JOC The Journal of Organic Chemistry

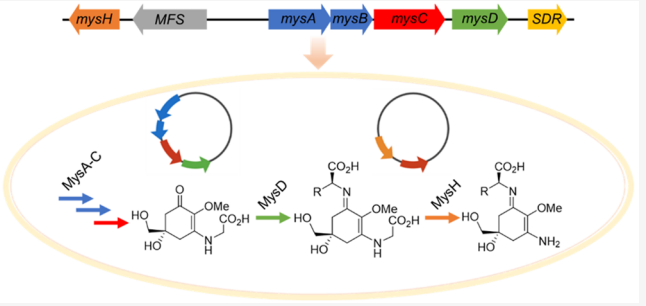
pubs.acs.org/joc Article

Biosynthesis and Heterologous Production of Mycosporine-Like Amino Acid Palythines

Manyun Chen, Garret M. Rubin, Guangde Jiang, Zachary Raad, and Yousong Ding*

Cite This: J. Org. Chem. 2021, 86, 11160−11168 Read Online

ABSTRACT: Mycosporine-like amino acids (MAAs) are a family of natural products that are produced by a variety of organisms for protection from ultraviolet damage. In this work, we combined different bioinformatic approaches to assess the distribution of the MAA biosynthesis and identified a putative gene cluster from *Nostoc linckia* NIES-25 that encodes a short-chain dehydrogenase/reductase and a nonheme iron(II)- and 2-oxoglutarate-dependent oxygenase (MysH) as potential new biosynthetic enzymes. Heterologous expression of refactored gene clusters in *E. coli* produced two known biosynthetic intermediates, 4-deoxygadusol and mycosporine-glycine, and three disubstituted MAA analogues, porphyra-334, shinorine, and mycosporine-glycine-alanine. Importantly, the disubstituted MAAs were converted into palythines by MysH. Furthermore, biochemical characterization revealed the substrate preference of recombinant MysD, a D-Ala-D-Ala ligase-like enzyme for the formation of disubstituted MAAs. Our study advances the biosynthetic understanding of an important family of natural UV photoprotectants and opens new opportunities to the development of next-generation sunscreens.



INTRODUCTION

Skin cancers are among the most common cancer types in the United States with about 1.2 million Americans living with melanoma and 3 million more affected by nonmelanoma skin cancers.[1,2] Solar radiation, especially ultraviolet (UV) radiation, is an established risk factor of skin cancers,[3] as more than 90% of melanoma in some populations is linked to sunlight exposure.[4] UV rays, mainly UVA (315−400 nm) and UVB (280−315 nm), induce a variety of damages on the biomolecules (e.g., DNA and proteins) of living organisms on earth.[5] In addition to behavioral changes, proper skin protection from excessive sun exposure has proven to be effective in reducing skin cancers.[6] In this regard, many organic and inorganic compounds have been developed to dissipate the energy of UV rays and/or directly block their reach on the skin, and some have been used as active ingredients of commercial sunscreens.[7] However, there are increasing concerns on the potential negative health impact of synthetic sunscreens (e.g., endocrine disruption, neurotoxicity, and systemic absorption),[8−10] while multiple organic UV filters are accumulated in almost all water sources globally and may be potential contributors to coral reef bleaching, raising a severe environmental concern about their use.[11] These concerns highlight a need for new safer, biodegradable, and environmentally friendly sunscreen compounds. On the other hand, natural organisms have developed multiple effective UV mitigation strategies when utilizing solar energy, including the biosynthesis of diverse natural products as photoprotectants.[12,13] These natural products (e.g., flavonoids, phenols, terpenoids, and polyketides) absorb UV radiation and release energy through thermal de-excitation, similar to synthetic chemical UV filters, while providing additional protection from UV-induced damages with other biological functions, e.g., antioxidant, anti-inflammation, and immunomodulation.[14] These compounds provide important inspirations for the development of a new generation of sunscreens.[15]

Mycosporine-like amino acids (MAAs) are a family of natural, thermally and photochemically stable UV protectants (Figure 1A).[16] Originally isolated from terrestrial fungal species, over 30 MAA analogues have been identified from taxonomically diverse marine and terrestrial organisms, e.g., cyanobacteria, eukaryotic algae, corals, plants, and vertebrates, and possess various functional groups at the C1 and to a lesser extent the C3 of the characteristic cyclohexenimine core (Figure 1A).[16−18] Indeed, the majority of MAAs carries a C3-L-Gly moiety, though L-Ala, L-Glu, and other amine-containing components also appear. Common amino acid building blocks at the C1 include L-Ser (shinorine), L-Thr (porphyra-334), L-

Special Issue: Natural Products: An Era of Discovery in Organic Chemistry

Received: February 14, 2021
Published: May 18, 2021

https://doi.org/10.1021/acs.joc.1c00368
J. Org. Chem. 2021, 86, 11160−11168

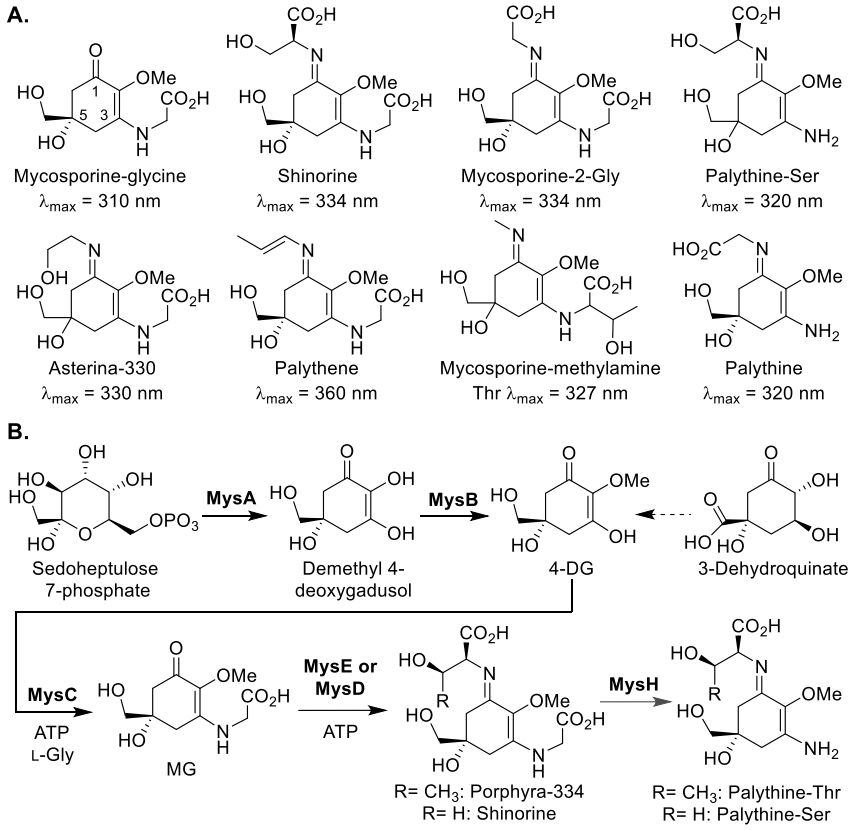


Figure 1. (A) Chemical structure and maximal absorbance of representative mycosporine-like amino acid analogues. (B) The biosynthetic pathway of shinorine, porphyra-334, palythine-Ser, and palythine-Thr. The last arrow indicates a new biosynthetic step uncovered in this work.

