Biosynthesis of polybrominated aromatic molecules by marine bacteria

A dissertation submitted in partial satisfaction of the requirements for the Degree Doctor of Philosophy in

Marine Biology

by

Abraham Abbas El Gamal

Committee in charge:

Professor Bradley S. Moore, Chair
Professor Lihini I. Aluwihare
Professor Peter C. Dorrestein
Professor Amro Hamdoun
Professor Chambers C. Hughes

2016
The dissertation of Abrahim Abbas El Gamal is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

________________________________________

________________________________________

________________________________________

________________________________________

Chair

University of California, San Diego

2016
Dedication

For nana
# Table of Contents

Signature Page ........................................................................................................... iii
Dedication ................................................................................................................... iv
Table of Contents ....................................................................................................... v
List of Figures ............................................................................................................. vi
List of Tables ............................................................................................................... x
Acknowledgements .................................................................................................. xi
Vita .............................................................................................................................. xiii
Abstract of the Dissertation ...................................................................................... xiv
Chapter 1: Introduction ............................................................................................. 1
Chapter 2: Biosynthesis of polybrominated aromatic molecules by marine bacteria .......................................................................................... 25
Chapter 3: Biosynthesis of coral settlement cue tetrabromopyrrole in marine bacteria by a uniquely adapted brominase-thioesterase enzyme pair .............................................................................. 91

Chapter 4: An alkylhydroperoxidase-like debrominase is an enzymatic switch in the biosynthesis of pentabromopseudilin .................................................................................................................. 124
Chapter 5: Opportunities in halopyrrole biosynthesis ............................................ 170
References .................................................................................................................. 186
# List of Figures

Figure 1.1: Halogens matter. Examples of natural products whose biological is demonstrated to be influenced by halogenations .................................................. 2
Figure 1.2: Schematic for MCD assay for detection of electrophilic halogen species. ... 5
Figure 1.3: Representative diversity of halogenated natural products ........................................ 7
Figure 1.4: Flavin-dependent halogenases ........................................................................ 9
Figure 1.5: Mechanistic proposal for halogenation-mediated atropo-selective N-C coupling in the biosynthesis of Marinopyrrole A ................................................. 17
Figure 1.6: A modular polybrominated pyrrole/phenol (Bmp) biosynthetic pathway from marine bacteria .................................................................................. 18
Figure 1.7: A dehalogenase switch in pentabromopseudilin biosynthesis ...................... 20
Figure 1.8: A metagenomics approach toward elucidation of pyrrole-imidazole alkaloid biosynthesis .................................................................................. 22
Figure 1.9: Bi-modular scheme for the biosynthesis of polybrominated marine natural products by the bmp pathway ......................................................... 24

Supplementary Figure 2.1: Production of signature polybrominated molecules 1–3 by marine bacteria .................................................................................. 45
Supplementary Figure 2.2: High-resolution mass spectra for selected polybrominated metabolites from an organic extract of P. luteoviolacea 2ta16 ....................... 46
Supplementary Figure 2.3: Chemical logic for the biosynthesis of halogenated pyrrole moieties in natural products ................................................................. 47
Supplementary Figure 2.4: pGAL-bmp1–7 yeast/E. coli shuttle vector ............................ 48
Supplementary Figure 2.5: Chemotype for bmp7 deletion ........................................ 49
Supplementary Figure 2.6: Plasmids for co-expression of bmp1 with bmp3 ............... 50
Supplementary Figure 2.7: MS2 product ions corresponding to phosphopantetheine, phosphopantetheine acylated with pyrrole, and subsequent brominated states of the pyrrole ................................................................. 51
Supplementary Figure 2.8: Enzymatic synthesis of pyrrolyl-S-Bmp1(ACP) by Bmp3 and Bmp4 .................................................................................................. 52
Supplementary Figure 2.9: 1H NMR spectrum for 16 ................................................... 53
Supplementary Figure 2.10: Hydrolysis of model esterase substrate p-nitrophenyl acetate (pNPA) by Bmp1(TE) ................................................................. 54
Supplementary Figure 2.11: Chemotype for heterologous expression of bmp1–7 omitting bmp8 .................................................................................................. 55
Supplementary Figure 2.12: Spectral characteristics of CYP450 Bmp7 and spectrophotometric determination of substrate tolerance ........................................ 56
Supplementary Figure 2.13: *In vitro* Bmp7 reaction products with 6 (black, top curve) and 7 (bottom, red curve) as substrates. 58
Supplementary Figure 2.14: Characterization of 2 generated by Bmp7 *in vitro* reaction with 6 as a substrate. 59
Supplementary Figure 2.15: Characterization of 7 generated by Bmp7 *in vitro* reaction with 6 as a substrate. 61
Supplementary Figure 2.16: Purification of recombinant Bmp5 and time course for the change in concentrations of the substrate, intermediate and product species as shown in Fig. 4a. 62
Supplementary Figure 2.17: Mass spectra for 8 and 4, and in vivo production of 4 and 5 by Bmp5. 63
Supplementary Figure 2.18: Negative control reactions demonstrate the requirement of KBr and NADPH for catalysis by Bmp5. 64
Supplementary Figure 2.19: Comparison of retention times and NMR spectra of polybrominated biphenyl products 1 and 10 generated by Bmp7 *in vitro* with authentic synthetic standards. 65
Supplementary Figure 2.20: Mass spectrometric analysis of the polybrominated biphenyls, and OH-BDE products generated by Bmp7. 67
Supplementary Figure 2.21: 1H NMR spectrum for 11. 68
Supplementary Figure 2.22: Identification of 2′-OH-BDE-68 (13) as a reaction product generated by Bmp7. 69
Supplementary Figure 2.23: HPLC detection of products generated by Bmp7 using 5 as a substrate. 70
Supplementary Figure 2.24: MS and MS/MS profiles for 14, produced enzymatically by reaction of Bmp7 with 5. 71
Supplementary Figure 2.25: Mass spectrometric analysis of the biphenyl and hydroxylated diphenyl ether products formed by Bmp7 reaction with 17. 72
Supplementary Figure 2.26: Characterization of 3 generated by Bmp7 in an *in vitro* reaction with 4 and 6 as substrates. 73
Supplementary Figure 2.27: Characterization of 15 generated by Bmp7 in an *in vitro* reaction with 4 and 6 as substrates. 75
Supplementary Figure 2.28: Production of iodinated bromophenols by heterologous host E. coli expressing bmp1–8 grown in the presence of KI. 77
Supplementary Figure 2.29: Mass spectrometry based characterization of a polybrominated phenol-pyrrole ether product generated by Bmp7. 79
Supplementary Figure 2.30: Primary sequence analysis identifies Bmp5 to be homologous to flavin-dependent monoxygenases. Primary sequence of Bmp5 was provided as an input to the NCBI BLAST program and analyzed against the ‘Non-redundant protein sequence (nr)’ dataset. All homologs of Bmp5. 81
Figure 3.1: Structures and biosynthesis of halopyrrole natural products. 93
Figure 3.2: A genetic basis for production of 1. 93
Figure 3.3: Enzymatic synthesis of 1. 94
Figure 3.4: Preparation of 8 and *in vitro* conversion to 1 by Bmp2 and Bmp1(TE). 94
Figure 3.5: Scheme for bromination-dependent pyrrole offloading and decarboxylation. 95
Figure 3.6: Structural basis for tetrahalogenation activity of Bmp2

Figure S3.1: LC/MS comparisons of in vivo and in vitro EICs to those of authentic standards of 1, 6, and 7

Figure S3.2: Confirmation of CoA extension and loading of pyrrole substrates onto ACPs

Figure S3.3: Bioinformatic identification and in vitro confirmation of catalytic Ser202 of Mm_Bmp1(TE) with a model chromogenic substrate, p-nitrophenyl acetate (pNPA).

Figure S3.4: Non-enzymatic hydrolysis of proposed ACP-bound tetrabrominated intermediate yields 1

Figure S3.5: Interrogation of the potential of 2,3,4-tribromopyrrole 6 and 7 to serve as intermediates in the biosynthesis of 1

Figure S3.6: Reductive debromination of 1 by NADPH yields 6

Figure S3.7: FAD-bound holo structures of pyrrolyl-S-ACP flavin-dependent halogenases from this study

Figure S3.8: Lack of chlorination activity for Bmp2-TM

Figure S3.9: In vivo coexpression of Mm_bmp1, Mm_bmp3, and Mm_bmp4 with wildtype Mm_bmp2 (Mm_bmp2-wt) or Mm_bmp2 triple mutant (Mm_bmp2-TM)

Figure S3.10: $^1$H NMR spectrum of 1 (CDCl₃, 500 MHz)

Figure S3.11: $^1$H NMR spectrum of 7 (CDCl₃, 500 MHz)

Figure S3.12: $^{13}$C NMR spectrum of 7 (CDCl₃, 125 MHz)

Figure S3.13: Scheme for the synthesis of 9

Figure S3.14: $^1$H NMR spectrum of Boc-protected S1 (CDCl₃, 500 MHz)

Figure S3.15: $^1$H NMR spectrum of 9 (d₆-DMSO, 500 MHz)

Figure S3.16: $^{13}$C NMR spectrum of 9 (d₆-DMSO, 150 MHz)

Figure S3.17: DQF-COSY spectrum of 9 (d₆-DMSO, 600 MHz)

Figure S3.18: HSQC spectrum of 9 (d₆-DMSO, 600 MHz)

Figure S3.19: HMBC spectrum of 9 (d₆-DMSO, 600 MHz)

Figure S3.20: Sequence alignments of Bmp2 homologs

Figure 4.1: Bmp biosynthetic pathway

Figure 4.2: Debromination of bromopyrroles by Bmp8

Figure 4.3: Redox state of Bmp8-CXXC

Figure 4.4: Biochemical interrogation of Bmp8-CXXC

Figure 4.5: Biochemical interrogation of the conserved CXYCXXH of Mm_Bmp8 and a distant homolog 2ta16_Bmp8-H

Figure 4.6: Bmp8 mechanistic proposal

Figure S4.1: Affinity purification of N-His6-Bmp8 by co-expression with Bmp1(TE), and cleavage of N-His6 affinity tag

Figure S4.2: Dechlorination of 2,3,4,5-tetrachloropyrrole catalyzed by Bmp8

Figure S4.3: Bmp8 homologs from (meta)genomic datasets

Figure S4.4: Bmp8-CXXC is reduced in enzyme as-purified

Figure S4.5: Conserved Bmp8 Cys57 and Cys163 do not play a direct role in Bmp8 reaction mechanism
List of Tables

Supplementary Table 2.1: High-resolution MS used for determination of molecular formulae of polybrominated metabolites from extract of *P. luteoviolacea* 2ta16.... 83

Supplementary Table 2.2: *P. luteoviolacea* 2ta16 bmp gene cluster annotation......... 84

Supplementary Table 2.3: Primers used to generate overlapping fragments for yeast assembly of yeast/E. coli shuttle vector pGAL-bmp1–7........................................ 85

Supplementary Table 2.4: Primers used to generate gene deletions from pETDuet-bmp. ................................................................................................................ 86

Supplementary Table 2.5: Co-transformations of E. coli BL21(DE3) for heterologous expression of bmp cluster and derivatives....................................................... 87

Table S3.1: Bmp homology................................................................. 120

Table S3.2. Data collection and refinement statistics for Mpy16, Bmp2, and Bmp2-TM structures.......................................................... 121

Table S4.1. Mass shift calculation from whole-protein mass spectrometry analysis…. 164
Acknowledgements

Many thanks to my advisor Bradley S. Moore, for taking me on as a Ph.D. student and providing me with a rigorous and forgiving environment in which to grow as a researcher, mentor, and person. Many thanks to Kazuya Yamanaka who was my first teacher of molecular biology in the laboratory and who provided me with the foundations and tools to become an independent researcher. Many thanks to Vinayak Agarwal who was my first teach in biochemistry in the laboratory, and with whom I have enjoyed a close and productive collaboration for the major part of my Ph.D. Bradley Moore, Kazuya Yamanaka, and Vinayak Agarwal have truly lead by example with their patience and remarkable mentoring skills, and have made this Ph.D. an enriching, stimulating, and worthwhile experience.

Chapter 2, in full, is a reprint of materials as it appears in “Biosynthesis of polybrominated aromatic molecules by marine bacteria” in Nature Chemical Biology, 2014, Agarwal V., El Gamal A., Yamanaka K., Poth D., Kersten R. D., Schorn M., Allen E. E., and Moore, B. S., The dissertation author was one of two equally contributing primary investigators and authors of this paper.

Chapter 3, in full, is a reprint of materials as it appears in “Biosynthesis of a coral chemical cue tetrabromopyrrole in marine bacteria by a uniquely adapted brominase-thioesterase enzyme pair” in Proceedings of the National Academy of Sciences U.S.A., 2016. El Gamal A., Agarwal V., Diethelm S., Rahman I., Schorn M., Sneed J.M., Louie G.V., Whalen, K.E., Mincer T.J., Noel J.P., Paul V.J., and Moore B.S. The dissertation author was one of two equally contributing primary investigators and authors of this paper.
Chapter 4, in full, is currently being prepared for submission of the material.

El Gamal A., Agarwal V., Rahman I., and Moore B.S. The dissertation author is the primary investigator and author of this paper.
Vita

2011 – 2016  Ph.D., University of California, San Diego
2009 – 2011  M.S., Stanford University
2005 – 2009  B.S., Stanford University

Publications


Abstract of the Dissertation

Biosynthesis of polybrominated aromatic molecules by marine bacteria

by

Abraham Abbas El Gamal

Doctor of Philosophy in Marine Biology

University of California, San Diego, 2016

Professor Bradley S. Moore, Chair

Nature produces a plethora of bioactive secondary metabolites with diverse hydrocarbon scaffolds and bioactivities. These natural products serve as signaling and defense molecules in their physiological settings, and as inspiration for therapeutics in the clinic. Natural product scaffolds are elaborated with functional groups that influence their biological activities. One such chemical functionality found in thousands of natural product molecules is halogenation. Reflective of the structural diversity of halogenated natural products (halometabolites), biology has evolved an equally diverse set of enzyme catalysts using numerous strategies for activating halides for addition onto various oxidation states of carbon. The last two decades have witnessed a renaissance in the discovery of mechanistically diverse halogenating
enzymes due to the unprecedented opportunity to study them in the context of halometabolite biosynthetic pathways afforded by advancements in nucleic acid sequencing technologies and computational biology. Despite the fact that the majority of halometabolites are brominated compounds derived from the marine environment, the majority of halogenting enzymes characterized to date are involved in chlorination of terrestrial microbial metabolites. Chapter 2 of this thesis describes the biosynthesis of the highly brominated pyrrole-phenol marine bacterial metabolite pentabromopseudilin by the brominated marine pyrrole/phenol (Bmp) biosynthetic pathway, which revealed the first two examples of brominating enzymes—pyrrole and phenol halogenases—from a confirmed biosynthetic context. Chapter 3 provides an in-depth investigation of regiopromiscuous pyrrole halogenase Bmp2 in contrast to canonical regioselective pyrrole chlorinases. In addition to enzymes involved in the addition of halogen atoms to the aromatic building blocks, the Bmp pathway was found to contain a dehalogenating enzyme (Bmp8) that partially undoes the work of the Bmp2 to allow a free bromopyrrole moiety to participate in the final step of pentabromopseudilin biosynthesis. Chapter 4 describes the activity and mechanism of Bmp8 as the first example of a dehalogenating tailoring enzyme from the confirmed context of a halometabolite biosynthetic pathway. Chapter 5 of this thesis concludes with an exploration of additional opportunities in the study of natural products containing halogenated pyrrole moieties, describing a strategy for the elucidation of the biosynthesis of the bioactive pyrrole-imidazole alkaloid class of sponge secondary metabolites.
Chapter 1: Introduction

1.1 Halogenated natural products: questions and opportunities

Nature synthesizes a remarkable diversity of bioactive halogenated organic compounds ranging from simple aliphatic and aromatic molecules to complex terpenoids, polyketides, oligopeptides, and alkaloids\(^1\). Halogenation is of critical importance to the bioactivities of natural products used in the clinic, under clinical trial, and in their native physiological contexts\(^3\)\(^-\)\(^6\). For example, the “antibiotic of last resort” vancomycin exhibits a drop to 70\% of its activity against a model strain of *Bacillus subtilis* following the removal of one of the two chlorines decorating its glycopeptide core, and a 50\% reduction in activity when both chlorines are eliminated (Fig. 1.1)\(^4\). In a natural physiological context, tetraiodinated thyroid hormone (T\(_4\)), a modulator of mammalian metabolism, is activated upon enzymatic deiodination to T\(_3\), and is rendered inactive by a second round of enzymatic deiodination\(^7\). Halogenation plays a direct role in the mechanism of action of the γ-lactam-β-lactone natural product salinosporamide A in clinical trial for the treatment of cancer, which forms a covalent adduct with the active site threonine residue of the 20S proteasome facilitated by a chlorine leaving group\(^6\)\(^,\)\(^8\). In addition to halogenation, halogen identity also affects the potency of halogenated natural products. For example, replacing the two chlorines of balhimycin, a differently glycosylated variant of vancomycin, to bromines profoundly alters its antimicrobial profile\(^5\). Furthermore, fluorination, while exceedingly rare in biology is a well-proven strategy for increasing the efficacy of pharmaceuticals\(^9\)\(^,\)\(^10\).
Figure 1.1: Halogens matter. Examples of natural products whose biological is demonstrated to be influenced by halogenations.

Despite the wealth of halogenated natural products discovered to date (>4,000 compounds) and their recognized importance as potential therapeutic agents, an understanding of the biosynthesis of these compounds has lagged behind their discovery largely because relatively few biosynthetic gene loci encoding their biosyntheses have been identified and characterized. However, with the rapid growth in genetic information made available by advancements in nucleic acid sequencing, along with the development of increasingly sophisticated tools to interpret this genetic information, understanding of bacterial natural product biosynthesis is on an unprecedented upswing\textsuperscript{11,12}. Paralleling the structural diversity of organohalogen natural products, the study of their biosynthesis has, in short order, revealed an equally diverse array of halogenation biocatalysts\textsuperscript{13-15}. Unlike traditional synthetic approaches for halogenation that utilize harsh conditions, and often add halogens in a non-regioselective manner, halogenation enzyme catalysts operate in aqueous solution under ambient conditions and can be highly regioselective, and even halogen specific\textsuperscript{16,17}. 
From the discovery of the earliest halogenation biocatalyst, a promiscuous fungal chloroperoxidase (CPO) fifty years ago, to the more recent discovery of exquisitely $\text{O}_2$-dependent halogenases over thirty years later, halogenation biocatalysts have been enthusiastically explored for their potential biotechnological applications through extensive study of their substrate scope and the scalability of the reactions they catalyze$^{18,19}$. An additional lesser-explored area of biotechnological interest from both the perspectives of synthetic chemistry and bioremediation of organohalogen pollutants, is the discovery of novel biocatalysts that perform dehalogenation either in the context of biosynthetic tailoring enzymes, or from the degradation pathways of natural organohalogens$^{20-22}$. This chapter will provide a brief overview of the state of the art of understanding of organohalogen biosynthesis with an emphasis on the questions and opportunities arising from the study of enzymatic transformations involving halogens from within biosynthetic contexts, and concluding with motivations for the study of the biosynthesis of polybrominated aromatic molecules by marine bacteria that is the subject of this thesis.

1.2 The inorganic halogen pool

Halogens are biologically available in the environment as halides (Cl$^-$, Br$^-$, F$^-$, and I$^-$). The majority of organohalogen natural products are brominated or chlorinated with fewer examples of iodinated compounds, and even rarer examples of fluorinated compounds$^2$. While chlorination is represented in both marine and terrestrial environments, bromination is almost exclusively a feature of marine natural products due to the relatively high bioavailability of bromide in seawater$^{2,23}$. Remarkably, despite the fact that brominated natural products outnumber their chlorinated
counterparts, chloride occurs in seawater at a concentration nearly three orders of magnitude greater than that of bromide, hinting at the predominance of oxidative strategies in organohalogen natural product biosynthesis as bromide is more readily oxidized than chloride\textsuperscript{15,24}. Moreover, biological fluorination is exceedingly rare despite the great natural abundance fluorine likely due to the extreme electronegativity of fluorine\textsuperscript{24}. On the other hand, the scarcity of organoiodine natural products (~110 reported), that still outnumber their organofluorine counterparts by over tenfold, is explained by the extremely low natural availability of iodide rather than redox potential\textsuperscript{10,24,25}. In line with the widespread distribution of natural organohalogenes, emerging studies have suggested a native role for microbial dehalogenation as part of a biogeochemical halogen cycle in marine and terrestrial environments\textsuperscript{20-22}. However, in contrast to the abundance of reported halogenated natural products, the biochemical transformations leading to their natural formation and degradation remain largely unexplored, providing a vast untapped source of novel biocatalysts.

### 1.3 Biological halogenation strategies: as diverse as the scaffolds they modify

The first halogenating biocatalyst reported nearly fifty years ago was an H\textsubscript{2}O\textsubscript{2}/heme-dependent chloroperoxidase (CPO) secreted from the terrestrial fungus \textit{Caldariomyces fumago}, and implicated in the biosynthesis of halogenated cyclo pentanediol fungal halometabolite caldariomycin\textsuperscript{26-28}. CPO has since been thoroughly studied and even marketed as an electrophilic halogenating reagent with broad substrate specificity\textsuperscript{29}. The initial characterization of CPO introduced a spectrophotometric assay to monitor electrophilic halogenation activity utilizing a
chromophoric analog of caldariomycin, monochlorodimedone (MCD), that exhibits a loss in absorption at 277 nm following electrophilic halogenation\textsuperscript{26}. The MCD assay relies on the ability of a candidate enzyme to oxidize a halide substrate (X:\ Cl\textsuperscript{-}, Br\textsuperscript{-}, and I\textsuperscript{-}) to form a diffusable hypohalite species (i.e., utilizing an oxidative ‘\textit{X}\textsuperscript{+}\textit{'} strategy), and hence inherently selects for promiscuous oxidative halogenation enzyme catalysts (Fig. 1.2)\textsuperscript{30}. Utilizing the MCD assay, numerous haloperoxidases were discovered, including an additional class of vanadium-dependent haloperoxidases found in seaweeds some twenty years after the initial report of the first CPO\textsuperscript{31}.

Figure 1.2: Schematic for MCD assay for detection of electrophilic halogen species. The MCD assay monitors a decrease in the absorbance at 277 nm upon electrophilic addition of a halogen to MCD to form the dihalogenated species.

The paucity of alternative biological halogenation enzymes and strategies discovered over the decades that followed the discovery of haloperoxidases, led to the assumption that haloperoxidases were the predominant means of halogenation underlying the biosynthesis of halometabolites—in other words, absence of evidence became evidence for absence\textsuperscript{32}. Indeed, at the time of the discovery of the first CPO, this argument was helped by the fact that biological halogenation was largely considered to be an accident of biology, as only thirty halogenated natural products had been described\textsuperscript{33}. However, with the subsequent discovery of thousands of bioactive natural products bearing halogens attached to both electron rich and
aliphatic scaffolds in apparently regiospecific configurations, it has become clear that halogenation is not a mere accident of Nature (Fig. 1.3)\textsuperscript{1,2}. Moreover, the discovery of fluorinated natural products, despite the inability of any known haloperoxidase to accept fluoride as a substrate further challenged the notion that these enzymes were the sole strategy for biological halogenation\textsuperscript{10}. Hence, the broad structural diversity of organohalogen natural products signaled a wealth of substrate-specific biocatalysts utilizing alternative halogenation strategies beyond promiscuous haloperoxidases\textsuperscript{30}. One of the main obstructions to the discovery new halogenation biocatalysts has been the inability to link enzymes with their natural substrates. However, disruptive advancements in genome sequencing and computational biology over the last two decades have lead to a renaissance in the discovery of natural product biosynthetic pathways and tailoring enzymes\textsuperscript{34}. 
Figure 1.3: Representative diversity of halogenated natural products. Halogens decorate a diversity of natural product scaffolds that foretell diverse biological halogenation strategies.

Nearly forty years after the discovery of the first haloperoxidase, a new class of regioselective $\text{O}_2$/flavin-dependent electrophilic halogenases was described$^{35}$. This enzyme, PrnA from the biosynthesis of the anti-fungal pyrrolonitrin, was shown to catalyze a regioselective chlorination at the 7-position of L-tryptophan. PrnA is related to two-component flavin-dependent monooxygenases, so-named because they require an external NAD(P)H-dependent flavin reductase redox partner (Fig. 1.4A,B)$^{35-37}$. Based on structural studies of tryptophan 7-halogenases PrnA and RebH from the biosynthesis of the indolocarbazole natural product rebeccamycin, a mechanism for flavin-dependent halogenases akin to their monooxygenase counterparts has been proposed. The reaction sequence is initiated by the attack of molecular oxygen on the $\text{C}_{4\text{a}}$ position of isoalloxazine moiety of the flavin co-factor.
to generate a flavin peroxide (FAD-C_{4a}-O-OH) which is then attacked by a halide (X^{-}) to release water and form a flavin hypohalite (FAD-C_{4a}-O-X) that is subsequently resolved to a diffusable hypohalous acid, the same reactive intermediate proposed for haloperoxidases\textsuperscript{36}. Subsequent to its release, the hypohalous acid is proposed to react directly with a conserved lysine residue amine side chain to form a chloramine intermediate that confers control to the halogenation reaction taking place in the substrate-binding pocket\textsuperscript{36,38}. Despite strong evidence for formation of an apparent “long-lived” chloramine intermediate, the identity of this adduct is disputed on the grounds that such an intermediate might be too unreactive to halogenate tryptophan\textsuperscript{39}. In addition to tryptophan 7-halogenases, tryptophan 5- and 6- halogenases from the biosyntheses of the bacterial natural products pyrrindomycin and thienodolin, respectively, have also been described (Fig. 1.4B)\textsuperscript{40,41}. More recently a pair of flavin-dependent halogenases, KtzQ and KtzR, were shown to perform serial halogenations, respectively, of the 6- and 7- positions of L-tryptophan in the biosynthesis of the cyclic depsipeptide kutznerides (Fig. 1.4B)\textsuperscript{42}. The discovery of a series of tryptophan halogenases catalyzing a range of regioselective halogen additions has provided an opportunity for studying the molecular basis of halogenation regioselectivity. Indeed, comparative structure-guided mutagenesis of tryptophan halogenases exhibiting different regioselectivities has lead to the identification a molecular basis for the exquisite regioselectivity exhibited by these halogenating biocatalysts\textsuperscript{43}. 
Figure 1.4: Flavin-dependent halogenases. (A) The catalytic cycle for typical 2-component flavin-dependent halogenases (Nu = nucleophile, FR = flavin-reductase, NR = NAD(P)+-reductase). (B)-(D) Transformations catalyzed by (B) Tryptophan halogenases, (C) pyrrolyl-S-carrier protein (CP) halogenases, and (D) tyrosyl-S-CP halogenases. (E) Left hand side shows ribbon structure of CndH in blue with “insertions” from PrnA/RebH shown in red and green, and the right hand side shows primary amino acid sequence alignments of tyrosyl-S-CP halogenases (top two lines) with RebH and PrnA (bottom two lines) from ref. 49.

Following the discovery of flavin-dependent halogenases that act on a free amino (i.e. L-tryptophan), regioselective flavin-dependent halogenases acting on L-proline-derived pyrrolyl- and L-tyrosine-derived tyrosyl-S-carrier-protein substrates were also described, adding to the versatility of flavin-dependent halogenases (Fig. 1.4C,D). Based on comparison of the crystal structure of tryptophan halogenase PrnA to that of a putative tyrosyl-S-carrier protein halogenase (CndH from the biosynthesis of chondrochlorens), a preliminary means of distinguishing between
flavin-dependent halogenases acting on free versus thiotemplated substrates was proposed based on primary sequence motifs (Fig. 1.4E)\(^{49}\). Prior to the work described in this thesis, several examples of flavin-dependent pyrrole 4,5-dichlorinases\(^{44-46}\) and one example of a pyrrole 5-chlorinase\(^{47}\) acting on carrier protein-bound substrates were described. Chapters 2 and 3 of this thesis characterize a pyrrole brominase that exhibits total loss of halogenation regiocontrol to afford fully brominated 2,3,4,5-tetrabromopyrrole via decarboxylative tetrabromination of a carrier protein-bound pyrrole. The availability of pyrrole halogenases displaying a series of regioselectivies from mono- through tetra- halogenation affords a unique opportunity to investigate a molecular basis for regiocontrol in halogenases acting on thiotemplated substrates, which is exploited in Chapter 3 of this thesis.

Incidentally, the scope of reactions catalyzed by flavin-dependent halogenase may extend beyond halogenation for halogenation’s sake. A pair of putative flavin-dependent halogenases, Mpy10 and Mpy11, was implicated in the atropo-selective N-C coupling reaction of two monomers of monodeoxypyoluteorin in the final step of the biosynthesis of the marine natural product marinopyrrole A\(^{45}\). While this mechanism has not yet been demonstrated, Yamanaka et al. propose that biaryl coupling is activated by cryptic C or N halogenation resulting in the loss of a halide leaving group (Fig. 1.5). Notably the marinopyrrole biosynthetic pathway also contains a flavin-dependent pyrrole dichlorinase, which regiospecifically installs two chlorine atoms on an ACP-bound pyrrole substrate. Hence, marinopyrrole A biosynthesis may provide an example of the use of halogenation by flavin-dependent enzyme catalysts as both a means to an ends and means in and of itself.
Figure 1.5: Mechanistic proposal for halogenation-mediated atropo-selective N-C coupling in the biosynthesis of Marinopyrrole A. The coupling of monodeoxypyoluteorin is proposed to occur via either C- or N- halogenation involving putative flavin-dependent halogenases Mpy10 and Mpy11. Scheme adapted from ref. 45 (Yamanaka et al., J. Am. Chem. Soc. 2012).

The discovery of flavin-dependent halogenases led to reactionary claims that this class of enzyme rather than the haloperoxidases was the truly dominant regioselective electrophilic halogenation catalyst\(^{30,37}\). While numerous examples of natural products gene clusters encoding flavin-dependent halogenases have been reported\(^{19,44-47,50,51}\), these absolutist claims have been quickly tempered by the subsequent discovery of bacterial H\(_2\)O\(_2\)/vanadium-dependent haloperoxidases catalyzing highly selective chlorination reactions in confirmed biosynthetic contexts\(^{52-55}\). These enzymes are the first examples of vanadium-dependent peroxidases of bacterial origin and appear to have evolved from bacterial acid phosphatases\(^{52}\). The first bacterial vanadium-dependent haloperoxidase characterized was the chloroperoxidase NapHI from the biosynthesis of the polyketide-terpenoid (meroterpenoid) antibiotic napyradiomycin that catalyzes a site-specific chlorination-cyclization reaction\(^{54}\). Another recently described vanadium-dependent
chloroperoxidase Mcl24 catalyzes a regioselective chlorination/oxidative
dearomatization/cyclization reaction sequence in the biosynthesis of the
meroterpenoids merochlorins A and B\textsuperscript{55}. Notably, the use of an electrophilic
halogenation strategy would not have been immediately apparent from inspection of
the structures of these meroterpenoid natural products, which bear halogens on
aliphatic carbons. Hence, the discovery of regiospecific haloperoxidases delineates
the importance of studying halogenating enzymes in their biosynthetic contexts.

The existence of organohalogen natural products bearing halogens on
unactivated carbons foretold a radical halogenation strategy\textsuperscript{56,57}. Several years after
the discovery of flavin-dependent halogenases, another $\text{O}_2$-dependent class of
halogenase related to $\text{Fe}^{II}/\alpha$-ketoglutarate($\alpha$-KG)-dependent dioxygenases was
reported from the biosyntheses of the non-ribosomal peptides syringomycin A and the
barbamides, and the non-proteinogenic amino acid armentomycin\textsuperscript{58-60}. These non-
heme $\text{O}_2$/Fe\textsuperscript{II}/$\alpha$-KG-dependent halogenases generate a halogen radical from chloride
(i.e. an ‘$X^\prime$’ strategy) to halogenate methyl groups of carrier protein-tethered amino
acids\textsuperscript{13}. Interestingly the kutzernide biosynthetic pathway contains both flavin-
dependent halogenases and a non-heme $\text{O}_2$/Fe\textsuperscript{II}/$\alpha$-KG-dependent halogenases. As
previously mentioned, the former are involved in the tandem dichlorination of
tryptophan, while the latter was shown to catalyze a “cryptic” C-H activating
chlorination reaction leading to cyclopropane ring formation in the biosynthesis of the
non-proteinogenic amino acid building block derived from L-isoleucine\textsuperscript{61}. Hence the
kutzernide biosynthetic pathway demonstrates Nature’s capacity to combine multiple
halogenation strategies, using halogens to manipulate both reactivity and bioactivity.
Likely due to the electronegativity of fluorine, no known oxidative or radical strategy for its incorporation into natural product scaffolds has been discovered. Hence, the halogenation strategy implied by the existence of fluorinated natural products is a nucleophilic halogenation strategy that directly utilizes fluoride. Indeed, shortly after the discovery of the first flavin-dependent halogenases, a S-adenosyl methionine (SAM)-dependent nucleophilic fluorinase was reported from the soil bacterium *Streptomyces catteleya* known to produce the fluorometabolites fluoroacetate and 4'-fluorothreonine. An additional SAM-dependent nucleophilic halogenase was later described from the biosynthesis of the chlorinated marine natural product salinosporamide A. SAM-dependent halogenases directly utilize halides (i.e. $X^-$: F$^-$ or Cl$^-$) which perform a substitution reaction at the C-5’ position of SAM, attacking the nucleophilic sulfonium center, and releasing a neutral L-methionine leaving group. The resultant halogenated nucleoside is then metabolized to precursors that are incorporated into the final natural product. As such, nucleophilic halogenation by SAM-dependent halogenases is a remarkable feat of Nature, but is also a highly specialized halogenation strategy. Nonetheless, fluorine, which is commonly used to enhance efficacy of clinical drugs, is by far the most well-represented halogen in pharmaceuticals, and hence novel fluorinated natural products and associated fluorinating strategies should continue to be sought.

By the start of the decade four additional classes of halogenating enzymes utilizing halides via three distinct strategies—oxidative, radical, and nucleophilic—had been discovered. Remarkably, in a paradigm shift from the biased MCD assay used to screen for haloperoxidases, the true substrates for all of these new
halogenation biocatalysts are known thanks to their systematic study in their biosynthetic contexts. Moreover, the improved capacity for prediction of biosynthetic gene loci from natural product scaffolds has enabled the appreciation of halogenating biocatalysts in their physiological contexts, and afforded the opportunity for engineering of these biocatalysts for altered regioselectivity and substrate scope. The fact that halogenating enzymes like all enzymes of secondary metabolism evolved from primary metabolism hints that many more unique halogenation biocatalysts remain to be discovered. Indeed, halogenated natural products are already being linked to genetic contexts which lack annotations for known halogenating enzymes. For example, a putative N-oxidase-like enzyme-BrtJ has been proposed to be the halogenase involved in the biosynthesis of the recently described cyanobacterial aromatic glycolipid natural product bartolosides. Further, the extensive natural product cataloging efforts of prior decades have endowed enzymologists with an inventory of natural products as a springboard for generating biosynthetic hypotheses, while sequencing, computational biology, molecular biology, metabolic engineering, analytical chemistry, and synthetic chemistry have teamed up to provide a direct basis for interrogation of these hypotheses.

1.4 Comment on the “haloperoxidase doctrine”

In light of the renaissance in halogenation enzymology, it is curious to note the persistence of the notion in the scientific literature that haloperoxidases are responsible for the bulk of marine halogenation. Apart from a few recent examples in which bacterial haloperoxidases found in the contexts of biosynthetic gene clusters have been shown to be directly involved in specific reactions, the main
drivers for the “haloperoxidase doctrine” are studies that rely on transformations of putative precursors for natural products for which no genetic information is known. An additional *ad ignorantium* argument that has persisted in the literature is the notion that haloperoxidases are responsible for the majority of brominated natural products due to the prevalence of vanadium-dependent bromoperoxidases found in red algae that are also an abundant source of bromophenol metabolites. Indeed, the first two examples of marine brominases from a physiological context described in chapters 2-3 of this thesis are substrate-specific flavin-dependent halogenases and not haloperoxidases. Moreover, another notion challenged by this thesis is the dichotomy that flavin-dependent halogenases are regioselective, while haloperoxidases are regiopromiscuous. This thesis describes an “aberrant” regiopromiscuous flavin-dependent halogenation that exhibits complete loss of halogenation regiocontrol with respect to an ACP-bound pyrrole substrate. This loss in regiocontrol in turn serves as a means for elimination of the proline-derived α-carboxylate, which is otherwise retained in lesser halogenated halopyrrole-containing natural products. Indeed, it seems the only generalization that can be drawn about halogenating enzymes, or for that matter any biosynthetic tailoring enzyme, is that they are uniquely evolved to influence functional outcomes specified in the contexts of biosynthetic pathways.

