SUPPLEMENTARY INFORMATION FOR:

Genetic and biochemical reconstitution of bromoform biosynthesis in *Asparagopsis* lends insights into seaweed ROS enzymology

Hem R. Thapa,[1] Zhenjian Lin,[2] Dongqi Yi,[1] Jennifer E. Smith,[3] Eric W. Schmidt,[2] Vinayak Agarwal[1,4,*]

[1]School of Chemistry and Biochemistry, Georgia Institute of Technology

[2]Department of Medicinal Chemistry, University of Utah

[3]Center for Marine Biodiversity and Conservation, Scripps Institution of Oceanography, University of California, San Diego

[4]School of Biological Sciences, Georgia Institute of Technology

*Correspondence: vagarwal@gatech.edu; Ph: 404-385-3798

MATERIALS AND METHODS

Sample collection

Samples of the filamentous, sporophyte phase of the red macroalga *Asparagopsis taxiformis* were collected from five different locations: Guam, California, Hawaii, Florida Keys, and Fiji. Samples of *Macrocystis pyrifera* were collected from the southern California coast (San Diego). Algal biomass for nucleic acid preparation was stored in RNA*later* solution (3.54 M ammonium sulfate, 16.7 mM sodium citrate, 13.3 mM ethylenediaminetetraacetic acid (EDTA), pH 8.0). Biomass for chemical analysis was stored without any treatment. Both types of biomass samples were kept in a cooler during field collection and samples once brought to the lab were stored at -80 °C until further use.

Culturing of *Asparagopsis taxiformis*

Upon collection of the filamentous phase of *A. taxiformis* in California, samples were cultured in the lab at Scripps Institution of Oceanography to allow for further experimentation. Specifically, samples were sorted and cleaned under stereomicroscope to isolate *A. taxiformis* from any other taxa growing within the algal matrix. Roughly 1 g of wet weight was placed in each of several 2-liter Erlenmeyer flasks filled with autoclaved and UV sterilized seawater from the Scripps Institution of Oceanography's running seawater system. Flasks were sealed with stoppers and all samples were gently aerated with ambient air to allow for a gentle tumble. Seawater was replaced weekly and sample biomass to volume ratios were kept below 5g/L. All flasks were cultured on a 12hr day/light cycle at 50 mE/cm^2/s at ambient temperature (roughly 22 °C).

GC-MS analysis of chemical extracts

Algal biomass collected either from the field or laboratory were lyophilized for 16 h. Freeze-dried biomass was used for extraction with MeOH. Biomass was soaked in MeOH for several hours before extraction with vigorous agitation on vortex mixer, centrifuged at 16,000×g for 30 min to remove debris, and an aliquot of supernatant was analyzed by GC-MS (1260G with 7890a MS; Agilent Technologies) in electron ionization (70 eV) mode using a DF-5ms ultra inert GC column (30 m length, 0.25 mm width and 0.5 µM film thickness). Bromoform production quantitation for *A. taxiformis* and *M. pyrifera* was quantified based on calibration curves generated from a bromoform standard. The column temperature conditions were as follows: 40 °C for 3 min, increased to 200 °C at 10 °C/min and held for 1 min with total run time of 20 min. Injection port, interface and ion source were kept at 250 °C, 300 °C and 230 °C, respectively. Helium was used as carrier gas at a flow rate of 0.9 mL/min.

LC-MS analysis

Chemical extracts of *A. taxiformis* prepared above were analyzed using 1290 Infinity II UHPLC system (Agilent Technologies) coupled to a high-resolution Impact II Q-TOF mass spectrometer (Bruker Daltonics). Mass spectrometry data were collected in the negative ionization mode in the mass range *m/z* 100-2000 Da, and the data collected were deposited in the MassIVE repository (ID# MSV000084456). Samples were analyzed using Kinetex™ 1.7 µm C18 reversed phase UHPLC column (50 × 2.1 mm) at a flow rate of 0.5 mL/min and the chromatographic separation was achieved using two solvents; solvent A (H_2O, 0.1% formic acid) and solvent B (MeCN, 0.1% formic acid). The chromatography elution profile was as follows: 5% solvent B from 0-5 min, linear gradient to 100% solvent B from 5-20 min, 100% solvent B from 20-25 min, linear gradient to 5% solvent B from 25-26 min, 5% solvent B from 26-28 min, linear gradient to 100% solvent B from 28-29 min, 100% solvent B from 29-33 min, linear gradient to 5% solvent B from 33-34 min, and 5% solvent B from 34-35 min.

MS^2 and MS^3 characterization of molecule **1** (1,1,1,5,5,5-hexabromo-2,4-dione) was achieved by direct infusion into the Bruker amaZon SL ion-trap mass spectrometer in a negative ionization mode. Algal extract was prepared and enriched for **1** prior to mass spectrometry-based characterization. Freeze-dried algal biomass (55 g) was extracted twice with 100 mL MeOH by stirring at room temperature for 16 h, extracts pooled, filtered, and concentrated using rotary evaporator. The concentrated extract was resuspended in small volume of hexane:EtOAc (40:1) before application to a silica gel chromatography column equilibrated in the same solvent. Elution was done with a gradient of hexane:EtOAc (40:1→10:1→8:1→4:1→2:1). The hexane:EtOAc (2:1) fraction containing molecule **1** was concentrated and then infused in to the mass spectrometer at a flow rate of 4 µL/min using a metered peristaltic pump. Signal was allowed to stabilize before isolation of MS^1 ion of 572.49 and then fragmented to obtain MS^2 ions *m/z* 318.74 and *m/z* 250.75. MS^2 ion *m/z* 318.74 was isolated and fragmented further to obtain MS^3 spectra.

Genomic DNA extraction

DNA from algal samples was extracted following the protocol described in previous study.[1] Algal biomass samples stored in RNAlater were thawed at 4 °C and washed twice with TE buffer (10 mm Tris-HCl (pH 7.5), 1 mM Na-EDTA) to remove excess salts. Biomass was frozen with liquid nitrogen and freeze-dried overnight in a lyophilizer. Dried biomass was ground to fine powder with mortar and pestle using liquid nitrogen. 500 mg of crushed powder was resuspended in 10 mL of CTAB buffer (3% w/v cetyltrimethylammonium bromide (CTAB), 1.4 M NaCl, 20 mM Na-EDTA, 100 mM Tris-HCl (pH 8.0),

0.2% polyvinylpolypyrrolidone, 0.2% β-mercaptoethanol, and 0.2 mg/mL proteinase K). Samples were incubated at 55 °C for 2 h with gentle mixing every 15 min and then centrifuged at 14,000×g for 15 min at room temperature to remove cellular debris. To the supernatant was added 2 mL of 5 M Na-acetate (pH 8.0), cooled on ice for 30 min, and centrifuged at 14,000×g for 15 min at 4 °C. The aqueous layer (supernatant) was extracted with equal volume of phenol:chloroform:isoamyl alcohol ((25:24:1) saturated with 10 mM Tris-HCl (pH 8.0), 1 mM Na-EDTA) and the mixture was centrifuged at 12,000×g for 5 min at 4 °C. The supernatant thus obtained was extracted with equal volume of chloroform and then centrifuged at 12,000×g for 5 min at 4 °C. To the supernatant thus obtained was added equal volume of ice-chilled isopropanol, mixed by inverting the tube, and centrifuged at 12,000×g for 15 min at 4 °C. The supernatant was removed and the pellet containing nucleic acids was washed with 75% (v/v) ethanol, centrifuged as before, supernatant removed, and the pellet was dried in a speedvac to remove any residual ethanol. The dried pellet was resuspended in 0.5 mL of 10 mM Tris-HCl (pH 8.0) containing 0.6 mg/mL of RNase A, incubated at room temperature for 1 h and then incubated overnight at 4 °C to digest any contaminating RNA. The RNase treated sample was extracted with equal volume of phenol:chloroform:isoamyl alcohol followed by chloroform extraction as described before. Na-acetate (pH 5.2) was added to the supernatant to a final concentration of 0.3 M followed by addition of three volumes of ice-chilled ethanol, and then incubated at -20 °C for 2 h to precipitate DNA. The sample was then centrifuged at 16,000×g for 20 min at 4 °C, the pellet washed with 75% (v/v) ethanol, centrifuged again, and the pellet was dried in a speedvac. The DNA pellet was resuspended in desired volume of 10 mM Tris-HCl (pH 8.0) and the quality of genomic DNA was analyzed by gel electrophoresis and nanodrop.

Phylogenetic analysis

Phylogenetic assignment of *Asparagopsis* samples was achieved by PCR amplification and amplicon sequencing of the *18S rRNA* gene. Universal primers, as described in literature, (forward primer 566:5'-CAGCAGCCGCGGTAATTCC-3' and reverse primer 1200:5'-CCCGTGTTGAGTCAAATTAAGC-3') were used for PCR amplification, and covered variable region v4-v5 of *18S rRNA* gene.[2] Intraspecific relationship of *A. taxiformis* samples were inferred by amplification of mitochondrial marker *cox2-cox3* spacer that spans the intergenic region between two conserved gene pair cytochrome oxidase subunit 2 (*cox2*) and cytochrome oxidase subunit 3 (*cox3*). Degenerate primers, as described before, (forward primer cox2:5'-GTACCWTCTTTDRGRRKDAAATGTGATGC-3' and reverse primer cox3:5'-GGATCTACWAGATGRAAWGGATGTC-3') was used for amplification of *cox2-cox3* spacer.[3] PCR reactions were done in total volume of 25 µL and contained 20 ng of genomic DNA, 0.4 µM primers, 2.5 mM dNTPs, 1× Phusion buffer and 0.5 unit of Phusion-high fidelity DNA polymerase. PCR reactions were

carried out in a thermocycler using the following program: initial denaturation at 98 °C for 2 min, 33 cycles of 98 °C for 30 sec, 55 °C for 30 sec, 72 °C for 60 sec for *18S rRNA* gene and 72 °C for 30 sec for *cox2-cox3* spacer, and final extension at 72 °C for 10 min. The PCR amplicons were purified using DNA Clean and Concentrator Kit (Zymo Research). An 'A' nucleotide was added to blunt-ends of PCR product using GoTaq polymerase (Promega). The amplicons were then cloned into pGEM-T Easy vector as per manufacturer's instructions followed by transformation into *Escherichia coli* (DH5α) under appropriate antibiotic selection. Individual colonies were selected for growth and plasmid DNA extracted using standard miniprep protocols followed by Sanger sequencing of the inserts. Sequences were deposited in the GenBank database.

GenBank accession numbers for algal *18S rRNA* region are as follows:

Guam (MN547333)

California (MN547332)

Hawaii (MN547334)

Florida (MN547336)

Fiji (MN547335)

GenBank accession numbers for algal *cox2-cox3* spacer are as follows:

Guam (MN563722)

California (MN563723)

Hawaii (MN563724)

Florida (MN563725)

Fiji (MN563726)

Phylogenetic tree was constructed using the *cox2-cox3* spacer sequence information of five *A. taxiformis* samples from our study and sequences of other *A. taxiformis* samples available in the GenBank database. Tree was constructed in MEGA7 using default parameters of Maximum Likelihood method based on Tamura-Nei model.[3, 4]

Preparation of algal protein extract

Algal biomass for protein extract preparation were harvested, snap frozen in liquid nitrogen and stored at -80 °C until further use. 100 mg of frozen biomass was added to 2 mL safe-lock tube containing 0.8 mL of homogenization buffer (100 mM HEPES-Na (pH 7.6), 10% glycerol) and 1.1 g of 1.4 mm ceramic beads. Samples were homogenized at 4 °C using a Bullet Blender Storm 24 (Next Advance) for 10 min at a power of '10'. The lysate thus obtained was centrifuged at 9,000×g for 10 min to obtain the total protein extract as the supernatant. Typical protein preparation utilized 400-600 mg of total algal biomass which resulted in 2-3 mL of supernatant. The 9,000×g supernatants were pooled and divided equally into two aliquots and then dialyzed separately in dialysis buffer A (100 mM HEPES-Na (pH 7.6), 10% glycerol, 50 mM KCl and 10 µM sodium orthovanadate) to yield holo-enzyme extract and in buffer B (100 mM HEPES-Na (pH 7.6), 10% glycerol, 50 mM KCl and 1 mM Na-EDTA) to yield apo-enzyme extract. Dialysis with Na-EDTA will result in apo-enzyme extract as EDTA is known to form stable complex with vanadate.[5] Dialysis was done using Slide-A-Lyzer dialysis cassette (3.5K MWCO, 3 mL) overnight at 4 °C in 1 L of dialysis buffer A or B. Dialysis was repeated in fresh buffer for another 4 h before using the protein extract for enzyme assays.

Enzyme assays for bromoform production

Enzyme assays were typically done in 1 mL total volume at 30 °C and contained 100 mM HEPES-Na (pH 7.6), 50 mM KBr, 10 µM sodium orthovanadate, 1 mM substrate (**5** or **6**) and 50 µg of protein extract or 1 µM recombinant protein. Reactions were initiated by addition of 1 µL of 1 M H_2O_2, and same volume of H_2O_2 was added every 10 min for a total reaction time of 90 min. Assay with apo-enzyme extract was done in absence of sodium orthovanadate. Reactions were quenched by addition of 1 mL of GC-grade diethyl ether and 100 µL brine. Reaction products were then extracted by vigorous agitation on a vortex mixer, centrifuged at 1,500×g to obtain phase separation, and the top organic layer removed. Extraction was repeated once again with 1 mL of diethyl ether, organic extracts pooled, concentrated using rotary evaporator, and aliquots of extracts were analyzed by GC-MS using conditions described above.

RNA Isolation

Total RNAs from algal sample were isolated as reported before.[6] Algal biomass stored in RNAlater was washed with TE buffer as described earlier and then frozen in liquid nitrogen. 200 mg of frozen tissue was pulverized in liquid nitrogen using mortar and pestle. Frozen powder was added to 1 mL of TRIzol (Invitrogen) and mixed briefly using vortex to obtain homogenous solution. Sample incubated at room

temperature for 5 min, mixed again, and centrifuged at 10,000 x g for 5 min at room temperature to pellet cell debris. Centrifugation repeated when necessary to obtain the clear supernatant. The supernatant mixed with 0.2 mL of chloroform, mixed using vortex, incubated at room temperature for 5 min, and then centrifuged at 12,000×g for 15 min at 4 ºC to obtain phase separation. The supernatant (aqueous phase) containing RNA was carefully removed to a new tube, 0.5 mL of isopropanol added, and mixed by invert mixing. Sample incubated at room temperature for 10 min followed by centrifugation at 12,000×g for 15 min at 4 ºC to pellet RNA. The supernatant was removed, RNA pellet washed with 1 mL of 75% ethanol by inverting mixing and centrifuged at 12,000×g for 15 min at 4 ºC. The supernatant was removed, RNA pellet was partially dried using speed vacuum, and further processed to ensure removal of contaminating polysaccharides. This was achieved by resuspension of the RNA pellet in 2 M LiCl at room temperature followed by centrifugation at 12,000×g for 15 min at 4 ºC. This process was repeated until the size of the RNA pellet did not reduce further. The RNA pellet was then dissolved in 0.5 mL TE buffer, extracted with equal volume of phenol/chloroform/isoamyl alcohol ((25:24:1) saturated with 10 mM Tris-HCl (pH 8.0), 1 mM Na-EDTA), and centrifuged for phase separation. The aqueous phase was further extracted with equal volume of chloroform and centrifuged at 12,000×g for 15 min at 4 ºC. The supernatant containing RNA was treated with 0.1 volume of 3 M Na-acetate (pH 5.2) and 2.5 volumes of ethanol and incubated overnight at -80 ºC to precipitate RNA. The sample was centrifuged at 12,000×g for 15 min at 4 ºC, the supernatant removed, and the RNA pellet was washed with 75% (v/v) ethanol. The pellet was dried in a speedvac followed by resuspension in desired volume of nuclease free water. The isolated total RNA was treated with DNA-free™ DNA removal kit (Invitrogen) to remove any contaminating genomic DNA. Several micrograms of good quality RNA were obtained with A_{260}/A_{280} and A_{260}/A_{230} values greater than 2.0. Analysis of total RNA by agarose gel did not show any genomic DNA contamination and was further confirmed by reverse transcriptase-PCR (RT-PCR) analysis.