Gly (mycosporine-2-Gly), and L-Ala.[16−18] These moieties at the C1 and C3 can likely be converted into other functional groups, including amino alcohol (e.g., asterina-330), enaminone (e.g., palythene), methyl amine (e.g., mycosporine-methylamine-Thr), or an amine group (e.g., palythine and palythine-Ser),[17,19] while some glycosylated MAAs have also been produced in a variety of organisms.[20,21] Of note, except a few analogues (e.g., mycosporine-glycine, porphyra-334, palythene, and palythine),[22−25] the absolute configuration of the majority of MAAs, particularly the C5, has not been fully elucidated. Despite notable structural diversity, these MAA analogues display absorption maxima between 310 and 362 nm and possess extinction coefficients of up to 50 000 $M^{-1}\cdot cm^{-1}$.[16,17] They are among the strongest UV-absorbing compounds, and the cyclohexenimine core is critical for the dissemination of UV energy. Furthermore, accumulated evidence demonstrates the antioxidative, anti-inflammatory, and antiaging properties of MAAs, providing another mechanism of photoprotection.[14] The superior UV protection properties of MAAs can lead to the development of next-generation sunscreens for broad cosmetic applications if the low quantity available from natural resources or the lack of efficient synthetic preparation is properly addressed.[25,26]

Recently, several initial biosynthetic steps of MAAs have been elucidated through biochemical and genetic studies. Their biosynthesis starts from the production of 4-deoxygadusol (4-DG) from sedoheptulose 7-phosphate, an intermediate of the pentose phosphate pathway, by a dimethyl 4-degadusol synthase (DDGS; MysA) and an *O*-methyltransferase (*O*-MT; MysB) (Figure 1B).[27] In some microbes, 4-DG may also be produced from 3-dehydroquinate of the shikimate pathway through incompletely defined enzymatic steps.[28] Interestingly, the Mahmud group reported that some vertebrates use a 2-epi-5-epi-valiolone synthase (EEVS) and a bifunctional enzyme to convert sedoheptulose 7-phosphate into gadusol as a sunscreen compound.[29] Both EEVS and DDGS are members of the sugar phosphate cyclase superfamily,[30] while the bifunctional enzyme carries an N-terminal MT domain and a C-terminal NAD^+-dependent oxidoreductase domain. 4-DG is next converted into mycosporine-glycine (MG) by an ATP-grasp enzyme MysC through introducing an amino acid moiety, primarily L-Gly, at the C3 (Figure 1B). Intriguingly, the Walsh group revealed that MysC from the cyanobacterium *Anabaena variabilis* ATCC 29413 phosphorylates 4-DG rather than L-Gly, typical of other ATP-grasp enzymes.[27] MG is the direct biosynthetic precursor of disubstituted MAAs (e.g., porphyra-334) with an amino acid moiety at the C1 (Figure 1B). The Walsh group biochemically confirmed that a nonribosomal peptide synthetase (NRPS)-like enzyme MysE, which contains an adenylation (A), a thiolation (T), and a thioesterase (TE) domain, catalyzes this step in the biosynthesis of shinorine.[27] On the other hand, an MAA biosynthetic gene cluster (BGC) from the cyanobacterium *Nostoc punctiforme* ATCC 29133 has no NRPS gene but a D-Ala-D-Ala ligase-like enzyme gene *mysD*.[31] The heterologous expression of this BGC in *E. coli* produces three MAA analogues, shinorine (the major product), porphyra-334, and mycosporine-2-Gly, confirming MysD's involvement in the MAA biosynthesis. However, the following biosynthetic route from disubstituted MAAs to other MAA analogues remains completely unknown.

https://doi.org/10.1021/acs.joc.1c00368
J. Org. Chem. 2021, 86, 11160−11168

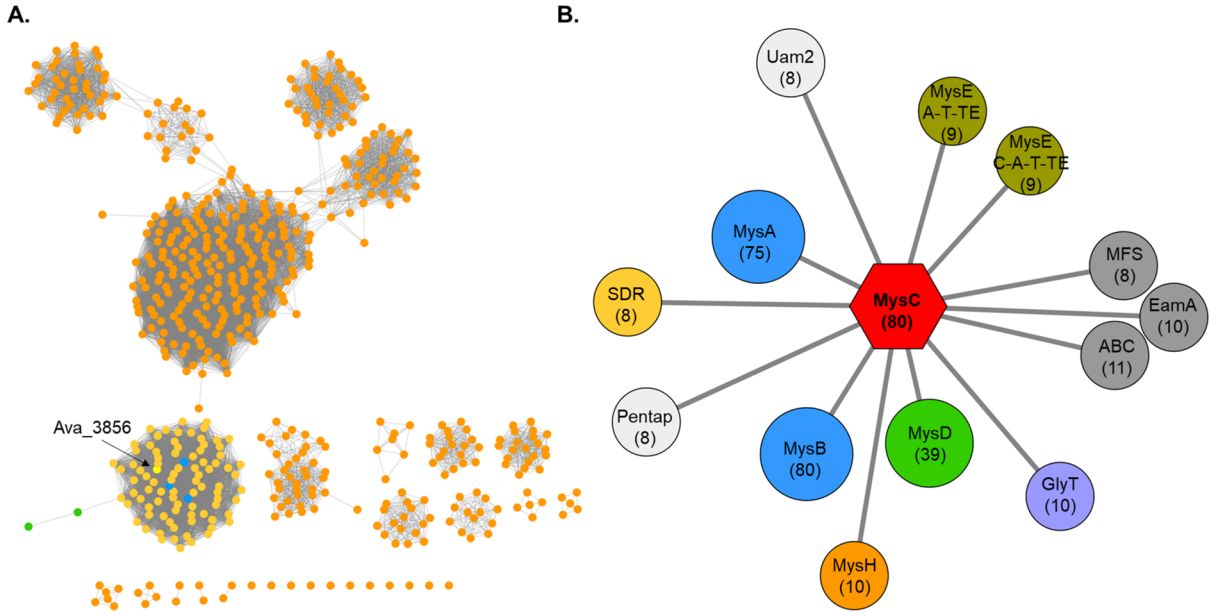


Figure 2. (A) Sequence similarity network (SSN) of one cluster with 585 members in Figure S1 with >45% protein sequence identity. One cluster was formed by 92 MysC homologues including Ava_3856 highlighted in yellow. Blue and green dots represent homologues from α-proteobacteria and eukaryotes, respectively. (B) Genome neighborhood network (GNN) analysis identified enzymes with 8 times or more of co-occurrence within ten open reading frames upstream or downstream of 80 MysC homologues. The occurrence times of each enzyme group were labeled. GlyT: glycosyltransferase; Pentap: pentapeptide repeats; Uam2: putative restriction endonuclease.

We report herein the heterologous production of serial MAA analogues including palythines in *E. coli*. Our sequence similarity network (SSN) and genome neighborhood network (GNN) analyses of known MAA biosynthetic enzymes identified a putative *mysD*-containing BGC in the genome of *Nostoc linckia* NIES-25 that is adjacent to a short-chain dehydrogenase/reductase (SDR) gene and a nonheme iron-(II)- and 2-oxoglutarate-dependent (Fe/2OG) oxygenase gene *mysH*.[32] Heterologous expression of multiple refactored MAA BGCs in *E. coli* produced MAA analogues and demonstrated the direct conversion of disubstituted MAAs into palythines by the Fe/2OG enzyme MysH. Furthermore, biochemical characterization of its recombinant MysD supported its role in the formation of porphyra-334, shinorine, and other MAA analogues. This work opens new opportunities for the development of next-generation sunscreens via synthetic biology and biocatalysis approaches.

RESULTS AND DISCUSSION

Distribution of MAA BGCs in Microbial Genomes. Genome mining has become a powerful approach for the discovery of new natural products and enzymology,[33] supported by the exponential growth of genomic sequence data. To probe the distribution of MAA BGCs, we first used MysC (Ava_3856) from *A. variabilis* ATCC 29413 as the query to mine its homologues in the UniRef50 database that includes all proteins with at least 50% sequence identity to and 80% overlap with the longest sequence in the family.[27,34] This analysis revealed that MysC belongs to the protein family #02655 (ATP_Grasp_3, PF02655) in the Pfam database,[35] which includes 8435 ATP grasp-enzyme homologues (Oct 2020). Our subsequent SSN analysis of this family identified 22 distinct clusters with a sequence identity of >35% (Figure S1). One cluster of 585 members was reanalyzed to separate homologues with >45% protein sequence identity into 15 clusters and 11 singletons, including one cluster formed by 92 MysC homologues (Figure 2A, Table S1). Except for three MysC homologues from α-proteobacteria (e.g., *Mycobacterium* sp.) and two from eukaryotes (e.g., *Chromera velia*), the rest all are from cyanobacteria. The result of this bioinformatics analysis suggested that MAAs are mainly used by cyanobacteria for photoprotection. On the other hand, the increasing availability of eukaryotic genomes (e.g., fungi, corals, and macroalgae) will lead to a more complete understanding of the MAA genomic distribution.