### 1.5 Biological dehalogenation: a new frontier in natural products biosynthesis

While the major part of this chapter has discussed Nature’s strategies for adding halogens to hydrocarbon scaffolds, another frontier for biocatalysis is the
discovery of enzymes that remove halogen atoms from organic substrates. Enzymatic dehalogenation is of particular interest in the remediation of highly brominated persistent pollutants of which many have analogs in natural products\textsuperscript{70}. While studies have been performed to probe the capacities of microbial communities for dehalogenation of man-made compounds, thus far none has looked toward the biosyntheses and degradation of highly halogenated microbial natural products as an inspiration for the discovery for new dehalogenating biocatalysts\textsuperscript{71}. Apart from the few examples of dehalogenations taking place in the course of transformations involving cryptic halogenation reactions, in which a halogen installed by a halogenating enzyme serves as a leaving group, no example of a dedicated dehalogenase tailoring enzyme has been described from the context of a natural product biosynthetic pathway\textsuperscript{45,61,72,73}. Moreover, despite a recognition of Nature’s capacity to degrade halogenated natural products, no enzyme involved in the natural degradation of organohalogen natural products has been identified from an environmental context\textsuperscript{20-22}. An illustrative example of the wealth of enzymology arising from the formation and metabolism of a halogenated natural product comes from the study of thyroid hormone biosynthesis, which has afforded three new classes of enzymes—a halogenating enzyme and two classes of deiodinases (flavoprotein and selenoprotein) involved in activation-inactivation of thyroid hormone and reclamation of scarce iodide\textsuperscript{7,74}. Chapter 4 of this thesis describes the discovery and elucidation of a novel class of dehalogenase, Bmp8, providing the first example of a dehalogenase tailoring enzyme from the context of a natural product biosynthetic pathway.

1.6 In this thesis: **motivations for the study of the biosynthesis of pentabromopseudilin**
Figure 1.6: Pentabromopseudilin. The compact highly brominated structure of the marine microbial natural product pentabromopseudilin is packed with biosynthetic opportunity.

Pentabromopseudilin is the first reported marine microbial natural product, isolated in 1966 by Burkholder and colleagues from the marine bacterium *Pseudoalteromonas bromoutilis*\(^{75}\). Apart from its place in history, the compact structure of pentabromopseudilin which consists of a tribromopyrrole moiety coupled to a dibromophenol moiety via a C-C bond is packed with opportunity from a biosynthetic standpoint (Fig. 1.6). From the perspective of novel enzymology, no brominating enzyme from the confirmed context of a natural product biosynthetic pathway had been reported when this study was begun five years ago—the structure of pentabromopseudilin suggests at least two brominating enzymes. In addition to motivations deriving from biosynthetic novelty, our attention was drawn to the biosynthesis of pentabromopseudilin by analogy of both of its brominated aromatic moieties to polybrominated aromatic bipyrrrole and biphenyl compounds that accumulate up the marine food chain\(^{70}\). While bioaccumulative polyhalogenated bipyrrroles are strictly of natural origin, the hydroxylated polybrominated diphenyl ethers (OH-BDEs) are thought to be of mixed natural and anthropogenic origin\(^2\). However, claims of the natural origin of OH-BDEs at the time of this study were based exclusively on circumstantial evidence\(^{76,77}\). Hence, an unfulfilled need remained for a genetic or biosynthetic logic with which to mine environmental
metagenomes as means of understanding the distribution of sources of natural OH-BDEs. Critically, the ability to identify a molecular origin for the biosynthesis of OH-BDEs has the potential to dramatically influence the discussion regarding our small molecule interaction with the environment, as most biomonitoring programs ignore OH-BDEs altogether, assuming them to be metabolic products of anthropogenic polybrominated diphenyl ethers (PBDEs)\textsuperscript{78,79}. The remainder of this thesis will describe the enzymatic and chemical diversity afforded by the total \textit{in vitro} reconstitution of the biosynthesis of pentabromopseudilin, as well as questions and opportunities arising from the study thereof. This introductory chapter concludes with a synopsis of the remaining chapters.

\subsection*{1.6.1 Chapter 2: Biosynthesis of polybrominated aromatic molecules by marine bacteria}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure17.pdf}
\caption{A modular polybrominated pyrrole/phenol (Bmp) biosynthetic pathway from marine bacteria. Pentabromopseudilin biosynthesis is encoded by the modular Bmp biosynthetic gene cluster that additionally affords combinatorically C-C and C-O coupled homo- and hetero- coupled bromo- pyrroles and phenols.}
\end{figure}

This published chapter (Agarwal and El Gamal \textit{et al.}, \textit{Nat. Chem. Biol.}, 2014) consists of the initial description of the brominated marine pyrrole/phenol (Bmp) biosynthetic pathway found across several species and genera of host-associated marine \(\gamma\)-proteobacteria (Fig. 1.7). In this study several genomes of marine \(\gamma\)-
proteobacteria confirmed to produce pentabromopseudilin were sequenced, and interrogated using genes from the biosynthesis of the halopyrrole moiety of pyoluteorin\textsuperscript{44}. This query lead to the identification of a modular biosynthetic gene locus, which was heterologously expressed and functionally interrogated \textit{in vivo} and \textit{in vitro}. This study confirmed the hypothesis that the bromopyrrole moiety of pentabromopseudilin arises from a conserved halopyrrole biosynthetic logic. Additionally, this chapter describes the first two examples of marine brominating enzymes from the confirmed context of a natural product biosynthetic pathway. The pyrrole halogenase Bmp2 is homologous to previously described flavin-dependent pyrrole halogenases and is demonstrated to catalyze an unprecedented three brominations on an ACP-bound pyrrole substrate. The phenol halogenase Bmp5 employs a decarboxylative halogenation mechanism to convert $p$-hydroxybenzoic acid to 2,4-dibromophenol. The architecture of Bmp5 is unique among flavin-dependent halogenases characterized to date in that its protein sequence resembles that of single-component monooxygenases, meaning that it does not require an external NAD(P)H-dependent flavin reductase and contains sequence motifs for binding both flavin and a NAD(P)H. It was also revealed that these marine brominating enzymes are selective for bromide and exclude chloride, providing a basis for comparing flavin-dependent chlorinases (accept Cl, Br) to brominases (accept on Br) toward understanding the fundamental question of halogen specificity of halogenating enzymes. Also characterized in the chapter is the versatile cytochrome P450 Bmp7 that forms the basis for combinatorial C-C and C-O coupling of the bromopyrrole and bromophenol building blocks produced by the Bmp pathway.
to afford an astonishing chemical diversity consisting of fifteen metabolites, several of which are new chemical entities. Notably, OH-BDEs were found to be among the metabolites synthesized by the Bmp pathway. This initial report provided the near complete in vitro reconstitution of the pentabromopseudilin biosynthetic pathway, but left unanswered the enzymes and reaction sequence leading to the offloading and decarboxylation of the bromopyrrole moiety. The elucidation of these transformation and associated enzymes form the basis for chapter 3 and 4 of this thesis.

1.6.2 Chapter 3: Biosynthesis of the coral larval settlement cue tetrabromopyrrole in marine bacteria by a uniquely adapted brominase-thioesterase enzyme pair

![Figure 1.8: A modular polybrominated pyrrole/phenol (Bmp) biosynthetic pathway from marine bacteria. Pentabromopseudilin biosynthesis is encoded by the modular Bmp biosynthetic gene cluster that additionally affords combinatorically C-C and C-O coupled homo- and hetero- coupled bromo-pyrroles and phenols.]

This published chapter (El Gamal and Agarwal et al., Proc. Natl. Acad. Sci. USA, 2016) focuses on the unusual mechanism of offloading and decarboxylation of the bromopyrrole moiety of pentabromopseudilin. This study was originally motivated by another related bromopyrrole natural product, tetrabromopyrrole, which is a coral larval settlement cue produced by marine bacteria that form biofilms on coral substrates. In the initial elucidation of the the biosynthesis of
pentabromopseudilin, asymmetric tribromopyrrole—not tetrabromopyrrole—was identified as the substrate for the bromo-phenol/pyrrole coupling enzyme Bmp7. However, this finding stood at odds with a symmetric pyrrole intermediate implied by shuffling with respect to C-C coupling in earlier isotope feeding studies of pentabromopseudilin biosynthesis. Hence, studying the biosynthesis of tetrabromopyrrole had the potential to address two biosynthetic questions. Firstly, the presence of a bromine in place of the anticipated proline-derived α-carboxylate in tetrabromopyrrole implied a brominative decarboxylation mechanism. Secondly, the elucidation of the tetrabromopyrrole biosynthetic pathway offered to address the question of the identity of the symmetric intermediate implied in the biosynthesis of pentabromopseudilin. Sequencing of two tetrabromopyrrole-producing bacteria and querying with Bmp biosynthetic genes revealed that only the bromopyrrole module of the Bmp pathway was present in their genomes. Further, this study revealed that the pyrrole front-end of the Bmp pathway produced tetrabromopyrrole as its sole enzymatic product. Moreover, it was demonstrated that the pyrrole halogenase Bmp2 catalyzes a fourth halogenation of the carrier protein-bound pyrrole scaffold that triggers the thioesterase-mediated release of the pyrrole, and leads to spontaneous decarboxylation to tetrabromopyrrole (Fig. 1.8). Hence, the decarboxylative halogenation mechanism of Bmp2 mirrors that of the p-hydroxybenzoic acid brominase/decarboxylase Bmp5. Notably, Bmp2 provides the first example of a flavin-dependent halogenase exhibiting a complete loss of regiocontrol. While previously characterized pyrrole halogenases catalyze up to two regiospecific halogenation on carrier protein-bound pyrrole scaffolds, Bmp2 catalyzed an
unprecedented four halogenations. To interrogate a molecular basis for halogenation regiospecificity, high-resolution crystal structures were generated for the pyrrole dichlorinase Mpy16 from the biosynthesis of marinopyrrole A$^{45}$ and tetrabrominase Bmp2. Structure-guided mutagenesis of the active site of Bmp2 to resemble that of Mpy16 lead to a reduction in the degree of halogenation catalyzed from four down to one. The discovery of tetrabromopyrrole as the enzymatic product of the bromopyrrole biosynthetic “module” of the Bmp pathway, leads to the natural follow-on question as to the mechanism for the conversion of symmetric tetrabromopyrrole to the asymmetric Bmp7-substrate tribromopyrrole in the biosynthesis of pentabromopseudilin. The elucidation of an unusual dehalogenase tailoring enzyme-Bmp8 catalyzing this transformation forms the basis for Chapter 4 of this thesis.

1.6.3 Chapter 4: An alkylhydroperoxidase-like debrominase is an enzymatic switch in the biosynthesis of pentabromopseudilin

![Diagram](image)

**Figure 1.9: A dehalogenase switch in pentabromopseudilin biosynthesis.** Dehalogenase-Bmp8 utilizes a cofactor-independent redox thiolate mechanism to transforms 2,3,4,5-tetrabromopyrrole (Br$_4$Py) to 3,4,5-tribromopyrrole (Br$_3$Py) to permit coupling with 2,4-bromophenol in the formation of pentabromopseudilin.

This chapter describes a new class of co-factor independent reductive-dehalogenases. We show that Bmp8 catalyzes the transformation of tetrabromopyrrole to the Bmp7-substrate tribromopyrrole, reconciling the
pentabromopseudilin biosynthetic scheme with earlier isotope feeding studies and the biosynthesis of tetrabromopyrrole. This chapter describes the mass-spectrometry-driven elucidation of Bmp8, which revealed that Bmp8 utilizes a redox active cysteine residue side chain thiol in its reaction mechanism. Mining publically available databases with the primary sequence of Bmp8 revealed hundreds of homologs from different biosynthetic contexts bearing a simple conserved motif that could be used to predict for the ability to dehalogenate bromopyrroles. Homology-guided mutagenesis was further performed to identify residues in the motif that are critical to the activity of Bmp8. To the best of our knowledge, Bmp8 is the first reported example of a dehalogenating tailoring enzyme from the context of a biosynthetic pathway. Additionally, Bmp8 is only the second example of a reductive dehalogenase that utilizes a cofactor-independent mechanism (the other is a mammalian selenoprotein deiodinase that uses a similar mechanism that likely exemplifies a case of convergent evolution), as well as the only other physiologically confirmed dehalogenating enzyme described in biology (Fig. 1.9)\textsuperscript{7}. Hence, this chapter concludes the total \textit{in vitro} characterization of the biosynthesis of pentabromopseudilin.

1.6.4 Chapter 5: Opportunities in halopyrrole biosynthesis
Figure 1.10: A metagenomics approach toward elucidation of pyrrole-imidazole alkaloid biosynthesis. A conserved halopyrrole biosynthetic logic may serve as a handle to mine for pyrrole-imidazole alkaloid biosynthetic gene loci, unlocking the biosynthesis of a diverse class of bioactive natural products.

One of the key challenges in the identification of gene clusters encoding the biosyntheses of non-canonical natural products (i.e. neither NRPS nor PKS) is the lack of an appropriate query with which to search genomic datasets. One strategy for going after chemical moieties that lack biosynthetic precedents is to identify natural products that wed the familiar to the exotic. For example, in the case of the hybrid pyrrole-phenol structure of pentabromopeudilin (Chapter 2), the familiar halopyrrole moiety served as a biosynthetic hook with which to extract the biosynthesis of the bromophenol moiety. Chapter 5 explores the application of the same halopyrrole biosynthetic probe toward the elucidation of the biosynthesis of the pyrrole-imidazole alkaloid (PIA) class of sponge natural products. This intriguing class of bioactive natural products consists of over 150 chemical structures ranging in complexity from the simplest achiral building block oroidin to the awe-inspiring oroidin dimer palau’amine whose hexacyclic core consists of eight contiguous stereocenters. This chapter will describe an approach toward the identification of a genetic basis for PIA biosynthesis from sponge metagenomes, along with insights gained from an initial application of this approach to two metagenomes of sponges harboring PIAs (Fig. 1.10).
Chapter 2: Biosynthesis of polybrominated aromatic molecules by marine bacteria
Biosynthesis of polybrominated aromatic organic compounds by marine bacteria

Vinayak Agarwal1,2, Abraham A El Gamal1,4, Kazuya Yamanaka2,4, Dennis Poth1, Roland D Kersten1,5, Michelle Schorn1, Eric E Allen1,2 & Bradley S Moore1,2,4

Polybrominated diphenyl ethers (PBDEs) and polybrominated bipyrroles are natural products that bioaccumulate in the marine food chain. PBDEs have attracted widespread attention because of their persistence in the environment and potential toxicity to humans. However, the natural origins of PBDE biosynthesis are not known. Here we report marine bacteria as producers of PBDEs and establish a genetic and molecular foundation for their production that unifies paradigms for the elaboration of bromophenols and bromopyrroles abundant in marine biota. We provide biochemical evidence of marine brominases revealing decarboxylative-halogenation enzymology previously unknown among halogenating enzymes. Biosynthetic motifs discovered in our study were used to mine sequence databases to discover unrealized marine bacterial producers of organobromine compounds.

Although bromide is present in only trace amounts, its greater natural abundance in seawater provides a substrate reservoir for halogenating enzymes evolved to selectively activate and transfer bromide to organic molecules. Marine organisms, including bacteria, algae and invertebrates, are prolific sources of organobrominated compounds that number over 2,000 reported molecules with a wide range of biological properties1,2. These brominated natural products range from simple volatile bromoalkanes to highly complex alkaloids3,4,5. Yet, our understanding of the biosynthesis of brominated natural products has lagged far behind their discovery, largely because few biosynthetic gene loci have been identified and characterized. Herein we report a pervasive marine bacterial pathway to polybrominated aromatic compounds that accounts for a diverse suite of common polybrominated pyrrole and phenol-based natural products.

In their hydroxylated (OH-BDE) and methoxylated (MeO-BDE) forms, PBDEs are abundant across all trophic levels of marine life ranging from marine plants6, algae7, and invertebrates8,9 to marine mammals at the apex of the food chain10,11. OH-BDEs and MeO-BDEs detected in marine biota were once thought to be derived from chemical transformation of anthropogenically produced polybrominated flame retardant chemicals of similar structure12. However, derivatives such as 3,3’,5,5’-tetrabromo-2,2’-biphenylidine (1) (ref. 18), hexabromo-2,2’-bipyrole (2) (ref. 19) and the hybrid bromophenol-bromopyrrole pentabromoposeudinil (3) (refs. 19,20) (Fig. 1a), by Pseudolateromonas spp., marine γ-proteobacteria often associated with eukaryotic hosts12. Although 1 and 2 are not industrially synthesized, their methylated analogs are extensively detected in marine mammals12,21 (Fig. 1a). Notably, in contrast to the microbial biosynthesis of polybrominated biphenyls and bipyrroles, production of PBDEs has not been confirmed from marine bacterial sources.

Herein we report the discovery of a conserved biosynthetic gene cluster in marine bacteria responsible for the synthesis of widespread polybrominated aromatic compounds and describe two flavin-dependent brominases involved in the synthesis of universal polybromophenol- and polybromopyrrole-based metabolites. For what is to our knowledge the first time, our findings establish marine bacteria as sources of OH-BDEs, and we provide the first report and biochemical characterization of a flavin-dependent decarboxylative-brominase enzyme using an inferred enzyme architecture not previously realized among halogenases.

RESULTS

A genetic basis for bromination in marine bacteria

In our search for marine prokaryotic sources of polybrominated molecules, we focused on the cosmopolitan marine genus Pseudolateromonas spp., previously shown to produce 1–3. Specifically, we queried the small-molecule natural products synthesized by coral-associated Pseudolateromonas laterale isolated in the Florida Keys13 and by planktonic Pseudolateromonas phenolica O-B6 isolated off the coast of Japan14. We analyzed organic extracts of liquid cultures of these marine bacteria grown in the presence of bromide by high performance LC/MS/MS. Among numerous polybrominated molecules present in the culture extract of both bacteria, we confirmed the production of 1–3 by MS and NMR spectroscopy of isolated compounds (Fig. 1b and Supplementary Results, Supplementary Figs. 1 and 2).

1Center for Oceans and Human Health, Scripps Institution of Oceanography, University of California-San Diego, La Jolla, California, USA. 2Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California-San Diego, La Jolla, California, USA. 3Division of Biological Sciences, University of California-San Diego, La Jolla, California, USA. 4Scripps Institution of Oceanography, University of California-San Diego, La Jolla, California, USA. 5Present address: JNC Corporation, Yokohama Research Center, Yokohama, Japan (K.Y.); Skaggs Institute for Biological Studies, San Diego, California, USA (R.D.K.). *These authors contributed equally to this work. *e-mail: bs.moore@ucsd.edu
Supplementary Table 1). Additionally, we confirmed the production of a large series of bromophenol and bromopropyrole monomeric molecules, including 2,4-dibromophenol (4), 2,4,6-tribromophenol (5), 2,3,4,5-tetramethylpyrrole (6), and 2,3,4,5-tetramethylpyrrole (7) on the basis of their MS spectra and comparison to synthetic standards (Supplementary Fig. 2 and Supplementary Table 1).

Having established experimentally tractable marine bacterial sources for the production of 1-7, we next sought to determine the genetic basis for the production of these polybrominated small-molecule marine natural products. Toward this aim, we sequenced the genomes of P. luteoviolacea 21a16 and P. phenaeca O-BC30, resulting in draft 6.36-Mb and 4.76-Mb genomes, respectively. We hypothesized that the bromopropyroles and, by extension, the bromopyrrole moieties of heterodimeric molecules such as 3 are derived in a manner analogous to the biosynthesis of the dichlorinated pyrrole moiety of the antifungal natural product pyriteoterin,21 consistent with prior isotope feeding studies tracing the source of the pyrrole moiety of 3 to 1-proline (Supplementary Fig. 3). Hence, we queried the genomes of P. luteoviolacea 21a16 and P. phenaeca O-BC30 for the proline dehydrogenase (PilE) and the pyrrole halogenase (PLA) enzymes from pyriteoterin biosynthesis. This analysis successfully led to the identification of the brominated marine pyrrole/phenols (brompyroles) biosynthetic gene loci in both bacteria. Homologs of pilE and pilA (brompyroles and brompyroles, respectively; Fig. 1c) were found clustered together with genes putatively encoding an acyl carrier protein (ACP)-thioesterase (TE) di domain protein (brompyrole), proline dehydrogenase (brompyrole), flavin-dependent oxygenase (brompyrole), chorismate lyase (brompyrole), cytochrome P450 (CYP450, brompyrole), ferredoxin reductase (brompyrole), and ferredoxin reductase NADPH (brompyrole). The primary sequence similarity of brompyroles to pyriteoterin biosynthetic enzymes is provided in Supplementary Table 2.

To confirm the contribution of the brompyrole gene cluster to the biosynthesis of 1-7, we constructed an expression vector and heterologously expressed P. luteoviolacea 21a16 genes bnap-1-8 in Escherichia coli (Supplementary Fig. 4 and Supplementary Table 3). In the presence of bromide, we observed the heterologous production of 1-7 (Fig. 2a,b). Notably, the complexity of the culture extract precluded a concomitant visual representation of all of the polybrominated molecules within a single total ion chromatogram (TIC). We next explored the individual functions of the bnap-1-8 genes by expressing the bnap cluster with individual gene deletions (Supplementary Table 5). As postulated by homology to the pyrrole pyrrole dichlorinated enzyme PRA and consistent with its assignment as a pyrrole brominase, deletion of bnap2 abolished the production of bromopropyroles 2, 3, 6 and 7 (Fig. 2a). In contrast, deletion of bnap5 eliminated 1, 3 and all of the other bromopropyrole-containing species, leading to the proposed role of bnap5 as a phenol brominase. We further identified bnap7 as a coupling enzyme as its deletion led to the abolishment of the dimers 1-3, whereas monomers 4-6 could still be detected. 7 was not a notable exception (Supplementary Fig. 5). These experiments established what is to our knowledge the first genetic link to the production of polybrominated aromatic molecules in nature.

We next analyzed genomic data sets for the presence of orphan gene loci related to the bnap gene cluster. Indeed, we identified a highly conserved gene cluster in the melanogenic marine bacterium Marinomonas mediterranea MM6-1 (ref. 26), with an insertion of a putative brominase between bnap7 and bnap5 (Fig. 1c). As expected, we confirmed the production of brominated molecules 1-3 by M. mediterranea MM6-1 (Supplementary Fig. 1), providing halogenated natural products from this genus that commonly associates with sea grass. Indeed, the methylated variant of 7 (2,3,4,5-tetramethylylpyrrole) and MeO-BDEs such as 2'-MeO-BDE-68 (described below) have been detected in the sea grass Halophila ovalis and obligate marine herbivores such as dungenos.7 We also identified a truncated gene cluster in the Lissoclinum patella tunicate metagenomic data set for the marine cyanobacterial symbiont Prochloron didemni P-2 fijii7 (Fig. 1c).

Additionally, we found a highly similar gene locus in the sulfur-oxidizing 3-proteobacterium Thioalkalivibrio thiocyanoxidans ARH 4 (ref. 28) that, notably, was missing CYP450 bnap7 and CYP450 electron transport partners bnap9 and bnap10 (Fig. 1c).
**Figure 2 | Genetic basis of the Bmp pathway.** (a) TICs for organic extracts of *E. coli* expressing the *P. labilis* BmpA/BmpB gene cluster and selected gene deletions. 2 is present in high abundance in the absence of bmpD as *E. coli* expressing both bmpA and bmpB preferentially produce I and other coupled bromochromenes. Retention times of 5 and 7 are nearly identical; hence, for the sake of clarity, the more abundant species is highlighted in chromatograms where both are present (i.e., bmpD− and bmpD−). In addition to 5−7, the bmpD− extract contains other brominated phenol species. Peaks labeled with the solid circle and solid triangle were later correlated to OH-BDEs 11 and 12 (described below). TIC for an organic extract of *E. coli* expressing bmpD− failed to detect the absence of bromide is shown as a negative control. (b) Chemical structures of bromophenol and bromopyrene monomers 4−7.

**Brominase Bmp2 catalyzes pyrrole tribromination**

We next sought biochemical confirmation for individual enzymatic transformations in the Bmp pathway, suggested by proline-based halopyrrole biosynthetic logic and gene deletion experiments. As characterized for the biosynthesis of pyrrole moieties in numerous alkaloid natural products, the proline adenylyltransferase BmpA, together with the flavin-dependent dehydrogenase BmpB, catalyzed the near-stoichiometric conversion of homo-Bmp1 to pyrrolyl-5-Bmp1 (Supplementary Figs. 6−8; detailed analytical protocols are described in Supplementary Note 1). Enzymatic conversion was abolished in the absence of either ATP or L-proline, whereas reduced conversion was observed in the absence of MgCl₂. These observations were consistent with the proposed reaction scheme for adenylyltransferases proceeding via an adenylated amino acid intermediate that is subsequently transferred to the phosphopentathine sulhydryl of the ACP with the concomitant release of AMP.

Bmp2 is homologous to canonical flavin-dependent halogenases that use ACP-tethered substrates (Supplementary Note 3). These enzymes require an exogenous supply of reduced flavin cofactor (FADH₂), typically provided in situ by a flavin reductase that catalyzes the reduction of the flavin cofactor with the concomitant oxidation of NAD(P)H. In the presence of E. coli flavin reductase SsuE, NADPH and bromide, Bmp2 catalyzed the bromination of pyrrolyl-5-Bmp1 (ACP) to mono-, di-, and tribromopyrrolyl-5-Bmp1 (ACP) (Fig. 3). This result established the in vitro reconstitution of a flavin-dependent halogenase with physiological bromination activity involved in natural product biosynthesis. Assignment of Bmp2 as a brominase is further supported by the observation that no incorporation of chloride was detected.
Although numerous bromopyrrole containing marine natural products have been identified\cite{1,2,3}, and 3 bear a distinct chemical signature in that all three available positions on the pyrrole ring are brominated, as opposed to mono- or dibromination, which is more commonly observed in brominated alkaloids with the 3-position of the pyrrole being universally unmodified.

Chemical logic for the biosynthesis of 2 and 3 using enzyme functionalities present within the bmp gene locus dictates that an ACP-bound tribromopyrrole intermediate is released by the TE domain of Bmp1 and decarboxylated by Bmp8 to give 6. Although the esterase activity of Bmp1 TE domain could indeed be confirmed by the hydrolysis of the model ester substrate p-nitrophenylacetate (Supplementary Fig. 10). Biochemical investigations into the formation of 6 by Bmp8 were precluded by our inability to generate soluble recombinant Bmp8 protein. However, the expression of bmp1-7 in E. coli expressing bmp8 resulted in altered levels of bromopyrroles, thus supporting a role of Bmp8 in bromopyrrole biosynthesis (Fig. 2a and Supplementary Fig. 11).

Bmp5, a decarboxylating phenol brominase

In contrast to assembly of bromopyrroles, the biosynthetic logic for the assembly of bromophenols and thereby of PBDEs and polybrominated biphenyls could not be readily anticipated. Previous isotope feeding experiments\cite{4} and the presence of a chorismate lyase (Bmp6) in the gene cluster suggested 4-hydroxybenzoic acid (4-HBA) as the precursor for the bromophenol moiety of 3. Although the bmp5 deletion experiments postulated its role as the monobromonaphthalene synthase, the lack of sequence similarity of the flavoenzyme Bmp5 to canonical flavin-dependent halogenases, which includes Bmp2, led to uncertainty in this assignment. Moreover, incorporation of 4-HBA into bromophenols necessitates decarboxylation, a transformation not previously associated with flavin-dependent halogenases. Hence, we undertook a detailed in vitro characterization of the activity of Bmp5. Upon incubation of 4-HBA with recombinant Bmp5 in the presence of bicarbonate, NADPH, and FAD, we observed a time-dependent formation of one major product via a stable intermediate (Fig. 4a and Supplementary Fig. 16). The identity of the intermediate as 3-bromo-4-hydroxybenzoic acid (8) and of the product as 4 was confirmed by MS and retention time comparison to authentic synthetic standards (Fig. 4a and Supplementary Fig. 17). Bmp5 could use 9 as a substrate for the synthesis of 4, further establishing 8 as the physiological intermediate for the reaction (Fig. 4b). Bmp5 also generated 5 when expressed in E. coli in the presence of bromide (Supplementary Fig. 17). Although production of 5 could not be detected in vitro, this observation is consistent with the acceptance of 3,5-dibromo-4-hydroxybenzoic acid (9) as a poor substrate by Bmp5 as compared to 4-HBA and 8 (Fig. 4b). Bmp5 activity was abolished in the absence of either bromide or NADPH, whereas a partial reduction in activity was observed upon omission of FAD, consistent with the partial occupancy of FAD in Bmp5 as purified (Supplementary Fig. 18). Consistent with prior characterization of halide specificity of aromatic halogenases\cite{5}, Bmp5 could use iodide in vivo, leading to the formation of iodo-phenols, whereas no chloride incorporation was observed (Supplementary Fig. 28).

In contrast to the flavin-dependent halogenases described to date, Bmp5 did not require the addition of a flavin reductase to regenerate FADH₂ in situ. This finding is consistent with sequence homology of Bmp5 to single-component flavin-dependent oxygenases (Supplementary Fig. 30). Consistent with the formation of electrophilic peroxy intermediates by flavin-dependent oxygenases and halomunium intermediates flavin-dependent halogenases\cite{6,7}, a putative reaction scheme catalyzed by Bmp5 can be discerned (Fig. 4c). To the best of our knowledge, Bmp5 represents the first example of a decarboxylating flavin-dependent halogenase. The stability of benzoxic acids distinguishes the Bmp5 decarboxylation activity from the extensively sampled marine haloperoxidases\cite{8,9} that participate in the deoxyhalosative degradation of halogenated derivatives of 3-oxo carbonylic acids and acyl homoserine lactones\cite{10,11}.

Bmp7 catalyzed coupling of bromophenols and bromopyrroles

We next investigated whether the bromophenol products of Bmp5 could be accepted as substrates by CYP450 Bmp7. In the presence of ferredoxin (Bmp9), ferredoxin reductase (Bmp10) and NADH, Bmp7 catalyzed the coupling of 4 to at least six distinct products that could be identified by LC/MS/MS (Fig. 5a). Characterization of Bmp7 products provided by MS/MS. NMR and comparison to authentic standards is detailed in Supplementary Note 2. We confirmed polybrominated biphenyls previously isolated from \textit{Pseudomonas stutzeri} spp., 1 (ref. 18) and 3,5,5'-tribromo-2,2'-biphenylidol (10) (ref. 10) as Bmp7 products (Fig. 5 and Supplementary Figs. 19 and 20). Additionally, a tribromo-biphenylidol isomer of 10 (Fig. 5a) was detected as a Bmp7 product but could not be isolated in sufficient quantities for comprehensive structure elucidation.
Guided by differential MS/MS fragmentation of polybrominated biphenyls and OH-BDEs, we further characterized three OH-BDE products, namely 2,6-dibromo-4-(2,4-dibromophenox) phenol (11), 2-bromo-4-(2,4-dibromophenox)phenol (12) and 4,4,6,8-tetrabromo-2-(3,4,5,6-tetrabromobenzofuro)[3,2-b] pyrrole (15) by comparison to authentic standards isolated from P. luteoviolacea 21t16 (Fig. 6a and Supplementary Figs. 26–27). Additionally, we detected an isomer of 3 as a minor product generated by Bmp7 during the in vitro coupling of 4 and 6, which could also be detected in the culture extracts of the three producer strains. Though a comprehensive structural characterization by NMR spectroscopy could not be realized, a characteristic hydroxytribromopyrrole MS/MS product ion suggests a polybrominated phenol-pyrrrole ether, a class of marine natural product not previously described (Supplementary Fig. 29).

DISCUSSION

Although several Bmp pathway products identified in this study were previously isolated from marine bacteria (1–3 (refs. 18–20), 7 (ref. 19), 10 and 15 (ref. 40), with 3 being reported in 1966, the genetic and molecular bases for their biosyntheses have remained elusive until now. In this study, we identified a conserved gene locus in marine γ-proteobacteria encoding the biosynthesis of at least 15 polybrominated aromatic marine natural products containing ubiquitous bromopyrrole and bromophenol building blocks. A near-complete in vitro reconstitution of all of the enzymatic steps within the Bmp pathway provides the basis for a modular biosynthetic scheme (Fig. 6a).

The bromopyrrole and bromophenol biosynthetic modules function independently and are united through the action of CYP450 Bmp7. CYP450-catalyzed phenol coupling reactions have extensive biochemical precedents, and the formation of polybrominated biphenyls 1 and 10 and OH-BDEs 11–14 are expected to follow accepted biradical mechanisms (Fig. 6b). A similar mechanism can be envisaged for coupling of bromopyrroles, consistent with CYP450-catalyzed coupling in indolecarboxylic acid antibiotics. The modularity of the Bmp pathway is reflected in the preservation of only the bromopyrrole biosynthetic module in the marine symbiont cyanobacterium P. didemni P2-F1 (Fig. 1c). As the bromophenol module is absent, it is likely that P. didemni P2-F1 produces only halogenated pyrroles. Furthermore, a homologous gene locus identified in T. thiooxydans B4A lacks a gene coding for a coupling CYP450 (Fig. 1c). It is conceivable that an even wider applicability of the chemical logic described here for the biosynthesis of marine bromophenols and bromopyrroles is limited only by the availability of genomic data.

The discovery of flavoenceyme Bmp7 as a phenol brominase was unexpected. Bmp7 is homologous to flavin-dependent oxygenses rather than flavin-dependent halogenases. Primary sequence motifs for binding both the flavin cofactor as well as the nicotinamide electron donor can be readily discerned within the primary sequence of Bmp5 (Supplementary Fig. 30). Indeed,
Figure 6 | Bi-modular scheme for the biosynthesis of polybrominated marine natural products by the bmp pathway. (a) Chorinate, the precursor for bromophenols, is converted to 4-HBA by chorinate lyase (CL) Bmp1 and then to 4 and 5 by flavin-dependent halogenase (Hal) Bmp2. The CyPA50 coupling (C) enzyme Bmp7 generates a suite of diverse polybrominated biphenyls (1 and 10) and OH-BDEs (11-14) from 4 and 5. The electron transfer partners for Bmp7, Bmp9 and Bmp10 are omitted for clarity. The bromopropones are derived from L-proline. Acylation of L-proline to the ACP domain of Bmp7 by the proline adenyltransferase (A) Bmp4 initiates its oxidation by the flavin-dependent dehydrogenase (DH) Bmp3 and tribromination by the flavin-dependent halogenase (Hal) Bmp2. The TE domain of Bmp7 most likely catalyzes the offloading of a carboxylic acid intermediate that is decarboxylated (D) by carboxyvinylolactone decarboxylase homolog Bmp8 to 6. 6 can be dimerized by Bmp7 to generate 2, or it can be dimerized with 4 to generate heterodimers such as 3 and 15. During the homodimerization of 6, we also observed the formation of 7. The coloring scheme is consistent with that in Figure 5c. (b) Proposed steps for radical generation (i), rearrangement (ii-iv) and coupling of 4 by Bmp7 to generate biphenyls and OH-BDEs. bmp7 homologs in the biosynthetic gene clusters identified in M. mediterranea MMB-1 and T. hioflexus strain AEB-4 are annotated as flavin-dependent oxidases (Fig. 1c). Although decarboxylation was until now an unrealized activity within the catalytic repertoire of flavin-dependent halogenases, ortho-hydroxylation of 4-HBA as well as para-decarboxylative hydroxylation of 4-HBA have been reported for flavin-dependent oxidases, leading to a proposal for a two-step reaction scheme for Bmp5 (Fig. 4c). The collective action of an electron-donating phenol hydroxyl and electron-withdrawing carboxyl group would first direct electrophilic bromination at the ortho position, and subsequent proton abstraction by a catalytic base would result in the formation of 8. This intermediate would then undergo a second bromonium addition para to the hydroxyl, followed by decarboxylation to generate 4. The Bmp5 reaction is regiospecific in that ortho bromination is the first half-reaction, followed by decarboxylative bromination at the para position. This assertion for the strictly directed two-step Bmp5 reaction scheme is supported by 4-bromophenol not being detected as an intermediate or as a product of the Bmp5 reaction (Fig. 4a). Furthermore, the Bmp5 active site most likely sterically occludes bromonium addition at one of the ortho positions of the aryl ring, thus biasing the second half-reaction toward decarboxylative-bromination at the para position, rather than a second ortho bromination to yield 9 as a second intermediate. This postulate is supported by the observation that under identical experimental conditions, rates of decrease of 4-HBA and 8 in an in vitro reaction with Bmp5 are nearly identical, whereas 9 is a poor substrate for Bmp5 (Fig. 4b). Consequently, 5 is a minor product generated by Bmp5 that could not be detected in vitro but is only detected to be produced in vivo when Bmp5 is heterologously expressed in E. coli in the presence of bromide in the culture medium (Supplementary Fig. 17). Additionally, Bmp5 could use iodide but not chloride in vivo (Supplementary Fig. 28).

