C_{18} reversed-phase column (100 × 4.6 mm).

Maintain a constant flow rate of 0.3 mL/min using the following solvents for the mobile phase:

solvent A: H_2O + 0.1 % (v/v) formic acid; solvent B: MeCN + 0.1 % (v/v) formic acid. The following elution profile can be used (with empirical adjustments as needed):

0–5 min: The chromatography elution profile was as follows: 5 % solvent B.

5–18 min: the linear gradient to 100 % solvent B.

18–22 min: 100 % solvent B.

22–24 min: the linear gradient to 5 % solvent B.

24–30 min: 5 % solvent B from 24 to 30 min

Acquire MS data between 5 min and 24 min. For MppE core peptides, the $[M+2H]^{2+}$ parent ions dominate the MS^1 spectra. The area under the extracted ion chromatograms (EICs) for the $[M+2H]^{2+}$ MS^1 ions can be used to quantify the relative amounts of the substrate and product peptides.

4.3 Data analysis

The procedure described above involves coexpression of the gene encoding the precursor peptide, *mppE*, with genes encoding the lanthionine synthetase (*mppM*) and the flavin-dependent halogenase (*mppI*). The MppE core peptide bears one Cys and one Ser residue that conceivably could be involved in lanthionine ring formation by MppM (Fig. 7) (Repka et al., 2017). This transformation will entail a mass loss of 18 Da as compared to unmodified MppE core peptide and can be detected by the mass spectrometer. Similarly, halogenation at either or both Trp side chains by MppI would entail mass increases that will be detected by the mass spectrometer. Halogenation can proceed before or after lanthionine ring formation and can entail chlorination and/or bromination as the *E. coli* culture media has been supplemented with 1 g/L KBr (step 4, Section 4.2.2). Note that per definition, physiological chlorinases will chlorinate as well as brominate, but physiological brominases will only brominate and not chlorinate (Blasiak & Drennan, 2009). Chlorinated and brominated products possess distinctive isotopic signatures that are easily detected even by modest resolution mass spectrometers. The conceivable chemical changes upon modification of MppE by MppM and MppI are illustrated in Fig. 7. Note that LC/MS characterization described in Section 4.2.3 proceeds after removal of the leader peptide by the LahT150 peptidase (Bobeica et al., 2019).

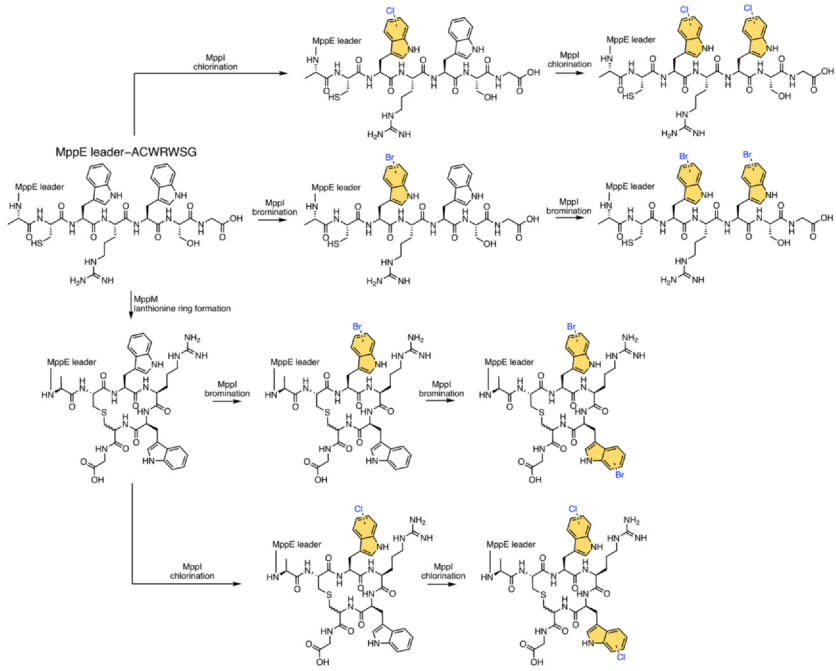

Fig. 7 Scheme for modification of the MppE precursor peptide by lanthionine synthetase MppM and flavin-dependent halogenase MpplI. It is not known, *a priori*, whether MpplI is a chlorinase or a brominase.