We next performed the GNN analysis of the MysC cluster to identify enzymes with high co-occurrence frequency within ten open reading frames upstream or downstream of MysC (Figure 2B). A total of 12 MysC homologues had no nearby open reading frames in the GNN analysis and were removed from further analysis. These homologues are all predicted from unassembled whole-genome shotgun sequencing projects, likely reflecting poor sequencing quality. As expected, homologues of MysA (75), MysB (80), NRPS MysE (18), and MysD (39) were frequently colocalized with 80 MysC homologues to form the MAA BGC (Figure 2B). In addition to MysEs with the A-T-TE domain organization,[27] nine enzymes carry an additional condensation (C) domain.[21] Furthermore, we observed the high occurrence of transporters (29) including ABC, EamA-like,[36] and major facilitator superfamily (MFS) transporters, though almost all known MAAs have been extracted only from biomasses and might be located in the extracellular matrix.[17,21,26,37,38] A recent study also found the frequent presence of a transporter gene within the MAA BGCs in the microbial mat communities of Shark Bay, Australia.[39] Importantly, the GNN analysis revealed three enzyme groups that may contribute to the structural diversity of MAAs, including glycosyltransferases (10), phytanoyl-CoA dioxygenases (10), and short-chain dehydrogenases/reductases (SDRs, 8). Although many glycosylated MAA analogues have been reported, the corresponding glycosyltransferases remain unidentified.[21] Phytanoyl-CoA dioxygenases belong to the

https://doi.org/10.1021/acs.joc.1c00368
J. Org. Chem. 2021, 86, 11160−11168

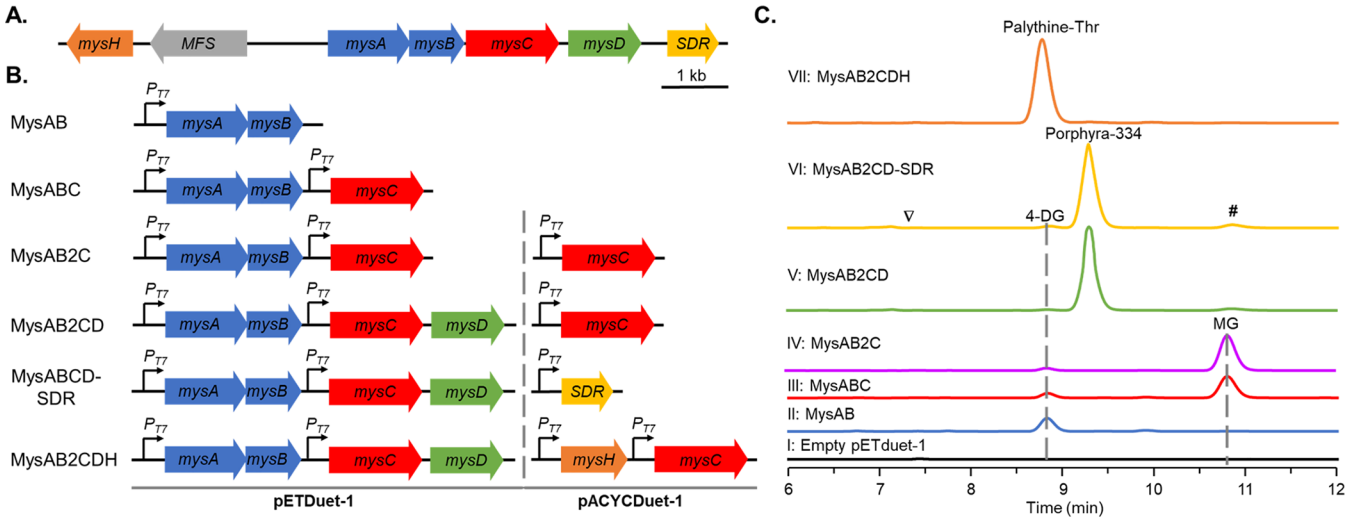


Figure 3. (A) Gene organization of the MAA gene cluster from *Nostoc linkia* NIES-25. (B) Representative refactored MAA clusters cloned into pETDuet-1 and pACYCDuet-1. (C) HPLC traces of crude extracts of *E. coli* cells expressing refactored MAA clusters. All products were detected at 310 nm. ∇ and # indicate shinorine and MG-Ala, respectively.

Fe(II)/2OG enzyme family, and the 10 enzymes colocalized with MysCs all carry the catalytically essential 2-His-1-carboxylate facial triad for coordinating Fe(II) (Figure S2).[40] Phytanoyl-CoA dioxygenases catalyze the α-hydroxylation of phytanoyl-CoA in the degradation of phytanic acid.[41] On the other hand, members of the Fe(II)/2OG enzyme family are known to catalyze a wide range of reactions, e.g., hydroxylation, decarboxylation, dehydration, oxidation, reduction, isomerization, ring formation, and expansion,[42] some of which may lead to the production of MAA analogues (Figure 1A). We here name these phytanoyl-CoA dioxygenases related to the MAA biosynthesis as MysHs. Similar to Fe(II)/2OG enzymes, SDRs form a large protein superfamily that demonstrates a broad substrate range and rich function diversity.[43] Two other protein groups that frequently co-occur with MysC are restriction endonucleases and pentapeptide repeats, whose roles in the biosynthesis of MAAs are unclear.

Heterologous Expression of Refactored MAA BGCs from *Nostoc linkia* NIES-25 in *E. coli*. Based on the results of the above bioinformatics studies, we sought to characterize new MAA biosynthetic enzymes. Specifically, we selected a putative 9.6-kb MAA BGC from a 1.78-Mb plasmid (GenBank: AP018223.1) in *Nostoc linkia* NIES-25, which encodes MysA-D (NIES25_64130 to NIES25_64160), a phytanoyl-CoA dioxygenase (MysH, NIES25_64110), an MFS transporter (NIES25_64120), and an SDR (NIES25_64170) (Figure 3A, Table S2). To examine the expression of this cluster in *N. linkia* NIES-25, we cultured this strain in BG-11 medium at 26 °C for 21 days. However, HPLC analysis of methanolic extracts of pelleted cells and lyophilized culture medium failed to identify any peak with maximal absorbance between 310 and 360 nm. On the other hand, extracted ion chromatogram (EIC) extraction of LC high-resolution (HR) MS data of methanolic extracts of pelleted cells revealed a peak corresponding to the parental ion of porphyra-334 (HRMS (ESI) *m/z*: $[M + H]^+$ calcd for $C_{14}H_{23}N_2O_8^+$ 347.1449, found 347.1444, Figure S3), whose selective MS/MS fragmentation ions further suggested the production of porphyra-334. EIC analysis further suggested a putative peak corresponding to shinorine (HRMS (ESI) *m/z*: $[M + H]^+$ calcd for $C_{13}H_{21}N_2O_8^+$ 333.1292, found 333.1400, Figure S3A), but its low abundance yielded only a low-quality MS/MS spectrum, preventing a reliable structural identification. A peak for putative MG-Ala (HRMS (ESI) *m/z*: $[M + H]^+$ calcd for $C_{13}H_{21}N_2O_7^+$ 317.1343) was not observed in EIC analysis (Figure S3A). Nonetheless, this study suggested that the MAA cluster in *N. linkia* NIES-25 is active under the laboratory culturing conditions.

To further characterize the MAA biosynthesis in *N. linkia* NIES-25, we designed multiple refactored BGCs for heterologous expression in *E. coli* BL21-Gold (DE3) (Figure 3B). The coexpression of *mysAB* under the control of the T7 promoter in pETDuet-1 led to the production of 4-DG (Figure 3C, **II**), which showed maximal absorbance at 294 nm and a protonated ion of *m/z* 189.0751 (HRMS (ESI) *m/z*: $[M + H]^+$ calcd for $C_8H_{13}O_5^+$ 189.0757, Figure S4), agreeing with reported data.[26] 4-DG was only detected from the methanolic extract of cell pellets, the same for all other MAAs described below. No 4-DG was detected in the control transformed with the empty pETDuet-1 (Figure 3C, **I**). When *mysC* was expressed along with *mysAB* in pETDuet-1, we observed the production of MG in *E. coli* (Figure 3C, **III**) as confirmed by its maximal absorbance at 310 nm and protonated ion of *m/z* 246.0963 (HRMS (ESI) *m/z*: $[M + H]^+$ calcd for $C_{10}H_{16}NO_6^+$ 246.0972, Figure S4). A small quantity of 4-DG was still observed (Figure 3C, **III**), suggesting the imbalanced catalytic activity of MysC compared with MysAB. Indeed, when one additional copy of *mysC* was coexpressed in a middle-copy number vector pACYCDuet-1 (Figure 3B, *mysAB2C*), the peak area of MG was improved by about 1.5 times, while that of 4-DG was decreased by about 50% (Figure 3C, **IV**). Next, we examined the catalytic function of MysD in the production of disubstituted MAAs by coexpressing its gene with *mysAB2C* in *E. coli* (Figure 3B). HPLC analysis of the methanolic extract of *E. coli* pellets expressing *mysAB2CD* revealed one new major peak with the retention time of 9.3 min and one new minor peak at 10.8 min, while 4-DG was still found (Figure 3C, **V**). These new peaks showed the same maximal absorbance at around 334 nm (Figures S5–S6). HRMS and MS/MS analysis indicated the production of porphyra-334 as the major peak (HRMS (ESI) *m/z*: $[M + H]^+$ calcd for $C_{14}H_{23}N_2O_8^+$

