

# Biosynthetic origin of natural products isolated from marine microorganism–invertebrate assemblages

T. Luke Simmons\*, R. Cameron Coates\*, Benjamin R. Clark\*, Niclas Engene\*, David Gonzalez†, Eduardo Esquenazi‡, Pieter C. Dorrestein<sup>†§¶||</sup>, and William H. Gerwick\*<sup>¶||</sup>

\*Scripps Institution of Oceanography, Departments of †Chemistry and Biochemistry, ‡Pharmacology, and †Biology, and ¶Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California at San Diego, La Jolla, CA 92093

Edited by Jerrold Meinwald, Cornell University, Ithaca, NY, and approved December 10, 2007 (received for review October 16, 2007)

In all probability, natural selection began as ancient marine microorganisms were required to compete for limited resources. These pressures resulted in the evolution of diverse genetically encoded small molecules with a variety of ecological and metabolic roles. Remarkably, many of these same biologically active molecules have potential utility in modern medicine and biomedical research. The most promising of these natural products often derive from organisms richly populated by associated microorganisms (e.g., marine sponges and ascidians), and often there is great uncertainty about which organism in these assemblages is making these intriguing metabolites. To use the molecular machinery responsible for the biosynthesis of potential drug-lead natural products, new tools must be applied to delineate their genetic and enzymatic origins. The aim of this perspective is to highlight both traditional and emerging techniques for the localization of metabolic pathways within complex marine environments. Examples are given from the literature as well as recent proof-of-concept experiments from the authors' laboratories.

Extensive drug discovery and small molecule screening programs over the past two decades have shown marine organisms to be rich sources of structurally diverse and highly bioactive natural products (1). The molecular architectures of marine metabolites are distinct from those of their terrestrial relatives in that the physicochemical requirements of adaptation to an aqueous world, the biosynthetic pathways used, and even the elements employed in crafting their arsenal of defensive molecules are quite different (2). As a consequence of their structural diversity and uniqueness, marine natural products are providing a prominent share of the recent clinical and preclinical lead compounds for the treatment of various diseases, most prominently cancer (3).

At first analysis, the sources of these important leads are taxonomically diverse and include sponges, tunicates, corals, mollusks, fungi, and sediment-derived bacteria. However, there is growing recognition that the "collected source" for these molecules is not necessarily the "metabolic source." That is, marine bacteria and cyanobacteria, either assimilated by an invertebrate grazer (e.g., sea hare grazing on cyanobacteria) or growing in association with an invertebrate host in a symbiotic or commensal relationship, are frequently the true origin of these molecules (4–6). The growing appreciation of microbial metabolism in the production of many of the marine pharmaceutical leads comes from evidence in several sectors. Ultimately, firm knowledge of the natural product-producing organism, the "holder of the genes," is essential to numerous branches of science and technology, including ecology, biosynthesis, and natural product drug development.

Of 20 compounds deriving from (or inspired by) marine natural products that are currently or were recently in clinical trials for the treatment of cancer (5), 15 were isolated from sponges, tunicates, and mollusks, with only 5 coming directly from a microorganism. However, based on biosynthetic parallels, distribution in taxonomically diverse organisms, and, in a few cases, subsequent direct isolation from a producing microorganism, 16 of these anticancer molecules actually derive from microbial sources and only 4 derive from macroorganisms. For example, dolastatin 10 (1; Fig. 1), originally isolated from the marine gastropod *Dolabella auricularia*, was later shown to originate from cyanobacteria of the genera *Symploca* and *Lyngbya* upon which they feed (7–10). The powerful PKC-activating cancer cell toxin bryostatin (2) was isolated from the Californian bryozoan *Bugula neritina*, but the biosynthetic capacity to make this unique macrolide has been localized to an associated bacterium *Endobugula sertula* (11). Although the ecological roles of most marine natural products are not well understood, most appear to be involved in defense against predation or competition for space and nutrients (12–14).

It has been surprisingly difficult to demonstrate that an invertebrate-associated microorganism is unequivocally responsible for the production of a given secondary metabolite (6). However, recent advances on a number of fronts are providing new and powerful methods for investigating these relationships. For example, in the plant-endophyte arena, the anticancer compounds podophyllotoxin and camptothecin were recently discovered as products of endophytic microbes (15, 16). New advances in cell sorting, cul-

turing methods, and methods of chemical detection are enabling the physical separation of associated species for subsequent chemical analysis. Additionally, mass spectral imaging is proving exceptionally powerful for identifying the location of natural products in mixed species assemblages. Finally, gene probing techniques, such as catalyzed reporter deposition (CARD)-FISH or *in situ* hybridization, can identify those organisms with the genetic capacity to produce specific molecules of interest. This article is a review of advances made in each of these three areas as reported in the literature and provides some new and original results from our laboratories on these subjects.

## Results and Discussion

**Microorganism Isolation Approaches.** Confronted with a mixture of species that produces important secondary metabolites, the simplest approach is to separate cell types and then analyze these for the compound(s) of interest. It is plausible, however, that the natural product under study is produced in one species, excreted, and then assimilated by a second species. For example, efforts to localize peptides in the tunicate *Lissoclinum patella* showed that they occur predominately in the ascidian

Author contributions: T.L.S., P.C.D., and W.H.G. designed research; T.L.S., R.C.C., B.R.C., N.E., D.G., E.E., and P.C.D. performed research; T.L.S., R.C.C., B.R.C., N.E., E.E., P.C.D., and W.H.G. analyzed data; and T.L.S., R.C.C., P.C.D., and W.H.G. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

¶To whom correspondence may be addressed. E-mail: pdorrest@ucsd.edu or wgerwick@ucsd.edu.

This article contains supporting information online at [www.pnas.org/cgi/content/full/0709851105/DC1](http://www.pnas.org/cgi/content/full/0709851105/DC1).

