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

Abrahim El Gamal[a,2], Vinayak Agarwal[a,2], Stefan Diethelm[a], Imran Rahman[a], Michelle Schorn[a], Jennifer M. Sneed[b], Gordon V. Louie[c], Kristen E. Whalen[d], Tracy J. Mincer[d], Joseph P. Noel[c], Valerie J. Paul[b], and Bradley S. Moore[a,e,1]

[a]Center for Oceans and Human Health, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093, USA

[b]Smithsonian Marine Station at Fort Pierce, Fort Pierce, Florida 34949, USA

[c]Howard Hughes Medical Institute, Jack H. Skirball Center for Chemical Biology and Proteomics, The Salk Institute for Biological Studies, La Jolla, California 92037, USA

[d]Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

[e]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, California 92093, USA

[1]To whom correspondence should be addressed: bsmoore@ucsd.edu, Tel: 858-822-6650, Fax: 858-534-1318

[2]These authors contributed equally to this work.

Classification: Biological Sciences, Biochemistry

Keywords: Natural products, Biosynthesis, Halogenation, Enzymology

Manuscript information: 15 pages, 6 Figures.

Word and character count: 250 words in abstract, total character count ~ 39000

Refs: 33

ABSTRACT

Halogenated pyrroles (halopyrroles) are common chemical moieties found in bioactive bacterial natural products. The halopyrrole moieties of mono- and di- halopyrrole-containing compounds arise from a conserved mechanism in which a proline-derived pyrrolyl group bound to a carrier protein is first halogenated then elaborated by peptidic or polyketide extensions. This paradigm is broken during the marine pseudoalteromonad bacterial biosynthesis of the coral larval settlement cue tetrabromopyrrole (**1**), 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 **1**, we sequenced two *Pseudoalteromonas* genomes and identified a conserved four-gene locus encoding the enzymes involved its complete biosynthesis. Through total *in vitro* reconstitution of the biosynthesis of **1** using purified enzymes and biochemical interrogation of individual biochemical steps, we show that all four bromine atoms in **1** are installed by the action of a single flavin-dependent halogenase- Bmp2. Tetrabromination of the pyrrole induces a thioesterase-mediated offloading reaction from the carrier protein and activates the biosynthetic intermediate for decarboxylation. Insights into the tetrabrominating 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 **1**, and sets a firm foundation for querying the biosynthetic potential for the production of **1** in marine (meta)genomes.

SIGNIFICANCE STATEMENT

The majority of pharmaceuticals are inspired by natural product scaffolds that are functionalized by tailoring enzymes, such as halogenases. Degree of halogenation is an important determinant of natural product bioactivity, yet little is known regarding the molecular basis for the exquisite control exhibited by tailoring halogenases. Known pyrrole halogenases commonly perform up to two halogenations on the pyrrole. Our study of tetrabromopyrrole biosynthesis revealed a uniquely adapted halogenase-thioesterase enzyme pair that catalyzes an unprecedented series of halogenations on a pyrrole. Structural comparison of the pyrrole tetrahalogenase to a pyrrole dihalogenase revealed key residues involved in controlling degree of halogenation. Our findings provide fundamental insights that might be applied in the rational design of biocatalysts toward directed biosynthesis of new chemicals.

\body

INTRODUCTION

Marine bacteria of the genus *Pseudoalteromonas* 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 *Pseudoalteromonas* 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 organohalogens (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 halopyrrole is extended using modular assembly lines to yield non-ribosomally synthesized peptide, or polyketide natural products (e.g., **2–4**, **Fig. 1**) (7-10). During the biosynthesis of **2–4**, and other related natural products, the halopyrrole carboxylic acid is transthioesterified to downstream ACPs during modular elongation reactions. Molecules arising from these modular biosynthetic pipelines are characterized by the preservation of the prolyl alpha-carboxyl 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-carboxyl carbon atom is lost during the biosynthesis of **1** and the cytotoxic marine bacterial natural product pentabromopseudilin (**5**, **Fig. 1**) (4, 11). Uniquely, **5** is biosynthesized via the coupling of 2,3,4-tribromopyrrole (**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-carboxyl carbon *en route* the biosyntheses 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. While all previously described pyrrole halogenases 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 **1**, has not been characterized. These open biosynthetic questions, together with the ecological significance of **1** 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 **1** by the total *in vitro* enzymatic reconstitution of its biosynthetic machinery. We show that a single flavin-dependent brominase installs an unprecedented four halogens on an ACP-bound pyrrole required for the progression of thioesterase (TE)-mediated offloading and decarboxylation reactions. We also exploit an opportunity to structurally characterize and investigate via site-directed mutagenesis the molecular basis for differential number of halogen additions catalyzed by flavin-

dependent halogenases on aromatic substrates. Our structural comparison of a highly homologous pyrrole *di*halogenase to the pyrrole *tetra*halogenase reveals subtle variations in biocatalyst design that lead to divergent molecular outcomes.

RESULTS

Genetic basis for the biosynthesis of 1

We recently reported that **1** produced by the marine bacterium *Pseudoalteromonas* sp. PS5 induces settlement of larvae associated with several Caribbean coral species (3). To investigate the genetic basis for the biosynthesis of **1**, we sequenced and assembled a 5.08 Mbp draft genome for *P.* sp. PS5. Querying the draft genome for the presence of the halopyrrole biosynthetic genes, we identified a gene locus in *P.* sp. PS5 with high homology to a subset of bromopyrrole biosynthetic genes present in the previously reported *bmp* gene locus from marine bacteria of the genera *Pseudoalteromonas* and *Marinomonas* (11) (**Fig. 2A**, **Table S1**). The organization of the *bmp* homologs identified in the genome of *P.* sp. PS5 (*PS5_bmp*) is identical to that of the *bmp* gene cluster found in *Marinomonas mediterranea* MMB-1 *bmp* (*Mm_bmp*) and others (11). Notably, the *PS5_bmp* gene locus maintains the bromopyrrole biosynthetic module *Mm_bmp1–4* and the bromopyrrole/phenol coupling cytochrome P450 (CYP450) accessory genes *Mm_bmp9–10*, but lacks the genes associated with bromophenol biosynthesis (*Mm_bmp5–6*) and the CYP450-*Mm_bmp7* (**Fig. 2A**). In addition, we also sequenced and assembled a 5.13 Mbp draft genome for *Pseudoalteromonas* sp. A757 that also produces **1** (12). Querying the genome of *P.* sp. PS5 as before revealed a gene cluster with high homology to *PS5_bmp* (*A757_bmp*), providing a second example of a stand-alone bromopyrrole biosynthetic pathway from the genus *Pseudoalteromonas* (**Fig. 2A**, **Table S1**)

To evaluate the role of *PS5_bmp1–4* genes in the production of **1**, we cloned and heterologously expressed *PS5_bmp1–4* in *Escherichia coli*. Only in the presence of bromide in the culture media, we observed robust heterologous production of **1** along with a minor amount of **6**, consistent with the product profile observed in organic extracts of cultures of *P.* sp. PS5 and *P.* sp. A757 (**Fig. 2B**). While these results demonstrate that *bmp1–4* genes are necessary for the production of **1**, the molecular detail 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-carboxyl was not discernible at this stage. Hence, we next examined the biosynthetic pathway by total *in vitro* enzymatic reconstitution of the production of **1** using purified enzyme catalysts.

Decarboxylative bromination of thiotemplated pyrrole

In order to examine the individual roles of Bmp1–4, we performed the total *in vitro* reconstitution of the biosynthesis of **1** (**Fig. 3A**). Due to high sequence similarity of PS5_Bmp1–4 to Mm_Bmp1–4 (**Table S1**), and

production of **1** by both *P.* sp PS5 and *M. mediterranea* MMB-1 (11), we used recombinant Mm_Bmp1–4 proteins in our *in vitro* investigations. Bmp2 and Bmp4 were individually purified as N-His_6-tagged proteins, while N-His_6-tagged apo-Bmp1 was purified in complex with untagged Bmp3 and converted to its pantetheine-loaded holo-form by the promiscuous *Bacillus subtilis* phosphopantetheinyl transferase Sfp (11, 13) (**Fig. 3A**. Incubation of L-proline and bromide with purified Bmp1–4 enzymes, along with flavin and nicotinamide cofactors (FAD and $NADP^+$), and cofactor regeneration enzymes (PtdH and SsuE) led to the formation of **1** as the major product (**Fig. 3**; see SI Appendix, **Fig. S1**). The production of **1** was abolished in the absence of ATP, which is required for the loading of L-proline on to the ACP-domain of Bmp1 by Bmp4 (**Fig. 3B**). By comparison to authentic synthetic standards, we also confirmed the production of **6** and 3,4,5-tribromo-pyrrole-2-carboxylic acid (**7**) as minor products of the reaction (**Fig. 3B**).

Having established that Bmp1–4 enzymes are sufficient for the production of **1,** we next queried the timing and mechanism of the terminal bromination, offloading, and decarboxylation reactions. We previously established that Mm_Bmp2 converts pyrrolyl-*S*-Bmp1(ACP) to tribromopyrrolyl-*S*-Bmp1(ACP) (**Fig. 3A**), where the Bmp1(ACP) domain comprises residues 1–77 of Bmp1 (11). Hence, we rationalized that 3,4,5-tribromopyrrolyl-*S*-Bmp1(ACP) (**8**, **Fig. 4A**) might be an intermediate *en route* to **1**. To test this hypothesis, we synthesized **8** following a recently described chemoenzymatic method (14). Briefly, synthetic **7**, generated by the bromination of pyrrole-2-carboxylic acid, was sequentially ligated to cysteamine and then to pantothenic acid to afford 3,4,5-tribromopyrrolyl-*S*-pantetheine **9** (**Fig. 4A**, see Supporting Information for synthesis protocols and product characterization data). Compound **9** was then extended by the *E. coli* coenzyme A (CoA) biosynthetic enzymes CoaA, CoaD, and CoaE to generate 3,4,5-tribromopyrrolyl-*S*-CoA, that was used as the substrate for the *in situ* transfer of the 3,4,5-tribromopyrrolyl-*S*-phosphopantetheine moiety to the serine side chain hydroxyl of apo-Bmp1(ACP) by Sfp to yield **8** (**Fig. 4A**; see SI Appendix, **Fig. S2**). We then expressed and purified the Bmp1 thioesterase domain, Bmp1(TE), whose activity we previously confirmed using a model esterase 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**, **Figs. 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 Ser202 and confirmed that its mutation to alanine resulted in loss of esterase activity for the *p*-nitrophenylacetate substrate (see SI Appendix, **Fig. S3**). Substitution of Bmp1(TE) with catalytically inactive Bmp1(TE)S202A mutant enzyme led to significantly reduced production of **1** (**Fig. 4B**; see SI Appendix, **Fig. S4**). Finally, exclusion of Bmp2 from the reaction completely abolished the production of **1** (**Fig. 4B**; see 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 (see SI Appendix, **Fig. S5**).

These results provide two important findings. First, both catalysts, Bmp1(TE), and the halogenase Bmp2 are required to convert the intermediate **8** to **1**. Second, *two* enzymatic activities, that is, hydrolysis via thioesterase Bmp1(TE) and bromination via Bmp2 lead to *three* chemical events – 1) offloading of the pyrrole from Bmp1(ACP), 2) the fourth bromination on the 2'-position of the pyrrole, and 3) decarboxylative loss of the L-proline derived alpha-carboxylate carbon atom.

We next evaluated the timing of formation of **6** relative to the formation of **1**. We thus incubated **6** 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 non-physiological substrate for Bmp2 (see SI Appendix, **Fig. S5**). Hence, we propose that **6** might be a non-enzymatic reductive degradation byproduct of **1**. Indeed, incubation of synthetically prepared **1** with NADPH led to the conversion to **6** at levels comparable to total *in vitro* reconstitution reactions (see 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 following formation of **1** (**Fig. 5**).

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

Non-enzymatic hydrolysis of **i** may also lead to release of **iii** to afford **1**, consistent with our observation of low levels of **1** from reactions excluding Bmp1(TE) and with inactive Bmp1(TE)S202A. The observation of similar levels of **7** in all reactions further supports direct hydrolytic bypass of TE-mediated offloading. Furthermore, the fourth bromination of **8** by Bmp2, followed by transesterification and decarboxylation of **i** neatly differentiates the biosynthesis of **1** from non-ribosomal peptide synthetase and polyketide synthase derived pathways involving halopyrrole units in which the L-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 **1** is in stark contrast to all previously characterized pyrrole halogenases that halogenate the pyrrole ring only once or twice. In order 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 to the FAD-bound holo-structure of dihalogenase Mpy16 participating in the biosynthesis of **3** that we determined at 1.95 Å resolution (**Fig. 6B**, **Table S2**) (8). Bmp2 and Myp16 structures are highly homologous with all secondary structural elements in the vicinity of the active site strictly conserved between the two enzymes (see SI Appendix, **Fig. S7**). Furthermore, the FAD cofactor isoalloxazine rings and the postulated active site lysine residues (15) (K74 for Bmp2, K72 for Mpy16, **Fig. 6A**,**B**) are superimposable. Though we could not determine a substrate pyrrolyl-*S*-ACP co-crystal structure for either Bmp2 or Mpy16, the position of the pyrrole binding sites for Bmp2 and Myp16 was inferred by a structural alignment of Bmp2 and Mpy16 with the crystal structures of substrate-bound forms of the flavin-dependent tryptophan-7-chlorinases PrnA (16) and RebH (17). Fittingly, the postulated pyrrole-binding site in Bmp2 and Mpy16 (**Fig. 6A**,**B**) is in close proximity to the side chain of the catalytic lysine residue. An examination of the amino acids lining this putative pyrrole binding site in Bmp2 reveals a lack of conservation of three key sites (Y302, F306, and A345) that are otherwise strictly conserved among all pyrrolyl-*S*-ACP halogenases that selectively catalyze one (HrmQ (10)) or two (PltA (7), Mpy16 (8), Clz5 (9), and Pyr29 (18)) halogenations on the pyrrole ring (**Fig. 6D**). We thus mutated these three residues in Bmp2 to the corresponding residues in Mpy16 (**Fig. 6B**) to generate Bmp2-Y302S/F306V/A345W triple mutant enzyme (henceforth referred to as Bmp2-TM). We next investigated the affect of the triple mutation on the *in vivo* heterologous production of **1** in *E. coli*. Co-expression of *Mm_bmp1*, *Mm_bmp3*, and *Mm_bmp4*, with *Mm_bmp2-TM* in *E. coli* led to a complete loss in the *in vivo* production of **1** in contrast to robust production with wild-type *Mm_bmp2* (see SI Appendix, **Fig. S9**).