https://doi.org/10.1021/acs.joc.1c00368
J. Org. Chem. 2021, 86, 11160−11168

347.1449, found 347.1436, Figure S5), whose yield reached 3.5 ± 0.2 mg/L. The minor peak showed the protonated ion of *m/z* 317.1332 (HRMS (ESI) *m/z*: $[M + H]^+$ calcd for $C_{13}H_{21}N_2O_7^+$ 317.1343, Figure S6), and HRMS/MS analysis indicated it to be MG-Ala.[37] As shinorine is commonly isolated along with porphyra-334, a careful search of the LC and LC-MS spectra led to the identification of shinorine with a retention time of 7.3 min (Figure 3C, **V**), a protonated ion of *m/z* 333.1279 (HRMS (ESI) *m/z*: $[M + H]^+$ calcd for $C_{13}H_{21}N_2O_8^+$ 333.1292), and an expected MS/MS fragmentation (Figure S7). The production of these three disubstituted MAAs demonstrates that MysD from *N. linkia* NIES-25 functionalizes the C1 of MG, using multiple amino acids as substrates, with a strong preference to L-Thr. Substrate promiscuity of MysD has previously been observed in the heterologous expression of the MAA BGC from *N. punctiforme* ATCC 29133 and *Actinosynnema mirum* DSM 43827 in *E. coli* and *Streptomyces avermitilis* SUKA22, respectively.[31,37] In both cases, shinorine was the dominant product, suggesting different substrate preferences of MysD of different origins.

The successful production of disubstituted MAAs by expressing *mysA-D* from *N. linkia* NIES-25 in *E. coli* prompted us to characterize the functions of two other biosynthetic genes in the cluster. Coexpression of *sdr* on pACYCDuet-1 (Figure 3B) had no obvious change on the product profile of *E. coli* expressing *mysABCD* on pETDuet-1 (Figure 3C, **VI**), suggesting the unclear enzymatic function of SDR for the MAA biosynthesis. Similarly, the *sdr* gene is adjacent to the MAA BGC in *Scytonema* cf. *crispum* UCFS15, and its coexpression with the cluster produces only shinorine in *E. coli*.[21] In contrast, when *mysH* was cloned alone or with the second copy of *mysC* in pACYCDuet-1 (Figure 3B) and expressed in *E. coli* transformed with *mysABCD*, we observed a new major peak with the retention time of close to 8.8 min, concurrent with the almost complete disappearance of porphyra-334 (Figure 3C, **VII**, Figure S8). The content of the new peak showed maximal absorbance at 320 nm, and its molecular formula was established as $C_{12}H_{20}N_2O_6$ based on a protonated ion of *m/z* 289.1382 (HRMS (ESI) *m/z*: $[M + H]^+$ calcd for $C_{12}H_{21}N_2O_6^+$ 289.1394, Figure S9). HRMS/MS analysis of the parent molecular ion generated multiple fragment ions (e.g., *m/z* 245.112, 186.099, and 172.083), suggesting the peak content as palythine-Thr (Figure S9),[20,44] whose yield in *E. coli* transformed with *mysAB2CDH* reached 2.7 ± 0.3 mg/L. To further elucidate its structure, we purified about 1 mg of this compound for 1D and 2D NMR analysis (Table S3, Figures S10−12). Comparison of its ^1H and ^{13}C chemical shifts to those of palythine-Thr in a recent report allowed the assignment of 3-aminocyclohexenimine (C1, 2, 3, 4, 5, and 6) and Thr (C9, 10, 11, and 12, Table S3).[45] Furthermore, the assignment of the Thr moiety was supported by C12−H/C11−H/C9−H COSY correlations and HMBC correlations from C12−H to C9/C11 and from C9−H to C10 (Figure 4). The presence of a 3-aminocyclohexenimine moiety was supported by the HMBC correlations from C4−H to C2/C3/C5/C6, from C6−H to C1/C2/C5, and from C7−H to C4/C5/C6. The connectivity of the Thr and 3-aminocyclohexenimine moieties was further confirmed by the HMBC correlation from C9−H to C1 (Figure 4). Additionally, the HMBC correlation from C8−H to C2 supported the presence of a methoxy group at the C2 (Figure 4). Collectively, the combination of HRMS and NMR analyses indicates the production of palythine-Thr in *E. coli*, expressing

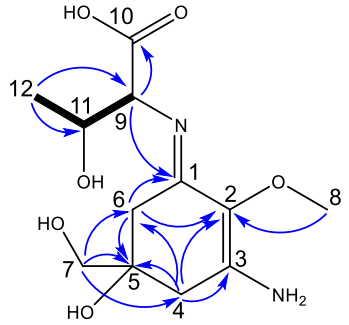


Figure 4. $^1H-^1H$ COSY (bold) and selected HMBC (H → C) correlations of isolated palythine-Thr.

mysAB2CDH from *N. linkia* NIES-25. Importantly, these results support the direct conversion of porphyra-334 into palythine-Thr catalyzed by MysH (Figures 3 and 4), an advancement in understanding the MAA biosynthesis. Given the same biosynthetic origin, palythine-Thr should share the same C5-*S* configuration as porphyra-334.

Currently known palythines include palythine, palythine-Ser, palythine-Thr, and their derivatives produced by corals, cyanobacteria, and other organisms (Figure 1A).[16,46] Similar to the biosynthesis of palythine-Thr, palythine and palythine-Ser may be converted directly from the corresponding mycosporine-2-Gly and shinorine by MysH homologues (Figure S13A) and retain the same C5-*S* configuration (Figure 1A). To our knowledge, the direct conversion of the L-Gly moiety into the amine is a new reaction to the Fe(II)/2OG enzyme family.[42] One potential reaction path is that MysH catalyzes an α-hydroxylation on the C3-L-Gly moiety, followed by automatic hydrolysis to release palythines and glyoxylic acid (Figure S13A). The C3-amine of palythines can be further methylated by an *N*-methyltransferase to produce MAA analogues carrying a C3-methylamine (e.g., mycosporine-methylamine-Thr, Figure 1A).[16] Since *E. coli* expressing *mysAB2CD* produced porphyra-334, shinorine, and MG-Ala (Figure 3C, **V**), we intended to check the formation of palythine-Ser and palythine-Ala in the crude extract of *E. coli* cell pellets expressing *mysAB2CDH*. Indeed, we identified expected *m/z* values 275.1227 and 259.1288 for these two palythines (HRMS (ESI) *m/z*: $[M + H]^+$ calcd for $C_{11}H_{19}N_2O_6^+$ 275.1238 for palythine-Ser; HRMS (ESI) *m/z*: $[M + H]^+$ calcd for $C_{11}H_{19}N_2O_5^+$ 259.1288 for palythine-Ala, Figures S14−15), indicating the substrate promiscuity of MysH. Palythine-Ser showed maximal absorbance at 320 nm, and HRMS/MS fragmentations of both compounds suggested their structure assignment (Figure S14−15). Further biochemical and mechanistic studies of MysH are underway. Finally, we coexpressed both *mysH* and *SDR* with *mysABCD* in *E. coli* and observed the same product profile as that of the coexpression of *mysABCDH* (Figure S8), indicating that SDR may not take any palythines as the substrate.