Transcriptome sequencing, assembly and analysis

Short-gun sequencing was performed using an Illumina HiSeq 2000 sequencer with a ~350 bp inserts and 125 bp paired-end runs, or an Illumina NovaSeq sequencer with a ~450 bp inserts and 150 bp paired-end runs at the Huntsman Cancer Institute's High Throughput Genomics Center at the University of Utah. Raw reads were trimmed and adaptors removed by Trimmomatic-0.39[7] with parameters (PE -phred33 ILLUMINACLIP TruSeq3-PE-2.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:(80 or 150)). The trimmed reads were merged using BBMerge[8] and then assembled using rnaSPAdes (mink=21 maxk=121 step=10)[9] with standard parameters in the Center for High Performance Computing at the University of Utah.

The eukaryotic genes predicted by AUGUSTUS were used as reference genes, trimmed reads of each sample were multi-mapped to the reference using Salmon6 with parameters (salmon index -t -i index -k 31; salmon quant --index --validateMappings --libType A --dumpEq -r). The identical predicted genes from both samples were hierarchically clustered into clusters, and then cluster count was summarized using Corset 1.047 according to shared reads information with parameters (-f true -g -n -i salmon_eq_classes). Transcript abundance of each gene was estimated using EdgeR8 by normalized counts per million (cpm).

RT-PCR and cloning

All genes from *A. taxiformis* that are characterized in this study were cloned using cDNA as template. GeneBank accession numbers for *A. taxiformis* ACP and FabH are AOM65890.1 and AOM66007.1, respectively. The cDNA was prepared using SuperScript IV first-strand synthesis system (Thermo Fisher). The first step of cDNA synthesis reaction was set up in 13 µL reaction volume containing 600 ng of total RNA, 1 µL of 10 mM dNTP mix, and 1 µL of 50 µM oligo dT_{20} primer. The reaction was incubated at 65 °C for 5 min, incubated in ice for 2 min, and then added to a tube containing 4 µL of 5x superscript buffer, 1 µL of 100 mM DTT, 1 µL of ribonuclease inhibitor, and 200 unit of Superscript IV enzyme. The reaction mix was incubated at 55 °C for 10 min followed by heat inactivation at 80 °C for 10 min. The reaction mix was further treated with 2 unit of RNase H by incubating at 37 °C for 20 min. The product (3 µL) from cDNA synthesis reaction was used as template for PCR with high fidelity Phusion DNA polymerase and gene specific primers. Negative control reaction for RT-PCR was identical to other reactions except omission of the reverse transcriptase enzyme. RT-PCR products were cleaned with DNA clean and concentrator kit, nucleotide A added using GoTaq DNA polymerase, and then cloned into pGEM-T Easy vector as described above followed by transformation into *E. coli* DH5α, colony selection and miniprep. The pGEM-T constructs were used as template for cloning into desired expression vectors.

Cloning and protein purification

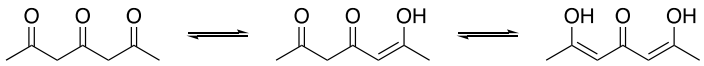
For characterization of *C. crispus* VHPOs[10], three of the five VHPOs with highest sequence identity to *A. taxiformis* Mbb4 protein were chosen, gene fragments optimized for expression in *E.coli* were synthesized (Twist Biosciences), and cloned into pGEM-T Easy vector. The *Cc*VHPOs were arbitrarily numbered. The GenBank accession numbers of *Cc*VHPO1, *Cc*VHPO2 and *Cc*VHPO3 are XP_005714237.1, XP_005719435.1 and XP_005710894.1, respectively. The pGEM-T gene construct for each gene was used as a template for PCR with high fidelity Phusion DNA polymerase. The MatB gene construct as described in previous study[11] was a gift from Dr. Keatinge-Clay (UT Austin) and was used as a template for cloning

into pET28MBP vector. The PCR amplicons for each gene were cloned into expression vector using NEBuilder HiFi DNA Assembly master mix following the manufacturer's instructions. The *E. coli* expression constructs used in this study are listed in Supplementary Table 3. The expression constructs for algal genes were transformed into *E. coli* (Rosetta) cells, whereas MatB and Sfp constructs were transformed into *E. coli* (BL21DE3) cells. In a typical protein expression experiment, 10-mL overnight culture was used to inoculate two 1-liter terrific broth media (1-liter for MatB and Sfp) containing appropriate antibiotic (kanamycin at 50 mg/L, ampicillin at 100 mg/L and chloramphenicol at 34 mg/L). *E. coli* cultures were grown at 30 ºC with shaking at 180 rpm until OD_{600} of 0.8, cultures cooled down to 18 ºC, incubated for another 1 h at 18 ºC, and then induced with 0.3 mM IPTG. All subsequent steps for purification were done on ice or at 4 ºC using the standard protocol unless mentioned otherwise. Cultures were harvested by centrifugation, pellet resuspended in lysis buffer (20 mM Tris pH 8.0, 500 mM NaCl, 1 mM EDTA), cells lysed by sonication, centrifuged at 30,000×g for 40 min, and the supernatant loaded into 5 mL His-Trap Ni-NTA column using AKTA Prime FPLC system. Column was washed extensively with wash buffer (20 mM Tris pH 8.0, 500 mM NaCl, 30 mM imidazole) and protein eluted using linear gradient to 100% elution buffer (20 mM Tris pH 8.0, 500 mM NaCl, 250 mM imidazole) for a total volume of 40 mL. Purity of eluant fractions were checked using SDS-PAGE, and fractions containing protein of interest were pooled and concentrated to 2.5 mL using Amicon centrifugal filters. Protein samples were desalted using PD-10 column and eluted in storage buffer (20 mM HEPES pH 7.5, 50 mM KCl, 10% glycerol). Purified proteins were stored as small aliquots at -80 °C and fresh aliquots of each protein were used for enzyme assays. None of the proteins characterized in this study showed loss of activity during storage at -80 °C for several months.

Monocholorodimedone (MCD) Assay

Halogenation activity of Mbb enzymes and *Cc*VHPOs were determined using MCD assay.[5] Reactions were performed in a quartz cuvette at 25 ºC, and the decrease in absorbance of MCD substrate were monitored spectrophotometrically at 290 nm every 30 sec for 15 min using NanoDrop OneC UV-Vis spectrophotometer. Reactions were initiated by adding the recombinant enzymes in a 0.75 mL total reaction volume containing 100 mM Na-citrate (pH 6.5), 100 mM KBr or KCl, 100 µM MCD, 10 µM sodium orthovanadate, and 2 mM hydrogen peroxide. Assays were done in triplicate for total of six enzymes at a varied concentration, Mbb1-250 nM, Mbb3-1 µM, Mbb4-100 nM, *Cc*VHPO1-50 nM, *Cc*VHPO2-50 nM, and *Cc*VHPO3-250 nM. Enzyme concentrations used in the assays were empirically determined due to the difference in the activity of individual enzymes.

Synthesis of heptane-2,4,6-trione (6)



6a **6** **6b**

6 was prepared as described before.[12] Commercial 2,6-dimethyl-γ-pyrone (500 mg, 4.03 mmol) was dissolved in 2.5 mL ethanol and treated with 0.5 mL of 16 M NaOH. The reaction mixture was stirred at 60 ºC for 5 h, and then heated at 100 ºC for another 1 h. Precipitate formed during the reaction was filtered and washed with diethyl ether. The crystals were dissolved in water followed by the addition of 5 mL of 3 M HCl solution. The aqueous phase was extracted thrice with diethyl ether. The organic layer was dried with anhydrous Na-sulfate and concentrated under vacuum which afforded the target compound as a yellow solid. Tautomeric equilibrium of product in chloroform-*d* was 68.8% **6**, 24.6 % **6b** and 6.6 % **6a**. **^{1}H-NMR** (400 MHz, $CDCl_3$) δ 15.21 (s, 1H), 14.18 (s, 1H), 5.56 (s, 1H), 5.14 (s, 1H), δ 3.70 (s, 4H), δ 3.40 (s, 2H), 2.25 (s, 6H), 2.08 (s, 3H), 1.98 (s, 3H). **^{13}C-NMR** (101 MHz, $CDCl_3$) δ 202.07, 193.88, 191.23, 187.02, 178.65, 101.21, 98.76, 57.85, 54.12, 30.52, 24.76, 21.99.

Synthesis of malonyl-CoA

Previously described enzyme MatB was used to synthesize malonyl-CoA.[11] MatB enzyme assay was performed in a 100 µL reaction volume containing 100 mM HEPES-Na (pH 7.6), 20 mM $MgCl_2$, 20 mM Na-malonate, 20 mM ATP, 5 mM TCEP, 5 mM CoA-SH, 15% (v/v) glycerol and 5 µM MatB. Reaction was incubated at 30 ºC for 23 h. Control reaction omitted MatB. An aliquot of reaction was quenched with equal volume of solvent (MeOH + 1% trifluoroacetic acid (TFA) for HPLC analysis and MeCN + 0.2% formic acid for LC-MS/MS analysis), precipitate removed by centrifugation, and supernatant analyzed by HPLC system and LC-MS/MS using Phenomenex Luna C8 column 5 µm (250 × 4.6 mm) column. LC-MS/MS analysis was done using Agilent 1290 Infinity II UHPLC system coupled to a high-resolution Bruker Impact II Q-TOF mass spectrometer. MS data were collected in the positive ionization mode from *m/z* 100-2000 Da. HPLC analysis used solvent A (100% H_2O, 0.1% TFA) and solvent B (100% MeOH, 0.1% TFA). LC-MS analysis used solvent A (100% H_2O, 0.1% formic acid) and solvent B (100% MeCN, 0.1% formic acid). Samples were analyzed at a flow rate of 0.5 mL/min, and elution profile for both HPLC and LC-MS/MS analysis were: 5% solvent B from 0-3 min, linear gradient to 25% solvent B from 3-8 min, linear gradient to 50% solvent B from 8-20 min, linear gradient to 100% solvent B from 20-23 min, 100% solvent B from 23-26 min, linear gradient to 5% solvent B from 26-27 min, 5% solvent B from 27-28 min, linear gradient to 100% solvent B from 28-29 min, 100% solvent B from 29-30 min, and linear gradient to 5% solvent B from 30-31 min.

Synthesis of malonyl-*S*-ACP

Malonyl-CoA synthesized above was used as the substrate to modify ACP using Sfp enzyme.[13] Assays were performed at 30 ºC for 3 h in a 100 µL total reaction volume containing 100 mM HEPES-Na (pH 7.6), 500 µM ACP, 50 µL reaction product from MatB assay, 10 mM $MgCl_2$, and 2 µM Sfp. Control reaction omitted the Sfp enzyme. An aliquot of reaction was quenched by adding equal volume of MeCN + 0.2% formic acid, precipitate observed was removed by centrifugation, and supernatant was analyzed by LC-MS/MS using Aeris 3.6 µm widepore XB-C18 (200 Å 250 × 4.6 mm) column at a flow rate of 0.5 mL/min. MS data were collected in the positive ionization mode from *m/z* 100-2000 Da, and data analysis for ACP proteins (with or without acyl-groups) were done following the method described in previous study.[14] The elution profile for LC-MS/MS analysis was as follows: 5% solvent B from 0-5 min, linear gradient to 70% solvent B from 5-30 min, linear gradient to 95% solvent B from 30-31 min, 95% solvent B from 31-35 min, linear gradient to 5% solvent B from 35-36 min, 5% solvent B from 36-38 min, linear gradient to 95% solvent B from 38-39 min, 95% solvent B from 39-42 min, linear gradient to 5% solvent B from 42-43 min, and 5% solvent B from 43-46 min.

Synthesis of acetoacetyl-*S*-ACP (8)

Synthesis of **8** was performed at 30 ºC for 3 h in a 100 µL total reaction volume containing 100 mM HEPES-Na (pH 7.6), 50 µL reaction product from Sfp loading assay, 1 mM acetyl-CoA, 1 mM TCEP and 2 µM FabH. After 3 h incubation, an aliquot of reaction was quenched with equal volume of MeCN + 0.2% formic acid, precipitate removed by centrifugation and analyzed by LC-MS/MS using the procedure described above for malonyl-*S*-ACP.

Preparative scale synthesis of 8 and enzyme assay

2×2.5 mL assays for synthesis of malonyl-CoA were set up as described above. After incubation at 30 ºC for 23 h, both samples were pooled, and reaction product was used to set up 4×2.5 mL assays for synthesis of malonyl-*S*-ACP using the reaction conditions described above. After incubation of Sfp loading reaction at 30 ºC for 3 h, precipitate was removed by centrifugation at 1500×g for 10 min, and the supernatant was used to set up 8×2.5 mL assays with the FabH enzyme. Reaction conditions for synthesis of acetoacetyl-*S*-ACP were same as described above. After incubation at 30 ºC for 3 h, samples were pooled, centrifuged, and the supernatant was concentrated to 2.5 mL using Amicon centrifugal filters (3 kDa MWCO). Concentrated sample was loaded onto a desalting PD-10 column, and acetoacetyl-*S*-ACP was eluted using the buffer 20 mM HEPES-Na (pH 7.5), 50 mM KCl, 10% glycerol. Aliquots of acetoacetyl-*S*-ACP were

stored at -80 °C. Production of malonyl-CoA, malonyl-*S*-ACP and acetoacetyl-*S*-ACP in each reaction step was verified using LC-MS/MS. Enzyme assays for bromoform synthesis were performed at 30 ºC for 1.5 h in a 1 mL reaction volume containing 100 mM HEPES-Na (pH 7.6), 50 mM KBr, 10 µM sodium orthovanadate, 0.67 mM acetoacetyl-*S*-ACP and 1 µM Mbb1 or Mbb4 enzymes. Reactions were initiated by addition of 1 µL of 1 M H_2O_2. H_2O_2 was replenished by adding 1 µL of 1 M H_2O_2 after every 10 min through the course of reaction. After 90 min, reactions were quenched by addition of equal volume of diethyl ether and 100 µL brine. Reactions were extracted by vigorous agitation, centrifuged at 1,500×g for 10 min and the top organic layer removed. Extraction was repeated thrice. Organics extracts were pooled, concentrated, and 10 µL of concentrated extract was analyzed by GC-MS using manual injection using the conditions described above.

Metagenome sequencing, assembly, binning and annotation.

Short-gun sequencing was performed using an Illumina NovaSeq sequencer with a ~450 bp inserts and 150 bp paired-end runs at the Huntsman Cancer Institute's High Throughput Genomics Center at the University of Utah. Raw reads were trimmed and adaptors removed by Trimmomatic-0.39[7] with parameters (PE -phred33 ILLUMINACLIP TruSeq3-PE-2.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:150). The trimmed reads were merged using BBMerge[8] and then assembled using metaSPADES (mink=21 maxk=121 step=10)[9] with standard parameters in the Center for High Performance Computing at the University of Utah. Bacterial contigs were binned out of metagenome assembly using Autometa.[15] The eukaryotic contigs were annotated using AUGUSTUS[16] with transcriptome assembly as training data.