1. Per the scheme illustrated in Fig. 7, tabulate the $[M+nH]^{n+}$ (n = 1,2,3) monoisotopic masses for all chemical species illustrated in Fig. 7, considering that the peptide bond between the MppE leader and the Ala residue in the MppE core has been hydrolyzed. Due to the presence of the basic Arg side chain and the primary amine that is unmasked by LahT150 catalyzed hydrolysis of the peptide bond, it is likely that the mass spectrometer would detect multiply charged species of the (modified)MppE core peptide in high abundance. Note the isotopic signature of the chlorinated and brominated molecules.
2. Deconvolute the total ion chromatogram (TIC) recorded by the mass spectrometer during the LC/MS experiment into an MS^1-only base peak chromatogram (BPC). Deconvoluting the TIC into the BPC reduces the data complexity by disregarding the recorded MS^2 fragmentation spectra which makes preliminary data analysis much easier.
3. Manually cycle through the BPC to detect MS^1 ions corresponding to the $[M+nH]^{n+}$ (n = 1,2,3) monoisotopic masses calculated in step 1, above.

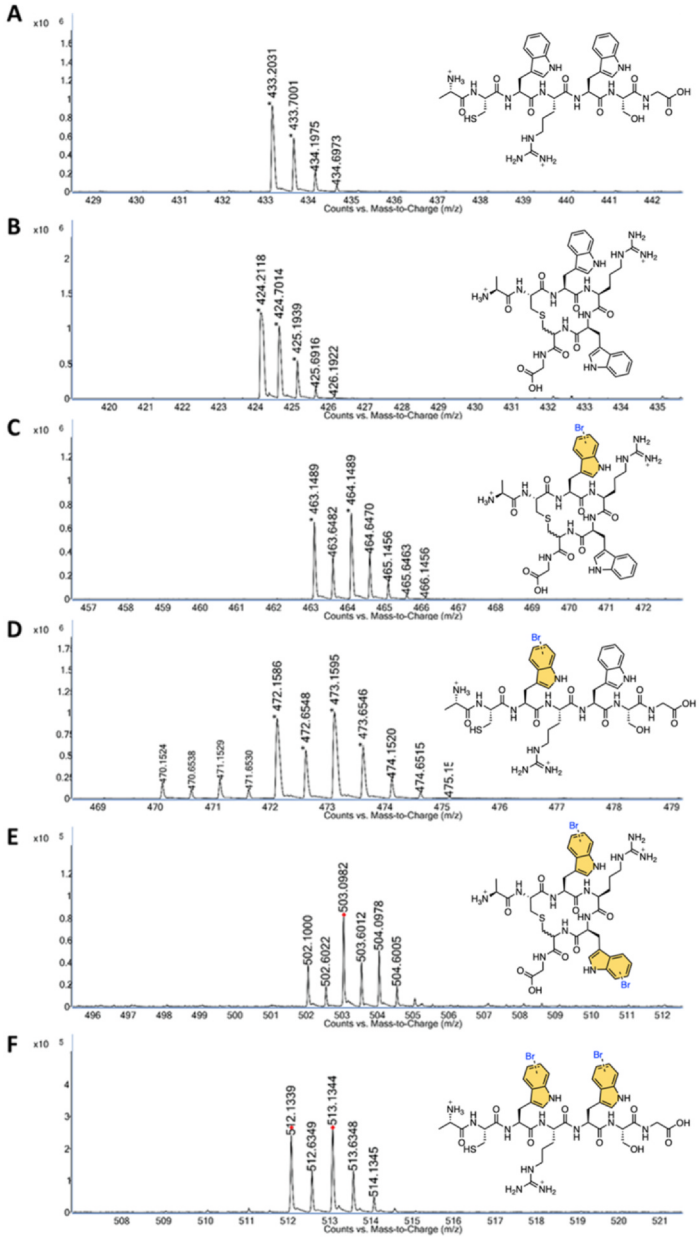

Fig. 8 The isotopic distribution of $[M+2H]^{2+}$ ions corresponding to the LahT150-digested MppE peptides: (A) unmodified MppE core, (B) MppM-modified macrocyclic MppE core, (C) macrocyclized and monobrominated MppE core, (D) linear and monobrominated MppE core, (E) macrocyclized and dibrominated MppE core, and (F) linear and dibrominated MppE core. These products were observed when *mppE* was coexpressed with *mppM* and *mppI* in *E. coli*. Note that no chlorinated products were detected.