It is notable that neither bromophenol monomers nor OH-BDEs had previously been isolated from marine bacterial sources. Bromophenols and OH-BDEs identified in this study (and their methylated derivatives) have instead been extensively detected in marine eukaryotes, ranging from autotrophs and invertebrates, to apex predators. Although sponge-associated symbiotic cyanobacteria were hinted as producers of OH-BDEs, a definite genetic basis for OH-BDE biosynthesis had not been established. Furthermore, in the absence of definitive genetic evidence for algal biosynthesis of OH-BDEs, a contribution of associated bacteria toward the production of OH-BDEs sourced from marine algae cannot be discounted. It is thus noteworthy that all three marine bacteria harboring the Bmp pathway as identified in this study also belong to genera commonly associated with marine eukaryotes. Identification of the bmp gene locus should henceforth aid the computational mining of marine metageneomic data sets for the identification of bacterial as well as potential eukaryotic OH-BDE producers using similar metrics. Together with 1 (Fig. 1a), the methylated derivative for 13 (2'-MeO-BDE-68; Fig. 5b) is extensively detected in marine mammals and even humans, suggesting that bromophenols and OH-BDEs identified in this study bioaccumulate in the marine food web and are transferred to humans via trophic connections. Furthermore, methylation of OH-BDEs and, by extension, of polybrominated biphenyls and bipyroles (Fig. 1a) seems to be a facile biochemical transformation consistent with a positive correlation between the existence of OH-BDEs and MeO-BDEs in marine environmental matrices.
OH-BDEs, such as those characterized in this study, are commonly found two distinct chemical signatures in which the hydroxy group is positioned either ortho or para to the ether linkage. Of the two isomers, para-OH-BDEs are potent inhibitors of thyroid hormone signaling, with 14 having a higher affinity for binding to the thyroid hormone transport proteins than its physiological substrate. Hence, the detection of para-OH-BDEs such as 11 and 12 and, specifically, 14 puts into perspective the toxic potential of the metabolites generated by the Bmp pathway. As the synthesis of biphenyls and diphenyl ethers could also be confirmed for Bmp7 using polychlorinated and polycyclic phenolic substrates (Supplementary Figs. 25 and 28), other naturally occurring polyhalogenated aryl compounds may also be derived from analogous enzymatic pathways with differing halogen specificities of the halogenases. As the industrial production of polybrominated flame retardant chemicals continues to phased out in recognition of their persistence and toxicity, we anticipate that naturally produced marine bacterial PBDEs will take a center stage in informing our small molecule-mediated interactions with the environment.

Received 15 December 2013; accepted 19 May 2014; published online 29 June 2014

METHODS

Accession codes. GenBank. Sequence data for P. atrovocilis 21a16 and P. phenolica O-BC30 has been deposited under accession codes AUSV0000000 (accession code for contig containing the Bmp locus: AUSVI0000133.1) and KFS40211 (Bmp locus accession code KFS40211.1), respectively.

References

Acknowledgments
We thank our colleagues E. Trinh for preliminary in vitro studies with Brmp6, R.M. Driggers for assistance in NMR data collection, P.A. Jordan for NMR data analysis, T. So for MS data collection and L. Altorres for useful discussions. This work was jointly supported by the US National Science Foundation (CHE-1115747) and the US National Institute of Environmental Health Sciences (PO1 ES021921) through the Oceans and Human Health program, the Gordon and Betty Moore Foundation Marine Microbial Sequencing Project, the Helen Hay Whitney Foundation postdoctoral fellowship to V.A., the US National Institutes of Health (N01-HV-68553) Marine Biobank grant, the National Institutes of Health (N01-HV-68581) Marine Biobank grant, and the National Institutes of Health (N01-HV-68583) Marine Biobank grant.

Author contributions

Competing financial interests
The authors declare no competing financial interests.

Additional information
Supplementary information and chemical compound information is available in the online version of the paper. Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Correspondence and requests for materials should be addressed to R.M.D.
Supplementary Information for

Biosynthesis of polybrominated aromatic organic compounds by marine bacteria

Vinayak Agarwal, Abrahim A. El Gamal, Kazuya Yamanaka, Dennis Poth, Roland D. Kersten, Michelle Schorn, Eric E. Allen, Bradley S. Moore
SUPPLEMENTARY RESULTS

Supplementary Note 1

Detection of dehydrogenation and halogenation of L-proline acylated to Bmp1.

Trypsin-digested N-His8c-Bmp1(ACP) was analyzed by LC-MS/MS. 10 μL of the trypsin-digested assay mixture was injected onto a reverse phase C4 column (Vydac 5 μm, 4.6 mm × 250 mm). Water + 0.1% formic acid was used as solvent A, and MeCN + 0.1% formic acid was used as solvent B. The elution profile was as follows (0.7 mL/min): 10% B for 10 min, linear increase to 30% B across 5 min, linear increase to 70% B across 40 min, linear decrease to 10% B across 5 min, linear increase to 100% B across 1 min followed by 2 min at 100% B, decrease to 10% B across 1 min, 10% B for 2 min and 5 min of post-time equilibration. The assay relies on the detection of MS2 product ions i–v as shown in Supplementary Fig. 7. Briefly, ion i would be generated by collision-induced disassociation for holo-N-His8c-Bmp1(ACP) and ion ii by pyrrolyl-S-N-His8c-Bmp1(ACP). Ions iii–v correspond to brominated variants for phosphopantetheinyalted halo-pyrrole ions as previously described in the literature1,2. Expression of N-His8c-Bmp1(ACP) in E. coli should provide a mixture of apo-N-His8c-Bmp1(ACP), and holo-N-His8c-Bmp1(ACP) as a result of crosstalk with fatty acid biosynthesis. The side chain hydroxyl of the serine residue at the carboxy terminus of the α-2 helix for the Bmp1(ACP) should be post-translationally esterified with phosphopantetheine for holo-N-His8c-Bmp1(ACP), but not for apo-N-His8c-Bmp1(ACP). By sequence homology, this serine residue was identified to be Ser35 for Bmp1. Holo-N-His8c-Bmp1(ACP) will not participate further in the assay, as acylation by Sfp (B. subtilis phosphopantetheinyll transferase) requires the presence of a non-esterified Ser35 side chain. The presence of holo-N-His8c-Bmp1(ACP) was confirmed by detection of MS2 ion i. As a second means of independent verification, the MS2 ion i was traced back to a MS1
peptide fragment of N-His$_5$-Bmp1(ACP) that bears the Ser25 residue. For peptide mapping, the mass of the holo-peptide fragment was calculated based on the mass of the monoisotopic peak for the peptide, and its charge state ($z$). For calculation of the mass of the apo-peptide fragment, the mass corresponding to ion vi was subtracted from it. The mass of the apo-peptide fragment thus determined could be traced to the peptide fragment generated by trypsin digestion of His$_5$-Bmp1(ACP) as shown in Fig. 3 and Supplementary Fig. 8. The equations for calculation of the apo-peptide fragment mass corresponding to holo-N-His$_5$-Bmp1(ACP) as identified in Supplementary Fig. 8 are as follows:

$$z = \frac{1}{(mass\ difference\ between\ successive\ isotope\ peaks)}$$

$$apo\ peptide\ mass = [(parent\ ion \times z) - (z \times 1.00728)] - 340.0858$$

Upon transfer of 16 to apo-N-His$_5$-Bmp1(ACP), MS2 ion ii could be detected, and upon mono-, di-, and tri-halogenation by Bmp2, ejection ions corresponding to iii-v could be detected. The apo-peptide fragments generating these MS2 ions were mapped back to Bmp1(ACP) by subtraction of the mass of ions viii-x (Supplementary Fig. 7) from the respective holo-peptide masses.

Using this assay methodology, we could generate N-His$_5$-pyrrolyl-S-Bmp1(ACP) by loading 16 onto the apo-N-His$_5$-Bmp1(ACP). Upon addition of Bmp2, SsuE (E. coli flavin reductase), NADPH and KBr, mono-, di-, and tri-brominated Bmp1(ACP) peptides could be detected, thus establishing tri-bromination of the pyrrole ring by Bmp2. No halogenation was observed when KBr was replaced with KCl or KI. Regiospecificity for bromination of the pyrrole ring by Bmp2 is assumed to be similar to the pyrrole chlorinase PltA due to high sequence similarity between the two enzymes.
Supplementary Note 2

Identification and characterization of polybrominated biphenyl and OH-BDE products generated by Bmp7 in vitro, and a mass spectrometry-based detection of OH-BDEs in culture extracts.

MS1 for the Bmp7 in vitro reaction analyte (Supplementary Fig. 20, panel a) led to determination of the molecular weights of different product species, and together with the typical isotopic distribution afforded by bromination, molecular formulae of the products could be discerned. With 4 as a substrate, three isomers each for the molecular formulae C_{12}H_{8}O_{2}Br_{4} (Supplementary Fig. 20, panel b) and C_{12}H_{7}O_{2}Br_{3} (Supplementary Fig. 20, panel c) could be reproducibly detected as the major products of the reaction. The most abundant isotope ion for each formula was then selected for targeted MS/MS by collision induced disassociation. An identical approach was used to analyze products of the reaction with 2,4-dichlorophenol (17) (described later). The molecular formulae of the major products, C_{12}H_{6}O_{2}Br_{4} and C_{12}H_{7}O_{2}Br_{3}, could be putatively assigned to biphenyl products 1 and 10, respectively, as both molecules have previously been isolated from Pseudoalteromonas species. Upon chromatographic analysis of authentic synthetic standards of 1 and 10 (Supplementary Fig. 19), we could assign one of the isomers for C_{12}H_{6}O_{2}Br_{4} as 1, and one of the isomers for C_{12}H_{7}O_{2}Br_{3} as 10. These assignments were further confirmed by preparative scale purification of these two products from large scale in vitro Bmp7 reactions, and comparison of ¹H NMR spectra against values reported in literature⁵⁻⁴. While the mechanism for synthesis of 1 from 4 is expected to be analogous to bi-radical aryl coupling reactions⁵⁻⁶, the production of 10 would additionally entail dehalogenation that has been reported in literature for aryl-coupling enzymes employing radical intermediates⁷.
MS/MS analysis for 1 and 10 showed [M-Br\(^-\)] and [M-2Br\(^-\)] MS2 product ions (Supplementary Fig. 20, panels d–e, and panel f respectively), as would be expected for polybrominated biphenyls. Another isomer for C\(_{12}\)H\(_5\)O\(_2\)Br\(_3\) also showed [M-Br\(^-\)] MS2 product ions (Supplementary Fig. 20, panel f), leading to its assignment as a biphenyl (denoted by * in Fig. 5a). However, the molecule could not be isolated from large-scale Bmp7 reactions in sufficient quantities for comprehensive structure elucidation by NMR spectroscopy. As this species is isomeric with 10, and is putatively a polybrominated biphenyl, we postulate its structure to be 3,3',5-tribromo-[1,1'-biphenyl]-2,4'-diol.

At this point, the identities of the other two isomers of C\(_{12}\)H\(_5\)O\(_2\)Br\(_4\) and one isomer for C\(_{12}\)H\(_5\)O\(_2\)Br\(_5\) were unknown. One of the isomers of C\(_{12}\)H\(_5\)O\(_2\)Br\(_4\) (retention time ~ 28.2 min) showed MS2 product ions corresponding to dibromobenzoquinone (Supplementary Fig. 20, panel g). This ion could not be generated by a polybrominated biphenyl species, as scission of the carbon-carbon bond between the two aryl rings would lead to the formation of a dibromophenol ion, regardless of the positions of the bromines on the benzene ring. However, an OH-BDE molecule, in which scission of the ether bond occurs prior to the loss of bromine during MS/MS can generate a MS2 ion corresponding to dibromobenzoquinone as observed in panel g, and also reported in literature for para-OH-BDEs\(^8\). Hence, we putatively assigned this species as a para-OH-BDE. This hypothesis was supported by the observation that an isomeric OH-BDE to 1 should have a greater retention time than 1, as one of the hydrophilic hydroxyl groups in 1 would be involved in ether bond formation in the OH-BDE. Preparative scale purification of this species from large scale Bmp7 reactions, and structure elucidation by NMR led to assignment of the structure as 2,6-dibromo-4-(2,4-dibromophenoxy)phenol (11) (Supplementary Fig. 21). Structure assignments for OH-BDEs are based upon \(^1\)H NMR, HSQC, HMBC, and H\(_2\)BC experiments. \(^{13}\)C NMR shifts, deduced from HSQC, HMBC and H\(_2\)BC correlations, are listed in the figure below. Specifically, 11
was isolated by preparative HPLC separation of products from ethyl acetate extracts of large scale Bmp7 reactions with 4. The solvent was removed in vacuo, and dried overnight to yield a white residue. The residue was dissolved in d$_6$-DMSO and NMR spectra collected using 600 MHz Varian NMR microprobe. Deduced structure of the molecule with proton shifts listed in red, and carbon shifts as deduced from HSQC and HMBC couplings listed in green is shown in panel a below. HSQC correlations are shown as thick bonds, HMBC correlations are shown as red arrows, and H2BC$^{13}$ correlations are shown as blue arrows in panel b below.

Note that the symmetry of one aryl ring is supported by the presence of a singlet in the $^1$H NMR (Supplementary Fig. 21), and the HMBC coupling shown as a bold arrow that is identical to a HSQC coupling. Symmetry for the aryl rings is also supported by MS/MS fragmentation consistent with para-OH-BDE$^8$ (Supplementary Fig. 20). The isomeric ortho-OH-BDE (2'-'OH-BDE-68) would have no symmetrical aryl ring. The $^1$H and $^{13}$C shifts
reported here for the non-symmetrical aryl ring are in agreement with those reported in literature\textsuperscript{12}.

One of the isomers for C\textsubscript{12}H\textsubscript{7}Br\textsubscript{3} exhibiting MS2 product ions corresponding to monobromobenzoquinone (Supplementary Fig. 20, panel h), was isolated and its structure characterized by NMR spectroscopy to be 2-bromo-4-(2,4-dibromophenoxy)phenol (12). Specifically, 12 was isolated by preparative HPLC purification of ethyl acetate extracts of large scale Bmp7 reactions with 4. The solvent was removed \textit{in vacuo}, and dried overnight to yield a white residue. The residue was dissolved in \textit{d}\textsubscript{6}-DMSO and NMR spectra collected using 600 MHz Varian NMR microprobe. Deduced structure of the molecule with proton shifts listed in red and carbon shifts listed in green is shown panel a below. HSQC correlations are shown as thick bonds, HMBC correlations are shown as red arrows, and H\textsubscript{2}BC\textsuperscript{13} correlations are shown as blue arrows in panel b below.
12: $^1$H NMR (600 MHz, DMSO) $\delta$ 7.93 (d, $J = 2.3$ Hz, 1H), 7.52 (dd, $J = 8.8, 2.4$ Hz, 1H), 7.20 (d, $J = 2.9$ Hz, 1H), 6.98 (d, $J = 8.8$ Hz, 1H), 6.91 (dd, $J = 8.8, 2.8$ Hz, 1H), 6.86 (d, $J = 8.8$ Hz, 1H). The $^1$H and $^{13}$C shifts reported here for the aryl ring bearing two bromine atoms are in agreement with those reported in literature. Overall, the NMR assignments are in agreement with previous report of 12 isolated from marine algae. Structure assignment is also supported by MS/MS fragmentation consistent with para-OH-BDE (Supplementary Fig. 20).

Hence, as this point, all three isomers for C$_{12}$H$_5$O$_2$Br$_3$ and two isomers for C$_{12}$H$_6$O$_2$Br$_2$ had been identified. Both 11 and 12 are para-OH-BDEs, in that the ether linkage is positioned para to the phenolic hydroxyl. The third, and as yet unidentified isomer for C$_{12}$H$_5$O$_2$Br$_2$, was also postulated to be an OH-BDE due to its higher retention time than 1, and similar retention time as the OH-BDE 11. However, this species demonstrated no major MS2 product ions. Envisaging a OH-BDE isomer for 11, we rationalized the structure for this species to be 2,4-dibromo-6-(2,4-dibromophenoxy)phenol (2'-OH-BDE-68) (13). This hypothesis was confirmed by the successive addition of increasing amounts of commercially available synthetic 13 (AccuStandard, HBDE-4006S-CN) to the Bmp7 reaction extracts and by observation of an identical retention time of the Bmp7 reaction product to synthetic 13 standard (Supplementary Fig. 22). Hence, by a combination of LC-MS/MS, NMR spectroscopy and comparison to synthetic standards, all major products produced by the Bmp7 reaction could be identified. Notably, 13 is an ortho-OH-BDE, in that the ether linkage resides ortho to the phenolic hydroxyl. Differential MS/MS fragmentation of ortho- and para-OH-BDEs has been reported in literature. The dibromobenzoquinone and monobromobenzoquinone MS2 ions for 11 and 12 respectively were used as signatures to mine the LC-MS/MS data for bacterial extracts, leading to the identification of 11 and 12 in the extracts of producer strains. It should be noted that the culture extracts contained more
than fifteen polybrominated aromatic organic molecules, in addition to other small molecules. This precluded a facile prior identification of OH-BDE molecules, as done for 1–3 from the culture extracts. Additionally, identification of ortho-OH-BDE relies on comparison to authentic synthetic standards as they do not possess a distinct MS/MS signature.

5 did not demonstrate a typical absorbance shift associated with substrate binding for Bmp7 (Supplementary Fig. 12). However, as 5 is a product generated by Bmp5 (Supplementary Fig. 17) and is present in the Δbmp7 extracts (Supplementary Fig. 5), we explored whether 5 could be used by Bmp7 as a substrate. Indeed, Bmp7 could use 5, though substrate consumption was significantly lower as compared to that for 4 (Supplementary Fig. 23). A singular major product could be identified, with the MS1 predicted formula C_{12}H_{13}Br_{2}O_{2} (Supplementary Fig. 24, panel a). As a dibromobenzoquinone MS2 product ion was observed (Supplementary Fig. 24, panel b), we postulated that the product of the reaction to be a para-OH-BDE. Preparative-scale purification of the product from large scale Bmp7 reactions with 5, and subsequent NMR characterization led to the identification of the chemical structure of the product as 2,6-dibromo-4-(2,4,6-tribromophenoxy)phenol (14). Specifically, 14 was isolated by preparative HPLC separation of products from ethyl acetate extracts of large scale Bmp7 reactions with 5. The solvent was removed in vacuo, and dried overnight to yield a white residue. The residue was dissolved in d_{6}-DMSO, NMR spectra collected using 600 MHz Varian NMR microprobe. Final deduced structure of the molecule, with the proton shifts listed in red, and the carbon shifts listed in green is shown in panel a below. HSQC correlations are shown as thick bonds, HMBC correlations in red arrows are shown in panel b below. No H2BC correlations were observed.
14: $^1$H NMR (600 MHz, DMSO) δ 8.08 (s, 1H), 7.06 (s, 1H). Uncertainty in the assignment of two HMBC couplings is shown as dashed arrows, and is likely caused by the highly similar $^{13}$C shifts of the two carbon atoms in the symmetrical ring. Note that the symmetry of both aryl rings is supported by the presence of only two singlets in the $^1$H NMR, and the HMBC couplings shown in bold arrows that are identical to HSQC couplings. Symmetry for both rings is also supported by MS/MS fragmentation consistent with para-OH-BDEs\(^8\) (Supplementary Fig. 24). A hypothetical isomeric ortho-OH-BDE would have only one symmetrical aryl ring.

Using the above discussion for products generated by Bmp7 using 4 and 5 as substrates, we could rationalize the products generated by Bmp7 using 17 as a substrate (Supplementary Fig. 25, panel a). It should be noted that 17 is not a physiological product generated by Bmp5 as Bmp5 does not incorporate chloride. LC-MS/MS analysis identifies three major isomeric product peaks with the molecular formula $C_{12}H_{14}ClO_2$ (Supplementary Fig. 25, panel b). The first product peak showed [M-Cl]$, [M-2Cl]$, and [M-3Cl]$\ M$:S2 product ions (Supplementary Fig. 25, panels c, d and e, respectively). This lead to the assignment of this species as 4,4',6,6'-tetrachloro-2,2'-biphenol (18), a chlorinated analog of 1. The second isomeric product peak demonstrated MS2 product ions corresponding to
dichlorobenzoquinone (Supplementary Fig. 25, panel f), leading to a putative assignment of its structure as 2,6-dichloro-4-(2,4-dichlorophenoxy)phenol (19). The third isomeric product did not show any major MS2 product ions. By analogy to Bmp7 brominated products, the structure for this species can be postulated to be 2,4-dichloro-6-(2,4-dichlorophenoxy)phenol (20). Though 18–20 are not as yet reported natural products, several million tons of 17 are industrially synthesized each year, and also generated by the degradation of the herbicide 2,4-dichlorophenoxyacetic acid. Hence, it is conceivable that 17 could exist in the marine microbial metabolome, owing to anthropogenic origins, or enzymatically synthesized by homologs of the Bmp5 enzymes. Consequently, 18–20, and other reaction products of Bmp7 with 17 may already exist as halogenated marine natural products.
Supplementary Figures

**Supplementary Fig. 1. Production of signature polybrominated molecules 1–3 by marine bacteria.** Extracted ion chromatograms (EICs) for extracts of *P. luteoviolacea* 2ta16, *P. phenolica* O-BC30, and *M. Mediterranea* MMB-1 analyzed by LC-MS/MS demonstrating the production of 1 (red curves), 2 (violet curves), and 3 (orange curves). Bacteria were grown in Difco 2216 Marine Broth without additional supplementation of bromide. Extracts dissolved in MeOH were injected onto a reverse phase C18 column (Phenomenex Luna, 5 μm, 4.6 × 100 mm) operating on an Agilent 1260 HPLC in tandem to an Agilent 6530 Accurate Mass Q-TOF mass spectrometer. The solvents used for HPLC were water + 0.1% formic acid (A) and MeCN + 0.1% formic acid (B). The HPLC elution profile was as follows (flow rate 0.5 mL/min): 10% B for 5 min, linear gradient to 70% B across 10 min, linear increase to 100% B across 30 min, 100% B for 3 min, linear decrease to 10% B across 2 min, 10% B for 5 min. An identical LC-MS/MS profile has been used for all analytical scale heterologous culture extract analysis, Bmp7 reaction extracts, and for synthetic standards.
Supplementary Fig. 2. High-resolution mass spectra for selected polybrominated metabolites from an organic extract of *P. luteoviolacea* 2ta16. For accurate mass determination, an organic extract of *P. luteoviolacea* 2ta16 was analyzed by LC-ESI-MS in negative polarity on an Agilent 6230 Accurate Mass TOF-MS at the UCSD Chemistry and Biochemistry Molecular Mass Facility (La Jolla, CA). Isotope patterns for 1–7, 10 and 15 identified in the culture extract of *P. luteoviolacea* 2ta16 are shown. See also Supplementary Table 1 for error estimation in mass spectra.
Supplementary Fig. 3. Chemical logic for the biosynthesis of halogenated pyrrole moieties in natural products. The scheme for the generation of dichlorinated pyrroles, tethered to an ACP molecule is shown as characterized for pyoluteorin biosynthesis. The ACP (PttL) is phosphopantetheinylated to yield a free sulphydryl moiety, which is then acylated to an amino acid (L-proline) by virtue of a thioester bond by an adenyltransferase PttF. This reaction requires the hydrolysis of a stoichiometric amount of ATP to AMP and PPI. Subsequent to acylation, the pyrrolidine ring is dehydrogenated by PttE, with concomitant reduction of two FAD molecules to FADH₂. The pyrrole ring is then dichlorinated by the chlorinase PttA (5-position followed by 4-position). Halogenation requires reduced flavin cofactor, and appropriate halide ion in the presence of molecular oxygen. In the pyoluteorin biosynthetic scheme, the dichlorinated pyrrolyl-5-ACP then undergoes three rounds of polyketide extension, and subsequent release and aromatization. The dichlorinated pyrrole moiety in pyoluteorin is highlighted in red. The canonical four helix-bundle architecture of ACP (PttL) is shown in the red cartoon.
Supplementary Fig. 4. pGAL-bmp1–7 yeast/E. coli shuttle vector. pGAL-bmp1–7 yeast/E. coli shuttle vector was assembled in S. cerevisiae BJ5465 from PCR amplified fragments (for primers see Supplementary Table 3).
Supplementary Fig. 5. Chemotype for bnp7 deletion. Deletion of bnp7 from the heterologous expression vector pETDuet-bnp1–7 results in accumulation of pyrrole and phenol monomers 4–6 in E. coli. Notably absent is 7, which likely arises as a byproduct of dehalogenative radical coupling catalyzed by Bnp7. The figure shows EICs for the most abundant ions corresponding to each compound (the y-axis scale is identical for all plots).
Supplementary Fig. 6. Plasmids for co-expression of *bmp1* with *bmp3*. Upon coexpression of the above two plasmids in *E. coli* BL21Gold(DE3), Bmp1(ACP) could be purified in complex with Bmp3. Purified Bmp1(ACP)-Bmp3 complex was subsequently used for assays with Bmp4 as described in Supplementary Fig. 8.
Supplementary Fig. 7. MS2 product ions corresponding to phosphopantetheine, phosphopantetheine acylated with pyrrole, and subsequent brominated states of the pyrrole. Note that the sites for the mono- and di-bromination for the pyrrole have been assumed as characterized for the pyrrole chlorinase PhtA (Supplementary Fig. 3). The masses for the MS2 ions are listed. Under the experimental conditions in this study, MS2 product ions I–V were observed.
Supplementary Fig. 8. Enzymatic synthesis of pyrrolyl-S-Bmp1(ACP) by Bmp3 and Bmp4. (a) 40 μM of Bmp1(ACP)-Bmp3 complex, purified as described in Supplementary Fig. 6 was incubated with 1 μM purified Bmp4, 4 mM L-proline, 2 mM ATP, 2 mM DTT and 10 mM MgCl₂ for 3 h at 30 °C in 100 mM potassium phosphate (pH 7.6) buffer.
Reaction components (b) ATP, (c) Bmp4, (d) L-proline, and (e) MgCl₂ were individually deleted and replaced with buffer. The assays were analyzed by LC-MS/MS. A peptide fragment corresponding to holo-Bmp1(ACP) (shaded blue) was identified by a diagnostic MS2 ion corresponding to species i as shown in Supplementary Fig. 7, while a peptide fragment corresponding to pyrrolyl-S-Bmp1(ACP) (shaded yellow) was identified by the diagnostic MS2 ion corresponding to species ii as shown in Supplementary Fig. 7. Both peptides were found to bear +8 charges as determined by their MS1 isotopic mass distribution. For illustrative purposes, MS1 spectra were summed over an identical 4 min time window during which both holo-Bmp1(ACP) and pyrrolyl-S-Bmp1(ACP) peptides were found to elute.

In the presence of all reaction components, Bmp4 and Bmp3 catalyzed the conversion of holo-Bmp1(ACP) to pyrrolyl-S-Bmp1(ACP) (panel a). Elimination of ATP, Bmp4 or L-proline abolished the formation of pyrrolyl-S-Bmp1(ACP). However, elimination of MgCl₂ (panel e) only led to a partial reduction in activity of Bmp4. These findings are consistent with the well-accepted reaction mechanism of adenyltransferase enzymes, such as Bmp4, in which the amino acid (L-proline for Bmp4) is activated by adenylation (derived from ATP), and subsequently transferred on to the terminal sulfhydryl of the phosphopantetheine arm of the holo-ACP.
Supplementary Fig. 9. $^1$H NMR spectrum for 16.

$^1$H NMR (600 MHz, D$_2$O) δ 8.67 – 8.63 (m, 1H), 8.33 (s, 1H), 7.16 – 7.12 (m, 1H), 7.15 (d, $J$ = 7 Hz, 1H), 6.29 – 6.26 (m, 1H), 6.18 (d, $J$ = 10 Hz 1H), 4.63-4.62 (m, 1H), 4.30 – 4.23 (m, 2H), 4.04 (s, 1H), 3.88 – 3.84 (m, 2H), 3.61 – 3.57 (m, 2H), 3.48 – 3.46 (m, 1H), 3.45-3.43 (m 1H) 3.41 (m, 1H), 3.23 (s, 1H), 3.12 (m, 1H), 2.44 (t, $J$ = 10 Hz, 2H), 0.94 (s, 3H), 0.80 (s, 3H).
Supplementary Fig. 10. Hydrolysis of model esterase substrate p-nitrophenyl acetate (pNPA) by Bmp1(TE). 5 μM of purified Bmp1(TE) was used as the catalyst for the hydrolysis of 250 μM freshly prepared pNPA in 100 mM potassium phosphate (pH 7.6) buffer at room temperature. The reaction was monitored spectrophotometrically at wavelength of 400 nm in a quartz cuvette. Observed absorbance is plotted against time. Upon addition of enzyme, we observed significant increase in the hydrolysis of pNPA (denoted by the absorbance increase due to production of p-nitrophenol), implying that Bmp1(TE) catalyzed the hydrolysis of pNPA.
Supplementary Fig. 11. Chemotype for heterologous expression of *bmp1–7* omitting *bmp8*. Expression in *E. coli* of *bmp1-7* omitting *bmp8* results in decreased levels of 3 and 6 with a concomitant increase in 7; 1 is shown as a control. Cultures of *E. coli* expressing the full *bmp* cluster or full *bmp* cluster excluding *bmp8* were grown in triplicate, extracted, and analyzed by LC-MS/MS as previously described. Levels of 3, 6, and 7 were quantified by peak integration of extracted ion chromatograms (EICs) corresponding to the most abundant ion for each compound. Means were calculated from each of three replicates for integrals taken from EICs for each ion and plotted, with the error bars representing standard deviations.
Supplementary Fig. 12. Spectral characteristics of CYP450 Bmp7 and spectrophotometric determination of substrate tolerance. (a) An absorbance scan for 5 μM purified Bmp7 in 100 mM potassium phosphate buffer (pH 7.6) was performed using a Cary-40 UV-Vis spectrophotometer in a quartz cuvette. Bmp7 demonstrates a typical CYP450 absorbance curve, with a Soret peak at 415 nm (black curve) upon addition of 1 mM imidazole (red curve). (b) 6 was synthesized according to published protocols. Difference absorbance spectra upon titration of Bmp7 with 200 μM 6 or 7 identify binding of 6 in a manner characteristic for substrates of CYP450 enzymes; note that 7 does not yield a similar absorbance change. (c) Similarly, a difference absorbance spectra upon titration of Bmp7 with 4 or 5 identifies binding of 4 characteristic for substrates of CYP450 enzymes.
Supplementary Fig. 13. *In vitro* Bmp7 reaction products with 6 (black, top curve) and 7 (bottom, red curve) as substrates. With 6 as a substrate, Bmp7 generated three distinct products, 2, 7, and an uncharacterized species corresponding to pentabromo-bipyrrrole (marked by *, having a molecular formula C₉H₃Br₃N₂ as predicted from MS1 spectrum). This product could not be generated in quantities sufficient for complete structure elucidation. The pentabromo-bipyrrrole product can be rationalized on the basis of dehalogenated-aryl coupling polybrominated biphenyl and OH-BDE products 10, 12 and 14 (*vide infra*). Upon reaction with 7, no product peaks could be observed.
Supplementary Fig. 14. Characterization of 2 generated by Bmp7 in vitro reaction with 6 as a substrate. Extract of Bmp7 reaction (black curves) was compared to an authentic NMR characterized sample of 2 (red curves) isolated from P. luteoviolacea 2ta16. (a, b) Comparison of retention times. MS1 EIC corresponding to the most abundant isotope for the molecular formula C₆H₂Br₆N₂ (m/z = 604.51, 10 ppm tolerance) for the Bmp7 reaction extract (a) and authentic standard of 2 (b) analyzed under identical chromatographic conditions. (c, d) Comparison of MS1 profiles. MS1 mass spectrum for species at retention time 23.4 min from the Bmp7 reaction extract (c) and authentic standard of 2 (d). (e, f) Comparison of MS/MS profiles. MS2 profile of the product ions generated from most abundant isotope MS1 ions corresponding to the molecular formula C₆H₂Br₆N₂ at identical retention times from the Bmp7 reaction extract (e) and authentic standard of 2 (f). Most abundant MS2 ions corresponded to the [M-Br]⁻. Identical retention times, MS1 and MS/MS profiles as compared to a NMR characterized authentic standard lead to the characterization of the Bmp7 reaction product as 2.
NMR spectra for 2 purified from preparative scale cultures of *P. luteoviolacea* 2ta16 and used as authentic standard in panels b, d and f above:

2. $^1$H NMR: (CDCl$_3$) $\delta$ [ppm] 9.87 (bs, 1H); $^{13}$C NMR: (CDCl$_3$) $\delta$ [ppm] 121.60, 115.37, 102.76, 100.98
Supplementary Fig. 15. Characterization of 7 generated by Bmp7 \textit{in vitro} reaction with 6 as a substrate. Extract of Bmp7 reaction (black curves) was compared to an authentic commercial synthetic standard of 7 (Sigma-Aldrich, L165042-50MG) (red curves). (a, b) Comparison of retention times. MS1 EIC corresponding to the most abundant isotope for the molecular formula C$_4$HBr$_4$N (m/z = 381.67, 10 ppm tolerance) for the Bmp7 reaction extract (a) and authentic standard of 7 (b) analyzed under identical chromatographic conditions. (c, d) Comparison of MS1 profiles. MS1 mass spectrum for species at retention time 20.8 min from the Bmp7 reaction extract (c) and authentic standard of 7 (d). (e, f) Comparison of MS/MS profiles. MS2 profile of the product ions generated from most abundant isotope MS1 ions corresponding to the molecular formula C$_4$HBr$_4$N at identical retention times from the Bmp7 reaction extract (e) and authentic standard of 7 (f). Most abundant MS2 ions corresponded to the [M-Br$^-$]. Identical retention times, MS1 and MS/MS profiles as compared to an authentic synthetic standard lead to the characterization of the Bmp7 reaction product as 7.
Supplementary Fig. 16. Purification of recombinant Bmp5 and time course for the change in concentrations of the substrate, intermediate and product species as shown in Fig. 4a. (a) Affinity chromatography purified Bmp5 is analyzed by SDS-PAGE, along with molecular weight marker (Fisher Scientific BP3602) with molecular weights listed in red in kDa. The expected molecular weight for N-His₆-Bmp5 is 61 kDa (b) Area under the substrate, intermediate and product peaks as shown in Fig. 4a were integrated, and mean values from three independent experiments were plotted against reaction time with error bars representing standard deviation. The intermediate and product were identified to be 8 and 4 respectively by high-resolution mass spectrometry and comparison of retention times to authentic synthetic standards.
Supplementary Fig. 17: Mass spectra for 8 and 4, and in vivo production of 4 and 5 by Bmp5. The mass spectra for (a) Bmp5 intermediate 8 and (b) Bmp5 product 4 generated by in vitro reaction of Bmp5 with 4-HBA led to the assignment of the chemical formulae as C7H7O3Br and C8H7OBr2 respectively. (c) MS1 EIC for m/z = 248.85 (10 ppm tolerance) and m/z = 326.766 (10 ppm tolerance) showing the production of 4 (top) and 5 (bottom) when Bmp5 was expressed in E. coli in LB media supplemented with 1 g/L KBr. Assuming identical ionization for 4 and 5, it can be discerned that 4 is the major product generated by Bmp5.
Supplementary Fig. 18. Negative control reactions demonstrate the requirement of KBr and NADPH for catalysis by Bmp5. No conversion of 4-HBA to 8 and 4 was observed when Bmp5, NADPH or KBr were omitted from the Bmp5 in vitro assay. Omission of FAD did not abolish enzyme activity, but lowered substrate turnover.
Supplementary Fig. 19. Comparison of retention times and NMR spectra of polybrominated biphenyl products 1 and 10 generated by Bmp7 in vitro with authentic synthetic standards. Extract of Bmp7 reaction with 4 (red curve) was analyzed using identical chromatographic conditions as for authentic synthetic standards of 1 (black curve) and 10 (blue curve). Note that retention times for authentic standards of 1 and 10 match those for two separate product peaks from the enzymatic reaction leading to the identification of the chemical structures of these species. Synthetic standard of 1 was obtained from Sigma-Aldrich (S842753-50MG) and dissolved in 1 mL DMSO. Analysis by LC-MS/MS identified two major peaks in roughly 80:20 relative abundance with respective molecular formulae as C_{12}H_{10}O_{4}Br_{4} and C_{13}H_{11}O_{4}Br_{3}. Both products were purified by semi-preparative scale HPLC. After drying in vacuo, 500 μg of each product was dissolved in 50 μL DMSO, and 1H NMR spectra were acquired. By molecular formulae, MS/MS fragmentation patterns, and comparison of NMR spectra to literature\textsuperscript{3-4}, the identity of the synthetic standard of 1 and for the synthetic standard of 10 can be discerned as shown above. This lead to the identification of 1 and 10 as products generated by Bmp7.
The $^1$H NMR chemical shifts of the synthetic standards of 1 and 10, which are in good agreement to literature values, are as follows:

1: $^1$H NMR (600 MHz, CDCl$_3$) δ 7.65 (d, $J = 2.3$ Hz, 1H), 7.31 (d, $J = 2.3$ Hz, 1H).

10: $^1$H NMR (600 MHz, CDCl$_3$) δ 7.67 (d, $J = 2.3$ Hz, 1H), 7.39 (dd, $J = 8.6$, 2.4 Hz, 1H), 7.37 (d, $J = 2.3$ Hz, 1H), 7.34 (d, $J = 2.4$ Hz, 1H), 6.90 (d, $J = 8.6$ Hz, 1H).
Supplementary Fig. 20. Mass spectrometric analysis of the polybrominated biphenyls, and OH-BDE products generated by Bmp7. (a) Total ion chromatogram (TIC) for the Bmp7 reaction extract. Peaks corresponding to the three isomers each for the major products corresponding to the molecular formulae C_{12}H_{2}Br_{3}O_{2} and C_{13}H_{2}Br_{3}O_{2} are indicated. (b) MS1 extracted ion chromatogram (EIC) corresponding to the most abundant ion for molecular formula C_{12}H_{2}Br_{3}O_{2} (m/z = 500.69, 10 ppm tolerance) identifies a set of three distinct isomeric product peaks. Chemical structure for 1 is shown on the right. (e) MS1 EIC corresponding to the most abundant ion for molecular formula C_{13}H_{2}Br_{3}O_{2} (m/z = 420.79, 10 ppm tolerance) identifies a second set of three distinct isomeric product peaks. The chemical structure for 10 is shown on the right. (d) EIC corresponding to MS2 product ion [M-Br]^− from C_{12}H_{2}Br_{3}O_{2} (m/z = 421.78, 10 ppm tolerance). (e) EIC corresponding to MS2 product ion [M-2Br]^− from C_{12}H_{2}Br_{3}O_{2} (m/z = 339.85, 10 ppm tolerance). (f) EIC corresponding to MS2 product ion [M-Br]^− from C_{13}H_{2}Br_{3}O_{2} (m/z = 340.86, 10 ppm tolerance). (g) EIC corresponding to dibromobenzoquinone MS2 product ion generated from C_{12}H_{2}Br_{3}O_{2} (m/z = 265.84, 10 ppm tolerance). Chemical structure for 11 with the loss of dibromobenzene (in red) is shown on the right. (h) EIC corresponding to bromobenzoquinone MS2 product ion generated from C_{12}H_{2}Br_{3}O_{2} (m/z = 185.93, 10 ppm tolerance). Chemical structure for 12 with the loss of dibromobenzene (in red) is shown on the right. Note that para-OH-BDEs (11–12) possess a MS/MS signature typified the detection of a bromobenzoquinone MS fragment ion, while ortho-OH-BDEs (13) are not fragmented.
Supplementary Fig. 21. $^1$H NMR spectrum for 11.