Quantitative RT-PCR

The total RNA from each sample were isolated and cDNA synthesized as above. The product from cDNA synthesis reaction was diluted five-fold in nuclease free water and 3 µL was used as template in a 20 µL PCR reaction containing 10 µL of PowerUp SYBR green master mix (ThermoFisher) and 0.5 µM primers. The qRT-PCR reactions were carried out on a StepOnePlus Real-Time PCR system (Applied Biosystems) using following program: initial denaturation at 95 °C for 2 min, 40 cycles of (95 °C for 15 sec, 55 °C for 15 sec and 72 °C for 60 sec. Expression level for each gene were calculated relative to actin gene using the $2^{-\Delta\Delta CT}$ method.[17] Primers used for qRT-PCR are listed on supplementary Table 4.

Cloning of *yno1*, *mbb2* and *CcNOX (Ccmbb2)* in yeast expression plasmids

The *yno1*/pYES2 plasmid and BY4741 yeast strain (MATa, *his3*Δ1, *leu2*Δ0, *met15*Δ0, *ura3*Δ0, *yno1Δ*), which has a deletion of *yno1* gene, was a gift from Reddi laboratory at Georgia Tech. Synthetic clones for *mbb2* and *CcNOX* with sequences optimized for expression in yeast were obtained commercially. All genes, *yno1*, *mbb2*, and *CcNOX (Ccmbb2)*, were amplified using high fidelity Phusion DNA polymerase and PCR amplicons were cloned into pESC-URA expression vector using NotI and SpeI restriction sites. Gene constructs were transformed into *yno1Δ* yeast strain using standard lithium acetate protocol.[18] Transformants were selected and patches of positive colonies were maintained on synthetic dropout medium (SC-URA) supplemented with 2% glucose.

Dihydroethidium (DHE) Assay

Reactive oxygen species production by yeast cells were measured using DHE as a fluorescent probe and following the methodology as described before.[19] Yeast cultures were maintained in SC-URA supplemented with 2% raffinose. Triplicate cultures for each gene constructs were used for DHE assay. Several time points and galactose concentration were screened to determine optimal signal for DHE assay. In brief, 25 µL of overnight cultures was used to inoculate 5 mL of media containing 10% galactose. The cultures were grown at 30 ºC by shaking at 200 rpm for 43 h, harvested by centrifugation, washed, and resuspended in distilled water. Cells corresponding to 6.25 OD_{600} units were harvested by centrifugation at 6,000 x g and washed three times with PBS buffer. Cells were resuspended in 300 µL PBS buffer, DHE added to final concentration of 10 µM, and then incubated in dark at 30 ºC for 10 min with invert mixing every two min. After incubation, cells were harvested by centrifugation, washed at least three times with PBS buffer, pellet resuspended in 300 µL PBS buffer, 100 µL of cell suspension aliquoted to black microwell plate, and fluorescence measured using Biotek Synergy Mx microplate reader with an excitation wavelength of 518 nm and an emission wavelength of 605 nm.

SUPPLEMENTARY FIGURES

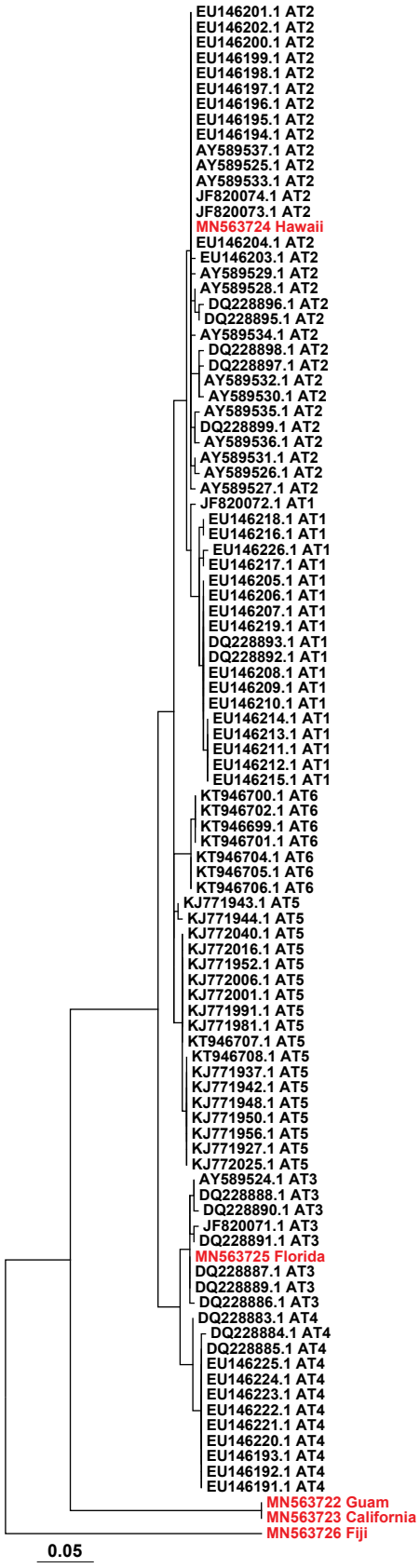


Figure S1: Maximum Likelihood phylogenetic tree of *Asparagopsis taxiformis* using *cox2-cox3* spacer. GenBank accession number followed by lineage information is shown for all *cox2-cox3* spacer sequences. Samples from this study are highlighted in red. The evolutionary history was inferred by using the Maximum Likelihood method based on the Tamura-Nei model.[4] The tree with the highest log likelihood (-1151.26) is shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood[10] approach, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 100 nucleotide sequences. All positions containing gaps and missing data were eliminated. There was a total of 267 positions in the final dataset. Evolutionary analyses were conducted in MEGA7.[20]

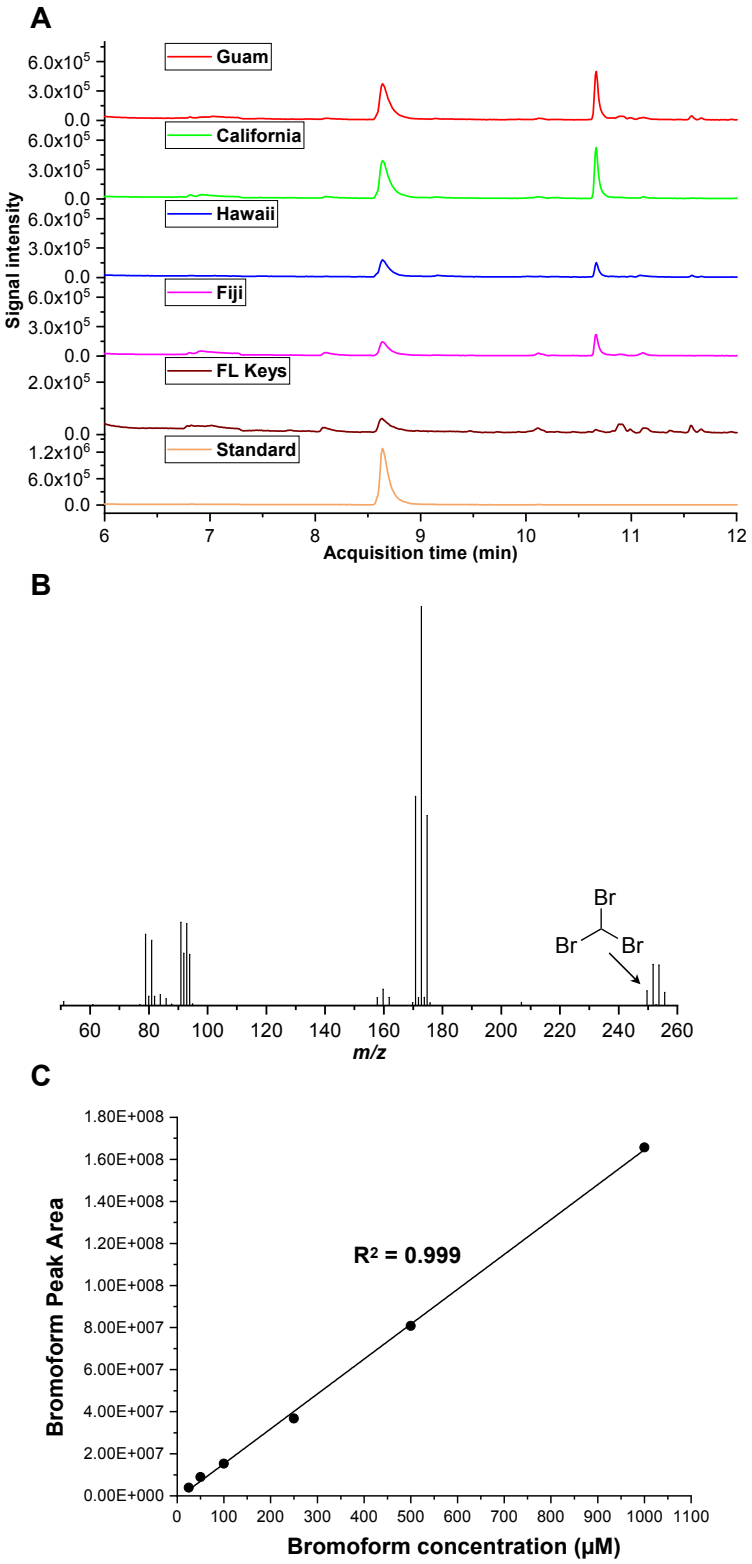


Figure S2: Bromoform production in red macroalga *Asparagopsis taxiformis*. (**A**) GC-MS chromatograms for MeOH extracts demonstrating bromoform production by *A. taxiformis* in geographically disperse samples. (**B**) MS spectrum of bromoform. Arrow corresponds to monoisotopic ion. (**C**) Calibration curve generated based on GC-MS analysis of different concentrations of bromoform standard. 25 mg freeze dried algal biomass for each sample was soaked in 625 µL MeOH for several hours, metabolites extracted using vortex mixer, and then centrifuged. The supernatant (MeOH extracts) was diluted 25-fold in DCM and 1 µL was analyzed by GC-MS. Bromoform peak area for each sample was determined and then used for quantification using bromoform standard calibration curve. Representative calculation: peak area 46,655,308 for California sample corresponds to 294.405 µM of bromoform. When adjusted for dilution, peak area is equivalent to 46.5 µg of bromoform in 1 mg of dry biomass (4.65% dry weight).

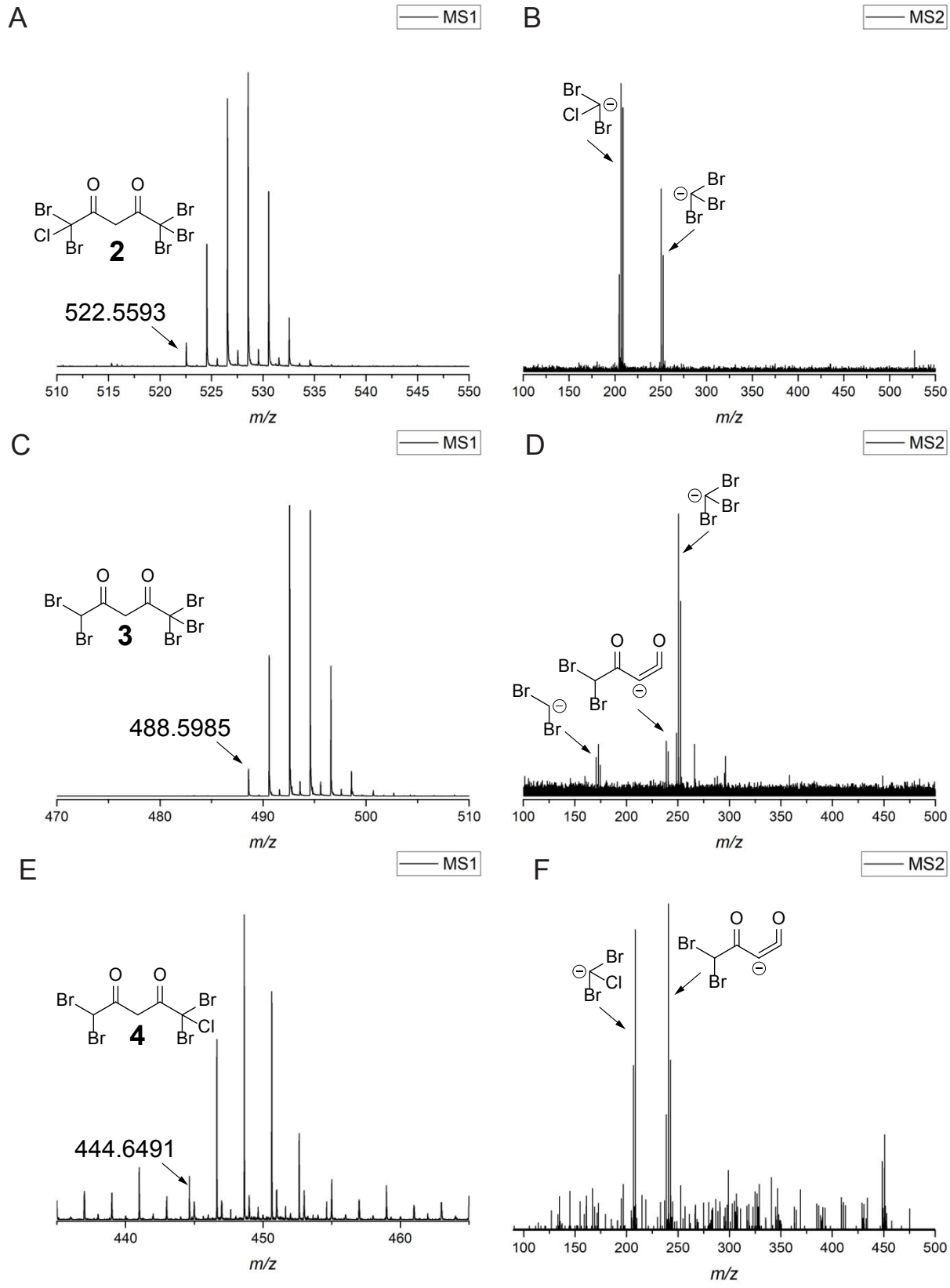


Figure S3: MS^1 and MS^2 spectra for **2**–**4**.