Using this approach, it could be discerned that MS^1 ions corresponding to any of the chlorinated species illustrated in Fig. 7 could not be detected leading to characterization of MppI as a physiological brominase (Saha et al., 2025). The *mpp* BGC is derived from a marine cyanobacterium and other marine cyanobacterial halogenases have similarly been identified as physiological brominases with no chlorination activity (Lukowski et al., 2023; Thapa & Agarwal, 2021).

4. Curate the detected MS^1 ions against the theoretical isotopic signatures and the expected error tolerance for the mass spectrometer. The isotopic signatures for the MppE core peptides detected when *mppE* was coexpressed with *mppM* and *mppI* in *E. coli* are illustrated in Fig. 8. For MppE core peptides, the $[M+2H]^{2+}$ ions were observed in higher abundance than other charged species.
5. Alternatively, extracted ion chromatograms (EICs) can be generated that can aid in this search. Note that EIC generation should proceed within the error tolerance for the mass spectrometer and the ions detected that

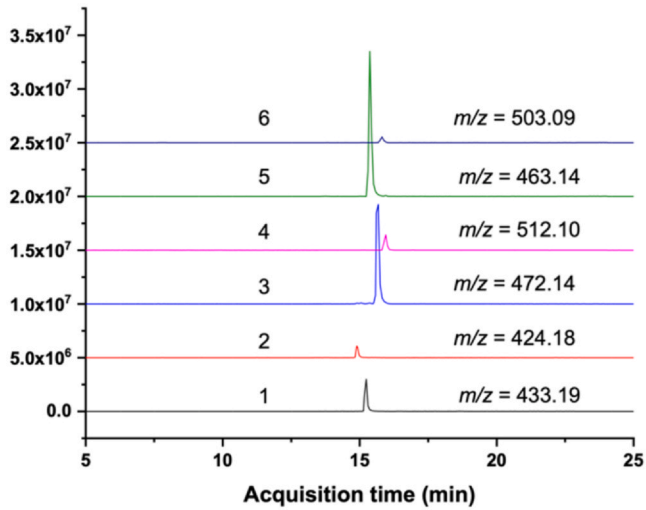

Fig. 9 EICs for $[M+2H]^{2+}$ ions demonstrating the presence of MppE-derived RiPPs when *mppE* was coexpressed with *mppM* and *mppI* in *E coli*. From bottom to top, the following chromatograms are shown: 1–linear unmodified MppE core, 2–MppM-modified macrocyclic MppE core, 3–linear and monobrominated MppE core, 4–linear and dibrominated MppE core, 5–macrocyclized and monobrominated MppE core, and 6–macrocyclized and dibrominated MppE core. It is evident that the monobrominated products predominate, and that while this experiment establishes the bromination activity of MppI, it does not inform whether the physiological substrate for MppI is the linear unmodified MppE peptide, or MppM-modified macrocyclic MppE peptide.

constitute high abundance areas within the EIC should be compared to the theoretical isotopic signatures. The EICs for MppE core peptides when *mppE* was coexpressed with *mppM* and *mppI* in *E. coli* are illustrated in Fig. 9. With the assumption that the ionization efficiencies for differently modified core peptides are identical, the EICs can also inform the relative abundances of different RiPPs that are detected in a single LC/MS experiment which can be aided by the generation of calibration curves to correlate mass spectrometric abundances and/or area under the EIC to absolute peptide abundance in the sample.