To explore the molecular basis for the *in vivo* loss in production of **1** by the Bmp2-TM enzyme, we compared the *in vitro* enzymatic activities of Bmp2, Mpy16, and Bmp2-TM enzymes. We chemoenzymatically synthesized pyrrolyl-*S*-Bmp1(ACP) and pyrrolyl-*S*-Mpy15, the requisite substrates for Bmp2 and Mpy16, respectively (8, 11, 14). Using a previously reported mass spectrometry-based proteomic assay that relies on detection of acyl-(cyclo)pantetheine MS2 product ions (14), we verified that wild-type Bmp2 enzyme catalyzed three brominations on the pyrrole ring (**Fig. 6E**), and that Mpy16 catalyzed no more than two chlorinations on the pyrrole ring acylated to Mpy15 (**Fig. 6F**) (14). Note that although Bmp2 tetrabrominates the pyrrole ring as established by our previous findings, the assay employed here detects only up to tribromination of pyrrolyl-*S*-Bmp1(ACP). In contrast to the wild-type Bmp2, Bmp2-TM could only catalyze a single bromination on the pyrrole ring (**Fig. 6G**; see SI Appendix, **Fig. S8**), thereby precluding the synthesis of **1**.

Lastly, to eliminate the possibility that the reduced degree of halogenation catalyzed by Bmp2-TM was due to disruption of the enzyme active site by the three mutations, we determined the crystal structure of holo-Bmp2-TM at a resolution of 1.98 Å (**Fig. 6C**). Structural comparison of Bmp2-TM to Bmp2 and Mpy16, confirmed that the mutation of the three Bmp2 active site residues did not alter the positioning of the FAD isoalloxazine ring or the catalytic lysine side chain.

DISCUSSION

In addition to characterizing the biosynthesis of a microbially produced coral settlement cue, our study reveals novel enzymology stemming from a biosynthetic motif ubiquitous among halopyrrole-containing natural products. While our previous work had established tribromination of pyrrolyl-*S*-Bmp1(ACP) by Bmp2, the enzymatic activities responsible for the addition of the fourth bromine atom in **1** 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 tetrabromination 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 **1**, we posit **1** is an intermediate *en route* to the production of **5**, reconciling a previous report implicating a L-proline-derived symmetrical pyrrolic intermediate in the biosynthetic scheme for the production of **5** (19). Previously, based on primary sequence homology, we postulated that the enzyme Bmp8 (**Fig. 2A**) participates in the decarboxylation of **7** to produce **6** (11), a hypothetical route that is likely not operative in light of the biochemical data presented above. Hence, the physiological transformations for the production of **6** *en route* **5** remain to be elucidated. Furthermore, the previously reported production of **1** during the CYP450-Bmp7 mediated biradical homodimerization of **6** is likely an *off-pathway* route, with Bmp1–4 being the primary players in the production of **1**. Successive biochemical studies promise to address these open questions to characterize all steps in the biosynthesis of **5** and assign physiological roles for each of the Bmp enzymes.

The complete *in vitro* reconstitution of the production of **1** using purified enzyme catalysts suggests a mechanism for the acyl-phosphopantetheine 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-PltL (57% amino acid similarity to Mm_Bmp1(ACP)) participating in the biosynthesis of the dichloropyrrole-containing natural product pyoluteorin (7, 20). By analogy, in the biosynthesis of **1**, it is possible that the terminal halogenation on the pyrrole partially liberates the sequestered substrate from the ACP, allowing access to the thioesterase. Hence, the terminal fourth bromination by Bmp2 serves the dual role of triggering the release of the sequestered pyrrole moiety by presumably 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 have been demonstrated in the biosynthesis of NRPS-PKS hybrid curacin A, in which a cryptic halogenation catalyzed by α-ketoglutarate-dependent halogenase-CurA promotes cyclopropane ring formation (21), and in the biosynthesis of meroterpenoid merochlorins A and B, in which a dearomatization/terpene cyclization reaction cascade is initiated by a chlorination catalyzed by vanadium-dependent haloperoxidase- Mcl24 (22). Most analogously, paralleling the decarboxylative bromination strategy employed by Bmp2, flavin-dependent brominase- Bmp5 catalyzes a bromination reaction that drives decarboxylation in the conversion of free *p*-hydroxybenzoic acid to 2,4-dibromophenol, the bromophenol building block of **5** (11).

The tetrabromination activity of Bmp2 is unprecedented among the numerous flavin-dependent halogenases described to date in the natural product literature 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, in the process determining the first crystal structures of *biochemically characterized* flavin-dependent halogenases acting on ACP-bound substrates. While the previously reported 2.1 Å crystal structure of putative tyrosyl-*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 catalyst, Bmp2-TM, exhibiting an altered halogenation profile. Notably, the active site of Bmp2-TM exhibited no perturbation of the bound FAD co-factor with respect to wild-type 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 (17, 24-25). Our findings extend this observation to flavin-dependent halogenases that catalyze halogenation of aromatic substrates acylated to ACPs in demonstrating that side chains of residues lining the putative halogenase 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 Mpy16, efforts to alter the putative substrate-binding cavity to resemble that of Bmp2 resulted in insolubility of the mutant Mpy16 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 appear that Nature has taken note through the evolution of highly specialized halogenation catalysts.

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 acyl-*S*-ACP substrates. In addition, the primary question across all classes of halogenating enzymes is the structural determinant for halide specificity among these catalysts. While Bmp2 and Mpy16 demonstrate different halide specificities, the amino acid side chains and the positioning of the FAD isoalloxazine ring relative to these side chains is 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 enthalpic penalty associated with desolvation of the halide ion in the halogenase active site prior to $2e^-$ oxidation to generate the electrophilic halonium. Furthermore, the contribution of the redox potential of the flavin isoalloxazine ring necessary to oxidize the halide, and the stability of the activated lysine-amine intermediate generated *en route* to 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 underscoring 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 engineer-able toolkits to tailor the biosynthesis of halogenated natural products (30).

MATERIALS AND METHODS

Detailed materials and methods are provided in the *SI Materials and Methods*.

Database deposition information

The sequences for *P.* sp. PS5 and *P.* sp. A757 derived bmp gene clusters are deposited in GenBank under accession numbers KR011923 and KT808878, respectively. Structures for Mpy16, Bmp2, and Bmp2-TM are deposited in the Protein Data Bank (www.rcsb.org) under accession numbers 5BUK, 5BVA, and 5BUL, respectively.

ACKNOWLEDGEMENTS

We thank our colleague B.M. Duggan at the University of California-San Diego Skaggs School of Pharmacy and Pharmaceutical Sciences NMR facility (La Jolla, CA) for assistance in acquiring NMR data. This work was jointly supported by the US National Science Foundation (OCE-1313747) and the US National Institute of Environmental Health Sciences (P01-ES021921) through the Ocean and Human Health Program to B.S.M., and the US National Institute of Allergy and Infectious Disease R01-AI47818 to B.S.M. and R21-

AI119311 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 to V.J.P., the Howard Hughes Medical Institute to J.P.N., the US National Institutes of Health (NIH) Marine Biotechnology Training Grant predoctoral fellowship to A.E. (T32-GM067550), the Helen Hay Whitney Foundation postdoctoral fellowship to V.A., and a Swiss National Science Foundation (SNF) postdoctoral Fellowship to S.D. During the review and publication of this manuscript, crystal structure of the flavin-dependent halogenase PltA was reported in literature (31).

AUTHOR CONTRIBUTIONS

A.E., V.A., and B.S.M. designed research. A.E., V.A., S.D., I.R., and M.A.S. performed research. J.M.S., G.V.L., K.E.W., T.J.M., J.P.N., and V.J.P contributed analytical reagents and tools. A.E., V.A., and B.S.M. analyzed data and wrote the manuscript with input from all authors.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

MAIN TEXT FIGURES & LEGENDS

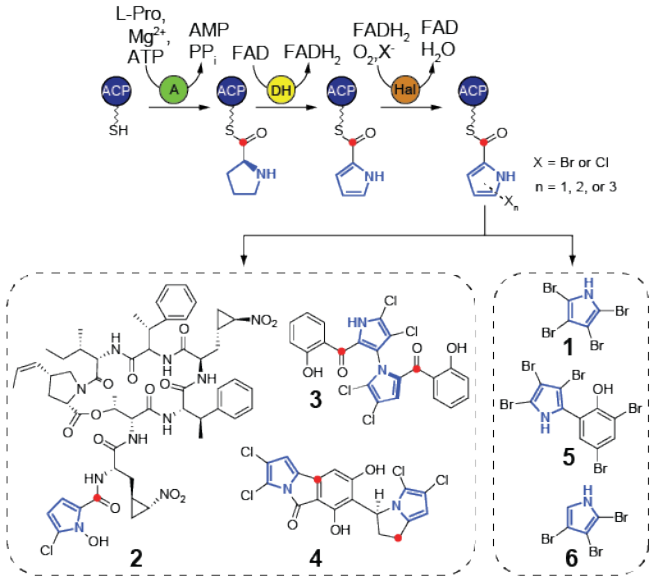

Fig. 1. Structures and biosynthesis of halopyrrole natural products. Halopyrrole biosynthesis starts from the loading of L-proline on to an ACP phosphopantetheine thiol by an adenylation (A) enzyme, followed by a $4e^-$ oxidation of the prolyl ring to a pyrrole (in blue) catalyzed by a dehydrogenase (DH) enzyme. Subsequent halogenation by a flavin-dependent halogenase (Hal) then installs one, two, or three halogens on the pyrrole ring. At this stage, the thiotemplated halopyrrole can proceed down assembly line biosynthetic routes to afford natural products such as hormaomycin (**2**), marinopyrrole A (**3**), and chlorizidine (**4**) (8-10). However, the biosyntheses of **1** and **5–6** do not entail modular assembly line extension of the L-proline-derived halopyrrole, but rather involve the loss of the prolyl alpha-carboxylate carbon atom (in red) via an unexplained mechanism.

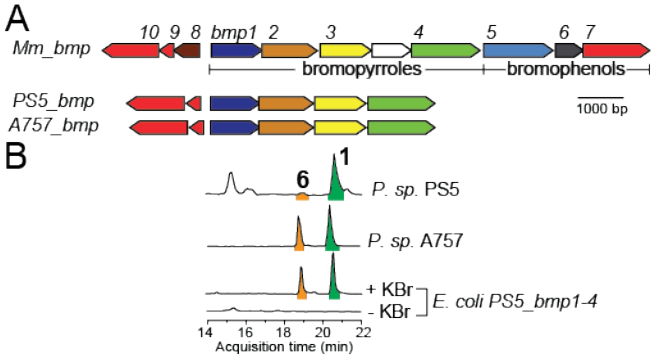

Fig. 2. A genetic basis for production of 1. (**A**) *Mm_bmp*, *PS5_bmp, and A757_bmp* gene clusters with bromopyrrole and bromophenol biosynthetic genes indicated below the *Mm_bmp* gene cluster; a putative

permease (uncolored) is inserted between *Mm_bmp3* and *Mm_bmp4*. Mm_Bmp1–4 are colored per their catalytic roles shown in **Fig. 1**. Note that *bmp1* encodes a di-domain protein with an ACP domain at the N-terminus followed by a TE domain (11). (**B**) LC/MS extracted ion chromatograms (EICs) for $[M-H]^{1-}$ ions corresponding to **1** and **6** in organic extracts of *P.* sp. PS5, *P.* sp. A757, and *E. coli* expressing *PS5_bmp1–4* grown in media with (+) or without (-) bromide.

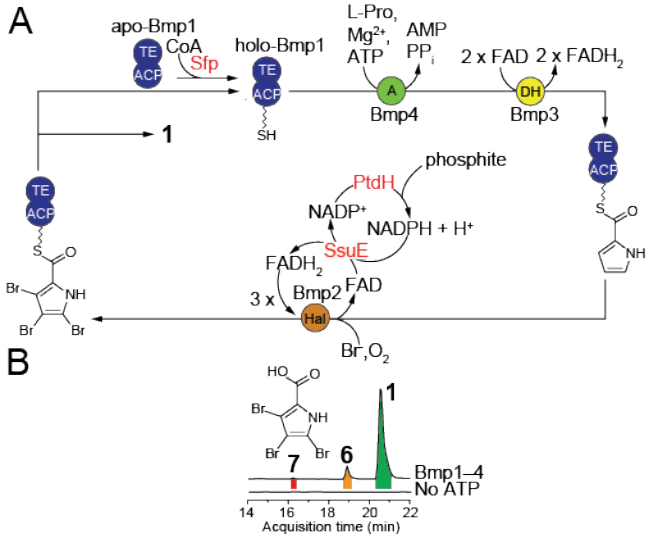

Fig. 3. Enzymatic synthesis of 1. (**A**) Scheme for the total *in vitro* enzymatic synthesis of **1** starting from L-proline. Accessory enzymes not derived from the *Mm_bmp* gene cluster are indicated in red. $FADH_2$ required as a cofactor for Bmp2 was regenerated *in situ* by *E. coli* flavin reductase SsuE (32) which in turn oxidizes NADPH to $NADP^+$. $NADP^+$ was recycled to NADPH *in situ* using $NAD(P)^+$ reductase/phosphite dehydrogenase- PtdH (33). (**B**) EICs for $[M-H]^{1-}$ ions corresponding to **1**, **6**, and **7** for organic extracts of reactions described in panel **A**. Note that the production of **1**, **6**, and **7** was abolished in the absence of ATP.