Biochemical Characterization of Recombinant MysD. The current and previous heterologous expression studies supported the function of MysD in the biosynthesis of disubstituted MAAs (Figure 3).[31,37] To further characterize its catalytic properties, we prepared recombinant His_6-tagged MysD of *N. linkia* NIES-25 from *E. coli* after a single affinity purification (Figure S16). The enzyme reaction was performed with MysD (0.5 μM), purified MG (50 μM), and L-Thr (1 mM) in the presence of ATP (1 mM) and Mg^{2+} (10 mM) at

https://doi.org/10.1021/acs.joc.1c00368
J. Org. Chem. 2021, 86, 11160−11168

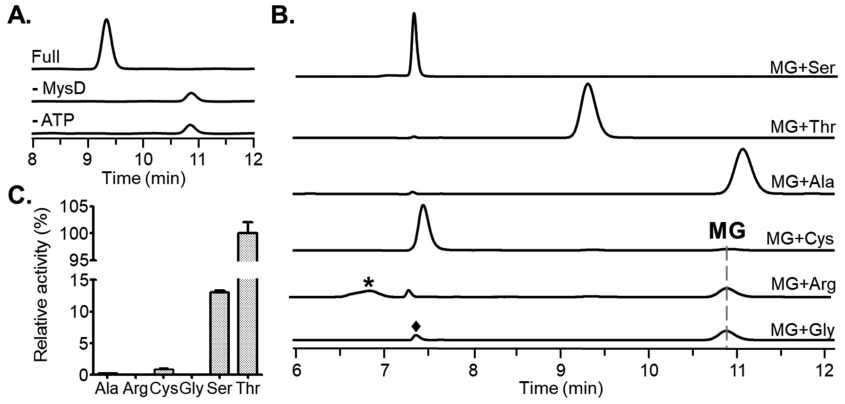


Figure 5. (A) HPLC traces of the MysD reactions with MG and L-Thr as substrates. Porphyra-334 was produced in the full reaction, which was missed in the control reaction without MysD or ATP. (B) HPLC analysis indicated that MysD accepted L-Ala, L-Arg, L-Cys, L-Gly, L-Ser, and L-Thr as its amino acid substrate. * and ♦ indicate MG-Arg and MG-2-Gly, respectively. The detection wavelength was 334 nm. (C) Relative activities of six amino acid substrates in the MysD reaction. The formation of porphyra-334 in the MysD reaction containing L-Thr after 8 min was determined in HPLC analysis. The corresponding MG consumption level was set as 100% to normalize the relative MG consumption levels in five other reactions that were performed for 30 min to allow the quantitation of corresponding disubstituted MAAs. Data represent mean ± s.d. of two independent experiments.

room temperature for 2 h. HPLC analysis of the reaction mixture identified the formation of porphyra-334 (Figure 5A), which showed the same maximal absorbance and MS spectrum as that from the heterologous production (Figure 3C, Figure S5). No product was formed in the control reactions without enzyme or ATP (Figure 5A). The requirement of ATP for the MysD reaction supports its prediction as the D-Ala-D-Ala ligase-like enzyme of the ATP grasp superfamily.[31] We further determined the optimal temperature and pH of its reaction at 37 °C and pH = 8.5 (Figure S17). Under these optimal reaction conditions, we screened all 20 natural amino acids (5 mM) along with MG (50 μM) in the MysD reaction. HPLC analysis found that MysD was able to accept six amino acids as its substrate, including L-Ala, L-Arg, L-Cys, L-Gly, L-Ser, and L-Thr (Figure 5B, Figures S18A and S18C). LC-HRMS and MS/MS analysis indicated the formation of their corresponding disubstituted MAAs, MG-Ala, MG-Arg (HRMS (ESI) m/z: $[M + H]^+$ calcd for $C_{16}H_{28}N_5O_7^+$ 402.1983, found 402.1977), MG-Cys (HRMS (ESI) m/z: $[M + H]^+$ calcd for $C_{13}H_{21}N_2O_7S^+$ 349.1064, found 349.1059), mycosporine-2-Gly (HRMS (ESI) m/z: $[M + H]^+$ calcd for $C_{12}H_{19}N_2O_7^+$ 303.1187, found 303.1182), shinorine, and porphyra-334 (Figures S19–21). L-Ser and L-Thr led to the complete consumption of MG in the MysD reactions after 3 h, followed by L-Cys. The retention times of MG-Ala and MG were very close, and the left shoulder of the MG-Ala peak at 310 nm suggested a small amount of MG left in the reaction (Figure S18B). Nonetheless, the result of this biochemical study well agreed with the production of porphyra-334 along with small amounts of shinorine and MG-Ala in the above heterologous expression study (Figure 3C). Analogous to MysC,[27] two potential reaction paths of MysD for the formation of disubstituted MAAs are depicted in Figure S13B. To further understand MysD's substrate preference, we lowered the enzyme concentration to 0.25 μM. Under these conditions, MysD showed the highest activity toward L-Thr, converting about 40% MG into porphyra-334 in 8 min. The consumed MG level in this reaction was set as 100% to normalize its level in the five other reactions, which was determined from the concentrations of produced disubstituted MAAs in the reactions (Figure 5C). This quantitative analysis showed that the consumption level of MG in the MysD reaction containing L-Ser was about 12.7% to that with L-Thr, followed by L-Cys (0.9%) and L-Ala (0.4%), and two other amino acids (about 0.06%). Together, the results of these biochemical studies highlight the broad substrate scope of MysD and its strong preference toward L-Thr.

Recent advances in bioinformatics and synthetic biology tools have unleashed the potential of all organisms for the discovery of new natural products and new enzymology for a variety of applications.[47] In the search for new MAA analogues, we identified a group of Fe(II)/2OG enzymes that frequently co-occur with the known MAA biosynthetic enzymes. Refactoring such an MAA BGC from *N. linkia* NIES-25 for the heterologous expression in *E. coli* interrogated the catalytic functions of MysA, MysB, MysC, MysD, MysH, and one SDR for the biosynthesis of MAA analogues. The direct conversion of disubstituted MAAs into corresponding palythines by MysH filled a critical gap in the biosynthetic understanding of many MAA analogues produced by a variety of prokaryotic and eukaryotic organisms. Furthermore, this work provided the first biochemical insights into the substrate preference of MysD. The study presented here has thus set the stage for future synthetic biology and biocatalysis studies to produce new MAA analogues to drive the development and applications of this important family of natural photoprotectants.

EXPERIMENTAL PROCEDURES

General Experimental Procedures. Molecular biology reagents and chemicals were purchased from Thermo Scientific, NEB, Fisher Scientific, or Sigma-Aldrich. GeneJET Plasmid Miniprep Kit and GeneJETGel Extraction Kit (Thermo Scientific) were used for plasmid preparation and DNA purification, respectively. *E. coli* DH5α (Agilent) was used for routine cloning studies, and *E. coli* BL21-gold(DE3) (Agilent) was used for protein expression and heterologous production. The cyanobacterial strain *Nostoc linkia* NIES-25 was obtained from the National Institute for Environmental Studies, Japan. DNA sequencing was performed with GENEWIZ or Eurofins. A Shimadzu Prominence UHPLC system (Kyoto, Japan) coupled with a PDA detector was used for HPLC analysis. NMR spectra were recorded in D_2O on a Bruker 600 MHz spectrometer located in the AMRIS facility at the University of Florida, Gainesville,

https://doi.org/10.1021/acs.joc.1c00368
J. Org. Chem. 2021, 86, 11160−11168

FL, USA. Spectroscopy data were collected using Topspin 3.5 software. HRMS data were generated on a Thermo Fisher Q Exactive Focus mass spectrometer equipped with an electrospray probe on a Universal Ion Max API source. Structural assignments were made with additional information from gCOSY, gHSQC, and gHMBC experiments.

Bioinformatics Analysis. The SSN of ATP-grasp ligases (ATP_Grasp_3, PF02655) was generated by the EFI-Enzyme Similarity Tool (https://efi.igb.illinois.edu/efi-est/) with an ∼35% cutoff threshold.[32] The identified MysC-containing cluster (585 homologues) was further reanalyzed with an ∼45% cutoff threshold. The resultant MysC-containing cluster was submitted for GNN analysis (https://efi.igb.illinois.edu/efi-gnt/) with a neighborhood size set at 10 and a co-occurrence lower limit set at 10%. All the SSNs and GNN were visualized in Cytoscape.[48] The amino acid sequences of mined MysH homologues were aligned by the ClustalW algorithm.[49]

Construction of Refactored BGCs. The MAA biosynthetic genes were amplified from isolated genomic DNA of *Nostoc linkia* NIES-25. The *mysAB* together were amplified and cloned into pETDuet-1 *Nco*I/*Pst*I sites to give pETDuet-1-mysAB. The *mysC* or *mysCD* was then cloned into the *Kpn*I/*Xho*I site of pETDuet-1-mysAB to give pETDuet-1-mysABC and pETDuet-1-mysABCD. The *sdr* was cloned into the *Nde*I/*Xho*I site of pACYCDuet-1, and the *mysH* was cloned into the *Nco*I/*Pst*I site of pACYCDuet-1 or pACYCDuet-1-sdr. The *mysC* was then cloned into the *Kpn*I/*Xho*I site of pACYCDuet-1 or pACYCDuet-1-mysH. All oligonucleotide primers (Table S4) used were ordered from Sigma-Aldrich. The resultant constructs were transformed or cotransformed into *E. coli* BL21-gold(DE3). After appropriate antibiotics selection, positive clones were used for fermentation.

