pubs.acs.org/acschemicalbiology

Articles

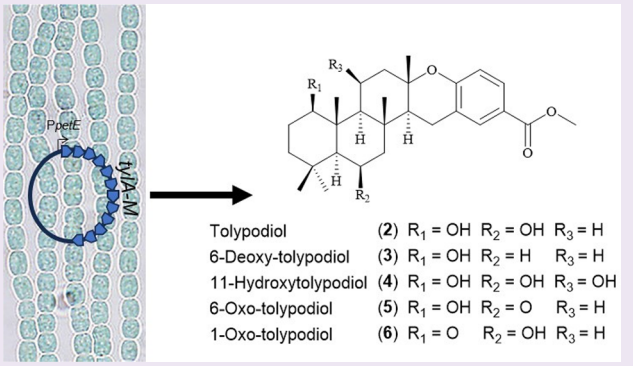
Identification, Heterologous Expression, and Characterization of the Tolypodiol Biosynthetic Gene Cluster through an Integrated Approach

Daniel Back,[#] Timothy J. O'Donnell,[#] Kyle K. Axt, Joshua R. Gurr, Juan M. Vanegas, Philip G. Williams,* and Benjamin Philmus*

Cite This: https://doi.org/10.1021/acschembio.3c00225

Read Online

ABSTRACT: Cyanobacteria are tremendous producers of biologically active natural products, including the potent anti-inflammatory compound tolypodiol. However, linking biosynthetic gene clusters with compound production in cyanobacteria has lagged behind that in other bacterial genera. Tolypodiol is a meroterpenoid originally isolated from the cyanobacterium HT-58-2. Here we describe the identification of the tolypodiol biosynthetic gene cluster through heterologous expression in *Anabaena* and in vitro protein assays of a methyltransferase found in the tolypodiol biosynthetic gene cluster. We have also identified similar biosynthetic gene clusters in cyanobacterial and actinobacterial genomes, suggesting that meroterpenoids with structural similarity to the tolypodiols may be synthesized by other microbes. We also report the identification of two new analogs of tolypodiol that we have identified in both the original and heterologous producer. This work further illustrates the usefulness of *Anabaena* as a heterologous expression host for cyanobacterial compounds and how integrated approaches can help to link natural product compounds with their producing biosynthetic gene clusters.



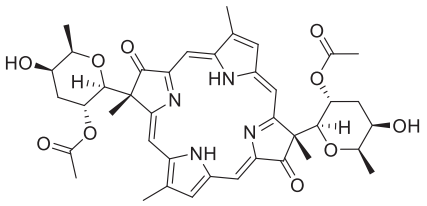
INTRODUCTION

Cyanobacteria, which are Gram-negative photosynthetic bacteria, are rich sources of specialized secondary metabolites (aka natural products) with a wide range of biological activities including antiviral, antibacterial, and anticancer.[1−4] Peptides (linear and cyclic), depsipeptides, polyketides, polyketide-(depsi)peptide hybrids, meroterpenoids, and terpenes have all been isolated from cyanobacteria.[5−10] To date, few cyanobacterial compounds have been assessed in preclinical trials, but two structural classes have entered clinical trials (cryptophycins and dolastatins), and brentuximab vedotin (trade name: Adcetris), an antibody−drug conjugate (ADC) containing monomethyl auristatin E (dolastatin analog), was approved by the FDA in 2011 for the treatment of relapsed Hodgkin lymphoma and systemic anaplastic large cell lymphoma.[11−13] More recently, other ADCs containing the monomethyl auristatin E warhead have been approved for the treatment of other cancers (e.g., Tisotumab vedotin approved in 2021).

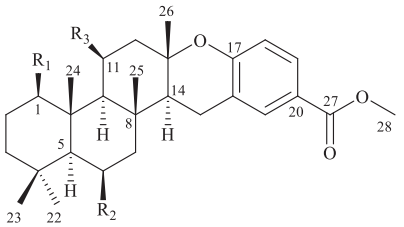
The cyanobacterial strain HT-58-2, originally classified as a *Tolypothrix* sp., but subsequently shown to align more closely with *Brasilonema* sp., is known to produce two unique classes of natural products, the tetrapyrrole macrocyclic tolyporphins (e.g., tolyporphin A, **1**)[14,15] and the meroterpenoid tolypodiol (**2**).[16] Tolypodiol was reported to have potent anti-inflammatory activity in a mouse ear edema assay, with the reported activity similar to hydrocortisone.[16] More recently, we reported the isolation, structure determination, and biological activity of 6-deoxy-tolypodiol (**3**) and 11-hydroxytolypodiol (**4**) from the extract of *Brasilonema* sp. HT-58-2.[17]

Here, we report a biosynthetic gene cluster (BGC) for the production of tolypodiol, with successful heterologous expression of the tolypodiol (*tyl*) BGC in *Anabaena* sp. UTEX 2576, and propose a biosynthetic pathway. We also identified similar BGCs in both cyanobacterial and actinobacterial genomes suggesting that these molecules play an undefined role in the life of the producing organisms. Furthermore, we demonstrate that the methyltransferase, TylH, is responsible for converting a carboxylic acid to a methyl ester, most likely as one of the final steps in the tolypodiol biosynthetic pathway. Lastly, we report two new analogs of tolypodiol, 6-oxo-tolypodiol (**5**) and 1-oxo-

Received: April 19, 2023
Accepted: July 3, 2023



Tolyporphin A (**1**)



Tolypodiol	(**2**)	R_1 = OH	R_2 = OH	R_3 = H
6-Deoxytolypodiol	(**3**)	R_1 = OH	R_2 = H	R_3 = H
11-Hydroxytolypodiol	(**4**)	R_1 = OH	R_2 = OH	R_3 = OH
6-Oxo-tolypodiol	(**5**)	R_1 = OH	R_2 = O	R_3 = H
1-Oxo-tolypodiol	(**6**)	R_1 = O	R_2 = OH	R_3 = H

tolypodiol (**6**), that were isolated from biomass of HT-58-2 after extraction of its lyophilized material as previously reported.[17] The structures of **5** and **6** were determined largely by spectroscopic and spectrometric experiments, such as LC-MS and 1D and 2D NMR.

MATERIALS AND METHODS

General. Optical rotation measurements were taken on a JASCO DIP-370 Digital Polarimeter at the sodium line (589 nm) at 22 °C. Electronic circular dichroism (ECD) and ultraviolet−visible (UV−vis) measurements for **5** and **6** were taken on a Chirascan Circular Dichroism Spectrometer with the sample dissolved in methanol and placed in a 1 cm quartz cuvette with a solvent subtraction for baseline correction. Infrared (IR) spectra were recorded on a CaF_2 salt plate by using a Shimadzu IRAffinity-1 Fourier-Transform Infrared Spectrophotometer. All NMR spectra for initial structural characterization of tolypodiol analogs (**5** and **6**) were acquired on an Agilent Technologies 600 MHz DD2 spectrometer operating at 600 MHz for ^{1}H and 150 MHz for ^{13}C NMR or on a Varian Unity Inova 500 MHz spectrometer operating at 500 MHz for ^{1}H or 125 MHz for ^{13}C. The appropriate residual solvent signal was used as an internal reference. All NMR samples were placed in 3 mm Shigemi tubes for analysis. NMR spectra for tolypodiol and its hydrolysis products were recorded on a Bruker 700 MHz spectrometer with a 5 mm carbon cryoprobe with a *z*-axis gradient. ^{1}H NMR and ^{13}C NMR chemical shifts (δ) are expressed in ppm values. $CDCl_3$ was used as the NMR solvent (δ_H = 7.26 ppm; δ_C = 77 ppm). High-pressure liquid chromatography (HPLC) high-resolution electrospray ionization mass spectrometry (HRESIMS) was conducted on one of two instruments: System 1, Agilent 6545 (Quadrupole-Time of Flight (Q-ToF) mass spectrometer with electrospray ionization (ESI) source was used in the positive mode to acquire high-resolution mass spectrometry (HRMS) data with a C18 Agilent Eclipse Plus (2.1 × 50 mm, 1.8 μm) column (Method A); System 2, an Agilent 6545 Q-ToF equipped with a Jet Stream ion source downstream of an Agilent 1260 Infinity HPLC system consisting of a degasser, quaternary pump, autosampler (held at 12 °C), and column compartment (held at 40 °C). Separation was achieved with a Kinetex C18 column (2.1 × 50 mm, 2.6 μm, Phenomenex, Torrance, CA) (Method B). The Q-ToF machines were operated using MassHunter software and data was processed offline using MassHunter Qualitative software. HPLC separations were performed using a Shimadzu system consisting of LC-20AB binary high-pressure gradient solvent delivery unit, SPD-M20A Photodiode Array Detector, and a CBM-20A System Controller. Enzymatic assays were analyzed by HPLC using an Agilent 1100 HPLC instrument consisting of a vacuum degasser, quaternary pump, autosampler (cooled to 4 °C), column thermostat (maintained at 30 °C), and diode array detector.

All chemicals and solvents were purchased either from Sigma-Aldrich, Alfa Aesar, or Tokyo Chemical Industry Co., Ltd. (TCI). *S*-Adenosyl-L-methionine (SAM) was obtained as *S*-adenosyl-L-methionine disulfate tosylate (MilliporeSigma, Burlington, MA) and used without purification. NdeI, XhoI, T4 DNA ligase, Phusion DNA polymerase, *Escherichia coli* NEB5α, and *Escherichia coli* BL21(DE3)-T1R were purchased from New England Biolabs (Ipswich, MA) and used per the manufacturer's recommendations. Lysogeny broth, Miller (LB) was purchased from MilliporeSigma. GenCatch Plasmid DNA Mini-Prep Kit and EconoSpin All-In-One Mini Spin Columns were purchased from Epoch Life Sciences (Sugar Land, TX). PrimeSTAR GXL DNA polymerase was purchased from Takara Bio (Mountain View, CA) and oligonucleotides were purchased from Integrated DNA technologies with standard desalting (Coralville, IA). The UV absorptions were measured on an Eppendorf BioSpectrometer Kinetic and the cell used was an Eppendorf μCuvette G1.0 (Eppendorf, New Haven, CT). All microbial procedures were performed using good microbial practices and aseptic techniques. Kanamycin (50 μg/mL) and spectinomycin (100 μg/mL) for selection of plasmids in *E. coli*, while spectinomycin and streptomycin (2.5 μg/mL each) were used for the selection of *Anabaena* sp. UTEX 2576.