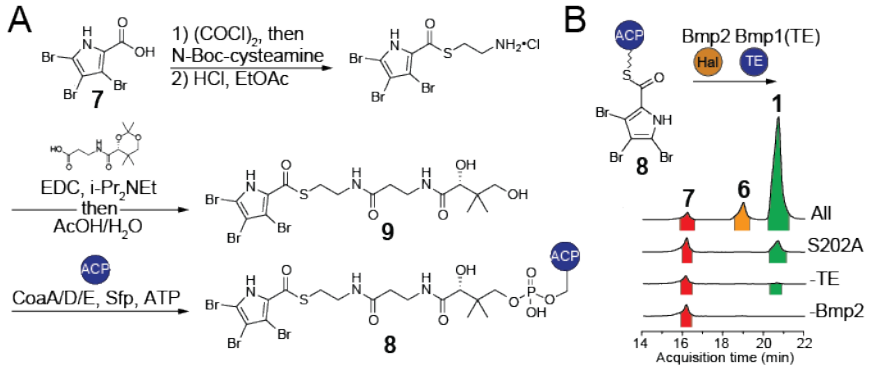

Fig. 4. Preparation of 8 and *in vitro* conversion to 1 by Bmp2 and Bmp1(TE). (**A**) Scheme showing generation of **8** by one-pot *in vitro* extension of synthetically prepared **9** using *E. coli* CoA biosynthetic enzymes, and loading onto holo-Bmp1(ACP) by Sfp. (**B**) Scheme for conversion of **8** to **1**, **6**, and **7** by Bmp2 and Bmp1(TE). Omitted for clarity are the cofactor regeneration systems for Bmp2 shown in **Fig. 3**. Only trace, or no production of **1** was observed when either Bmp2 or Bmp1(TE) were omitted from the *in vitro* enzymatic reaction, or when the catalytically inactive Bmp1(TE)S202A enzyme was used.

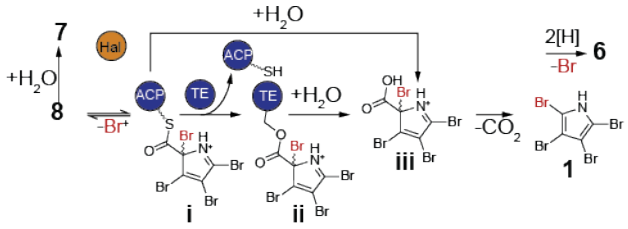

Fig. 5. Scheme for bromination-dependent pyrrole offloading and decarboxylation. Proposed reaction mechanisms for terminal bromination, offloading, and decarboxylation reactions with **8**; inferred intermediates are indicated by bold Roman numerals.

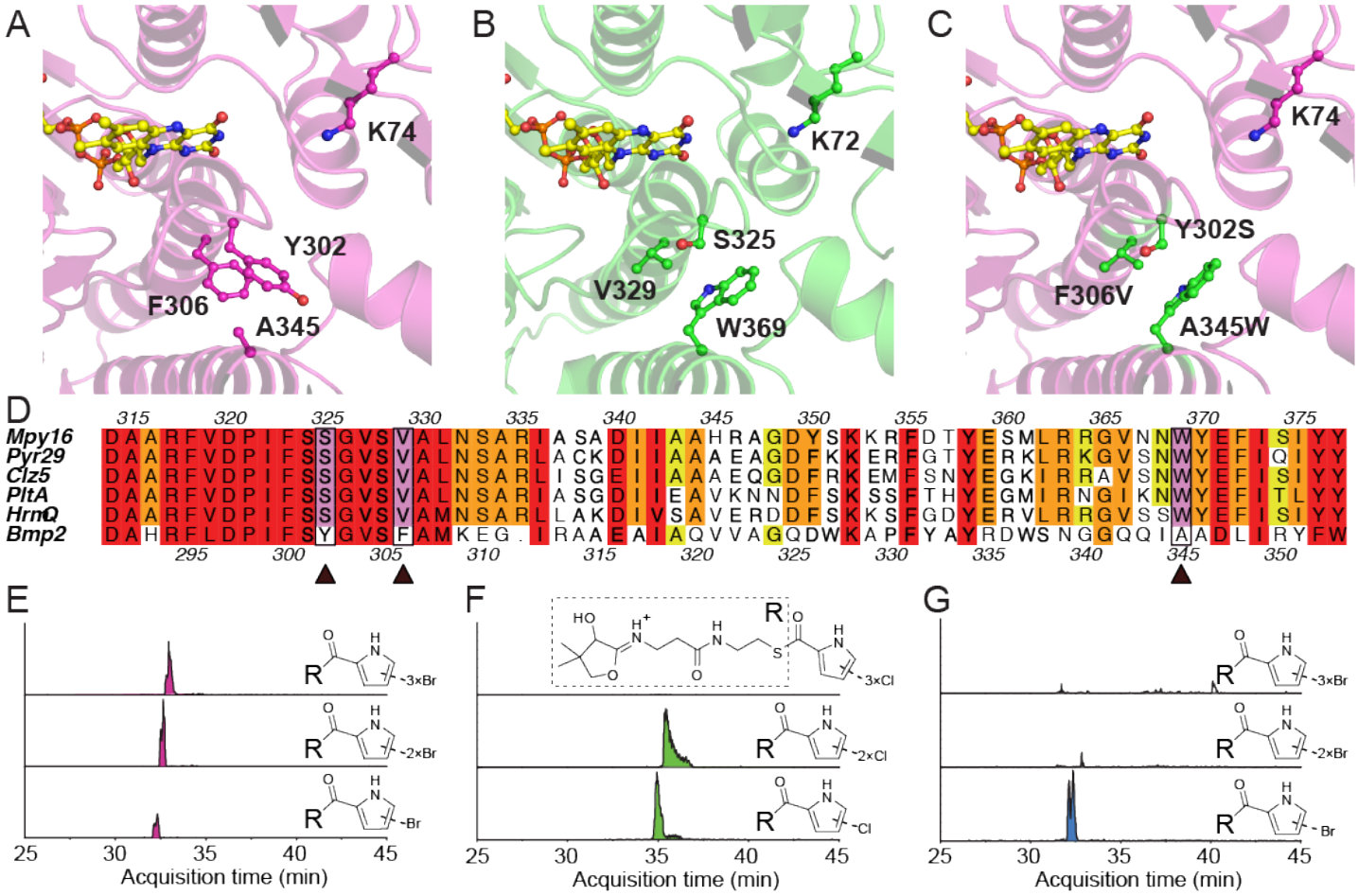

Fig. 6. Structural basis for tetrahalogenation activity of Bmp2. Comparison of the active sites of (**A**) holo-Bmp2, (**B**) holo-Mpy16, and (**C**) holo-Bmp2-TM. Note that side chains of Bmp2-TM residues S302, V306, and W345 in panel **C** structurally map to the corresponding side chains of Mpy16 residues S325, V329, and W369 shown in panel **B** without perturbation of the FAD isoalloxazine ring or the proposed catalytic lysine side chain. (**D**) Primary sequence alignments of tetrahalogenase Bmp2 with mono- and di- halogenating pyrrolyl-*S*-ACP halogenases demonstrates that the amino acids that were mutated on the basis of structural comparison to form Bmp2-TM (indicated by ▲) are strictly conserved in all pyrrolyl-*S*-ACP halogenases but Bmp2 (see SI Appendix, **Fig. S20** for alignment of Bmp2 homologs from different bacterial species). EICs for acyl-(cyclo)pantetheine MS2 product ions demonstrate that (**E**) wild type Bmp2 generates mono-, di- and tri-

bromopyrrolyl-*S*-Bmp1(ACP) products, while (**F**) Mpy16 generates only mono- and di-chloropyrrolyl-*S*-Mpy15, and (**G**) Bmp2-TM generates only mono-bromopyrrolyl-*S*-Bmp1(ACP) product.

References

1. Holmstrom C & Kjelleberg S (1999) Marine *Pseudoalteromonas* species are associated with higher organisms and produce biologically active extracellular agents. *FEMS Microbiol Ecol* 30(4):285-293.
2. Tebben J, *et al.* (2011) Induction of larval metamorphosis of the coral *Acropora millepora* by tetrabromopyrrole isolated from a *Pseudoalteromonas* bacterium. *PLoS One* 6(4):e19082.
3. Sneed JM, Sharp KH, Ritchie KB, & Paul VJ (2014) The chemical cue tetrabromopyrrole from a biofilm bacterium induces settlement of multiple Caribbean corals. *Proc Biol Sci* 281(1786).
4. Neu AK, Mansson M, Gram L, & Prol-Garcia MJ (2014) Toxicity of bioactive and probiotic marine bacteria and their secondary metabolites in *Artemia* sp. and *Caenorhabditis elegans* as eukaryotic model organisms. *Appl Environ Microbiol* 80(1):146-153.
5. Gribble GW (2010) *Naturally Occurring Organohalogen Compounds – a Comprehensive Update* (Springer).
6. Walsh CT, Garneau-Tsodikova S, & Howard-Jones AR (2006) Biological formation of pyrroles: nature's logic and enzymatic machinery. *Nat Prod Rep* 23(4):517-531.
7. Dorrestein PC, Yeh E, Garneau-Tsodikova S, Kelleher NL, & Walsh CT (2005) Dichlorination of a pyrrolyl-*S*-carrier protein by $FADH_2$-dependent halogenase PltA during pyoluteorin biosynthesis. *Proc Natl Acad Sci U S A* 102(39):13843-13848.
8. Yamanaka K, Ryan KS, Gulder TA, Hughes CC, & Moore BS (2012) Flavoenzyme-catalyzed atropo-selective N,C-bipyrrole homocoupling in marinopyrrole biosynthesis. *J Am Chem Soc* 134(30):12434-12437.
9. Mantovani SM & Moore BS (2013) Flavin-linked oxidase catalyzes pyrrolizine formation of dichloropyrrole-containing polyketide extender unit in chlorizidine A. *J Am Chem Soc* 135(48):18032-18035.
10. Hofer I, *et al.* (2011) Insights into the biosynthesis of hormaomycin, an exceptionally complex bacterial signaling metabolite. *Chem Biol* 18(3):381-391.
11. Agarwal V, *et al.* (2014) Biosynthesis of polybrominated aromatic organic compounds by marine bacteria. *Nat Chem Biol* 10(8):640-647.
12. Whalen KE, Poulson-Ellestad KL, Deering RW, Rowley DC, & Mincer TJ (2015) Enhancement of antibiotic activity against multidrug-resistant bacteria by the efflux pump inhibitor 3,4-dibromopyrrole-2,5-dione isolated from a *Pseudoalteromonas* sp. *J Nat Prod* 78(3):402-412.
13. Quadri LE, *et al.* (1998) Characterization of Sfp, a *Bacillus subtilis* phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. *Biochemistry* 37(6):1585-1595.
14. Agarwal V, *et al.* (2015) Chemoenzymatic synthesis of acyl coenzyme A substrates enables *in situ* labeling of small molecules and proteins. *Org Lett* 17(18):4452-4455.
15. Yeh E, Blasiak LC, Koglin A, Drennan CL, & Walsh CT (2007) Chlorination by a long-lived intermediate in the mechanism of flavin-dependent halogenases. *Biochemistry* 46(5):1284-1292.
16. Dong C, *et al.* (2005) Tryptophan 7-halogenase (PrnA) structure suggests a mechanism for regioselective chlorination. *Science* 309(5744):2216-2219.
17. Zhu X, *et al.* (2009) Structural insights into regioselectivity in the enzymatic chlorination of tryptophan. *J Mol Biol* 391(1):74-85.
18. Zhang X & Parry RJ (2007) Cloning and characterization of the pyrrolomycin biosynthetic gene clusters from *Actinosporangium vitaminophilum* ATCC 31673 and *Streptomyces* sp. strain UC 11065. *Antimicrob Agents Chemother* 51(3):946-957.
19. Peschke JD, Hanefeld U, & Laatsch H (2005) Biosynthesis of the marine antibiotic pentabromopseudilin. 2. The pyrrole ring. *Biosci Biotech Bioch* 69(3):628-630.
20. Jaremko MJ, Lee DJ, Opella SJ, & Burkart MD (2015) Structure and substrate sequestration in the pyoluteorin type II peptidyl carrier protein PltL. *J Am Chem Soc* 137(36):11546-11549.
21. Gu L, *et al.* (2009) Metamorphic enzyme assembly in polyketide diversification. *Nature* 459(7247):731-735.

22. Diethelm S, Teufel R, Kaysser L, & Moore BS (2014) A multitasking vanadium-dependent chloroperoxidase as an inspiration for the chemical synthesis of the merochlorins. *Angew Chem Int Ed Engl* 53(41):11023-11026.
23. Buedenbender S, Rachid S, Muller R, & Schulz GE (2009) Structure and action of the myxobacterial chondrochloren halogenase CndH: a new variant of FAD-dependent halogenases. *J Mol Biol* 385(2):520-530.
24. Lang A, *et al.* (2011) Changing the regioselectivity of the tryptophan 7-halogenase PrnA by site-directed mutagenesis. *Angew Chem Int Ed Engl* 50(13):2951-2953.
25. Shepherd SA, *et al.* (2015) Extending the biocatalytic scope of regiocomplementary flavin-dependent halogenase enzymes. *Chem Sci* 6(6):3454-3460.
26. Eustaquio AS, *et al.* (2003) Clorobiocin biosynthesis in *Streptomyces*: identification of the halogenase and generation of structural analogs. *Chem Biol* 10(3):279-288.
27. Harris CM, Kannan R, Kopecka H, & Harris TM (1985) The role of the chlorine substituents in the antibiotic vancomycin - preparation and characterization of monodechlorovancomycin and didechlorovancomycin. *J Am Chem Soc* 107(23):6652-6658.
28. Neumann CS, Fujimori DG, & Walsh CT (2008) Halogenation strategies in natural product biosynthesis. *Chem Biol* 15(2):99-109.
29. Blasiak LC & Drennan CL (2009) Structural perspective on enzymatic halogenation. *Acc Chem Res* 42(1):147-155.
30. Teufel R, Agarwal V, & Moore BS (2016) Unusual flavoenzyme catalysis in marine bacteria. *Current Opinion in Chemical Biology* 31:31-39.
31. Pang AH, Garneau-Tsodikova S, & Tsodikov OV (2015) Crystal structure of halogenase PltA from the pyoluteorin biosynthetic pathway. *J Struct Biol* 192(3):349-357.
32. Eichhorn E, van der Ploeg JR, & Leisinger T (1999) Characterization of a two-component alkanesulfonate monooxygenase from *Escherichia coli*. *J Biol Chem* 274(38):26639-26646.
33. Costas AMG, White AK, & Metcalf WW (2001) Purification and characterization of a novel phosphorus-oxidizing enzyme from *Pseudomonas stutzeri* WM88. *J Biol Chem* 276(20):17429-17436.