© 2008 by The National Academy of Sciences of the USA

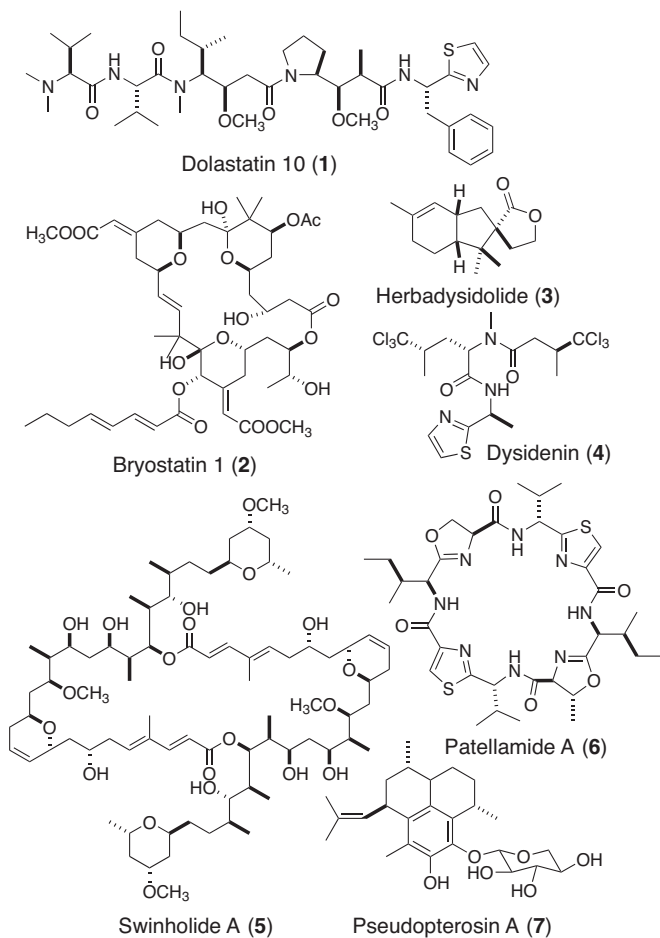


Fig. 1. Structures 1–7.

tunic and not in the symbiotic cyanobacterium *Prochloron* sp. (17). However, subsequent heterologous expression and genome sequence analysis indicated that these peptides are indeed genetically encoded for and biosynthesized by the cyanobacterium (18, 19). In a few cases, and despite considerable technological difficulties, successful segregation of cell types and their cultivation has allowed isotope incorporation studies to demonstrate new biosynthesis and, hence, identification of the producing organism (4, 6, 20).

Sponge tissues are extremely rich in bacteria that can comprise up to 40% of their biomass (21). Early efforts to localize natural products to specific cell types used formalin or glutaraldehyde-fixed sponge-cyanobacterial tissues from the tropical sponge *Dysidea herbacea* (Fig. 2). The tissues were disrupted, and individual cells were separated by using a fluorescence-activated cell sorter (FACS). Chemical analysis of the sorted cell types showed sesquiterpenoid compounds (e.g., 3) to be physically associated with the sponge cells, whereas chlorinated peptides (e.g., 4) were local-

ized to the associated cyanobacterium, *Oscillatoria spongeliae* (22). In studies of the cellular localization of several complex metabolites isolated from the sponge *Theonella swinhoei*, including swinholide A (5) and the peptide theopalauamide, a combination of dissection and differential centrifugation gave cell preparations that were chemically analyzed. Curiously, theopalauamide was found in a filamentous bacterial population morphologically related to *Beggiatoa*, whereas swinholide A (6) was localized to a mixed unicellular bacterial fraction (23, 24). This latter finding is at odds with our subsequent isolation of swinholide A from a free-living marine cyanobacterium, *Geitlerinema* sp. (25).

The microbial communities associated with various sponges have been investigated by using both culture-dependent and culture-independent methodologies (4). Recently, a culturable  $\alpha$ -proteobacterial symbiont was obtained from several sponges, including *Mycale laxissima* (26). A fluorescently labeled 16S rRNA gene probe was prepared for FISH analysis to visualize gene distribution. Intriguingly, the bacteria were concentrated on the

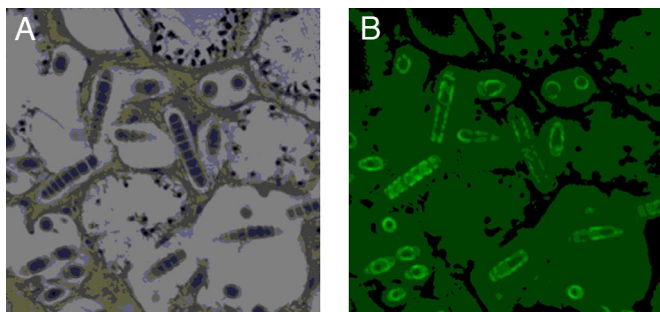
surfaces of the sponge larvae, suggesting vertical transmission of symbionts between sponge generations. However, no natural products have been reported from this bacterium to date (27).

In another recent example, the spatial distribution of bacteria within the sponge *Tethya aurantium* was mapped by using denaturing gradient gel electrophoresis (DGGE) and 16S rDNA clone library analysis. These studies showed that a new phylotype of *Flexibacteria* occurred in the sponge cortex whereas the endosome contained the cyanobacterium *Synechococcus* sp. Interestingly, 16S rDNA sequences for a  $\beta$ -proteobacteria were found throughout the endosome and cortex (27).

Various gene probing methods (e.g., FISH, PCR screening, Southern analysis, and genome sequencing) can demonstrate the occurrence of a biosynthetic gene in a particular isolated cell type. For example, the Haygood laboratory demonstrated a reduction in the abundance of a ketosynthase gene from the putative bryostatin gene cluster in *B. neritina* larvae that had been treated with antibiotics (28). In a second example, a genome sequence of the cyanobacterium *Prochloron* sp., a symbiont in the tunicate *Lissoclinum patella*, identified the cyanobacterium as the biosynthetic source of the ribosomally encoded peptides patellamides A (6) and C (18, 19).

Eukaryotic dinoflagellates are common coral symbionts, and soft corals in the Gorgonacea are particularly well studied for their bioactive diterpenoids. The pseudopterosins (e.g., 7) for example are potent antiinflammatory terpenes obtained from the soft coral *Pseudopteroergorgia elisabethae* and are used in commercially available skin care products. Investigation of the site of biosynthesis of the pseudopterosins by using differential centrifugation produced a 99% pure preparation of the dinoflagellate *Symbiodinium* sp. This macroalgal preparation was found to contain compound 7, and provision of either [ $^{14}\text{C}$ ]NaHCO<sub>3</sub> or  $^3\text{H}$ -geranylgeraniol pyrophosphate led to radioactive HPLC peaks corresponding to the pseudopterosins (29). From these and other data, it was concluded that the dinoflagellate was responsible for pseudopterosin biosynthesis. However, a subsequent patent has appeared that identifies associated bacteria as the biosynthetic source of the pseudopterosins, pointing out the danger of concluding too much from a cell separation followed by chemical analysis or unlabeled radiolabeling experiments (30).