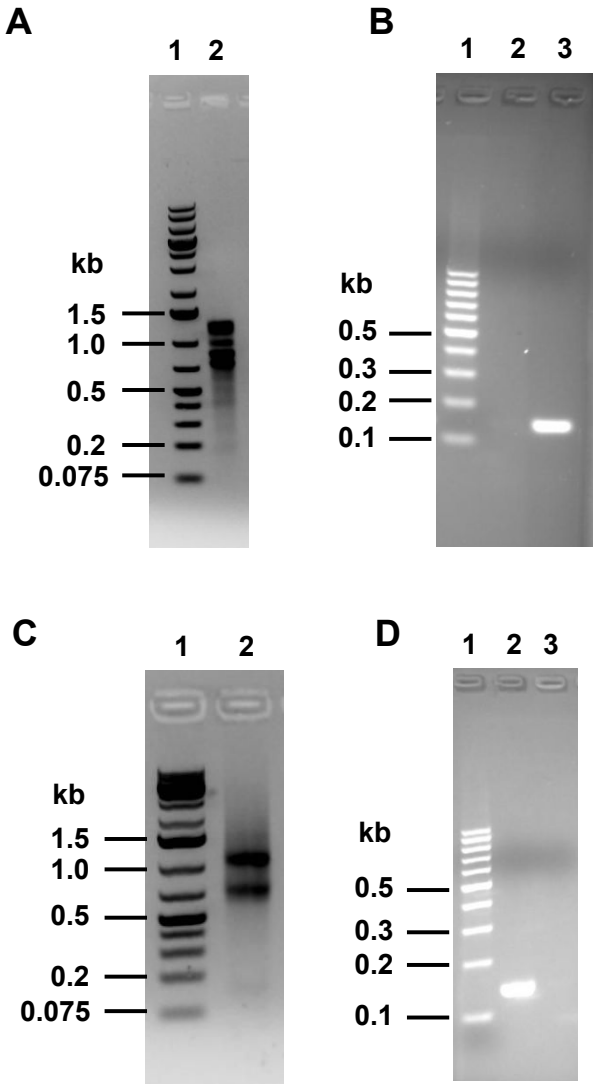


Figure S4: Total RNA and RT-PCR analysis by gel electrophoresis. Agarose gel pictures showing total RNA isolated from *A. taxiformis* collected from Guam (panel **A**, lane 2) and California (panel **C**, lane 2). RT-PCR reaction was done to amplify RuBisCo large subunit gene fragment using isolated RNA samples. Panel **B**; RT-PCR with Guam RNA in absence (lane 2) or presence (lane 3) of RT enzyme. Panel **D**; RT-PCR with California RNA in presence (lane 2) or absence (lane 3) of RT enzyme. Lane 1 in each panel corresponds to DNA ladder.

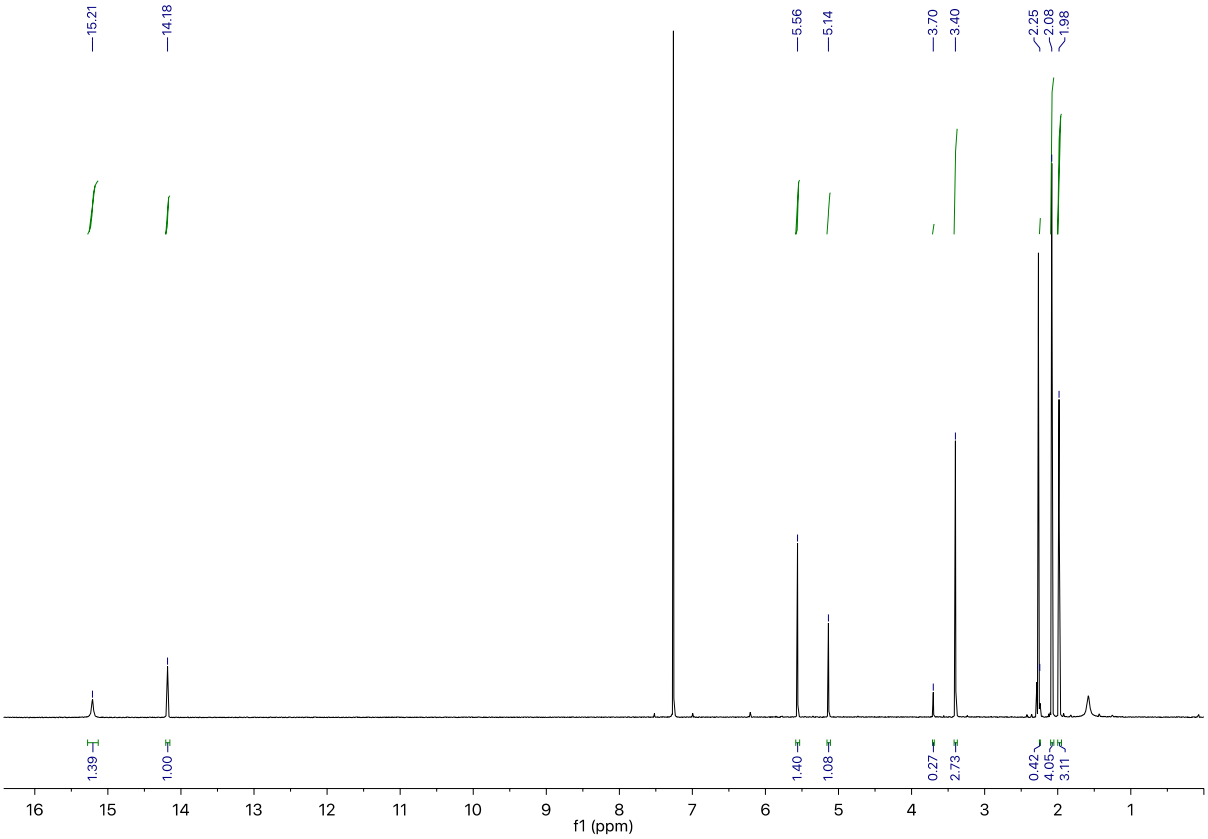


Figure S5: ^{1}H-NMR (400 MHz, $CDCl_3$) spectrum for **6**.

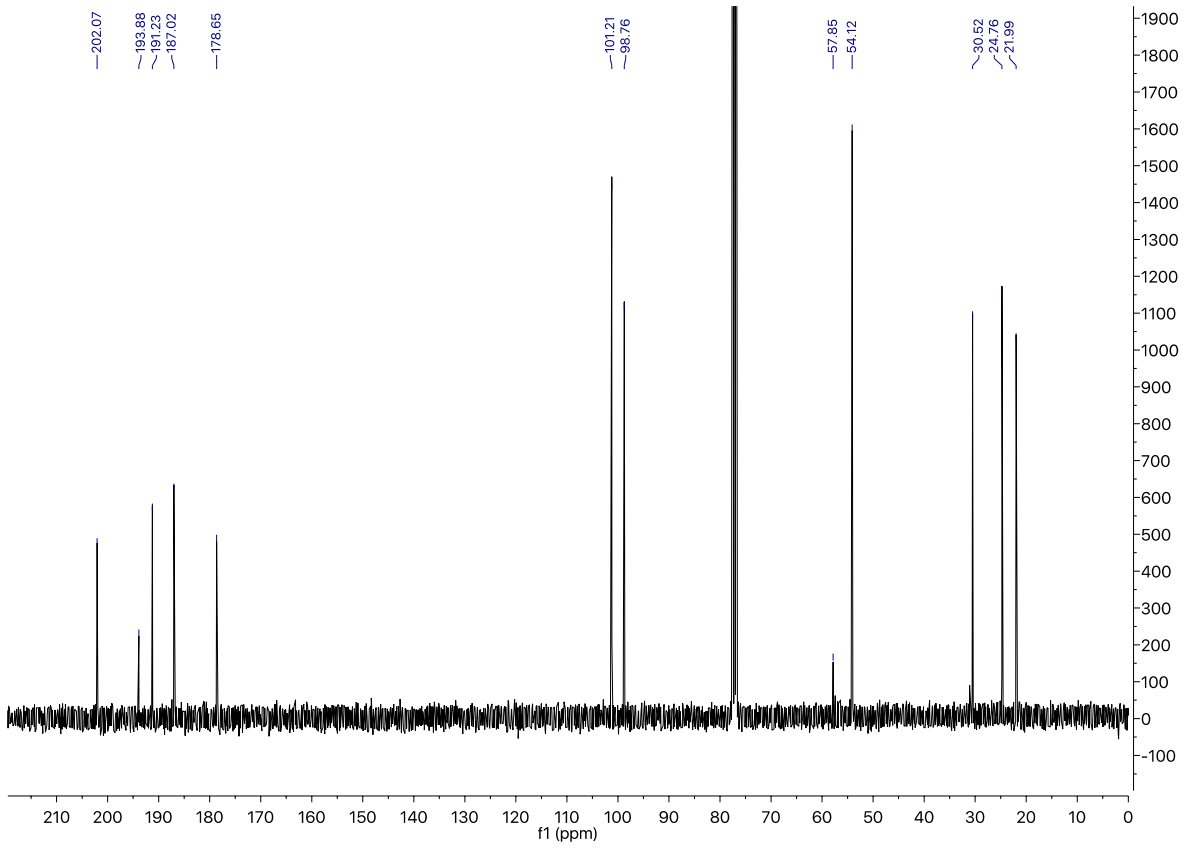


Figure S6: ^{13}C-NMR (101 MHz, $CDCl_3$) spectrum for **6**.

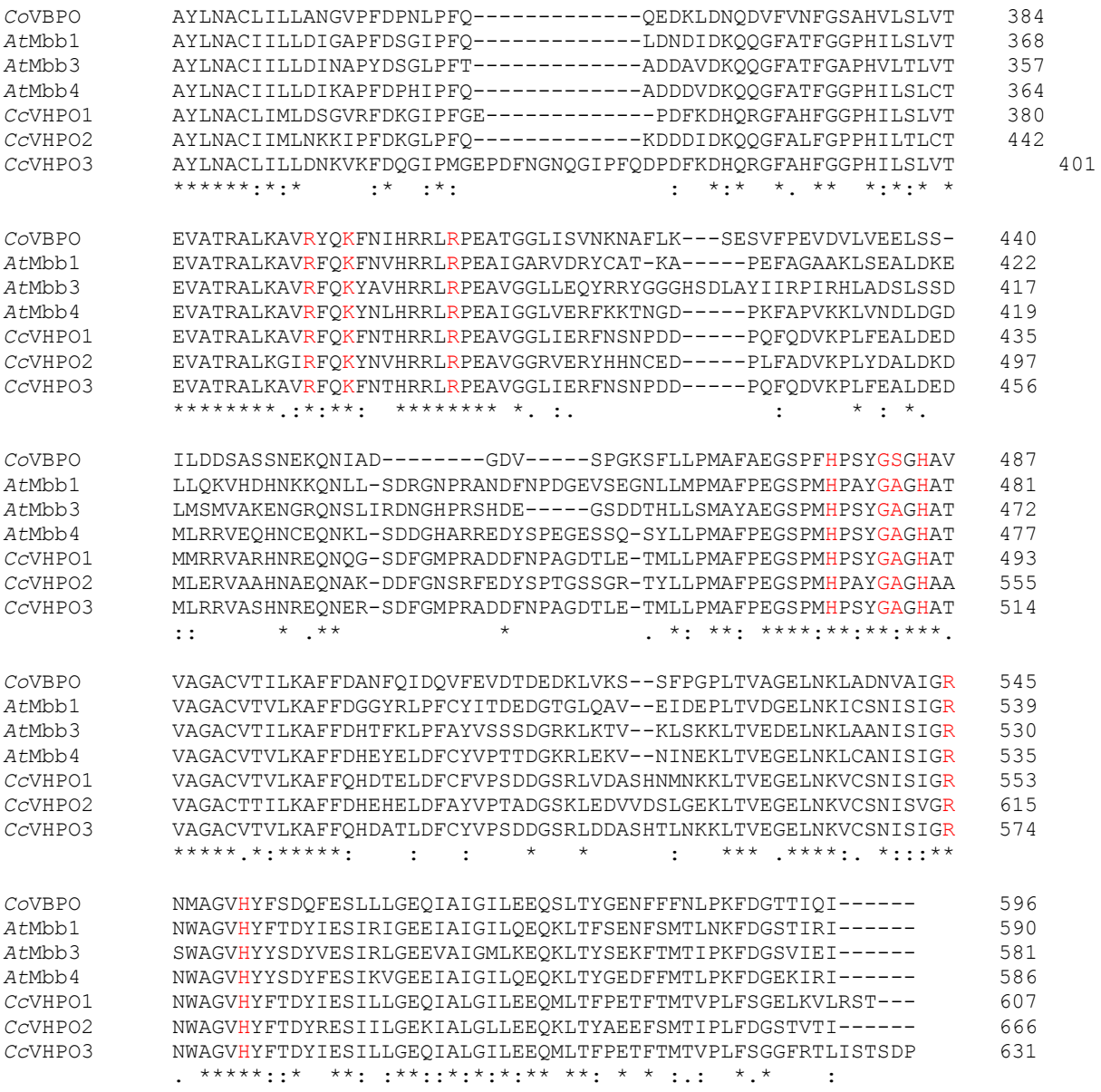


Figure S7: **Amino acid sequence alignment of VHPO and Mbb proteins from marine red macroalgal species.** Only partial amino acid sequences for each protein is shown and alignment was generate using Clustal Omega.[21] All vanadate binding residues except for Ser483 in *Co*VBPO is strictly conserved between *At*Mbb and *Cc*VHPO proteins. Accession numbers and percent identity of proteins with *Co*VBPO are shown in parenthesis: *Corallina officinalis Co*VBPO (PDB 1QHB_A; 100%), *At*Mbb1 (50.44%), *At*Mbb3 (47.53%), *At*Mbb4 (52.24%), *C. crispus Cc*VHPO1 (Genbank XP_005714237.1, 50.53%), *Cc*VHPO2 (GenBank XP_005719435.1, 47.83%), and *Cc*VHPO3 (GenBank XP_005710894.1, 50.26%).

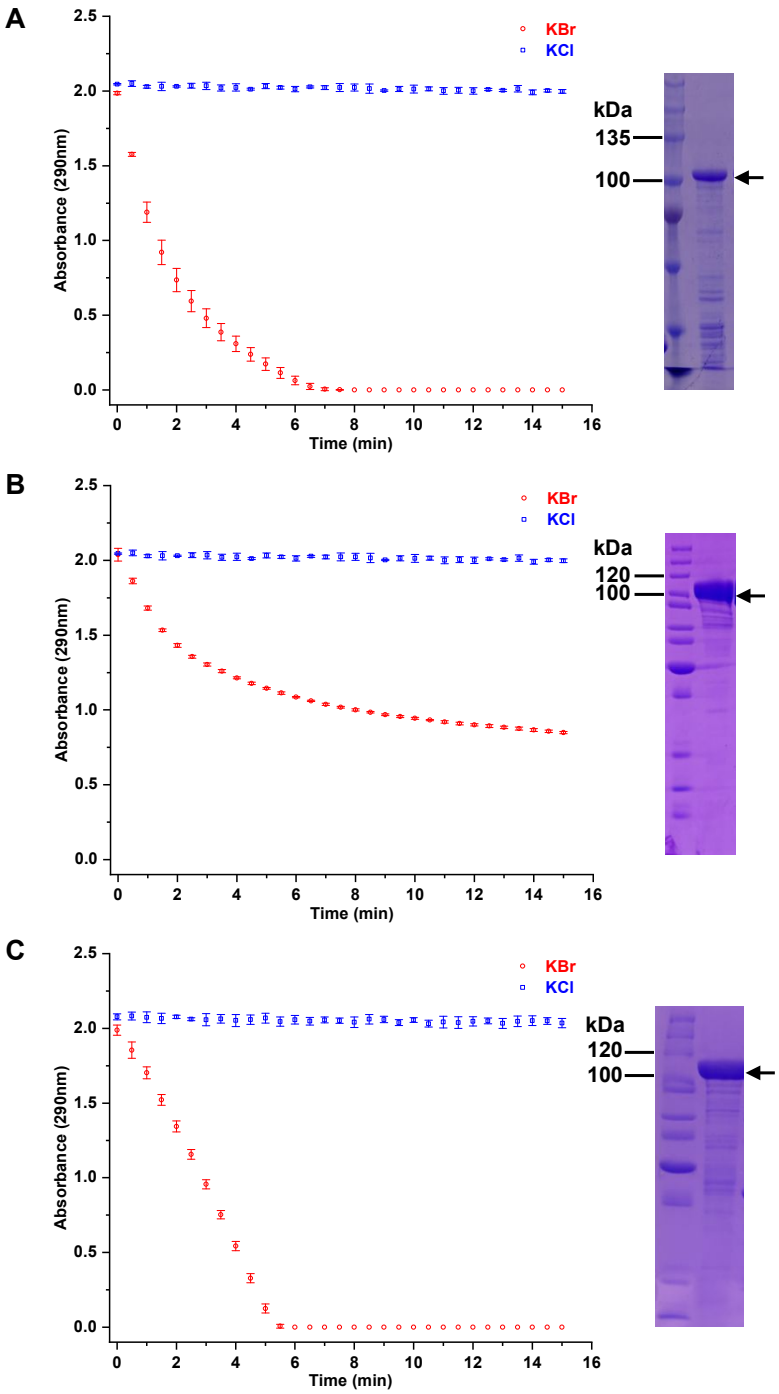


Figure S8: **Monochlorodimedone (MCD) assay with recombinant Mbb VHPOs using KBr and KCl as halide sources.** Change in absorbance of MCD substrate at 290nm over the time when incubated with Mbb enzymes in a MCD assay: (**A**) Mbb1 (250 nM), (**B**) Mbb3 (1 µM), and (**C**) Mbb4 (100 nM). Concentration for each enzyme used here was empirically determined due to difference in the activity of individual enzymes. SDS-PAGE gel on right panel of each graph demonstrates the purity of Mbb proteins: His-MBP-Mbb1 (110.6 kDa), His-MBP-Mbb3 (110.9 kDa) and His-MBP-Mbb4 (111.8 kDa).