APPENDIX: SUPPLEMENTARY INFORMATION

Supplementary information for

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

Abrahim 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 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 ºC glycerol stocks and grown overnight at 30 ºC 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 PGM Template OT2 400 bp Library Preparation Kit (Life Technologies). The library was sequenced on an Ion Torrent PGM platform (Life Technologies) using an Ion Torrent 318v2 sequencing chip (Life Technologies).

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 SacI/NotI (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* (3) (Genbank accession code: NC_015276.1) was amplified from genomic DNA and cloned into NdeI/XhoI digested pRSFDuet (Novagen) using the same protocol as before. Thus constructed pETduet-*PS5_bmp1–4* and pRSFduet-*acpS* 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-*acpS* grown at 30 ºC. The cultures were grown at 37ºC for 2 h with shaking, then transferred to 30 ºC and induced after 15 min with 50 µM IPTG (final concentration). Cultures were incubated for an additional 16 h at 30 ºC 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 ºC glycerol stocks of *P.* sp. PS5 or *P.* sp. A757 and grown overnight at 30 ºC. 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 ºC 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 ºC 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 C_{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. 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 $[M-H]^{1-}$ ± *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 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-His_6-*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* Bl21-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 $MgCl_2$, 50 mM KBr, 2 mM $NADP^+$, 0.1 mM FAD, 10 mM phosphite, 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 ºC 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 F_{254} 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 C_{18} 5µ 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_2CO_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, N*H*); HRMS (ESI) *m/z* calculated for C_4NBr_4 ($[M-H]^{1-}$) 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_4NBr_3 ($[M-H]^-$) 299.7665, found 299.7663. **7**: (300 mg, 0.86 mmol, 10%). R_f 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, COO*H*), 11.71 (bs, 1H, N*H*); ^{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 $C_5H_3NO_2Br_3$ ($[M+H]^-$) 345.7708, found 345.7674.

Chemical synthesis of 3,4,5-tribromopyrrolyl-*S*-pantetheine 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_2Cl_2 (10 mL). Oxalyl chloride (0.15 mL, 1.7 mmol, 2 equiv.) was added slowly and the mixture was allowed to stir for 1 h. The solvent was removed and the residue was taken up in CH_2Cl_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%). R_f 0.30 (hexanes/EtOAc 4:1); ^{1}H-NMR (500 MHz, $CDCl_3$, n.b.- 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 $C_{12}H_{15}N_2O_3SBr_3Na$ ($[M+Na]^+$) 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)[1]. The solution was allowed to warm to ambient temperature and stirred for 15 min before 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_2Cl_2 (2 mL) was added followed by diisopropylethylamine (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_2O (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. The combined organic phases were washed with brine and dried over $MgSO_4$. The solvent was removed and the residue was purified by flash column chromatographs (CH_2Cl_2/MeOH 20:1) to give diol **9:** (7.5 mg, 0.012 mmol, 80%) R_f 0.45 (CH_2Cl_2/MeOH 9:1); ^{1}H-NMR (500 MHz, d_6-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); ^{13}C-NMR (150 MHz, d_6-DMSO): δ

[1] To a mixture of EtOAc (2.4 mL) and EtOH (0.72 mL) at 0 °C was slowly added AcCl (0.85 mL).

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 $C_{16}H_{22}N_3O_5SBr_3Na$ ($[M+Na]^+$) 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 $MgCl_2$, and 9 mM ATP in 50 mM HEPES (pH 7.9) buffer were incubated at 30ºC 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 add 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 C_4 column (Vydac 5 µm, 4.6 mm × 250 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. Loading was assessed by extraction of the predicted acyl-(cylco)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 peptidic 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 µM **8**, 50 mM KBr, 0.1 mM FAD, 2 mM $NADP^+$, SsuE, PtdH, 10 mM phosphite, 50 µM Mm_Bmp2, and 25 µM Mm_Bmp1(TE) or Mm_Bmp1(TE):S202A for 12 h at 30 ºC 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 Mm_Bmp1-4 *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 OD_{600} ~ 0.6 at 30 ºC and protein expression was induced by the addition of IPTG to a final concentration of 0.3 mM. The temperature was adjusted to 18 ºC, 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 isocratically. 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% (w/v) PEG8000, 5% (v/v) glycerol. The crystals were briefly dipped in 100 mM Tris-HCl (pH 8.0), 10% (w/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 3E1T 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 LB 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 OD_{600} ~ 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*-ACPs 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-phosphite, 5 mM TCEP, 10 µM PtdH and SsuE 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

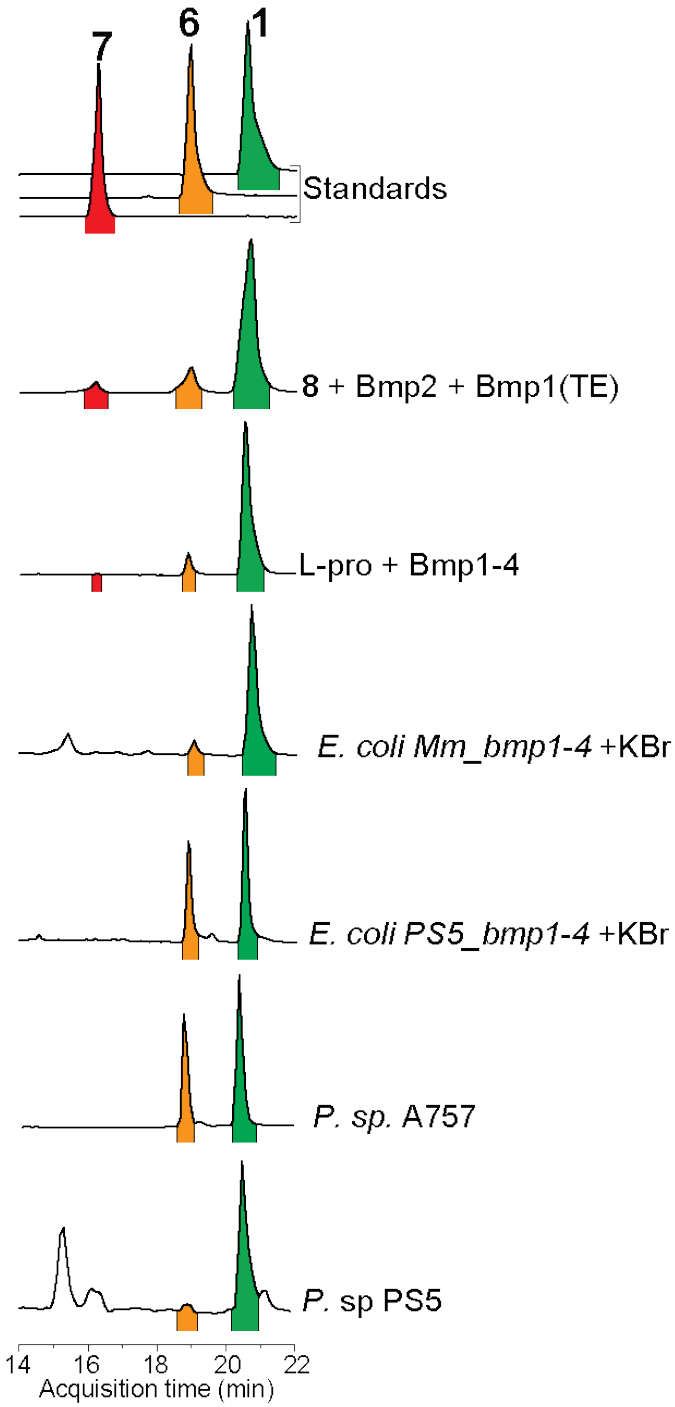

Figure S1. LC/MS comparisons of *in vivo* and *in vitro* EICSs to those of authentic standards of 1, 6, and 7. EICs for predicted $[M-H]^{1-}$ for the most abundant isotopic masses of **1** (*m/z* 381.67), **6** (*m/z* 301.76), and **7** (*m/z* 345.75). All traces are self-normalized.

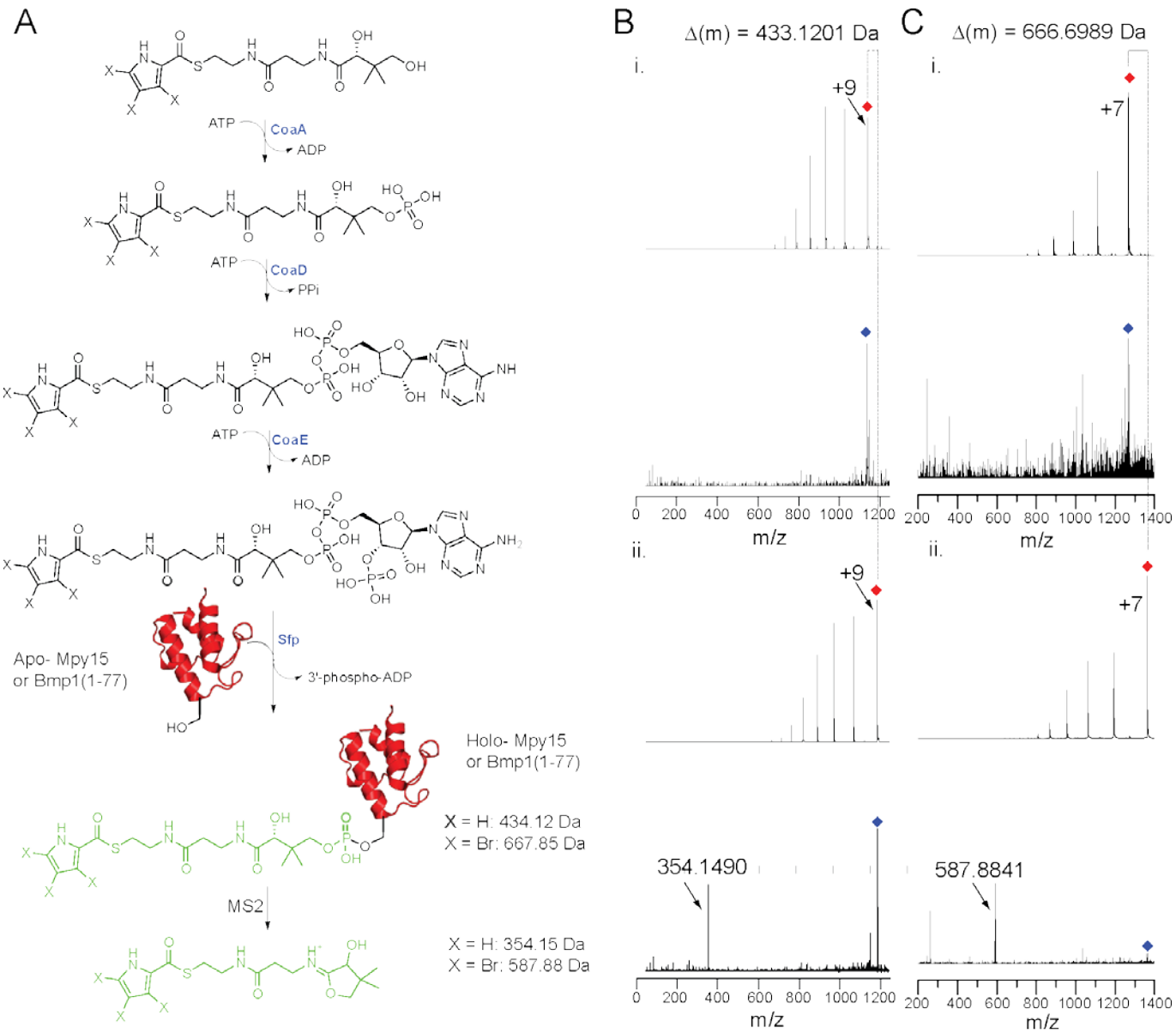

Figure S2. Confirmation of CoA extension and loading of pyrrole substrates onto ACPs. A. Scheme for CoA *in vitro* extension of chemically synthesized pyrrolyl-*S*-pantetheine precursor to pyrrolyl-*S*-CoA by *E. coli* CoA biosynthetic enzymes CoaA, CoaD, and CoaE. Upon completion of CoA synthesis, the 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**.

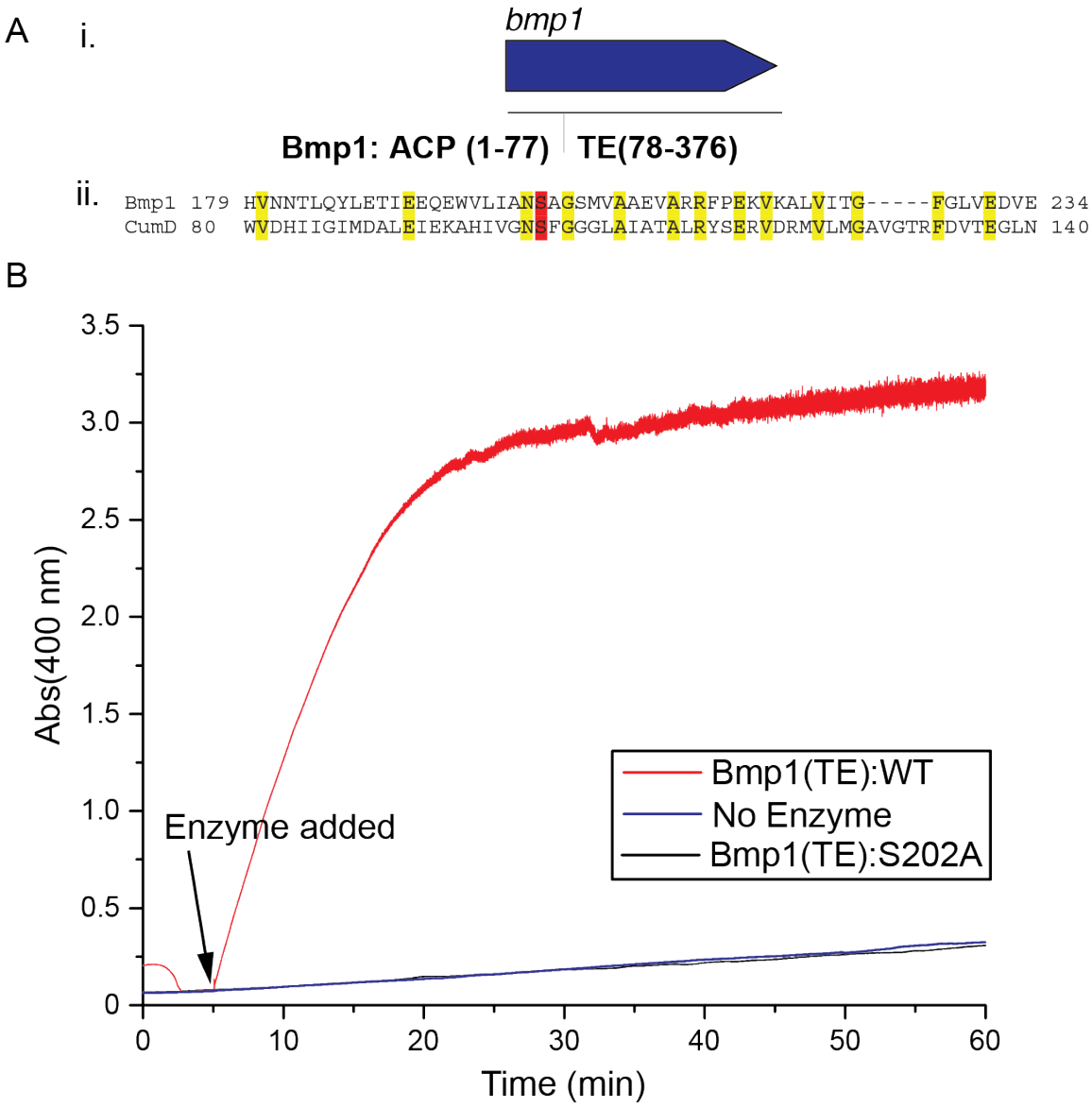

Figure S3. Bioinformatic identification and *in vitro* confirmation of catalytic Ser202 of Mm_Bmp1(TE) with a model chromogenic substrate, p-nitrophenyl acetate (pNPA).