Although most marine microorganisms are not easily separated from complex assemblages by direct manipulation, filamentous cyanobacteria grow to dimensions more amenable to manual isolation



**Fig. 2.** Photomicrographs of a thin section of sponge tissue showing embedded cyanobacterial cells. (A) Cross-section of the sponge *Dysidea herbacea* stained with O-toluidine showing composition of cells types. (B) Autofluorescence (522 nm) of cyanobacterial cells due to chlorophyll. [Reprinted with permission from ref. 92 (Copyright 2005, Springer).]

(e.g., *Lyngbya majuscula* filaments are up to 100  $\mu\text{m}$  wide and several centimeters long). However, the tough polysaccharide sheath surrounding cyanobacterial filaments can harbor a multitude of heterotrophic bacteria that are difficult to remove (31–33) (Fig. 3). We and others find that the dominating heterotrophic bacteria associated with these sheaths are gram-negatives from the phyla *Proteobacteria* and *Bacteroidetes*. These include the  $\alpha$ -proteobacteria (*Caulobacter*, *Thalassobius*, and *Rhodospirillaceae*),  $\beta$ -proteobacteria (*Achromobacter*),  $\gamma$ -proteobacteria (*Pseudomonas* and *Oceanospirillum*), and *Bacteroidetes* (*Flavobacterium* and *Flexibacter*) (34, 35). Interestingly, many of these same bacterial groups are also commonly associated with sponges and tunicates (36, 37).

Creating axenic cultures of cyanobacteria is extraordinarily difficult, and relatively few are reported in the literature. Previous investigations proposed that axenic cultures were produced by using irradiation and antibiotics; however, proof of their monoculture state used lack of bacterial growth on multiple media types as evidence (31–33, 38). Because the media, nutrient, and environmental combinations were not exhaustive, doubt remains as to whether these cultures were truly free of bacteria. Many oligotrophic organisms grow very slowly and to low cell densities; therefore, their presence can be easily overlooked by classic culture-dependent tests for axenicity (39). Hence, it remains uncertain whether sheath-associated bacteria are involved in the production or modification of natural products reported from cyanobacteria.

**Chemical Detection Methods. Microanalysis by mass spectrometry (MS).** In conjunction with the physical and genetic approaches described above, chemical detection methods have improved significantly, in particular MS, and these are becoming highly useful in locating the biosynthetic origin of specific metabolites. Particu-

larly noteworthy improvements in mass spectrometry include the ability to analyze small quantities of material while increasing both resolution and mass accuracy. These developments have been largely driven by the proteomics community, usually in the context of discovering disease biomarkers (40–43). These efforts have spurred considerable development of new pre-separation methods, new ionization approaches, improved mass filters, faster scan rates for improved time scale detection, and new detection methods, only some of which have found their way into the routine analysis of natural products. These developments in MS forecast that soon we may be able to perform rigorous structure elucidation on submicrogram quantities within certain metabolite classes. In the following sections, we present several examples from the authors' laboratories that demonstrate how some of these very sensitive MS technologies can be used in the discovery of new marine natural products, and how variations in these methods can help in identifying the true biosynthetic sources of these compounds.

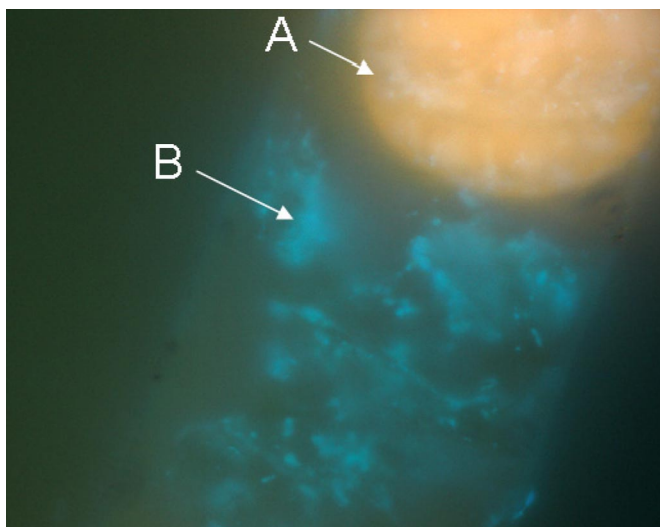
Electrospray ionization (ESI) and MALDI MS are revolutionary soft ionization methods that enable MS measurements of chemically sensitive biomolecules (44). However, for either of these techniques to be effective, there is an absolute requirement for ionization. ESI has become the most common approach of ionization in the natural product chemist's arsenal because it can be interfaced with LC capabilities, thereby connecting molecular mass data with elution characteristics of a particular molecule. These data greatly expedite the assignment of a natural product as either novel or known (dereplication).

In addition to traditional HPLC-ESI-MS, capillary LC-ESI-MS has proven extraordinarily useful in proteomic research. Capillary LC requires extremely small sample sizes (<1.0  $\mu\text{g}$ ) with the ability to

interface with many types of chromatographic supports such as reverse phase (C18, C4, etc.), hydrophilic interaction liquid chromatography (HILIC), cationic exchange, or even combinations of multiple resins [see supporting information (SI) Fig. 8] (45, 46). For example, we detected (and subsequently isolated and confirmed by NMR) the anticancer lead compound iejimalide A (**8**; Fig. 4) from a 0.1- $\mu\text{g}$  crude cyanobacterial extract. This was significant because this macrolide had only previously been found from an Okinawan tunicate (47, 48). Detection was achieved by using a 100- $\mu\text{m}$  fused silica capillary column loaded with 10 cm of C18 resin and a 5- $\mu\text{m}$  tip. The estimated detection limit of iejimalide A by this technique is in the picomole-to-femtomole range. By contrast, traditional LC-MS with an LCQ instrument and 4.6-mm-diameter HPLC column required >100  $\mu\text{g}$  of iejimalide-containing crude extract for detection (estimated detection limit in the micromole-to-nanomole range).