11: $^1$H NMR (600 MHz, DMSO) $\delta$ 7.95 (d, $J = 2.4$ Hz, 1H), 7.56 (dd, $J = 8.8, 2.3$ Hz, 1H), 7.25 (s, 2H), 7.00 (d, $J = 8.7$ Hz, 1H).
Supplementary Fig. 22. Identification of 2′-OH-BDE-68 (13) as a reaction product generated by Bmp7. Bmp7 produces three isomeric products of molecular formula C_{12}H_{6}Br_{x}O_{2}. Products 1 and 11 had been identified by comparison to authentic standard for 1 (Supplementary Fig. 19), and comprehensive structural characterization by NMR spectroscopy for 11 (Supplementary Note 2 and Supplementary Fig. 21). Extract of in vitro reaction of Bmp7 with 4 (red trace) was supplemented with a synthetic standard of 13 in increments of 0.25 µg each. Note that of the three isomeric peaks with C_{12}H_{6}Br_{x}O_{2} molecular formula, supplementation with authentic standard identifies the third peak with retention time ~ 27.8 min as 13.
Supplementary Fig. 23. HPLC detection of products generated by Bmp7 using 5 as a substrate. Curve in black represents a negative control reaction in which Bmp7 was omitted. Only a single brominated product was identified (red curve). Note that the substrate (5) has only been partially consumed.
Supplementary Fig. 24. MS and MS/MS profiles for 14, produced enzymatically by reaction of Bmp7 with 5. (a) MS1 isotope distribution for 14. The mass and isotope distribution corresponds to the molecular formula C_{12}H_{18}Br_{2}O_{2}. (b) MS2 ions corresponding to the scission of ether bond in the structure of 14. Two distinct MS2 ions can be detected, corresponding to dibromobenzoquinone and tribromobenzene as shown. Detection of the dibromobenzoquinone MS2 ion is consistent with 14 being a para-OH-BDE^8, as discussed in the Supplementary Note 2 and Supplementary Fig. 20. The final deduced structure of 14 is shown with the appropriate fragment ions.
Supplementary Fig. 25. Mass spectrometric analysis of the biphenyl and hydroxylated diphenyl ether products formed by Bmp7 reaction with 17. (a) Total ion chromatogram (TIC) for the Bmp7 reaction extract. Peaks for the three isomeric major products corresponding to the molecular formulae C_{12}H_{10}ClO_2 are indicated. (b) MS1 extracted ion chromatogram (EIC) corresponding to the most abundant ion for molecular formula C_{12}H_{10}ClO_2 (m/z = 322.90, 10 ppm tolerance) identifies a set of three distinct isomeric product peaks. The postulated chemical structure for 18 is shown on the right. (c) MS2 EIC corresponding to [M-Cl]^- (m/z = 286.92, 10 ppm tolerance). (d) MS2 EIC corresponding to [M-2Cl]^- (m/z = 250.95, 10 ppm tolerance). (e) MS2 EIC corresponding to [M-3Cl]^- (m/z = 222.95, 10 ppm tolerance). (f) MS2 EIC corresponding to dichlorobenzoquinone (m/z = 175.94, 10 ppm tolerance). The postulated chemical structure for 19 with the loss of dichlorobenzene (in red) is shown on the right. Note that of the three isomers, one isomers demonstrated the successive loss of chlorine atoms, leading to its postulated structure assignment as a polychlorinated biphenyl analogous to 1. One of the isomers demonstrates a dichlorobenzoquinone MS2 product ion, leading to its postulated structure assignment as a para-hydroxyl diphenyl ether, analogous to 11. The third isomer, with similar retention time to the para-hydroxy diphenyl ether, and not demonstrating any MS2 fragment ions is postulated to be an ortho-hydroxy diphenyl ether, analogous to 13.
Supplementary Fig. 26. Characterization of 3 generated by Bmp7 in an in vitro reaction with 4 and 6 as substrates. Extract of Bmp7 reaction (black curves) was compared to an authentic NMR characterized sample of 3 (red curves) isolated from P. luteoviolacea 2ta16. (a, b) Comparison of retention times. MS1 EIC corresponding to the most abundant isotope for the molecular formula C_{19}H_{15}Br_5NO (m/z = 551.60, 10 ppm tolerance) for the Bmp7 reaction extract (a) and authentic standard of 3 (b) analyzed under identical chromatographic conditions. (c, d) Comparison of MS1 profiles. MS1 mass spectrum for species at retention time 25.2 min from the Bmp7 reaction extract (c) and authentic standard of 3 (d). (e, f) Comparison of MS/MS profiles. MS2 profile of the product ions generated from most abundant isotope MS1 ions corresponding to the molecular formula C_{19}H_{15}Br_5NO at identical retention times from the Bmp7 reaction extract (e) and authentic standard of 3 (f). Most abundant MS2 ions corresponded to the [M-Br]^+. Identical retention times, MS1 and MS/MS profiles as compared to a NMR characterized authentic standard lead to the characterization of the Bmp7 reaction product as 3.
NMR spectrum for 3 purified from preparative scale cultures of *P. luteoviolacea* 2ta16 and used as authentic standard in panels b, d and f above:

3: $^1$H NMR: (CDCl$_3$) δ [ppm] 9.54 (bs, 1H), 8.09 (d, $J = 2.4$ Hz, 1H), 7.58 (d, $J = 2.4$ Hz, 1H), 6.10 (bs, 1H)
Supplementary Fig. 27. Characterization of 15 generated by Bmp7 in an in vitro reaction with 4 and 6 as substrates. Extract of Bmp7 reaction (black curves) was compared to an authentic NMR characterized sample of 15 (red curves) isolated from P. luteoviolacea 2ta16. (a, b) Comparison of retention times. MS1 EIC corresponding to the most abundant isotope for the molecular formula C_{10}H_{17}Br_{4}NO (m/z = 471.68, 10 ppm tolerance) for the Bmp7 reaction extract (a) and authentic standard of 15 (b) analyzed under identical chromatographic conditions. (c, d) Comparison of MS1 profiles. MS1 mass spectrum for species at retention time 29.8 min from the Bmp7 reaction extract (c) and authentic standard of 15 (d). (e, f) Comparison of MS/MS profiles. MS2 profile of the product ions generated from most abundant isotope MS1 ions corresponding to the molecular formula C_{10}H_{17}Br_{4}NO at identical retention times from the Bmp7 reaction extract (e) and authentic standard of 15 (f). Most abundant MS2 ions corresponded to the [M-Br']'. Identical retention times, MS1 and MS/MS profiles as compared to a NMR characterized authentic standard lead to the characterization of the Bmp7 reaction product as 15.
$^1$H NMR spectrum for 15 purified from preparative scale cultures of *P. luteoviolacea* 2ta16 and used as authentic standard in panels b, d and f above:

15: $^1$H NMR (CDCl$_3$) $\delta$ [ppm] 8.79 (bs, 1H), 7.57 (d, $J = 1.8$ Hz, 1H), 7.54 (d, $J = 1.8$ Hz, 1H)
Supplementary Fig. 28. Production of iodinated bromophenols by heterologous host E. coli expressing bmp1–8 grown in the presence of KI. When E. coli expressing bmp1–8 was grown in the presence of 1 g/L KI (no KBr was present), we could detect MS1 signals corresponding to the molecular formulae C_{12}H_{6}I_{4}O_{2} (top) (expected: 688.6474, found: 688.6472) and C_{12}H_{7}I_{3}O_{2} (bottom) (expected: 562.7507, found: 562.7508). The molecular formula C_{12}H_{6}I_{4}O_{2} corresponds to the iodinated analog of 1 (putative structure shown in top
panel), while C$_{12}$H$_7$I$_2$O$_2$ corresponds to the iodinated analog of 10 (putative structure shown in bottom panel). This result implies that Bmp5 could indeed accept iodide for incorporation to 4-HBA. Together with the dimerization of 17 as shown in Supplementary Fig. 25, this result demonstrates that Bmp7 is non-specific for the identity of halogens atoms present on the phenol ring to catalyze their coupling. Also note that during this experiment, 10 g/L chloride was present in the *E. coli* growth media. However, we could not detect the formation of any chlorinated phenol monomers or dimers.
Supplementary Fig. 29. Mass spectrometry based characterization of a polybrominated phenol-pyrrole ether product generated by Bmp7 in an in vitro reaction with 4 and 6 as substrates, which is also detected in the culture extracts of *P. luteoviolacea* 2ta16. Extract of Bmp7 reaction (black curves) was compared to an extract of *P. luteoviolacea* 2ta16 (red curves). (a, b) Comparison of retention times. MS1 EIC corresponding to the most abundant isotope for the molecular formula C_{10}H_{13}Br_{5}NO (m/z = 551.60, 10 ppm tolerance) for the Bmp7 reaction extract (a) and a MS2 EIC corresponding to m/z = 314.78, 10 ppm tolerance (vide infra) for an extract of *P. luteoviolacea* 2ta16 (b) analyzed under identical chromatographic conditions. Note that in (a), two isomeric molecules corresponding to the molecular formula C_{10}H_{13}Br_{5}NO could be detected. Peak denoted by corresponds to 3, as shown in Supplementary Fig. 26. The peak corresponding to is an isomer of 3. (c, d) Comparison of MS1 profiles. MS1 mass spectrum for species at retention time 27.9 min from the Bmp7 reaction extract (c) and an extract of *P. luteoviolacea* 2ta16 (d). (e, f) Comparison of MS/MS profiles. MS2 profile of the product ions generated from most abundant isotope MS1 ions corresponding to the molecular formula C_{10}H_{13}Br_{5}NO at identical retention times.
from the Bmp7 reaction extract (e) and an extract of *P. luteoviolacea* 2ta16 (f). Most abundant MS2 ions (cluster around m/z = 314.78) corresponded to a hydroxytribromopyrrole moiety (shown in a dashed box in (e)), while a [M-Br] MS2 product ion can also be detected. Detection of a hydroxytribromopyrrole MS2 ion leads us to postulate that the structure of this molecule could correspond to a polybrominated phenol-pyrole ether as shown in (a).
Supplementary Fig. 30. Primary sequence analysis identifies Bmp5 to be homologous to flavin-dependent monooxygenases. Primary sequence of Bmp5 was provided as an input to the NCBI BLAST program and analyzed against the ‘Non-reduant protein sequence (nr)’ dataset. All homologs of Bmp5 identified using this analysis (apart P. phenolica O-BC30 and M. mediterranea MMB-1, as described in this study) were annotated as flavin-dependent dimethylamine monooxygenases. However, as most of the sequences returned by the BLAST search were also annotated as hypothetical or predicted proteins with uncharacterized biochemical activities, a second BLAST search was performed against the ‘Protein Data Bank (pdb)’ database. The results of this search were more definitive, and are discussed here. The primary structural homologs for Bmp5 that had the most primary sequence coverage identified were:

1. PDB: 2GV8 and related structures. Flavin-dependent monooxygenase from *Schizosaccharomyces pombe*\(^4\).
2. PDB: 2XVI and related structures. Flavin-dependent monooxygenase from *Methylophaga aminisulfidivorans*.

Further analysis reveals that the for both 2GV8 and 2XVI, homology to Bmp5 is divided into two distinct regions. The first region comprises of residues from the N-terminus of Bmp5 to residue 200 (identity/similarity = 28%/43% to 2GV8 and 27%/48% to 2XVI; henceforth referred to as *Motif 1*). The second region comprises of residues 315 to 426 (identity/similarity = 25%/47% to 2GV8 and 29%/46% to 2XVI; henceforth referred to as *Motif 2*). Upon mapping both regions back on to the structure 2GV8 (shown in cartoon representation above), it is evident that *Motif 1* and *Motif 2* (colored red in cartoon representation) comprise of all core secondary structural elements constituting the flavin binding and the nicotinamide binding domains. This includes the flavin binding GAGLSG sequence motif (colored blue), and the nicotinamide binding GLGESAD sequence motif (colored blue) identified in the primary sequence of Bmp5. Hence, the core structural elements in the structure of 2GV8 for both flavin binding, as well as nicotinamide binding can be identified in the primary sequence for Bmp5.

It is also noteworthy that the BLAST search for the primary sequence of Bmp5 against the PDB database did not return any hits for flavin-dependent halogenases, either of the PrnA/RebH/PyrH (and an unpublished crystal structure PDB: 2PYX) type that utilize free substrates, or of the CmlS/CndH type that utilize substrates tethered to acyl carrier proteins. On the flip side, an identical BLAST search against the PDB database for the primary sequence of Bmp2 returns homology to CndH (31%/47%), CmlS (30%/45%) and an unpublished crystal structure PDB: 3NIX (28%/44%). A similar result is obtained for the following halogenases: chlorinase PIIA (CmlS: 30%/45%; CndH: 28%/46%; 3NIX: 27%/45%), chlorinase SgcC3 (CmlS: 36%/52%; CndH: 53%/66%; 3NIX: 32%/44%), chlorinase Mpy16 (CmlS: 30%/43%; CndH: 32%/49%; 3NIX: 26%/44%), chlorinase ClzS (CmlS: 30%/43%; CndH: 32%/49%; 3NIX: 26%/44%), chlorinase Pyr20 (CmlS: 31%/45%; CndH: 30%/47%; 3NIX: 28%/44%), among other flavin-dependent halogenases that have been described in literature. Noteworthy here is the chlorinase ChlA. Activity for ChlA was demonstrated for a substrate molecule not tethered to an ACP molecule. However, ChlA demonstrates greatest homology to halogenases requiring ACP tethered substrates (CmlS: 25%/40%; CndH: 20%/38%; 3NIX was not returned as a homolog by BLAST). Expectedly, tryptophan chlorinases KizQ and KizR display homology to PrnA/RebH/PyrH.
Supplementary Tables

Supplementary Table 1. High-resolution MS used for determination of molecular formulae of polybrominated metabolites from extract of *P. luteoviolacea* 2ta16.

<table>
<thead>
<tr>
<th>Molecule</th>
<th>Molecular Formula</th>
<th>[M-H]^- calculated (m/z)</th>
<th>[M-H]^- measured (m/z)</th>
<th>Error (mmu)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>C_{12}H_{18}Br_{2}O_{2}</td>
<td>496.7029</td>
<td>496.7029</td>
<td>0.0</td>
</tr>
<tr>
<td>2</td>
<td>C_{12}H_{18}Br_{2}N_{2}</td>
<td>598.5245</td>
<td>598.5244</td>
<td>0.1</td>
</tr>
<tr>
<td>3</td>
<td>C_{10}H_{18}Br_{5}NO</td>
<td>547.637</td>
<td>547.6137</td>
<td>0.0</td>
</tr>
<tr>
<td>4</td>
<td>C_{12}H_{18}Br_{2}O</td>
<td>248.8556</td>
<td>248.8554</td>
<td>0.2</td>
</tr>
<tr>
<td>5</td>
<td>C_{12}H_{18}Br_{2}O</td>
<td>326.7661</td>
<td>326.7660</td>
<td>0.1</td>
</tr>
<tr>
<td>6</td>
<td>C_{12}H_{18}Br_{3}N</td>
<td>299.7665</td>
<td>299.7663</td>
<td>0.2</td>
</tr>
<tr>
<td>7</td>
<td>C_{12}H_{18}Br_{3}N</td>
<td>377.6770</td>
<td>377.6765</td>
<td>0.5</td>
</tr>
<tr>
<td>9</td>
<td>C_{12}H_{18}Br_{3}O_{2}</td>
<td>418.7923</td>
<td>418.7922</td>
<td>0.1</td>
</tr>
<tr>
<td>14</td>
<td>C_{10}H_{18}Br_{5}NO</td>
<td>467.6875</td>
<td>467.6874</td>
<td>0.1</td>
</tr>
<tr>
<td>Gene</td>
<td>Locus tag</td>
<td>Size [aa]</td>
<td>Proposed function</td>
<td>Homolog (% Similarity/% Identity) <em>functionally characterized</em></td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------</td>
<td>-----------</td>
<td>------------------------------------------------</td>
<td>---------------------------------------------------------------</td>
</tr>
<tr>
<td>bmp1</td>
<td>Pl2ta16_01235</td>
<td>77 (1-77)</td>
<td>Prolyl acyl carrier protein (N-terminal domain)</td>
<td>Pttl, peptidyl carrier protein [Pseudomonas protegens Pf-5] (60/34)*</td>
</tr>
<tr>
<td>bmp1</td>
<td>Pl2ta16_01235</td>
<td>280 (78-358)</td>
<td>Thioesterase (C-terminal domain)</td>
<td>α/β-hydrolase fold protein [Krypdia tusciae DSM 2912] (47/28)</td>
</tr>
<tr>
<td>bmp2</td>
<td>Pl2ta16_01236</td>
<td>406</td>
<td>Pyrrole halogenase</td>
<td>PtlA, FADH2-dependent halogenase [Pseudomonas protegens Pf-5] (44/29)*</td>
</tr>
<tr>
<td>bmp3</td>
<td>Pl2ta16_01237</td>
<td>380</td>
<td>Proline dehydrogenase</td>
<td>PtlE, acyl-CoA dehydrogenase [Pseudomonas protegens Pf-5] (62/45)*</td>
</tr>
<tr>
<td>bmp4</td>
<td>Pl2ta16_01238</td>
<td>518</td>
<td>L-Proline adenyl transferase</td>
<td>PtlF, prollyl-AMP ligase [Pseudomonas protegens Pf-5] (55/40)*</td>
</tr>
<tr>
<td>bmp5</td>
<td>Pl2ta16_01239</td>
<td>515</td>
<td>Flavin-dependent oxygenase</td>
<td>Dimethylnline monoxygenase (N-Oxide forming) [Myotis brandii] (50/33)</td>
</tr>
<tr>
<td>bmp6</td>
<td>Pl2ta16_01240</td>
<td>202</td>
<td>Chorismate lyase</td>
<td>4-hydroxybenzoate synthetase [Beggiaota alba] (60/41)</td>
</tr>
<tr>
<td>bmp7</td>
<td>Pl2ta16_01241</td>
<td>491</td>
<td>Cytochrome P450</td>
<td>Cytochrome P450 [Phalacrocorax carbo] (50/30)</td>
</tr>
<tr>
<td>bmp8</td>
<td>Pl2ta16_01234</td>
<td>196</td>
<td>Carboxymuconolactone decarboxylase</td>
<td>Carboxymuconolactone decarboxylase [Blastococcus saxobsidens DD2] (58/42)</td>
</tr>
<tr>
<td>bmp9</td>
<td>Pl2ta16_01233</td>
<td>106</td>
<td>Ferredoxin</td>
<td>Ferredoxin [Idiomarina lothiensis L21R] (73/52)</td>
</tr>
<tr>
<td>bmp10</td>
<td>Pl2ta16_01232</td>
<td>415</td>
<td>Ferredoxin reductase</td>
<td>Ferredoxin reductase [Pseudomonas sp. 19-rlm] (70/53)</td>
</tr>
</tbody>
</table>
Supplementary Table 3. Primers used to generate overlapping fragments for yeast assembly of yeast/E. coli shuttle vector pGAL-bmp1-7. All primer sequences are listed from 5’ to 3’.

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Primer Sequence</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>pGALbmp1F</td>
<td>AAGGTAACAGGTAAGCTCGGTAATAACTG ATATAATTTAATTGAG</td>
<td>pGAL-MF region1</td>
</tr>
<tr>
<td>pGALbmp1F</td>
<td>TGCAGTCTCTTGAATAAATTCACTGACTAGTGACGGTC TGTTAA</td>
<td>pGAL-MF region1</td>
</tr>
<tr>
<td>pGALbmp2F</td>
<td>TTAACGGGACTACAGTGCACAAAATTCATCAAGA GACTGCA</td>
<td>pGAL-MF region2</td>
</tr>
<tr>
<td>pGALbmp2F</td>
<td>CTAGTCTGGCAATAATTGTTAAACTAGTGACTGCCTCGGTC</td>
<td>pGAL-MF region2</td>
</tr>
<tr>
<td>pGALbmp3F</td>
<td>ACCGAGCGACGAGGACTAGTGTTCAACAATA TTGCGAGACTAGA</td>
<td>bmp cluster region1</td>
</tr>
<tr>
<td>pGALbmp3F</td>
<td>TGGGCCAGGTAACCCAATTAGCATCAGTCCGTTGAGCATA</td>
<td>bmp cluster region1</td>
</tr>
<tr>
<td>pGALbmp4F</td>
<td>GTATGCCACTACAGAATGCTAAATTCGCTACATCTGGCCA</td>
<td>bmp cluster region2</td>
</tr>
<tr>
<td>pGALbmp4F</td>
<td>TGATCTTGTACTCAGTGGCAATGTTCGTA TGTCAGTGACCCAA</td>
<td>bmp cluster region2</td>
</tr>
<tr>
<td>pGALbmp5F</td>
<td>GTACGATACGAAAACATGGCAAACTGAGTACA TAGTCACA</td>
<td>bmp cluster region3</td>
</tr>
<tr>
<td>pGALbmp5F</td>
<td>AATTAATTATACAGTTATTACCGAGCTCTAC TGTTCGTTACGGGT</td>
<td>bmp cluster region3</td>
</tr>
</tbody>
</table>
Supplementary Table 4. Primers used to generate gene deletions from pETDuet-\textit{bmp}.
All primer sequences are listed from 5' to 3'.

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Primer Sequence</th>
<th>ORF deleted</th>
</tr>
</thead>
<tbody>
<tr>
<td>bmp2KO FWD</td>
<td>CATGGCATGAAGGAAAAGCACCTGCGCTTTTTCTCTTCTTCTTCTGGAAGGATTGTAGTATTG</td>
<td>bmp2</td>
</tr>
<tr>
<td>bmp2KO REV</td>
<td>TGCTTTTCCCTCTCATGCCCAGTCCACCTAGTTAAGAGTCTTCAACAAACCCACAC</td>
<td></td>
</tr>
<tr>
<td>bmp5KO FWD</td>
<td>TGGTGCGTAAATAGTACGAGAAGGACATGATTGCATTGACTTCAATGAAAGCCTGT</td>
<td>bmp5</td>
</tr>
<tr>
<td>bmp5KO REV</td>
<td>CTCGTAATCTTTACGCACCACTTTGCACAGTATTACCTCCGTGGGCTGGGAGGT</td>
<td></td>
</tr>
<tr>
<td>bmp7-XmaF</td>
<td>ACACCAGAATTATCAACCTTTAGTGCACTGGATTGATTGACCCGCGGTACCGGTAACTGATGCGC</td>
<td>bmp7</td>
</tr>
<tr>
<td>bmp7-XmaR</td>
<td>TAAGGATGAGTGGCAAAAACCATCCTAAAGTGGAGTGGCCTTCCCCTGGGAACCTATGAGCTCGAGCCA</td>
<td></td>
</tr>
</tbody>
</table>
Supplementary Table 5. Co-transformations of *E. coli* BL21(DE3) for heterologous expression of *bmp* cluster and derivatives.

<table>
<thead>
<tr>
<th>Strain name</th>
<th>Co-transformed plasmid 1</th>
<th>Co-transformed plasmid 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>bmp1–8</td>
<td>pETDuet-bmp1–7</td>
<td>pCOLADuet-bmp8-ppT</td>
</tr>
<tr>
<td>Abmp2</td>
<td>pETDuet-bmp1–7:Abmp2</td>
<td>pCOLADuet-bmp8-ppT</td>
</tr>
<tr>
<td>Abmp5</td>
<td>pETDuet-bmp1–7:Abmp5</td>
<td>pCOLADuet-bmp8-ppT</td>
</tr>
<tr>
<td>Abmp7</td>
<td>pETDuet-bmp1–7:Abmp7</td>
<td>pCOLADuet-bmp8-ppT</td>
</tr>
<tr>
<td>Abmp8</td>
<td>pETDuet-bmp1–7</td>
<td>pCOLADuet-ppT</td>
</tr>
</tbody>
</table>
SUPPLEMENTARY REFERENCES


Chapter 2, in full, is a reprint of materials as it appears in “Biosynthesis of polybrominated aromatic molecules by marine bacteria” in Nature Chemical Biology, 2014, Agarwal V., El Gamal A., Yamanaka K., Poth D., Kersten R. D., Schorn M., Allen E. E., and Moore, B. S., The dissertation author was one of two equally contributing primary investigators and authors of this paper.

V.A., A.A.E., E.E.A. and B.S.M. designed research; V.A., A.A.E. and K.Y. performed genetic experiments; V.A., A.A.E. and R.D.K. performed in vitro experiments; M.S. and E.E.A. generated sequencing data; D.P. contributed new analytical reagents; and V.A., A.A.E., E.E.A. and B.S.M. analyzed data and wrote the manuscript.
Chapter 3: Biosynthesis of coral settlement cue
tetrabromopyrrole in marine bacteria by a uniquely
adapted brominase-thioesterase enzyme pair
Biosynthesis of coral settlement cue tetrabromopyrrole in marine bacteria by a uniquely adapted brominase–thioesterase enzyme pair

Abraham El Gamal1,8, Vinayak Agarwal1,8, Stefan Dieleman1,8, Imran Rahman1,8, Michelle A. Schorn1,8, Jennifer M. Sneed1,8, Gordon V. Louie1,11, Kristen E. Whalen1,11, Tracy J. Mincer1,8, Joseph P. Noel1,11, Valerie J. Paul1,8, and Bradley S. Moore10,11,8

*Center for Ocean and Human Health, Unisys Institution of Oceanography, University of California, San Diego, La Jolla, CA 92093; **National Marine Station at Fort Pierce, Fort Pierce, FL 34949; ***Howard Hughes Medical Institute, The Salk Institute for Biological Studies, La Jolla, CA 92037; **Jack H. Skaggs Center for Chemical Biology and Proteomics, The Salk Institute for Biological Studies, La Jolla, CA 92037; *Marine Chemistry and Geochemistry, Wood Hole Oceanographic Institution, Woods Hole, MA 02543; and *Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, CA 92093

Edited by Michael A. Fischbach, University of California, San Francisco, CA, and accepted by the Editorial Board February 24, 2016 (received for review October 4, 2015)

Halogenated pyrroles (halopyrroles) are common chemical moieties found in biologically active natural products. The halopyrrole moieties of mono- and dihalopyrrole-containing compounds arise from a conserved mechanism in which a proline-derived pyrrolyl group bound to a carrier protein is first halogenated and then elaborated by peptidyl or polyketide extensions. This paradigm is broken during the marine Pseudomonas monodon bacterial biosynthesis of the coral larval settlement cue tetrabromopyrrole (T), which arises from the substitution of the proline-derived carboxylate by a bromine atom. To understand the molecular basis for decarboxylative bromination in the biosynthesis of T, we sequenced two Pseudomonas genomes and identified a conserved four-gene locus encoding the enzymes involved in its complete biosynthesis. Through total in vitro reconstitution of the biosynthesis of T using purified enzymes and biochemical interrogation of individual biochemical steps, we show that all four bromine atoms in T are installed by the action of a single flavin-dependent halogenase: Bmp2. Tetrahaloboration of the pyrrole induces a thioesterase-mediated offloading reaction from the carrier protein and activates the biosynthetic intermediate for decarboxylation. Insights into the tetrahaloboration activity of Bmp2 were obtained from the high-resolution crystal structure of the halogenase contrasted against structurally homologous halogenase Mpy16 that forms only a dihalogenated pyrrole in marinopyrrole biosynthesis. Structure-guided mutagenesis of the proposed substrate-binding pocket of Bmp2 led to a reduction in the degree of halogenation catalyzed. Our study provides a biogenetic basis for the biosynthesis of T and sets a firm foundation for querying the biosynthetic potential for the production of T in marine (met)genomes.

Marine bacteria of the genus Pseudomonas produce numerous small molecule natural products with varied roles in marine chemical ecology (1–4). Recently, tetrabromopyrrole (1, Fig. 1) was established as a chemical cue produced by Pseudononas that induces larval settlement and metamorphosis in reef-building Caribbean coral species (3). Additionally, the chemical structure of 1 is notable for its degree of halogenation (one-to-one carbon-to-halogen ratio), unique among naturally occurring aromatic organic halogehnes (5).

Halopyrrole-containing microbial natural products are biosynthesized starting from L-proline. A highly conserved halopyrrole biosynthetic route involves oxidation of an acyl carrier protein (ACP)-loaded prolyl side chain and subsequent halogenation by a flavin-dependent halogenase (Fig. 1) (6). At this stage, two biosynthetic routes are possible. In the first route, the ACP-loaded halogenase is extended using modular assembly lines to yield nonribosomally synthesized peptide, or polyketide natural products (e.g., 2–4, Fig. 1) (7–11). During the biosynthesis of 2–4, and other related natural products, the halopyrrole carboxylic acid is transhalogenated to downstream ACPs during modular elongation reactions. Molecular scaffolds from these modular biosynthetic pipelines are characterized by the preservation of the prolyl alpha-carbon-carbon atom as a carbonyl, or embedded in the final natural product skeleton in varying oxidation states (in red, Fig. 1). In a recently discovered alternate route, the prolyl alpha-carbon-carbon atom is lost during the biosynthesis of 1 and the cytotoxic marine bacterial natural product pentabromopyrrolequin (5, Fig. 1) (4, 11). Uniquely, 5 is biosynthesized via the coupling of 2,3,4-trihalopyrrole (6, Fig. 1) to 2,4-dibromophenol rather than via a modular assembly line. The molecular mechanism for the pyrrole offloading from the ACP and elimination of prolyl alpha-carbon-carbon en route to the biosynthesis of 1 and 5–6 has not been determined.

The structure of 1 poses an additional biosynthetic challenge due to the presence of a bromine atom in place of an acyl side chain. Although all previously described pyrrole halogenases natural products | biosynthesis | halogenation | enzyamochemistry

Significance
This paper is the result of a collaborative effort among scientists at the University of California, San Diego, and researchers from the National University of Singapore. The authors declare no conflict of interest.
catalyze one, two, or three halogen additions upon the ACP-loaded pyrrole ring, a halogenase capable of four halogenations on the pyrrole ring, as implied by the structure of I, has not been characterized. These open biosynthetic questions, together with the ecological significance of I in coral larval settlement, motivated us to explore the genetic and molecular logic for its biosynthesis in marine bacteria.

Herein, we establish the biosynthesis of I by the total in vitro enzymatic reconstitution of its biosynthetic machinery. We show that a single flavin-dependent brominase installs an unprecedented four halogenations on an ACP-bound pyrrole required for the progression of diastereoselective TE-mediated offloading and decarboxylation reactions. We also exploit an opportunity to structurally characterize and investigate via site-directed mutagenesis the molecular basis for a differential number of halogen additions catalyzed by flavin-dependent halogenases on aromatic substrates. Our structural comparison of a highly homologous pyrrole diohalogenase to the pyrrole tetrahalogenase reveals subtle variations in bioisosterism design that lead to divergent molecular outcomes.

Results
Genetic Basis for the Biosynthesis of I. We recently reported that I produced by the marine bacterium Pseudoalteromonas sp. PSS induces settlement of larval species with several Caribbean coral species. To investigate the genetic basis for the biosynthesis of I, we sequenced and assembled a 5.03-Mbp draft genome for P. sp. PSS. Querying the draft genome for the presence of the halogenase biosynthetic genes, we identified a gene locus in P. sp. PSS with high homology to a subset of bromopyrrole biosynthetic genes present in the previously reported bnap gene locus from marine bacteria of the genera Pseudoalteromonas and Marinomonas. The organization of the bnap homologs identified in the genome of P. sp. PSS (PSS_bnap) is identical to that of the bnap gene cluster found in Marinomonas mediterranea MMB-1 bnap (MMB-bnap) and others (11). Indeed, the PSS_bnap gene locus maintains the bromopyrrole biosynthetic module Mmbm-bnap-4 and the bromopyrrole/phenol coupling cytochrome P450 (CYP450) accessory genes Mmbm-bnap-10, but lacks the genes associated with bromophenol biosynthesis (Mmbm-bnap-5-6) and the CYP450 Mmbm-bnap7 (Fig. 2A). In addition, we also sequenced and assembled a 5.13-Mbp draft genome for Pseudoalteromonas sp. A757 that also produces I (12). Querying the genome of P. sp. PSS as before revealed a gene cluster with high homology to PSS_bnap (A757_bnap), providing a second example of a stand-alone bromopyrrole biosynthetic pathway from the genus Pseudoalteromonas (Fig. 2A and SI Appendix, Table S1).

To evaluate the role of PSS_bnap-4 genes in the production of I, we cloned and heterologously expressed PSS_bnap-4 in Escherichia coli. Only in the presence of bromide in the culture media, we observed robust heterologous production of I along with a minor amount of 6, consistent with the product profile observed in organic extracts of cultures of P. sp. PSS and P. sp. A757 (Fig. 2B). Although these results demonstrate that bnap-4 genes are necessary for the production of I, the molecular details underlying the offloading of the l-proline-derived pyrrole moiety from the ACP, its timing relative to the bromination events, and the chemical logic for the loss of l-proline derived alpha-carbonyl was not discernible at this stage. Therefore, we next examined the biosynthetic pathway by total in vitro enzymatic reconstitution of the production of I using purified enzyme catalysts.

Decarboxylative Bromination of Thioester-Linked Pyrrole. To examine the individual roles of Bmp4-4, we performed the total in vitro reconstitution of the biosynthesis of I (Fig. 3D). Due to the high sequence similarity of PSS_bnap-4 to Mmbm-bnap-4 (SI Appendix, Table S1), and production of I by both P. sp. PSS and M. mediterranea MMB-1 (11), we used recombinant Mmbm-bnap-4 proteins in our in vitro investigations. Bmp2 and Bmp4 were individually purified as N-His6-tagged proteins whereas N-His6-tagged Bmp1 was purified in complex with untagged Bmp2 and converted to its pantetheine-loaded holo-form by the promiscuous Bacillus subtilis phosphopantetheinyl transferase Sfp (11, 13) (Fig. 34). Incubation of l-proline and bromide with purified Bmp1-4 enzymes, along with flavin and NADP+ cofactors (FAD and NADP+) and cofactor regeneration enzymes (PufD and SsdE), led to the formation of I as the major product (Fig. 3 and SI Appendix, Fig. 2).

Fig. 2. A genetic basis for production of I. (A) Mmbm_bnap, PSS_bnap, and A757_bnap genes cluster with bromopyrrole and bromophenol biosynthetic genes indicated below the Mmbm_bnap gene cluster; a putative peroxidase (uncolored) is inserted between Mmbm_bnap and Mmbm_bnap4. Mmbm_bnap-4 are colored per their catalytic roles shown in Fig. 1. Note that bnap encodes a d-isoform protein with an ACP domain at the N terminus followed by a TE domain (11). (B) IONS extracted ion chromatograms (EIC) for [M+H]+ ions corresponding to 1 and 6 in organic extracts of P. sp. PSS, P. sp. A757, and E. coli expressing PSS_bnap-4 grown in media with (+) or without (−) bromide.
generated by the bromination of pyrrole-2-carboxylic acid, was sequentially ligated to cysteine and then to pantethenic acid to afford 3,4,5-tribromopyrrol-5-panethenine 9 (Fig. 4A and SI Appendix for synthesis protocols and product characterization data). Compound 9 was then extended by the E. coli CoA biosynthetic enzymes CoA, CoA, and CoA to generate 3,4,5-tribromopyrrol-5-panethenine 10, which was used as the substrate for the in situ transfer of the 3,4,5-tribromopyrrol-5-panethenine moiety to the serine side chain hydroxyl of apo-Bmp1(ACP) by Sfp to yield 8 (Fig. 4A and SI Appendix, Fig. S2). We then expressed and purified the Bmp1 thioesterase domain, Bmp1(TE), whose activity we previously confirmed using a model ester substrate, p-nitrophenylacetate (11). Incubation of 8 with Bmp2, Bmp1(TE), bromide, NADP+, FAD, and cofactor regeneration components led to formation of 1 as the major product, together with minor production of 6 and 7 (Fig. 4B and SI Appendix, Fig. S1). Exclusion of Bmp1(TE) from the reaction led to trace production of 1 (Fig. 4B). Additionally, we mapped the active site serine residue of Bmp1(TE) to Ser192 and confirmed that its mutation to alanine resulted in loss of esterase activity for the p-nitrophenylacetate substrate (SI Appendix, Fig. S3). Substitution of Bmp1(TE) with catalytically inactive Bmp1(TE)E520A mutant enzyme led to significantly reduced production of 1 (Fig. 4B and SI Appendix, Fig. S4). Finally, exclusion of Bmp2 from the reaction completely abolished the production of 1 (Fig. 4B and SI Appendix, Fig. S4). Similar levels of 7 were observed across all reactions, suggesting that it is likely an "off-pathway" product resulting from the hydrolysis of 8 (Fig. 4B). In support of 7 as a hydrolytic shunt product, no conversion of 7 to 1 or 6, or to any new products, was observed upon its incubation with Bmp2, Bmp1(TE), and bromide (SI Appendix, Fig. S5). These results provide two important findings. First, both catalytic Bmp1(TE) and the halogenase Bmp2, are required to convert the intermediate 8 to 1. Second, two enzymatic activities (i.e., hydrolysis via thioesterase Bmp1(TE) and bromination via Bmp2) lead to three chemical events: (i) offloading of the pyrrole from Bmp1(ACP), (ii) the fourth bromination on the 2-position of the pyrrole, and (iii) decarboxylation of the 1-proline-derived alpha-carboxylic acid carbon atom.