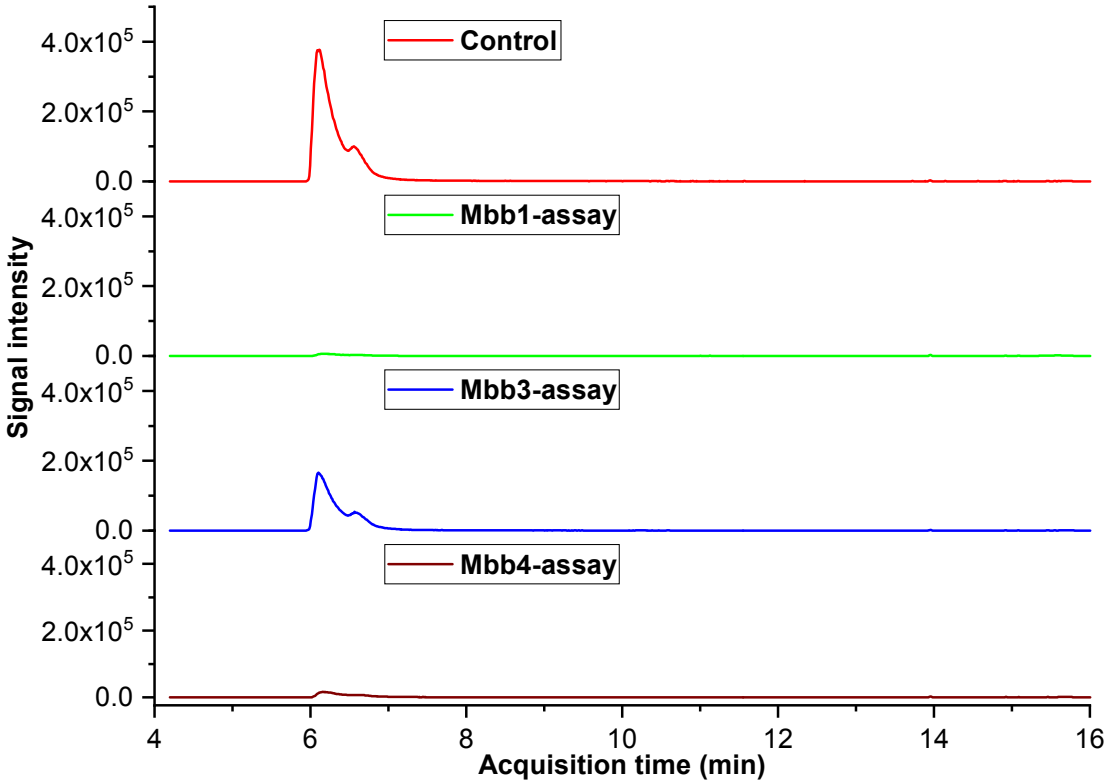


Figure S9: **GC-MS analysis of enzyme assay products showing turnover of substrate 5**. Assays were conducted with recombinant Mbb VHPOs. Extracted ion chromatograms for molecular ion *m/z* 100.05 for **5**.

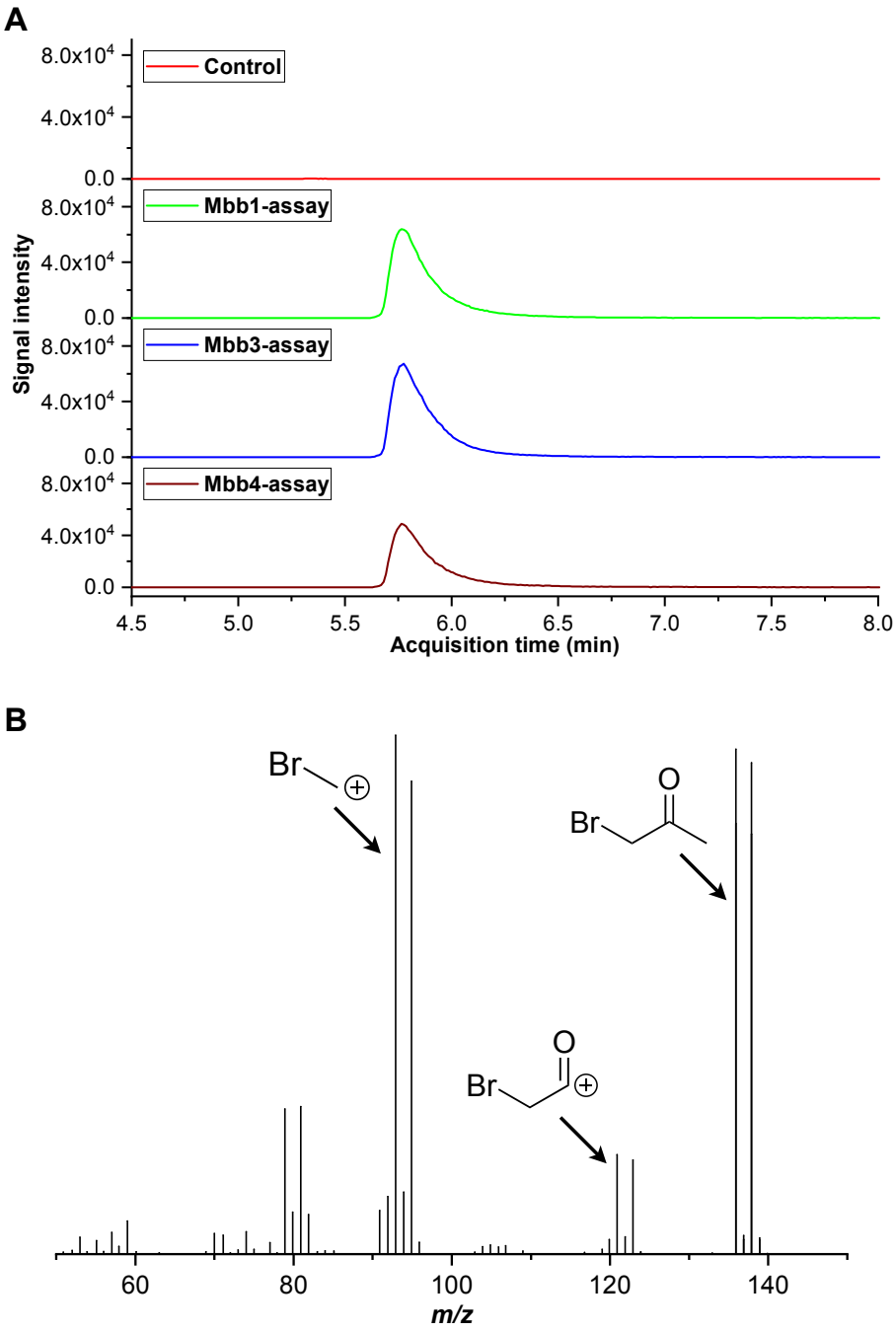


Figure S10: **GC-MS analysis of enzyme assay products.** Assays were conducted with recombinant Mbb VHPOs using **5** as substrate. (**A**) Extracted ion chromatograms for most abundant isotopic ion *m/z* 135.95 for monobromoacetone. (**B**) MS spectrum supports the structural assignment for monobromoacetone by annotation of fragment ions. Arrow corresponds to monoisotopic ion of each molecule.

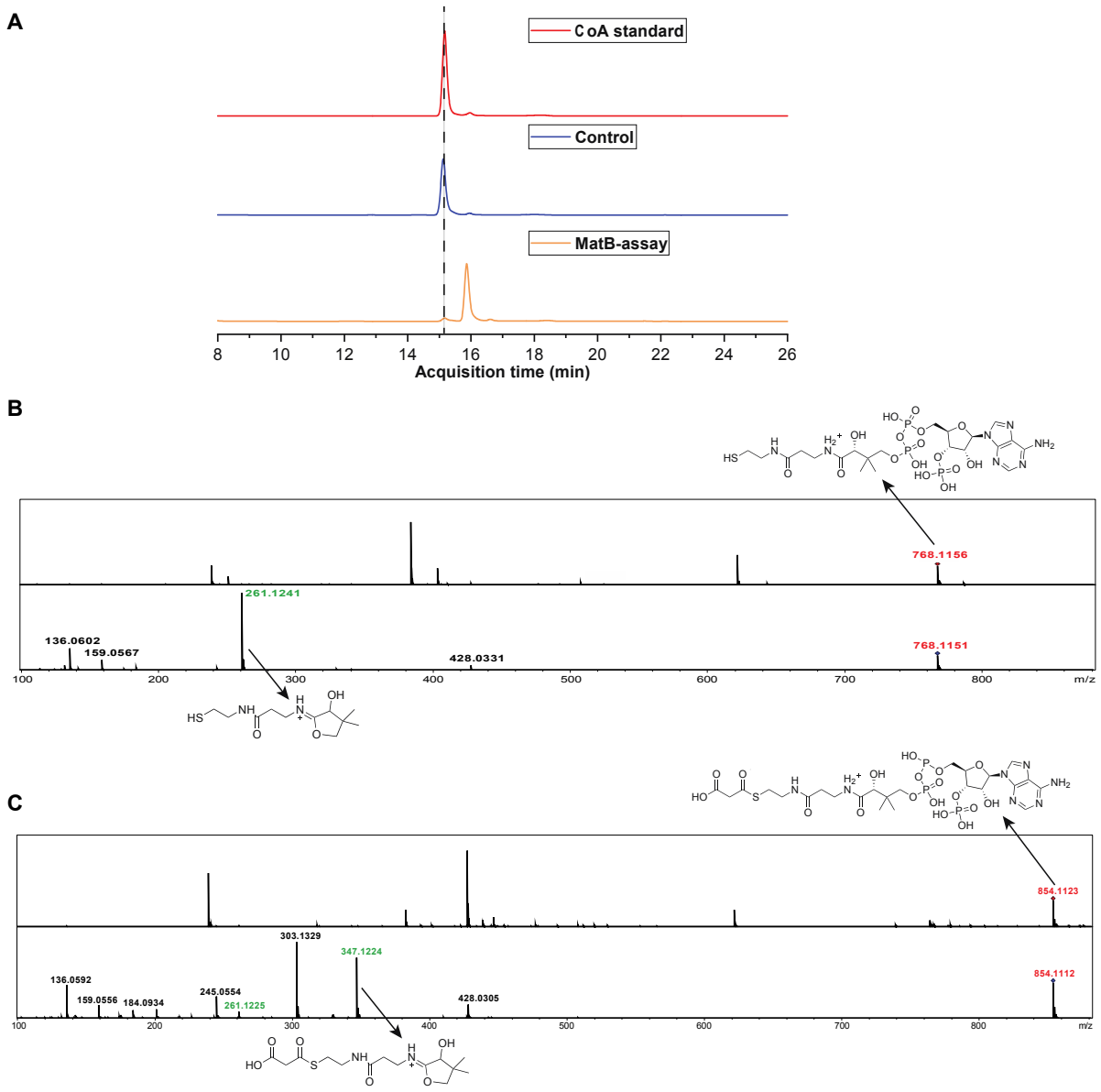


Figure S11: **Enzymatic synthesis of malonyl-CoA.** (**A**) HPLC chromatogram showing synthesis of malonyl-CoA when CoA-SH and malonic acid were used as substrates in an ATP/Mg^{+2} dependent assay with MatB enzyme.[11] Control reaction was done in the absence of MatB. (**B**) MS^1 (top) and MS^2 (bottom panel) spectra for CoA-SH. The characteristic (cyclo)pantetheine MS^2 product ion is observed in the MS^2 spectra.[22] (**C**) MS^1 (top) and MS^2 (bottom panel) spectra for malonyl-CoA. The characteristic malonyl-*S*-(cyclo)pantetheine MS^2 product ion is observed in the MS^2 spectra.

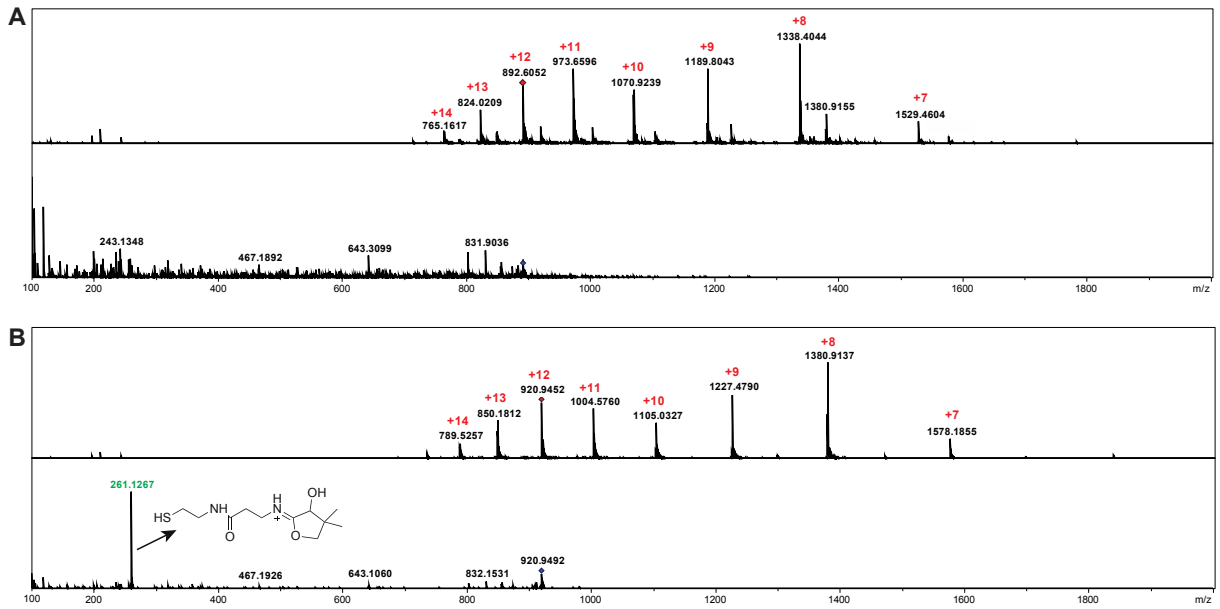


Figure S12: **Mass spectrometric characterization of apo-ACP and holo-ACP.** (**A**) MS^1 (top) and MS^2 (bottom panel) spectra corresponding to apo-ACP. Multiple charge states are observed in MS^1 spectrum. The (cyclo)pantetheine ejection ion is not observed for apo-ACP in the MS^2 spectrum. (**B**) MS^1 (top) and MS^2 (bottom panel) spectra corresponding to holo-ACP. Multiple charge states of peptides are observed in MS^1 spectrum and MS^2 spectrum of holo-ACP shows a characteristic (cyclo)pantetheine ejection ion at *m/z* 261.12.

