Comparative Profiling and Discovery of Novel Glycosylated Mycosporine-Like Amino Acids in Two Strains of the Cyanobacterium Scytonema cf. crispum

Paul M. D’Agostino,a,b Paul M. D’Agostino, Vivek S. Javalkote,a,c Rabia Mazmouz,a Russell Pickford,d Pravin R. Puranik,c Brett A. Neilana
School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, NSW, Australiaa; Department of Chemistry, Technische Universität München, Garching, Germanyb; School of Life Sciences, North Maharashtra University, Umavinagar, Jalgaon, Maharashtra, India; Bioanalytical Mass Spectrometry Facility, University of New South Wales, Sydney, NSW, Australiac

ABSTRACT

The mycosporine-like amino acids (MAAs) are a group of small molecules with a diverse ecological distribution among microorganisms. MAAs have a range of physiological functions, including protection against UV radiation, making them important from a biotechnological perspective. In the present study, we identified a putative MAA (mys) gene cluster in two New Zealand isolates of Scytonema cf. crispum (UCFS10 and UCFS15). Homology to “Anabaena-type” mys clusters suggested that this cluster was likely to be involved in shinorine biosynthesis. Surprisingly, high-performance liquid chromatography analysis of S. cf. crispum cell extracts revealed a complex MAA profile, including shinorine, palythine-serine, and their hexose-bound variants. It was hypothesized that a short-chain dehydrogenase (UCFS15_00405) encoded by a gene adjacent to the S. cf. crispum mys cluster was responsible for the conversion of shinorine to palythine-serine. Heterologous expression of MysABCE and UCFS15_00405 in Escherichia coli resulted in the exclusive production of the parent compound shinorine. Taken together, these results suggest that shinorine biosynthesis in S. cf. crispum proceeds via an Anabaena-type mechanism and that the genes responsible for the production of other MAA analogues, including palythine-serine and glycosylated analogues, may be located elsewhere in the genome.

IMPORTANCE

Recently, New Zealand isolates of S. cf. crispum were linked to the production of paralytic shellfish toxins for the first time, but no other natural products from this species have been reported. Thus, the species was screened for important natural product biosynthesis. The mycosporine-like amino acids (MAAs) are among the strongest absorbers of UV radiation produced in nature. The identification of novel MAAs is important from a biotechnological perspective, as these molecules are able to be utilized as sunscreens. This study has identified two novel MAAs that have provided several new avenues of future research related to MAA genetics and biosynthesis. Further, we have revealed that the genetic basis of MAA biosynthesis may not be clustered on the genome. The identification of the genes responsible for MAA biosynthesis is vital for future genetic engineering.

Cyanobacteria are photosynthetic prokaryotes with a global distribution ranging from meltwater ponds to hot springs, from fresh to hypersaline waters, and from hot to cold deserts. Scytonema cf. crispum UCFS10 (= S. cf. crispum CAWBG524) and S. cf. crispum UCFS15 (= S. cf. crispum CAWBG72) were recently isolated in New Zealand and are notable for their distinct biosynthesis of paralytic shellfish toxins (1, 2). Both species are filamentous and form blackish-green mats that can survive severe exposure to sunlight (3).

Depletion of ozone in the atmosphere has occurred because of anthropogenic activities, resulting in an increase in UV radiation (UVR) on the earth’s surface. It is reported that a 1% decrease in the ozone layer causes a 1.3 to 1.8% increase in UVB penetration (4). UVR has adverse biological effects because of its photochemical absorption by biologically significant molecules such as proteins and nucleic acids (5). Absorption of UVR by these biomolecules can lead to the formation of photoproducts similar to cyclobutyl-type dimers, pyrimidine adducts, so-called “spore photoproducts,” pyrimidine hydrates, and DNA protein cross-links, resulting in cell dysfunction, damage, and ultimately cell death (6). In humans, overexposure to UVR can result in sunburn and skin cancer. Therefore, the identification of naturally derived UVR-absorbing molecules is highly significant from a biotechnological viewpoint (7).

The production of sunscreen molecules by certain species of cyanobacteria is a well-described response to UVR (8). These compounds include both the yellow lipid-soluble dimeric pigment scytonemin and the water-soluble MAAs, which have been shown to interact with UVR (9–11). The MAAs are a large family of secondary metabolites and have been identified in a diverse
range of organisms, including fungi, eukaryotic algae, and cyanobacteria (12, 13). Structurally, MAAs contain an amino-cyclohexenone ring or an amino-cyclohexenimine ring and have absorbance maxima ($\lambda_{\text{max}}$) ranging from 310 to 365 nm. Amino-cyclohexenimine MAAs possess a basic cyclohexenimine ring attached to the amino acid glycine at the C-3 position and another amino acid, an amino alcohol, or an enaminone system attached at the C-1 position (Fig. 1). Structural diversity among MAAs is achieved via the attachment of different amino acids to the core, which can then be further modified by decarboxylation or demethylation, resulting in alteration of the $\lambda_{\text{max}}$ (14).