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).

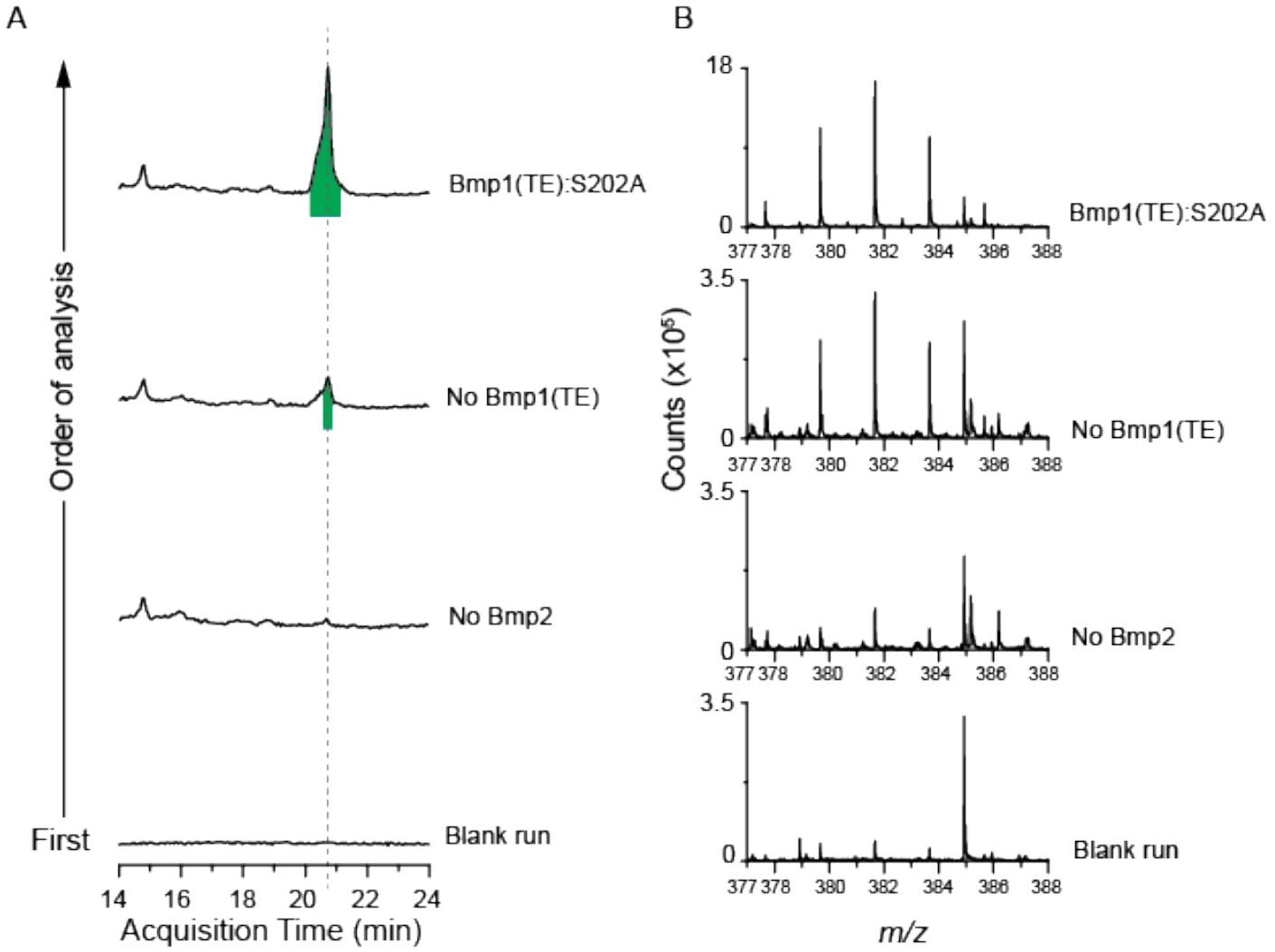

Figure S4. Non-enzymatic hydrolysis of proposed ACP-bound tetrabrominated intermediate yields 1. Extracted ion chromatograms (EICs) for 2,3,4,5-tetrabromo-1H-pyrrole (*m/z* 381.67) from incubations of 3,4,5-tribromopyrrolyl-*S*-ACP with various combination of Bmp1(TE) and Bmp2. All EICs are normalized 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**.

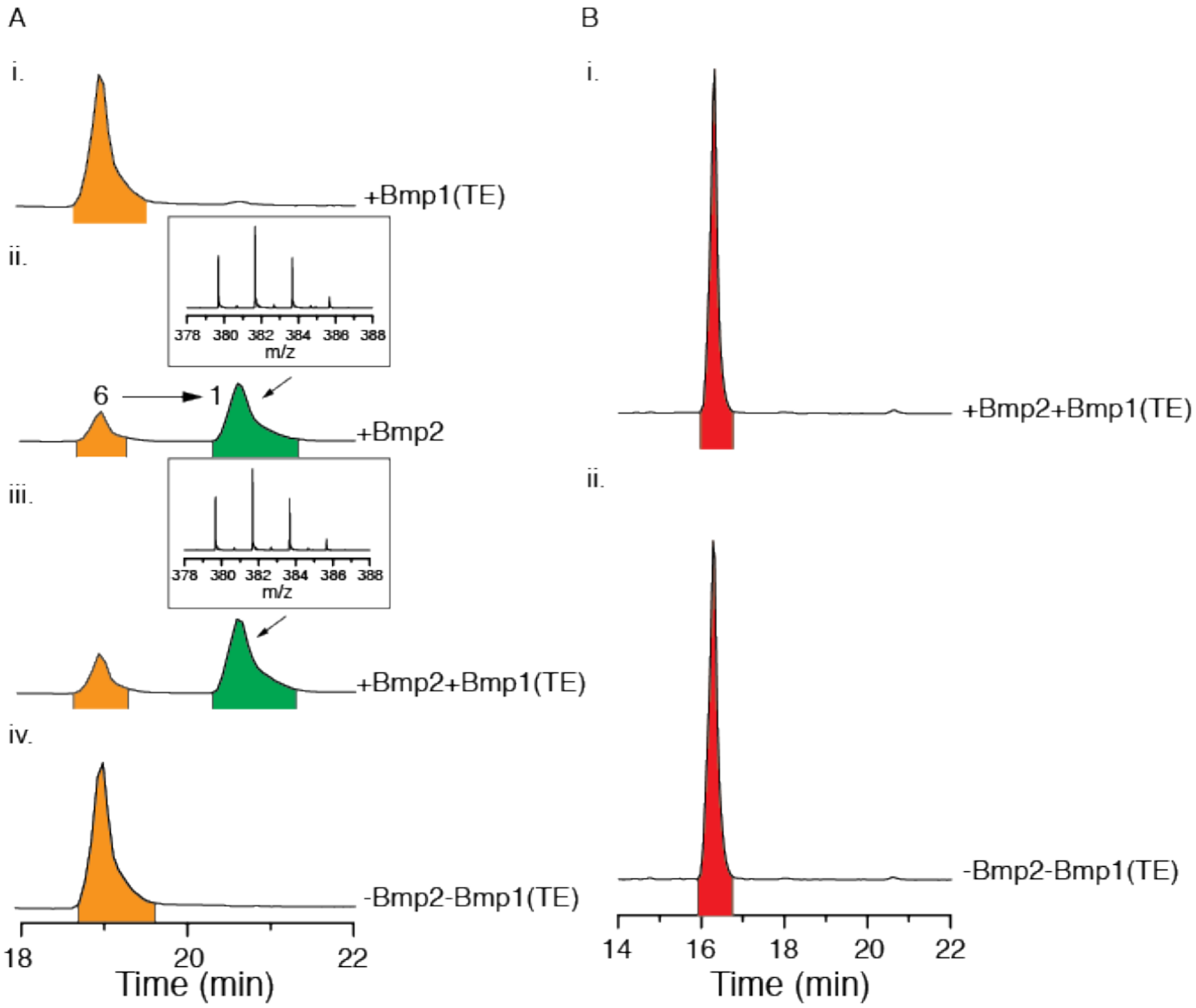

Figure S5. Interrogation of the potential of 2,3,4-tribromopyrrole 6 and 7 to serve as intermediates in the biosynthesis of 1. One mL reactions consisting of 25 µM **2** or **7** were incubated with 50 mM KBr, 0.1 mM FAD, 2 mM $NADP^+$, SsuE, PtdH, 10 mM phosphite including or excluding 50 µM Bmp2 and 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 ºC. 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 'i".

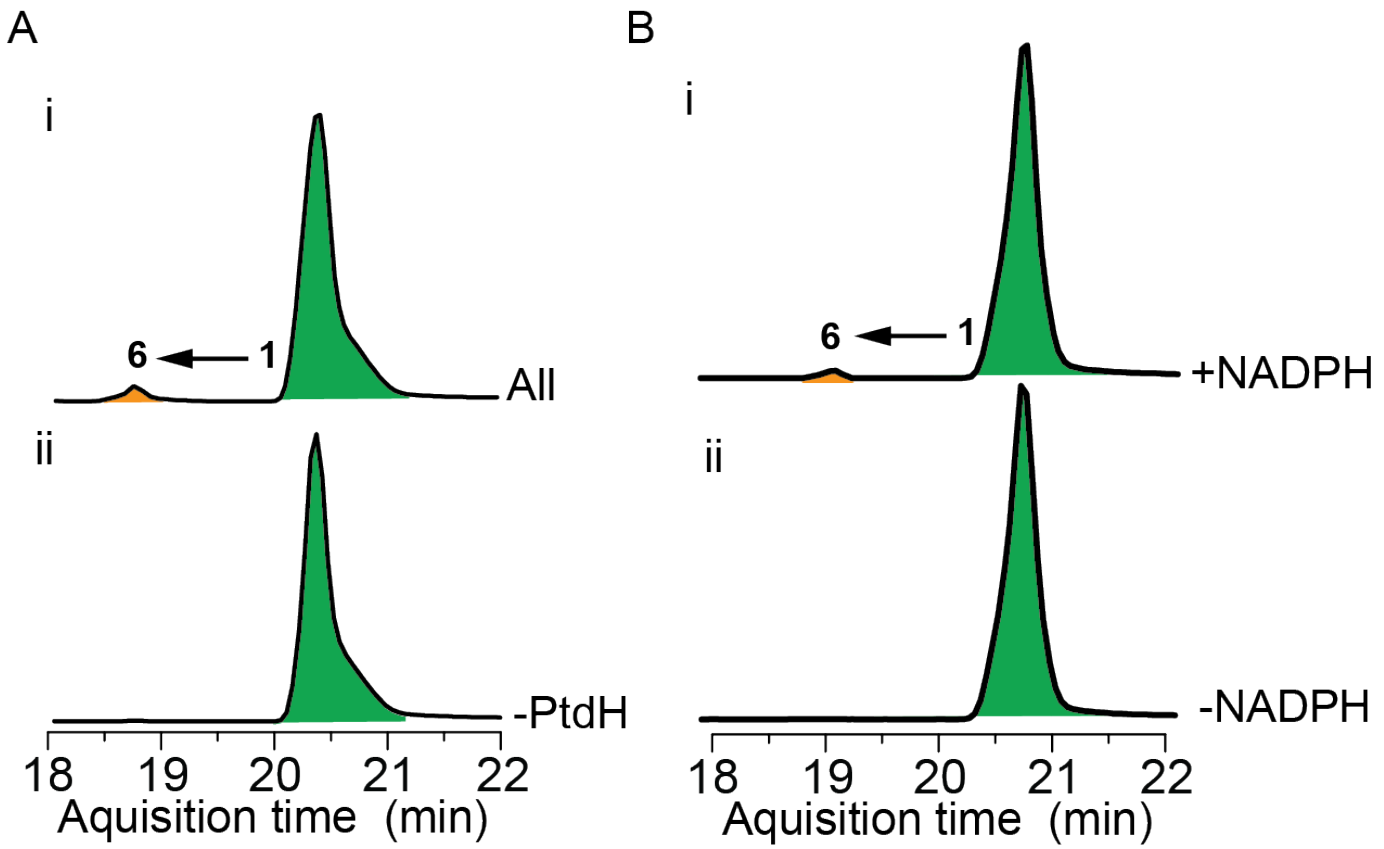

Figure S6. Reductive debromination of 1 by NADPH yields 6. A) EICs for **1** (*m/z* 381.67, shaded green) and **6** (*m/z* 301.76, shaded orange) for the organic extracts of 1 mL reactions incubated in 20 mM Tris-HCl (pH 8) with 50 mM KCl and 10% (v/v) glycerol for 12 h at 30 ºC consisting of 50 µM **1** with 2 mM $NADP^+$, 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; $NADP^+$ 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 ºC 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.