Strain Source. *Anabaena* sp. UTEX 2576 was acquired from the UTEX Culture Collection of Algae, University of Texas at Austin (Austin, TX, 78705). *Brasilonema* sp. HT-58-2 cultures were revived from a cryostorage maintained by the University of Hawaii Cyanobacterial Culture Collection. *Escherichia coli* strains NEB10β and BL21(DE3) were obtained from New England Biolabs.

Cultivation of *Anabaena* sp. UTEX 2576. *Anabaena* sp. UTEX 2576 was cultivated in a Hoffman incubation chamber (Hoffman Manufacturing, Inc., Corvallis, OR, USA) at 28 °C under 24 h of constant light illumination. The light was supplied by two GE Lighting 49893 F40/PL/AQ Plant and Aquarium Tube bulbs with a light intensity of 25 microeinsteins/m^2/s, and the atmosphere contained 1% CO_2.

Cultivation and Extraction of the Cyanobacterium HT-58-2. Cultures were grown in BG-11(Nitrogen+) media. Over time, they were scaled up to 20 L Pyrex carboys with aeration at 4−5 L/min flow rates and under continuous fluorescent illumination. The BG-11 medium with sodium nitrate as the nitrogen source (BG-11(Nit+)) formulation was composed of the following: $NaNO_3$ (1.5 g/L), $K_2HPO_4 \times 3H_2O$ (0.04 g/L), $MgSO_4 \times 7H_2O$ (0.075 g/L), Na_2CO_3 (0.02 g/L), $CaCl_2 \times 2H_2O$ (0.036 g/L), Na_2EDTA (0.001 g/L), ferric ammonium citrate (0.006 g/L) and citric acid (0.006 g/L) with trace metals H_3BO_3 (2.86 mg/L), $MnCl_2 \times 4H_2O$ (1.81 mg/L), $ZnSO_4 \times 7H_2O$ (0.22 mg/L), $Na_2MoO_4 \times 2H_2O$ (0.39 mg/L), $CuSO_4 \times 5H_2O$ (0.08 mg/L), and $Co(NO_3)_2 \times 6H_2O$ (0.05 mg/L). The phylogenetic characterization and genomic sequencing of this strain has been reported previously.[18] Cell material was harvested by filtration through the equivalent of a cheesecloth after 45 days of growth in 20 L carboys and lyophilized prior to extraction. The freeze-dried HT-58-2 (43.35 g) cell mass was exhaustively extracted in 1:1 CH_2Cl_2:*i*-PrOH to produce the crude extract (2.33 g) after solvent evaporation.

The HT-58-2 strain was originally described in 1992[14] but was later subjected to 16S rRNA analysis revealing that it was more aligned with strains of *Brasilonema*,[18] a genus formally described in 2007,[19] rather than *Tolypothrix* as originally identified from morphological features. The strain has been maintained in cryostorage at University of Hawaii since the original report,[16] periodically removing cryovials to grow the strain.

Isolation of 6-Oxo-tolypodiol (5) and 1-Oxo-tolypodiol (6). A modified Kupchan partition was used to separate crude oil into hexanes, dichloromethane, and aqueous methanolic fractions. The

https://doi.org/10.1021/acschembio.3c00225
ACS Chem. Biol. XXXX, XXX, XXX−XXX

dichloromethane fraction yielded the new compounds after C8 silica gel fractionation using a methanol in water gradient (25%, 50%, 75%, and 100% MeOH). Along with tolypodiol, the 75% MeOH fraction yielded both 6-oxo-tolypodiol (**5**) (0.8 mg) and 1-oxo-tolypodiol (**6**) (0.5 mg) after preparative HPLC (Phenomenex Gemini 5u C18 110A Axia Pac, 250 × 21.20 mm, Torrance, CA) purification using acetonitrile and water with 0.1% formic acid in a linear gradient from 70% to 100% acetonitrile over 30 min followed by 100% acetonitrile for 15 min. The fractions collected at 12−14 min were combined and further purified on a semipreparative scale (Phenomenex Luna PFP(2) 100A, 250 × 10 mm) with an 80%/100% methanol:water (0.1% formic acid) gradient over 30 min (**5**, t_R 18.5 min; **6**, t_R 24.0 min) with purity of 93% and 90%, respectively, as estimated by ^{1}H NMR.

Assembly and Cloning of the *tyl* Biosynthetic Gene Cluster. The meroterpenoid BGC sequence originated from *Brasilonema* sp. HT-58-2 was codon optimized using the Build OptimizatiOn Software Tools (BOOST)[20] and an *Anabaena* sp. PCC 7120 codon frequency table to remove all synthesis constraints. (Refactored sequences are listed in Table S1 as Batch202D_p002.) Overlapping synthetic DNA fragments were obtained from Twist Bioscience and assembled into the pPJAV550 destination vector[21] (Batch202D_p009 and Batch202D_p010) using yeast Transformation-Associated Recombination (TAR) cloning.[22] Vector sequences are listed in Table S2.

All yeast transformations were performed with an auxotrophic strain CEN.PK2 (Euroscarf, Germany) using the YEAST1 yeast transformation kit (MilliporeSigma) and the manufacturer's protocol. Yeast colonies were then collected from the single-dropout selective plates and plasmid DNA were harvested and transformed into *E. coli* DH10B cells. The resulting colonies were sequence verified using the PacBio sequencing platform.

During the refactoring of the BGCs, all internal SrfI restriction sites were removed, and a unique SrfI site was reintroduced at the 5′ ends of each BGCs. These SrfI sites were then used to clone two promoters by using the NEBuilder HiFi DNA Assembly kit (New England Biolabs) to generate a total of 12 BGC constructs (Batch202D_p007 and p018 in Table S1). Again, all construct sequences were verified by PacBio sequencing.

Heterologous Expression of Tolypodiol in *Anabaena* sp. UTEX 2576. The plasmids pMero_HT583 and pMero_HT583_PpetE were introduced into *Anabaena* sp. UTEX 2576 through triparental mating with *E. coli* TOP10 cells containing either pMero_HT583 and pMero_HT583_PpetE and *E. coli* JCM113 according to our previously published protocol.[21,23] The exoconjugates were selected on BG11(N+) media supplemented with spectinomycin and streptomycin (2.5 μg/mL each) for 4 weeks at 28 °C with 24 h illumination in a 1% CO_2 atmosphere. Single colonies were restreaked onto BG11(N+) media supplemented with spectinomycin and streptomycin. The colonies were then cultivated for an additional 5 weeks followed by collection of the cells by scraping them from the plate. The cells were added to a tared 20 mL scintillation vial, lyophilized, and then extracted with 20 mL ethyl acetate overnight with stirring provided by a magnetic stir bar. The cellular debris were removed via filtration through a Buchner funnel, and the extract was concentrated to near dryness *in vacuo* using a 30 °C water bath. The extract was transferred to a 11 mL glass vial and concentrated to dryness in a miVac Quattro Concentrator (ATS Life Sciences Scientific Products, Warminster, PA). The extract was redissolved in HPLC grade methanol at a concentration of 1 mg mL^{-1} and then diluted to a concentration of 0.2 mg mL^{-1} for LCMS analysis.

LCMS Analysis of Tolypodiol. *Method A.* An Agilent 6545 LC-MS Q-Tof with ESI ionization was used in the positive mode to acquire high-resolution mass spectrometry (HRMS) data with a C18 Agilent Eclipse Plus (2.1 mm × 50 mm, 1.8 μm) column. A flow rate of 0.5 mL/min with a water (A)−acetonitrile (B) mobile phase system containing 0.1% (v/v) formic acid was used. The column was pre-equilibrated with 95% A/5% B. Upon injection, the mobile phase composition was maintained for 0.75 min followed by changing the mobile phase to 0% A/100% B over 9.25 min using a linear gradient, and 100% B was maintained for an additional 7 min.

The following parameters were used with positive polarity: Gas Temp, 350 °C; Drying gas, 10 L/min; Nebulizer, 55 psi; Sheath gas temp, 400 °C; Sheath gas flow, 10 L/min; VCap, 3500 V; Nozzle voltage (Expt), 1000 V; Fragmentor, 175 V; Skimmer, 65 V; Oct 1 RF Vpp, 750 V; Mass range, 100−3000 m/z; Acquisition rate, 10 spectra/s; Time, 100 ms/spectrum. The AutoMS/MS settings were as follows: Mass range, 100−3000 m/z; 1 spectra/s; 1000 ms/spectrum; Transients/spectrum, 5889.

Method B. High-resolution mass spectrometry was obtained using an Agilent 1260 HPLC upstream of an Agilent 6545 Q-ToF system and downstream of an Agilent 1260 Infinity HPLC system. Separation was achieved using a Kinetex C18 column (50 × 2.1 mm, 2.6 μm, Phenomenex, Torrance, CA) at a flow rate of 0.4 mL/min and the following gradient. Line A was water with 0.1% (v/v) formic acid, and line B was acetonitrile with 0.1% (v/v) formic acid. The column was pre-equilibrated with 90% A/10% B. Upon injection the mobile phase composition was maintained for 0.75 min followed by changing the mobile phase to 10% A/90% B over 7.5 min using a linear gradient and then switching the mobile phase to 0% A/100%B over 0.1 min. The mobile phase is then maintained at 0%A/100%B for 3.65 min.