We next evaluated the timing of formation of 6 relative to the formation of 1. We thus incubated 8 with Bmp2, with and without Bmp1(TE), and cofactor-regenerating enzymes. We observed trace conversion of 6 to 1 only after prolonged reaction times, consistent with our hypothesis that 6 is a nonphysiological substrate for Bmp2 (SI Appendix, Fig. S3). Therefore, we propose that 6 might be a nonenzymatic reductive degradation byproduct of 1. Indeed, incubation of synthetically prepared 8 with NADPH led to the conversion of 6 to 1 at levels comparable with that in vitro.
reconstitution reactions (SI Appendix, Fig. S6). Together with our previous results, we propose that 7 and 6 are both off-pathway products arising, respectively, at stages preceding and after formation of 1 (Fig. 5).

In sum, our data support the Bmp2-catalyzed bromination of 8 leading to transient intermediate 1 as shown in Fig. 5, which would undergo transsterification to generate a Bmp1(TE) S202 side-chain-bound oxoester ii, thus offloading the pyrrole from Bmp1(ACP). Hydrolytic offloading from Bmp1(TE) would generate the stable alpha-brsmo acid iii. Driven by the rearrangement of the pyrrole ring, iii would spontaneously decarboxylate to I. Direct mass spectrometric evidence for I proved elusive, suggesting that ii rapidly reverts back to 8 via spontaneous debromination in the absence of Bmp1(TE). Therefore, the role of the Bmp1(TE) could be to trap the tetranoylated pyrrole species by means of an irreversible transsterification step whereas the fourth bromination by Bmp2 activates the substrate for decarboxylation subsequent to hydrolytic offloading from Bmp1(TE).

Nonenzymatic hydrolysis of I may also lead to release of iii to afford I, consistent with our observation of low levels of I from reactions excluding Bmp1(TE) and with inactive Bmp1(TE)S202A. The observation of similar levels of I in all reactions further supports direct hydrolytic bypass of TE-mediated offloading. Furthermore, the fourth bromination of 8 by Bmp2, followed by transsterification and decarboxylation of I, nearly differentiates the biosynthesis of I from nonribosomal peptide synthetase- and polyketide synthase-derived pathways involving halopentapeptides in which the δ-proline-derived alpha-carboxylate carbon atom is preserved in the mature natural product (Fig. 1).

Structural Basis for Unique Tetrahalogenating Activity of Bmp2. To realize the scheme shown in Fig. 5, the flavin-dependent brominase Bmp2 would need to halogenate all four positions of the pyrrole ring. This mechanistic requirement for the production of I is in stark contrast to all previously characterized pyrrole halogenases that halogenate the pyrrole ring only once or twice. To understand the structural basis for the tetrahalogenating activity of Bmp2, we determined the high-resolution crystal structure of FAD-bound, holo-Bmp2 at a limiting resolution of 1.87 Å (Fig. 6A) and compared it with the FAD-bound holo-structure of dihalogenase Mpy16 participating in the biosynthesis of 3 that we determined at 1.95 Å resolution (Fig. 6B and SI Appendix, Table S2) (8). Bmp2 and Mpy16 structures are highly homologous with all secondary structural elements in the vicinity of the active site strictly conserved between the two enzymes (SI Appendix, Fig. S7). Furthermore, the FAD cofactor isalloxazine rings and the postulated active site lysine residues (15) (S27 for Bmp2, K72 for Mpy16) (Fig. 6A and B) are superimposable. Although we could not determine a substrate pyrrolyl-S-ACP covalent structure for either Bmp2 or Mpy16, the position of the pyrrole binding sites for Bmp2 and Mpy16 was inferred by a structural alignment of Bmp2 and Mpy16 with the crystal structures of substrate-bound forms of the flavin-dependent tryptophan-7-chloromutase PmA (16) and RobH (17). Fittingly, the postulated pyrrole-binding site in Bmp2 and Mpy16 (Fig. 6A and B) is in close proximity to the side chain of the catalytic lysine residue. In contrast, the amino acid residue not considered to be within the active site, the tryptophan residue (19) (W97 for Bmp2, W101 for Mpy16) (Fig. 6A and B) is spatially far from the catalytic site.

Discussion

In addition to characterizing the biosynthesis of a microbially produced ecdaloid setonin cue, our study reveals previously unidentified enzymology stemming from a biosynthetic motif ubiquitous among halopentapeptide-containing natural products. Although our previous work had established trichromination of pyrrolyl-S-Bmp1(ACP) by Bmp2, the enzymatic activities responsible for the addition of the fourth bromine atom to I and offloading and decarboxylation of the pyrrolyl moiety from the ACP were unknown (11). Our current work demonstrates that a single halogenase, Bmp2, catalyzes the unprecedented tetrahalogenation of the ACP-bound pyrrole, which subsequently undergoes Bmp1(TE)-mediated offloading from the ACP and spontaneous decarboxylation. In light of the elucidation of the molecular and genetic details for the construction of I, we posit that I is an intermediate en route to the production of 6, reconciling a previous report implicating an ε-proline-derived symmetrical pyrrole intermediate in the biosynthetic scheme for the production of 6 (19). Previously, based on primary sequence homology, we postulated that the enzyme Bmp4 (Fig. 2A) participates in the decarboxylation of 7 to produce 6 (11), a hypothetical route that is likely not operable in light of the biochemical data presented above. Therefore, the physiological transformations for the production of 6 on route to 8 remains to be elucidated. Furthermore, the previously reported production of 1 during the CYF450-Bmp7-encoded binadial homodimerization of 6 is likely an off-pathway route, with Bmp1-4 being the primary players in the production of I. Successive
biochemical studies promise to address these open questions to characterize all steps in the biosynthesis of S and assign physiological roles for each of the Bmp enzymes.

The complete in vitro reconstitution of the production of I using purified enzyme catalysts suggests a mechanism for the acylphosphopantetheine thioester to be sequestered and thereby inaccessible to Bmp1(TE) until the terminal fourth bromination. Sequestration of a pyrrole tethered to a type II peptidyl carrier protein (PCP) was recently demonstrated for a highly homologous PCP-PtL (57% amino acid similarity to Ms_Bmp1(ACP)) participating in the biosynthesis of the dithiopyrrolo-containing natural product pyrovanillin (7, 20). By analogy, in the biosynthesis of I, it is possible that the terminal halogenation on the pyrrole partially liberates the sequestered substrate from the ACP, allowing access to the thioesterase. Therefore, the terminal fourth bromination by Bmp2 serves the dual role of triggering the release of the sequestered pyrrole moiety by presumptively making the thioester susceptible to hydrolysis, in addition to activating the substrate for elimination of the l-proline-derived alpha-carboxyl. Activating halogenation reactions in natural product biosynthetic pathways has been demonstrated in the biosynthesis of NRPS-PRK hybrid curacin A, in which a cryptic halogenation catalyzed by α-ketoglutarate–dependent halogenase–CoA promotes cyclopentane ring formation (21), and in the biosynthesis of meroterpenoid meroclorins A and B, in which a deaminatization/ketopine cyclization reaction cascade is initiated by a chlorination catalyzed by vanadium-dependent haloperoxidase-Mc24 (22). Most analogously, paralleling the decarboxylative bromination strategy used by Bmp2, flavin-dependent brominase–Bmp5 catalyzes a bromination reaction that drives dehydrobromination in the conversion of free p-hydroxybenzoic acid to 2,4-dibromophenol, the bromophenol building block of S (11).

The tetrahalogenation activity of Bmp2 is unprecedented among the flavin-dependent halogenases described to date that selectively add a specific number of halogen atoms to an aromatic substrate. In this study, we embraced the opportunity to interrogate the structural basis for this halogenation control in flavin-dependent pyrrolyl-S-ACP halogenases. Although the previously reported 2.1-A crystal structure of putative pyrrolyl-S-ACP flavin-dependent halogenase CndH highlighted the general differences in the architectures of flavin-dependent halogenases acting on free versus ACP-bound substrates (23), our study sheds mechanistic insight into the tuning of the enzyme active site in the context of confirmed biochemistry. Comparison of the putative substrate-binding cavities of the highly conserved structures of Bmp2 and Mpy16 resulted in a catalytically active site of Bmp2–TM exhibiting no perturbation of the bound FAD cofactor with respect to WT Bmp2, nor did it lead to any apparent change in halogen binding properties of the enzyme, as demonstrated by conservation of its specificity for bromide. Already, for flavin-dependent tryptophan halogenases, it has been demonstrated that amino acid side chains that constitute the substrate binding site control the regiochemical outcomes for halogen additions (15, 24, 25). Our findings extend...
this observation to flavin-dependent halogenases that catalyze halogenation of aromatic substrates asayed to ACCPs in demonstrating that side chains of residues lining the putative active site play a role in controlling substrate access, and potentially in specifying the positions on the pyrrole ring that are accessible to halogenation. Unfortunately, in the case of Mvl6, efforts to alter the putative substrate-binding cavity to resemble that of Bmp2 resulted in insolubility of the mutant Mvl6 enzymes. Nonetheless, our results demonstrate that the active sites of otherwise highly homologous flavin-dependent halogenases are uniquely evolved to afford distinct product profiles. Biological activities of natural products are influenced by the number of halogens decorating their organic scaffolds (26, 27), and it would seem that Nature has taken note.

In light of the existing literature and the findings of this study, several open questions remain regarding enzymatic halogenation (28). Of note is the question of the binding mode of both the ACP and the aromatic substrate by flavin-dependent halogenases that require aky-5-ACP substrates. In addition, the primary question is how different classes of halogenating enzymes is the structural determinant for halide specificity among these catalysts. Although Bmp2 and Mvl6 demonstrate different halide specificities, the amino acid side chains and the positioning of the FAD isoalloxazine ring relative to these side chains are remarkably conserved. Furthermore, the conservation of halide specificity in Bmp2-TM with respect to the WT enzyme raises the possibility that halide specificity in flavin-dependent halogenases is dictated not only by steric factors, such as halide ion radii, but also by the magnitude of the cationic propensity associated with desolvation of the halide ion in the active site of halogenase before 2° oxidation to the hypohalite halide. Furthermore, the contribution of the redox potential of the flavin isoalloxazine ring necessary to oxidize the halide, and the stability of the activated pyrimidine-intermediate generated en route to the transfer of the halonium to the aromatic substrate should be considered (15). As such, the currently proposed halide binding site for flavin-dependent halogenases is itself debatable (29), thus underscores the challenges associated with teasing apart the role of the enzymatic halogenase catalyst in each of the requisite mechanistic steps. Discovery and characterization of additional flavin-dependent halogenases promises to provide opportunities to answer these mechanistic questions, ultimately leading to engineeable tools to tailor the biosynthesis of halogenated natural products (30).

Materials and Methods
Detailed material and methods are provided in SI Appendix, Materials and Methods.

Database Deposition Information. The sequences for P. sp. PSS-5 and P. sp. A757-derived bim gene clusters have been deposited in GenBank under accession numbers KR119232 and KTB087878, respectively. Structures for Mvl6, Bmp2, and Bmp2-TM are deposited in the Protein Data Bank (www.rcsb.org) under accession numbers 1BLK, SBVA, and 1BLU, respectively.

Note. During the review and publication of this manuscript, the crystal structure of the flavin-dependent halogenase PsaA was reported in the literature (31).

ACKNOWLEDGMENTS. We thank our colleague B. M. Duggan at University of California, San Diego for assistance in acquiring HMR data. This work was jointly supported by US National Science Foundation Grant OCE-1021374 (to B.S.M.) and US National Institutes of Health (NIH Grants PO1 ES021821 and RO1- AI47178 [to B.S.M.] and R21-A1139131 [to K.E.W. and T.J.M.]) the Mote Protect Our Reef Grant Program (POR-2012-3), the Dart Foundation, the Smithsonian Competitive Grants Program for Science (J.P.), the Howard Hughes Medical Institute (J.P.), NIH Marine Biotechnology Training Grant Predoctoral Fellowship T32-GM07350 (to A.E.), a Helen Hay Whitney Foundation Postdoctoral Fellowship (to V.A.), and a Swede National Science Foundation Predoctoral Fellowship (to S.D.)
Supplementary information for

**Biosynthesis of coral settlement cue tetrabromopyrrole in marine bacteria by a uniquely adapted brominase-thioesterase enzyme pair**

Abraham El Gamal, Vinayak Agarwal, Stefan Diethelm, Imran Rahman, Michelle Schorn, Jennifer M. Sneed, Gordon V. Louie, Kristen E. Whalen, Tracy J. Mincer, Joseph P. Noel, Valerie J. Paul, and Bradley S. Moore

This supplementary information section contains

Materials & Methods

Supplementary Figures S1–S20

Supplementary Tables S1 & S2

Supplementary References
MATERIALS AND METHODS

Genome sequencing
Genomic DNA was isolated from 5 mL overnight liquid cultures of *P*. sp. PS5 and *P*. sp. A757 inoculated in 5 mL Difco 2216 Marine Broth (BD 212185) from -80 °C glycerol stocks and grown overnight at 30 °C with shaking at 200 rpm. Genomic DNA was isolated using a DNeasy Blood & Tissue Kit (Qiagen) following the spin column protocol for extraction of genomic DNA from gram-negative bacteria. Next generation sequencing libraries were constructed from genomic DNA of *P*. sp. PS5 and *P*. sp. A757 using the Ion Plus Fragment Library Kit (Thermo Fisher Scientific). Sample preparation was done with the Ion PGM 400 OT2 Kit (Thermo Fisher Scientific). The library was sequenced on an Ion Torrent PGM platform (Thermo Fisher Scientific) using the Ion PGM Hi-Q Sequencing Kit on an Ion Torrent 318v2 sequencing chip (Thermo Fisher Scientific).

Genome assembly, annotation and verification of Bmp genes

*De novo* assemblies of raw reads of *P*. sp. PS5 and *P*. sp. A757 were performed using CLC Genomics Workbench (Qiagen) and SPAdes (1) assemblers, respectively, using default parameters for Ion Torrent PGM sequencing. Annotation of the draft genomes of *P*. sp. PS5 and *P*. sp. A757 was performed using the RAST server (2). The draft genome was queried with translated sequences of *Mm Bmp1–10* using BLAST in the SEED Viewer genome browser interface. *PS5 bmp* and *A757 bmp* gene sequences were verified by Sanger sequencing (SeqXcel, La Jolla, CA).

Cloning and heterologous expression of *PS5_bmp1–4*

*PS5_bmp1–4* was PCR amplified from genomic DNA using PrimeStar Max DNA polymerase (Takara) standard protocol, and ligated into Sacl/NotIl (NEB) digested pETDuet (Novagen) using T4 DNA Ligase (NEB), followed by propagation in *E. coli* DH5α (NEB), Qiagen Miniprep plasmid extraction, and Sanger sequencing verification (SeqXcel, La Jolla, CA). An open reading frame encoding an ACP synthase (AcpS) from *M. mediterranea* (1) (Genbank accession code: NC_015276.1) was amplified from genomic DNA and cloned into Ndel/Xholl digested pRSFDuet (Novagen) using the same protocol as before. Thus constructed pETduet-*PS5_bmp1–4* and pRSFduet-ACP5 were co-transformed into *E. coli* BL21-Gold(DE3) (Agilent) for expression. Five mL LB containing 50 µg/mL kanamycin and 100 µg/mL carbenicillin to maintain plasmids with and without 1 g/L KBr (final concentration) were each inoculated with 100 µL of overnight culture of *E. coli* BL21-Gold(DE3)/pETDuet-*PS5_bmp1–4*/pRSF-ACP5 grown at 30 °C. The cultures were grown at 37°C for 2 h with shaking, then transferred to 30 °C and induced after 15 min with 50 µM IPTG (final concentration). Cultures were incubated for an additional 16 h at 30 °C with shaking at 200 rpm. Cultures were subsequently extracted with two volumes of EtOAc. The organic layer was collected and solvent was removed *in vacuo*. The resulting residue was dissolved in 100 µL MeOH for LC/MS/MS analysis.

Analytical scale culturing and extraction of wildtype tetrabromopyrrole-producing bacterial strains

*P*. sp. PS5 and *P*. sp. A757 were cultured as previously described (4-5). Briefly, a Difco 2216 Marine Broth (BD 212185)/agar plates were inoculated from a -80 °C glycerol stocks of *P*. sp. PS5 or *P*. sp. A757 and grown overnight at 30 °C. Five mL Difco 2216 liquid broth were inoculated with a single colonies of *P*. sp. PS5 or *P*. sp. A757 and grown overnight at 30 °C with shaking at 200 rpm. Fifteen mL Difco 2216 liquid broth supplemented with an additional 1 g/L KBr (final concentration) to increase production of brominated secondary metabolites (6) were inoculated with 100 µL of overnight cultures and incubated overnight at 30 °C with shaking at 200 rpm. The cultures were subsequently extracted with two volumes of EtOAc. The organic layers were collected, and the solvent was removed *in vacuo*. The residues were dissolved in 100 µL MeOH each for LC/MS/MS analysis.

LC/MS/MS elution profile used for analysis of *in vivo* and *in vitro* extracts

Extracts were analyzed by LC/MS/MS using a reverse phase C18 column (Phenomenex Luna, 5 µm, 4.6 × 100 mm) operating on an Agilent 1260 HPLC in tandem to an Agilent 6530 Accurate Mass Q-TOF mass
spectrometer. Mass spectra were acquired in negative ionization mode. Data was acquired for blank runs between samples to prevent and exclude carry-over of bromopyrroles from one sample to the next. HPLC solvents used were water + 0.1 % formic acid (A) and MeCN + 0.1 % formic acid (B). The HPLC elution profile was as follows: (initial flow rate 0.5 mL/min) 10% B for 5 min, linear gradient to 70% B over 10 min, linear increase to 80% B over 10 min, (flow rate changed to 0.7 mL/min) linear increase to 100% B over 0.5 min, hold at 100% B for 3 min, linear decrease to 10% B over 0.5 min. Identical injection volumes were used within a given experiment.

**Data analysis and plotting**

LC/MS/MS data was analyzed using the Agilent MassHunter software package. Data was plotted using Microsoft Excel and OriginPro (OriginLab). Extracted ion chromatograms were called using the [M-H]⁻ + m/z 0.5 for the most abundant isotope as follows: 1 (m/z 381.67), 6 (m/z 301.76), and 7 (m/z 345.75). Chromatograms corresponding to sets of experiments (i.e., within a given figure panel) are normalized to the largest peak among spectra, while individual chromatograms are normalized to the largest peak within that chromatogram.

**Cloning, protein expression and purification for in vitro assays**

Cloning and expression of Mm_Bmp1-Bmp3 complex, Mm_Bmp2, Mm_Bmp4, Mm_Bmp1(ACP), and Mm_Bmp1(TE) were performed in a manner identical to previously reported (3). *B. subtilis* Sfp and *P. stutzeri* PtdH were also purified in an identical manner to that previously described (3). Expression of *E. coli* SsuE was performed according to a published protocol (7).

Mm_Bmp1(TE)-S202A is new to this study, and was purified in an identical manner to that reported for wildtype recombinant proteins (3). Point mutagenesis of Mm_bmp1(TE) to generate the Ser202 to Ala mutant was performed by PCR mutagenesis using PrimeStar Max DNA Polymerase (Takara) standard protocol with pET28-N-His6-Mm_bmp1(TE) as the template. Primers were designed with the motif 5'-[20 nucleotide sequence-modified overlap] [35 nucleotide primer region]-3' followed by treatment with DpnI exonuclease, propagations in *E. coli* DH5α (NEB), Sanger sequencing verification (SeqXcel, La Jolla, CA), and transformation into *E. coli* B21-Gold(DE3) (Agilent) for expression.

**Total in vitro enzymatic synthesis of 1 from L-proline**

One mL reactions consisting 2 mM L-proline, 0.5 mM CoA, 2 mM MgCl₂, 50 mM KBr, 2 mM NAD⁺, 0.1 mM FAD, 10 mM phosphate, 10 μM Mm_Bmp1-Bmp3 complex, 25 μM Mm_Bmp2, 1 μM Mm_Bmp4, 2 μM Sfp, SsuE, and PtdH were incubated for 12 h at 30 °C with and without ATP in buffer consisting of 20 mM Tris-HCl (pH 8), 50 mM KCl, 10 % (v/v) glycerol. Reactions were extracted with 1.6 mL EtOAc. The organic layer was collected and solvent was removed in vacuo. The resulting residue was dissolved in 100 μL MeOH, and 30 μL were injected for LC/MS/MS analysis in the same manner as described for analysis of *in vivo* extracts.

**Chemical synthesis**

*Chemicals and solvents.* All chemicals were purchased from Acros, Aldrich, FluKa, or Alfa Aesar and used as such unless stated otherwise. For flash chromatography technical grade solvents were used without further purification. For reactions analytical grade solvents were purchased and used without further purification. Deuterated solvents were obtained from Sigma-Aldrich.

*Reactions.* All non-aqueous reactions were carried out using oven-dried glassware under an atmosphere of nitrogen unless otherwise stated. Reactions were magnetically stirred and monitored by TLC unless otherwise stated. Chromatographic purification was performed as flash chromatography (Alfa Aesar silica gel, 60 Å pore size) using the solvents indicated as eluent with 0.3-0.5 bar pressure. Thin layer chromatography (TLC) was performed on Merck silica gel 60 F₂₅₄ TLC glass plates and visualized with UV light or stained in ceric ammonium molybdate or potassium permanganate solutions. The yields given refer to chromatographically purified and spectroscopically pure compounds unless otherwise stated.
Analysis. $^1$H- and $^{13}$C- NMR spectra were recorded on VARIAN Inova (500 MHz) or BRUKER Avance (600 MHz) spectrometers in the solvents indicated. All signals are reported in ppm with the internal chloroform signal at 7.26 ppm or 77.0 ppm, or the internal DMSO signal at 2.50 ppm or 39.52 ppm as standard. The data is being reported as (s=singlet, d=doublet, t=triplet, q=quadruplet, m=multiplet or unresolved, br=broad signal, coupling constant(s) in Hz, integration). HiRes-MS was carried out on an Agilent 1100 Series Instrument with diode-array and MS detectors on a Phenomenex Luna C18 5µm 100 x 4.6 mm column.

Chemical syntheses of 1, 6, and 7

Synthetic standards of 1, 6, and 7 were prepared according to reported protocols. 1 was prepared according to literature procedures by halogenation of pyrrole using NBS (8). 2,3,4-tribromo-1H-pyrrole 6 was prepared according to a modified literature procedure (9). For the TIPS protection of pyrrole, sodium hydride was employed as a base instead of n-BuLi. 7 was prepared according to a slightly modified literature procedure (10): To a suspension of pyrrole-2-carboxylic acid (1g, 9 mmol) in CHCl$_3$ (10 mL) and AcOH (2 mL) was slowly added bromine (1.6 mL, 31.5 mmol). The reaction was warmed to 50 °C and stirred for 4 h. The mixture was diluted with CHCl$_3$ and water. The phases were separated and the organic phase was washed with water. The carboxylic acid was then extracted from the organic phase using 10% K$_2$CO$_3$. The basic solution was first washed twice with CHCl$_3$ and then acidified using 4M HCl. The resulting precipitate was collected by filtration and dried under vacuum to give 7. 1: $^1$H-NMR (500 MHz, CDCl$_3$) for full spectrum see Fig. S10: δ 8.52 (bs, 1H, NH); HRMS (ESI) m/z calculated for C$_7$H$_5$N$_2$Br$_2$ [(M+H)$^+$] 377.6770, found 377.6759. 6: $^1$H-NMR (600 MHz, CDCl$_3$) values in good agreement with literature (11): δ 8.46 (bs, 1H, NH), 6.86 (d, J = 3.1 Hz, 1H); HRMS (ESI) m/z calculated for C$_9$H$_7$N$_2$Br$_2$ [(M+H)$^+$] 299.7665, found 299.7663. 7: (300 mg, 0.86 mmol, 10%). Rf 0.15 (hexanes/EtOAc 4:1); $^1$H-NMR (500 MHz, d$_6$-DMSO, for full spectrum see Fig. S11): δ 12.97 (bs, 1H, COOH), 11.71 (bs, 1H, NH); $^{13}$C-NMR (150 MHz, d$_6$-DMSO, for full spectrum see Fig. S12): δ 159.4, 122.6, 106.7, 104.3, 103.4; HRMS (ESI) m/z calculated for C$_{13}$H$_{10}$NO$_2$Br$_2$ [(M+H)$^+$] 345.7708, found 345.7674.

Chemical synthesis of 3,4,5-trihydropyrrylol-5-pantethene probe 9

A synthetic scheme for 9 along with NMR spectra for intermediates and 2D NMR datasets for 9 are provided in the supplementary section (Figs. S13-S19). A flask was charged with previously prepared 7 (300 g, 0.86 mmol, 1 equiv.) and CH$_2$Cl$_2$ (10 mL). Oxalyl chloride (0.15 mL, 1.7 mmol, 2 equiv.) was added slowly and the residue was allowed to stir for 1 h. The solvent was removed and the residue was taken up in CH$_2$Cl$_2$ (10 mL). N-Boc-2-amino ethanethiol (228 mg, 1.3 mmol, 1.5 equiv.) followed by triethylamine (0.24 mL, 1.7 mmol, 2 equiv.) were added and the reaction was allowed to stir at ambient temperature overnight. After removal of the solvent, the residue was subjected to flash column chromatography (hexanes/EtOAc 9:1 → 4:1) to give the respective thioester (260 mg, 0.51 mmol, 59%). Rf 0.30 (hexanes/EtOAc 4:1); $^1$H-NMR (500 MHz, CDCl$_3$), n.d. Due to rotamers at the Boc carbamate, some peaks are split. NH peaks are not reported. A full 1 H spectrum is given in Fig. S13): δ 3.45-3.14 (m, 3H), 2.78 (t, J = 6.6 Hz, 1H), 1.54-1.39 (m, 9H); HRMS (ESI) m/z calculated for C$_{12}$H$_{14}$N$_2$O$_4$SBr$_2$Na [(M+Na)$^+$] 526.8246, found 526.8210. To a solution of this thioester (260 mg, 0.51 mmol) in EtOAc (4 mL) at 0 °C was added freshly prepared HCl in EtOAc (4 mL, 3 M in EtOAc). The solution was allowed to warm to ambient temperature and stirred for 15 min under use. The mixture was stirred at 0 °C for 30 min whereupon a white precipitate had formed. The solvent was removed under vacuum to afford hydrochloride salt S1 (quant.). A flask was charged with acetonide S2 (11) previously prepared (12) (64 mg, 0.25 mmol, 1.1 equiv.) and EDC (66 mg, 0.35 mmol, 1.5 equiv.). A suspension of S1 (100 mg, 0.23 mmol, 1 equiv.) in CH$_2$Cl$_2$ (2 mL) was added followed by diisopropyl ethylamine (80 µL, 0.46 mmol, 2 equiv.). The mixture was stirred at ambient temperature overnight. The solvent was evaporated and the residue was directly subjected to flash column chromatography (hexanes/EtOAc 1:1 → EtOAc) to afford to corresponding amide product (64 mg, 0.10 mmol, 43%). The acetonide (10 mg, 0.015 mmol) was dissolved in AcOH/H$_2$O (1 mL, 2:1), and the solution was stirred for 5 h. The mixture was poured into a separatory funnel containing sat. NaHCO$_3$ and EtOAc. The phases were separated and the aqueous phase was extracted three times with EtOAc.

1 To a mixture of EtOAc (2.4 mL) and EtOH (0.72 mL) at 0 °C was slowly added AcCl (0.85 mL).
The combined organic phases were washed with brine and dried over MgSO4. The solvent was removed and the residue was purified by flash column chromatography (CH2Cl2/MeOH 20:1) to give diol 9: (7.5 mg, 0.012 mmol, 80%) Rr 0.45 (CH2Cl2/MeOH 9:1); 1H-NMR (500 MHz, d6-DMSO): δ 8.16 (t, J = 5.7 Hz, 1H, NH), 7.70 (t, J = 5.9 Hz, 1H, NH), 5.36 (bs, 1H), 4.45 (bs, 1H), 3.69 (s, 1H), 3.36-3.20 (m, 4H), 3.16 (d, J = 10.4 Hz, 1H), 3.09 (t, J = 6.7 Hz, 1H), 2.31-2.22 (m, 2H), 0.79 (s, 3H), 0.77 (s, 3H); 13C-NMR (150 MHz, d6-DMSO): δ 178.9, 173.0, 170.8, 129.3, 109.9, 104.4, 104.0, 75.1, 68.2, 48.7, 39.1, 38.2, 35.2, 34.9, 21.0, 20.4; HRMS (ESI) m/z calculated for C10H22N2O3S2Br2Na ([M+Na]+) 627.8723, found 627.8719.

**Preparation and purification of 3,4,5-tribromopyrrolyl-S-Bmp1(ACP) 8 and pyrrolyl-S-Mpy15**

*E. coli* CoA biosynthetic enzymes pantothenate kinase CoaA, phosphopantetheine adenyl transferase CoaD, and dephospho-CoA kinase CoaE were expressed and purified as previously described (12). To generate 8, 10 reactions of 500 µL each consisting of 2 µM CoaA, 2 µM CoaD, 2 µM CoaE, 2 µM Sfp, 250 µM Bmp1(ACP), 1 mM 3,4,5-tribromopyrrolyl-S-pantetheine 9, 10 mM MgCl2, and 9 mM ATP in 50 mM HEPES (pH 7.9) buffer were incubated at 30°C for 12 h. As a negative control, ATP was substituted with an appropriate volume of buffer in a single 100 µL scale reaction. Reactions were pooled and purified by size exclusion chromatography on a Superdex 75 10/300 GL (GE Life Sciences) column eluted isocratically with 20 mM HEPES-Na (pH 7.5) 100 mM KCl. Glycerol was added to pooled fractions containing product to a final concentration of 10% (v/v) followed by concentration. An identical procedure was used for generation and purification of pyrrolyl-S-Mpy15 with Mpy15 switched for Bmp1(ACP) and 9 switch for pyrrolyl-S-pantetheine synthesized as previously reported (12).

**LC-MS/MS confirmation of chemo-enzymatically prepared pyrrolyl-S-ACPs**

Loading reactions were confirmed by LC/MS/MS as previously described (3). Reactions with and without ATP were injected onto a C4 column (Vydac 5 µm, 4.6 mm × 250 mm) operating on an Agilent 1100 HPLC in tandem to an Agilent 6530 Accurate Mass Q-TOF mass spectrometer. Mass spectra were acquired in positive ionization mode. HPLC solvents used were water + 0.1% formic acid (A) and MeCN + 0.1% (B). The elution profile was as follows: (flow rate: 0.7 mL/min): 10% B for 10 min, linear increase to 30% B over 5 min, linear increase to 70% B over 40 min, linear decrease to 10% B over 5 min, linear increase to 100% B over 1 min followed by 2 min at 100% B, decrease to 10% B over 1 min, 10% B for 2 min and 5 min of post-time equilibration. Loading was assessed by extraction of the predicted acyl-(cyclo)pantetheine MS2 fragments corresponding to the pantetheine arm of the holo-ACP as previously described (3). The mass and charge of the parent peptide were determined from the corresponding deconvoluted peptide MS1 spectra, and the mass difference was confirmed between apo-ACP from the no ATP negative control and holo-ACP from the reaction with ATP (Fig. S3).

**In vitro enzymatic synthesis of 1 from 8**

1 mL reactions consisting of 25 mM 8, 50 mM KBr, 0.1 mM FAD, 2 mM NADP+, SsuE, PudH, 10 mM phosphate, 50 µM Mm Bmp2, and 25 µM Mm Bmp1(TE) or Mm Bmp1(TE):S202A for 12 h at 30 °C in buffer consisting of 20 mM Tris-HCl (pH 8), 50 mM KCl, and 10% (w/v) glycerol. For reactions where enzymes were excluded (-Bmp2 and -TE) an equivalent volume of buffer was substituted. Reactions were extracted with 1.6 mL EtOAc. The organic layer was collected and solvent was removed in vacuo and the resulting residue was dissolved in 100 µL MeOH. Samples (30 µL injections) were analyzed by LC/MS/MS using the same conditions as described for in vivo and Mmp1-4 in vitro extracts. To ensure no carryover from one injection to the next, reactions were analyzed in the order ‘-Bmp2’, ‘-TE’, ‘S202A’, and ‘All’, with data gathered for blank injections preceding and following each reaction extract.

**Cloning, expression, purification, crystallization, and structure determination for Mm_Bmp2 and Mm_Bmp2-TM**

Cloning, expression, and purification. *P. sp.* PS-5 Bmp2 was cloned into the pET28b(+) (Novagen) plasmid vector utilizing the NdeI and XhoI restriction sites to afford a thrombin cleavable hexa-histidine tag at
the N-terminus of the recombinant protein. Selection pressure was maintained by inclusion of 50 μg/mL kanamycin (final concentration) in all solid and liquid growth media. The insert was verified by restriction analysis and Sanger sequencing (SeqXcel, La Jolla, CA), and transformed in to E. coli BL21-Gold(DE3) (Agilent) cells for protein expression. 1 L terrific broth culture was grown to an OD600 = 0.6 at 30 °C and protein expression was induced by the addition of IPTG to a final concentration of 0.3 mM. The temperature was adjusted to 18 °C, and the culture was allowed to incubate for an additional 18 h. Cultures were harvested by centrifugation, the supernatant discarded, and the pellet resuspended in 20 mM Tris-HCl (pH 8.0), 500 mM NaCl, 10% glycerol buffer, and lysed by sonication. The supernatant was clarified by centrifugation and loaded on to a 5 mL His-Trap Ni-NTA column (GE Biosciences) equilibrated in harvest buffer. The column was extensively washed with 20 mM Tris-HCl (pH 8.0), 1 M NaCl, 30 mM imidazole buffer, and eluted by a linear gradient to 20 mM Tris-HCl (pH 8.0), 1 M NaCl, 200 mM imidazole buffer across 20 column volumes. Purity of eluted proteins was checked by SDS-PAGE, and thrombin was added to a final concentration of 1 unit/mg recombinant protein. Thus purified protein was then dialyzed overnight against 20 mM Tris-HCl (pH 8.9), 50 mM KCl buffer at 4 °C. The dialyzed protein was applied to a 5 mL ion exchange Q Sepharose FF column (GE Biosciences) equilibrated in dialysis buffer, and eluted using a linear gradient to 20 mM Tris-HCl (pH 8.9), 1 M KCl buffer.

**Activity assays.** For activity assays, protein was concentrated after ion exchange using Amicon centrifugal filters to a final volume of 2.5 mL and buffer was exchanged using PD-10 size exclusion columns (GE Biosciences) to 20 mM Tris-HCl (pH 8.0), 10% glycerol buffer. The protein concentration was measured using the Bradford assay, and protein was stored at -80 °C in 100 μL aliquots.

**Crystallization.** For crystallization, 4 mL of ion exchange purified protein was applied to a Sephadex-75 size exclusion chromatography column (GE Biosciences) equilibrated in 20 mM HEPES-Na (pH 7.5), 100 mM KCl buffer and eluted isotropically. Protein purity was verified by SDS-PAGE, and pure fractions were pooled and concentrated using Amicon centrifugal filters to a final concentration of 100 μM. Sparse-matrix screening for crystal growth was performed using commercially available crystallization screens in hanging-vapor drop format at 9 °C. Initial conditions were optimized at 9 °C in hanging-vapor drop format to yield the following reproducible crystal growth condition for wild type Bmp2: 100 mM HEPES-Na (pH 7.5), 200 mM NaCl, 25% (w/v) PEG3350. Crystals appeared within 2 days and reached their maximum size by 5 days. The crystals were harvested and briefly dipped in 100 mM HEPES-Na (pH 7.5), 200 mM NaCl, 25% (w/v) PEG3350, 20% (v/v) ethylene glycol cryo-protectant prior to vitrification in liquid nitrogen.

**Structure determination.** X-ray diffraction were scaled and indexed using the HKL-2000 package (13) and unit cell composition determined (14). Bmp2 structure was determined by molecular replacement using Phaser (15) by a sequence-alignment based Ala model (16) of the PDB: 3E1T (17). Initial solution was extended manually using Coot (18) and computationally using ARP/wARP (19) in an iterative manner. Refinement was carried out using Phenix (20). Although clear density for FAD could be observed prior to refinement, the cofactor was manually built into the model only after the free R factor (21) dropped below 30%.