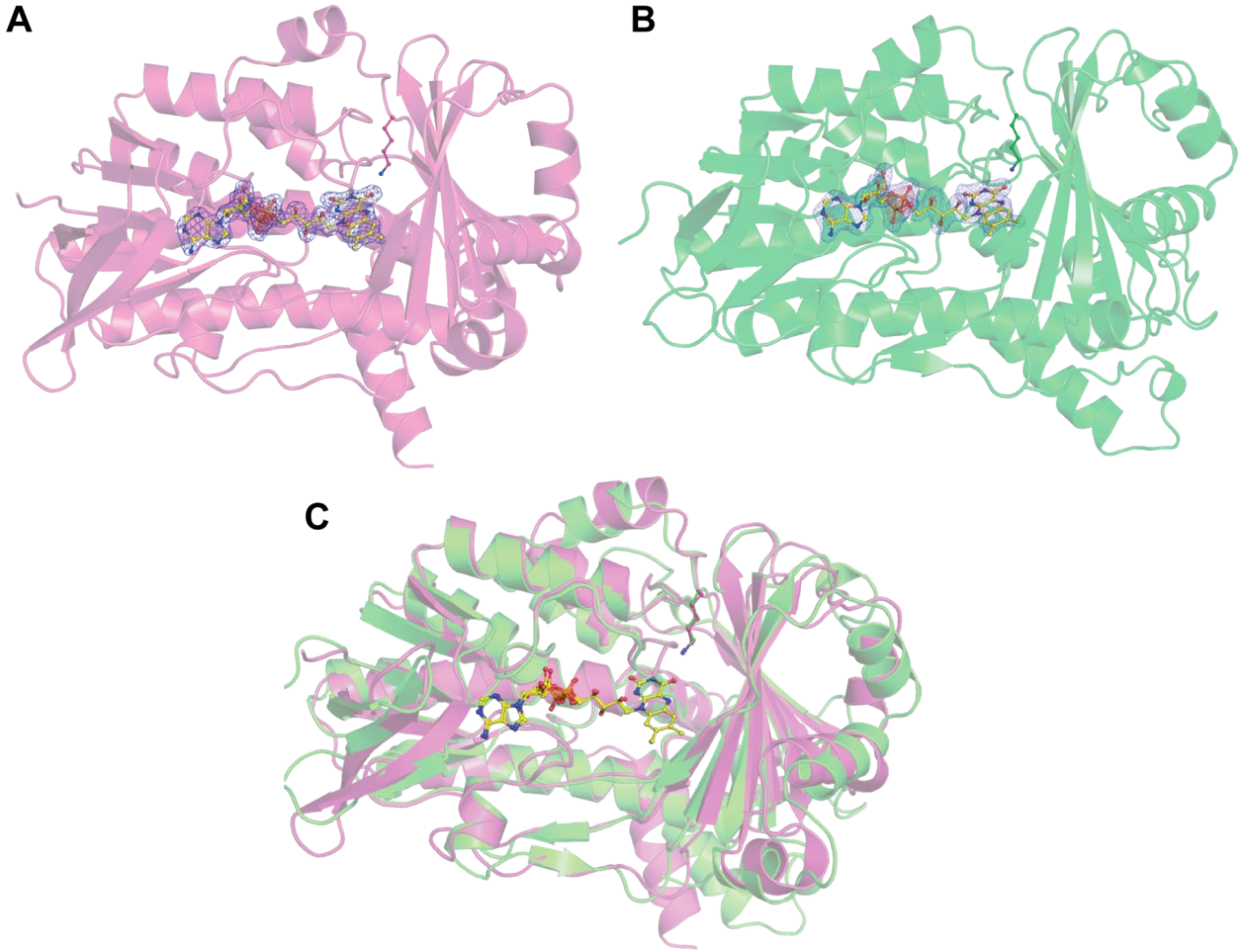

Figure S7. FAD-bound holo structures of pyrrolyl-*S*-ACP flavin-dependent halogenases from this study. (**A**) holo-Mm_Bmp2 structure. (**B**) holo-Mpy16 structure. (**C**) Superimposed Bmp2 (in pink) and Mpy16 (in green) structures. The flavin cofactors and the side chain of the catalytic lysine residues are shown in stick-ball representation.

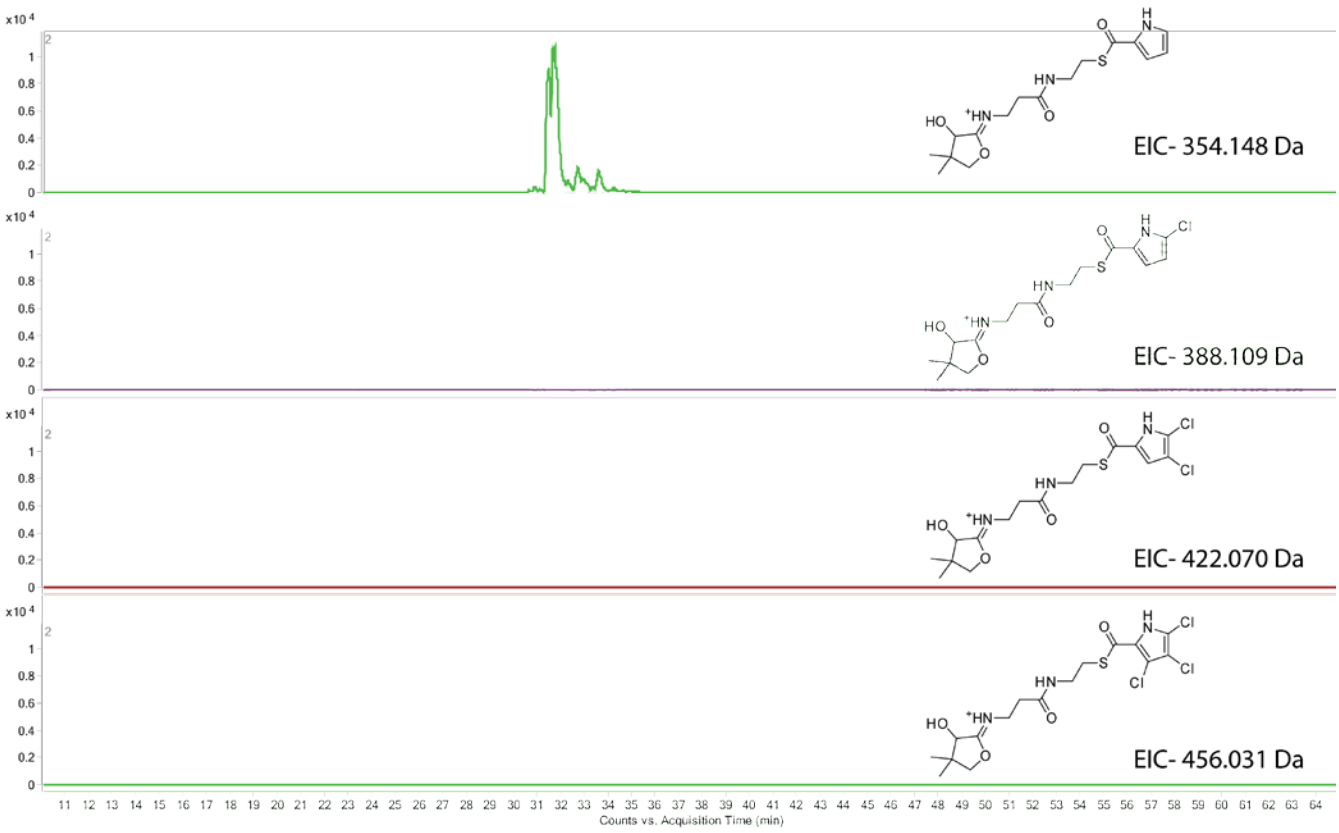

Figure S8. Lack of chlorination activity for Bmp2-TM. Halogenation reactions for Bmp2-TM were carried out as before, but in the absence of bromide, and in the presence of 200 mM chloride in the reaction. As can be seen from the EICs, no mono-, di-, or tri-chlorinated pyrrolyl-*S*-Bmp1(ACP) products can be detected.

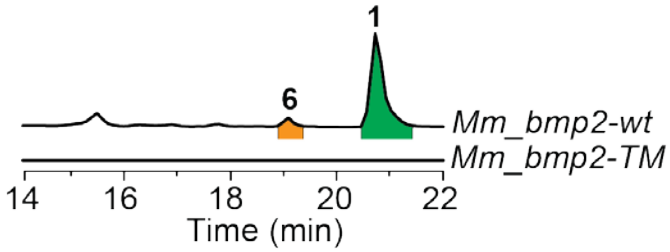

Figure S9. *In vivo* coexpression of *Mm_bmp1*, *Mm_bmp3*, and *Mm_bmp4* with wildtype *Mm_bmp2* (*Mm_bmp2-wt*) or *Mm_bmp2* triple mutant (*Mm_bmp2-TM*). Combined extracted ion chromatograms at $[M-H]^{1-}$ *m/z* 381.67 (**1**) and *m/z* 301.76 (**6**) corresponding to the dominant ions for the molecular formulae C_4NHBr_4 and $C_4NH_2Br_3$, 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).

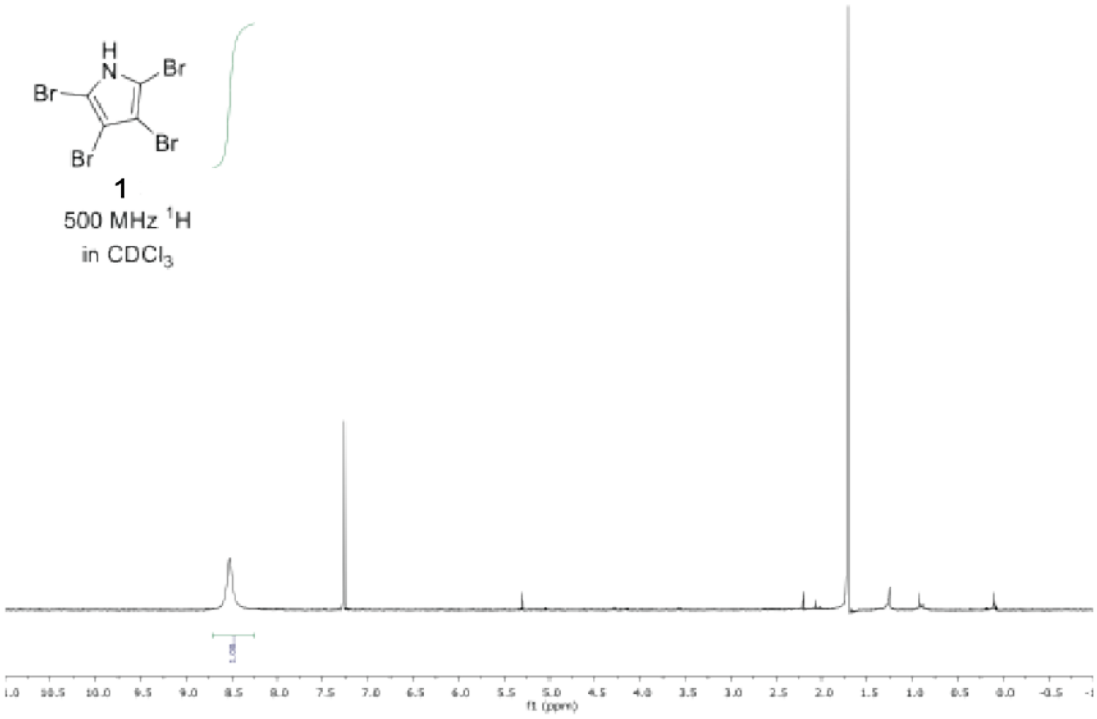

Figure S10. ^{1}H NMR spectrum of **1** ($CDCl_3$, 500 MHz).

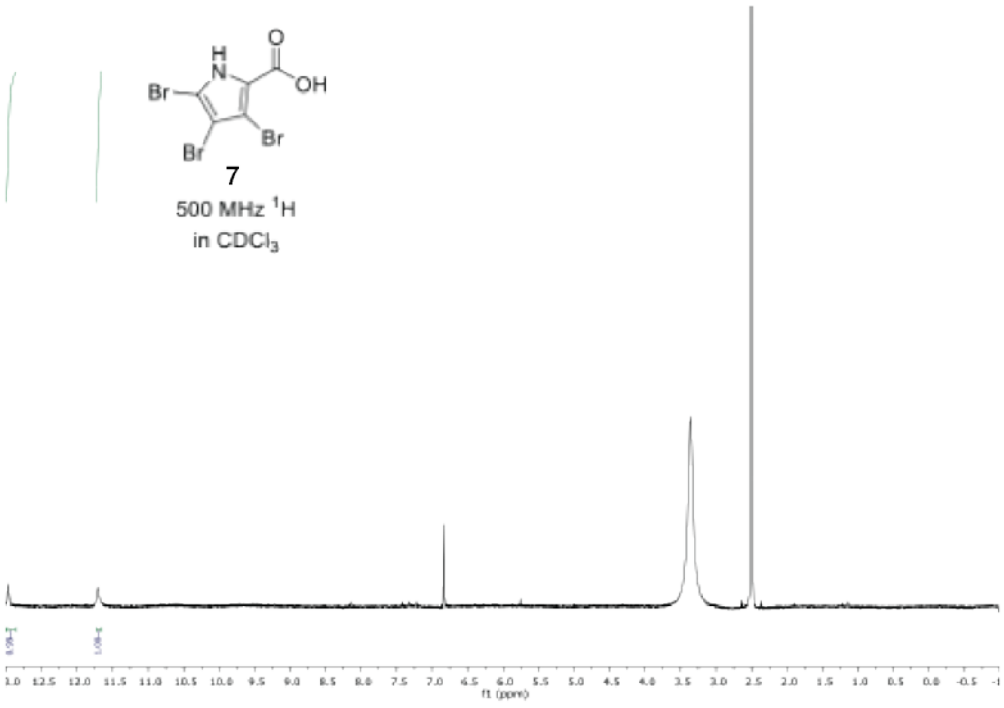

Figure S11. ^{1}H NMR spectrum of **7** ($CDCl_3$, 500 MHz).