MAAs have a high molar extinction coefficient ($\varepsilon \approx 28,000$ to 50,000 liters · mol$^{-1}$ · cm$^{-1}$) in the UVA and UVB regions and protect cyanobacteria by absorbing UVR energy, which is subsequently released in the form of heat, without generating reactive oxygen species (10, 14, 15). Apart from protection against UVR, MAAs are involved in different processes such as antioxidant activity, osmotic regulation, and protection against thermal stress and desiccation (16). They can also act as accessory photosynthetic pigments and intracellular nitrogen stores (16). Glycosylated MAAs found in cyanobacteria are believed to be associated with the cyanobacterial polysaccharide sheath (17).

Initially, it was believed that the aromatic amino acid-synthesizing shikimate pathway was associated with MAA biosynthesis and that 3-dehydroquininate (DHQ) was converted to 4-deoxygudasol (4-DG) prior to the formation different MAAs (18). However, heterologous expression of the MAA biosynthetic (mys) gene cluster by Balskus et al. (19) and targeted pathway inhibition and gene knockouts by Pope et al. (20) have revealed that both the pentose phosphate and shikimate pathways are involved in MAA biosynthesis, with the O-methyltransferase (OMT; MysB) being essential. MAA biosynthesis begins with a dehydroquininate synthase (DHQS; MysA) homolog and MysB, which produce the MAA precursor 4-DG from sedoheptulose 7-phosphate (SH 7-P) (19). Next, MysC, an enzyme belonging to the ATP-grasp superfamily, attaches glycine to 4-DG, forming mycosporine-glycine. The three genes (mysABC) present in both *Anabaena variabilis* ATCC 29413 (Ava_3856 to Ava_3858) and *Nostoc punctiforme* ATCC 29133 (NpR5598 to NpR5600) are thought to be essential for mycosporine-glycine biosynthesis, while additional genes are required to produce a range of MAA analogues (21). The final step involved in the biosynthesis of the common MAA shionorine can occur via two distinct enzymes, which catalyze functionally identical reactions via the addition of serine to mycosporine-glycine. One mechanism involves a nonribosomal peptide synthetase (NRPS) (Ava_3855) encoded in what is known as an “*Anabaena*-type” cluster (19). Alternatively, a *d*-Ala-*d*-Ala ligase (NpF5597; MysD) may add serine to mycosporine-glycine to form shionorine, as occurs in *N. punctiforme* ATCC 29133 and *Aphanotoche halophytica*. The genes involved in this pathway are located within so-called “Nostoc-type” *mys* clusters (21–23). However, the genes responsible for the formation of many other MAA analogues have yet to be identified.

The present study was initiated to explore the MAA chemical diversity and biosynthesis in *S. cf. crispum* strains UCFS10 and UCFS15. Genome mining identified a putative *Anabaena*-type *mys* gene cluster with high homology to the *mys* cluster of *A. variabilis* ATCC 29413. While preliminary bioinformatic predictions suggested that the *S. cf. crispum mys* cluster is involved in shionorine biosynthesis, the discovery of an adjacent short-chain dehydrogenase gene (*UCFS15_00405*) led us to hypothesize that other MAA analogues (palystine-serine, in particular) may be produced via modification of shionorine. The identification of novel MAAs and their associated genes is an important step forward in genetic engineering and the development of next-generation sunscreens.

**MATERIALS AND METHODS**

**Cyanobacterial cultures.** The cyanobacterial strains *S. cf. crispum* UCFS10 and UCFS15 were isolated from two different freshwater lakes in New Zealand (2) and are available from the Cawthron Institute Culture Collection of Microalgae (24). Both strains were cultivated in BG11 medium at 25°C under white fluorescent light at 50 μmol · m$^{-2}$ · s$^{-1}$ with a 12-h–12-h light–dark cycle.

**MAA extraction and partial purification.** Cyanobacterial or *Escherichia coli* biomass was washed with 1 volume of 20 mM NaH$_2$PO$_4$ and centrifuged at 5,000 × g for 15 min at 4°C. Cyanobacterial cells were resuspended in 100% high-performance liquid chromatography (HPLC)-grade methanol and stored overnight at 4°C. *E. coli* cells were resuspended in HPLC-grade methanol and lyzed on ice by sonication with five 20-s pulses at 40% amplitude with 1 min of recovery time between pulses (Branson Digital Sonifier M450, 3-mm probe). Lysed cellular debris was removed by centrifugation (5,000 × g for 15 min at 4°C). Cyanobacterial and *E. coli* methanolic extracts were desiccated with a rotary vacuum evaporator at 35°C (Buchi rotavapor R-210; Switzerland). Desic-