Fermentation, Extraction, and Isolation. To characterize MAA production in its native producer, *Nostoc linkia* NIES-25 was cultured in 300 mL of BG-11 medium (Sigma-Aldrich) at 26 °C. The culture was air bubbled and received a lighting cycle of 16 h/8 h (light/dark) with the illumination of 2000−2500 lx. After 21 days, the cells were pelleted down by centrifugation (4500 rpm, 15 min). The cyanobacterial cell pellet was lysed by sonication in ice-cold methanol (10 s pulse and 20 s rest, 2 min pulse in total). After centrifugation (4500 rpm, 30 min), the clear supernatants of lysates were collected and evaporated under reduced pressure. The dried extracts were resuspended in water (1 mL) for HPLC and LC-HRMS analysis. Following the same procedure, the expensed culture medium was lyophilized and redissolved in water (1 mL) for HPLC and LC-MS analysis.

To characterize the heterologous expression of the MAA BGC from *Nostoc linkia* NIES-25, *E. coli* strains carrying refactored gene clusters were cultured in 2 × 50 mL of Luria−Bertani broth supplemented with 50 μg/mL of ampicillin and/or chloramphenicol (37 °C, 225 rpm). When the cell culture OD_{600} reached 0.5, IPTG (final concentration 0.1 mM) was added to the culture to induce gene expression (18 °C, 180 rpm, 20 h). The cells were harvested by centrifugation (4500 rpm, 10 min), and collected cell pellets were extracted twice by 1 mL of methanol. The methanolic extracts were dried in the speed vacuum concentrator and resuspended in water (300 μL) for HPLC and LC-MS analysis.

For the large-scale production of palythine-Thr, *E. coli* expressing *mysAB2CDH* was cultured in 8 × 1 L of Luria−Bertani broth using the same expression conditions as described above. After expression, the cells were harvested by centrifugation (6000 rpm, 20 min) and lysed by sonication in 2 × 30 mL of ice-cold methanol (10 s pulse and 20 s rest, 8 min pulse in total). The cell lysates were centrifuged (4500 rpm, 10 min), and the clear supernatants were evaporated under reduced pressure. The dried methanolic extracts were resuspended in 1 mL of water and were first purified on an Agilent Zorbax SB-C18 column (9.4 × 250 mm, 5 μm) using 0.1% formic acid in water and 2% methanol as mobile phases. Corresponding fractions were collected (maximal absorption at 320 nm), combined, evaporated to remove organic solvents, and then lyophilized. The residues were resuspended in water (200 μL) and further purified on a Phenomenex Luna C8 column (4.6 × 250 mm, 5 μm) using the same mobile phases as above. Palythine-Thr fractions were collected, combined, evaporated to remove organic solvents, and lyophilized. About 1 mg of palythine-Thr was purified for NMR analysis.

Palythine-Thr: white solid. HRMS (ESI) m/z: $[M + H]^+$ calcd for $C_{12}H_{21}N_2O_6^+$ 289.1394, found 289.1382. ^{1}H NMR (600 MHz, D_2O): δ 4.32 (m, 1H), 4.08 (d, J = 4.6 Hz, 1H), 3.69 (s, 3H), 3.58 (s, 2H), 2.97 (d, J = 17.1 Hz, 1H), 2.93 (d, J = 17.5 Hz, 1H), 2.77 (dd, J = 17.5, 1.3 Hz, 1H), 2.71 (dd, J = 17.1, 1.4 Hz, 1H), 1.26 (d, J = 6.5 Hz, 3H). ^{13}C{1H} NMR (151 MHz, D_2O): δ 177.90, 163.8, 163.8, 127.7, 74.2, 70.9, 70.2, 67.4, 62.0, 38.6, 36.6, 22.2.

MysD Expression and Purification. The *mysD* gene was amplified from the isolated genomic DNA of *Nostoc linkia* NIES-25 and inserted into the *Nde*I/*Xho*I sites of pET28b, and the resultant construct pET28b-mysD was transformed into *E. coli* BL21-gold-(DE3) for the expression of recombinant N-His_6-tagged MysD. Protein expression was carried out in 500 mL of Luria−Bertani broth supplemented with 50 μg/mL of kanamycin (37 °C, 225 rpm). When the cell culture OD_{600} reached 0.5, IPTG (final concentration 0.1 mM) was added to the culture to induce gene expression (18 °C, 180 rpm, 20 h). The cells were harvested by centrifugation (6000 rpm, 20 min), and collected cell pellets were resuspended in the lysis buffer (25 mM Tris-Cl, pH 8.0, 100 mM NaCl, 1 mM β-mercaptoethanol, and 10 mM imidazole) and lysed by sonication on ice (10 s pulse and 20 s rest, 1 min in total). Following centrifugation (15000 rpm, 4 °C, 30 min), recombinant MysD was purified by the HisTrap Ni-NTA affinity column (GE Healthcare). N-His_6-tagged MysD was eluted using a 0−100% B gradient in 15 min at the flow rate of 2 mL/min, using A buffer (25 mM Tris-Cl, pH 8.0, 250 mM NaCl, 1 mM β-mercaptoethanol, and 30 mM imidazole) and B buffer (25 mM Tris-Cl, pH 8.0, 250 mM NaCl, 1 mM β-mercaptoethanol, and 300 mM imidazole). Fractions with recombinant MysD were collected, concentrated, and buffer-exchanged into storage buffer (50 mM Tris-Cl, pH 8.0, 10% glycerol). The purity of the recombinant protein was analyzed on SDS-PAGE, and the concentration was determined by its calculated extinction coefficient and absorbance at 280 nm as measured by NanoDrop.

MysD Reactions. MG was purified from extracts of *E. coli* expressing *MysAB2C* by HPLC and used as the substrate for the MysD reactions. The quantity of MG was calculated based on its reported extinction coefficient (28 100 M^{-1} cm^{-1}). The initial MysD reactions included MG (50 μM), L-Thr (1 mM), Mg^{2+} (10 mM), and ATP (1 mM) in 100 mM Tris-Cl, at pH 7.5. The reactions were initiated by adding MysD (0.5 μM) and then incubated at room temperature for 2 h. The control reactions omitted MysD or ATP. All reactions were quenched by heat inactivation at 95 °C for 10 min. After centrifugation at 20 000g for 15 min, the clear supernatants were collected for HPLC and LC-HRMS analysis. To determine the optimal reaction conditions, the MysD reaction was performed in 100 mM buffer with a pH of 6.5 to 11 at 16 to 60 °C for 6 min. To explore the substrate scope of MysD, all 20 natural amino acids (5 mM) were screened in the above reaction mixtures under the optimal conditions for 3 h. The reactions were terminated and then analyzed in the HPLC and/or LC-MS analysis. To determine the relative activity of six identified amino acids as MysD's substrates, we used a two-step strategy. First, the reactions were performed with 0.25 μM MysD for 8 min, which led to no more than 50% consumption of MG into porphyra-334 with the best substrate L-Thr and into shinorine with L-Ser. For the other four amino acids, the levels of their corresponding disubstituted MAAs were determined after the reaction time was elongated to 30 min. All reactions were performed in at least two independent replicates.