The Agilent Q-ToF mass spectrometer was equipped with an Agilent JetSpray source operated by the following parameters: Auto MS/MS mode, Positive polarity; Gas Temp, 325 °C; Drying gas, 7 L/min; Nebulizer, 20 psi; Sheath gas temp, 270 °C; Sheath gas flow, 12 L/min; VCap, 4000 V; Nozzle voltage (Expt), 600 V; Fragmentor, 175 V; Skimmer, 65 V; Oct 1 RF Vpp, 750 V; Mass range, 100−3000 m/z; Acquisition rate, 10 spectra/s; Time, 100 ms/spectrum. The AutoMS/MS settings were as follows: Mass range, 50−3000 m/z; 10 spectra/s; 100 ms/spectrum; Transients/spectrum, 429; Isolation Width, Medium (~4 m/z). The precursor selection parameters were as follows: 5 Max Precursor Per Cycle; Abs. Threshold 1000 counts, Rel. Threshold (%) 0.01%, Excluded after 4 spectra; Released after 0.01 min.

Computational Docking and MD Simulations of TylF. The AlphaFold predicted structure of TylF was downloaded from the AlphaFold Protein Structure Database (Locus BZZ01_00205).[24] Next, the protein's orientation in the membrane was calculated using the PPM 2.0 web server.[25] Ligands were docked into the protein using AutoDock Vina[26] in UCSF Chimera.[27] Ligand binding was restricted to within the inside of the protein near the hypothetical active site, and conformations were picked, which placed the conserved residues in proximity to the predicted site of initialization of cyclization in the ligand.

Molecular dynamics (MD) simulations of *TylF* embedded in a lipid bilayer were carried out with the GROMACS simulation package[28] v2022 using the CHARMM36 force-field (FF) for lipids[29,30] and the CHARMM36m FF for proteins.[31−33] Water was simulated by using CHARMM TIP3 parameters. The glutamate residue in the predicted active site (E51) was modified to be protonated, and ligands with carboxylic acid were specified to be deprotonated. The membranes were composed of POPG, DGDG, MGDG, and SQDG to roughly match the experimental values measured for cyanobacteria.[34] The upper and lower membrane leaflets were symmetrical and contained 4 POPG lipids, 10 DGDG lipids, 18 MGDG lipids, and 10 SQDG lipids. Potassium and chloride ions were added to a concentration of 150 mM KCl. The combined protein−membrane system was set up using the CHARMM-GUI Web server.[35,36] Two different simulations were set up with the 3-geranylgeranyl-4-hydroxybenzoic acid and compound **10** bound in the active site. FF parameters for the two substrates were generated using the CHARMM-GUI interface and CGenFF.[37−39] Systems were simulated for 100 ns after initial energy minimization and equilibration using position restraints. Equations of motion were integrated with a leapfrog algorithm using a 2 fs time step. Van der Waals interactions were computed by using a force-switched Lennard-Jones potential between 1.0 and 1.2 nm. Electrostatic interactions were computed using the particle-mesh Ewald method with a real space cutoff of 1.2 nm and a Fourier grid spacing of 0.12 nm. Temperature was held constant using a velocity-rescaling

https://doi.org/10.1021/acschembio.3c00225
ACS Chem. Biol. XXXX, XXX, XXX−XXX

algorithm[40] with a time constant of 1 ps, and pressure was held constant with a stochastic cell-rescaling algorithm[41] using a time constant of 5 ps. Particle positions were saved in 5 ps intervals for trajectory analysis.

Synthesis of Desmethyltolypodiol (12). In order to hydrolyze the methyl ester of tolypodiol, a protocol using lithium hydroxide in aqueous methanol was adapted.[42] In brief, approximately 1 mg of tolypodiol was dissolved in LCMS grade MeOH (0.7 mL) and H_2O (0.2 mL) in a 5 mL round-bottomed flask to which a few drops of 0.1 N LiOH dissolved in LCMS grade water was added. The reaction was placed in a 40 °C water bath and stirred overnight while connected to a reflux condenser. The reaction was quenched by acidification with a few drops of 0.1 N HCl. This reaction was done twice to secure sufficient material. The reaction mixtures were combined, dried down under a gentle stream of air, reconstituted in a minimal amount of LCMS methanol/water (8:2), and then purified by HPLC (Phenomenex Luna 5u C18(2) 150 × 4.6 mm; A: H_2O, B: MeOH (0.1% formic acid in both); gradient of 70% B (0−2 min), 70−100% B (2−17 min), 100% B (17−22 min); flow rate = 0.6 mL/min) to yield 0.7 mg of the hydrolyzed product and 0.6 mg of recovered tolypodiol. For NMR data, see Figures S1−2 and Table S3. HRESIMS m/z 443.2791 $[M + H]^+$ (calc. for $C_{27}H_{39}O_5^+$, 443.2792, 0.2 ppm error).

Creation of Expression Vector for Overproduction of TylH. The DNA fragment corresponding to TylH (protein accession number ARV57258) was amplified from *Brasilomena* sp. strain HT-58-2 genomic DNA by PCR using PrimeSTAR GXL DNA polymerase and the oligonucleotides TylH_Fwd-NdeI (5′-*ATCTCTAAG*CATATGTCTAAATGGAATGCACAAGATTACC-3′) and TylH_Rev-XhoI (5′-*ATATAGAT*CTCGAGAGACACCGACAGTATGAGTGAAG-3′) with an annealing temperature of 58 °C using an extension time of 2:00 min and 35 cycles. The restriction sites of NdeI and XhoI are indicated by the underlined bases, while the italicized bases indicate random bases added to increase restriction enzyme efficiency. The PCR amplicon was purified using the EconoSpin All-In-One Mini Spin Column followed by digestion with NdeI and XhoI. The restricted fragment was purified using the EconoSpin All-In-One Mini Spin Column followed by ligation into pET28a, which had been previously digested with NdeI and XhoI and purified using the EconoSpin All-In-One Mini Spin Column using T4 DNA ligase. The ligation mixture was then introduced into chemically competent *E. coli* NEB5α cells. Plasmid containing cells were then selected for on LB agar plates supplemented with kanamycin after overnight incubation at 37 °C. Random colonies were selected and the plasmids isolated with the GenCatch Plasmid DNA Mini-Prep Kit. The DNA inserts were sequenced with Sanger sequencing at the Center for Qualitative Life Sciences at Oregon State University to ensure insert sequence fidelity. The resulting sequence verified plasmid was named pET28-TylH-1.

Overexpression and Purification of TylH. Chemically competent *E. coli* strain BL21(DE3) cells were transformed with pET28-TylH-1, and the transformation mixture was plated on LB agar plates supplemented with kanamycin. A single colony of *E. coli* BL21(DE3) cells containing pET28-TylH-1 was picked and grown overnight at 37 °C in LB media containing kanamycin (25 mL) and cultured in a shaker at 200 rpm and 37 °C overnight. A portion of the seed culture (5 mL) was transferred into LB medium (2 × 500 mL) in 2 × 2 L Erlenmeyer flasks containing kanamycin and grown at 37 °C until the OD_{600} reached 0.6 (approximately 3.5 h). Then, the temperature was reduced to 15 °C, and after 1 h of adaptation, isopropyl β-D-1-thiogalactopyranoside (IPTG, 0.1 mM final concentration) was added to each flask. After further growth overnight at 200 rpm (approximately 16 h), the cells were harvested by centrifugation (3200 × *g*, 20 min, 4 °C), the media was decanted, and the cell pellet was stored at −20 °C until use.

The frozen cells from 1 L of culture were thawed on ice and then suspended in 30 mL of lysis buffer [Tris (50 mM, pH 8), NaCl (300 mM), imidazole (10 mM)]. The suspension was sonicated using a Qsonica sonicator (9 mm probe, 30% amplitude, 45 s, five times) on ice, then cell debris were removed by centrifugation at (13,751 × *g*, 4 °C for 30 min). The supernatant was applied to a Ni-NTA resin (750 μL, BioRad, Hercules, CA) by batch binding at 4 °C for 1 h followed by pelleting the resin by centrifugation at 3200 ×*g* for 30 s. The supernatant was discarded, and the resin was washed with lysis buffer (2.5 mL) followed by washing two times with 1 mL of Wash buffer [Tris (50 mM, pH 8), NaCl (300 mM), imidazole (25 mM), β-mercaptoethanol (5 mM)]. The resin was then washed twice with 1.5 mL of elution buffer [Tris-HCl (50 mM, pH 8), NaCl (300 mM), imidazole (250 mM), and β-mercaptoethanol (5 mM)] followed by another wash with elution buffer (0.5 mL). The fractions from the washing buffer contained the highest concentration of TylH (determined by Bradford stain and SDS-PAGE, Figure S3). The protein solution was desalted using an Econo-Pac 10DG Desalting column (BioRad) equilibrated with desalting buffer [Tris (50 mM, pH 8), glycerol (10% w/v), NaCl (50 mM), and β-mercaptoethanol (5 mM)]. The protein concentration of the fractions collected from the washing buffer was determined using an Eppendorf BioSpectrometer at (280 nm, 1 mm cell) using a calculated extinction coefficient of 29,910 cm^{-1} M^{-1}.

In Vitro Enzymatic Activity Assays. Reaction mixtures containing 100 mM Tris buffer (pH 7.5), 5 mM SAM dissolved in DI water (stock: 10 mg mL^{-1}), 1 mM substrate dissolved in MeOH (stock: 10 mg mL^{-1}), 10 μM TylH (stock: 0.7 mg mL^{-1} in desalting buffer), and 1 mM β-mercaptoethanol in a final volume of 100 μL were mixed gently by pipetting and incubated at 28 °C overnight. The protein was removed by the addition of acetonitrile (50 μL) followed by centrifugation (15,871 × *g*) for 3 min. All negative control reactions were performed in the same manner by replacing the omitted reagent with the respective solvent. A portion of the quenched reaction (15 μL) was then analyzed by HPLC as described below.