6. Topology of macrocyclic ring formation can be discerned by annotation of MS^2 fragmentation spectra. In some instances, particularly for linear peptides, the site for halogen addition can also be discerned from the MS^2 fragmentation spectra. If the post-translational modifications cannot be discerned using fragmentation spectra alone, complementary

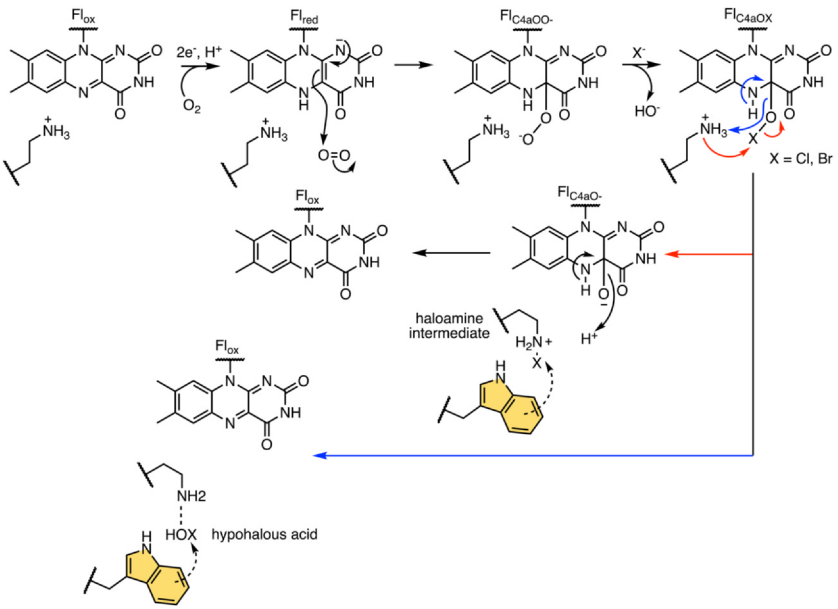

Fig. 10 Proposed routes for oxidative halogenation by flavin-dependent halogenases. The oxidized flavin cofactor (Fl_{ox}) is reduced, and the reduced flavin cofactor (Fl_{red}) coordinates with molecular oxygen to generate the flavin-C4a peroxide (Fl_{C4aOO-}). The addition of halide displaces one of the oxygen atoms, leading to the formation of a Fl_{C4aOX} intermediate which is resolved either by haloamine intermediate formation (red arrows) or by displacement of hypohalous acid in which the Lys side chain acts as the catalytic base to protonate the hypohalite. In either scenario, the Lys side chain delivers the halonium to the hydrocarbon binding site that is distal to the flavin cofactor.

approaches such as *de novo* structure determination by NMR spectroscopy and chemical degradation can be used. Structure determination for the MppE-derived RiPPs has been recently reported (Saha et al., 2025).

The procedures described above allow for biochemical activity reconstitution of peptide halogenases mined from sequence databases. Full structure elucidation of the halogenated peptide product is out of the scope of a single protocol.

5. Conclusions

With the preliminary results illustrated in Fig. 5, it is evident that several putative peptide halogenases can be detected by genome mining that are encoded within RiPP BGCs. Protocols described herein can facilitate the rapid determination of the halogenation bioactivity of these peptide halogenases, which can then be supplemented with more in depth biochemical and structural investigations. As a starting point, leveraging *E. coli* as a heterologous expression system can lead to prioritizing BGCs with functional enzymes for further studies.

The flavin-dependent halogenases all possess a conserved catalytic Lys residue, the side chain of which is postulated to coordinate to the oxidized halogen atom either by the formation of a covalent haloamine or by hydrogen bonding to the hypohalous acid (Fig. 10) (Dong et al., 2005; Flecks et al., 2008; Phintha et al., 2021; Yeh et al., 2006). These catalytic Lys residues are conserved within peptide halogenases and well, and mutagenesis of these Lys residues leads to a complete loss in enzyme activity (Nguyen et al., 2021; Ortega et al., 2017; Saha et al., 2025). For small molecule halogenases, the binding site of the electron rich substrate is distant to the flavin cofactor; the catalytic Lys residues are implicated in bridging the site of halide oxidation proximal to the flavin isoalloxazine ring and the hydrocarbon substrate binding site distal to the flavin cofactor (Fig. 10) (Agarwal et al., 2017). While the substrate binding site for peptide halogenases has not yet been identified, the catalytic Lys residues in MibH, SrpI, MppI, and ChlH could be playing similar roles.

Another open question is the mode of leader peptide engagement for peptide halogenases that participate in RiPP biosynthesis. Prior work from van der Donk and Nair has demonstrated that MibH does not need the leader peptide in the corresponding precursor peptide for activity.