---

**FIG 1** Structural diversity of MAAs. The structures of 4-deoxygudasol, shionorine, and palystine-serine are shown. Amino acid linkages to C-1 and C-3 of 4-deoxygudasol and alterations in functional groups provide structural diversity for MAAs. For example, glycine and serine bound to the MAA core results in the formation of shionorine. Decarboxylation and demethylation are proposed to result in the formation of palystine-serine from shionorine.
cated extracts were dissolved in 1 mL of Milli-Q water and mixed vigorously with 500 µl of chloroform to remove pigments. Thereafter, the aqueous phase was carefully aspirated and filtered through a Millipore syringe-driven 0.22-µm PES membrane filter (Millipore, USA) to obtain partially purified MAAs. Shinorine and porphyra-334 were extracted from Helioguard 365 (Mibelle Biochemistry, Switzerland) and used as analytical standards (see the supplemental material).

**HPLC analysis.** Aqueous partially purified MAA extracts were analyzed by HPLC on an HP Agilent Series 1100 apparatus equipped with continuous degassing system J7P3006883, binary pump DE72001480, autosampler ALS DE72003901, column oven Colcomp DE72003, and photodiode array detector DAD DE7200619. A 50-µl sample was injected into a Sphereclone 5-µm ODS (2) C18 column (250-mm length by 4.60-mm inside diameter; Phenomenex) protected with a guard column of similar material (20-mm length by 4.60-mm inside diameter). A wave-length of 330 nm was used for MAA detection, and a photodiode array detector (PDA) UV spectrum of 200 to 400 nm was obtained for each peak. HPLC chromatographic separation was adapted from the method developed by Ingalls et al. (25). Buffer A was Milli-Q water, buffer B was methanol, and the flow rate was 1 ml/min. The gradient employed for separation was 0% B at the start of the run, 20% B at 7 min, 50% B at 9 min, 80% B at 12 min, 100% B at 13 min, where the gradient was held for 10 min, followed by a 10-min equilibration at 0% B prior to the next injection.

**RPLC-MS analysis.** Reverse-phase liquid chromatography-mass spectrometry (RPLC-MS) was performed on an LCQ DECA XP Plus ion trap mass spectrometer (Thermo Fisher Scientific) interfaced with a Surveyor HPLC, autosampler, and PDA detector via an electrospray interface operating in the positive-ion mode. Chromatography was performed on a Phenomenex Luna C18 column (2.1 by 150 mm). Solvent A was Milli-Q water with 0.1% formic acid, and solvent B was methanol. Solvent was delivered at 300 µl/min with a gradient program identical to that used for HPLC analysis. Full-scan mass spectra were acquired over a mass range of m/z 200 to 800 with an automated, data-dependent tandem scan performing MS (tandem mass spectrometry [MS²]) of the most intense ions in the full scan. An exclusion list was used to ignore high-intensity background ions.

**HILIC-MS² analysis.** MAAs were further analyzed by hydrophilic interaction liquid chromatography (HILIC) with the same instrumentation as RPLC-MS, including an autosampler, an LC pump, a Surveyor PDA detector, and an LCQ Deca XP Plus ion trap mass spectrometer (Thermo Fisher Scientific). A Zic-HILIC 3.5-µm (100 by 2.1 mm); Merck) column protected with a matching guard column was used for separation. The PDA scan was performed from 190 to 400 nm with 330 nm being monitored for the analytes. Mobile phase A was 5 mM ammonium formate and 0.1% formic acid in water, and mobile phase B was 0.01% formic acid in acetonitrile. The system was run at a flow rate of 200 µl/min. The gradient employed for separation was 95% B at the beginning of the run and held for 5 min, 5% B at 22 min and held for a further 2 min, and 95% B at 25 min and held for a further 15 min. Ions were generated by positive electrospray and monitored over the mass range m/z 200 to 800 by automated, data-dependent MS² of the most intense ions in the full scan. An exclusion list was used to ignore high-intensity background ions. MS² files were analyzed with the Thermo Xcalibur software package (version 2.2 SP1.AB). MS² and in silico fragmentation predictions were obtained with the Thermo Mass Frontier package (version 7.0.5.9 SR3).