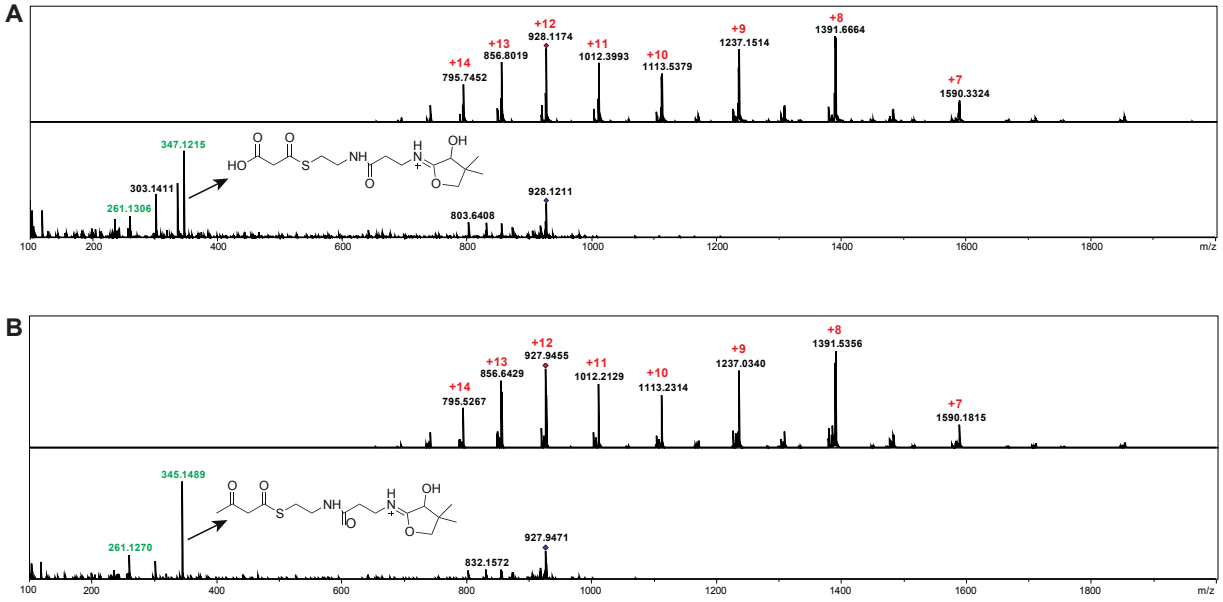


Figure S13: **Mass spectrometric characterization of malonyl-*S*-ACP and acetoacetyl-*S*-ACP (8).** (**A**) MS^1 (top) and MS^2 (bottom panel) spectra corresponding to malonyl-*S*-ACP. Multiple charge states of peptides are observed in MS^1 spectrum and MS^2 spectrum of malonyl-*S*-ACP shows a characteristic malonyl-*S*-(cyclo)pantetheine ejection ion at *m/z* 347.12. (**B**) MS^1 (top) and MS^2 (bottom panel) spectra corresponding to acetoacetyl-*S*-ACP. Multiple charge states of peptides are observed in MS^1 spectrum and MS^2 spectrum of acetoacetyl-*S*-ACP shows a characteristic acetoacetyl-*S*-(cyclo)pantetheine ejection ion at *m/z* 345.14.

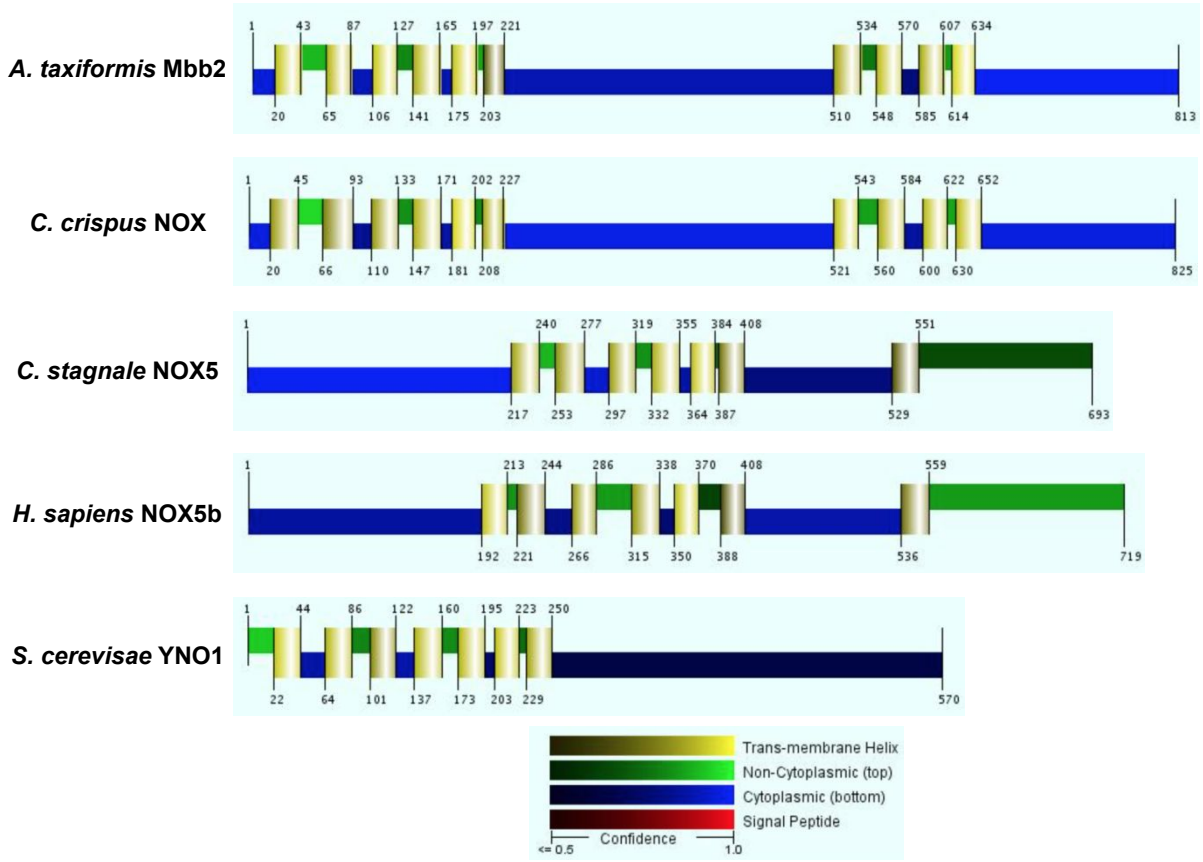
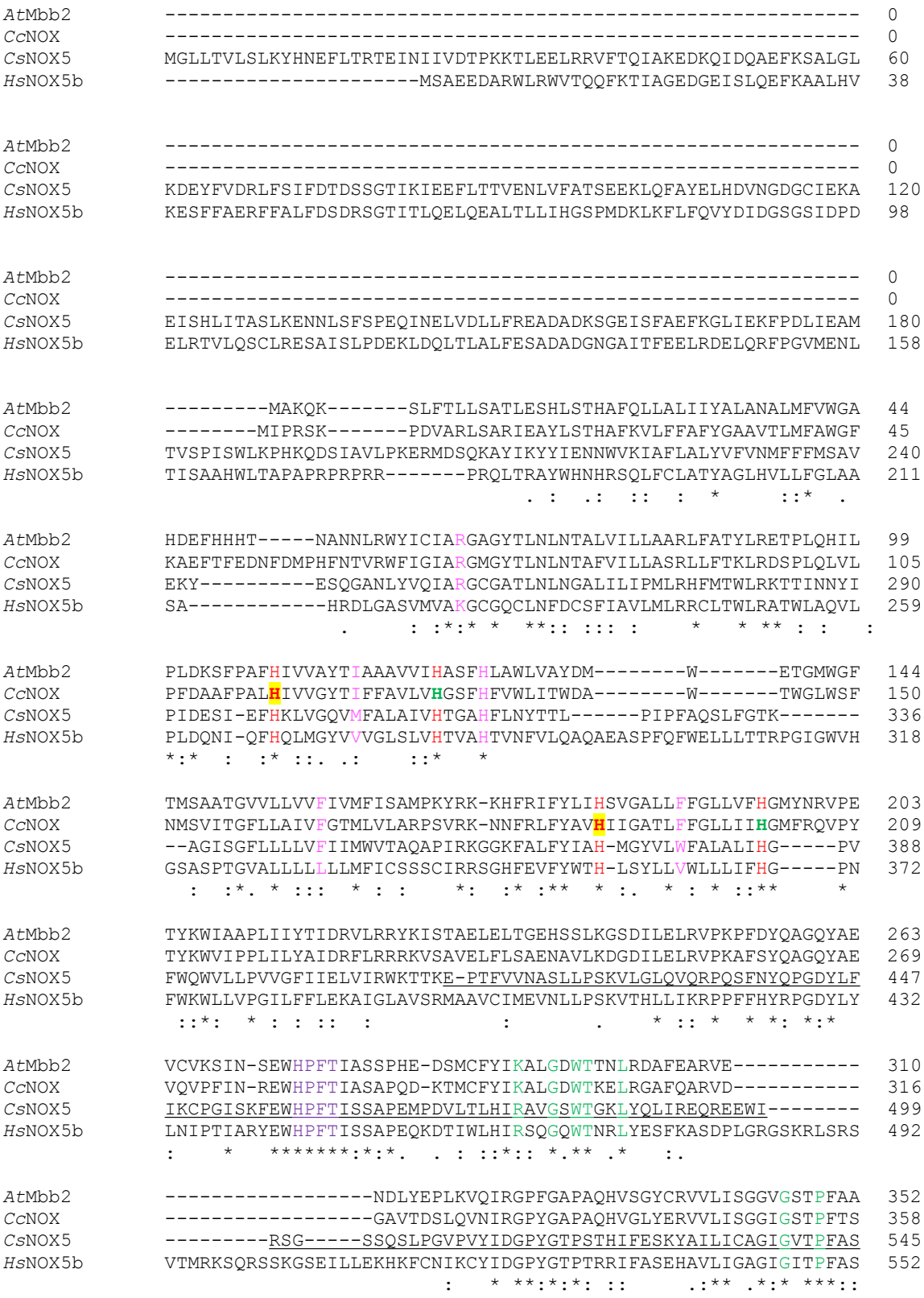


Figure S14: Transmembrane domain prediction for NOX protein sequences using the Phillius transmembrane prediction server.[23] Genbank accession numbers and percent identity of proteins with Mbb2 is shown in parenthesis: Red macroalga *Asparagopsis taxiformis* Mbb2 (100%), red macroalga *Chondrus crispus* NOX (Genbank XP_005719187; 48.81%), cyanobacterium *Cylindrospermum stagnale* (Genbank WP_015206836; 27.27%), human NOX5b (Genbank AAK57193; 30.51%), and yeast *Saccharomyces cerevisiae* YNO1 (Genbank NP_011355; 15.78 %). Human NOX5b is used for comparison as it showed highest identity to Mbb2 when compared to all seven human NOX proteins.[24] Cyanobacterial NOX5 is used for comparison as its crystal structure has been elucidated.[25] Although yeast YNO1 showed low identity to Mbb2 when compared to other NOX proteins, it is shown here to highlight the difference in organization of transmembrane domains between NOX protein sequences. Furthermore, yeast with deletion of *yno1* gene is used for *in vivo* characterization of *Atmbb2* and *CcNOX* genes.