**Mm_Bmp2-TM.** Mm_Bmp2-TM was generated by PCR site-directed mutagenesis using bmp2-wt pET28b(+) plasmid vector as template. Mutations were verified using Sanger sequencing (SeqXcel, La Jolla, CA), and recombinant Mm_Bmp2-TM protein was expressed and purified in a manner identical to that described above. Mm_Bmp2-TM was prone to aggregation, and care was taken to minimize the time between purification and crystallization trials. Mm_Bmp2-TM crystallized in conditions similar to Bmp2-wt, albeit with different space group and unit cell dimensions, with the final optimized mother liquor being 100 mM HEPES-Na (pH 7.5), 200 mM ammonium sulfate, 20% (w/v) PEG3350, 5% (v/v) ethylene glycol. Crystals were briefly dipped in 100 mM HEPES-Na (pH 7.5), 200 mM ammonium sulfate, 20% (w/v) PEG3350, 25% (v/v) ethylene glycol cryo-protectant prior to vitrification in liquid nitrogen. Structure of Mm_Bmp2-TM was determined using the structure of Bmp2-wt as a molecular replacement search model.

**Cloning, expression, purification, crystallization, and structure determination for Mpy16**

**Cloning, expression, and purification.** Mpy16 was amplified from the genomic DNA of Streptomyces CNQ-418 (22) and cloned into a modified pET28-MBP vector that we have described previously (23). This vector affords a N-terminal hexa-histidine tag, followed by a maltose binding protein (MBP), and a TEV
protease cleavage site downstream of the MBP protein, followed by the translated recombinant protein. Culture
growth conditions, and protein purification protocols are identical to those described above (TEV was used in
place of thrombin to remove the purification tag), with an additional subtractive Ni-NTA chromatography step
before ion exchange employed to remove the hexa-histidine bearing MBP tag subsequent to TEV digestion.

**Crystallization and structural determination.** Initial crystallization conditions for Mpy16 were identified
as before for Bmp2, and optimized to 100 mM Tris-HCl (pH 8.0), 10% (v/v) PEG8000, 5% (v/v) glycerol. The
crystals were briefly dipped in 100 mM Tris-HCl (pH 8.0), 10% (v/v) PEG8000, 30% (v/v) glycerol cryo-
protectant prior to vitrification in liquid nitrogen. X-ray diffraction data was indexed and scaled as before for
Bmp2, and molecular replacement solution identified using SOLVE as molecular replacement search model.
Model building and refinement was carried out in a manner identical to that described above.

**Co-expression of Mm_bmp2 with Mm_bmp1, Mm_bmp3, and Mm_bmp4**

A two-plasmid system in *E. coli* BL21 Gold (DE3) for the coexpression of *Mm_bmp1*, *Mm_bmp3*, and
*Mm_bmp4* has been described previously for the in vivo production of pyrrolyl-S-Bmp1 (3). To this system was
introduced a third plasmid, pETDuet (Novagen), bearing *Mm_bmp2* or *Mm_bmp2-TM* in the second cloning site
(MCS-2) of the plasmid vector. 5 mL liquid cultures were grown in the presence of 50 µg/mL kanamycin,
50 µg/mL streptomycin, and 100 µg/mL ampicillin (final concentration) and supplemented with 1 g/L (final
concentration) of KBr. At OD600 ~ 0.6, protein expression was induced by the addition of 0.3 mM IPTG
(Promega, PAV3951) and growth was allowed at 30 °C for an additional 18 h. The culture was extracted twice
with EtOAC and solvent removed in vacuo. The residue was dissolved in methanol, and analyzed LC/MS/MS as
described previously.

**In vitro assay of wild type Mm_Bmp2, Mm_Bmp2-TM, and Mpy16 with pyrrolyl-S-ACP s as substrates**

Pyrrolyl-S-Bmp1(ACP) and pyrrolyl-S-Mpy15 were prepared as described previously (12) and used as
substrates in in vitro reactions. 50 µM pyrrolyl-S-ACP substrate was incubated with 20 µM halogenase
catalyst in 20 mM HEPES-Na (pH 7.9) reaction buffer containing 200 mM halide (KCl or KBr), 1 mM NADP+, 20 µM
FAD, 10 mM freshly made Na-phosphate, 5 mM TCEP, 10 µM Pdh and SufE enzymes, and 10% v/v glycerol.
Reactions were incubated at 30 °C for 3 h, and 100 µL reaction was quenched by the addition of 50 µL of 2%
formic acid. Precipitate was removed by centrifugation and assays analyzed by LC/MS/MS as reported
previously (12). Abundance of acyl-(cyclo) pantetheine MS2 product ions were individually normalized for each
assay as reported in Figures 6E-G.
SUPPLEMENTARY FIGURES

Figure S1. LC/MS comparisons of in vivo and in vitro EICs to those of authentic standards of 1, 6, and 7. EICs for predicted [M-H]⁺ for the most abundant isotopic masses of 1 (m/z 381.67), 6 (m/z 301.76), and 7 (m/z 345.75). All traces are self-normalized.
Figure S2. Confirmation of CoA extension and loading of pyrrole substrates onto ACPs. A. Scheme for CoA in vitro extension of chemically synthesized pyrrolyl-S-pantetheine precursor to pyrrolyl-S-CoA by E. coli CoA biosynthetic enzymes CoaA, CoaD, and CoaE. Upon completion of CoA synthesis, the substrate is loaded onto the catalytic serine of apo-ACP via B. subtilis phosphopantetheine transferase Sfp to generate “holo” pyrrolyl-S-ACP, resulting in a gain in mass specified to the right. To confirm substrate generation, the whole protein was analyzed by LC/MS/MS, resulting in a characteristic pantetheine MS2 fragments. Comparison to apo-ACP from a negative control reaction in which ATP required for CoA enzymatic steps is excluded, demonstrates the expected mass shift between the parent ion and for apo- and holo- forms. B. i. MS1 and M2 for apo-Mpy15, ii. MS1 and MS2 for pyrrolyl-S-Mpy15. Red diamonds indicated the MS1 peptide masses while blue diamonds indicated peptide mass sampled for MS2. The pantetheine fragment mass is indicated in the MS2 spectrum. Charge is indicated to the right of the parent peptide. The corresponding mass shift between apo- and holo- Mpy15 is indicated by the dotted lines. C. Analogous scheme to ‘B’ for generation of 3,4,5-tribromopyrrolyl-S-Bmp1(ACP) 8 from chemically synthesized 3,4,5-tribromopyrrolyl-S-pantetheine 9.
Figure S3. Bioinformatic identification and in vitro confirmation of catalytic Ser202 of Mm_Bmp1(TE) with a model chromogenic substrate, p-nitrophenyl acetate (pNPA).

A. i) Architecture of Bmp1 ACP-TE didomain, ii) BLAST query of the PDB database with the amino acid sequence of Mm_Bmp1 gives closest functionally and structurally characterized homolog CumD (44% similarity, 25% identity, 65% query coverage, e-value 7x10^{-14}), a meta-cleavage product hydrolase from *Pseudomonas fluorescens* IP01 involved in the degradation cumene (isopropyl benzene); Bmp1:Ser202 aligns with confirmed catalytic Ser103 of CumD (24). B. 250 µM freshly prepared pNPA in 20 mM Tris-HCl (pH ~ 8), 50 mM KCl, 10% (v/v) glycerol was incubated with 5 µM Bmp1(TE), Bmp1(TE):S202A, or no enzyme. Upon addition of fresh substrate in DMSO to the buffer solution, the reaction was monitored spectrophotometrically at wavelength of 400 nm in a plastic UV cuvette. Enzyme was added 5 minutes after addition of the substrate. Measured absorbance is plotted versus time. Upon addition of Bmp1(TE), a steep absorbance increase was observed due to production of hydrolysis product p-nitrophenol. Loss of hydrolytic activity, identical to incubation without enzyme (blue curve), was observed when pNPA was incubated with Bmp1(TE):S202A (black curve).
Figure S4. Non-enzymatic hydrolysis of proposed ACP-bound tetrabrominated intermediate yields 1. Extracted ion chromatograms (EICs) for 2,3,4,5-tetrabromo-1H-pyrrole (m/z 381.67) from incubations of 3,4,5-tri bromopryrrolyl-S-ACP with various combination of Bmp1(TE) and Bmp2. All EICs are normalized to the largest peak (highlighted in green) in ‘Bmp1(TE):S202A’ EIC. Green highlight indicates presence of 1. LC/MS analyses were performed sequentially with blank runs before and after each run in order to discount carry-over of substrate. Only the first blank run is shown, as 1 was not observed in the ‘no Bmp2’ reaction, and levels increased in the subsequent reactions. B. Mass spectral pattern at location of the dashed line in A; ‘No Bmp1(TE)’ and ‘Bmp1(TE):S202A’ reactions show m/z and isotopic distribution consistent with 1.
Figure S5. Interrogation of the potential of 2,3,4-tribromopyrrole 6 and 7 to serve as intermediates in the biosynthesis of 1. One mL reactions consisting of 25 μM 2 or 7 were incubated with 50 mM KBr, 0.1 mM FAD, 2 mM NADP+, SsuE, PtdH, 10 mM phosphate including or excluding 50 μM Bmp2 and 25 μM Bmp1(TE, 78-376) in 20 mM Tris-HCl (pH 8) with 50 mM KCl and 10% (v/v) glycerol for 12 h at 30 °C. Reactions were extracted with 1.6 mL EtOAc. Solvent was removed in vacuo and the resulting residue was dissolved in 100 μL MeOH. The extracts were analyzed by LC-MS/MS as previously described. A. Combined EICs for m/z 301.76 (6) and m/z 381.67 (1) reactions with substrate 6 (highlighted in orange) with i) Bmp1(TE) ii) Bmp2, iii) Bmp2 and Bmp1(TE), and iv) neither Bmp1(TE) nor Bmp2. Incomplete conversion to 1 (highlighted in green) was observed in ii)-iii). Mass spectra for products are shown above EICs for ii)-iii). All EICs are normalized to ‘i’. B. Combined EICs for m/z 345.75 (7) and m/z 381.67 (1) for reactions with 7 (highlighted in red) as substrate with i) Bmp2 and Bmp1(TE) and ii) neither Bmp2 nor Bmp1(TE). No conversion was observed. EICs are normalized to ‘1’.
Figure S6. Reductive debrumination of 1 by NADPH yields 6. A) EICs for 1 (m/z 381.67, shaded green) and 6 (m/z 301.76, shaded orange) for the organic extracts of 1 mL reactions incubated in 20 mM Tris-HCl (pH 8) with 50 mM KCl and 10% (v/v) glycerol for 12 h at 30 °C consisting of 50 μM 1 with 2 mM NADPH, 10 mM phosphite, and including or excluding PtdH (i.e., NADPH generation system with and without PtdH). (i) In the presence of the full NADPH generation system modest conversion of 1 to 6 is observed. (ii) When PtdH is eliminated from the reaction, no conversion of 1 to 6 is observed; NADPH is not reduced to NADPH in the absence of PtdH. B) EICs for 1 (m/z 381.67, shaded green) and 6 (m/z 301.76, shaded orange) for the organic extracts of 1 mL reactions incubated in 20 mM Tris-HCl (pH 8) with 50 mM KCl and 10% (v/v) glycerol for 12 h at 30 °C consisting of 50 μM 1 with or without 1 mM NADPH. (i) In the presence of NADPH modest conversion of 1 to 6 is observed, consistent with the experiment shown in panel ‘A’ with the complete NADPH generation system. (ii) In the absence of NADPH no conversion of 1 to 6 is observed.
Figure S7. FAD-bound holo structures of pyrolyl-S-ACP flavin-dependent halogenases from this study. (A) holo-Mm_Bmp2 structure. (B) holo-Mpy16 structure. (C) Superimposed Bmp2 (in pink) and Mpy16 (in green) structures. The flavin cofactors and the side chain of the catalytic lysine residues are shown in stick-ball representation.
Figure S8. Lack of chlorination activity for Bmp2-TM. Halogenation reactions for Bmp2-TM were carried out as before, but in the absence of bromide, and in the presence of 200 mM chloride in the reaction. As can be seen from the EICs, no mono-, di-, or tri-chlorinated pyrrolyl-S-Bmp1(ACP) products can be detected.
Figure S9. *In vivo* coexpression of *Mm_bmp1*, *Mm_bmp3*, and *Mm_bmp4* with wildtype *Mm_bmp2* (*Mm_bmp2-wt*) or *Mm_bmp2* triple mutant (*Mm_bmp2-TM*). Combined extracted ion chromatograms at [M-H]^−m/z 381.67 (1) and m/z 301.76 (6) corresponding to the dominant ions for the molecular formulae C₇N₁H₁Br₁ and C₆N₁H₂Br₁, respectively. Production of 1 and 6 can be observed when *Mm_bmp2-wt* is coexpressed with *Mm_bmp1*, *Mm_bmp3*, and *Mm_bmp4*, together with previously reported *M. mediterranea* MMB-1 phosphopantetheinyl transferase (GenBank locus tag: NC_015276.1) (3). However, neither production of 1 nor 6 is observed with *bmp2-wt* is substituted by *Mm_bmp2-TM* (expressing Bmp2-Y302S, F306V, A345W triple mutant).
Figure S10. $^1$H NMR spectrum of 1 (CDCl$_3$, 500 MHz).

Figure S11. $^1$H NMR spectrum of 7 (CDCl$_3$, 500 MHz).
Figure S12. $^{13}$C NMR spectrum of 7 (CDCl$_3$, 125 MHz).

Figure S13. Scheme for the synthesis of 9.
Figure S14. $^1$H NMR spectrum of Boc-protected S1 (CDCl$_3$, 500 MHz).

Figure S15. $^1$H NMR spectrum of 9 (d$_6$-DMSO, 500 MHz).
Figure S16. $^{13}$C NMR spectrum of 9 (d$_6$-DMSO, 150 MHz).

Figure S17. DQF-COSY spectrum of 9 (d$_6$-DMSO, 600 MHz).
Figure S18. HSQC spectrum of 9 (d$_6$-DMSO, 600 MHz).

Figure S19. HMBC spectrum of 9 (d$_6$-DMSO, 600 MHz).
<table>
<thead>
<tr>
<th>1</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
<th>30</th>
<th>35</th>
<th>40</th>
<th>45</th>
<th>50</th>
<th>55</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSS_Bmp2</td>
<td>MNQFTYDVMVIGSPFAGLCGECCRKKGSYLCIEKDEQPRFHIGESTLNGAQGRD</td>
<td>MNQFTYDVMVIGSPFAGLCGECCRKKGSYLCIEKDEQPRFHIGESTLNGAQGRD</td>
<td>MNQFTYDVMVIGSPFAGLCGECCRKKGSYLCIEKDEQPRFHIGESTLNGAQGRD</td>
<td>MNQFTYDVMVIGSPFAGLCGECCRKKGSYLCIEKDEQPRFHIGESTLNGAQGRD</td>
<td>MNQFTYDVMVIGSPFAGLCGECCRKKGSYLCIEKDEQPRFHIGESTLNGAQGRD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MNQFTYDVMVIGSPFAGLCGECCRKKGSYLCIEKDEQPRFHIGESTLNGAQGRD</td>
<td>MNQFTYDVMVIGSPFAGLCGECCRKKGSYLCIEKDEQPRFHIGESTLNGAQGRD</td>
<td>MNQFTYDVMVIGSPFAGLCGECCRKKGSYLCIEKDEQPRFHIGESTLNGAQGRD</td>
<td>MNQFTYDVMVIGSPFAGLCGECCRKKGSYLCIEKDEQPRFHIGESTLNGAQGRD</td>
<td>MNQFTYDVMVIGSPFAGLCGECCRKKGSYLCIEKDEQPRFHIGESTLNGAQGRD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MNQFTYDVMVIGSPFAGLCGECCRKKGSYLCIEKDEQPRFHIGESTLNGAQGRD</td>
<td>MNQFTYDVMVIGSPFAGLCGECCRKKGSYLCIEKDEQPRFHIGESTLNGAQGRD</td>
<td>MNQFTYDVMVIGSPFAGLCGECCRKKGSYLCIEKDEQPRFHIGESTLNGAQGRD</td>
<td>MNQFTYDVMVIGSPFAGLCGECCRKKGSYLCIEKDEQPRFHIGESTLNGAQGRD</td>
<td>MNQFTYDVMVIGSPFAGLCGECCRKKGSYLCIEKDEQPRFHIGESTLNGAQGRD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MNQFTYDVMVIGSPFAGLCGECCRKKGSYLCIEKDEQPRFHIGESTLNGAQGRD</td>
<td>MNQFTYDVMVIGSPFAGLCGECCRKKGSYLCIEKDEQPRFHIGESTLNGAQGRD</td>
<td>MNQFTYDVMVIGSPFAGLCGECCRKKGSYLCIEKDEQPRFHIGESTLNGAQGRD</td>
<td>MNQFTYDVMVIGSPFAGLCGECCRKKGSYLCIEKDEQPRFHIGESTLNGAQGRD</td>
<td>MNQFTYDVMVIGSPFAGLCGECCRKKGSYLCIEKDEQPRFHIGESTLNGAQGRD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MNQFTYDVMVIGSPFAGLCGECCRKKGSYLCIEKDEQPRFHIGESTLNGAQGRD</td>
<td>MNQFTYDVMVIGSPFAGLCGECCRKKGSYLCIEKDEQPRFHIGESTLNGAQGRD</td>
<td>MNQFTYDVMVIGSPFAGLCGECCRKKGSYLCIEKDEQPRFHIGESTLNGAQGRD</td>
<td>MNQFTYDVMVIGSPFAGLCGECCRKKGSYLCIEKDEQPRFHIGESTLNGAQGRD</td>
<td>MNQFTYDVMVIGSPFAGLCGECCRKKGSYLCIEKDEQPRFHIGESTLNGAQGRD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MNQFTYDVMVIGSPFAGLCGECCRKKGSYLCIEKDEQPRFHIGESTLNGAQGRD</td>
<td>MNQFTYDVMVIGSPFAGLCGECCRKKGSYLCIEKDEQPRFHIGESTLNGAQGRD</td>
<td>MNQFTYDVMVIGSPFAGLCGECCRKKGSYLCIEKDEQPRFHIGESTLNGAQGRD</td>
<td>MNQFTYDVMVIGSPFAGLCGECCRKKGSYLCIEKDEQPRFHIGESTLNGAQGRD</td>
<td>MNQFTYDVMVIGSPFAGLCGECCRKKGSYLCIEKDEQPRFHIGESTLNGAQGRD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure S20. Sequence alignments of Bmp2 homologs.** Sequence alignment of Bmp2 homologs from *Pseudoalteromonas* sp. PSS (PSS_Bmp2), *Marinomonas mediterranea* MMB-1 (MM_Bmp2), *Pseudoalteromonas* sp. A757 (A757_Bmp2), *Pseudoalteromonas luteoviolacea* 2t16 (2t16_Bmp2), and *Pseudoalteromonas phenolica* O-BC30 (O-BC30_Bmp2) showing conservation of Y302, F306, and A345 residues (numbering according to PSS_Bmp2 sequence).
### Table S1. Bmp homology.

<table>
<thead>
<tr>
<th>Protein name</th>
<th>Annotation</th>
<th>PS5 X Mn (% amino acid similarity/ identity)</th>
<th>PS5 X A757 (% amino acid similarity/ identity)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bmp1</td>
<td>Acyl carrier protein(ACP)-thioester didomain</td>
<td>83/70</td>
<td>98/94</td>
</tr>
<tr>
<td>Bmp2</td>
<td>FADH$_2$-dependent pyrrolyl-S-ACP brominase</td>
<td>88/83</td>
<td>99/95</td>
</tr>
<tr>
<td>Bmp3</td>
<td>FADH$_2$-dependent prolyl-S-ACP dehydrogenase</td>
<td>84/76</td>
<td>97/93</td>
</tr>
<tr>
<td>Bmp4</td>
<td>L-proline adenylation domain</td>
<td>67/51</td>
<td>93/85</td>
</tr>
<tr>
<td>Bmp5</td>
<td>FADH$_2$-dependent p-hydroxbenzoate brominase/decarboxylase</td>
<td>No PS5 homolog found</td>
<td>-</td>
</tr>
<tr>
<td>Bmp6</td>
<td>Chorismate lyase</td>
<td>No PS5 homolog found</td>
<td>-</td>
</tr>
<tr>
<td>Bmp7</td>
<td>Bromo- pyrrole/phenol coupling cytochrome P450</td>
<td>No PS5 homolog found</td>
<td>-</td>
</tr>
<tr>
<td>Bmp8</td>
<td>Carboxymuconolactone decarboxylase/peroxidase</td>
<td>No PS5 homolog found</td>
<td>-</td>
</tr>
<tr>
<td>Bmp9</td>
<td>Ferredoxin</td>
<td>73/64</td>
<td>98/93</td>
</tr>
<tr>
<td>Bmp10</td>
<td>Ferredoxin reductase</td>
<td>61/42</td>
<td>82/69</td>
</tr>
</tbody>
</table>
Table S2. Data collection and refinement statistics for Mpy16, Bmp2, and Bmp2-TM structures.

<table>
<thead>
<tr>
<th>Data collection</th>
<th>Mpy16</th>
<th>Bmp2</th>
<th>Bmp2-TM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>P2₁, 2, 2</td>
<td>P2₁, 2, 2</td>
<td>P₂₁</td>
</tr>
<tr>
<td>a, b, c (Å)</td>
<td>62.83, 73.28, 189.72</td>
<td>82.97, 90.48, 57.92</td>
<td>49.23, 51.53, 77.06</td>
</tr>
<tr>
<td>α, β, γ</td>
<td>90.00, 90.00, 90.00</td>
<td>90.00, 90.00, 90.00</td>
<td>90.00, 100.89, 90.00</td>
</tr>
<tr>
<td>Rmerge (%)</td>
<td>0.09 (0.64)¹</td>
<td>0.08 (0.94)</td>
<td>0.06 (0.49)</td>
</tr>
<tr>
<td>I/σ(I)</td>
<td>37.6 (3.0)</td>
<td>24.9 (1.9)</td>
<td>30.0 (8.5)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>100.0 (99.8)</td>
<td>100.0 (100.0)</td>
<td>98.8 (99.5)</td>
</tr>
<tr>
<td>Total reflections</td>
<td>64779</td>
<td>36287</td>
<td>26147</td>
</tr>
<tr>
<td>Redundancy</td>
<td>12.0 (11.7)</td>
<td>7.0 (7.0)</td>
<td>6.3 (5.9)</td>
</tr>
<tr>
<td>Refinement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>38.08 – 1.95</td>
<td>19.23 – 1.87</td>
<td>32.97 – 1.98</td>
</tr>
<tr>
<td>R&lt;sub&gt;work&lt;/sub&gt;/R&lt;sub&gt;free&lt;/sub&gt;</td>
<td>0.172/0.212</td>
<td>0.188/0.224</td>
<td>0.179/0.215</td>
</tr>
<tr>
<td>Number of atoms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>6825</td>
<td>3148</td>
<td>3096</td>
</tr>
<tr>
<td>Solvent</td>
<td>428</td>
<td>274</td>
<td>219</td>
</tr>
<tr>
<td>FAD</td>
<td>106</td>
<td>53</td>
<td>53</td>
</tr>
<tr>
<td>Average B values</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>31.53</td>
<td>27.75</td>
<td>26.00</td>
</tr>
<tr>
<td>Solvent</td>
<td>35.77</td>
<td>33.99</td>
<td>31.77</td>
</tr>
<tr>
<td>FAD</td>
<td>23.14</td>
<td>19.49</td>
<td>16.09</td>
</tr>
<tr>
<td>R.m.s. deviations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.007</td>
<td>0.007</td>
<td>0.007</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.036</td>
<td>1.097</td>
<td>1.085</td>
</tr>
</tbody>
</table>

¹Values for the highest resolution shell are reported in parentheses.
References

Chapter 3, in full, is a reprint of materials as it appears in “Biosynthesis of a coral chemical cue tetrabromopyrrole in marine bacteria by a uniquely adapted brominase-thioesterase enzyme pair” in Proceedings of the National Academy of Sciences U.S.A., 2016. El Gamal A., Agarwal V., Diethelm S., Rahman I., Schorn M., Sneed J.M., Louie G.V., Whalen, K.E., Mincer T.J., Noel J.P., Paul V.J., and Moore B.S. The dissertation author was one of two equally contributing primary investigators and authors of this paper.

Chapter 4: An alkylhydroperoxidase-like debrominase is an enzymatic switch in the biosynthesis of pentabromopseudilin

4.1 Introduction

Halogenated organic compounds (organohalogens) are pervasive in the environment\textsuperscript{2,23}. Efforts toward the bioremediation of anthropogenic organohalogen toxins have led to the identification of microbes and associated enzymes capable of dehalogenation through to complete mineralization of a wide range of man-made pollutants\textsuperscript{71,83,84}. In particular, cobalamin-dependent reductive dehalogenases have been the focus of much study due to their ability to couple dehalogenation with microbial respiration in contaminated sediments\textsuperscript{85}. Though considerable progress has been made in the study of their structure and mechanism, these oxygen-sensitive and often membrane-associated proteins have proven to be largely intractable to direct interrogation\textsuperscript{86}. In addition to man-made organohalogens, is an astonishing diversity of organohalogens of natural origin ranging from simple aliphatic and aromatic molecules that resemble structures and sub-structures of man-made pollutants\textsuperscript{23,70} to complex natural product scaffolds such as terpenoids, oligopeptides, alkaloids, and polyketides with diverse biological activities\textsuperscript{1}. In line with the widespread distribution of natural organohalogens, emerging studies have suggested a native role for microbial dehalogenation as part of a biogeochemical halogen cycle in marine and terrestrial environments\textsuperscript{20-22}. However, in contrast to the abundance of reported halogenated natural products (> 4,000 compounds)\textsuperscript{2}, the biochemical transformations leading to their natural formation and degradation remain largely unexplored, providing a potential untapped source of novel biocatalysts\textsuperscript{2,20,23}. Moreover, an
example of a dedicated dehalogenase from the context of a confirmed microbial natural product biosynthetic pathway has yet to be reported.

The terminal reaction in biosynthesis of the highly brominated bacterial marine natural product pentabromopseudilin (1) by the Bmp pathway involves the coupling of asymmetric 2,3,4-tribromopyrrole (2) to 2,4-dibromophenol catalyzed by cytochrome P450(CYP450)-Bmp7 (Fig. 4.1A,B)\(^{50}\). This biosynthetic scheme is at odds with earlier isotope feeding studies that implied a symmetric intermediate in the biosynthesis of 1 due to shuffling of the \(^{13}\)C-labeled 5-position of L-proline with respect C-C coupling of the L-proline-derived bromopyrrole moiety of 1 (Fig. 4.1B)\(^{81}\). In corroboration of these isotope experiments, we recently confirmed that symmetric 2,3,4,5-tetrabromopyrrole (3) (Fig 4.1B), which is not a substrate for Bmp7, is an intermediate in the biosynthesis of 1\(^{16,50}\). Therefore, in order to reconcile these observations, 3 must undergo a dehalogenation to 2 via an unknown mechanism. Here we demonstrate that the transformation of 3 to 2 is catalyzed by a dedicated dehalogenase enzyme-Bmp8 encoded in the Bmp biosynthetic pathway utilizing a cofactor-independent thiolate redox mechanism. We probed (meta)genomic datasets and found hundreds of Bmp8 homologs from diverse genera and varied genetic contexts bearing a conserved catalytic triad, and demonstrate that a distant homolog of Bmp8 is also able to catalyze the dehalogenation of 3 via an analogous mechanism. To best of our knowledge Bmp8 is the first reported example of a bacterial cofactor-independent dehalogenase, as well as the first example of a dehalogenase enzyme from the context of a confirmed natural product biosynthetic pathway.
While we previously showed that degradation of \( \textbf{3} \) to \( \textbf{2} \) occurs in aqueous solution under reducing conditions (Fig. S4.15)\(^{16} \), the fact that \( \textbf{1} \) is a major product of the Bmp pathway suggested a direct enzymatic route for the conversion of \( \textbf{3} \) to Bmp7 substrate \( \textbf{2} \). In search of a dehalogenase catalyzing the conversion of \( \textbf{3} \) to \( \textbf{2} \), we re-examined the Bmp biosynthetic gene cluster for open reading frames (ORFs) possibly encoding dehalogenating enzymes. The Bmp biosynthetic gene cluster consists of modularly encoded bromopyrrole (Bmp1–4, Bmp_BrPy) and bromophenol (Bmp5–6, Bmp_BrPh) enzymes united by coupling CYP450-Bmp7 (Fig. 4.1A,B). In turn these modules are flanked by genes encoding Bmp7 redox partners Bmp9–10 (Fig. 4.1A). Upstream of \( \texttt{bmp9–10} \) is an additional ORF \( \texttt{bmp8} \) (Fig. 4.1A) putatively encoding a \( \sim 21 \) kDa protein that annotates as a member of the carboxymuconolactone decarboxylase superfamily (CMD), which takes its name from decarboxylases participating in the degradation of monocyclic aromatic molecules, but also includes alkylhydroperoxidase subunit D (AhpD) involved in intracellular defense against oxidative stress\(^{87} \). Bacteria that produce \( \textbf{3} \) and not \( \textbf{1} \) such as the biofilm-associated marine bacterium \textit{Pseudoalteromonas} sp. PS5 possess only Bmp_BrPy and vestigial \( \texttt{bmp9–10} \) (Bmp7 is not present in these bacteria), while \( \texttt{bmp8} \) is notably absent in the genomes of these bacteria (Fig. 4.1A). Our previous studies confirmed the functions of Bmp1–7 along with Bmp9–10, but the role of Bmp8 remained speculative\(^{50} \). On the basis of primary sequence annotation as a “decarboxylase” we initially postulated that Bmp8 might play a role in the elimination of proline-derived alpha-carboxylate of a tribromopyrrolyl-S-acyl carrier protein(ACP) intermediate leading to \( \textbf{1} \) (Fig. 4.1C)\(^{50} \). Involvement of Bmp8 in the bromopyrrole biosynthesis was supported by
our observation that *in vivo* expression of *bmp8* with *bmp1–7* in *Escherichia coli* led to an increase in the heterologous production of 2 and 1 with a concomitant decrease in the level of 3. 50. However, this proposal was challenged by our more recent finding that elimination of the prolyl alpha-carboxylate is mediated by a brominative mechanism catalyzed by pyrrole halogenase-Bmp2 and leading to 3 (Fig. 4.1C) 16. Hence, we were presented with an alternate hypothesis that Bmp8 might be a reductive dehalogenase that converts 3 to 2 (Fig. 4.1B), consistent with its *in vivo* activity and distant homology to AhpD redox enzymes. Therefore, we sought to purify recombinant Bmp8 to interrogate its function *in vitro.*

**Figure 4.1:** Bmp biosynthetic pathway. (A) Bmp gene clusters from *M. mediterranea* MMB-1 (*Mm_bmp*) and *P. sp*. PS5 (*PS5_bmp*); bromopyrrole (BrPy) and bromophenol (BrPh) biosynthetic modules are labeled below; *bmp9–10* encode CYP450-Bmp7 redox partners ferredoxin (Fd) and ferredoxin reductase (Fd Red), respectively. *Mm_bmp* features a putative permease (grey) inserted between *bmp3* and *bmp4* of the BrPy module. (B) Biosynthetic scheme for 1 including the hypothesized Bmp8-catalyzed dehalogenation of 3. The position of 13C-isotope label from Ref. 14 is indicated by the blue circle at the 5-position L-proline and then again in 1. (C) Terminal reaction sequence leading to 3; 2 arises as an off-pathway reductive degradation product of 3.

### 4.2 Results and Discussion
Insolubility of Bmp8 had precluded previous functional characterization. However, the fact that Bmp8 demonstrated in vivo activity in E. coli heterologously expressing the Bmp pathway implied that its soluble expression might rely on the co-expression of another component of the Bmp pathway. We recently showed that thioesterase domain of the ACP-thioesterase(TE) didomain-Bmp1, Bmp1(TE), catalyzes offloading of a tetrabrominated pyrrole from Bmp1(ACP) en route to 3 (i.e., the reaction sequence directly preceding proposed dehalogenation of 3) (Fig. 4.1C). Fortuitously, co-expression of N-His$_6$-tagged-Bmp8 with untagged Bmp1(TE), lead to soluble expression of Bmp8 from the Marinomonas mediterranea MMB-1 Bmp gene cluster (Fig. 4.1A) as a stand-alone recombinant protein not in complex with Bmp(TE) following purification by affinity chromatography (Fig. S4.1). We subsequently performed proteolytic cleavage of the N-terminal His$_6$-tag prior to biochemical characterization of Bmp8.

With purified soluble Bmp8 in hand, we tested the hypothesis that Bmp8 catalyzes the dehalogenation of 3 to 2 (Fig. 4.2A, S4.10). Incubation of synthetically prepared 3 with equimolar Bmp8 as-purified lead to rapid and complete turnover to 2 (Fig. 4.2B). Pre-incubation of Bmp8 with EDTA did not alter activity, implying that Bmp8 does not utilize a metal cofactor (Fig. 4.2B). Addition of 3 in 2-fold excess to Bmp8 led to incomplete conversion to 2, while turnover was achieved by the addition of glutathione, suggesting that Bmp8 undergoes oxidation in the course of the debromination of 3 and is rescued by an exogenous reducing agent (Fig. 4.2A,B). Slight conversion of 3 to 2 was also observed in the presence of glutathione and absence of Bmp8 (Fig. 4.2B), consistent with our previous observation that 3 is
labile to aqueous reductive degradation. A minor product with predicted molecular formula C₄NH₃Br₂ confirmed to be symmetric 3,4-dibromopyrrole (4) by ¹H-NMR was additionally observed in all reactions where Bmp8 was present indicating that 2 might also be a substrate for Bmp8 (Figs. 4.2A,B and S4.10). Consistent with this observation, incubation of Bmp8 with a 2-fold excess of 2 and glutathione lead to partial conversion of 2 to 4, in contrast to complete conversion of 3 to 2 for the same incubation time (Figs. 4.2B and S4.11). Hence, Bmp8 regioselectively dehalogenates the symmetric 2/5-position of 3. The loss in efficiency exhibited by the second dehalogenation is likely due to a change in the reactivity of the pyrrole upon elimination of an electron-withdrawing bromine atom. To explore halogen specificity, we incubated Bmp8 with the chlorinated analog of 3, 2,3,4,5-tetrachloropyrrole, and glutathione and observed partial conversion to a product with molecular formula C₄NH₂Cl₃ as determined by high resolution mass spectrometry (Figs. S4.2 and S4.9). The difference in reactivity between the brominated and chlorinated substrates is likely explained by significant differences in bond dissociation energies between C-Cl and C-Br bonds (ΔΔH_f ~ 120 KJ/mol)⁸⁸. Consistent with this observation, non-enzymatic dechlorination of 2,3,4,5-tetrachloropyrrole by glutathione was not observed (Fig. S4.2). Taken together these results indicate that Bmp8 is a cofactor-independent reductive dehalogenase that catalyzes a series of regioselective dehalogenations of halopyrroles that appears to be oxidized in the course of dehalogenation (Fig. 4.2A). In further support of loss of activity of Bmp8 due to oxidation, we showed that Bmp8 exhibits a loss of activity following extended pre-incubation in air (Fig. S4.13). Additionally, we showed that Bmp8 dialyzed after
treatment with excess 1 exhibits a loss of activity that can be recovered upon subsequent dialysis with a reducing agent (Fig. S4.12). We next investigated the mechanism of dehalogenation employed by Bmp8.

Figure 4.2: Debromination of bromopyrroles by Bmp8. (A) Scheme for reductive dehalogenation 3 and 2 and recycling of Bmp8 by glutathione (GSH). (B) Combined extracted ion chromatograms (EICs) for m/z’s corresponding to [M-H]⁻ ions for 3 (shaded green), 2 (shaded orange), and 4 (shaded red) for organic extracts of Bmp8 incubated with 3 or 2 at various ratios of substrate to enzyme with and without additives for a fixed reaction time (30 min); remaining starting material is indicated by an asterisk (*).

Due to the fact that the transformation of 3 to 2 is a net 2e⁻ reduction reaction, we inspected the primary sequence of Bmp8 for redox motifs. Bmp8 bears homology to the CMD superfamily member AhpD, whose primary function is that of a thioredoxin that reduces the side chain thiolate of the catalytic cysteine residue of the alkylperoxidase AhpC, but has also been shown to possess alkylhydroperoxidase
activity (ROOH to ROH and H₂O)⁸⁷,⁸⁹. AhpD contains a redox motif found among diverse classes of enzymes catalyzing disulfide exchange reactions (i.e., thioredoxins) consisting of two cysteine residues separated by two variable amino acid residues (CXXC)⁹⁰. Consistent with its homology to AhpD, Bmp8 contains a CXXC motif (Bmp8-CXXC) comprised of amino acid residues Cys82 and Cys85 (Fig. 4.3A). Additionally, our query of publically available (meta)genomic datasets with the primary amino acid sequence of Bmp8 lead to the discovery of an expanded conserved CXYCXXH motif present among hundreds of homologs from genomic and marine environmental metagenomic sequences (Fig. S4.3). The Bmp8 homolog sequences for which genomic information was available represent diverse genera of marine bacteria, and the vast majority of these sequences are derived from genetic contexts other than the Bmp pathway (Fig. S4.3). We next explored the functional significance of the conserved CXYCXXH motif identified in our homology search.