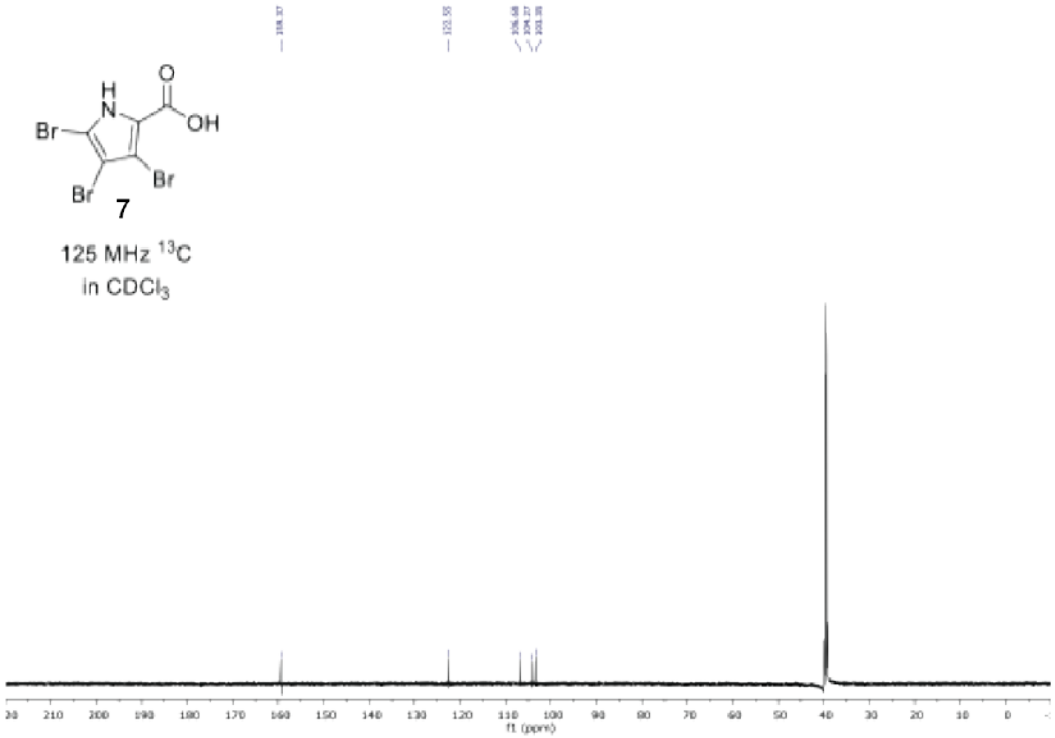

Figure S12. ^{13}C NMR spectrum of **7** ($CDCl_3$, 125 MHz).

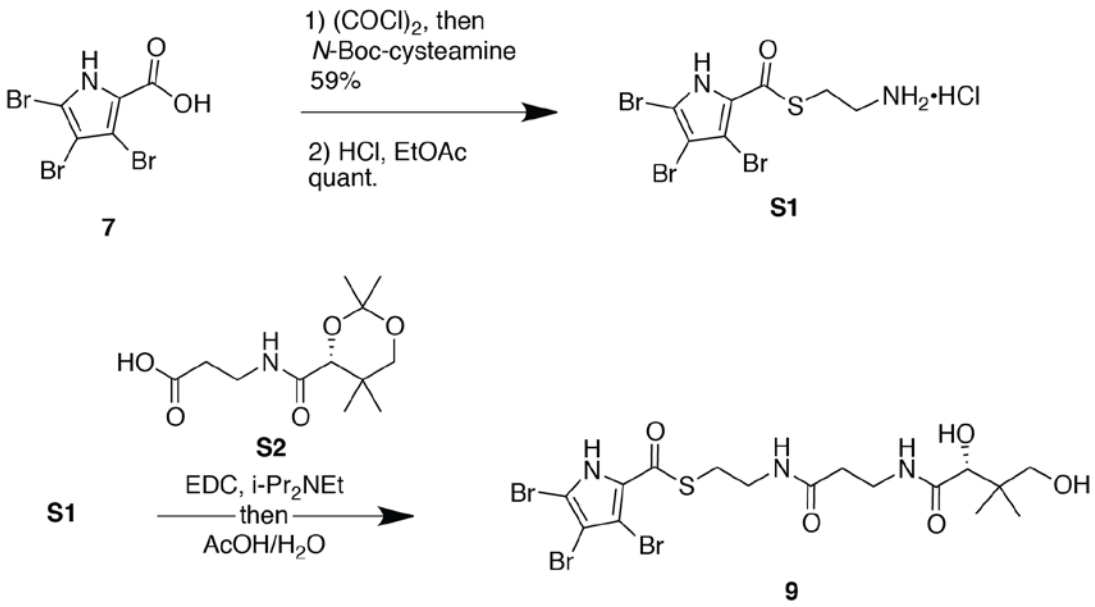

Figure S13. Scheme for the synthesis of **9**.

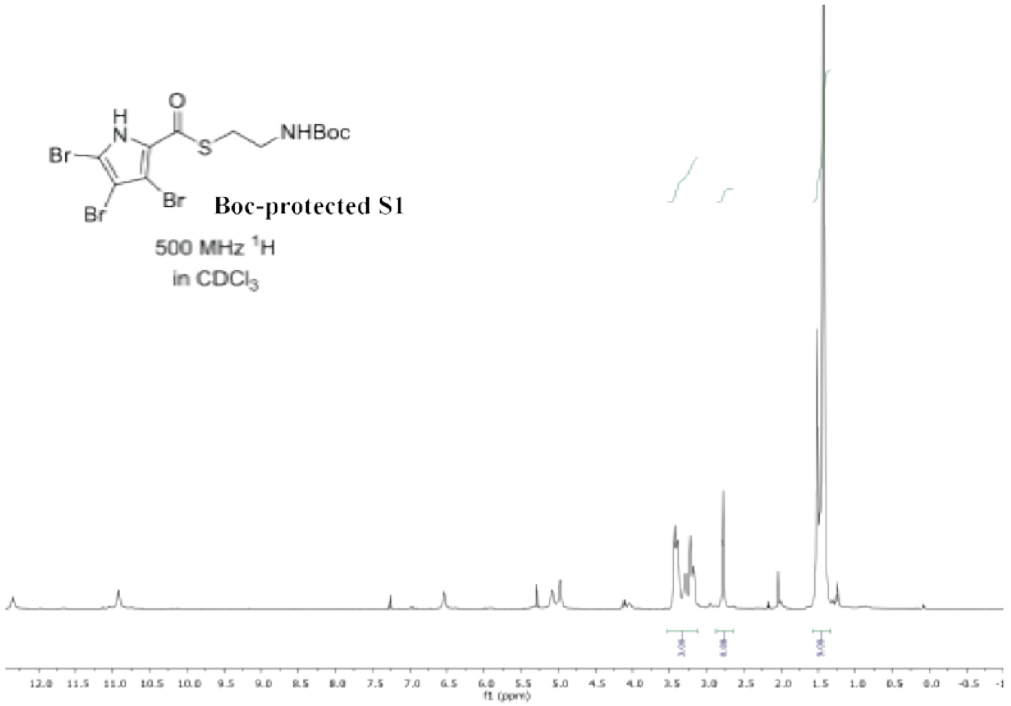

Figure S14. ^{1}H NMR spectrum of Boc-protected **S1** ($CDCl_3$, 500 MHz).

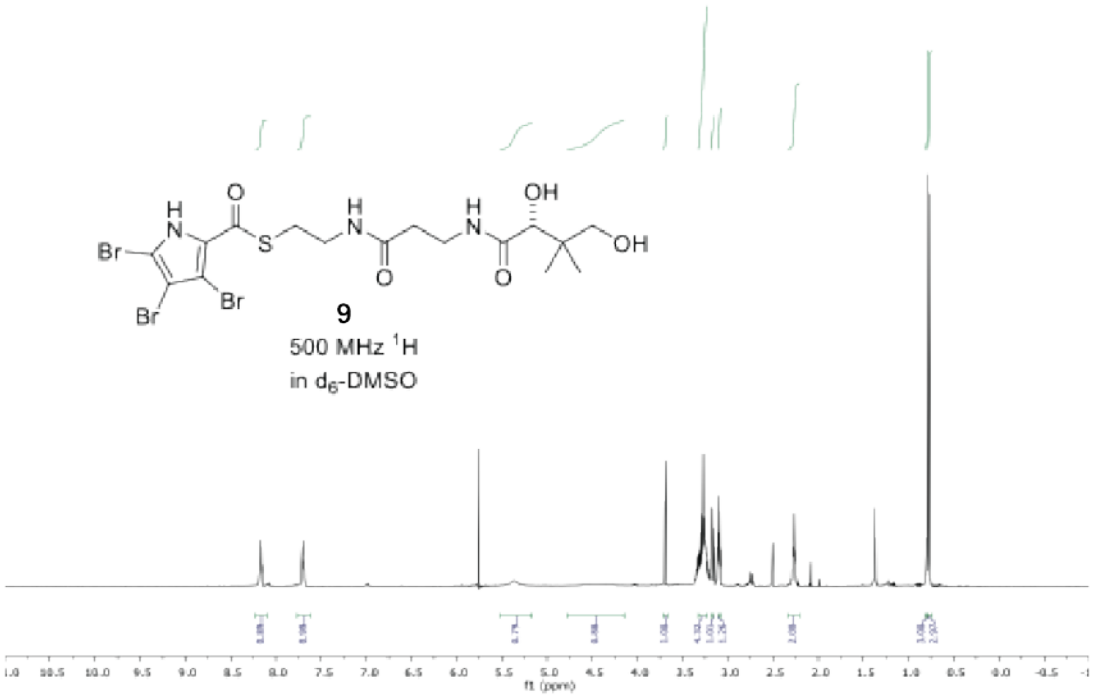

Figure S15. ^{1}H NMR spectrum of **9** (d_6-DMSO, 500 MHz).

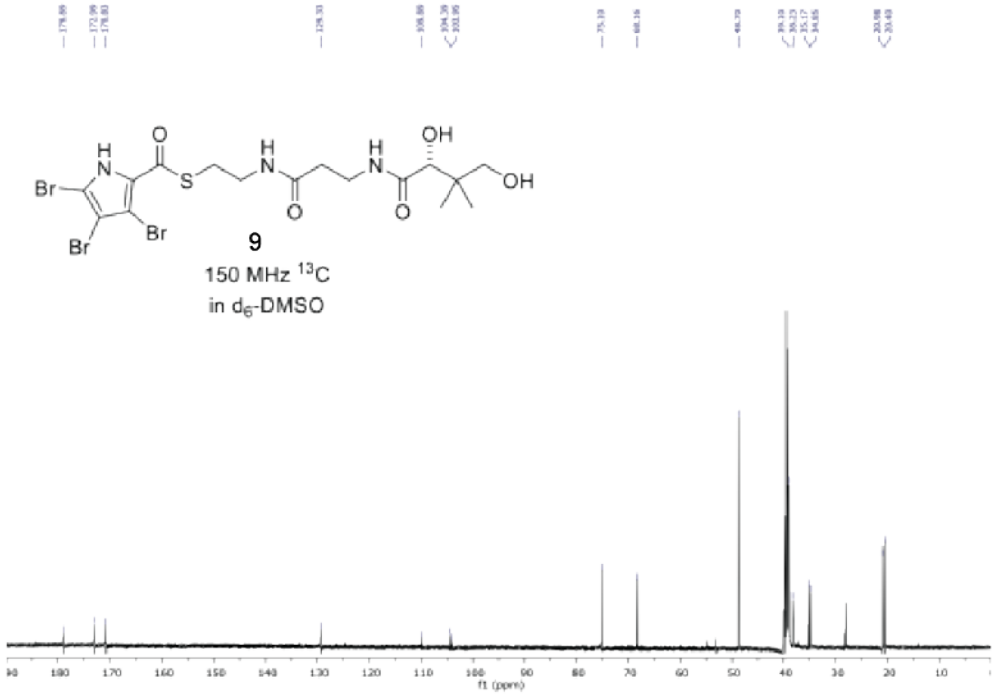

Figure S16. ^{13}C NMR spectrum of **9** (d_6-DMSO, 150 MHz).

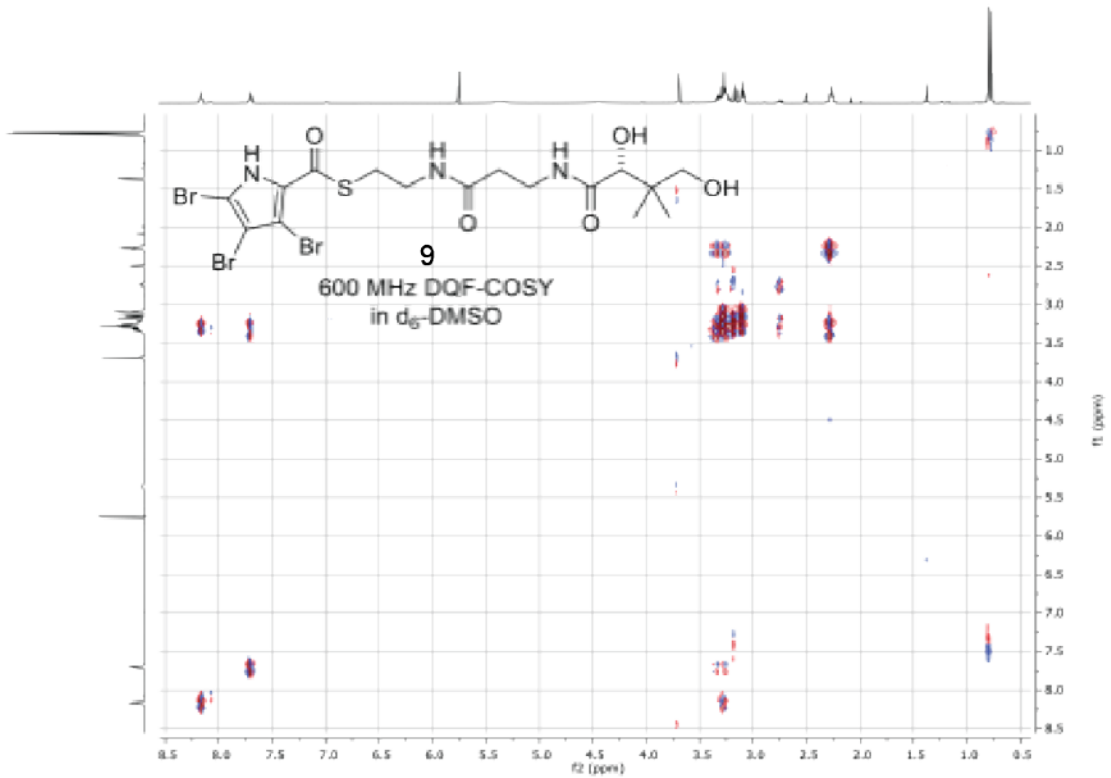

Figure S17. DQF-COSY spectrum of **9** (d_6-DMSO, 600 MHz).