HPLC Analysis of In Vitro TylH Reaction Mixtures. All enzymatic reactions were analyzed on an Agilent 1100 HPLC instrument consisting of a vacuum degasser, quaternary pump, autosampler (cooled to 4 °C), column thermostat (maintained at 30 °C), and diode array detector using a Luna C18 column (15 cm × 4.6 mm, 5 μm, Phenomenex, Torrance, CA) where line C was water +0.1% (v/v) formic acid and line D was acetonitrile with 0.1% (v/v) formic acid at a flow rate of 1 mL/min. The column was pre-equilibrated with 90% C/10% D and held at this composition for 1 min. The composition was changed to 100% D over the next 20 min using a linear gradient. The composition was then held at 100% D and held for 5 min. The mobile phase was then changed to 90% C/10% D over 8 min, and the composition was held for 8 min prior to the next injection.

RESULTS AND DISCUSSION

Identification of the Tolypodiol Biosynthetic Gene Cluster (BGC). A retrobiosynthetic analysis of tolypodiol predicted it to be biosynthesized from *p*-hydroxybenzoic acid and geranylgeranyl diphosphate as key building blocks. Given the structural similarities between merosterol A and tolypodiol, we first attempted to search the published genome of HT-58-2 by BLAST analysis using MstE as the probe, given its role as a terpene cyclase in the biosynthesis of the merosterol A.[43,44] While a similar protein was identified, BZZ01_14670, it did not appear to be in the proximity of other encoded proteins that were predicted to be involved in the biosynthesis of tolypodiol. We then decided to use chorismate lyase as a BLAST probe as this was expected to be unique to tolypodiol biosynthesis compared to the isoprenoid biosynthetic enzymes that could be found in multiple genetic loci. In bacteria, *p*-hydroxybenzoic acid is known to be synthesized by UbiC;[45] therefore, we used All0938 (UbiC from *Anabaena* sp. PCC 7120) as the BLAST probe sequence and identified two proteins with significant similarity in the genome of HT-58-2, BZZ01_27380 (87% identity/92% similarity) and

https://doi.org/10.1021/acschembio.3c00225
ACS Chem. Biol. XXXX, XXX, XXX−XXX

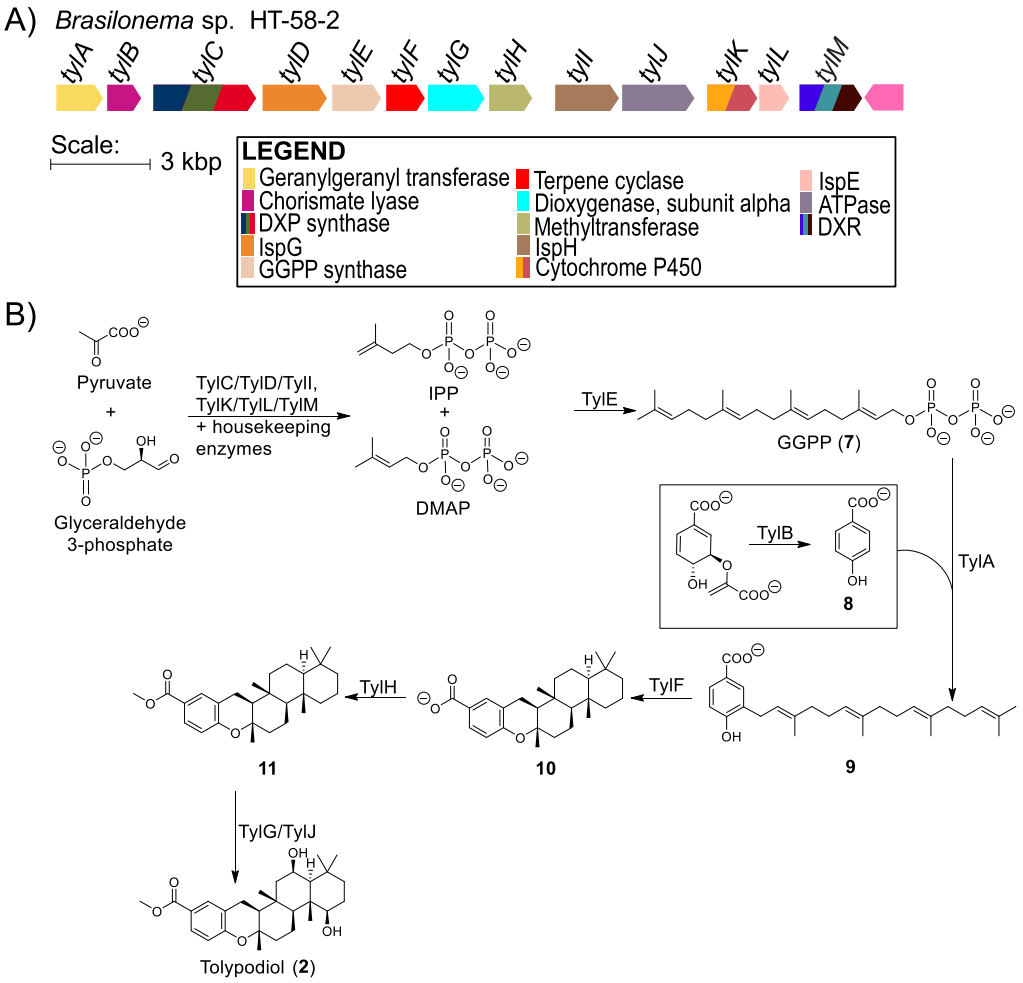


Figure 1. (A) Biosynthetic gene cluster for the biosynthesis of tolypodiols identified in the genome of *Brasilonema* sp. HT-58-2. (B) Predicted biosynthetic pathway for the formation of tolypodiol in HT-58-2.

BZZ01_00185 (73% identity/81% similarity). Examination of the genomic region near BZZ01_00185 showed the presence of encoded proteins that have the predicted activities necessary for the biosynthesis of tolypodiol (Figure 1, Table S4). This same genomic region was recently postulated to be involved in the production of the tolypodiols by the Lindsey group, although the involvement of TylF (BZZ01_00205) in terpene cyclization was not investigated.[18,46] We subsequently renamed BZZ01_00185 as *tylB*, and it is predicted to encode a chorismate lyase, which is responsible for converting chorismate to *p*-hydroxybenzoic acid. TylC, TylD, TylI, TylK, and TylM have similarity to enzymes involved in the formation of the isoprenoid precursors dimethylallyl diphosphate (DMAP) and isopentenyl diphosphate (IPP). TylE belongs to the polyprenyl synthase family and is presumably involved in the biosynthesis of geranylgeranyl diphosphate (GGPP). TylA is predicted to be 4-hydroxybenzoate solanesyltransferase, which is likely responsible for the transfer of the geranylgeranyl unit to the 3-position of 4-hydroxybenzoate. TylG has similarity to aromatic ring hydoxylases and is likely responsible for introduction of one of the hydroxy groups, while TylJ (annotated as a cytochrome P_{450}) is likely responsible for the installation of the second hydroxy group. TylH is predicted to be a SAM dependent methyltransferase, which is envisioned to methylate a carboxylic acid precursor to form the methyl ester seen in the tolypodiols. TylL is annotated as a member of the AAA ATPase protein family[47] and could be a regulator of the tolypodiol BGC expression or responsible for the transport of tolypodiol through membranes. No other transport proteins were identified in the vicinity of the *tyl* BGC. TylF is annotated as a hypothetical protein but is presumed to be the terpene cyclase (see the subsequent sections).

Heterologous Expression of the Tolypodiol (tyl) BGC in Anabaena sp. As genetic disruption of the tolypodiol producing *Brasilonema* sp. HT-58-2 has not been reported and the culture is unialgal and not axenic, we elected to use heterologous expression to prove that *tylA-tylM* was responsible for the biosynthesis of tolypodiol. We and others have previously used *Anabaena* sp. to successfully heterologously express cyanobacterial BGCs[21,23,48,49] and elected to express the *tyl* BGC in this cyanobacterial host. The *tyl* BGC was cloned into pPJAV550 after codon optimization for expression in *Anabaena*. We cloned two versions of the *tyl* BGC, one with the native promoter and one under the control of the P_{petE} promoter (see Methods and Materials for full description) to yield pHT583_Mero-1 and pHT583_Mero_PpetE-1 respectively.[50]

Introduction of pHT583_Mero-1 and pHT583_Mero_PpetE-1 to *Anabaena* sp. UTEX 2576 (=*Anabaena* sp. PCC 7120, herein *Anabaena*) through triparental mating with *Escherichia coli* JCM113 and NEB10*β* containing the plasmid

https://doi.org/10.1021/acschembio.3c00225
ACS Chem. Biol. XXXX, XXX, XXX−XXX

of interest,[23] resulted in exoconjugates after selection on BG-11(Nit+)$_{SpSm}$ agar plates (BG-11 media containing sodium nitrate as the nitrogen source, supplemented with spectinomycin and streptomycin (2.5 μg/mL each)). Colonies were selected, restreaked, and cultivated for an additional 5 weeks on BG-11(Nit+)$_{SpSm}$ agar plates. The cells were then collected, lyophilized, extracted with ethyl acetate, and analyzed by LC-HRMS. We were able to detect tolypodiol in the extract of pHT583_Mero_PpetE-1/*Anabaena* but not pHT583_Mero-1/*Anabaena* (Figure 2 and Figure S12). The produced