**DNA isolation, PCR amplification, and cloning.** Genomic DNA isolation from S. cf. crispum UCFS10 and UCFS15 was performed by the phenol-chloroform extraction method (26). PCR amplification of DNA was performed with a 50-µl reaction mixture containing 1X Velocity Hi-Fi buffer (Bioline, Australia), 1,000 µM deoxynucleoside triphosphates (Astral Scientific, Caringbah, Australia), 0.4 µM each gene-specific primer (Table 1, Integrated DNA Technologies, Singapore), 32 mM/µl Velocity DNA polymerase (Bioline, Australia), and sterile Milli-Q water. Thermal cycling was performed in a Bio-Rad 96-well iCycler (Bio-Rad, Hercules, CA) and began with an initial denaturation cycle of 98°C for 2 min, followed by 30 cycles of DNA denaturation at 98°C for 30 s, primer annealing at 55°C for 30 s, DNA amplification at 72°C for 30 s, and a 10 skb amplified, and a final extension at 72°C for 10 min.

Two expression constructs were generated. The first consisted of pET28::mysA/ABCE containing the mys cluster putatively responsible for shinorine biosynthesis, and the second construct, pET15b::UCFS15_00405, contained the coclustered short-chain dehydrogenase (UCFS15_00405; see the supplemental material). Primers incorporating restriction sites (NcoI/SalI) were designed to amplify the genes involved in the cluster, with subsequent ligation into the pET28b vector (Table 1). Primers incorporating restriction sites (NcoI/Xhol) were designed to amplify UCFS15_00405, with subsequent ligation into the pET15b vector. E. coli GB2005 competent cells were used for standard subcloning. E. coli BL21(DE3) cells (Novagen) were transformed with the PRARE plasmid used for heterologous expression. The possible expression of incompatible plasmids over short times has been reported in the past (27, 28). Plasmid compatibility for the duration of culture expression was confirmed by the identification of soluble protein products expressed from both coexpressed plasmids (see the supplemental material).

**Heterologous expression of the S. cf. crispum UCFS15 mys gene cluster and coclustered short-chain dehydrogenase (UCFS15_00405) in E. coli.** E. coli colonies freshly transformed with pET28::mysA/ABCE and pET15::empty or pET28::mysA/ABCE and pET15b::UCFS15_00405 were used to generate a 50-ml pre-expression culture grown in LB medium and incubated overnight at 37°C with shaking at 200 rpm. These pre-expression cultures were used to inoculate expression cultures (1% [vol/vol]) into 200 ml of LB medium. Expression cultures were supplemented with 50 µg/ml kanamycin and 100 µg/ml ampicillin to maintain the selection of both the pET28b and pET15 plasmids. Expression cultures were incubated at 30°C with shaking (200 rpm) until an optical density at 600 nm of 0.6 to 0.8 was reached. Cultures were then cooled on ice (to ~18°C), induced with isopropyl-β-D-thiogalactopyranoside (IPTG; final concentration, 500 µM), and incubated for 16 h at 18°C with shaking (200 rpm). To ensure successful expression of all cloned proteins, proteins were extracted from expression cultures, excised from an SDS-PAGE gel, and submitted for trypsinolysis and identification (see the supplemental material). MAAs were extracted from expression cultures with methanol as previously described.

**Accession number(s).** The nucleotide sequences of the UCFS10 and UCFS15 mys clusters are publicly available and have been submitted to GenBank under accession numbers KX021865 and KX021866, respectively.
RESULTS

Bioinformatic analysis. The S. cf. crispum UCFS10 and UCFS15 genomes were mined for the presence of mys clusters, revealing four genes with homology to mysA (DHQ), mysB (O-MT), mysC (ATP-Grasp), and the Anabaena variabilis gene cluster (ATP-Grasp). In silico analysis of the mysE adenylation domain with Antismash V3.0 (29) suggested that serine is likely to be incorporated at the substrate binding site during shinorine biosynthesis. A putative short-chain dehydrogenase (UCFS10_04340/UCFS15_00405) was discovered at the 3’ end of the mys cluster (in reverse orientation) in both strains (Fig. 2). This gene was most similar to a short-chain dehydrogenase (WP_039738519) from the cyanobacterium Hassallia byssoidea. The mys gene clusters from UCFS10 and UCFS15 had an overall nucleotide sequence identity of 99.9% (Table 2). Genomic screening of the UCFS10 and UCFS15 genomes revealed 84 and 85 putative glycosyltransferases not involved in cell wall biosynthesis, respectively. The flanking regions of all putative glycosyltransferase genes could not identify further mys-like genes.