HPLC and LC-MS Analysis. Samples were analyzed on a Shimadzu Prominence UHPLC system (Kyoto, Japan) coupled with a PDA detector. The compounds were separated on a Phenomenex Luna C8 column (4.6 × 250 mm, 5 μm) using the following HPLC program: 2% B for 15 min, 2−90% B gradient in 2 min, 90% B for 2 min, 90−2% in 2 min, and re-equilibration in 2% B for 6 min. The A phase was 0.1 M triethylammonium acetate pH 7.0, and the B phase was methanol. The flow rate was set at 0.5 mL/min. In the

https://doi.org/10.1021/acs.joc.1c00368
J. Org. Chem. 2021, 86, 11160−11168

quantitative analysis of relative activity of MysD with different amino acid substrates, water-containing 0.1% formic acid was used as phase A to fully separate MG with MG-Ala. LC-HRMS and HRMS/MS experiments were conducted on a Thermo Scientific Q Exactive Focus mass spectrometer with a Dionex Ultimate RSLC 3000 uHPLC system, equipped with the H-ESI II probe on an Ion Max API Source. Methanol (B)/water (A) containing 0.1% formic acid were used as mobile phases, and the same LC program was used as in the HPLC analysis. The eluents from the first 3 min were diverted to waste by a diverting valve. MS1 signals were acquired under the Full MS positive ion mode, covering a mass range of m/z 150−2000, with resolution at 35 000 and AGC target at 1×10^6. Fragmentation was obtained with the parallel reaction monitoring (PRM) mode using an inclusion list of calculated parental ions. The AGC target was set at 5×10^4 for MS2. Precursor ions were selected in the quadrapole typically with an isolation width of 3.0 m/z and fragmented in the HCD cell at a collision energy (CE) of 30. For some ions, the isolation width was 2.0 m/z, and stepwise CEs of 15, 20, and 25 were used.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.joc.1c00368.

Bioinformatics analysis, MysC homologues, HR-MS/MS analysis, TIC, EIC, 1D and 2D NMR spectra, and primers (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Yousong Ding − *Department of Medicinal Chemistry, Center for Natural Products, Drug Discovery and Development (CNPD3), University of Florida, Gainesville, Florida 32610, United States;* orcid.org/0000-0001-8610-0659; Email: yding@cop.ufl.edu

Authors

Manyun Chen − *Department of Medicinal Chemistry, Center for Natural Products, Drug Discovery and Development (CNPD3), University of Florida, Gainesville, Florida 32610, United States*

Garret M. Rubin − *Department of Medicinal Chemistry, Center for Natural Products, Drug Discovery and Development (CNPD3), University of Florida, Gainesville, Florida 32610, United States*

Guangde Jiang − *Department of Medicinal Chemistry, Center for Natural Products, Drug Discovery and Development (CNPD3), University of Florida, Gainesville, Florida 32610, United States*

Zachary Raad − *Department of Medicinal Chemistry, Center for Natural Products, Drug Discovery and Development (CNPD3), University of Florida, Gainesville, Florida 32610, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.joc.1c00368

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by startup funds provided by the University of Florida (Y.D.) and NIH (R35 GM128742 to Y.D.). We thank Prof. Guangrong Zheng for discussions and Mr. Dake Liu and Mr. Jim Rocca for technical support. We thank the National Institute for Environmental Studies, Japan, for providing *Nostoc linkia* NIES-25. A portion of this work was performed in the McKnight Brain Institute at the National High Magnetic Field Laboratory's AMRIS Facility, which is supported by the National Science Foundation Cooperative Agreement No. DMR-1644779 and the State of Florida. This work was supported in part by an NIH award, S10RR031637, for the 600 MHz Avance III magnetic resonance instrumentation.

■ REFERENCES

(1) Rogers, H. W.; Weinstock, M. A.; Feldman, S. R.; Coldiron, B. M. Incidence estimate of nonmelanoma skin cancer (keratinocyte carcinomas) in the U.S. population, 2012. *JAMA Dermatol.* **2015**, *151* (10), 1081−1086.

(2) Siegel, R. L.; Miller, K. D.; Fuchs, H. E.; Jemal, A. Cancer statistics, 2021. *Ca-Cancer J. Clin.* **2021**, *71* (1), 7−33.

(3) Moan, J.; Grigalavicius, M.; Baturaite, Z.; Dahlback, A.; Juzeniene, A. The relationship between UV exposure and incidence of skin cancer. *Photodermatol., Photoimmunol. Photomed.* **2015**, *31* (1), 26−35.

(4) Armstrong, B. K.; Kricker, A. How much melanoma is caused by sun exposure. *Melanoma Res.* **1993**, *3* (6), 395−401.

(5) Holick, M. F. Biological effects of sunlight, ultraviolet radiation, visible light, infrared radiation and vitamin D for health. *Anticancer Res.* **2016**, *36* (3), 1345−1356.

(6) Ghiasvand, R.; Weiderpass, E.; Green, A. C.; Lund, E.; Veierod, M. B. Sunscreen use and subsequent melanoma risk: A population-based cohort study. *J. Clin. Oncol.* **2016**, *34* (33), 3976−3983.

(7) Latha, M. S.; Martis, J.; Shobha, V.; Sham Shinde, R.; Bangera, S.; Krishnankutty, B.; Bellary, S.; Varughese, S.; Rao, P.; Naveen Kumar, B. R. Sunscreening agents: a review. *J. Clin. Aesthet. Dermatol.* **2013**, *6* (1), 16−26.

(8) Krause, M.; Klit, A.; Blomberg Jensen, M.; Søeborg, T.; Frederiksen, H.; Schlumpf, M.; Lichtensteiger, W.; Skakkebaek, N. E.; Drzewiecki, K. T. Sunscreens: are they beneficial for health? An overview of endocrine disrupting properties of UV-filters. *Int. J. Androl.* **2012**, *35* (3), 424−436.

(9) Ruszkiewicz, J. A.; Pinkas, A.; Ferrer, B.; Peres, T. V.; Tsatsakis, A.; Aschner, M. Neurotoxic effect of active ingredients in sunscreen products, a contemporary review. *Toxicol. Rep.* **2017**, *4*, 245−259.

(10) Matta, M. K.; Zusterzeel, R.; Pilli, N. R.; Patel, V.; Volpe, D. A.; Florian, J.; Oh, L.; Bashaw, E.; Zineh, I.; Sanabria, C.; Kemp, S.; Godfrey, A.; Adah, S.; Coelho, S.; Wang, J.; Furlong, L. A.; Ganley, C.; Michele, T.; Strauss, D. G. Effect of sunscreen application under maximal use conditions on plasma concentration of sunscreen active ingredients a randomized clinical trial. *JAMA* **2019**, *321* (21), 2082−2091.

(11) Schneider, S. L.; Lim, H. W. Review of environmental effects of oxybenzone and other sunscreen active ingredients. *J. Am. Acad. Dermatol.* **2019**, *80* (1), 266−271.

(12) Pandika, M. Looking to nature for new sunscreens. *ACS Cent. Sci.* **2018**, *4* (7), 788−790.

(13) Saewan, N.; Jimtaisong, A. Natural products as photoprotection. *J. Cosmet. Dermatol.* **2015**, *14* (1), 47−63.

(14) Kageyama, H.; Waditee-Sirisattha, R. Antioxidative, anti-inflammatory, and anti-aging properties of mycosporine-like amino acids: Molecular and cellular mechanisms in the protection of skin-aging. *Mar. Drugs* **2019**, *17* (4), 222.

(15) Losantos, R.; Funes-Ardoiz, I.; Aguilera, J.; Herrera-Ceballos, E.; Garcia-Iriepa, C.; Campos, P. J.; Sampedro, D. Rational design and synthesis of efficient sunscreens to boost the solar protection factor. *Angew. Chem., Int. Ed.* **2017**, *56* (10), 2632−2635.

(16) Carreto, J. I.; Carignan, M. O. Mycosporine-like amino acids: Relevant secondary metabolites. Chemical and ecological aspects. *Mar. Drugs* **2011**, *9* (3), 387−446.

(17) Bandaranayake, W. Mycosporines: are they nature's sunscreens? *Nat. Prod. Rep.* **1998**, *15* (2), 159−172.