**Nanospray Fourier transform (FT)-ion cyclotron resonance MS (ICRMS).** High-resolution MS is another extremely useful technique for analyzing crude mixtures for the presence of a specific natural product (49). We used a complement of nanospray and FT-ICRMS to analyze the crude extract of the iejimalide A-producing cyanobacteria from Papua New Guinea. Despite the fact that the crude extract showed >40 ions in the region  $m/z$  675–705, the sodiated iejimalide A peak at  $m/z$  701.418 (calculated  $m/z$  701.414) was well resolved and firmly identified the presence of iejimalide A (**8**; see SI Fig. 9). In total, this extract displayed several hundred unique molecular ions in a single spectrum (including several consistent with novel hydroxylated analogs of **8**), and these were only resolvable because of the high-resolution attributes of FT-ICRMS. It should be noted that additional characterization of these metabolites can be achieved by combining the high resolution of FT-ICRMS with tandem MS (49). This “snapshot” of the secondary metabolome of a cyanobacterium will surely become an important discovery tool as natural product scientists learn to process and use such rich datasets.

**MALDI-TOF-MS.** The species diversity associated with complex marine assemblages is well suited for analysis by another powerful MS technique, matrix-assisted laser desorption/ionization–time-of-flight MS (MALDI-TOF-MS). The utility of MALDI-MS for the analysis of biological samples (e.g., single cells and tissues) has been demonstrated for both biomedical and natural science applications. For example, MALDI-techniques were applied to investigate entire colonies of the cyanobacterium *Microcystis* to under-



**Fig. 3.** Trichome structure of the filamentous marine cyanobacterium *Lyngbya majuscula*-3L. Shown is DAPI staining and epifluorescent imaging at  $\times 1,000$  magnification using a Zeiss Axioskop (filter set #2, emission 420 nm+). A, cyanobacterial phycoerythrin in filament tip (orange); B, filament sheath-associated bacterial DNA (blue).

stand the diversity and distribution of hepatotoxic oligopeptides from these natural communities (50). Recently, “intact cell MALDI-TOF” (ICM) was used to assess the diversity of microbes from the tissues of marine invertebrates (e.g., sponges) by “proteomic clustering,” allowing both the dereplication of known secondary metabolites and providing a taxonomic characterization of the species present (51).

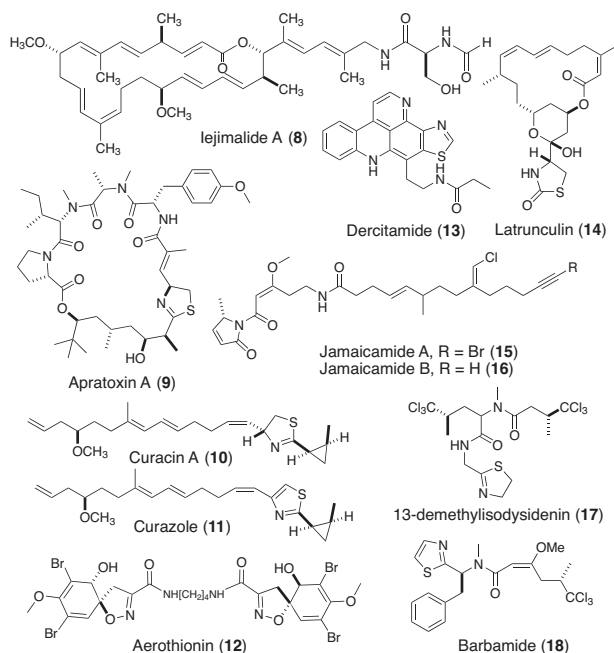
We have extended the utility of MALDI-TOF-MS to the analysis of single filaments of marine cyanobacteria

for the identification of known and new secondary metabolites. The analysis of intact filament assemblages enables the localization of specific secondary metabolites within specific filament types. MALDI-TOF-MS analysis was performed on fresh single filaments of a strain of *L. bouillonii* known to produce apratoxin A (9), a potent mammalian cell cytotoxin (52, 53). In this example, a single filament was placed on a MALDI target plate and coated with an  $\alpha$ -cyano-4-hydroxycinnamic acid matrix. Data analysis readily gave a strong

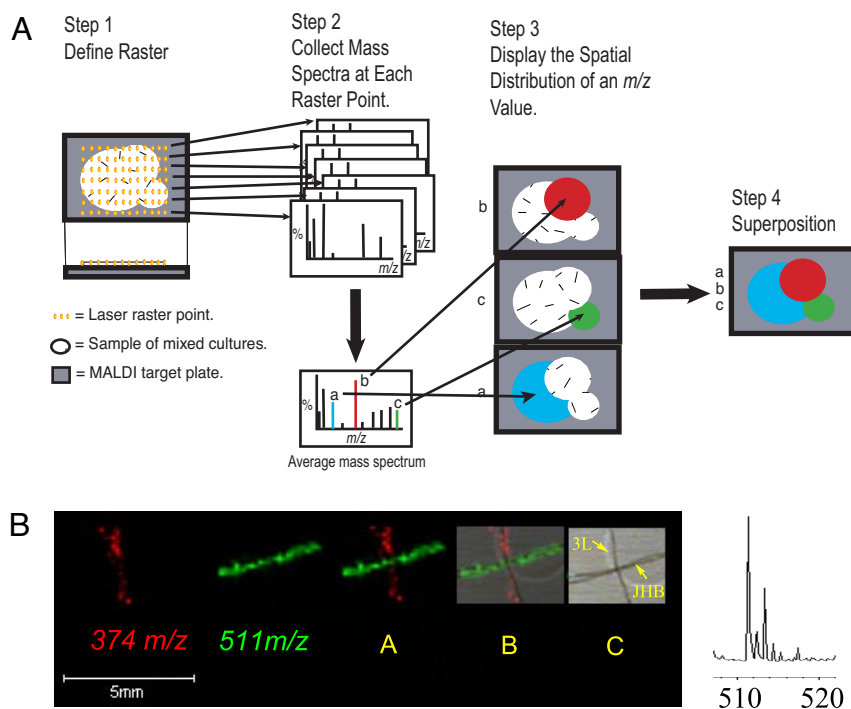
pseudomolecular ion at  $m/z$  840 (see SI Fig. 10), a value corresponding to the molecular formula of apratoxin A [ $C_{45}H_{69}N_5O_8S + H$ ].