Intriguingly, MibH could only chlorinate the peptide indolic side chain after the leader peptide has been removed (Ortega et al., 2017). In contrast, SrpI required the presence of the leader peptide, and specific peptide-protein intermolecular interactions were necessary to guide the activity of SrpI (Nguyen & Agarwal, 2023; Nguyen et al., 2024). Similar observations were also made for MppI recently, in which the leader peptide was an obligate requirement for halogenation activity (Saha et al., 2025). Thus, the leader peptide requirement is not generalizable, and needs to be biochemically evaluated for each peptide halogenase independently. From a biocatalysis point-of-view, a peptide halogenating enzyme that is broadly substrate promiscuous and whose activity can be guided by a relatively short leader peptide is desirable. Genome mining and biochemical characterization of peptidic halogenases can deliver these biocatalysts that can then interface with traditional methods in chemical synthesis to explore the pharmacological and biotechnological applications of hypermodified peptides.

Acknowledgment

We gratefully acknowledge support from the National Institutes of Health (NIH; R35GM142882 to V.A.).

References

Agarwal, V., Miles, Z. D., Winter, J. M., Eustaquio, A. S., El Gamal, A. A., & Moore, B. S. (2017). Enzymatic halogenation and dehalogenation reactions: Pervasive and mechanistically diverse. *Chemical Reviews, 117*(8), 5619–5674. https://doi.org/10.1021/acs.chemrev.6b00571.

Blasiak, L. C., & Drennan, C. L. (2009). Structural perspective on enzymatic halogenation. *Accounts of Chemical Research, 42*(1), 147–155. https://doi.org/10.1021/ar800088r.

Bobeica, S. C., Dong, S., Huo, L., Mazo, N., McLaughlin, M. I., Jiménez-Osés, G., ... van der Donk, W. A. (2019). Insights into AMS/PCAT transporters from biochemical and structural characterization of a double glycine motif protease. *eLife, 8*, 1–27.e42305. https://doi.org/10.7554/eLife.42305.

Böhringer, N., Kramer, J., Mora, E., Padva, L., Wuisan, Z. G., Liu, Y., ... Schäberle, T. F. (2023). Genome- and metabolome-guided discovery of marine BamA inhibitors revealed a dedicated darobactin halogenase. *Cell Chemical Biology, 30*(8), 943–952.e7. https://doi.org/10.1016/j.chembiol.2023.06.011.

Dong, C., Flecks, S., Unversucht, S., Haupt, C., van, Pee, K. H., & Naismith, J. H. (2005). Tryptophan 7-halogenase (PrnA) structure suggests a mechanism for regioselective chlorination. *Science, 309*(5744), 2216–2219. https://doi.org/10.1126/science.1116510.

Dumas, A., Spicer, C. D., Gao, Z., Takehana, T., Lin, Y. A., Yasukohchi, T., & Davis, B. G. (2013). Self-liganded Suzuki–Miyaura coupling for site-selective protein PEGylation. *Angewandte Chemie International Edition, 52*(14), 3916–3921. https://doi.org/10.1002/anie.201208626.

Flecks, S., Patallo, E. P., Zhu, X., Ernyei, A. J., Seifert, G., Schneider, A., ... van Pee, K. H. (2008). New insights into the mechanism of enzymatic chlorination of tryptophan. *Angewandte Chemie (International Ed. in English), 47*(49), 9533–9536. https://doi.org/10.1002/anie.200802466.

Foulston, L. C., & Bibb, M. J. (2010). Microbisporicin gene cluster reveals unusual features of lantibiotic biosynthesis in actinomycetes. *Proceedings of the National Academy of Sciences, 107*(30), 13461–13466. https://doi.org/10.1073/pnas.100828510.

Gao, Z., Gouverneur, V., & Davis, B. G. (2013). Enhanced aqueous Suzuki–Miyaura coupling allows site-specific polypeptide 18F-labeling. *Journal of the American Chemical Society, 135*(37), 13612–13615. https://doi.org/10.1021/ja4049114.

Gerlt, J. A., Bouvier, J. T., Davidson, D. B., Imker, H. J., Sadkhin, B., Slater, D. R., & Whalen, K. L. (2015). Enzyme function initiative-enzyme similarity tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, 1854*(8), 1019–1037. https://doi.org/10.1016/j.bbapap.2015.04.015.