Comparison of UCFS10 and UCFS15 MAA profiles. Partially purified MAAs from S. cf. crispum UCFS10 and UCFS15 were analyzed by HPLC and detected at an absorbance maximum (λmax) at 330 nm (Fig. 3). The MAA profile and the relative amounts of MAAs produced by each strain were consistent across multiple extractions and HPLC runs. HPLC analysis of UCFS10 revealed three major peaks at retention times of 2.24 min (I), 3.51 min (II), and 4.03 min (III). Similarly, UCFS15 showed four prominent peaks at retention times of 2.24 min (IV), 3.22 min (V), 3.51 min (VI), and 4.03 min (VII). The shinorine and porphyra-334 analytical standards had retention times of 3.46 and 3.57 min, respectively. These results suggest that both strains produce multiple MAA analogues. The most abundant peak associated with UCFS10 was peak III, followed by peak I, which displayed half the intensity. The most abundant peak associated with UCFS15 was peak V, followed by equal proportions of peaks IV and VI. UV spectrum analysis revealed that peak I and peak IV had a UV λmax of 330 nm while peaks II, III, V, VI, and VII had a UV λmax of 320 nm. To identify the molecule corresponding to each peak in the S. crispum MAA extract, each peak was purified by fractionation. Unfortunately, the quantity and purity of collected fractions were too low for structural elucidation by nuclear magnetic resonance. Thus, further structural analysis was conducted by high-resolution HILIC-MS2.

Analysis of fractionated MAAs by high-resolution HILIC-MS2. To identify the ions responsible for UV absorption, peaks I to VII were fractionated and analyzed by high-resolution HILIC-MS2 (Table 3; see the supplemental material). HILIC is a chromatographic method designed to separate polar (hydrophilic) molecules and has recently been used for the analysis of MAAs (13). HILIC-MS of peak I resulted in two ions at 21 min (m/z 495.1828 [M + H]+) and 22.22 min (m/z 495.1827 [M + H]+), designated peak Ia and peak Ib. MS2 of peak Ia and peak Ib m/z 495 precursor ions resulted in a major fragment with a m/z of 333.1296. The calculated mass difference between 495.1828 Da and 333.1296 Da is 162 Da, whereas the difference between 495.1827 Da and 333.1296 Da is 163 Da. This mass difference corresponds to the calculated mass of a hexose molecule (C6H10O5), similar to previous reports of glycosylated MAAs, including mycosporine-glutaminol-glucoside, mycosporine-glutaminol-glucoside, hexose-bound palythine-threonine, and penrose-bound shinorine (30–32). Further analysis of the peak Ia and peak Ib MS2 data showed a mass fingerprint with high similarity to shinorine, which has a m/z of 333 [M + H]+. Thus, it is proposed that the MAAs isolated in peak I represent a C6H10O5 (m/z 495) addition to shinorine and that two isomers may be present (on the basis of differing retention times but with identical mass and MS2 fingerprints). Thus, peak Ia and peak Ib are proposed to correspond to different isomers of a glycosylated shinorine analogue termed hexose-shinorine. HILIC-MS2 analysis of the purified peak IV fraction identified an ion with a m/z of 495.1830 [M + H]+ displaying a single retention time of 22 min, with an identical MS2 fragmentation pattern as peak I, suggesting that it also corresponds to shinorine with an attached hexose group. Peaks II and VI consisted of a major ion with a m/z of 275.1238 [M + H]+.

Table 2. Bioinformatic analysis of the UCFS10 and UCFS15 mys gene cluster

<table>
<thead>
<tr>
<th>Gene(s)</th>
<th>Enzyme family</th>
<th>Closest homologa</th>
<th>Name</th>
<th>Accession no.</th>
<th>% similarity</th>
<th>% nucleotide sequence similaritya</th>
</tr>
</thead>
<tbody>
<tr>
<td>mysA</td>
<td>3DHQ</td>
<td>Calothrix parietina</td>
<td>WP_015199132</td>
<td>99</td>
<td>99.9</td>
<td></td>
</tr>
<tr>
<td>mysB</td>
<td>MT</td>
<td>Calothrix parietina</td>
<td>WP_015199131</td>
<td>93</td>
<td>99.9</td>
<td></td>
</tr>
<tr>
<td>mysC</td>
<td>ATP-Grasp</td>
<td>Calothrix parietina</td>
<td>WP_015199130</td>
<td>99</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>mysE</td>
<td>NRPS</td>
<td>Chlorogloeopsis fritschi</td>
<td>WP_016876762</td>
<td>99</td>
<td>99.8</td>
<td></td>
</tr>
<tr>
<td>UCFS10_04340, UCFS15_00405</td>
<td>Short-chain dehydrogenase</td>
<td>Hassallia byssoidea</td>
<td>WP_039738519</td>
<td>92</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