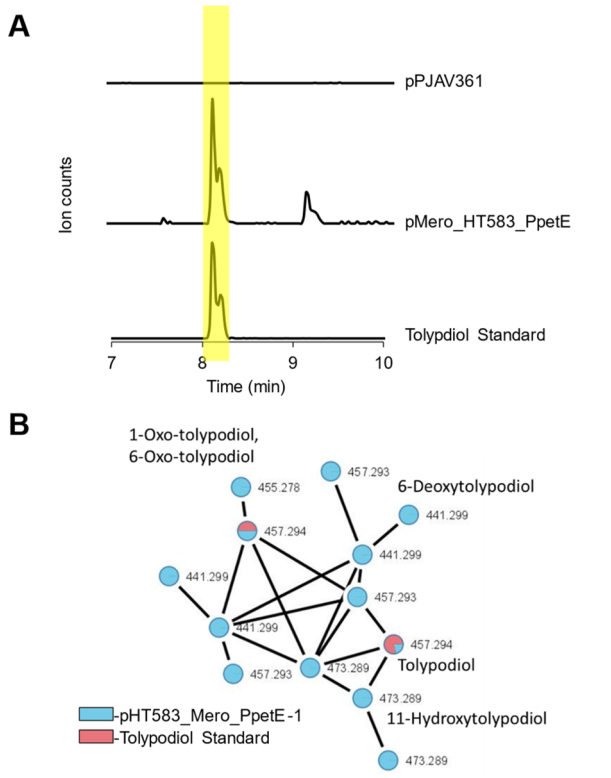


Figure 2. Production of tolypodiol and related compounds by expression of the *tyl* BGC in *Anabaena*. (A) Extracted ion chromatogram for tolypodiol (457.2940−457.2960 *m*/*z*) in *Anabaena* containing pPJAV361 (empty vector), pMero_HT583_PpetE (*tyl* BGC), and purified tolypodiol. (B) Molecular network of metabolites found in *Anabaena* containing pMero_HT583_PpetE. The full GNPS network can be seen in Figure S14. Additional nodes with *m*/*z* values of 457.29 are likely stereoisomers of tolypodiol. These peaks were produced in insufficient quantities for NMR structural determination.

compound had a protonated molecular ion consistent with the molecular formula $C_{28}H_{41}O_5^+$ (M+H obs. 457.2940, calc. 457.2949, 2.0 ppm error) as well as an identical retention time (8.1 min) and MS/MS fragmentation pattern as the tolypodiol standard (Figures S13−14). We used molecular networking[51,52] to identify other analogs present in the extract (Figure 2B, Figure S15) and were able to observe 6-deoxytolypodiol (**3**), 11-hydroxytolypodiol (**4**), 6-oxo-tolypodiol (**5**), 1-oxo-tolypodiol (**6**), and a compound with the same mass and MS/MS fragmentation pattern as tolypodiol with a different chromatographic retention time, likely a conformer or stereoisomer of tolypodiol (Figures S12−14), in the extracts of the *Anabaena* strains containing pHT583_Mero_PpetE-1, thereby linking these metabolites to the tolypodiol BGC. The structural elucidation of compounds **5** and **6** is described in the section New Tolypodiol Structures. The presumed isomer of **2** was produced in quantities that precluded a structural and stereochemical determination.

TylF Is the Founding Member of a New Family of Bacterial Terpene Cyclases with Similarity to Fungal Terpene Cyclases. Using TylF as a query sequence for a BLAST analysis against the Swiss-Prot/UniProt database retrieves terpene cyclase proteins from the BGCs involved in the biosynthesis of a variety of fungal natural products (Table S5). The top hit was DpmpB, which is involved in the biosynthesis of decalin-containing diterpenoid pyrones in *Macrophomina phaseolina* MS6[53] (28% amino acid similarity/50% amino acid identity). Just outside of the top 100 sequences returned from the BLAST query was Trt1 (25% amino acid similarity/47% amino acid identity), which has been characterized as a terpene cyclase in the terretonin biosynthetic pathway. Trt1 was previously shown to cyclize 5-farnesyl-3,5-dimethylorsellinic acid methyl ester, but not the corresponding acid 5-farnesyl-3,5-dimethylorsellinic acid.[54]

TylF does not contain an aspartate-rich motif found in Class I (DDxxD) or Class II (DxDD) terpene cyclases.[55] Multiple sequence alignment of the proteins identified by BLAST identified only one conserved motif Lxx(D/N) xxWExx(Y/F) found in the aligned proteins (Figure S4), and we propose this motif to be involved in cyclization. When Trt1 is included in the multiple sequence alignment (MSA) the conserved motif was identified as (L/I) xx(D/N) xxWExx(Y/F) (Figure S4). Downloading the predicted protein structure of TylF from the AlphaFold Protein Structure Database (entry A0A1Y0RDB4)[24] allowed the prediction of the location of the conserved motif (Figure S5). The L^{44}SANIAWEFLF^{54} sequence was identified in a region that had very high prediction confidence (pLDDT > 90, Figure S6, green residues), and the conserved Trp and Glu residues are located inside a cavity that is presumed to be the active site (see below). The AlphaFill database entry (AF-A0A1Y0RDB4-F1-model_v4, accessed Jan 27, 2023) did not have any ligands, metal ions, or substrates predicted to be bound to the protein. We have also been unable to produce soluble protein in *E. coli*, most likely due to the fact that TylF has seven predicted transmembrane sections (Figure S7), and therefore, we have been unable to directly test the role of TylF as a terpene cyclase and the role of each individual residue in the L^{44}SANIAWEFLF^{54} motif.

While this manuscript was under revision, a paper describing key residues in fungal meroterpenoid cyclase AdrI was published.[56] In AdrI, the conserved glutamate residue (Glu63 in AdrI/Glu51 in TylF) was not invovled in cyclization, but a neighboring aspartate (Asp59 in AdrI) was found to be the catalytic acidic residue. In the same study, IndA7 was also studied, and IndA7 contained a asparagine at position 59. Substitution of Glu63 resulted in a loss of the cyclization activity in IndA7. TylF and all Tyl-like proteins identified in bacterial genomes contain an asparagine like IndA7 (Asn47), which provides support for our assignment of the conserved Glu residue in the TylF motif (L^{44}SANIAWEFLF^{54}) as the catalytic acidic residue that initiates meroterpenoid cyclization.

We turned to MD simulations to interrogate substrate binding to TylF as Trt1 was previously shown to accept methyl 5-farnesyl-3,5-dimethylorsellinate,[54] while MstE was shown to cyclize 3-geranylgeranyl-4,5-dihydroxybenzoate.[43,44] Therefore, the identity of the cyclization substrate was unclear, as TylF is

https://doi.org/10.1021/acschembio.3c00225
ACS Chem. Biol. XXXX, XXX, XXX−XXX

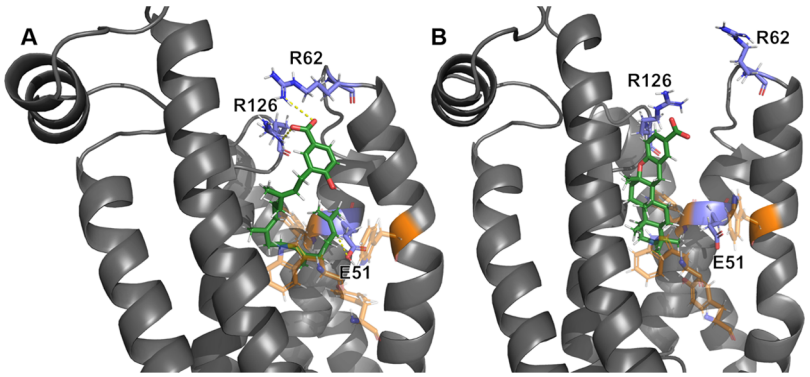


Figure 3. Protein models of TylF bound to substrates. (A) Binding of 3-geranylgeranyl-4-hydroxy benzoic acid (**9**) to TylF. The carboxylate of **9** is coordinated by Arg62 and Arg126 (shown in blue), while the geranylgeranyl side chain is contained in a cavity bound at the bottom by Trp20, Trp50, Trp72 (shown in orange), and Glu51 (modeled as protonated, shown in blue). (B) Cyclized compound **10** bound to TylF. The carboxylate of **10** is coordinated by Arg62/126 (shown in blue), while the cyclized B-E rings are contained within a pocket bordered at the bottom Trp20, Trp50, Trp72 (shown in orange), and Glu51 (modeled as deprotonated, shown in blue).

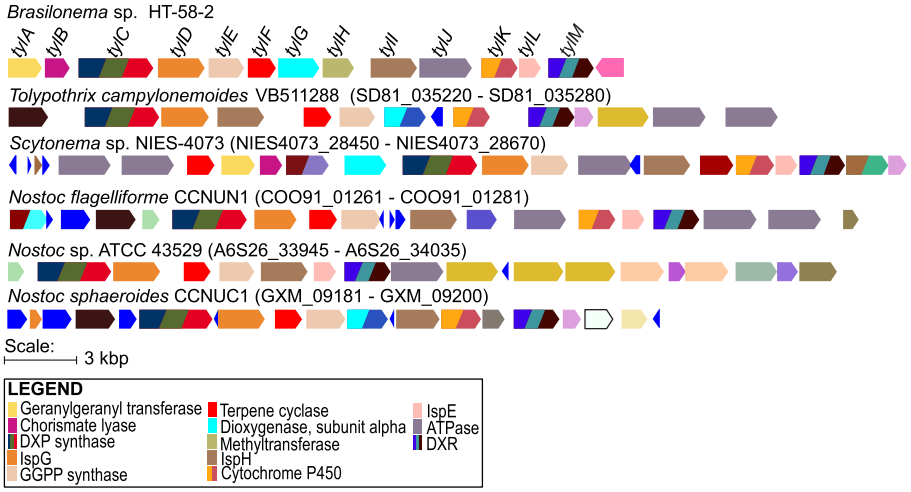


Figure 4. BGCs identified using TylF as the BLAST probe for an EFI-EST search in cyanobacterial genomes.

a cyanobacterial protein (as is MstE), but it is membrane bound (as is Trt1). We examined the predicted binding of 3-geranylgeranyl-4-hydroxybenzoic acid to determine if it is a substrate for cyclization. Docking of 3-geranylgeranyl-4-hydroxybenzoic acid followed by 100 ns of MD simulations resulted in a model with stable rmsd values (Figure S8). The 3-geranylgeranyl-4-hydroxybenzoate fit within a cavity that was bound at the bottom by three Trp residues (Trp20, Trp50, Trp72), and the geranylgeranyl side chain was located in proximity to the L[44] SANIAWEFLF[54] motif (Figure 3A). The top of the cavity was bordered by Arg62 and Arg126. The guanidino groups of the Arg residues fluctuated between 2.95 and 11.87 Å (Arg62) and between 3.17 and 11.11 Å (Arg126) from the carboxylate of **9** during the course of the MD simulation. This suggests that a hydrogen bond can be formed between the substate carboxylate and Arg residues and that **9** is the correct substrate. After cyclization, Arg62 and Arg126 are no longer closely associated with the carboxylic acid, and the top of the protein opens, providing a path for cyclized product **10** to exit the putative active site cavity (Figure 3B).