AtMbb2 ---------- 0
CcNOX ---------- 0
CsNOX5 MGLLTVLSLKYHNEFLTRTEINIIVDTPKKTLEELRRVFTQIAKEDKQIDQAEFKSALGL 60
HsNOX5b ----------MSAEEDARWLRWVTQQFKTIAGEDGEISLQEFKAALHV 38
AtMbb2 ---------- 0
CcNOX ---------- 0
CsNOX5 KDEYFVDRLFSIFDTDSSGTIKIEEFLTTVENLVFATSEEKLQFAYELHDVNGDGCIEKA 120
HsNOX5b KESFFAERFFALFDSDRSGTITLQELQEALTLLIHGSPMDKLKFLFQVYDIDGSGSIDPD 98
AtMbb2 ---------- 0
CcNOX ---------- 0
CsNOX5 EISHLITASLKENNLSFSPEQINELVDLLFREADADKSGEISFAEFKGLIEKFPDLIEAM 180
HsNOX5b ELRTVLQSCLRESAISLPDEKLDQLTLALFESADADGNGAITFEELRDELQRFPGVMENL 158
AtMbb2 ---------MAKQK-------SLFTLLSATLESHLSTHAFQLLALIIYALANALMFVWGA 44
CcNOX --------MIPRSK-------PDVARLSARIEAYLSTHAFKVLFFAFYGAAVTLMFAWGF 45
CsNOX5 TVSPISWLKPHKQDSIAVLPKERMDSQKAYIKYYIENNWVKIAFLALYVFVNMFFFMSAV 240
HsNOX5b TISAAHWLTAPAPRPRPRR-------PRQLTRAYWHNHRSQLFCLATYAGLHVLLFGLAA 211
AtMbb2 HDEFHHHT-----NANNLRWYICIARGAGYTLNLNTALVILLAARLFATYLRETPLQHIL 99
CcNOX KAEFTFEDNFDMPHFNTVRWFIGIARGMGYTLNLNTAFVILLASRLLFTKLRDSPLQLVL 105
CsNOX5 EKY----------ESQGANLYVQIARGCGATLNLNGALILIPMLRHFMTWLRKTTINNYI 290
HsNOX5b SA----------HRDLGASVMVAKGCGQCLNFDCSFIAVLMLRRCLTWLRATWLAQVL 259
AtMbb2 PLDKSFPAFHIVVAYTIAAAVVIHASFHLAWLVAYDM--------W-------ETGMWGF 144
CcNOX PFDAAFPALHIVVGYTIFFAVLVHGSFHFVWLITWDA--------W-------TWGLWSF 150
CsNOX5 PIDESI-EFHKLVGQVMFALAIVHTGAHFLNYTTL------PIPFAQSLFGTK------- 336
HsNOX5b PLDQNI-QFHQLMGYVVVGLSLVHTVAHTVNFVLQAQAEASPFQFWELLLTTRPGIGWVH 318
AtMbb2 TMSAATGVVLLVVFIVMFISAMPKYRK-KHFRIFYLIHSVGALLFFGLLVFHGMYNRVPE 203
CcNOX NMSVITGFLLAIVFGTMLVLARPSVRK-NNFRLFYAVHIIGATLFFGLLIIHGMFRQVPY 209
CsNOX5 --AGISGFLLLLVFIIMWVTAQAPIRKGGKFALFYIAH-MGYVLWFALALIHG-----PV 388
HsNOX5b GSASPTGVALLLLLLLMFICSSSCIRRSGHFEVFYWTH-LSYLLVWLLLIFHG-----PN 372
AtMbb2 TYKWIAAPLIIYTIDRVLRRYKISTAELELTGEHSSLKGSDILELRVPKPFDYQAGQYAE 263
CcNOX TYKWVIPPLILYAIDRFLRRRKVSAVELFLSAENAVLKDGDILELRVPKAFSYQAGQYAE 269
CsNOX5 FWQWVLLPVVGFIIELVIRWKTTKE-PTFVVNASLLPSKVLGLQVQRPQSFNYQPGDYLF 447
HsNOX5b FWKWLLVPGILFFLEKAIGLAVSRMAAVCIMEVNLLPSKVTHLLIKRPPFFHYRPGDYLY 432
AtMbb2 VCVKSIN-SEWHPFTIASSPHE-DSMCFYIKALGDWTTNLRDAFEARVE---------- 310
CcNOX VQVPFIN-REWHPFTIASAPQD-KTMCFYIKALGDWTKELRGAFQARVD---------- 316
CsNOX5 IKCPGISKFEWHPFTISSAPEMPDVLTLHIRAVGSWTGKLYQLIREQREEWI-------- 499
HsNOX5b LNIPTIARYEWHPFTISSAPEQKDTIWLHIRSQGQWTNRLYESFKASDPLGRGSKRLSRS 492
AtMbb2 ------------------NDLYEPLKVQIRGPFGAPAQHVSGYCRVVLISGGVGSTPFAA 352
CcNOX ------------------GAVTDSLQVNIRGPYGAPAQHVGLYERVVLISGGIGSTPFTS 358
CsNOX5 ---------RSG-----SSQSLPGVPVYIDGPYGTPSTHIFESKYAILICAGIGVTPFAS 545
HsNOX5b VTMRKSQRSSKGSEILLEKHKFCNIKCYIDGPYGTPTRRIFASEHAVLIGAGIGITPFAS 552

```
AtMbb2        ICKHLHHLNKSENHSKALESHA-SSK-----RMS--QVQWRIRDAISILFDVSLDDSRND  404
CcNOX         ICKDLHHHKVKENATSATGFEPSTST-----LLK--RIESRVSTAISTLYGVDISNAKDI  411
CsNOX5        ILKSILHRNQQNP---------------AKMPLKKVHFYWLNREQKAFEWFVELLSKIEA  590
HsNOX5b       ILQSIMYRHQKRKHTCPSCQHSWIEGVQDNMKLHKVDFIWINRDQRSFEWFVSLLTKLEM  612
              * : : : : ..                    :    .       :  : *.:    :

AtMbb2        EAANQQRRQQLADMLNMSPKGDVHNVSFKRTQSKRSDDFDSVEDSKQMNF----------  454
CcNOX         NQEEEEKRVYLANMLNLTAPGSGSSSGETTE-----LEVEMVDASKQADESSSDSDRSTS  466
CsNOX5        EDT-----------------------------------------------------------  593
HsNOX5b       DQAEEA--------------------------------------------------------  618
              :

AtMbb2        -------SLCRHDSKSSFNLDDCSDAIRYYNRPSFQFFSLSTRHVINLYEYRTRLLAFLH  507
CcNOX         MESYNVKNMLRKEQDEEYILDDIKNSAN--------ARRGNRERLSHLYEGRSKVLEFLH  518
CsNOX5        -----------------------------------------------------------  593
HsNOX5b       -----------------------------------------------------------  618

AtMbb2        TTRFTFALLLTLVARIVILCIVSIFNLGHIGLYN---SHIDATWVIATNSILGLILGVAL  564
CcNOX         TSRVNLLLLFVLIARIFFICISSIIKADYIMINAEPHAIESGLWIVIVDTVLSIIFAIVL  578
CsNOX5        -----------------------------------------------------------  593
HsNOX5b       -----------------------------------------------------------  618

AtMbb2        LTTILLEISFMRMRFFYRVWRCVDFFIFLPITFLCNISNFAS-WNGHAQPKFLIFLDLII  623
CcNOX         PLTIFLELSYMGSRFFRTVGRTLDFFVFLPLTITSASLGIKALVTERTDEQIVLFLHYIV  638
CsNOX5        ----------------------------------------------------NNLFDLNL  601
HsNOX5b       ----------------------------------------------------QYGRFLELHM  628
                                                                   ::.  :

AtMbb2        VLPIMLLLLCHRMYRSVGSRNLLDDTAECRQGCKCNKTIPDVDFVWTTPRDSDDEWLRNE  683
CcNOX         FLPTLFVLLAVRMYRALGKRTLLTDAP---CHCSHRDIVPNVDFVWTVPHENDDEWLRSE  695
CsNOX5        ----------YLTGAQQKSDMKSSTL----------FVAMD-----------------  622
HsNOX5b       ----------YMTSALGKNDMKAIGL----------QMALD-----------------  649
                         :  :  .  :                 :*

AtMbb2        LYPLATGTELRLHRYVTRENMADLEDPERFITTANAGRPQWDEIFAQIAEQTPSHSKIGV  743
CcNOX         LEPLADGTELKLHRYVTRAKEVDMEAGSEFITSSNTGRPEWDEIFGKIAAEAPSNSVVGV  755
CsNOX5        --L----------MH--QETKVDLITGLKSRTKTGRPDWEEIFKDVAKQHA-PDNVEV  665
HsNOX5b       --L----------LA--NKEKKDSITGLQTRTQPGRPDWSKVFQKVAAEK--KGKVQV  691
                           :   :  . :    : : :: ***:*.::* .:* :    . : *

AtMbb2        FFCGPHPMGAAVQKSMRKVEVMSNLRGSYLRKTESAVLVDDLILRDEGEVKLLREYGCNI  803
CcNOX         FFCGPHKMGDSVQSAMRRAEINSNLRGAYLRSTKEKTLMKDLGLPQRGLIKMLMGTGCSV  815
CsNOX5        FFCGPTGLALQ-----------------------------------------LRHLCTKY  684
HsNOX5b       FFCGSPALAKV-----------------------------------------LKGHCEKF  710
              ****   :.                                             *     .

AtMbb2        RFVFREENFS 813
CcNOX         RFVFREENFG 825
CsNOX5        GFGYRKENF- 693
HsNOX5b       GFRFFQENF- 719
               * : :***
```

Figure S15: Amino acid sequence alignment of NOX protein sequences using Clustal Omega.[21] Four NOX protein sequences, *Cylindrospermum stagnale* NOX (*Cs*NOX), *Asparagopsis taxiformis* NOX (*At*Mbb2), *Chondrus crispus* NOX (*Cc*NOX) and Human NOX5b, were used for alignment. Yeast NOX was excluded in alignment study due to its low amino acid identity to other NOX proteins. Residues numbers described

here correspond to *Cs*NOX protein and their structure to function relationship is inferred from study describing the crystal structures of transmembrane domain and dehydrogenase domain of *Cs*NOX protein.[25] Four histidine residues (His313 and His385 ligated to phorphyrin of extracytoplasmic heme, highlighted in green, and His299 and His372 ligated to phorphyrin of cytoplasmic heme, highlighted in red) involved in heme binding in *Cs*NOX5 protein are conserved in *At*Mbb2, *Cc*NOX and *Hs*NOX5b proteins. The two His residues mutated to Ala (data for loss of NOX activity shown in Figure 4B) are highlighted in bold. Similarly, other hydrophobic residues (Met306, Phe348 and Trp378) that intercalates between two heme groups and dioxygen substrate binding residues (His317 and Arg256) in *Cs*NOX5 are also conserved among other NOX proteins, and are highlighted in magenta letters. Residues corresponding to dehydrogenase domain of *Cs*NOX is underlined. The NADPH and FAD binding residues are highlighted in green and purple letters, respectively.

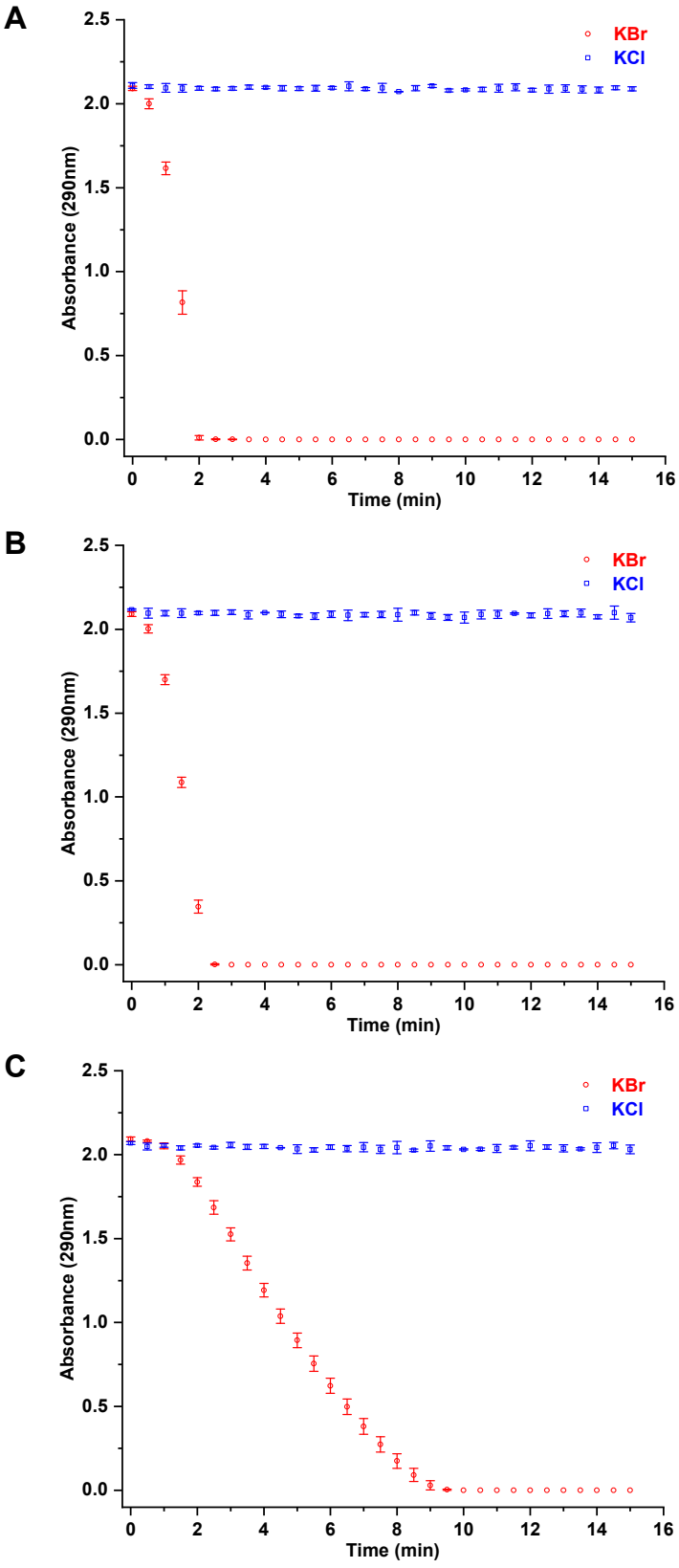


Figure S16: **Monochlorodimedone (MCD) assays with recombinant *Cc*VHPOs**. Change in absorbance of MCD at 290nm was recorded over the time in assays with (**A**) *Cc*VHPO1 (50 nM), (**B**) *Cc*VHPO2 (50 nM), and (**C**) *Cc*VHPO3 (250 nM). Two different halide sources, KBr and KCl, were tested in separate assays, and the enzyme concentrations used here were empirically determined.

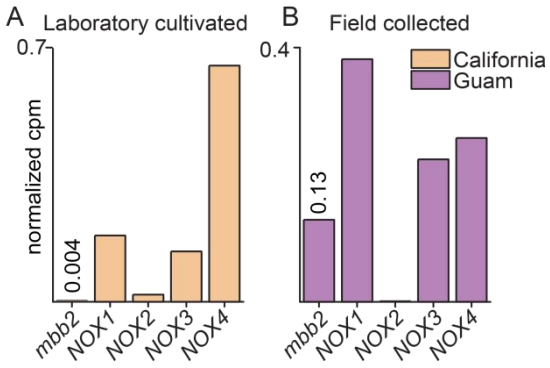
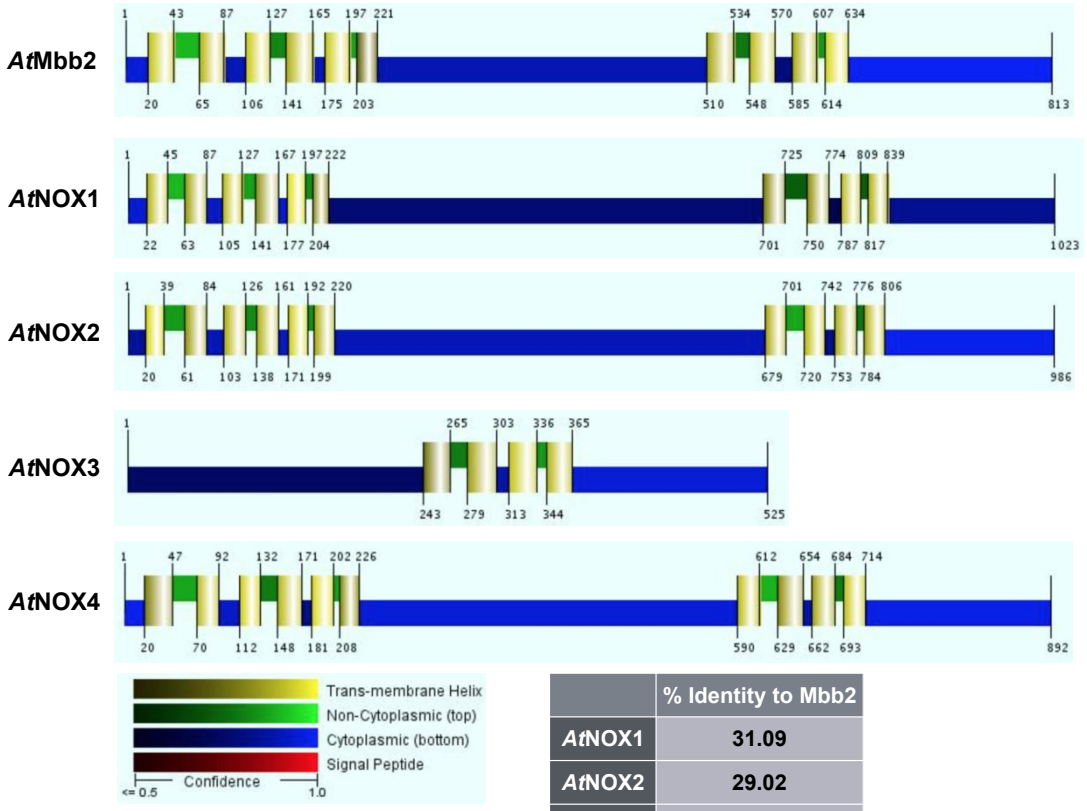


Figure S17. Individually normalized counts per million (cpm) transcript levels for NOX encoding genes in (**A**) laboratory cultivated (California) and (**B**) field collected (Guam) *A. taxiformis* samples.



	% Identity to Mbb2
*At*NOX1	31.09
*At*NOX2	29.02
*At*NOX3	25.32
*At*NOX4	58.37

Figure S18: Transmembrane domain prediction for *Asparagopsis taxiformis* NOX-like protein sequences using the Phillius transmembrane prediction server.[23] Percent identity of *At*NOX proteins to *At*Mbb2 is shown in table, and was determined based on protein sequence alignment generated using Clustal Omega.[21] GenBank accession numbers is shown in parenthesis: *At*NOX1 (GenBank MN911452), *At*NOX2 (GenBank MN911451), *At*Mbb2 (GenBank MN893468), *At*NOX3 (GenBank MN911453) and *At*NOX4 (GenBank MN911454).