Haft, D. H., Basu, M. K., & Mitchell, D. A. (2010). Expansion of ribosomally produced natural products: A nitrile hydratase- and Nif11-related precursor family. *BioMedicalCentral, Biology, 8*(70), 2–15. https://doi.org/10.1186/1741-7007-8-70.

Harris, L. A., Saad, H., Shelton, K. E., Zhu, L., Guo, X., & Mitchell, D. A. (2024). Tryptophan-centric bioinformatics identifies new lasso peptide modifications. *Biochemistry, 63*(7), 865–879. https://doi.org/10.1021/acs.biochem.4c00035.

Lessard, J. C. (2013). Chapter seven - molecular cloning. In J. Lorsch (Ed.). *Methods in Enzymology* (pp. 85–98)Academic Press.

Lin, Y., Schneider, F., Eberle, K. J., Chiodi, D., Nakamura, H., Reisberg, S. H., ... Baran, P. S. (2022). Atroposelective total synthesis of darobactin A. *Journal of the American Chemical Society, 144*(32), 14458–14462. https://doi.org/10.1021/jacs.2c05892.

Lukowski, A. L., Hubert, F. M., Ngo, T., Avalon, N. E., Gerwick, W. H., & Moore, B. S. (2023). Enzymatic halogenation of terminal alkynes. *Journal of the American Chemical Society, 145*(34), 18716–18721. https://doi.org/10.1021/jacs.3c05750.

Montalbán-López, M., Scott, T. A., Ramesh, S., Rahman, I. R., van Heel, A. J., Viel, J. H., ... van der Donk, W. A. (2021). New developments in RiPP discovery, enzymology and engineering. *Natural Product Reports, 38*(1), 130–239. https://doi.org/10.1039/D0NP00027B.

Montua, N., & Sewald, N. (2024). Perfect partners: Biocatalytic halogenation and metal catalysis for protein bioconjugation. *Chembiochem: a European Journal of Chemical Biology, 25*(23), 1–11.e202400496. https://doi.org/10.1002/cbic.202400496.

Moore, M. J., Qu, S., Tan, C., Cai, Y., Mogi, Y., Keith, D. J., & Boger, D. L. (2020). Next-generation total synthesis of vancomycin. *Journal of the American Chemical Society, 142*(37), 16039–16050. https://doi.org/10.1021/jacs.0c07433.

Nesic, M., Ryffel, D. B., Maturano, J., Shevlin, M., Pollack, S. R., Gauthier, D. R., ... Sarlah, D. (2022). Total synthesis of darobactin a. *Journal of the American Chemical Society, 144*(31), 14026–14030. https://doi.org/10.1021/jacs.2c05891.

Nguyen, N. A., & Agarwal, V. (2023). A leader-guided substrate tolerant RiPP brominase allows Suzuki–Miyaura cross-coupling reactions for peptides and proteins. *Biochemistry, 62*(12), 1838–1843. https://doi.org/10.1021/acs.biochem.3c00222.

Nguyen, N. A., Lin, Z., Mohanty, I., Garg, N., Schmidt, E. W., & Agarwal, V. (2021). An obligate peptidyl brominase underlies the discovery of highly distributed biosynthetic gene clusters in marine sponge microbiomes. *Journal of the American Chemical Society, 143*(27), 10221–10231. https://doi.org/10.1021/jacs.1c03474.

Nguyen, N. A., Vidya, F. N. U., Yennawar, N. H., Wu, H., McShan, A. C., & Agarwal, V. (2024). Disordered regions in proteusin peptides guide post-translational modification by a flavin-dependent RiPP brominase. *Nature Communications, 15*, 1265. https://doi.org/10.1038/s41467-024-45593-5.

Ortega, M. A., Cogan, D. P., Mukherjee, S., Garg, N., Li, B., Thibodeaux, G. N., ... van der Donk, W. A. (2017). Two flavoenzymes catalyze the post-translational generation of 5-chlorotryptophan and 2-aminovinyl-cysteine during NAI-107 biosynthesis. *ACS Chemical Biology, 12*(2), 548–557. https://doi.org/10.1021/acschembio.6b01031.