MstE utilizes a similar but distinct structural motif for binding its substrate, geranylgeranyl-3,4-dihydroxybenzoate.[44] Four aromatic side chains interact with the geranylgeranyl side chain (Phe49, Y100, Trp59, and Trp210), while Asp109 catalyzes the cyclization, similar to the situation envisioned for TylF. In contrast, the carboxylate of geranylgeranyl-3,4-dihydroxybenzoate is anchored through a hydrogen bond with Tyr157, while the 3′-OH group is bound by Glu339. While a structure of Trt1 is not available, the MSA shows that Trt1 contains an Ala and a Leu at the equivalent positions to Arg62 and Arg126 in TylF, which is consistent with binding a noncharged methyl ester.

Meroterpenoid BGCs Encoding TylF-like Proteins Are Widespread in Bacterial Genomes. We next investigated the prevalence of BGCs encoding TylF-like proteins in sequenced bacterial genomes. Using TylF as a BLAST probe for an EFI-EST[57,58] search (e-value cutoff, 5; maximum number of retrieved sequences, 1000), and creation of the clustered network with a cutoff score of 50 resulted in the identification of 30 clusters and 33 singletons (Figure S10). Proteins from cyanobacteria were identified in clusters 1 and 5 (hexagons, Figure S10). The TylF-like proteins in cluster 1 included cyanobacterial proteins from *Nostoc* sp. (cyanobiont of the lichen *Peltigera membranacea*), but the majority of these proteins were from other bacterial genera including *Lysinibacillus* (most prevalent), *Rufibacter*, etc. Examination of the genes upstream and downstream of the TylF-like protein encoding gene revealed that they were typically in small or nonexistent BGCs and are likely not responsible for the biosynthesis of natural products. Cluster 8 (Lavender, Figure

https://doi.org/10.1021/acschembio.3c00225
ACS Chem. Biol. XXXX, XXX, XXX−XXX

S10) contained fungal sequences including the known terpene cyclases NvfL from *Aspergillus novofumigatus* (strain IBT 16806), AtmB from *Byssochlamys spectabilis*, and DpmpB from *Macrophomina phaseolina* (strain MS6). Cluster 5 contained proteins encoded in the genomes of cyanobacteria (*Brasilonema*, *Tolypothrix*, *Scytonema*, and *Nostoc* but not the genome of *Anabaena*) and Actinobacteria (*Streptomyces*, *Actinoplanes*, *Saccharopolyspora*, and *Frankia*). Phylogenetic analysis of the TylF-like proteins encoded in bacterial genomes are distinct from those found in fungal genomes (Figure S31) and have between 28.9% and 100% identity to each other (Table S6).

Closer inspection of these genome neighborhoods revealed that these loci encoded Rieske 2Fe-2S proteins, polyprenyl synthase, terpene precursor biosynthetic proteins, prenyltransferases, and cytochrome P_{450} proteins and transporter proteins (e.g., ABC transporters) (Figure 4 and Figure S11). This suggests that these BGCs have the capacity to biosynthesize meroterpenoids and that TylF-like proteins are likely terpene cyclases. These data suggest that TylF is the first described member of a class/family of previously unrecognized bacterial terpene cyclases that are distributed across bacterial genera. The current products that these BGCs produce are under investigation in our laboratories.

Characterizing the Role of TylH in Tolypodiol Biosynthesis. To better understand tolypodiol biosynthesis and gather further evidence to link the tolypodiol BGC to compound production, we characterized the enzymatic activity of TylH. The encoding gene was expressed in *E. coli*, and TylH was purified using Ni-NTA chromatography. Enzymatic assays were then performed by incubating purified TylH with desmethyltolypodiol (**12**), which was obtained from hydrolysis of tolypodiol (**2**), SAM, and β-mercaptoethanol at RT overnight. Analysis of the enzymatic assay, by HPLC with standards, revealed almost complete conversion of **12** into tolypodiol (**2**) (Figure 5 and Figure S16). This conversion was not observed in any of the control reactions lacking TylH, SAM, or **12** (Figure S16).

TylH Substrate Specificity. The activity of TylH was then probed to determine if TylH has stringent specificity toward desmethyltolypodiol (**12**) or if it can be used for the *O*-methylation of a variety of carboxylic acids. To test this, an enzymatic assay was performed the same way as with **12** but using chroman-6-carboxylic acid (**13**) as the substrate. Chroman-6-carboxylic acid (**13**) was chosen as the substrate, as it is structurally identical to a portion of hydrolyzed tolypodiol containing the carboxylic acid that undergoes *O*-methylation. Analysis of the enzymatic assay by HPLC revealed that the substrate did not undergo conversion (Figure S17) into the corresponding methyl ester. In complementary experiments, TylH did not convert 4-hydroxybenzoic acid (**8**) to the corresponding methyl ester (data not shown), thus showing that TylH has high specificity toward substrates with a pentacyclic structure or that the C, D, or E rings play a role in substrate binding, although this needs to be confirmed with future experiments. The failure to convert chroman-6-carboxylic acid (**13**) and 4-hydroxybenzoic acid (**8**) into the corresponding methyl ester also suggests that TylH mediated *O*-methylation occurs after cyclization of **9**. We examined unpurified and fractionated extracts of *Brasilonema* sp. HT-58-2 but could not observe a signal with an m/z value consistent with **12**. This suggests that the conversion of the carboxylic acid to the methyl ester is not the final step in the biosynthetic pathway suggesting the pathway seen in Figure 1; however due

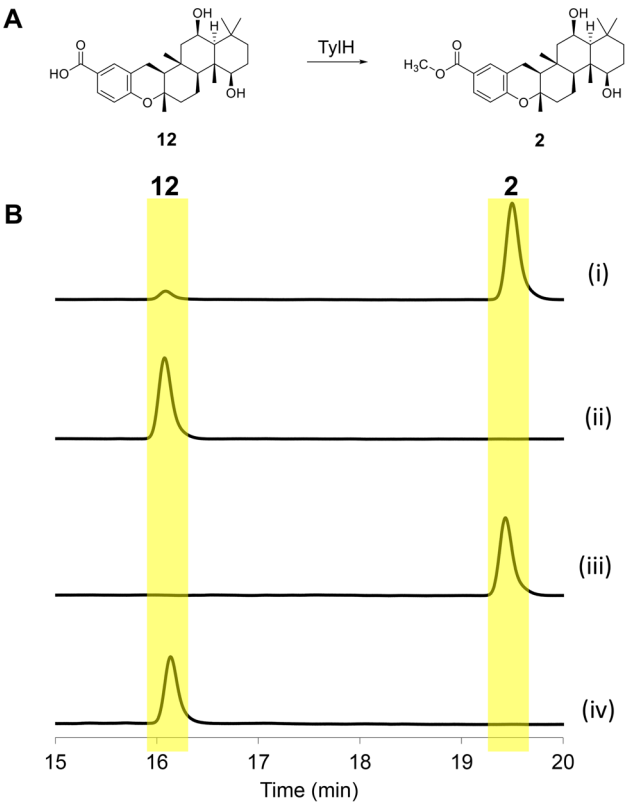


Figure 5. Conversion of desmethyltolypodiol (**12**) to tolypodiol (**2**) by TylH. (A) Reaction diagram of desmethyltopdiol (**12**) with TylH to produce tolypodiol (**2**). (B) HPLC chromatograms observing at 270 nm of (i) full reaction containing **12**, SAM, and TylH; (ii) reaction containing **12** and SAM; (iii) standard of tolypodiol; (iv) standard of desmethyltolypodiol. Full chromatograms are shown in Figure S13.

to the low ionization efficiency of the tolypodiols, this statement is made noting that caveat. In depth in vitro reconstitution experiments are required to determine the exact order of biosynthetic steps.

New Tolypodiol Structures. Simultaneously with the heterologous expression, we reanalyzed the extract of the original tolypodiol producer, *Brasilonema* sp. HT-58-2, and identified additional tolypodiol analogs, which were subsequently identified and characterized to be 6-oxo-tolypodiol (**5**) and 1-oxo-tolypodiol (**6**). The molecular formula of **5** was determined by HRESIMS to be $C_{28}H_{38}O_5$ based on the detection of a protonated molecule at m/z 455.2790 $[M + H]^+$ (calc for $C_{28}H_{39}O_5^+$, 455.2792; −0.4 ppm). It became clear early in the structure elucidation that **5** bore many similarities with tolypodiol, which is a well-characterized compound isolated from HT-58-2.[16] Specifically, the ^{1}H NMR spectrum revealed the 1,2,4-substituted benzene ring, methoxy functional group, and five methyl signals. After negotiating through the remainder of the ^{1}H and ^{13}C NMR spectra (Figures S18−19), it became apparent that the major difference was an oxidation at the C-6 position to a carbonyl (δ_C = 210.7) with regard to tolypodiol's hydroxy (δ_C = 68.8)[16] and a concomitant loss of the methine signal at 4.5 ppm, which would account for the observed mass difference of 2 amu. In addition, the diastereotopic protons at C-7 are further deshielded (δ_H **5**: 2.32, 2.20; tolypodiol: 2.04, 1.23) and appear as doublets, while H-5 exhibits a change in multiplicity from a doublet in

https://doi.org/10.1021/acschembio.3c00225
ACS Chem. Biol. XXXX, XXX, XXX−XXX

tolypodiol to a singlet in **5**. The remainder of the planar structure was confirmed with a set of two-dimensional NMR experiments (Figures S20−22). 2D-NOESY (Figures S23−24) along with relevant coupling constants were utilized to decipher the relative stereochemistry, which was determined to be the same as tolypodiol, except, of course, at the C-6 position, where a carbonyl replaced the secondary alcohol.