a Based on inferred peptide sequence similarity.
b Between UCFS10 and UCFS15.
FIG 3 Comparative HPLC chromatogram profiles of partially purified MAAs from *S. cf. crispum*. Shown are UCFS10 methanolic extracts (A) with peaks at 2.24 min (I), 3.51 min (II), and 4.03 min (III); UCFS15 methanolic extracts (B) with peaks at 2.24 min (IV), 3.22 min (V), 3.51 min (VI), and 4.03 (VII); and MAA standards (C) extracted from Helioguard 365 with peaks at 3.46 and 3.57 min. Inset graphs depict the UV absorption spectra of respective peaks from chromatograms and are represented by data labels of peak retention time and λ_{max}. Inset i shows absorption maxima at 333 nm (retention time, 2.24 min), 320 nm (retention time, 3.51 min), and 320 nm (retention time, 4.03 min). Inset ii shows absorption maxima at 333 nm (retention time, 2.20 min), 320 nm (retention time, 3.22 min), and 320 nm (retention time, 3.50 min). Inset iii represents an absorption maximum at 333 nm for both peaks with retention times of 3.46 min and 3.57 min of standard MAAs shinorine and porphyra-334, respectively. DAD, diode array detector; mAU, milli-absorbance units.
Tandem mass spectrometry of the m/z 275 ion of peaks II and VI yielded several fragment ions, which were previously identified in the MS and MS² spectra of palylthine-serine (33). Taken together, the UV absorbance maximum of 320 nm and the MS² data from the literature suggest that peaks II and VI correspond to palylthine-serine.

Peaks III, V, and VII contained an ion with m/z 437.2000 [M + H]⁺, 437.1767 [M + H]⁺, and 437.1766 [M + H]⁺, respectively. Peaks III, V, and VII displayed highly similar fragmentation patterns upon MS². Fragmentation of the m/z 437 ions resulted in a major fragment ion with m/z 275, indicating an m/z loss of 162. Calculation of the mass difference between the parent ion (m/z 437) and fragment ion (m/z 275) indicated a loss of the C₆H₁₀O₅ glucoside group, as suggested for peaks I and IV. Other fragmentation ions of peaks III, V, and VII were similar to palylthine-serine, indicating a glycosylated analogue of palylthine-serine, designated hexose-palythine-serine. While the high-resolution HILIC-MS² suggests that a glucosyl group similar to mycosporine-glutaminol-glucoside and mycosporine-glutamicol-glucoside is bound to shinorine, UV-LC-MS results indicated a loss of the C₆H₁₀O₅ glucoside group in shinorine, indicating an hexose-palythine-serine, designated hexose-palythine-serine.

**TABLE 3 Summary of HILIC-MS² fragment ions obtained from HPLC fractions**

<table>
<thead>
<tr>
<th>HPLC fraction</th>
<th>HILIC-MS RTᵃ</th>
<th>Parent ion</th>
<th>Theoretical</th>
<th>Fragment ionsᵇ</th>
<th>Proposed MAA</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ia</td>
<td>21.00</td>
<td>495.1828</td>
<td>495.1826</td>
<td>480, 449, 436, 392, 347, 333, 318, 315, 303, 300, 287, 274, 265, 256, 230, 214, 186</td>
<td>Hexose-shinorine</td>
<td>This study</td>
</tr>
<tr>
<td>Ib</td>
<td>22.22</td>
<td>495.1828</td>
<td>495.1826</td>
<td>480, 445, 436, 392, 375, 333, 318, 315, 303, 300, 287, 274, 265, 256, 230, 214, 186</td>
<td>Hexose-shinorine</td>
<td>This study</td>
</tr>
<tr>
<td>IV</td>
<td>22.00</td>
<td>495.1828</td>
<td>495.1826</td>
<td>480, 465, 436, 392, 333, 318, 315, 303, 300, 287, 274, 265, 245, 230, 188</td>
<td>Hexose-shinorine</td>
<td>This study</td>
</tr>
<tr>
<td>V</td>
<td>20.96</td>
<td>437.1767</td>
<td>437.1771</td>
<td>422, 407, 391, 378, 275, 257, 245, 239, 207</td>
<td>Hexose-palythine-serine</td>
<td>This study</td>
</tr>
<tr>
<td>VII</td>
<td>20.86</td>
<td>437.1766</td>
<td>437.1771</td>
<td>391, 275, 260, 257, 245, 242, 239, 229, 216, 213, 207</td>
<td>Palylthine-serine</td>
<td>This study</td>
</tr>
<tr>
<td>UCFS10</td>
<td>333ᵇ</td>
<td>333.1298</td>
<td>333.1298</td>
<td>317, 303, 300, 274, 265, 230, 211, 196, 186</td>
<td>Shinorine</td>
<td>46</td>
</tr>
<tr>
<td>UCFS15</td>
<td>333ᵇ</td>
<td>333.1298</td>
<td>333.1298</td>
<td>318, 303, 300, 274, 265, 230, 212, 186</td>
<td>Shinorine</td>
<td>46</td>
</tr>
<tr>
<td>Shinorine STD</td>
<td>333ᵇ</td>
<td>333.1298</td>
<td>333.1298</td>
<td>318, 303, 300, 286, 274, 265, 230, 211, 186</td>
<td>Shinorine</td>
<td>46</td>
</tr>
</tbody>
</table>

ᵃ RT, retention time (minutes).
ᵇ In bold are fragment ions that match the analytical standard or fragmentation fingerprint presented in the literature.
ᶜ Identified by LC-MS in methanolic extracts of UCFS10 and UCFS15 but not isolated by fractionation.