Figure S18. HSQC spectrum of **9** (d_6-DMSO, 600 MHz).

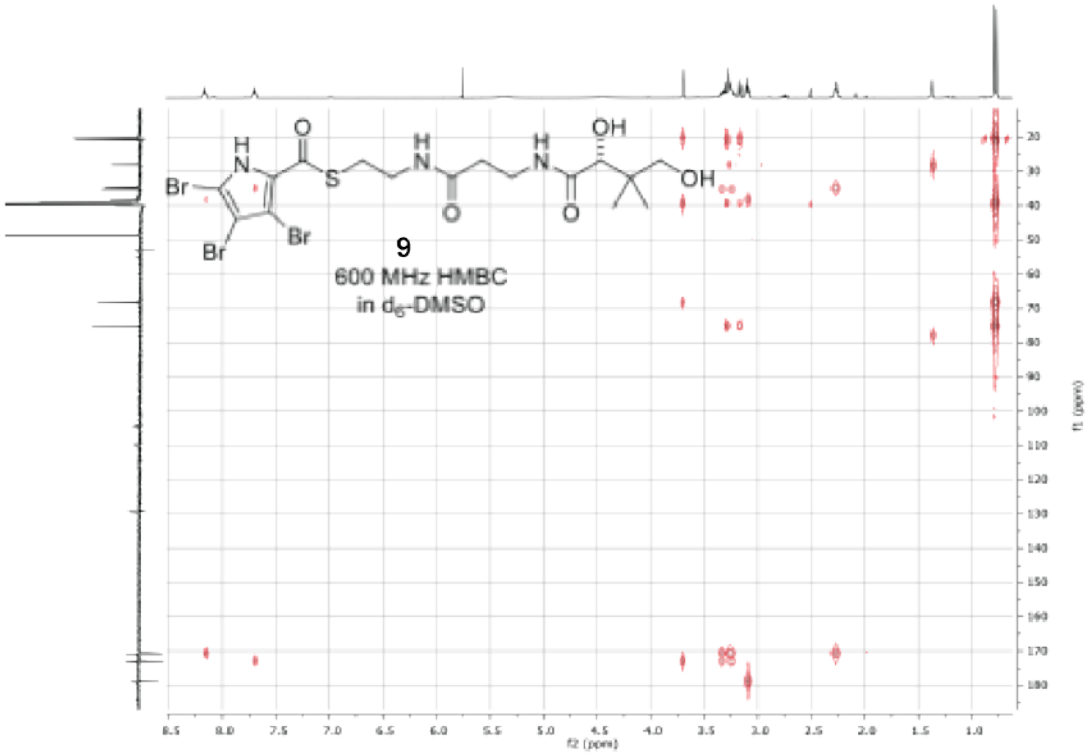

Figure S19. HMBC spectrum of **9** (d_6-DMSO, 600 MHz).

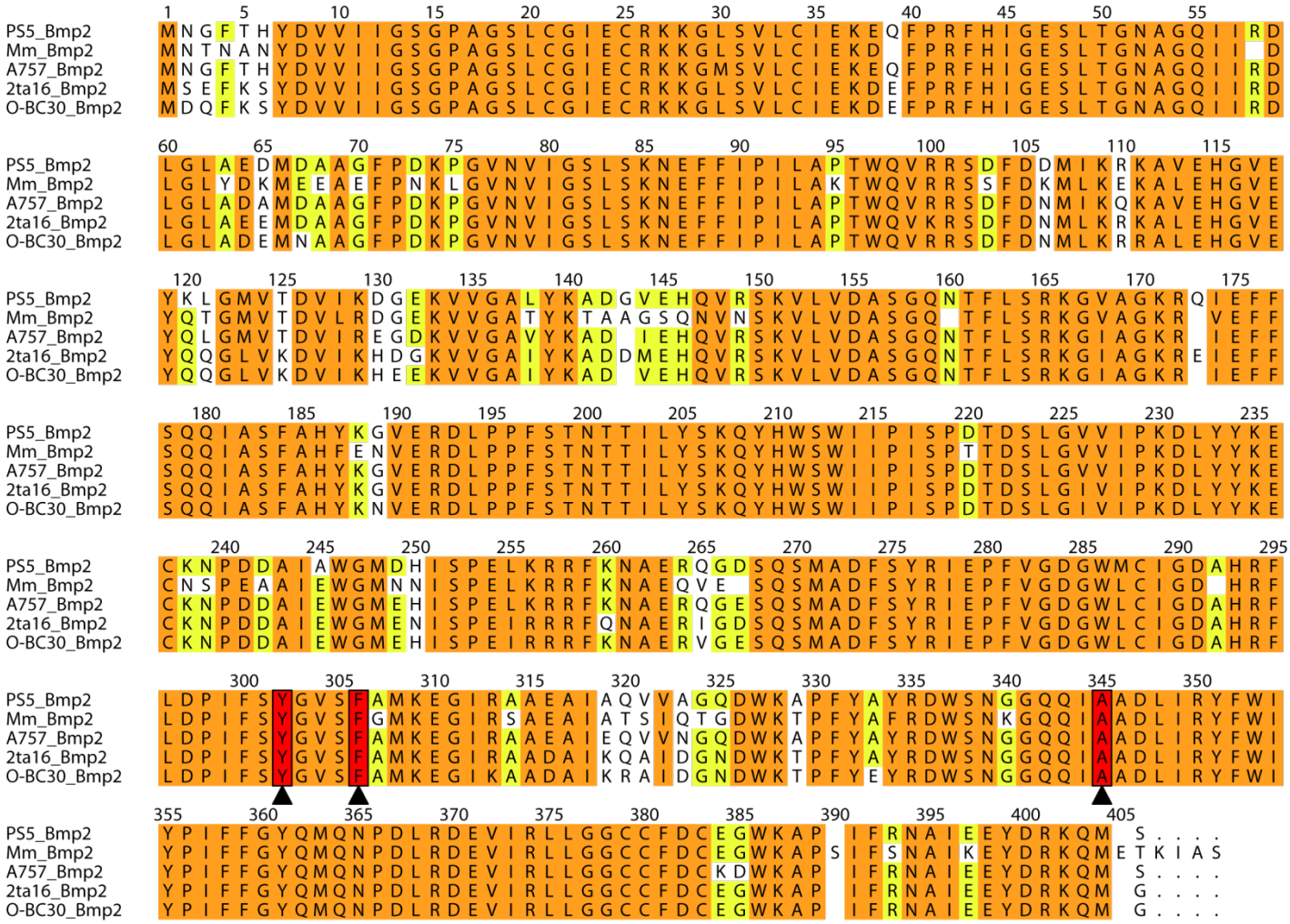

Figure S20. Sequence alignments of Bmp2 homologs. Sequence alignment of Bmp2 homologs from *Pseudoalteromonas* sp. PS5 (PS5_Bmp2), *Marinomonas mediterranea* MMB-1 (Mm_Bmp2), *Pseudoalteromonas* sp. A757 (A757_Bmp2), *Pseudoalteromonas luteoviolacea* 2ta16 (2ta16_Bmp2), and *Pseudoalteromonas phenolica* O-BC30 (O-BC30_Bmp2) showing conservation of Y302, F306, and A345 residues (numbering according to PS5_Bmp2 sequence).

SUPPLEMENTARY TABLES

Table S1. Bmp homology.

Protein name	Annotation	PS5 X Mm (% amino acid similarity/ identity)	PS5 X A757 (% amino acid similarity/ identity)
Bmp1	Acyl carrier protein(ACP)-thioesterase didomain	83/70	98/94
Bmp2	$FADH_2$-dependent pyrrolyl-*S*-ACP brominase	89/83	99/95
Bmp3	$FADH_2$-dependent prolyl-*S*-ACP dehydrogenase	84/76	97/93
Bmp4	L-proline adenylation domain	67/51	93/85
Bmp5	$FADH_2$-dependent p-hydroxbenzoate brominase/decarboxylase	No PS5 homolog found	-
Bmp6	Chorismate lyase	No PS5 homolog found	-
Bmp7	Bromo- pyrrole/phenol coupling cytochrome P450	No PS5 homolog found	-
Bmp8	Carboxymuconolactone decarboxylase/peroxidase	No PS5 homolog found	-
Bmp9	Ferredoxin	73/64	98/93
Bmp10	Ferredoxin reductase	61/42	82/69

Table S2. Data collection and refinement statistics for Mpy16, Bmp2, and Bmp2-TM structures.

	Mpy16	Bmp2	Bmp2-TM
Data collection			
Space group	$P2_1\,2_1\,2_1$	$P2_1\,2_1\,2$	$P2_1$
a, b, c (Å	62.83, 73.28, 189.72	82.97, 90.48, 57.92	49.23, 51.53, 77.06
α, β, γ	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 100.89, 90.00
Rmerge (%)	0.09 (0.64) [1]	0.08 (0.94)	0.06 (0.49)
I/σ(I)	37.6 (3.0)	24.9 (1.9)	30.0 (8.5)
Completeness (%)	100.0 (99.8)	100.0 (100.0)	98.8 (99.5)
Total reflections	64779	36287	26147
Redundancy	12.0 (11.7)	7.0 (7.0)	6.3 (5.9)
Refinement			
Resolution (Å)	38.08 – 1.95	19.23 – 1.87	32.97 – 1.98
R_{work}/R_{free}	0.172/0.212	0.188/0.224	0.179/0.215
Number of atoms			
Protein	6825	3148	3096
Solvent	428	274	219
FAD	106	53	53
Average B values			
Protein	31.53	27.75	26.00
Solvent	35.77	33.99	31.77
FAD	23.14	19.49	16.09
R.m.s. deviations			
Bond lengths (Å)	0.007	0.007	0.007
Bond angles (°)	1.036	1.097	1.085

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

References

1. Bankevich A, *et al.* (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19(5):455-477.
2. Aziz RK, *et al.* (2008) The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9:75.
3. Agarwal V, *et al.* (2014) Biosynthesis of polybrominated aromatic organic compounds by marine bacteria. *Nat Chem Biol* 10(8):640-647.
4. Sneed JM, Sharp KH, Ritchie KB, & Paul VJ (2014) The chemical cue tetrabromopyrrole from a biofilm bacterium induces settlement of multiple Caribbean corals. *Proc Biol Sci* 281(1786).
5. Whalen KE, Poulson-Ellestad KL, Deering RW, Rowley DC, & Mincer TJ (2015) Enhancement of antibiotic activity against multidrug-resistant bacteria by the efflux pump inhibitor 3,4-dibromopyrrole-2,5-dione isolated from a *Pseudoalteromonas* sp. *J Nat Prod* 78(3):402-412.
6. Feher D, Barlow R, McAtee J, & Hemscheidt TK (2010) Highly brominated antimicrobial metabolites from a marine *Pseudoalteromonas* sp. *J Nat Prod* 73(11):1963-1966.
7. Eichhorn E, van der Ploeg JR, & Leisinger T (1999) Characterization of a two-component alkanesulfonate monooxygenase from *Escherichia coli*. *J Biol Chem* 274(38):26639-26646.
8. Gilow HM & Burton DE (1981) Bromination and Chlorination of Pyrrole and Some Reactive 1-Substituted Pyrroles. *J Org Chem* 46(11):2221-2225.
9. John EA, Pollet P, Gelbaum L, & Kubanek J (2004) Regioselective syntheses of 2,3,4-tribromopyrrole and 2,3,5-tribromopyrrole. *J Nat Prod* 67(11):1929-1931.
10. Wang MZ, *et al.* (2011) Design, synthesis and antifungal activities of novel pyrrole alkaloid analogs. *Eur J Med Chem* 46(5):1463-1472.
11. Jansen PA, *et al.* (2013) Discovery of small molecule vanin inhibitors: new tools to study metabolism and disease. *ACS Chem Biol* 8(3):530-534.
12. Agarwal V, *et al.* (2015) Chemoenzymatic synthesis of acyl coenzyme A substrates enables *in situ* labeling of small molecules and proteins. *Org Lett* 17(18):4452-4455.
13. Otwinowski Z, Borek D, Majewski W, & Minor W (2003) Multiparametric scaling of diffraction intensities. *Acta Crystallogr A* 59(Pt 3):228-234.
14. Kantardjieff KA & Rupp B (2003) Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Sci* 12(9):1865-1871.
15. McCoy AJ, *et al.* (2007) Phaser crystallographic software. *J Appl Crystallogr* 40(Pt 4):658-674.
16. Stein N (2008) CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *J App Cryst* 41:641-643.
17. Buedenbender S, Rachid S, Muller R, & Schulz GE (2009) Structure and action of the Myxobacterial chondrochloren halogenase CndH: A new variant of FAD-dependent halogenases. *J Mol Biol* 385(2):520-530.
18. Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1):2126-2132.
19. Perrakis A, Sixma TK, Wilson KS, & Lamzin VS (1997) wARP: improvement and extension of crystallographic phases by weighted averaging of multiple-refined dummy atomic models. *Acta Crystallogr D Biol Crystallogr* 53(Pt 4):448-455.
20. Zwart PH, *et al.* (2008) Automated structure solution with the PHENIX suite. *Methods Mol Biol* 426:419-435.
21. Kleywegt GJ & Brunger AT (1996) Checking your imagination: applications of the free R value. *Structure* 4(8):897-904.
22. Yamanaka K, Ryan KS, Gulder TA, Hughes CC, & Moore BS (2012) Flavoenzyme-catalyzed atropo-selective N,C-bipyrrole homocoupling in marinopyrrole biosynthesis. *J Am Chem Soc* 134(30):12434-12437.
23. Lee J, *et al.* (2013) Structural and functional insight into an unexpectedly selective N-methyltransferase involved in plantazolicin biosynthesis. *Proc Natl Acad Sci U S A* 110(32):12954-12959.

24. Fushinobu S, *et al.* (2002) Crystal structures of a meta-cleavage product hydrolase from *Pseudomonas fluorescens* IP01 (CumD) complexed with cleavage products. *Protein Sci* 11(9):2184-2195.