HRESIMS was utilized to determine the molecular formula of **6** to be $C_{28}H_{38}O_5$ based on the identification of a protonated molecule at m/z 455.2788 $[M + H]^+$ (calc for $C_{28}H_{39}O_5^+$, 455.2792; −0.9 ppm). As in the case of **5**, the molecular formula was 2 amu less than tolypodiol, and the planar structure of **6** was recognized to be very similar to that of tolypodiol based on the 1D and 2D NMR spectra (Figures S25−29). It became evident that the major difference was an oxidation at the C-1 position from the hydroxy of tolypodiol to a carbonyl based on the lack of a proton resonance around 3.5 ppm that corresponded to the carbinol proton at C-1 of tolypodiol. An HMBC correlation indicated the presence of a carbonyl (δ_C = 214.7) at C-1 and the multiplicity of H-2a/b indicated coupling only to their diastereomeric partner and the adjacent diastereomeric protons, H-3a/b. Additionally, H2a/b were further deshielded (δ_H **5**: 3.15, 2.12) than those of tolypodiol (δ_H **2**: 1.68, 1.65) providing further evidence C-1 was oxidized. Analysis of the J_{H-H} coupling (H-5:2.4 Hz; H-6:2.4, 6.0 Hz) revealed that the relative stereochemistry at C-6 was the same as tolypodiol.

For the determination of absolute configuration, ECD spectra of **5** and **6** were collected and compared to those of tolypodiol (Figure S30). The negative Cotton effect at around 265 nm suggests that **5** and **6** have the same absolute configuration as tolypodiol, which is supported by the fact that the terpene cyclization is predicted to happen early in the biosynthetic pathway.

CONCLUSION

Cyanobacteria are a source of unique chemical structures, bioactivities, and enzymology, in which the tolypodiols provide all three. Here we describe the identification and successful heterologous expression of tolypodiol (**2**) and related compounds **3**, **4**, **5**, and **6**. Located in the biosynthetic gene cluster is a gene encoding a previously undescribed putative terpene cyclase, *tylF*, that evidence and our data suggest belongs to a new family of membrane-bound bacterial terpene cyclases with similarity to fungal meroterpenoid cyclases. We were also able to identify similar terpene cyclases in other bacterial genomes including cyanobacteria and actinobacteria. We identified a SAM-dependent methyltransferase, TylH, encoded in the biosynthetic gene cluster, and characterization by *in vitro* assays demonstrated that it was capable of methylating **12** to form **2**, but was not able to methylate chroman-6-carboxylic acid (**13**) or 4-hydroxybenzoic acid (**8**). While we were unable to detect desmethyltolypodiol (**12**) in either *Anabaena* sp. UTEX 2576 containing the *tyl* BGC or in the native producer *Brasilonema* sp. HT-58-2, the substrate specificity of TylH toward the cyclized substrate suggests that methylation occurs after cyclization of 3-geranylgeranyl-4-hydroxybenzoic acid (**9**). The identification of similar BGCs in other bacterial genomes suggests that other meroterpenoid compounds can be identified or linked to cryptic BGCs. Our successful heterologous expression opens the potential to investigate the role of the tolypodiols, and having a stable production host allows the in-depth investigation of the biosynthetic pathway.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acschembio.3c00225.

1D and 2D NMR spectra as well as MS and MS/MS spectra, SDS-PAGE gel, multiple sequence alignments, AlphaFold models, RMSD trajectory of residue distances, sequences of the BGC synthesized and vectors used, and table of predicted encoded protein functions found in the *tyl* BGC (PDF)

AUTHOR INFORMATION

Corresponding Authors

Philip G. Williams − *Department of Chemistry, University of Hawai'i at Mānoa, Honolulu, Hawaii 96822, United States;* orcid.org/0000-0001-8987-0683; Email: philipwi@hawaii.edu

Benjamin Philmus − *Department of Pharmaceutical Sciences, Oregon State University, Corvallis, Oregon 97331, United States;* orcid.org/0000-0003-2085-0873; Email: benjamin.philmus@oregonstate.edu

Authors

Daniel Back − *Department of Pharmaceutical Sciences, Oregon State University, Corvallis, Oregon 97331, United States*

Timothy J. O'Donnell − *Department of Chemistry, University of Hawai'i at Mānoa, Honolulu, Hawaii 96822, United States*

Kyle K. Axt − *Department of Pharmaceutical Sciences, Oregon State University, Corvallis, Oregon 97331, United States*

Joshua R. Gurr − *Department of Chemistry, University of Hawai'i at Mānoa, Honolulu, Hawaii 96822, United States*

Juan M. Vanegas − *Department of Biochemistry and Biophysics, Oregon State University, Corvallis, Oregon 97331, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acschembio.3c00225

Author Contributions

#D.B. and T.J.O'D. contributed equally.

Funding

We gratefully acknowledge the support from the National Institutes of Health/National Institute of General Medical Sciences (R15GM117541) to BP and National Institutes of Health/National Institute of Aging (5R01AG039468) to PW. D. Back was supported by a T32 Training grant from the National Institutes of Health/National Center for Complementary and Integrative Health (T32AT010131). The Oregon State University NMR Facility was partially funded by the National Institutes of Health, HEI Grant S10OD018518, and by the M. J. Murdock Charitable Trust grant #2014162. The LCMS system was purchased in part with funds from the OSU Research Office and the College of Pharmacy. Work at the Joint Genome Institute was funded by U.S. Department of Energy. The Joint Genome institute is a DOE Office of Science User Facility [contract no. DE-AC02-05CH11231]. Funds for the upgrades of the UH NMR instrumentation were provided by the CRIF program of the National Science Foundation (CH

E9974921), the Elsa Pardee Foundation, and the University of Hawaii at Manoa. The purchase of the Agilent QTOF LC-MS was funded by MRI grant 1532310 from the National Science Foundation.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We would like to thank M. Hamilton, I. Blaby, and J.-F. Cheng of the Joint Genome Institute for their assistance in DNA synthesis and cloning of gene clusters used in this manuscript. We thank W. Yoshida for his assistance with the NMR data acquisition.

■ REFERENCES

(1) Costa, M.; Costa-Rodrigues, J.; Fernandes, M. H.; Barros, P.; Vasconcelos, V.; Martins, R. Marine cyanobacteria compounds with anticancer properties: A review on the implication of apoptosis. *Mar. Drugs* **2012**, *10* (10), 2181−2207.

(2) Kehr, J.-C.; Gatte Picchi, D.; Dittmann, E. Natural product biosyntheses in cyanobacteria: A treasure trove of unique enzymes. *Beilstein J. Org. Chem.* **2011**, *7*, 1622−1635.

(3) Demay, J.; Bernard, C.; Reinhardt, A.; Marie, B. Natural products from cyanobacteria: Focus on beneficial activities. *Mar. Drugs* **2019**, *17* (6), 320.

(4) Pye, C. R.; Bertin, M. J.; Lokey, R. S.; Gerwick, W. H.; Linington, R. G. Retrospective analysis of natural products provides insights for future discovery trends. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (22), 5601−5606.

(5) Moore, R. E. Cyclic peptides and depsipeptides from cyanobacteria: A review. *J. Ind. Microbiol. Biotechnol.* **1996**, *16* (2), 134−143.

(6) Gerwick, W. H.; Tong Tan, L.; Sitachitta, N. Nitrogen-containing metabolites from marine cyanobacteria. In *The Alkaloids: Chemistry and Biology*, Cordell, G. A., Ed.; Academic Press: 2001; Vol. *57*, pp 75−184.

(7) Tan, L. T. Bioactive natural products from marine cyanobacteria for drug discovery. *Phytochemistry* **2007**, *68* (7), 954−979.

(8) Salvador-Reyes, L. A.; Luesch, H. Biological targets and mechanisms of action of natural products from marine cyanobacteria. *Nat. Prod. Rep.* **2015**, *32* (3), 478−503.

(9) Tidgewell, K.; Clark, B. R.; Gerwick, W. H. The natural products chemistry of cyanobacteria. In *Comprehensive Natural Products II*, Liu, H.-W.; Mander, L., Eds.; Elsevier: Oxford, 2010; Vol. *2*, pp 141−188.

(10) Niedermeyer, T. H. J. Anti-infective natural products from cyanobacteria. *Planta Med.* **2015**, *81* (15), 1309−1325.

(11) Singh, R. K.; Tiwari, S. P.; Rai, A. K.; Mohapatra, T. M. Cyanobacteria: an emerging source for drug discovery. *J. Antibio. (Tokyo)* **2011**, *64* (6), 401−412.

(12) Liu, Y. Renaissance of marine natural product drug discovery and development. *J. Mar. Sci. Res. Devel.* **2012**, *2*, No. e106.

(13) Vaklavas, C.; Forero-Torres, A. Safety and efficacy of brentuximab vedotin in patients with Hodgkin lymphoma or systemic anaplastic large cell lymphoma. *Ther. Adv. Hematol.* **2012**, *3* (4), 209−225.

(14) Prinsep, M. R.; Caplan, F. R.; Moore, R. E.; Patterson, G. M. L.; Smith, C. D. Tolyporphin, a novel multidrug resistance reversing agent from the blue-green alga *Tolypothrix nodosa*. *J. Am. Chem. Soc.* **1992**, *114* (1), 385−387.