SUPPLEMENTARY TABLES

Supplementary Table S1. Shortgun sequencing and assembly results.

	trimmed reads (pair)	contig #	n50 (bp)	longest contig (bp)	Total assembly size (bp)	Sequencing method	Library size	GC%	Read length (bp)
California gDNA	85,676,804	1104938	2261	2019246	1,517,284,223	Illumina NovaSeq	~450	51	151
Guam gDNA	105,547,008	980759	773	744171	765,833,334	Illumina NovaSeq	~450	46	151
California mRNA	44,450,663	20789	3018	20984	35,660,953	Illumina NovaSeq	~450	50	151
Guam mRNA	98,748,611	151422	783	38841	113,504,178	Illumina HiSeq 2000	~350	47	125

Supplementary Table 2: Amino acid and nucleotide sequence identity between Mbb proteins and *mbb* genes from Guam and California samples. Genes *mbb1*, *mbb3* and *mbb4* encodes for VHPO proteins.

	Amino acid similarity (%)	Nucleotide similarity (%)
Mbb1	96.949	96.221
Mbb2	98.770	97.052
Mbb3	96.558	95.246
Mbb4	96.587	95.060

Supplementary Table 3: List of *E. coli* expression constructs used in this study.

Gene	Plasmid	Fusion tag	Source organism	Reference
mbb1	pET-28MBP[26]	N-terminal 6xHis-MBP	*A. taxiformis* (Guam)	This study
mbb3	pET-28MBP	N-terminal 6xHis-MBP	*A. taxiformis* (Guam)	This study
mbb4	pET-28MBP	N-terminal 6xHis-MBP	*A. taxiformis* (Guam)	This study
acp	pET-28b	N-terminal 6xHis	*A. taxiformis* (Guam)	This study
fabH	pET-28b	N-terminal 6xHis	*A. taxiformis* (Guam)	This study
CcVBPO1	pET-28MBP	N-terminal 6xHis-MBP	*C. crispus*	This study
CcVBPO2	pET-28MBP	N-terminal 6xHis-MBP	*C. crispus*	This study
CcVBPO3	pET-28MBP	N-terminal 6xHis-MBP	*C. crispus*	This study
matB[11]	pET-28MBP[26]	N-terminal 6xHis-MBP	*Streptomyces coelicolor*	This study
sfp	pET-24b	C-terminal 6xHis	*Bacillus subtilis*	Agarwal et al. (2014)[27]; Quadri et al. (1998)[13]

Supplementary Table 4: **List of primers used for qRT-PCR**

Gene	Sample	qPCR-forward primer	qPCR-reverse primer
actin	Guam	GCATGACCCAGATCATGTTCGAG	CATAGATAGGGACGGTATGGGTGAC
mbb1	Guam	GAAGTTGGCGATGGATTCATCCAG	CTTCTTTGTACAATTCGCGACCGG
mbb2	Guam	GTGAGTGGCATCCGTTTACAATTG	CTTGAGCGGTTCGTACAGATCATTC
mbb3	Guam	GTCCAACAGATGCTGTTTGCCTC	GTGTACATCGGCGAGACCCTTG
mbb4	Guam	GAGATTACGGAGAGCTATTGGATGGC	CTTCATCGGTCAATGTTGACGGAC
actin	California	GCATGACCCAGATCATGTTCGAG	CATAGATAGGGACGGTATGGGTGAC
mbb1	California	GCAGAGTCCATCAATAACACCCAATG	CGACAAGGAGGAACTGTGACAGATAG
mbb2	California	GTGAGTGGCATCCGTTTACAATTG	GTGGTTCGTACAGATCATTCTCAACTCG
mbb3	California	CTGTGCACGACATGAATCATACCC	CTTTGTGCCGACAAGGAGAAACTG
mbb4	California	GAGATTACGGAGAGCTATTGGATGGC	CTTCATCGGTCAATGTTGACGGAC

All primers are listed from 5’ to 3’.

Supplementary Table 5: List of heme-peroxidase genes and their corresponding transcript level (counts per million) in California and Guam transcriptomes. Amino acid sequences for each heme-peroxidase gene is listed below.

	California cpm	Guam cpm
heme-peroxidase1	19,036,838	96,857,763
heme-peroxidase2	22,132,794	716,554,924
heme-peroxidase3	31,815,891	43,638,708
heme-peroxidase4	296,421	162,957,571
heme-peroxidase5	430,173,245	57,252,885

>Heme-peroxidase1
MSLIIHLLILSLLVVRHPAAAFPLCPARRTFSGACSSLFNPSWGAINTPHRRLIASPIPLPSDLPSPRLISNILCRHNDVKSTRYL
NELTTFFGQFVDHNIVLTPADGPLFPIPVPPSDPLFANFSNGLPFHRSRRVQAFTDHVFSERGPPDVFVPINVVSSVLDLSAVYGP
TISRVRALRGNGGKLKTSHNGRFLPFNHPHLDNEPTTGNIFFVAGDTRSNEHPVLTALHTLFVREHNDLVDELASKFSNYSDDWLF
QTARAINIAQFQKITLEHFYPAMTGRGLPPSSLFSDKYVDASIIDVFATAAFRVGHTMVNDVISPRDKHGPLASIKMSETFFQPGE
HMVHTDIDRFLRGAAWVRAQQVDLAVVDALRNHLFTSVRGEEGVDLVAMNIQRSRDNGLPSYNDIREAVGIPRASCFANISRKVSV
QTALSTAYGSVDRVEAWIGMVAEDHAPGAAMGETLIAVWELQFRKIRDGDRLYFRKPGLFPDVVTDQIARVRDLFYDSDTLRNIVL
RNTDIEDDELPRRMFFV

>Heme-peroxidase2
MISATILSLLLAASLTAALPSPAVHQDALTRVNCEISARFFDGTCTNPSNPAYGQATEAVFSYIPNLSSATFSSRGRPDARLISNA
VCDQDSPVFNDRNLNELLVFFGQFLDHDLVLSEAGDETVPIPVPADDPNTDVSSLPFRRNIRVSVPGEAPGTVRPENLLSQIIDLS
MVYGSDSGRANALRTFEEGRLKVSSDNLLPLNTEGLSNEPSTSENFFVAGDIRSNETPMLTVLHTIWVREHNNIIDELVQSRTFNR
LSDDGLYELARRINIAQFQKVVWEEFFPAIVGRSLPRYRGFNPNVNPTVSNIFATAAFRVGHTMVGDGVQVVQGSTRTVLPVTEMF
FREQSFFRSTGLDSFLLGAVTTRCQEIDNKVVNVLRNGLFSSIPGVSGFDLISLNIQRGRDHNIPSYNEVRELFGIERAQSFADIT
SDTDVQARLASVYDDIDDVDAFIGLISEEHVSGGSLGVTMRAIWTTEFTRLRDGDQFFYLRNGLFSSPLRREIRRVRMLRGRRGIF
RDIVLRNTGLTPSQVPRRPFFV

>Heme-peroxidase3
MTATPSNLLRPIFLVALLAAFATQAFSQAARGNGFGASVLQPIEPGVPEFRERKTDPCRDDYRTYDGSCNNKRMKLWGSAGLPHFS
YLPLRLSTKPKGRNLKSARQISNILSKQTTDIFNSRCLSEFFVFFGQFIDHTFAATPVENTKEFPIKIPADDPIFANFSGGVLPFE
RSRRGRVAGGLADLPINSVTSFLDLSSVYGSDDIRIQKLRTYKNGRMRTTKGNLLPLNTDSLRNAPTNGPMFFAAGDHRANEHPML
TSIHTLFVREHNSLADELRAKFPQWDDERLFQTARKIAIAEFQNIVFEGFFPAMTGRKLWEYRGYKRNVNPTLSDEFVTAAFRVGH
TMVGNEVKRAGPNNTPLSPISMKKMFFQPHTVMSRGVEQFLRGAIITPCQEVDVHVRNSLRDFLFTDIKEEKGFDLVALNLQRGRD
HGLPTYNELRVRFGRPPAFRFSEITRKRNLQSALANAYGNPNKVEAWIGLMCEDHIKGASIGKTLYRIWRREFRRMRAGDRFFYMV
PGLFEKEVRDKIQRVQDLFTDKDIMKGILLRNTKLTSEEIGESVWKADRCLANSN

>Heme-peroxidase4
MKFNLKFSFLSILALLSTSFALPTRLNTNVRQFNCQYQTRMLDGTCTNNIQPELGSTGRAVSSLLRNRSSKRPSSAKTTLPSARFI
SNTVSKQDGDILNDRGLNELVTFFGQFIDHTVVFTPVSETHMDIPIPEDDDIFANFTGGVLEFRRSERVPLAPGSRRERPVNELSA
ALDLASVYGVDQERNEELRTLVDGKLKTSPGNLLPLNTAGLPNAPSTGPNFFLAGDERANENPTLSALHTLFVLEHNNICDELKTN
FPSYTDEQLYETARKINIAEFQKITYEEFYPALTGRRIRRYRGFRATTDVGVLNEFATAGFRVGHTLVGNAIHRTGPGNSPRPDIP
FGEMFFRSAEVLQDNGIDEFIRGATQFEAQEVDLKVHDALRNGLFQGIPGEDAGFDLIALNLQRSRDHNLISYNQLRKRLGGGRAR
NFAQINRNVNIQNLLSTAYDGDVDKVEAWVGMMAERHERGGSFGPTLRKLWDKQFRMFRDGDRFFYLNDIFSTELRNAIPRIDALR
ADSETFRDIIIRNTDITDSELPRRIFFTQ

>heme-peroxidase5
MMSHIRSTALVAFVLLALSSFVVGMPPPQGRKYASRARIPTAQRQISCDIDTRSLDGQCTNPTDPSWSATDTAQFSYIEGHSSNIP
TGENLPSPRLISNTLCMQSEDLFNDREISEFFTFFGQLVDHDLYLTPTSEDEFPIPIGPEDEIFGMFPGDVLEFTRSERVPIVEGE
IAERPVNTVSGALDLSTVYGSSEERNAALREENSCKLLVSEGDLLPLNTMQIPNSPTTDPNMFVAGDTRSNEHPVLATMHTLFVRE
HNYICDQLAVLMPNTTAEEQYENARTINIAEFQDVVYDEFFPALIGRTLAPYEGFDPSVDPTPSNVYGAAGFRIGHTLVGNSLSRA
GPGNEPLEPLTMEEMFFRSTELLTDLGIEEFMRGSMQTTAQEVDIMIHDALRNFLFSEVEEEEGFDLIALNLQRGRDHAIPKFNEL
RVALNMEPLGSFAELTANEEVQAGMEEVYGTIDDVEAWIGMVAEDHMPDSSVGPTMGELWRTEYTRIRDGDMFFYQNSETFPAELA
ELPLITRLDEPGSVLRDIIIRNSEITEEEMNESPFFTS

SUPPLEMENTARY REFERENCES

1 Chekan, J. R. *et al.* (2019) Scalable biosynthesis of the seaweed neurochemical, kainic acid. *Angew. Chem.* 58, 8454-8457.

2 Hadziavdic, K. *et al.* (2014) Characterization of the 18S rRNA gene for designing universal eukaryote specific primers. *PLoS One* 9, e87624.

3 Zuccarello, G. C., Burger, G., West, J. A. & King, R. J. (1999) A mitochondrial marker for red algal intraspecific relationships. *Mol. Ecol.* 8, 1443-1447.

4 Tamura, K. & Nei, M. (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial-DNA in humans and chimpanzees. *Mol. Biol. Evol.* 10, 512-526.

5 Wever, R., Krenn, B. E. & Renirie, R. (2018) Marine vanadium-dependent haloperoxidases, their isolation, characterization, and application. *Method Enzymol.* 605, 141-201.

6 Thapa, H. R. *et al.*(2016) A squalene synthase-like enzyme initiates production of tetraterpenoid hydrocarbons in *Botryococcus braunii* Race L. *Nat. Commun.* 7, 11198.

7 Bolger, A. M., Lohse, M. & Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114-2120.

8 Bushnell, B., Rood, J. & Singer, E. (2017) BBMerge - Accurate paired shotgun read merging via overlap. *PLoS One* 12, e0185056.

9 Bankevich, A. *et al.* (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455-477.

10 Collen, J. *et al.* (2013) Genome structure and metabolic features in the red seaweed *Chondrus crispus* shed light on evolution of the Archaeplastida. *Proc. Natl. Acad. Sci. USA* 110, 5247-5252.

11 Hughes, A. J. & Keatinge-Clay, A. (2011) Enzymatic extender unit generation for in vitro polyketide synthase teactions: structural and functional showcasing of *Streptomyces coelicolor* MatB. *Chem. Biol.* 18, 165-176.

12 Maier, T. *et al.* (2007) PYRAZOLOPYRIMIDONES. PCT/EP2007/055846.

13 Quadri, L. E. N. *et al.* (1998) Characterization of Sfp, a *Bacillus subtilis* phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. *Biochemistry* 37, 1585-1595.

14 Thapa, H. R., Lail, A. J., Garg, N. & Agarwal, V. (2018) Chemoenzymatic synthesis of starting materials and characterization of halogenases requiring acyl carrier protein-tethered substrates. *Method Enzymol.* 604, 333-366.

15 Miller, I. J. *et al.* (2019) Autometa: automated extraction of microbial genomes from individual shotgun metagenomes. *Nucleic Acids Res.* 47, e57.

16 Stanke, M. & Morgenstern, B. (2005) AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* 33 W465-7.

17 Livak, K. J. & Schmittgen, T. D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* **25**, 402-408.

18 Gietz, R. D. & Schiestl, R. H. (1991) Applications of high-efficiency lithium-acetate transformation of Intact yeast-cells using single-stranded nucleic-acids as carrier. *Yeast* **7**, 253-263.

19 Montllor-Albalate, C. *et al.* (2019) Extra-mitochondrial Cu/Zn superoxide dismutase (Sod1) is dispensable for protection against oxidative stress but mediates peroxide signaling in *Saccharomyces cerevisiae*. *Redox Biol.* 21.

20 Kumar, S., Stecher, G. & Tamura, K. (2016) MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870-1874.

21 Sievers, F. & Higgins, D. G. (2018) Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci.* 27, 135-145.

22 Agarwal, V. *et al.* (2015) Chemoenzymatic synthesis of acyl Coenzyme A substrates enables in situ labeling of small molecules and proteins. *Org Lett* 17, 4452-4455.

23 Reynolds, S. M., Kall, L., Riffle, M. E., Bilmes, J. A. & Noble, W. S. (2008) Transmembrane topology and signal peptide prediction using dynamic bayesian networks. *Plos Comput. Biol.* 4, e1000213.

24 Panday, A., Sahoo, M. K., Osorio, D. & Batra, S. (2015) NADPH oxidases: an overview from structure to innate immunity-associated pathologies. *Cell Mol. Immunol.* 12.

25 Magnani, F. *et al.* (2017) Crystal structures and atomic model of NADPH oxidase. *Proc. Natl. Acad. Sci. USA* 114, 6764-6769.

26 Lee, J. *et al.* (2013) Structural and functional insight into an unexpectedly selective N-methyltransferase involved in plantazolicin biosynthesis. *Proc. Natl. Acad. Sci. USA* **110**, 12954-12959.

27 Agarwal, V. *et al.* (2014) Biosynthesis of polybrominated aromatic organic compounds by marine bacteria. *Nat. Chem. Biol.* 10.