We also used MALDI-MS in combination with ESI-FTMS, LC-ESI-MS, and NMR to detect and characterize a novel analog of curacin A (10) (54–56). Aside from two double-bond isomers, curacins B and C (57), and the 8-desmethyl analog curacin D (58), no other naturally occurring analogs have yet been isolated. The biosynthesis of the distinctive cyclopropyl ring has been the focus of detailed molecular genetic and mechanistic biochemistry investigations because it appears to involve the intermediacy of a cryptic chlorination reaction (59–61). MALDI analysis of a single *L. majuscula* filament from Curaçao revealed a pseudomolecular ion at  $m/z$  374  $[M+H]^+$ , consistent with the presence of curacin A. Another peak at  $m/z$  372  $[M+H]^+$  was also observed that did not correspond to any known cyanobacterial metabolite. FT-ICRMS of the extract from this strain allowed characterization of the 372 peak as exactly 2.015 Da less than curacin A, for a deduced molecular formula of  $C_{23}H_{33}NOS$ ; thus, it represented a potentially intriguing structural homolog of curacin A. The pure compound was isolated and characterized by NMR and other spectroscopic techniques and shown to be an oxidized form of curacin A with a thiazole rather than a thiazoline ring. Previously, we had produced analog 11 via semisynthesis from natural curacin A and given it the trivial name “curazole” (62). Based on detection of the  $[M+H]$  ion from intact cyanobacterial filaments and through direct analysis of unfractionated crude extracts, it seems certain that curazole (11) is a genuine natural product of this *L. majuscula* strain. Contrary to literature precedent (60, 63), our sequence analysis revealed that the oxidase required for the thiazoline-to-thiazole transformation is not present in the curacin A gene cluster, nor is it found immediately up- or downstream of the cluster (60). This suggests that this transformation of the thiazoline to the thiazole is not encoded by a dedicated protein or protein domain.

There are pitfalls in the direct MALDI analysis of biological tissues, as illustrated by our attempt to detect iejimalide A (8) from cyanobacterial filaments. As noted, iejimalide A had been previously detected by traditional organic extraction of the collected biomass followed by NMR and FT-ICRMS approaches. MALDI-TOF-MS of single filaments from this sample failed to detect any 8. Consistent with this finding, pure iejimalide A, once it became available, also failed to provide a detectable ion signal by MALDI-TOF, likely be-



**Fig. 4.** Structures 8–18.



**Fig. 5.** MALDI-TOF imaging to map the distribution of natural products in biological tissues. (A) Schematic for procedure of MALDI-TOF imaging. (B) From left to right, MALDI-TOF imaging of protonated curacin A (**12**) at  $m/z$  374 (red), sodiated jamaicamide B (**19**) at  $m/z$  511 (green), the simultaneous superposition of both  $m/z$  374 and 511, superposition of both masses plus photograph of filaments, photograph of both filaments identified as strains "3L" (= curacin A producer) and JHB [= jamaicamide A (**18**) and B (**19**) producer], and partial mass spectrum of JHB strain showing the molecular ion cluster of jamaicamide B with its distinctive chlorine isotope pattern.

cause of its lack of basic nitrogen atoms. This example demonstrates that MALDI and ESI are complementary approaches and that both should be used in natural product dereplication and discovery programs.

Other important MS advances stem from the development of novel ionization methods and new mass analyzers. For example, desorption electrospray ionization (DESI), electrospray-assisted laser desorption/ionization (ELDI), and direct analysis in real time (DART) are newly developed ionization methods that analyze samples directly and with minimal preparation (64–68). The new orbitrap Fourier transform (FT) mass analyzer can provide high resolution without the required expertise that an FT-ICR instrument requires (69, 70), a characteristic that has made this the instrument of choice in metabolomic studies (71). Although none of these more recent methods have yet been applied to the study of cyanobacterial natural products or complexes of coexisting species, it is likely that some will offer advantages over current methods.

**Imaging Techniques. Background.** In 1983, Faulkner and colleagues (72) used x-ray analysis coupled with transmission electron microscopy (TEM) to examine the

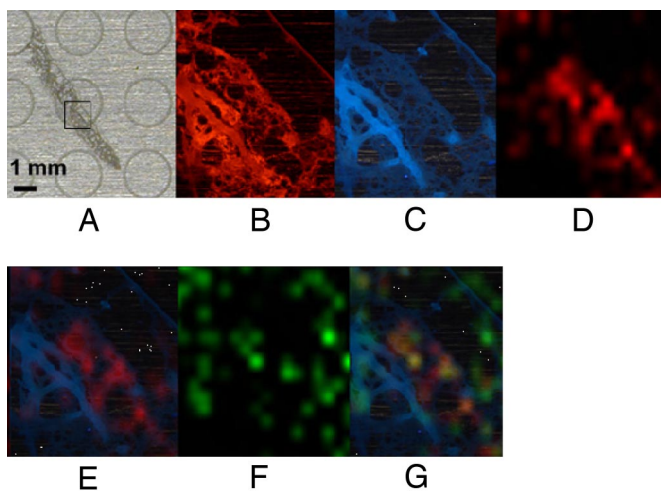
location of two brominated metabolites, arothionin (**12**) and homoarothionin, within a sponge–microorganism complex. Although these experiments recorded increased bromine in a particular cell type (spherulous sponge cells), it was not certain that the detected bromine was actually part of the target compounds. Subsequently, the location of the polyaromatic alkaloid dercitamide (**13**) was directly imaged by using epifluorescence and laser-scanning confocal microscopy in conjunction with TEM. Interestingly, this study localized the natural product exclusively to the sponge cells (*Oceanapia sagittaria*) and not to the associated microbes as was expected (73). In another early study, the imaging of latrunculin B (**14**) in the sponge *Negombata magnifica* used immunogold staining with light microscopy and TEM. These studies also revealed that sponge cells were the site of storage and presumed biosynthesis (74). In light of subsequent natural product and biosynthetic investigations, these results with latrunculin B are particularly intriguing because related compounds have been shown to be produced by microorganisms using assembly-line type syntheses typical of bacteria (2, 11).