(15) Prinsep, M. R.; Appleton, T. G.; Hanson, G. R.; Lane, I.; Smith, C. D.; Puddick, J.; Fairlie, D. P. Tolyporphin macrocycles from the cyanobacterium *Tolypothrix nodosa* selectively bind copper and silver and reverse multidrug resistance. *Inorg. Chem.* **2017**, *56* (10), 5577−5585.

(16) Prinsep, M. R.; Thomson, R. A.; West, M. L.; Wylie, B. L. Tolypodiol, an antiinflammatory diterpenoid from the cyanobacterium *Tolypothrix nodosa*. *J. Nat. Prod.* **1996**, *59* (8), 786−788.

(17) Gurr, J. R.; O'Donnell, T. J.; Luo, Y.; Yoshida, W. Y.; Hall, M. L.; Mayer, A. M. S.; Sun, R.; Williams, P. G. 6-deoxy- and 11-hydroxytolypodiols: Meroterpenoids from the cyanobacterium HT-58-2. *J. Nat. Prod.* **2020**, *83* (5), 1691−1695.

(18) Hughes, R. A.; Zhang, Y.; Zhang, R.; Williams, P. G.; Lindsey, J. S.; Miller, E. S. Genome sequence and composition of a tolyporphin-producing cyanobacterium-microbial community. *Appl. Environ. Microbiol.* **2017**, *83* (19), No. e01068-17.

(19) Fiore, M. F.; Sant'Anna, C. L.; Azevedo, M. T. d. P.; Komárek, J.; Kaštovský, J.; Sulek, J.; Lorenzi; Sturion, A. The cyanobacterial genus *Brasolonema*, gen. nov., a molecular and phenotypic evaluation. *J. Phycol.* **2007**, *43* (4), 789−798.

(20) Oberortner, E.; Cheng, J. F.; Hillson, N. J.; Deutsch, S. Streamlining the Design-to-Build Transition with Build-Optimization Software Tools. *ACS Synth. Biol.* **2017**, *6* (3), 485−496.

(21) Videau, P.; Wells, K. N.; Singh, A. J.; Gerwick, W. H.; Philmus, B. Assessment of *Anabaena* sp. strain PCC 7120 as a heterologous expression host for cyanobacterial natural products: Production of lyngbyatoxin A. *ACS Synth. Biol.* **2016**, *5* (9), 978−988.

(22) Larionov, V.; Kouprina, N.; Graves, J.; Chen, X. N.; Korenberg, J. R.; Resnick, M. A. Specific cloning of human DNA as yeast artificial chromosomes by transformation-associated recombination. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93* (1), 491−496.

(23) Videau, P.; Wells, K. N.; Singh, A. J.; Eiting, J.; Proteau, P. J.; Philmus, B. Expanding the natural products heterologous expression repertoire in the model cyanobacterium *Anabaena* sp. strain PCC 7120: Production of pendolmycin and teleocidin B-4. *ACS Synth. Biol.* **2020**, *9* (1), 63−75.

(24) Varadi, M.; et al. AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **2022**, *50* (D1), D439−D444.

(25) Lomize, M. A.; Pogozheva, I. D.; Joo, H.; Mosberg, H. I.; Lomize, A. L. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res.* **2012**, *40* (D1), D370−D376.

(26) Eberhardt, J.; Santos-Martins, D.; Tillack, A. F.; Forli, S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J. Chem. Inf Model* **2021**, *61* (8), 3891−3898.

(27) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera-a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25* (13), 1605−1612.

(28) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1−2*, 19−25.

(29) Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell, A. D., Jr; Pastor, R. W. Update of the CHARMM all-atom additive force field for lipids: Validation on six lipid types. *J. Phys. Chem. B* **2010**, *114* (23), 7830−7843.

(30) Pastor, R. W.; MacKerell, A. D. Jr., Development of the CHARMM force field for lipids. *J. Phys. Chem. Lett.* **2011**, *2* (13), 1526−1532.

(31) MacKerell, A. D., Jr; et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102* (18), 3586−3616.

(32) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D. Jr., Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain χ1 and χ2 dihedral angles. *J. Chem. Theory Comput.* **2012**, *8* (9), 3257−3273.

(33) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D. Jr., CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2017**, *14* (1), 71−73.

https://doi.org/10.1021/acschembio.3c00225
ACS Chem. Biol. XXXX, XXX, XXX−XXX

(34) Zienkiewicz, K. Lipid composition of cyanobacteria (spec). In *Encyclopedia of Lipidomics*, Wenk, M. R., Ed.; Springer Netherlands: Dordrecht, 2017; pp 1−4.
(35) Wu, E. L.; Cheng, X.; Jo, S.; Rui, H.; Song, K. C.; Dávila-Contreras, E. M.; Qi, Y.; Lee, J.; Monje-Galvan, V.; Venable, R. M.; Klauda, J. B.; Im, W. CHARMM-GUI Membrane Builder toward realistic biological membrane simulations. *J. Comput. Chem.* **2014**, *35* (27), 1997−2004.
(36) Lee, J.; et al. CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. *J. Chem. Theory Comput.* **2016**, *12* (1), 405−413.
(37) Kim, S.; Lee, J.; Jo, S.; Brooks, C. L., 3rd; Lee, H. S.; Im, W. CHARMM-GUI ligand reader and modeler for CHARMM force field generation of small molecules. *J. Comput. Chem.* **2017**, *38* (21), 1879−1886.
(38) Vanommeslaeghe, K.; et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2010**, *31* (4), 671−690.
(39) Yu, W.; He, X.; Vanommeslaeghe, K.; MacKerell, A. D. Jr., Extension of the CHARMM general force field to sulfonyl-containing compounds and its utility in biomolecular simulations. *J. Comput. Chem.* **2012**, *33* (31), 2451−2468.
(40) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126* (1), 014101.
(41) Bernetti, M.; Bussi, G. Pressure control using stochastic cell rescaling. *J. Chem. Phys.* **2020**, *153* (11), 114107.
(42) Dayal, B.; Salen, G.; Toome, B.; Tint, G. S.; Shefer, S.; Padia, J. Lithium hydroxide/aqueous methanol: mild reagent for the hydrolysis of bile acid methyl esters. *Steroids* **1990**, *55* (5), 233−237.
(43) Moosmann, P.; Ueoka, R.; Grauso, L.; Mangoni, A.; Morinaka, B. I.; Gugger, M.; Piel, J. Cyanobacterial *ent*-sterol-like natural products from a deviated ubiquinone pathway. *Angew. Chem., Int. Ed.* **2017**, *56* (18), 4987−4990.
(44) Moosmann, P.; Ecker, F.; Leopold-Messer, S.; Cahn, J. K. B.; Dieterich, C. L.; Groll, M.; Piel, J. A monodomain class II terpene cyclase assembles complex isoprenoid scaffolds. *Nat. Chem.* **2020**, *12* (10), 968−972.
(45) Meganathan, R. Ubiquinone biosynthesis in microorganisms. *FEMS Microbiol. Lett.* **2001**, *203* (2), 131−139.
(46) Jin, X.; Miller, E. S.; Lindsey, J. S. Natural product gene clusters in the filamentous Nostocales cyanobacterium HT-58-2. *Life* **2021**, *11* (4), 356.
(47) Snider, J.; Thibault, G.; Houry, W. A. The AAA+ superfamily of functionally diverse proteins. *Genome Biol.* **2008**, *9* (4), 216.
(48) Taton, A.; et al. Heterologous expression of cryptomaldamide in a cyanobacterial host. *ACS Synth. Biol.* **2020**, *9* (12), 3364−3376.
(49) Taton, A.; et al. Heterologous expression in *Anabaena* of the columbamide pathway from the cyanobacterium *Moorena bouillonii* and production of new analogs. *ACS Chem. Biol.* **2022**, *17* (7), 1910−1923.
(50) Buikema, W. J.; Haselkorn, R. Expression of the *Anabaena hetR* gene from a copper-regulated promoter leads to heterocyst differentiation under repressing conditions. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (5), 2729−2734.
(51) Aron, A. T.; et al. Reproducible molecular networking of untargeted mass spectrometry data using GNPS. *Nat. Protoc.* **2020**, *15* (6), 1954−1991.
(52) Wang, M.; et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **2016**, *34* (8), 828−837.
(53) Tsukada, K.; et al. Synthetic biology based construction of biological activity-related library of fungal decalin-containing diterpenoid pyrones. *Nat. Commun.* **2020**, *11* (1), 1830.
(54) Matsuda, Y.; Awakawa, T.; Itoh, T.; Wakimoto, T.; Kushiro, T.; Fujii, I.; Ebizuka, Y.; Abe, I. Terretonin biosynthesis requires methylation as essential step for cyclization. *ChemBioChem.* **2012**, *13* (12), 1738−1741.
(55) Christianson, D. W. Structural and chemical biology of terpenoid cyclases. *Chem. Rev.* **2017**, *117* (17), 11570−11648.
(56) Tang, J.; Matsuda, Y. Dissection of the catalytic mechanisms of transmembrane terpene cyclases involved in fungal meroterpenoid biosynthesis. *Angew. Chem., Int. Ed.* **2023**, No. e202306046.
(57) Gerlt, J. A.; Bouvier, J. T.; Davidson, D. B.; Imker, H. J.; Sadkhin, B.; Slater, D. R.; Whalen, K. L. Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochim. Biophys. Acta Proteins Proteom.* **2015**, *1854* (8), 1019−1037.
(58) Zallot, R.; Oberg, N.; Gerlt, J. A. The EFI web resource for genomic enzymology tools: Leveraging protein, genome, and metagenome databases to discover novel enzymes and metabolic pathways. *Biochemistry* **2019**, *58* (41), 4169−4182.

https://doi.org/10.1021/acschembio.3c00225
ACS Chem. Biol. XXXX, XXX, XXX−XXX