We first investigated the role of the Bmp8-CXXC motif in the redox mechanism of Bmp8. In mechanisms described for CXXC-containing reductases, the substrate is first attacked by an electron-donating cysteine residue side chain thiolate to form a thioether, which is then resolved by attack of the thioether bond by the second cysteine side chain thiolate resulting in release of the reduced product and formation of an intramolecular disulfide bridge at the enzyme active site⁹⁰. Hence we reasoned that the mechanism of Bmp8 might terminate in the formation of disulfide bridge between Bmp8-Cys82 and Bmp8-Cys85 (Fig. 4.3B). To investigate the role of Bmp8-CXXC in the Bmp8 reaction mechanism, we incubated fully reduced Bmp8 with a 2-fold excess of 3 or an equivalent volume of buffer and monitored the redox
state of the Bmp8-CXXC by mass spectrometric analysis of the trypsin-digested enzyme. For Bmp8 incubated with a 2-fold excess 3 we observed only one population corresponding to the oxidized Bmp8-CXXC-containing trypsin-digested fragment (−2H), while for the untreated sample we observed populations corresponding to both the oxidized and reduced fragments as anticipated due to increased susceptibility to oxidation of the cysteine residue side-chain thiols upon denaturation of the protein (Fig. 4.3C). To eliminate adventitious oxidation, we repeated the experiment, but quenched the intact enzyme with a thiol alkylating agent iodoacetamide prior to digestion. As expected, for the enzyme incubated with 3 we observed no alkylation of the trypsin-digested fragment containing the Bmp8-CXXC, while for the untreated protein we observed only the alkylated species (Fig. 4.3D). This binary result confirms oxidation of the Bmp8-CXXC in the course of dehalogenation of 3, and also demonstrates that the enzyme as-purified is initially in its fully reduced state, which we also confirmed by alkylation of the intact enzyme at the starting point of the reaction (Fig. S4.4). Bmp8 homologs derived from the Bmp pathway context contain an additional two conserved cysteine residues corresponding to Cys57 and Cys163 in M. mediterranea Bmp8 (Fig. S4.5). Hence we also investigated the potential of these additional conserved Cys residues to participate in the redox mechanism of Bmp8. While mutagenesis of either residue lead to insolubility of the protein, levels of alkylation of the trypsin-digested fragments containing either Cys57 and Cys163 were comparable between both 3-treated and untreated Bmp8, indicating that these additional conserved Cys residues are not involved in the redox reaction mechanism of Bmp8, but may instead serve a stabilizing role.
Figure 4.3: Redox state of Bmp8-CXXC. (A) Alignment (MAFFT) focusing on the region containing the conserved CXXC motif for Bmp8 from three marine bacteria harboring the Bmp pathway, *M. mediterranea* (Mm_Bmp8; ‘Bmp8’ in main text), *P. luteoviolacea* 2ta16 (2ta16_Bmp8), and *Pseudoalteromonas phenolica* O-BC30 (OBC30_Bmp8), in addition to functionally characterized bacterial alkylhydroperoxidase thioredoxin AhpD (*Mycobacterium tuberculosis* H37Rv, GenBank AAA86657.1); cysteine residues corresponding to the CXXC motif are highlighted in red with corresponding residue numbering indicated above for Mm_Bmp8 and below for AhpD. (B) Hypothesis scheme for redox state of the Bmp8-CXXC with and without treatment with a 2-fold excess of 3 and alkylation thereof with iodoacetamide (IAA). (C) EICs for the predicted m/z’s for [M+2H]^{2+} ions for the trypsin-digested peptide fragment containing Bmp8-CXXC in the reduced state (top, left) and oxidized state (bottom, left). Due to overlap between the oxidized and reduced peptide isotopic masses, the oxidized peptide mass extracts with the reduced mass, therefore the peak corresponding to the reduced peptide is highlighted in red for clarity. To the right of EICs, mass spectra averaged over EIC peaks corresponding to reduced and oxidized fragments with the -2H shift indicated for the oxidized peptide spectrum with respect to the reduced peptide spectrum (only the spectra corresponding to oxidized fragment for ‘-3’ is shown for the sake of space). Calculated isotope distributions are indicated by the red lines. (D) EICs for m/z’s for predicted [M+2H]^{2+} ions for alkylated and oxidized trypsin digested Bmp8-CXXC-containing peptide fragments with corresponding mass spectra averaged over EIC peaks shown to the right; predicted peptide distributions are indicated by the red lines.
Having identified the electron-donating role of Bmp8-CXXC, we next sought to determine the individual roles of Bmp8-Cys82 and Bmp8-Cys85 in the Bmp8 reaction mechanism. The mechanism for 2-Cys peroxidases such as AhpD involves an “peroxidatic” Cys residue that is critical to peroxidase activity and performs the initial attack of the peroxide (ROOH) to form a sulfinic acid species (Cys-S-OH), which is then resolved by a second non-essential Cys residue side chain thiolate releasing water and forming a disulfide bridge; the function of the resolving Cys can be complemented by exogenous nucleophiles in 1-Cys peroxidases that lack a resolving Cys residue. In the case of the peroxidase activity of Bmp8-homolog *M. tuberculosis* AhpD, Cys133 (corresponding to Bmp8-Cys85) serves the role of the peroxidatic residue, while Cys130 (corresponding to Bmp8-Cys82) provides the resolving thiolate and is not required for activity (Fig. 4.3A). To distinguish the roles of Bmp8-CXXC cysteine residues, Cys82 to Ala (Bmp8:C82A) and Cys85 to Ala (Bmp8:C85A) mutants were constructed and assayed for activity. Incubation of Bmp8:C82A with a 2-fold excess of 3 and glutathione lead to complete conversion to 2 and 4, while no conversion was observed for Bmp8:C85A treated in the same manner (Figs. 4.3A and 4.4A). Hence Cys85 is essential to activity, while Cys82 plays a non-essential role that appears to be complemented by an exogenous nucleophile.

To investigate the role of the glutathione thiolate as a complementary nucleophile, we monitored Bmp8:C82A incubated with glutathione by whole-protein spectrometry. Upon addition of a 2-fold excess of 3, we observed a mass shift consistent with glutathionylation of Bmp8:C82A. By contrast no glutathionylation
was observed Bmp8:C85A, indicating that the observed modification takes place on the side chain thiolate of Bmp8-Cys85 (**Figs. 4.4B and S4.6, Table S4.1**). Hence, the glutathione thiolate appears to complement for the resolving role of Bmp8-Cys82 side chain thiolate. Interestingly, addition of 3 to wildtype Bmp8 lead to positive mass shifts consistent with singly and doubly glutathionylated species, as well a negative mass shift that likely corresponds to the oxidized (-2H) form (**Figs. 4.4B and S4.6, Table S4.1**). This observation is consistent with the mechanism of disulfide bridge opening by glutathione, in which the glutathione thiolate would attack either the side chain sulfurs of Bmp8-Cys82 or Bmp8-Cys85 to form a disulfide bridge with one side chain thiol, while liberating the other. If the Bmp8-Cys82 side chain sulfur is the initial site of attack by the glutathione thiolate, then Bmp8-Cys85 would still be available to perform its catalytic function, leading to a second glutathionylation following an additional round of catalysis to afford doubly gluathionylated Bmp8, which we observed as the major gluathionylated species in the reaction with wildtype Bmp8 (**Figs. 4.4B and S4.6, Table S4.1**). On the other hand, if the Bmp8-Cys85 side chain were the initial site of attack, Bmp8 would have to be recycled back to its full reduced form before turnover could be restored, consistent with our observation of minor population of singly gluathionylated wildtype Bmp8 (**Figs. 4.4B and S4.6, Table S4.1**). Therefore, the fact that doubly gluathionylated species is the major species observed indicates a strong bias for nucleophilic attack on the side chain sulfur of Cys82, consistent with its resolving role. We further investigated whether adducts were formed upon incubation of Bmp8-Cys82 with 2 and 5, and observed no change in mass, consistent with the relatively unproductive interaction of these
substrates with Bmp8 (Fig. S4.16). Having confirmed the mechanism for complementation of Cys82 by glutathione, we further sought to explore an evolutionary justification for conservation of the CXXC motif among Bmp8 homologs.

The proposed role for the resolving Cys residue in 2-Cys peroxiredoxins is to protect that attacking Cys thiol from over-oxidation (i.e. to sulfinic acid, Cys-S-O2H), which leads to inactivation of the enzyme. Hence, we reasoned that Bmp8-Cys82 may serve the evolutionary function of protecting Bmp8-Cys85 from over-oxidation. Moreover, if Bmp8-Cys85 is susceptible to over-oxidation Bmp8:C82A should be capable of multiple turnovers in the absence of an exogenous reducing agent, where water rather than glutathione might serve as the exogenous resolving nucleophile. Indeed, incubation of Bmp8:C82A as-purified with a 2-fold excess of 3 in the absence of glutathione lead to complete conversion to 2 and 4 in contrast to only partial conversion observed for the wildtype enzyme treated with a 2-fold excess of 3 (Fig. 4.4A). To verify the formation of a sulfinic acid species with side chain thiol of Bmp8-Cys85, we monitored Bmp8:C82A by whole-protein mass spectrometry before and after the addition of 3 in the absence of an exogenous reductant. Addition of 3 to Bmp8:C82A in the absence of glutathione lead to an observed mass shift consistent with the formation of a sulfinic acid species (i.e., + 2[O]) (Figs. 4.4C and S4.6, Table S4.1). In contrast, pre-incubation of Bmp8:C82A with DTT (an alternative reductant to glutathione) lead to no change in mass upon addition of 3, further supporting that the mass shift observed in the absence of reductant is due to over-oxidation (Figs. 4.4A,C and S4.6, Table S4.1). We additionally showed that the
Bmp8 mechanism does not involve molecular oxygen by perform the reaction under anaerobic conditions; hence the apparent addition of oxygen is likely due to hydroxylation by water rather than addition of molecular oxygen (Fig. S4.14). To resolve the site of proposed hydroxylation, we additionally monitored Bmp8:C85A and wildtype Bmp8 before and after addition of 3 in the absence an exogenous reductant. Upon addition of 3, wildtype Bmp8 exhibited a subtle but significant downshift in mass likely due to disulfide bridge formation, while no mass shift was observed for Bmp8:C85A treated with 3, coincident with its lack of activity (Fig. 4.4C, Fig. S4.6, Table S4.1). Hence we deduced that over-oxidation in Bmp8:C82A takes place on the side chain thiolate of Bmp8-Cys85. Taken together these results indicate that Bmp8-Cys85 is essential for activity, while Bmp8-Cys82 protects Bmp8-Cys85 from over-oxidation. We next explored the role of the His and Tyr residues of the CXYCXXH motif conserved among Bmp8 homologs.
Figure 4.4: Biochemical interrogation of Bmp8-CXXC. (A) Combined EICs for predicted m/z’s of [M-H]- ions corresponding to 2 (highlighted orange), 3 (highlighted green), and 4 (highlighted red) for organic extracts of reactions consisting of Bmp8:C82A (C82A), Bmp:C85A (C85A), or wildtype Bmp8 (WT) incubated with a 2-fold excess of 3 and with reductants (GSH or DTT), and without reductants (-red); all reactions were quenched after a fixed incubation time of 30 min. (B)-(C) Mass spectra corresponding to the 21+ ion for Bmp8:C82A (C82A), Bmp:C85A (C85A), and wildtype Bmp8 (WT) incubated under various conditions before and after addition of a 2-fold excess of 3; red dashed lines indicated peak-to-peak alignments between treatments. Mass calculations for panels ‘B’ and ‘C’ are detailed in Fig. S6 and Table S1.

The YCXXH portion of the motif conserved among Bmp8 homologs suggests a potential Cys-Try-His catalytic triad essential to the reaction mechanism. To investigate the role of the conserved His residue, we performed site-directed mutagenesis of Bmp8-His88 to Val (Bmp8:H88V). Incubation of Bmp8:H88V with a 2-fold excess of 3 and glutathione led to no turnover, in support of its role in the Bmp8 reaction mechanism (Fig. 4.5A,B). Attempts to characterize the role of the conserved Tyr residue Bmp8-Tyr84 were precluded by insolubility of both Bmp8-Tyr84 to Phe and Ile mutants, perhaps supporting a stabilizing role for Tyr84. We further explored the relevance of the conserved CXYCXXH motif in predicting
bromopyrrole dehalogenation activity. Among the hits from our query of genomic
databases with the sequence of Bmp8 was a distant *Pseudoalteromonas luteoviolacea*
2ta16 homolog (2ta16_Bmp8-H; 55% Mm_Bmp8) from the genetic context of ORFs
putatively encoding enzymes involved in the electron transport, suggestive of a role
in scavenging reactive oxygen species⁹³ (Figs. 4.5A, S4.3, and S4.7). Notably, *P. luteoviolacea* 2ta16 also contains a functional Bmp8 homolog from the context of the
Bmp pathway (2ta16_Bmp8), which possesses even lower pairwise amino acid
similarity to 2ta16_Bmp8-H than does *M. mediterranea* MMB-1 Bmp8 (48.5% for
2ta16_Bmp8 as compared to 55% for Mm_Bmp8) (Fig. S4.8). We expressed and
purified ~22 kDa 2ta16_Bmp8-H and incubated the purified recombinant protein with
a 2-fold excess of 3 and glutathione. Incubation of the reaction with 2ta16_Bmp8-H
lead to partial conversion 3 to 2 for the same period in which full conversion was
observed for incubation with Bmp8, while extended incubation time led to complete
conversion of 3 to 2 and 4 (Figs. 4.5B and S4.7). Individual point mutagenesis of the
the Cys and His residues of the CXYCXXH motif of 2ta16_Bmp8-H led to the same
activity profile as exhibited for the corresponding Bmp8 mutants, implying an
analogous dehalogenation reaction mechanism involving these conserved residues
(Fig. 4.5A,B). Although lack of solubility precluded characterization of Bmp8-Y84F,
the corresponding Tyr to Phe mutant of 2ta16_Bmp8-H (Bmp8-H:Y80F) was soluble,
and demonstrated loss of activity with respect to dehalogenation of 3 (Fig. 4.5B). As
Tyr and Phe only differ by a hydroxyl group, this result suggests a stabilizing
hydrogen-bonding role for the conserved Tyr residue in 2ta16_Bmp8-H, and by proxy
in Bmp8.
Figure 4.5: Biochemical interrogation of the conserved CXYCXXH of Mm_Bmp8 and a distant homolog 2ta16_Bmp8-H. (A) Alignment (MAFFT) of M. mediterranea Bmp8 (Mm_Bmp8) with a distant P. luteoviolacea 2ta16 Bmp8 homolog (2ta16_Bmp8-H; GenBank WP_023399179.1) conserved residues corresponding to catalytic triad highlighted in colors with corresponding residues. (B) Combined EICs for predicted m/z’s of [M-H]⁻ ions corresponding to 2 (highlighted orange), 3 (highlighted green), and 4 (highlighted red) for Mm_Bmp8 (‘Bmp8’ in main text) and 2ta16_Bmp8-H (Bmp8-H, font colored red to distinguish from Bmp8) wild type enzymes (WT) and corresponding putative catalytic triad mutants for enzymes incubated with 3 and glutathione for a fixed reaction time (30 min); Bmp8-Y84F precluded is excluded due to insolubility of the recombinant enzyme.

Based on our mutagenesis experiments, we propose a mechanism for Bmp8 beginning with the shuttling of a proton from the thiol of Bmp8-Cys85 via Bmp8-His88 to the 2-position of 3 (Fig. 16). We next propose one of two mechanistic routes for dehalogenation 3 that each converge on the formation of a disulfide bridge between Bmp8-Cys82 and Bmp8-Cys85. In the first route (‘i’ in Fig. 4.6) the side chain thiolate of Bmp8-Cys85 would attack at the 2-position of 3, resulting in
displacement of bromide and the formation of a stable thioether bond with the pyrrole. Bmp8-Cys82 would then perform a nucleophilic attack on the thioether resulting in liberation of 2 and formation of a disulfide bridge. This route is similar to that employed by the CXXC-containing disulfide bond exchange protein DsbB, whose Cys44 side chain thiolate forms a thioether bond with its charge-transfer partner ubiquinone that is displaced by nucleophilic attack by the side chain thiolate of Cys41\textsuperscript{94}. This route is also analogous to that proposed for glutathione S-transferase reductive dehalogenases, which begins with a chloride-displacing glutathionylation of the substrate tetrachlorohydroquinone, followed by the release of the substrate via a disulfide-exchange reaction with a catalytic Cys residue in the enzyme active site\textsuperscript{95,96}.

The fact that glutathione is able to dehalogenate 3 non-enzymatically, may therefore lend support to this mechanistic route. The second route (‘ii’ in Fig. 4.6) is inspired by the mechanism for 2-Cys peroxiredoxins in which a sulfenic acid species (Cys-S-OH) is formed upon the initial attack of the catalytic cysteine residue side chain thiolate on the peroxide \textsuperscript{91}. In this alternate route, Bmp8-Cys85 would directly attack bromine to form a Cys-S-Br species, with the concomitant release of 2. Subsequent nucleophilic attack of the thioether by the Bmp8-Cys82 side chain thiolate would release bromide, leading to disulfide bond formation. An analogous mechanism to route ‘ii’ has been proposed in the mechanism peroxiredoxin-like selenoprotein thyroxine (thyroid hormone) deiodinases in which a redox active selenocysteine residue side chain selenol (in place of a cysteine side chain thiol) forms a putative transient selenyl-iodide adduct that is displaced by a resolving cysteine residue side chain thiolate\textsuperscript{97}. While a selenyl-iodide intermediate has never been detected on a
deiodinase, formation of a selynyl-iodide intermediate has been detected in a sterically stabilized model organoselenol compound. Unfortunately, we were unable to obtain evidence to decisively favor one route over another.

**Figure 4.6: Bmp8 mechanistic proposal.** Mechanistic proposal for dehalogenation of 3 by Bmp8; Mm_Bmp8 catalytic triad residues are color-coded according to panel ‘A’. Conserved residues are color-coded according to the alignment shown in Fig. 5A.

In characterizing Bmp8 we demonstrate a rare example of a cofactor-independent reductive dehalogenase from a confirmed physiological context. While dehalogenation has been implied in the course of biosynthesis of natural products involving cryptic substrate-activating halogenation reactions, to the best of our knowledge Bmp8 is the first example of a dedicated tailoring dehalogenase enzyme from the confirmed context of a natural product biosynthetic pathway. Moreover, while a variety of enzymes have been identified based on their ability to dehalogenate man-made organohalogens, the physiological nature of these transformations is speculative. To best of our knowledge, the only other examples of dehalogenases for which a physiological context is known are transmembrane flavoprotein and selenoprotein deiodinases involved in the activation-inactivation and catabolism of thyroid hormone and thyroid hormone iodotyrosine biosynthetic...
byproducts\textsuperscript{7,74,99}. Of relevance to Bmp8, the selenoprotein deiodinases utilize a cofactor-independent reaction mechanism akin to 2-Cys peroxiredoxins in which an active site selenocysteine residue serves as the electron donor in places of a redox active cysteine residue \textsuperscript{97}. Bmp8 bears no sequence homology to selenoprotein deiodinases, which are distantly related to 2-Cys peroxiredoxins, while Bmp8 bears homology to thioredoxins with peroxidase activity (e.g., AhpD). Therefore, the dehalogenation mechanism employed Bmp8 and selenoprotein deiodinases appears to be an example of convergent evolution. With respect to additional dehalogenases utilizing redox thiolate mechanisms, glutathione S-transferase-like enzymes involved in putative chlorophenol degradative pathway utilize a glutathione-dependent mechanism, as opposed to a mechanism that can be complemented by glutathione\textsuperscript{95,100-103}. It is worth noting here that glutathione may not be the physiological or sole means of regenerating Bmp8. Indeed, the fact that we were able to observe stable glutathionylated forms of Bmp8, suggest that its mechanism for recycling is inefficient. While glutathione occurs at high (mM) intracellular concentrations in bacterial cells\textsuperscript{104}, and has been suggested as the redox partner for 1-Cys peroxidases that lack a resolving cysteine residue\textsuperscript{91}, alternative regeneration systems may also be employed in the \textit{in vivo} regeneration of Bmp8. For example, peroxidase activity of \textit{M. tuberculosis} AhpD can be recycled by a surrogate electron donor NADPH-dependent flavoprotein thioredoxin reductase AhpF, an alternative electron donor for peroxiredoxin AhpC\textsuperscript{87}. Although no dedicated thioredoxin reductase is encoded in the Bmp gene cluster, homologs of AhpF are present in the
genomes of strains harboring the Bmp pathway, and therefore may serve as opportunistic redox partners for Bmp8.

The importance of dehalogenation biocatalysts in bioremediation invites speculation on the biotechnological potential of Bmp8. The major class of reductive aromatic organohalogen dehalogenases studied to-date are transmembrane oxygen-sensitive cobalamin-dependent enzymes that, despite decades of study and identification of hundreds of homologs, have proved experimentally intractable. By contrast, Bmp8 operates freely under aerobic conditions utilizing a simple cofactor-independent reaction mechanism that provides a promising platform for protein engineering. Indeed, we showed that a simple conserved CXYCXXH sequence motif is sufficient to predict bromopyrrole dehalogenase activity in a distantly related homolog, suggesting a potential for engineering of novel substrate scope by manipulation of the enzyme scaffold. Further, it has been shown that the simple alteration of the variable residues of CXXC motifs is sufficient for tuning the redox potential of the catalytic cysteine residue. Our discovery of a tailoring dehalogenase from a natural product biosynthetic context further brings to light elucidation of organohalogen natural product biosynthesis as a paradigm for the elucidation of novel dehalogenation enzymology acting on diverse substrate scaffolds.

Our initial exploration of Bmp8 was initially motivated by its biosynthetic role, which is essentially that of an enzymatic switch that repurposes an otherwise dead-end product into an intermediate in the biosynthesis of 1. In effect, Bmp8 delivers back-end regiocontrol following regiopromiscuous decarboxylative
halogenation by flavin-dependent halogenase-Bmp2. Further, given the well-defined ecological function of 3 as a chemoattractant settlement cue for coral larvae, it is interesting to consider the ecological consequence of Bmp8, which routes 3 into the considerably more energetically synthesis of 1\(^{80}\). Indeed, producers of 3 such as P. sp. PS5 seem to harbor only the Bmp bromopyrrole biosynthetic module, and lack Bmp8 and hence the capacity to synthesize 2 enzymatically\(^{16}\). While both 1 and 2 have been shown to possess feeding deterrent and antimicrobial activities\(^{75,106-108}\), it remains to be seen whether their ecological functions are redundant to or different from that of upstream 3—an intriguing case of biosynthesis motivating ecology.

4.3 Supplementary Information

4.4.1 Materials & Methods

**Preparation and characterization of halopyrroles.** *Preparation:* 2,3,4,5-tetrabromopyrrole (3) and 2,3,4-tribromopyrrole (2) were synthesized as previously described \(^{16}\). 3,5-dibromopyrrole (4) was prepared enzymatically. Briefly, ten 1 mL reactions consisting of 10 µM Bmp8, 100 µM 3, and 10 mM TCEP (to recycle Bmp8) were incubated in buffer (20 mM Tris-HCl pH 8 and 10% glycerol) for 12 h at 30 ºC. These conditions results in major production of a compound with molecular formula C\(_4\)H\(_3\)Br\(_2\)N as determined by HRMS with trace intermediary 2 remaining. 4 was purified by HPLC using a reverse-phase C\(_{18}\) column (Phenomenex Luna, 5 µM particles size, 100 Å pore size, 250 mm x 10 mm) the following gradient with Buffer ‘A’ (H\(_2\)O + 0.1% TFA), Buffer ‘B’ (MeCN + 0.1% TFA): (initial flow rate 3 mL/min) 90% to 30 % A over 5 min, hold at 30% A for 5 min, (increase flow rate to 4 mL/min) to 0% A over 15 sec, hold at 0% A for 2 min, to 90% A over 15 sec, hold
at 90% 3 min. All halopyrroles were stored in DMSO at -20 °C and were stable under repeated freeze-thaw. Characterization: The $^1$H-NMR spectrum of 4 confirmed its symmetry, and the $J$-coupling constant of the symmetric $CH$ of 4 matched to the literature value for the asymmetric $CH$ of 2, corroborating its assignment. **2,3,4,5-tetrachloropyrrole** was prepared in a manner analogous to 3, using NCS in lieu of NBS. The structure of the compound was confirmed by HRMS and $^1$H-NMR. 4: $^1$H-NMR (600 MHz, CDCl3) delta 8.32 (bs, 1H, NH), 6.81 (d, $J$ = 3 Hz, 2H), HRMS (ESI) [M-H]$^-$$m/z$ calculated for molecular formula C$_4$H$_3$Br$_2$N 221.8599, found 221.8599; **2,3,4,5-tetrachloropyrrole**: $^1$H-NMR (600 MHz, CDCl3): 9.06 (bs, 1H, NH), HRMS (ESI) [M-H]$^-$$m/z$ calculated for molecular formula C$_4$HCl$_4$N is 201.8797, found 201.8790. The mass of the single product observed for incubation of 2,3,4,5-tetrachloropyrrole with Bmp8 was measured to be [M-H]$^-$$m/z$ 167.9162 (calculated for molecular formula C$_4$H$_2$Cl$_3$N is 167.9180).

**Cloning, expression, and purification of Mm_Bmp8.** Expression of Bmp8 was achieved by co-expression of Marinomonas mediterranea MMB-1 bmp8 (Mm_bmp8) along with *M. mediterranea* MMB-1 bmp1(TE) (Mm_bmp1(TE)). Mm_bmp1(TE) comprises residues 78-376 of Mm_Bmp1, previously identified as the thioesterase domain of ACP-TE didomain Bmp1. PCR-amplified Mm_bmp8 was ligated into the NdeI/XhoI sites of pET28a, which incorporates a N-terminal His$_6$-tag. Mm_bmp1(TE) was cloned into MCS-2 of pCDFDuet, which does not incorporate a His-tag. pET28-N-His$_6$-Mm_bmp8 and pCDFDuet-MM_bmp1(TE) were co-transformed into *E. coli* BL21(GOLD). Expression was performed as by inoculation
of 1 L of terrific brother with 20 mL overnight culture with appropriate antibiotics. The culture was incubated for 6 h at 30 °C with shaking at 200 rpm. Temperature was reduced to 18 °C over 1 h followed by induced with 300 µL 1 M IPTG. The induced culture was incubated at 18 °C for an additional 14 h. Cultures were harvested by centrifugation, the and the pellet was resuspended 20 mM Tris-HCl (pH 8.0), 500 mM NaCl, 10% glycerol buffer, and lysed by sonication. The supernatant was clarified by centrifugation and loaded on to a 5 mL His-Trap Ni-NTA column (GE Biosciences) equilibrated with harves buffer. The column was extensively washed with 20 mM Tris-HCl (pH 8.0), 1 M NaCl, 30mM imidazole buffer, and eluted by a linear gradient to 20 mM Tris-HCl (pH ~ 8.0), 1 M NaCl, 250 mM imidazole buffer across 20 column volumes. N-His$_6$-Bmp8 purified as a stand-alone purified protein, while Bmp1(TE) was observed in the flow through and wash by SDS-PAGE analysis. Purity of eluted proteins was checked by SDS-PAGE. To cleave the N-His$_6$ tag thrombin was added to purified N-His$_6$-Bmp8 to a final concentration of 1 unit/mg recombinant protein. The cleavage reaction mixture was transferred into dialysis tubing, and incubated overnight in 2 L 20 mM Tris-HCl (pH ~ 8.9), 50 mM KCl, 3 mM DTT buffer at 4 °C. The dialyzed protein was applied to a 5 mL ion exchange Q Sepharose FF column (GE Biosciences) equilibrated in dialysis buffer, and eluted using a linear gradient to 20 mM Tris-HCl (pH ~ 8.9), 1 M KCl buffer. Glycerol was added to the ion exchange fraction to a final concentration of 10% (w/v). Thrombin cleavage was verified by SDS-PAGE analysis (N-His$_6$-cleaved protein retains N-terminal-GSH after cleavage). Purified Bmp8 was concentrated after ion exchange using an Amicon centrifugal filter with >10 kDA cut-off. Bmp8 mutants were
purified in the same way as for Bmp8. Bmp8-H was purified in the same manner as Bmp8, but without co-expression with Bmp1(TE).

**Construction, expression, and purification of Mm_Bmp8 mutants.** Expression constructs for Bmp8 mutants C57S, C82A, Y84F/I, C85A, H88V, and C163S were constructed by PCR point mutagenesis using pET28- *bmp8* as the template. Primers were designed with the motif 5′-[20 nucleotide sequence-modified overlap] [35 nucleotide primer region]-3′ followed by treatment with DpnI exonuclease, propagations in *E. coli* DH5α (NEB), Sanger sequencing verification (SeqXcel, La Jolla, CA), and transformation into *E. coli* BL21-Gold(DE3) (Agilent) for expression. Expression purification, and processing of Bmp8 mutants was performed in an identical manner as previously described for the purification of wild type Bmp8. Note that Bmp8 C57S, Y84F/I, and C163S were insoluble and therefore were not assayed.

**Identification, cloning, mutagenesis, and expression of 2ta16_Bmp8-H and mutants.** 2ta16_Bmp8-H (NCBI accession WP_023399179.1, 55% positives, expected value 2x10^{-35}) was identified in the genome *Pseudoalteromonas luteoviolacea* 2ta16 by a BLASTP search of the genome-sequenced strains available in-house using *Marinomonas mediterranea* MMB-1 Bmp8 (Mm_Bmp8) as a query. Analogous residues to the proposed active site of Bmp8 were identified via a pairwise amino acid alignment using the EMBOSS Needle (EMBL-EBI) web tool. PCR-amplified 2ta16_bmp8-H was cloned into the pET28a expression vector in identical manner to Mm_bmp8 to afford pET28-2ta16_bmp8-H. Expression constructs for
Bmp8-H C78A, Y80F, H84V, and C81A we constructed by PCR mutagenesis of pET28-\textit{bmp8-H} using the same methodology used for Bmp8 mutant constructs. N-His\textsubscript{6}-tagged 2ta16_Bmp8-H and mutants expressed as soluble recombinant proteins without need for co-expression with a helper, but were otherwise expressed, purified, and processed in the same manner as previously described for Mm_Bmp8.

\textit{In vitro} enzymatic activity assaying of Bmp8, 2ta16_Bmp8-H, and mutants. Freshly purified Bmp8, Bmp8-H, or mutants thereof were assayed in 1 mL reactions with an incubation time of 30 min at 30 °C in reaction buffer containing 20 mM Tris-HCl (pH 8), 50 mM KCl, and 10% glycerol. For reactions containing equimolar or excess substrate the enzyme concentration was set at 25 µM. To investigate metal-dependency, Bmp8 was pre-incubated with 1 mM EDTA at 30 °C prior to the addition of substrate. Reactions with enzyme not pre-treated with EDTA were also pre-incubated for 1 h at 30 °C prior to addition of substrate to account for loss of activity due to oxidation by molecular oxygen. Reactions with excess substrate were initiated by addition of 2 or 3 to a final concentration 50 µM (\~{}2-fold excess to enzyme). The volumes of no-enzyme negative control reactions were adjusted with an appropriate concentration of reaction buffer and substrate was added to the same concentration as for the corresponding enzymatic reaction. Enzymes were recycled by addition of glutathione to a final concentration of 1 mM (40-fold excess to enzyme). Reactions were quenched by addition of the whole reaction to 1.9 mL EtOAc followed by mixing by vortex mixing and separation by centrifugation. The organic layer was collected and the solvent removed \textit{in vacuo} at 30 °C. The resulting residue was dissolved in 100 µL MeOH for LC/MS analysis.
LC/MS analysis of *in vitro* enzyme assays. Extracts of *in vitro* enzyme assays were analyzed by LC/MS 30 µL of sample (in MeOH) on a reverse phase C\textsubscript{18} column (Phenomenex Luna, 5 µm, 4.6 × 100 mm) operating on an Agilent 1260 HPLC in tandem to an Agilent 6530 Accurate Mass Q-TOF mass spectrometer. Mass spectra were acquired in negative ionization mode. HPLC solvents used were water + 0.1 % formic acid (A) and MeCN + 0.1 % formic acid (B). The HPLC elution profile was as follows: (initial flow rate 0.5 mL/min) 10% B for 5 min, linear gradient to 70% B over 10 min, linear increase to 80% B over 10 min, (flow rate changed to 0.7 mL/min) linear increase to 100% B over 0.5 min, hold at 100% B for 3 min, linear decrease to 10% B over 0.5 min. Identical injection volumes were used within a given experiment.

Mass spectrometric analysis of trypsin-digested Bmp8-CXXC. Reactions consisting of Bmp8 (50 µM) and 3 (100 µM) or an equivalent of reaction buffer (20 mM Tris-HCl pH ~ 8, 50 mM KCl, 10 % glycerol) in a total volume of 100 µL reaction buffer. Reactions were incubated for 1 h at 30 ºC, flash frozen in an acetone/dry ice bath, and stored at -80 ºC prior to analysis. To minimize ambient oxidation, reactions were thawed immediately prior to injection of 15 µL of the sample onto a C\textsubscript{4} column (Higgins Analytical PROTO 300, 5 µm, 250 x 4.6 mm) operating on an Agilent 1260 HPLC in tandem to an Agilent 6530 Accurate Mass Q-TOF mass spectrometer. Mass spectra were acquired in positive ionization mode. HPLC solvents used were water + 0.1% formic acid (A) and MeCN + 0.1% (B). The
elution profile was as follows (flow rate: 0.7 mL/min): 10% B for 10 min, linear increase to 30% B over 5 min, linear increase to 70% B over 40 min, linear decrease to 10% B over 5 min, linear increase to 100% B over 1 min followed by 2 min at 100% B, decrease to 10% B over 1 min, 10% B for 2 min and 5 min of post-time equilibration. Fragment of interest were search using the extracted ion chromatogram feature in Agilent MassHunter software searching for MS1 masses predicted for predicted trypsin-digest fragments (ExPASy PeptideMass web tool) of interest in a range of ± m/z 0.1.

**Mass spectrometric analysis of trypsin-digested alkylated Bmp8-CXXC.** Reactions consisting of Bmp8 (50 μM) and 3 (100 μM) or an equivalent of reaction buffer (20 mM Tris-HCl pH ~ 8, 50 mM KCl, 10% glycerol) in a total volume of 100 μL reaction buffer. To check the initial redox state of the Bmp8-CXXC, one of two reactions without 3 was immediately quenched with 10 μL of 1 M iodoacetamide, incubated for 30 min at 30 ºC, then flash frozen in acetone/dry ice and store at -80 ºC prior to further work-up. The additional reactions with and without 3 were were incubated for 1 h at 30 ºC then quenched with 10 μL 1 M iodoacetamide, incubated for an additional 30 min at 30 ºC, flash frozen, and stored at -80 ºC. Thawed reactions were digested with 1 μg of proteomics grade trypsin (Sigma) for 1 h, then flash frozen and stored at -80 ºC prior to analysis. LC/MS analysis was performed as previously described above for the analysis of trypsin-digested Bmp8-CXXC without alkylation.
Whole-protein mass spectrometric analysis. Reactions consisting of wildtype Bmp8, Bmp8:C82A, or Bmp8:C85A (50 µM) in buffer (20 mM Tris-HCl pH ~ 8, 50 mM KCl, 10% glycerol, and 1 mM GSH or DTT for reactions containing reductants) were analyzed before and immediately after addition of 3 (100 µM). Additional reactions without 3 were incubated in parallel and analyzed at the end of each sequence to ensure that any modifications to the proteins were due to the addition of 3 rather than the result of extended incubation. Reactions were analyzed by direct injection of 60 µL of the reaction onto an Agilent 1260 HPLC in tandem to an Agilent 6530 Accurate Mass Q-TOF mass spectrometer at an injection speed of 1 mL/min maintaining a flow of water (+0.1 % formic acid) at flow rate of 0.7 mL/min held over 2 min.