**DISCUSSION**

*S. crispum* is a toxic filamentous cyanobacterium known to form heavily pigmented, UV-resistant mats (2, 3). Here we investigated the biosynthesis of MAA “sunscreens” in two New Zealand *S. crispum* isolates, UCFS10 and UCFS15. Bioinformatic analysis of the UCFS10 and UCFS15 genomes identified an *Anabaena*-type *mys* gene cluster (*mysABC*) and a clustered NRPS (*mysE*) in both strains, suggesting a genetic capacity for shinorine production. Homologous *mys* gene clusters have been identified in a variety of organisms, including, fungi, eukaryotic algae, and cyanobacteria (13, 14). Previous studies (19, 22, 23, 34, 35) have suggested that shinorine biosynthesis proceeds via a DHQS (encoded by *mysA*) and an O-MT (encoded by *mysB*), leading to the synthesis of 4-DG (20). 4-DG is then putatively linked to glycine by an ATP-grasp enzyme (encoded by *mysC*), resulting in the production of mycosporine-glycine. In the final step of shinorine biosynthesis, serine is attached to mycosporine-glycine by either an NRPS or a D-Ala–D-Ala ligase.

In addition to shinorine, a variety of other MAAs may be produced via the *mys* pathway. Different analogues are thought to arise via the addition of variable amino acid moieties onto the MAA core during the final step of biosynthesis (Fig. 1). For example, heterologous expression of the Nostoc-type *mysA*-to-*mysD*
cluster from *Actinosynema mirum* DSM 43827 resulted in the biosynthesis of three MAAs: the parent molecule shinorine, porphyra-334, and mycosporine-glycine-alanine (35). In this organism, biosynthesis of porphyra-334 and mycosporine-glycine-alanine occurred via the replacement of serine with threonine and alanine, respectively, by a promiscuous D-Ala–D-Ala ligase (35).

In the present study, HPLC and LC-MS analyses revealed the presence of multiple MAA analogues in the *S. crispum* study strains, including polythylene-serine, hexose-polythylene-serine, shinorine (a minor component), and hexose-shinorine. The mass spectral data also suggested that these MAAs were present in different ratios in each strain. However, heterologous expression of the *mys* cluster in *E. coli* resulted in the production of only two MAA analogues (at detectable levels): shinorine and its precursor mycosporine-glycine. These results are in agreement with previous heterologous-expression studies involving *Anabaena*-type *mys* clusters (*mysABCE*), which reported shinorine as the final product (19). These results also concur with the bioinformatic data available in the literature, which suggest that *mysE* NRPPs from group V cyanobacteria are predicted to preferentially bind serine (36). Thus, it is likely that substrate specificity is strict within *Anabaena*-type *mys* clusters, including the *S. crispum* cluster characterized in this study.

Coral species such as *Pocillopora capitata* and *Stylophora pistillata* and the dinoflagellates *Pseudo-nitzschia multiseries*, *Emiliania huxleyi*, and *Alexandrium catenella* have been shown to contain the MAAs shinorine, mycosporine-methylamine-serine, and polythylene-serine (33, 37, 38). On the basis of structure and the distribution of MAAs in corals, the biosynthesis of polythylene-serine is proposed to occur via the decarboxylation of the serine group on shinorine, resulting in the production mycosporine-methylamine-serine. This intermediate is then demethylated to form polythylene-serine (14). However, to date, there are no gene candidates for these tailoring reactions, nor is there any biochemical evidence to support this proposed pathway. However, it is important to note that polythylene-methylamine-serine was not detected in either UCFS10 or UCFS15 and polythylene-serine might actually be produced via an alternate pathway that bypasses polythylene-methylamine-serine as an intermediate.

Analysis of the regions flanking the *mys* cluster in UCFS10 and UCFS15 revealed the presence of a putative short-chain dehydrogenase gene (*UCFS15_00405*) transcribed in the reverse orientation to *mysABCE*. Enzymes belonging to the short-chain dehydrogenase family are known to catalyze a wide variety of reactions, including, but not limited to, decarboxylation, isomerization, epimerization, C=N bond reduction, dehydration, and carbonyl-alcohol oxidoreduction (39, 40). It was hypothesized that UCFS15_00405 may be responsible for the conversion of shinorine to polythylene-serine in *S. crispum*. Homologous dehydrogenase genes have also been observed adjacent to *mys* clusters in other cyanobacteria. For example, prephanate/arogenate dehydrogenases are encoded alongside the *mys* clusters of *Hapalosiophon welwitschiae* UH strain IC-52-3 and *Westiella intricata* UH strain HT-29-1 (36). In *W. intricata* UH strain HT-29-1, the dehydrogenase gene is cotranscribed with the *mys* cluster and the entire operon is induced by UV light (M. L. Micalef, personal communication). However, in the present study, heterologous expression of *S. crispum* *MysABCE* with the dehydrogenase did not result in the production of additional MAA analogues. In fact, UV absorbance spectra and LC-MS profiles were identical for both heterologous expression strains, regardless of the presence or absence of UCFS15_00405.