(18) Sinha, R. P.; Singh, S. P.; Hader, D. P. Database on mycosporines and mycosporine-like amino acids (MAAs) in fungi, cyanobacteria, macroalgae, phytoplankton and animals. *J. Photochem. Photobiol., B* **2007**, *89* (1), 29−35.
(19) Kicklighter, C. E.; Kamio, M.; Nguyen, L.; Germann, M. W.; Derby, C. D. Mycosporine-like amino acids are multifunctional molecules in sea hares and their marine community. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (28), 11494−11499.
(20) Nazifi, E.; Wada, N.; Yamaba, M.; Asano, T.; Nishiuchi, T.; Matsugo, S.; Sakamoto, T. Glycosylated porphyra-334 and palythine-threonine from the terrestrial cyanobacterium *Nostoc commune*. *Mar. Drugs* **2013**, *11* (9), 3124−3154.
(21) D'Agostino, P. M.; Javalkote, V. S.; Mazmouz, R.; Pickford, R.; Puranik, P. R.; Neilan, B. A. Comparative profiling and discovery of novel glycosylated mycosporine-like amino acids in two strains of the cyanobacterium *Scytonema cf. crispum*. *Appl. Environ. Microbiol.* **2016**, *82* (19), 5951−5959.
(22) Furusaki, A.; Matsumoto, T.; Tsujino, I.; Sekikawa, I. The crystal and molecular structure of palythine trihydrate. *Bull. Chem. Soc. Jpn.* **1980**, *53* (2), 319−323.
(23) Uemura, D.; Katayama, C.; Wada, A.; Hirata, Y. Crystal and molecule structure of palythiene possessing a novel 360 nm chromophore. *Chem. Lett.* **1980**, *9* (6), 755−756.
(24) Klisch, M.; Richter, P.; Puchta, R.; Häder, D.-P.; Bauer, W. The stereostructure of porphyra-334: An experimental and calculational NMR investigation. Evidence for an efficient 'proton sponge'. *Helv. Chim. Acta* **2007**, *90* (3), 488−511.
(25) White, J. D.; Cammack, J. H.; Sakuma, K.; Rewcastle, G. W.; Widener, R. K. Transformations of quinic acid. Asymmetric synthesis and absolute configuration of mycosporin I and mycosporin-gly. *J. Org. Chem.* **1995**, *60* (12), 3600−3611.
(26) Yang, G.; Cozad, M. A.; Holland, D. A.; Zhang, Y.; Luesch, H.; Ding, Y. Photosynthetic production of sunscreen shinorine using an engineered cyanobacterium. *ACS Synth. Biol.* **2018**, *7* (2), 664−671.
(27) Balskus, E. P.; Walsh, C. T. The genetic and molecular basis for sunscreen biosynthesis in cyanobacteria. *Science* **2010**, *329* (5999), 1653−1656.
(28) Pope, M. A.; Spence, E.; Seralvo, V.; Gacesa, R.; Heidelberger, S.; Weston, A. J.; Dunlap, W. C.; Shick, J. M.; Long, P. F. *O*-Methyltransferase is shared between the pentose phosphate and shikimate pathways and is essential for mycosporine-like amino acid biosynthesis in *Anabaena variabilis* ATCC 29413. *ChemBioChem* **2015**, *16* (2), 320−327.
(29) Osborn, A. R.; Almabruk, K. H.; Holzwarth, G.; Asamizu, S.; LaDu, J.; Kean, K. M.; Karplus, P. A.; Tanguay, R. L.; Bakalinsky, A. T.; Mahmud, T. *De novo* synthesis of a sunscreen compound in vertebrates. *eLife* **2015**, DOI: 10.7554/eLife.05919.
(30) Asamizu, S.; Xie, P.; Brumsted, C. J.; Flatt, P. M.; Mahmud, T. Evolutionary divergence of sedoheptulose 7-phosphate cyclases leads to several distinct cyclic products. *J. Am. Chem. Soc.* **2012**, *134* (29), 12219−29.
(31) Gao, Q.; Garcia-Pichel, F. An ATP-grasp ligase involved in the last biosynthetic step of the iminomycosporine shinorine in *Nostoc punctiforme* ATCC 29133. *J. Bacteriol.* **2011**, *193* (21), 5923−5928.
(32) 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.
(33) Challis, G. L. Genome mining for novel natural product discovery. *J. Med. Chem.* **2008**, *51* (9), 2618−2628.
(34) Suzek, B. E.; Wang, Y.; Huang, H.; McGarvey, P. B.; Wu, C. H.; UniProt, C. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* **2015**, *31* (6), 926−932.
(35) El-Gebali, S.; Mistry, J.; Bateman, A.; Eddy, S. R.; Luciani, A.; Potter, S. C.; Qureshi, M.; Richardson, L. J.; Salazar, G. A.; Smart, A.; Sonnhammer, E. L. L.; Hirsh, L.; Paladin, L.; Piovesan, D.; Tosatto, S. C. E.; Finn, R. D. The Pfam protein families database in 2019. *Nucleic Acids Res.* **2019**, *47* (D1), D427−D432.
(36) Franke, I.; Resch, A.; Dassler, T.; Maier, T.; Bock, A. YfiK from *Escherichia coli* promotes export of *O*-acetylserine and cysteine. *J. Bacteriol.* **2003**, *185* (4), 1161−1166.
(37) Miyamoto, K. T.; Komatsu, M.; Ikeda, H. Discovery of gene cluster for mycosporine-like amino acid biosynthesis from *Actinomycetales* microorganisms and production of a novel mycosporine-like amino acid by heterologous expression. *Appl. Environ. Microbiol.* **2014**, *80* (16), 5028−5036.
(38) Hu, C.; Voller, G.; Sussmuth, R.; Dittmann, E.; Kehr, J. C. Functional assessment of mycosporine-like amino acids in *Microcystis aeruginosa* strain PCC 7806. *Environ. Microbiol.* **2015**, *17* (5), 1548−1559.
(39) D'Agostino, P. M.; Woodhouse, J. N.; Liew, H. T.; Sehnal, L.; Pickford, R.; Wong, H. L.; Burns, B. P.; Neilan, B. A. Bioinformatic, phylogenetic and chemical analysis of the UV-absorbing compounds scytonemin and mycosporine-like amino acids from the microbial mat communities of Shark Bay, Australia. *Environ. Microbiol.* **2019**, *21* (2), 702−715.
(40) Hegg, E. L.; Que, L., Jr. The 2-His-1-carboxylate facial triad–an emerging structural motif in mononuclear non-heme iron(II) enzymes. *Eur. J. Biochem.* **1997**, *250* (3), 625−629.
(41) Mihalik, S. J.; Morrell, J. C.; Kim, D.; Sacksteder, K. A.; Watkins, P. A.; Gould, S. J. Identification of PAHX, a Refsum disease gene. *Nat. Genet.* **1997**, *17* (2), 185−189.
(42) Islam, M. S.; Leissing, T. M.; Chowdhury, R.; Hopkinson, R. J.; Schofield, C. J. 2-Oxoglutarate-dependent oxygenases. *Annu. Rev. Biochem.* **2018**, *87*, 585−620.
(43) Kavanagh, K. L.; Jornvall, H.; Persson, B.; Oppermann, U. Medium- and short-chain dehydrogenase/reductase gene and protein families: the SDR superfamily: functional and structural diversity within a family of metabolic and regulatory enzymes. *Cell. Mol. Life Sci.* **2008**, *65* (24), 3895−3906.
(44) Carignan, M. O.; Cardozo, K. H.; Oliveira-Silva, D.; Colepicolo, P.; Carreto, J. I. Palythine-threonine, a major novel mycosporine-like amino acid (MAA) isolated from the hermatypic coral *Pocillopora capitata*. *J. Photochem. Photobiol., B* **2009**, *94* (3), 191−200.
(45) Orfanoudaki, M.; Hartmann, A.; Ngoc, H. N.; Gelbrich, T.; West, J.; Karsten, U.; Ganzera, M. Mycosporine-like amino acids, brominated and sulphated phenols: Suitable chemotaxonomic markers for the reassessment of classification of *Bostrychia calliptera* (Ceramiales, Rhodophyta). *Phytochemistry* **2020**, *174*, 112344.
(46) Geraldes, V.; Jacinavicius, F. R.; Genuario, D. B.; Pinto, E. Identification and distribution of mycosporine-like amino acids in Brazilian cyanobacteria using ultrahigh-performance liquid chromatography with diode array detection coupled to quadrupole time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **2020**, *34*, No. e8634.
(47) Harvey, A. L.; Edrada-Ebel, R.; Quinn, R. J. The re-emergence of natural products for drug discovery in the genomics era. *Nat. Rev. Drug Discovery* **2015**, *14* (2), 111−29.
(48) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13* (11), 2498−504.
(49) Thompson, J. D.; Higgins, D. G.; Gibson, T. J. Clustal-W - Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **1994**, *22* (22), 4673−4680.

https://doi.org/10.1021/acs.joc.1c00368
J. Org. Chem. 2021, 86, 11160−11168