Data analysis and plotting. LC/MS data was analyzed and visualized using the Agilent MassHunter software package. For enzyme activity assays, extracted ion chromatograms were called using the calculated m/z for the [M-H]\(^-\) ± 0.1 for the most abundant M+2 isotope corresponding to a give halogenated pyrrole species. For peptides extracted ion chromatograms were extracted survey for the most abundant species predicted m/z for [M+ZH]\(^{Z^+}\) surveyed over a range of charges Z (integer > 0). Whole protein mass spectra were extracted by average over the elution peak over a fixed interval. Mass spectral isotope predictions were obtained using the Agilent Isotope Distribution Calculator tool, which is part of the MassHunter software suite. Data was exported from MassHunter and re-plotted in OriginPro (OriginLab), and figures were assembled in Adobe Illustrator (Adobe). Chromatograms corresponding
to sets of experiments are normalized to the largest peak among spectra, while
individual chromatograms are normalized to the largest peak within a given
chromatogram.
4.4.2 Supplementary Figures

Figure S4.1: Affinity purification of N-His$_6$-Bmp8 by co-expression with Bmp1(TE), and cleavage of N-His$_6$ affinity tag. Bmp8 analyzed by SDS-PAGE gel. Expected molecular weights are as follows: Bmp1(78-376, TE)-33.442 kDa, N-His$_6$-Bmp8-23.086 kDa, and N-GSH-Bmp8-21.566 kDa (N-terminal GSH remains post-cleavage). SDS-PAGE lanes are as follows: (A) Protein molecular weight ladder (Fisher Scientific BP3602), (B) Crude lysate of E. coli co-expressing N-His$_6$-Bmp8 (band boxed in blue) and Bmp1(TE) (band boxed in red), (C) Pre-wash flow-through from Ni$^{2+}$-affinity column, (D) Purified N-His$_6$-Bmp8, (E) Ion-exchange-purified N-His$_6$-tag-cleaved N-GSH-Bmp8.
Figure S4.2. Dechlorination of 2,3,4,5-tetrachloropyrrole catalyzed by Bmp8. (A) Scheme for dechlorination 5 to 2,3,4-trichloropyrrole (regiospecificity inferred from dehalogenation of 3). (B) Combined EICs for predicted m/z [M-H] ions corresponding to molecular formulae C₄NHCl₄ (5) and trichloropyrrole (C₄NH₂Cl₃) for reactions consisting of synthetically prepared 5 (100 µM) incubated with and without excess glutathione (2 mM) with and without Bmp8 (25 µM) in reaction buffer (20 mM Tris-HCl pH ~ 8, 50 mM KCl, 10% glycerol) for 30 min. at 30 ºC. (C) Mass spectra corresponding to the shaded peaks shown in panel ‘B’ exhibiting the anticipate isotopic distribution for tetra- and tri- chlorinated compounds, and illustrating the peak-to-peak mass difference corresponding to loss of chlorine.
Figure S4.3: Bmp8 homologs from (meta)genomic datasets. Alignment (MAFFT) focusing on conserved residues in the vicinity of the active site of Bmp8 (** above indicates residues conserved in this region for all sequences) with reference distance trees (Geneious, Jukes-Cantor, Neighbor-Joining) for top ~100 query hits (BLASTP, expected value cut-off of 10, default parameters) for query *M. mediterranea* MMB-1 Bmp8 (MM_Bmp8) from (A) GenBank non-redundant protein database (nr) and (B) Genbank environmental metagenomics database (env_nr). Bmp8 homologs found in the genetic context of the whole or partial Bmp gene locus are labeled in red and included for reference with the metagenomics hits in panel ‘B’. Bmp-harboring *P. luteoviolacea* 2ta16 includes a second non-Bmp Bmp8 homolog 2ta16_Bmp8-H highlighted in blue and also included in panel ‘B’. Sequences are labeled with GenBank accession numbers, and, in panel ‘A’ are labeled with derivative genera.
Figure S4.4: Bmp8-CXXC is reduced in enzyme as-purified. To verify that the initial redox state of Bmp8-CXXC, intact Bmp8 was treated with iodoacetamide then trypsin-digested, as described in the ‘Materials and Methods.’ (A) EICs for the trypsin digested fraction containing Bmp8-CXXC for the m/z predicted for the [M+2H]^{2+} of the oxidized, unalkylated/reduced, and alkylated fragments. Only the alkylated fragment was detected demonstrating that Bmp8-CXXC is in the fully reduced state as-purified, and that alkylation of the Bmp8-CXXC of intact Bmp8 goes to completion. (B) Mass spectrum corresponding to the depicted [M+2H]^{2+} of the alkylated Bmp8-CXXC-containing digested fragment averaged over the correspond EIC peak in panel ‘A’ (black). The predicted isotope distribution is indicated by the red lines.
Figure S4.5: Conserved Bmp8 Cys57 and Cys163 do not play a direct role in Bmp8 reaction mechanism. The state of the additional Bmp cysteine residues, Cys57 and Cys163 was investigated for Bmp8 incubated with and without 3, alkylated, and trypsin-digested as described in the ‘Materials and Methods.’ (A) Alignment (MAFFT) of Bmp8 homologs found in the genetic context of the Bmp pathway with conserved Mm_Bmp8-Cys57 and Cys163 indicated above. (B) EICs for the predicted m/z of the [M+2H]^2+ ion corresponding to the alkylated trypsin-digested fragment containing Bmp8-Cys57 for Bmp8 incubated with and without 3 with corresponding mass spectra averaged over EIC peaks (black) with predicted isotope distributions indicated by the red lines. (C) EICs for predicted m/z of the [M+H]^1+ ion corresponding to alkylated trypsin-digested fragment containing Bmp8-Cys163 for Bmp8 incubated with and without 3 and corresponding mass spectra averaged over EIC peaks (black) with predicted isotope distributions indicated by the red lines.
Figure S4.6: Whole-protein mass spectra for wildtype Bmp8 and Bmp8 Cys to Ala mutants under various conditions. (A)-(B) Mass spectra for Bmp8-WT (WT), Bmp8-C82A (C82A), and Bmp8-C85A (C85A) before and after treatment with 3. For spectra labeled with ‘start’ or ‘end’, ‘start’ refers to the sample before addition of 3, while ‘end’ refers to a separate reaction without 3 incubated in parallel as a negative control intended to demonstrate that observed mass shifts are due to the addition of 3 as opposed to extended incubation times. Charge states corresponding to each population are indicated at the top of panels ‘A’ and ‘B’, while masses predicted for the maximum isotope for each [M+ZH]^+ ion are indicated by the red lines for the ‘apo’ form, and blue lines for the ‘holo’ adduct-bound (+SG or +2[O]) forms; the +2SG predicted mass for ‘WT +3’ is indicated in green. (C) Calculation of the mean peak-to-peak difference ($\overline{\Delta_{max}}$) between ‘holo’ adduct-bound Bmp8-C82A and ‘apo’ Bmp8-C82A taken as the mean of differences for a given charge Z ($\Delta_{max,Z}$) over Z = 16 to 26; values are shown in Table S1. (D)-(E) Proposed adducts formed with Bmp8:C82A and Bmp8:WT for different treatments based on calculated $\overline{\Delta_{max}}$ in Table S1.
**Figure S4.7: A distant Bmp8 homolog from *P. luteoviolacea* 2ta16, 2ta16_Bmp-H.** 

(A) Genetic context for the 2ta16_Bmp8 homolog (2ta16_Bmp8-H) from *P. luteoviolacea* 2ta16 (GenBank WP_023399179.1). (B) 2ta16 Bmp8-H and Mm_Bmp8 were incubated with 3 as described in the ‘Materials and Methods’ and reactions were quenched by extraction with two volumes at EtOAc after incubation for 0.5 or 3 h. The figure shows combined EICs for predicted of \( m/z \)’s of [M-H]\(^{1-}\) ions for 2, 3, and 4 for extracts analyzed as described in the ‘Materials and Methods.’

**Figure S4.8: Functional characterization for Bmp8 homologs from Bmp pathways harbored by two additional marine bacteria.** (A) Bmp gene cluster from *P. luteoviolacea* 2ta16 (2ta16_bmp), *P. phenolica* O-BC30 (OBC30_bmp), and *M. mediterranea* (Mm_bmp). (B) Activity of Bmp8 homologs expressed and purified as described for Mm_Bmp8 in the materials and methods for enzyme incubated with ∼2-fold excess of 3 for a reaction time of 30 min. Traces represent combined EIC’s for the predicted [M-H]\(^{1-}\) for 3 (peak shaded green) and 2 (peak shaded orange).
Figure S4.9: Bmp8 selectively debrominates a differentially halogenated pyrrole substrate. (A) Route for the synthesis of 2-chloro-3,4,5-tribromopyrrole (Br$_3$ClPy) from 2. Briefly, to 2 (10 mg) in 15 mL THF was added 1.1 eq. NCS while stirring. The reaction was stirred for 3 h at room temperature, and solvent was removed *in vacuo*. The residue was dissolved in minimal hexane and purified by silica gel chromatography. The predicted $m/z$ for [M-1] ion for C$_4$NHBr$_3$Cl (Br$_3$ClPy) is 333.7275, measured is 333.72120. (B) EICs for reactions containing Bmp8 incubated with equimolar substrate (Br$_3$ClPy or 3 as a positive control) in the same manner as described for *in vitro* enzyme assays in ‘Materials and Methods’. The top two traces are combined EICs for predicted $m/z$ for [M-H]$^{-1}$ ions corresponding to 2, Br$_3$ClPy, and Br$_2$ClPy. The bottom two traces or EICs for predicted $m/z$ for [M-H]$^{-1}$ corresponding to 2 and 3. To the right of traces are mass spectra averaged over the respective highlighted peaks in EICs consistent with expected M+2 Cl/Br isotopic distributions. Only a single with $m/z$ 255.8165; predicted $m/z$ for the [M-H]$^{-1}$ ion for the singly debrominated product (i.e. with molecular formula C$_4$NH$_2$Br$_2$Cl) is 255.8169. (C) Scheme illustrating selective dehalogenation of Br$_3$ClPy by Bmp8.
Figure S4.10: Mass spectra for Bmp8 products. Mass spectra showing debromination of 3 (tetrabromopyrroles) to 2 (tribromopyrrole), and 2 to 4 (dibromopyrrole).
Figure S4.11: Regeneration of Bmp8. (A) Scheme for oxidation and reduction of Bmp8. i. Bmp8 as-purified was treated with an excess of tetrabromopyrrole (Br$_4$Py, 1) for 3 h, followed by ii. dialysis in buffer (20 mM Tris-HCl pH ~ 8, 50 mM KCl, 10% glycerol) containing dithiothreitol (+DTT) to regenerate the enzyme, and then iii. removal of DTT from the buffer by dialysis in the same buffer not containing DTT (-DTT). 2-chloro-3,4,5-tribromopyrrole (Br$_3$ClPy) was used as a substrate in case of carryover of Br$_4$Py used in excess to oxidize Bmp8, and, in principle, removed by dialysis. Br$_3$ClPy (equimolar to enzyme) was incubated with for 3 h at 30 °C with regenerated Bmp8 (‘+Bmp8 regenerated’), a blank buffer control treated in the same manner as regenerated Bmp8 (‘+DTT blank dialyzed’), dialyzed not regenerated Bmp8 (i.e., dialyzed twice in –DTT buffer instead of once in +DTT buffer and once in –DTT buffer, ‘+Bmp8$_{ox}$’), an as-purified Bmp8 positive control (‘+Bmp8$_{red}$’), or no Bmp8 (‘-Bmp8’). (B) Extracted ion chromatograms (EICs) for organic extract of reactions for predicted m/z’s of Br$_3$ClPy and ‘-Br’ 2-chloro-3,4-tribromopyrrole (Br$_2$ClPy). Reactions were extracted and analyzed in the same manner as described in the Materials and Methods section.
Figure S4.12: Oxygen sensitivity of Bmp8. Extracted ion chromatograms (EICs) for the predicted m/z’s [M-H]- for tetrabromopyrrole (Br4Py, 3) and tribromopyrrole (Br3Py, 2), and dibromopyrrole (Br2Py, 4). Reactions with equimolar Br4Py to with Bmp8 as-purified (‘D1’) and incubated overnight in a vial with the lid left open to allow for equilibration with air (‘D2’) with and without dithiothreitol (DTT). Reactions were incubated for 6h at 30 °C, and extracted and analyzed as previously described.
Figure S4.13: Reductive dehalogenation of tetrabromopyrrole (Br₄Py, 3) by Bmp8 under anaerobic conditions with and without dithiothreitol (DTT). Extracted ion chromatograms (EICs) for the predicted m/z’s [M-H] for Br₄Py and tribromopyrrole (Br₃Py, 2), and dibromopyrrole (Br₂Py, 4). Bmp8 and Br₄Py (~2-fold excess to enzyme) were incubated for 12 h in open vials in anaerobic chamber to evacuate the solutions of molecular oxygen. Solutions were then combined while still under anaerobic conditions and incubated for an additional 6 h at ambient temperature in the same anaerobic chamber (~25 ºC). Reactions were extracted in the anaerobic chamber using the same reaction procedure as described in the Materials and Methods sections. Reaction conditions and analysis are the same as described in Material and Methods section.
Figure S4.14: Non-enzymatic aqueous reductive dehalogenation of tetrabromopyrrole (Br₄Py, 3) by reducing agents. Extracted ion chromatograms (EICs) for the predicted m/z’s [M-H] for Br₄Py, tribromopyrrole (Br₃Py, 2), and dibromopyrrole (not observed). Br₄Py (100 µM) was incubated for 12 h at 3 °C in aqueous buffer (20 mM Tris-HCl pH ~ 8, 50 mM KCl, 10% glycerol) with 2 mM of reductants (glutathione, GSH; dithiotreitol, DTT; tris(2-carboxyethyl)phosphine, TCEP) or no reductant. Reactions were extracted and analyzed as described in the Material & Methods section (above).
Figure S4.15: Whole-protein mass spectrometry of Bmp8-C82A with other halogenated pyrrole substrates. Bmp8-C82A was incubated with additional Bmp8 substrates 2 and 5. A mass shift is only observed upon addition of 3, consistent with slow turnover of 2 and 5 by Bmp8. The procedure for whole-protein mass spectrometric experiments is described in the Material and Methods.
### 4.4.3 Supplementary Tables

**Table S4.1. Mass shift calculation from whole-protein mass spectrometry analysis.**

<table>
<thead>
<tr>
<th>Description</th>
<th>Calculated neutral protein mass (Da)*</th>
<th>Expected neutral mass for 'apo' protein (Da)</th>
<th>Difference from 'start' (Da)</th>
<th>Mass shift assignment</th>
<th>Expected shift (Da)</th>
<th>Mean peak-to-peak difference (Da)**</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bmp8:WT</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No reductant</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no 3 start</td>
<td>21,565.77±0.27</td>
<td>21,566.40</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>after 3</td>
<td>21,563.13±1.62</td>
<td>-</td>
<td>-2.64</td>
<td>-2H</td>
<td>-2.02</td>
<td>-2.64±1.57</td>
</tr>
<tr>
<td>no 3 end</td>
<td>21,565.57±0.26</td>
<td>21,566.40</td>
<td>0.20</td>
<td>No shift</td>
<td>0</td>
<td>-0.20±0.36</td>
</tr>
<tr>
<td><strong>GSH</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no 3 start</td>
<td>21,567.75±0.22</td>
<td>21,566.40</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>after 3</td>
<td></td>
<td>21,565.68±0.5</td>
<td>-</td>
<td>-2.07</td>
<td>-2H</td>
<td>-2.07±0.57</td>
</tr>
<tr>
<td>peak 1</td>
<td></td>
<td>21,872.66±1.37</td>
<td>-2.07</td>
<td>+GS</td>
<td>305.07</td>
<td>304.92±1.38</td>
</tr>
<tr>
<td>after 3</td>
<td></td>
<td>22,178.09±0.84</td>
<td>610.34</td>
<td>+2GS</td>
<td>610.14</td>
<td>610.34±0.83</td>
</tr>
<tr>
<td>peak 3</td>
<td></td>
<td>21,567.71±0.32</td>
<td>610.34</td>
<td>+2[O]</td>
<td>31.99</td>
<td>-0.04±0.34</td>
</tr>
<tr>
<td>no 3 end</td>
<td></td>
<td></td>
<td>610.34</td>
<td>+2[O]</td>
<td>31.99</td>
<td>-0.04±0.34</td>
</tr>
<tr>
<td><strong>Bmp8:C82A</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No reductant</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no 3 start</td>
<td>21,534.32±0.19</td>
<td>21,534.40</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>after 3</td>
<td>21,566.33±0.58</td>
<td>-</td>
<td>32.61</td>
<td>+2[O]</td>
<td>31.99</td>
<td>32.61±0.70</td>
</tr>
<tr>
<td>no 3 end</td>
<td>21,534.04±0.15</td>
<td>-</td>
<td>-0.28</td>
<td>No shift</td>
<td>0</td>
<td>-0.28±0.23</td>
</tr>
<tr>
<td><strong>GSH</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no 3 start</td>
<td>21,535.04±0.23</td>
<td>21,534.40</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>after 3</td>
<td>21,840.23±0.23</td>
<td>-</td>
<td>305.20</td>
<td>+GS</td>
<td>305.07</td>
<td>305.20±0.33</td>
</tr>
<tr>
<td>no 3 end</td>
<td>21,534.65±0.43</td>
<td>21,534.40</td>
<td>-0.38</td>
<td>No shift</td>
<td>0</td>
<td>-0.38±0.42</td>
</tr>
<tr>
<td><strong>DTT</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no 3 start</td>
<td>21,535.08±0.13</td>
<td>21,534.40</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>after 3</td>
<td>21,535.24±1.07</td>
<td>-</td>
<td>0.15</td>
<td>No shift</td>
<td>0</td>
<td>0.15±1.08</td>
</tr>
<tr>
<td>no 3 end</td>
<td>21,534.91±0.25</td>
<td>21,534.40</td>
<td>-0.18</td>
<td>No shift</td>
<td>0</td>
<td>-0.18±0.12</td>
</tr>
<tr>
<td><strong>Bmp8:C85A</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No reductant</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no 3 start</td>
<td>21,534.07±0.21</td>
<td>21,534.40</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>after 3</td>
<td>21,533.85±0.20</td>
<td>-</td>
<td>-0.22</td>
<td>No shift</td>
<td>0</td>
<td>-0.22±0.3</td>
</tr>
<tr>
<td>no 3 end</td>
<td>21,533.95±0.20</td>
<td>21,534.40</td>
<td>-0.13</td>
<td>No shift</td>
<td>0</td>
<td>-0.13±0.34</td>
</tr>
<tr>
<td><strong>GSH</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no 3 start</td>
<td>21,534.72±0.07</td>
<td>21,534.40</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>after 3</td>
<td>21,534.68±0.12</td>
<td>-</td>
<td>-0.04</td>
<td>No shift</td>
<td>0</td>
<td>-0.04±0.10</td>
</tr>
<tr>
<td>no 3 end</td>
<td>21,534.63±0.13</td>
<td>21,534.40</td>
<td>-0.09</td>
<td>No shift</td>
<td>0</td>
<td>-0.09±0.15</td>
</tr>
</tbody>
</table>

* Calculated as the average of Zx([M+ZH]^Z^-)-ZxH over Z=[16,26]. **Calculation shown in Fig. S6.
Chapter 4, in full, is currently being prepared for submission of the material.

El Gamal A., Agarwal V., Rahman I., and Moore B.S. The dissertation author is the primary investigator and author of this paper.

A.E. designed and carried out experiments, analyzed data, and wrote the paper; V.A., I.R., and B.S.M provided feedback on the paper and experimental design.
Chapter 5: Opportunities in halopyrrole biosynthesis

5.1 Lessons from pentabromopseudulalin biosynthesis: the halopyrrole biosynthetic handle

The elucidation of the brominated marine pyrrole/phenol (Bmp) biosynthetic pathway is a testament to the tremendous opportunities presented by the study of marine bromination. Indeed, awaiting discovery in the biosynthetic pathway of pentabromopseudulalin were three new enzymes catalyzing transformations involving halogens, including the first two examples of brominases from a confirmed biosynthetic context, as well as the first example of a tailoring dehalogenase. Reflecting on the process for the discovery of the Bmp gene cluster, it is worth highlighting how an established halopyrrole biosynthetic logic proved a powerful tool for teasing out a basis for associated bromophenol biosynthesis. One of the main challenges in mining for biosynthetic gene clusters encoding the biosyntheses of non-canonical natural products (i.e. neither NRPS nor PKS) is the lack of a biosynthetic “handle” for genome mining. Hence, the success in applying this halopyrrole probe begs the question of what other opportunities await. One area of great potential lies in unlocking the biosynthesis of the fabled bromopyrrole-containing pyrrole imidazole alkaloid (PIA) class of sponge natural products for which more than 150 structures have been described over the past forty years, and which have captivated the imaginations of synthetic chemists and pharmacologists alike.\(^\text{82}\). Feeding studies have shown incorporation of radiolabeled L-proline into the structure of the proposed PIA building block-oroidin. Hence, it would appear that the halopyrrole biosynthetic logic could strike again this time in unlocking the biosynthesis of the amino imidazole
moiety of oroidin, and thereby providing access to an entire class of natural products. The remainder of this chapter is dedicated to a description of the state-of-the-art and challenges in the elucidation of PIA biosynthesis along with an original example of sponge metagenome mining toward elucidation of a PIA gene locus.

5.2 A foray into pyrrole-imidazole alkaloid biosynthesis

5.2.1 Background and Introduction

Pyrrole-imidazole alkaloids (PIAs) are a diverse class of marine sponge secondary metabolites with promising bioactivities, including cytotoxic, anti-microbial, and anti-feedant properties. Among the over 150 PIAs described to date are the anti-microbial sceptrins, anti-fungal and anti-tumor palau’amine, and anti-tumor agelastatins, found in the sponge genera *Agelas, Stylissa, Stylotella, Axinella*, and *Hymeicaedon* (Fig. 5.1). Despite their promising potential as drug leads, access to complex PIAs has been a bottleneck to their development as drugs.
Figure 5.1: Packing a punch with parsimony. Complex PIAs are thought to arise from universal achiral building blocks.

Exemplifying biosynthetic parsimony, complex multi-stereocenter PIAs are proposed to originate from one of three achiral PIA building blocks, namely, oroidin, hymenidin, and clathrodin, which differ only in the nature of the substitution of the 4 and 5 positions of the pyrrole moiety (Fig. 5.1)\textsuperscript{110}. For example, palau’amine isolated from the sponge *Stylotella agminata*, whose hexacyclic structure features eight contiguous stereocenters, is a proposed dimer of achiral clathrodin (Figs. 5.1 and 5.2)\textsuperscript{111,112}. Moreover, biomimetic studies employing cell-free assays from crude sponge enzyme preparations from tissues of the Caribbean sponge species *Agelas sceptrum* and *Stylissa caribica* demonstrated oroidin as a precursor to the more complex bioactive PIA benzosceptrin C\textsuperscript{113}. Furthermore, the biogenetic precursors of the pyrrole moiety of oroidin is posited from feeding studies to be the amino acid L-proline, while L-homoarginine via L-lysine (or L-histidine) is proposed as the
precursor of the aminopropylimidazole moiety\textsuperscript{114}. As in the identification of the \textit{bmp} gene cluster (Chapter 2), the L-proline adenylation domain provides a facile basis for mining for potential PIA biosynthetic gene clusters.

\begin{center}
\includegraphics[width=0.5\textwidth]{figure5_2.png}
\end{center}

\textbf{Figure 5.2: Palau’amine: simplicity in complexity.} As per biosynthetic proposal, clathrodin moiety (red) and clathrodin sub-unit 3-amino-1-(2-amino imidazolyl) prop-1-ene moiety (green).

Understanding of PIA biosynthesis has progressed slowly over the last decades due to inherent limitations associated with traditional approaches of study. Indeed, the state-of-the-art in the study biosynthesis of PIAs has until now relied on indirect chemical techniques such as \textit{in vivo} isotope tracer experiments, and \textit{in vitro} crude enzyme assays\textsuperscript{110,113,115,116}. While these studies have suggested the chemical building blocks involved in higher PIA biosynthesis, they fall short in addressing the genetic and biochemical bases for their production. Moreover, despite monumental successes in the total syntheses of many complex PIAs such as palau’amine, synthetic chemistry has thus far proven an impractical mean for gaining access to PIA material\textsuperscript{117-120}. A conclusive molecular basis for PIA biosynthesis is key not only to the discovery of new enzymes, but also to gaining access to these valuable molecules through exploitation of nature’s synthetic toolkit.
5.2.2 Approach

![Figure 5.3: Scheme for metagenomic approach for elucidation of PIA biosynthesis.](above)

Experimental validation will afford a handle to accelerate the rate of discovery of new pathways.

Sponges harbor complex microbiomes that are thought to account for the majority of their secondary metabolism\(^{121}\). At least a portion of sponge microbiota are known to be vertically transferred from adult sponges to their offspring during spawning\(^{122}\). Hence, it is this core microbiome that likely gives rise to the highly reproducible chemotypes observed within sponge species\(^{123,124}\). Nonetheless, it cannot be excluded that the sponge itself might also possess the capacity to produce natural products, or at least modify the natural products produced by its microbiome. Hence, a deep-sequencing metagenomics approach which examines the entire genetic content of the sponge organism may provide the best first-pass approach for gaining a holistic understanding of the biosynthetic capacities of these organisms\(^{125}\). The initial workflow described in ‘Fig. 5.3’ (above) begins with identification and collection of sponges known to harbor PIAs, followed by chemical verification to confirm presence of PIAs. Validated specimens are then subjected to whole-animal metagenomic DNA extraction and deep-sequencing. Sequencing data is then assembled and annotated, and queried using candidate amino acid sequences based on retro-biosynthetic hypotheses. Upon identification of candidate gene loci, one might
proceed to validate targets through a combination of *in vivo* heterologous expression of a full gene cluster, and *in vitro* interrogation of individual biochemical transformations through the cloning and expression of individual enzymes. The verification of the first PIA biosynthetic pathway promises to accelerate the discovery of many more PIA pathways (Fig. 5.3). The remainder of this section will discuss preliminary findings utilizing this approach.

### 5.3 Preliminary work

#### 5.3.1 Collection of marine sponges harboring PIAs

The two fundamental questions motivating an initial study of PIA biosynthesis are, firstly, the origin of oroidin, and, secondly, whether more complex “higher” PIAs are derived from oroidin. Hence, we initially sought a sponge harboring oroidin, and another sponge harboring the simplest of the higher PIAs, the oroidin dimer sceptrin (Fig. 5.1). Moreover, sponges from distinct geographical locations were sought with the rationale that a core microbiome, independent of environment, might be derived via comparison of these two metagenomes thus focusing the search for candidate biosynthetic gene loci. To this end, the sponge *Stylissa massa* known to produce oroidin and cyclooroidin derivatives\(^ {126} \) was collected in Guam by a team lead by Dr. Vinayak Agarwal of the Moore Laboratory in the summer of 2014, while the Caribbean sponge *Agelas tubulata* shown to harbor sceptrins was obtained from collaborators at Sirenas Marine Discovery in La Jolla, CA. Hence, the rationale for choosing initial target sponges for investigation is born of a “simple-to-complex” strategy.
5.3.2 Chemical validation

Initial chemical validation of samples was performed by liquid chromatography tandem mass spectrometry (LC-MS) analysis of methanol extracts of *A. tubulata* and *S. massa*. Extracts of *S. massa* were confirmed to harbor masses consistent with anticipated oroidin and cyclooroidin derivatives (Figs. 5.4 and 5.5), while extracts of *A. tubulata* were confirmed to contain masses consistent with the oroidin dimer sceptrin (personal correspondence with Eduardo Esquinazi, Sirenas Marine Discovery).

Figure 5.4: Chemical diversity of *S. massa*. Representative structure of alkaloids furnished by extracts of the sponge *S. massa*.

---

**Oroidin**

**4.5-dibromopyrrole-2-carbonamide**

**Bromooidisine**

**Hymenin**

**10α-l-hymanaldisine**

**10β-hymanaldisine**

**10β-debrornolhymanaldisine**

**10α-debrornolhymanaldisine**
Figure 5.5: LC-MS validation of production of oroidin and cycooroidin compounds from *S. massa* collected in Guam. (A)-(E) Mass spectra with associated molecular formulae and proposed structures (also shown in Fig. 4) based on previous isolation work with *S. masssa*\textsuperscript{126}. Extracted ion chromatograms are shown to the tops of panels (A)-(B) for compounds with identical molecular formulae, and for which two peaks were present. Multiple known structures fit to the mass spectrum shown in panel ‘E’. The vertical axis is relative abundance. Structural assignments are entirely speculative based on predicted molecular formulae.

5.3.3 Metagenomes

Whole sponge metagenomic DNA was prepared from tissue of *A. tubulata* and *S. massa* using a protocol described by Brady *et al.*\textsuperscript{127} Metagenomes were deep-sequenced using the Illumina HiSeq sequencing platform. The resulting metagenomes were assembled and uploaded onto the JGI server for annotation. These metagenomes both exhibited remarkable microbial diversity. Comparison of the *S. massa* and *A. tubulata* prokaryotic diversity (Bacteria and Archaea) showed that the metagenomes were highly similar at a phylum level (Fig. 5.6).
Figure 2. Comparison of phylum-level prokaryotic diversity of *A. tubulata* and *S. massa*. Radial phylogenetic tree showing the phylum-level diversity based on 16S metagenomics reads. Bars associated with nodes show the abundance of operational taxonomic units (OTUs) for the respective metagenome.
5.3.4 Metagenomic query

Figure 5.7: Return of the halopyrrole biosynthetic handle. L-proline is thought to be the precursor in the biosynthesis of the halopyrrole moiety, therefore canonical halopyrrole biosynthetic motifs might be used to mine sponge metagenomes for PIA biosynthetic loci.

Based on reports of incorporation of radiolabeled L-proline into the structure of oroidin\textsuperscript{114}, it was hypothesized that the halopyrrole moiety of PIAs is derived via the same conserved halopyrrole biosynthetic logic used to mine bacterial genomes for the Bmp gene locus (Chapter 2) (Fig. 5.7). Therefore, proline adenylation domains and pyrrole halogenases were employed as a two-pronged hook to search for potential PIA clusters. Mining the metagenome of \textit{S. massa} lead to the identification of several short contigs containing putative proline adenylation domains, while no homologs to pyrrole halogenases could be identified. A search of the \textit{A. tubulata} metagenome similarly revealed several proline adenylation domains, but, likewise, their context could not be ascertained due the brevity of the contigs on which they were found. Unlike the metagenome of \textit{S. massa}, the metagenome of \textit{A. tubulata} returned some twenty homologs of flavin-dependent pyrrole halogenases. The most closely related homolog showed greater sequence homology to halogenases that act on thiotemplated rather than free substrates (Fig. 5.8). Further, a search of publically available genomic
databases using the putative pyrrole halogenase sequence from *A. tubulata* returned partial sequences for putative flavin-dependent halogenases amplified from metagenomes of sponges known to harbor PIAs in an earlier study.\(^{129}\)

![Phylogeny of putative *A. tubulata* pyrrole halogenase.](image)

**Figure 5.8: Phylogeny of putative *A. tubulata* pyrrole halogenase.** An unrooted distance tree was constructed (Jukes-Cantor, Neighbor-Joining, Geneious) for flavin-dependent halogenases including a pyrrole halogenase homolog from the metagenome of *A. tubulata* (‘Atub’, highlighted orange); the genetic context for ‘Atub’ is shown in ‘Fig. 25’.

Inspection of the gene neighborhood for the putative *A. tubulata* halogenase revealed a gene annotating as a CoA-ligase, which could serve in place of an adenylation domain (Fig. 5.9). Also present in the gene neighborhood were several genes encoding for enzymes catalyzing redox and radical chemistry (Fig. 5.9). Resistance mechanisms for natural products are often encoded their biosynthetic gene clusters.\(^{130}\) In addition to apparent biosynthetic enzymes, the cluster from *A. tubulata*
also encodes for putative topoisomerase IV. Interestingly, analogs of oroidin have been shown to be inhibitors of DNA topoisomerase IV\textsuperscript{131}. As a thought exercise, a cursory biosynthetic scheme for oroidin (the putative precursor of sceptrin harbored by \textit{A. tubulata}) was constructed based on the annotations found in the genetic context of the putative \textit{A. tubulata} halogenase (Fig. 5.9). Indeed, many biosynthetic steps leading to the formation of sceptrin, let alone oroidin, are apparently absent from the annotated genetic context captured by the \(\sim\) 21 kb contig shown in Figure 8. Additionally, the cluster contains many ORFs that lack annotations due dissimilarity from known enzymes. Therefore, before proceeding with detailed investigation of this gene cluster it the experimental validation of key transformations is advised. Several approaches for biochemical validation are described briefly in the following section.
Figure 5.9: Putative “PIA” cluster from the *A. tubulata* metagenome. The top of the figure shows the ~21 kb contig containing a putative *A. tubulata* pyrrole halogenase (orange) chosen for its promising genetic context. Annotations for apparent biosynthetic and resistance genes are shown according to the coloring scheme used in the gene map. ORFs shaded are annotated to encode “hypothetical proteins.” Based on these annotations a rough biosynthetic proposal was constructed for oroidin (note that *A. tubulata* harbors a putative oroidin dimer sceptrin). Red font indicates implicit transformations that could not be accounted for by gene annotations. The biosynthetic scheme is inspired by that shown in ref. 7. The transformation associated with the annotation “coproporphyrinogen III oxidase-like radical SAM enzyme” is shown in the dashed box (bottom, right). Note this biosynthetic scheme is included as a thought experiment, and is not experimentally validated.

5.3.4 Validation

Thus far a metagenomics approach has failed to reveal a viable PIA biosynthetic gene cluster for heterologous expression. Nonetheless, despite lack of sufficient genetic context in many cases, one might begin to screen candidate pyrrole halogenases using the methodology described in Chapter 3 of this thesis to study the reaction catalyzed by pyrrole brominase Bmp2. Moreover, the specificity of candidate proline adenylation domains can be assayed\(^\text{132}\).

5.4 Future directions

In light of the challenge thus far in identifying a high-confidence PIA biosynthetic gene locus via a metagenomics strategy, it is worth revisiting the
assumptions driving the present approach. The first assumption is that the halopyrrole moiety of PIAs derives from a conserved biosynthetic logic. This proposal is based on past studies showing incorporation of radiolabeled L-proline into oroidin\textsuperscript{114}. However, adherence to convention comes with the risk of turning a blind eye to alternative strategies. Indeed, the \textit{A. tubulata} metagenome-derived gene cluster used as an example in this chapter appears to possess a halogenase and CoA-ligase, but lacks an open reading frame (ORF) annotating as an acyl carrier protein (ACP) despite a substantial (~21 kb) genetic context. Hence, for the sake of argument, the biosynthesis of the halopyrrole moiety of oroidin might arise via an ACP-independent pathway, which is without precedent in halopyrrole biosynthesis\textsuperscript{133}. The second assumption is that the sponge microbiome, rather than the sponge itself is the producer of the major chemistry present in the sponge. The enormous capacity for secondary metabolism of host-associated microorganisms has been demonstrated from marine invertebrates\textsuperscript{134,135}. Nonetheless, if the sponge itself is responsible for whole or part of the biosynthesis of PIAs, a transcriptomic approach such as used to identify dispersed biosynthetic enzymes in plants by correlative gene expression might be more appropriate\textsuperscript{136}. As little is known about sponge enzymes, and relatively few genomes of marine invertebrates have been sequenced, relying on annotations to identify biosynthetic genes could pose an additional challenge\textsuperscript{137}. In light of the inherent risk involved in the pursuit of unknowns, the assumptions going into the approach to PIA biosynthesis should be thoroughly vetted. While the success of high risk research often hinges on a degree of chance, chance certainly favors a well-rationalized approach.
This chapter does not mean to suggest that a metagenomics approach to the elucidation of PIA biosynthesis is destined to failure, but that improvements can be made to increase the odds of success. It is apparent from initial sequencing efforts, that the metagenomes generated for *A. tubulata* and *S. massa* are exceedingly complex. Hence, a strategy that reduces the complexity of DNA samples prior to sequencing could dramatically increase the efficiency of downstream steps in the workflow. One strategy could be a single-cell genomics approach as has previously been used to identify sponge symbionts with enormous biosynthetic potential. Assuming limiting sequencing and computational resources, a single-cell approach would require a means of prioritizing amplified genomes for sequencing. In turn, this would hinge on confidence in the retrobiosynthetic proposal for oroidin which would word serve as the basis for designing degenerate nucleic acid probes for initial screening of single microbial cells. An additional strategy for simplifying metagenomes would be to sequence metagenomes of sponge eggs or larvae, which would ostensibly be enriched for vertically rather than horizontally transferred microbiota, hence increasing chances of capturing intact genomic information for the producing microbial symbiont. Despite evidence for vertical transfer of sponge symbionts, a metagenomics approach has never been taken to study this phenomenon. A traditional biochemical approach that casts a wider net in screening for crude enzyme activity coupled with a modern genomics approach could also serve to inform and specify a broader view on oroidin biosynthesis. Indeed, crude sponge protein extracts have been used to generate circumstantial evidence in support of the role of oroidin as a precursor to higher PIAs. Based on the dipeptide-
like structure of oroidin, it is reasonable to assume that its building blocks derive from amino acid precursors, hence, screening protein fractions for modification of a cocktail of amino acids utilizing a mass spectrometry networking approach is a highly feasible prioritization scheme toward the generation of protein sequence search tags. The study of the biosynthesis of PIAs is a daunting task, however any progress toward the identification of a genetic biosynthetic signature for a PIA molecule promises to crack open the biosynthesis of a class of over 150 natural products, whose bioactivities have amazed pharmacologists, and inspired synthetic chemists\textsuperscript{82}.

5.5 Acknowledgements

Many thanks to Dr. Vinayak Agarwal, Michelle Schorn, Dr. Valerie J. Paul (Smithsonian Marine Station), and Prof. Jason Biggs (University of Guam) for collecting \textit{Stylissa massa} from Guam, to Dr. Sheila Podell and Prof. Eric Allen for sequencing and assembly of the metagenome of \textit{A. tubulata}, to Dr. Zhenjian Lin and Prof. Eric Schmidt (University of Utah) for sequencing and assembly of the metagenome of \textit{S. massa}, and to Dr. Eduardo Esquinazi and colleagues at Sirenas Marine Discovery (La Jolla, CA) for collection and chemical analysis of \textit{A. tubulata}.
References


Technical Fact Sheet –Polybrominated Diphenyl Ethers (PBDEs) and Polybrominated Biphenyls (PBBs). (2014).

Kimbrough, K. L., Johnson, W. E., Lauenstein, G. G., Christensen, J. D. & Apeti, D. A. An Assessment of Polybrominated Diphenyl Ethers (PBDEs) in Sediments and Bivalves of the U.S. Coastal Zone. 87 (Silver Spring, MD, 2009).