A possible explanation for these results is that UCFS15_00405 is not active in *E. coli*. Alternatively, this enzyme may have a role outside MAA metabolism. *In vitro* characterization of the purified overexpressed dehydrogenase may shed further light on this subject. The fact that multiple MAAs are present in both *S. cf. crispum* study strains also raises the possibility that additional as-yet-unidentified shinorine tailoring genes are involved in MAA metabolism in this cyanobacterium, including those for polythylene-serine biosynthesis and MAA glycosylation. Genomic screening of both genomes revealed 84 and 85 putative glycosyltransferases in UCFS10 and UCFS15, respectively. Flanking regions of all glycosyltransferase genes did not reveal any *mys*-like coding regions. Thus, the *mys* cluster expressed in this study is the only identifiable genetic locus responsible for MAA biosynthesis and the genes that function to alter shinorine to further MAA analogues are likely to be located elsewhere in the genome.

In this study, we identified two novel glycosylated MAAs, hexose-shinorine and hexose-polythylene-serine, that were present in at least two isoforms, as evidenced by their distinct HPLC and HILIC retention times and identical masses. Surprisingly, a glycosyltransferase gene could not be found in the immediate vicinity of the *mys* cluster in either *S. crispum* strain. Glycosylated MAA derivatives are produced by a variety of organisms, including fungi and cyanobacteria. For example, mycosporine-glutaminol-glucoside is produced by the fungal species *Rhotorula minuta*, *R. sloofiae*, and *Xanthophyllozyma dendrorhous* (*Phaffia rhodozyma*), while mycosporine-glutamicol-glucoside is produced by the fungi *Sarcinomyces petricola* A95, *Coniosporium* sp. strain A28, *Botryosphaeria*-like AN13, and *Coniosporium*-like A148, as well as the rock-inhabiting cyanobacterium *Leptolyngbya foveolatum* (32, 41–43). In addition to its role in UVR protection, mycosporine-glutaminol-glucoside is believed to protect cells from antioxidant damage (43, 44), which may account for its production by microorganisms exposed to low levels of UVR.

Glycosylated MAAs, including 7-O-(β-arabinopyranosyl)-porphyra-334, hexose-bound porphyra-334, and pentose-bound shinorine, are also produced by the terrestrial cyanobacterium *Nostoc commune* (17, 30, 31). The production of glycosylated and hybrid MAAs with different absorption maxima provides broad-spectrum protection against UVR across the entire UVA and UVB regions (30). In *N. commune*, glycosylated MAAs are thought to be associated with the three-dimensional extracellular matrix, providing protection against UVR, as well as desiccation and oxidative stress (45). It has been suggested that glycosylated MAAs may be the physiologically inactive variant of algycone MAAs, since glycosylation can block degradation by hydrolases (41). However, it is also possible that glycosylated and algycone MAAs isoforms have different cellular functions. While the functions of hexose-shinorine and hexose-polythylene-serine identified in *S. crispum* remain unknown, it is likely that these MAAs are also associated with the extracellular matrix and involved in protection against UVR, desiccation, and/or oxidative stress. However, it is worth noting that while both *S. cf. crispum* strains were isolated from relatively similar environments with similar UVR exposure, they displayed distinct MAA profiles when grown in culture, highlighting the complexity of MAA metabolism in this cyanobacterium. Future investigations of the regulation of MAAs and the corresponding biosynthe-
sis genes under different stress conditions may help clarify the physiological roles of these compounds.

ACKNOWLEDGMENTS

We thank Susie Wood for donating the cyanobacterial cultures used in this study and Emily Balskus for advice regarding obtaining shinorine and palythine-334 analytical standards. Also, we thank Leanne Pearson and John Kalaitzis for editing the manuscript.

FUNDING INFORMATION

This work, including the efforts of Vivek S. Javalkote, was funded by Endeavour Research Fellowship: Australian Government. This work, including the efforts of Vivek S. Javalkote, was funded by Research Fellowship for Meritorious Students: University Grants Commission, India. Including the efforts of Vivek S. Javalkote, was funded by Research Fellowships: Australian Government. This work, in including the efforts of Vivek S. Javalkote, was funded by Research Fellowships: Australian Government. This work, in

REFERENCES