Phintha, A., Prakinee, K., Jaruwat, A., Lawan, N., Visitsatthawong, S., Kantiwiriyawanitch, C., ... Chaiyen, P. (2021). Dissecting the low catalytic capability of flavin-dependent halogenases. *Journal of Biological Chemistry, 296*(100068), 1–16. https://doi.org/10.1074/jbc.RA120.016004.

Pinchman, J. R., & Boger, D. L. (2013). Investigation into the functional impact of the vancomycin C-ring aryl chloride. *Bioorganic & Medicinal Chemistry Letters, 23*(17), 4817–4819. https://doi.org/10.1016/j.bmcl.2013.06.080.

Pinchman, J. R., & Boger, D. L. (2013). Probing the role of the vancomycin E-ring aryl chloride: Selective divergent synthesis and evaluation of alternatively substituted E-ring analogues. *Journal of Medicinal Chemistry, 56*(10), 4116–4124. https://doi.org/10.1021/jm4004494.

Repka, L. M., Chekan, J. R., Nair, S. K., & van der Donk, W. A. (2017). Mechanistic understanding of lanthipeptide biosynthetic enzymes. *Chemical Reviews, 117*(8), 5457–5520. https://doi.org/10.1021/acs.chemrev.6b00591.

Rogge, T., Kaplaneris, N., Chatani, N., Kim, J., Chang, S., Punji, B., ... Ackermann, L. (2021). C–H activation. *Nature Reviews Methods Primers, 1*, 43. https://doi.org/10.1038/s43586-021-00041-2.

Saha, N., Vidya, F. N. U., Luo, Y., van der Donk, W. A., & Agarwal, V. (2025). Transformation-guided genome mining provides access to brominated lanthipeptides. *Organic Letters, 27*(4), 984–988. https://doi.org/10.1021/acs.orglett.4c04529.

Saha, N., Vidya, F. N. U., Xie, R., & Agarwal, V. (2024). Halogenase-assisted alkyne/aryl bromide sonogashira coupling for ribosomally synthesized peptides. *Journal of the American Chemical Society, 146*(44), 30009–30013. https://doi.org/10.1021/jacs.4c12210.

Schmartz, P. C., Zerbe, K., Abou-Hadeed, K., & Robinson, J. A. (2014). Bis-chlorination of a hexapeptide–PCP conjugate by the halogenase involved in vancomycin biosynthesis. *Organic & Biomolecular Chemistry, 12*(30), 5574–5577. https://doi.org/10.1039/C4OB00474D.

Shi, Y., Yang, X., Garg, N., & van der Donk, W. A. (2011). Production of lantipeptides in Escherichia coli. *Journal of the American Chemical Society, 133*(8), 2338–2341. https://doi.org/10.1021/ja109044r.

Spicer, C. D., & Davis, B. G. (2011). Palladium-mediated site-selective Suzuki–Miyaura protein modification at genetically encoded aryl halides. *Chemical Communications, 47*(6), 1698–16700. https://doi.org/10.1039/C0CC04970K.

Spicer, C. D., Triemer, T., & Davis, B. G. (2012). Palladium-mediated cell-surface labeling. *Journal of the American Chemical Society, 134*(2), 800–803. https://doi.org/10.1021/ja209352s.

Thapa, H. R., & Agarwal, V. (2021). Obligate brominating enzymes underlie bromoform production by marine cyanobacteria. *Journal of Phycology, 57*(4), 1131–1139. https://doi.org/10.1111/jpy.13142.

Yeh, E., Cole, L. J., Barr, E. W., Bollinger, J. M., Ballou, J. D. P., & Walsh, C. T. (2006). Flavin Redox chemistry precedes substrate chlorination during the reaction of the flavin-dependent halogenase RebH. *Biochemistry, 45*(25), 7904–7912. https://doi.org/10.1021/bi060607d.

Zallot, R., Oberg, N., & Gerlt, J. A. (2019). The EFI web resource for genomic enzymology tools: Leveraging protein, genome, and metagenome databases to discover novel enzymes and metabolic pathways. *58*(41), 4169–4182. https://doi.org/10.1021/acs.biochem.9b00735.

Zhang, C., Vinogradova, E. V., Spokoyny, A. M., Buchwald, S. L., & Pentelute, B. L. (2019). Arylation chemistry for bioconjugation. *Angewandte Chemie International Edition, 58*(15), 4810–4839. https://doi.org/10.1002/anie.201806009.