**MALDI imaging.** It is clear that single-filament MALDI-TOF analysis is a

powerful method for directly observing both new and known secondary metabolites. Can MALDI-TOF also provide insight into the identity of metabolite-producing species found in complex environmental assemblages? As noted, this has been an extraordinarily difficult phenomenon to study because coexisting strains are difficult to grow as single pure strains, and many artifacts are introduced during sample preparation. An emerging strategy showing promise is MALDI-TOF imaging (75–78). MALDI imaging allows the spatial localization of precise molecular masses to particular cell types that are then observed through assignment of specific colors to a given mass (Fig. 5). Although this approach has not previously been applied to the analysis of natural products, it has been used to identify biomarkers of disease and the accumulation of therapeutics in whole-animal histological preparations (79, 80).

Here, we briefly demonstrate the power and utility of this technology in two preliminary studies (E.E., R.C.C., T.L.S., D.G., W.H.G., and P.C.D., unpublished work). First, two *L. majuscula* strains (JHB from Jamaica and 3L from Curaçao) were mixed and analyzed by MALDI-TOF imaging. It is known that the JHB strain produces jamaicamide B (**16**) and the 3L strain produces curacin A (**10**) (54, 62, 81). Sodiated jamaicamide B has an  $m/z$  of 511 and is colored green on the image, whereas protonated curacin A has an  $m/z$  of 374 and is shown in red (Fig. 5). The MALDI-TOF image clearly distinguishes the two producing strains by the secondary metabolites they contain. Moreover, it is clear that jamaicamide B is chlorinated from the observed  $^{35}\text{Cl}/^{37}\text{Cl}$  isotopic ratio.

In a second example, we analyzed the marine sponge *Dysidea herbacea*. Because of its relative availability, abundance of described natural products, and associated cyanobacterium (*Oscillatoria spongeliae*), *D. herbacea* has developed as a useful model for secondary metabolite and biosynthetic gene cluster localization studies (see below, *Gene-based imaging methods*). Here, we prepared thin sections of sponge–cyanobacterium tissue using a cryogenic microtome for analysis by epifluorescence microscopy and MALDI imaging. Fig. 6 shows the portion of a thin section of sponge–cyanobacterium tissue used in the imaging analysis by autofluorescence at two wavelengths and then by MALDI-TOF. The mass at  $m/z$  530 (Fig. 6 D, E, and G) and molecular ion isotope cluster (data not shown) are consistent with a hexachlorinated peptide such as 13-



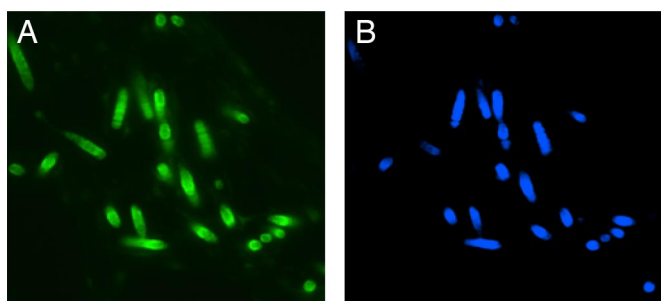
**Fig. 6.** MALDI-TOF imaging of the marine sponge *Dysidea herbacea*. (A) MALDI target plate with thin section in place. (B) Autofluorescence of *D. herbacea* thin section at 590 nm+. (C) Autofluorescence of *D. herbacea* thin section at 420 nm+. (D) MALDI image of  $m/z$  530. This mass and the complex molecular ion isotope cluster suggests its identity as the hexachlorinated peptide 13-demethylisodysidenin (20). (E) Autofluorescence (420 nm) image overlaid by MALDI image at  $m/z$  530 indicating localization of compound with this mass. (F) MALDI image of unknown compound  $m/z$  1,028. (G) Autofluorescence (420 nm) image overlaid with  $m/z$  530 (red) and  $m/z$  1,028 (green) showing the differential localization of these two molecules.

demethylisodysidenin (17). Another mass,  $m/z$  1,028 (Fig. 6 *F* and *G*), is for an unknown metabolite with a very different distribution in the sponge matrix and clearly illustrates the power of MALDI-TOF imaging in differentially localizing secondary metabolites from complex environments.

At present, MALDI imaging is limited by its spatial resolution, due in part to the requirement for a uniform crystalline matrix coating. The matrix coating allows for molecular diffusion of the analyte in the crystalline lattice, limiting the spatial resolution to the size of the crystals or the width of a typical  $N_2$  laser ( $\approx 100 \mu m$ ). With improvements in both matrix deposition and laser technologies, it is anticipated that resolution in the range of 7–10  $\mu m$  will become available (82). Finally, two recent advances that will allow structural charac-

terization in conjunction with imaging are MALDI imaging-FT-ICR and the LTQ imaging system with tandem MS capabilities (77, 83).

**Gene-based imaging methods.** The Haygood laboratory has examined the biosynthetic source of the anticancer bryostatins (e.g., 2). Originally isolated from the bryozoan *B. neritina*, the bryostatin structures are consistent with a microbial polyketide synthase origin. To investigate the metabolic origin of 2, this group (28) used degenerate primers to clone a component of the polyketide synthase, the ketosynthase (KS) domain, from both larval and adult forms of the bryozoan. A total of 9 unique KS sequences were recovered, and one pre-dominated in the larvae. This latter sequence was used to produce a biotin-labeled RNA gene probe that was subsequently visualized with an avidin-



**Fig. 7.** CARD-FISH analysis of *Dysidea herbacea* showing autofluorescence (522 nm) of cyanobacterial cells due to chlorophyll (A) and probe-specific hybridization with the complimentary oligonucleotide to *dysB1*, a homolog of the barbamide *barB1* biosynthetic gene (B). [Reprinted with permission from ref. 92 (Copyright 2005, Springer).]

horseradish peroxidase detection system. Larvae treated with this probe were labeled only in a thin band, the pallial sinus, which was shown to contain a new  $\gamma$ -proteobacterial strain, *Endobugula ser-tula*. Although it was not conclusively demonstrated at this point whether the KS was specifically involved in bryostatin biosynthesis, subsequent gene cloning work indicated that it very likely was (11). Our analysis shows a 98% sequence identity at the DNA base level between the original KS and the BryB KS present in the subsequently cloned and sequenced bryostatin gene cluster.

In 1996, we reported the structure of the chlorinated peptide, barbamide (18), from a Curaçao collection of the marine cyanobacterium *L. majuscula* (84). Barbamide is strongly molluscicidal to the intermediate snail host involved in schistosomiasis, *Biomphalaria glabrata*. The halogen atoms comprising the unique trichloromethyl group are located at a position without an adjacent activating functionality, and this suggested a novel route for their biochemical incorporation (85, 86). These barbamide halogenases and related enzymes have since been characterized as a new class of “radical halogenases” capable of direct halogenation of unreactive carbon atoms in natural products, such as methyl groups (87–89). Early on, we recognized the striking structural similarity of barbamide to a number of *Dysidea* “sponge” chlorinated peptides. We and others proposed that the cyanobacterial symbiont (*Oscillatoria spongeliae*) found growing among the sponge cells was likely responsible for chlorinated peptide production (90, 91). Cell sorting experiments, as noted earlier, supported this hypothesis as the peptides were recovered from the purified cyanobacterial fraction (90, 91).

The DNA sequences of the barbamide halogenases BarB1 and BarB2 allowed a gene-based approach to investigate and confirm the identity of the producing organism in this symbiosis (92). First, homology-based cloning from the mixed sponge/cyanobacterial tissue (Fig. 2) yielded a *barB1* homolog, *dysB1*, encoding a protein sequence with 93% amino acid identity to BarB1. The *dysB1* sequence in combination with cyanobacterial-specific 16S rDNA sequences was used in FISH experiments to probe thin sections of the sponge containing the associated cyanobacterial symbiont. The cyanobacteria were identified in the thin sections by their chlorophyll fluorescence and by localization of the fluorescently labeled cyanobacterial-specific 16S rDNA gene probe. The crucial experiment showed colocalization of the fluorescently labeled *dysB1* probe and

16S rDNA probe in cyanobacterial cells but not in sponge cells (Fig. 7). Cyanobacterial cells were thereby unequivocally shown to possess the halogenase mRNA necessary for chlorinated peptide biosynthesis.

## Conclusions

The presence of a natural product within a particular cell type can provide only circumstantial evidence as to its biosynthetic origin. Above, various older techniques for showing the presence of a natural product have been reviewed and newer experimental approaches discussed. Some of these are appropriate for analysis of isolated cells, including HPLC, NMR, and, most powerfully, MS. Other techniques are well suited for sample imaging, such as x-ray or fluorescence microscopy, and MALDI approaches. Nevertheless, all of these approaches fall short of providing unequivocal proof for the site of biosynthesis of a given compound.

Conceptually, demonstrating the biosynthesis of a natural product in a given cell type can be achieved on chemical, protein, or genetic levels. At the chemical level, demonstration of new compound synthesis in an isolated cell type is compelling and may involve any of a number of analytical chemical approaches, possibly involving isotope-labeled precursors. At the genetic level, DNA or RNA sequences encoding the biosynthetic enzymes can be localized, either in isolated cells or through imaging techniques. The CARD-FISH analysis of the sponge/cyanobacterium complex *Dysidea herbacea/Oscillatoria*

*spongelliae* provides an example of RNA detection (92); uncovering the gene for patellamide biosynthesis in isolated cells of *Prochloron* sp. from the tunicate *Lissoclinum patella* is an example of detection at the DNA level (18, 19). A considerable hurdle in applying most of these gene-based technologies is that the biosynthesis needs to be sufficiently understood such that the genes can be recognized with certainty. The standard methods for demonstrating that a given gene is involved in the biosynthesis of a compound involve either gene knockout experiments or heterologous expression of the biosynthetic pathway. Both of these methods require genetic manipulations that can be extremely challenging in a little studied marine organism. To date, the function of most marine natural product biosynthetic genes have been deduced through bioinformatic analysis.

At the protein level, little has been achieved because there is so little knowledge of biosynthetic proteins in general and even less so from members of complex marine environments. However, knowledge and techniques are emerging that will allow powerful interrogation of the site of biosynthesis using biosynthetic proteins. An increasing number of marine biosynthetic pathways are being cloned and characterized. Complementing this is a dramatic increase in the number and diversity of marine genome and metagenome sequences (93–95). In principle, this genetic sequence information can be used to identify diagnostic peptide fragments from putative biosynthetic enzymes that can be liberated by protease digestion

and localized by subsequent MALDI imaging.

Exciting hybrid technologies are becoming available that will increase the rigor and ease of demonstrating the site of biosynthesis in complex assemblages. Cell isolations and genome determinations are now possible on a single cell [e.g., using multiple displacement polymerase chain reaction technology (96)], from which biosynthetic clusters can be deduced and candidate structures generated by using established biosynthetic logic. Structures of sufficient structural or anticipated biological interest can be located in cultured or collected biomass with increasing facility (97). Alternatively, synthetic genes of substantial size (e.g., 50,000 bp) can be produced with optimized codon usage and promoter sequences for effective heterologous expression (98–100). Indeed, the coming years will witness amazing technological advances and allow us to ask and answer long standing, profound, and currently unimagined questions in chemical biology.

Details of the experiments described in this article can be found in *SI Methods and Data*.

**ACKNOWLEDGMENTS.** We thank P. Crews and K. Tenney (University of California, Santa Cruz) for samples of the sponge *Dysidea herbacea*, B. Palenik and B. Brahamsha for use of their microscopes, L. Gerwick for bioinformatics analysis of the curacin A gene cluster, R. Grindberg for analysis of BryB in the bryostatin gene cluster, T. Matainaho and the captains/crews of the Golden Dawn and Teleta Dive Boats for their assistance with cyanobacterial collections, and the governments of Papua New Guinea and Curaçao for permission to make these collections. This work was supported by National Institutes of Health Grants CA52955 and CA100851.

- Newman DJ, Cragg GM (2007) *J Nat Prod* 70:461–477.
- Ramaswamy AV, Flatt PM, Edwards DJ, Simmons TL, Han B, Gerwick WH (2006) in *Frontiers in Marine Biotechnology*, eds Prokisch P, Müller WEG (Horizon Bioscience, Norfolk, England), pp 175–224.
- Simmons TL, Andrianasolo E, McPhail K, Flatt P, Gerwick WH (2005) *Mol Cancer Ther* 4:333–342.
- Piel J (2006) *Curr Med Chem* 13:39–50.
- Simmons TL, Gerwick WH (2008) in *Oceans and Human Health*, eds Walsh P, Solo-Gabriele H, Fleming LE, Smith SL, Gerwick WH (Elsevier, New York), in press.
- Hildebrand M, Waggoner LE, Liu H, Sharp KH, Ridley CP, Haygood MG (2004) *Nat Prod Rep* 21:122–142.
- Pettit GR, Kamano Y, Herald CL, Tuinman AA, Boettner FE, Kizuh H, Schmidt JM, Baczynski L, Tomer K, Bontems RJ (1987) *J Am Chem Soc* 109:6883–6885.
- Pettit GR, Kamano Y, Dufresne C, Cerny RL, Herald CL, Schmidt JM (1989) *J Org Chem* 54:6005–6006.
- Leusch H, Moore RE, Paul VJ, Mooberry SL, Corbett TH (2001) *J Nat Prod* 64:907–910.
- Williamson RT, Chapin EL, Carr AW, Gilbert JR, Graupner PR, Lewer P, McKamey P, Carney JR, Gerwick WH (2000) *Org Lett* 2:289–292.
- Sudek S, Lopanik NB, Waggoner LE, Hildebrand M, Anderson C, Liu H, Patel A, Sherman DH, Haygood MG (2007) *J Nat Prod* 70:67–74.
- Paul VJ, Puglisi MP, Ritson-Williams R (2006) *Nat Prod Rep* 23:153–180.
- Pawlik JR (1993) *Chem Rev* 93:1911–1922.
- Faulkner DJ, Ghiselin MT (1983) *Mar Ecol Prog Ser* 13:295–301.
- Eyberger AL, Dondapati R, Porter JR (2006) *J Nat Prod* 69:1121–1124.
- Puri SC, Verma V, Amna T, Qazi GN, Spitteller M (2005) *J Nat Prod* 68:1717–1719.
- Salomon CE, Faulkner DJ (2002) *J Nat Prod* 65:689–692.
- Schmidt EW, Nelson JT, Rasko DA, Sudek S, Eisen JA, Haygood MG, Ravel J (2005) *Proc Natl Acad Sci USA* 102:7315–7320.
- Milne BF, Long PF, Starcevic A, Hranueli D, Jaspers M (2006) *Org Biomol Chem* 4:631–638.
- Kobayashi J, Ishibashi M (1993) *Chem Rev* 93:1753–1769.
- Friedrich AB, Merkert H, Fendert T, Hacker J (1999) *Mar Biol* 134:461–470.
- Ridley CP, Bergquist PR, Harper MK, Faulkner DJ, Hooper JN, Haygood MG (2005) *Chem Biol* 12:397–406.
- Bewley CA, Holland ND, Faulkner DJ (1996) *Experientia* 52:716–722.
- Schmidt EW, Bewley CA, Faulkner DJ, Haygood MG (1998) *J Org Chem* 63:1254–1258.
- Andrianasolo EH, Gross H, Goeger D, Musafija-Girt M, McPhail K, Leal RM, Mooberry SL, Gerwick WH (2005) *Org Lett* 7:1375–1378.
- Enticknap JJ, Kelly M, Peraud O, Hill RT (2006) *Appl Environ Microbiol* 72:3724–3732.
- Thiel V, Neulinger SC, Staufenberger T, Schmaljohann R, Imhoff JF (2007) *FEMS Microbiol Ecol* 59:47–63.
- Davidson SK, Allen SW, Lim GE, Anderson CM, Haygood MG (2001) *Appl Environ Microbiol* 67:4531–4537.
- Coleman AC, Kerr RG (2000) *Tetrahedron* 56:9569–9574.
- Kerr RG, Brueck T (2007) PCT Int Appl WO 20077037791.
- Adams DG (2001) *Microbiol Today* 28:131–133.
- Ferris MJ, Hirsch CF (1991) *Appl Environ Microbiol* 57:1448–1452.
- Gerwick WH, Roberts MA, Proteau PJ, Chen J (1994) *J Appl Phycol* 6:143–149.
- Paerl HW (1982) in *The Biology of Cyanobacteria*, eds Carr NG, Whitton BA (Univ of California Press, Berkeley), pp 441–461.
- Zehr JP, Mellon M, Braun S, Litaker W, Stepp T, Paerl HW (1995) *Appl Environ Microbiol* 61:2527–2532.
- Bewley CA, Faulkner DJ (1998) *Angew Chem Int Ed* 37:2162–2178.
- Perez-Matos AE, Rosado W, Govind NS (2007) *Antonie van Leeuwenhoek* 92:155–164.
- Kraus MP (1966) *Nature* 211:310–311.
- Giovannoni SJ, Stingl U (2005) *Nature* 437:343–348.
- Tyers M, Mann M (2003) *Nature* 422:193–197.
- Marshall AG, Hendrickson CL, Emmett MR, Rodgers RP, Blakney GT, Nilsson CL (2007) *Eur J Mass Spectrom* 13:57–59.
- Liu E, Gerszten RE (2006) *Cardiovasc Biomarkers* 575–586.
- Coon JJ, Syka JEP, Shabanowitz J, Hunt DF (2005) *Bio-Techniques* 38:519–523.
- Zhang J (2007) *Proteomics Clin Appl* 1:805–819.
- Boersema PJ, Divecha N, Heck AJR, Mohammed S (2007) *J Proteome Res* 6:937–946.
- Motoyama A, Venable JD, Ruse CI, Yates JR, III (2006) *Anal Chem* 78:5109–5118.

47. Nozawa K, Tsuda M, Ishiyama H, Sasaki T, Tsuruo T, Kobayashi J (2006) *Bioorg Med Chem* 14:1063–1067.
48. Kobayashi J, Cheng J, Ohta T, Nakamura H, Nozoe S, Hirata Y, Ohizumi Y, Sasaki T (1988) *J Org Chem* 53:6147–6150.
49. Feng X, Siegel MM (2007) *Anal Bioanal Chem* 389:1341–1363.
50. Welker M, Brunke M, Preussel K, Lippert I, von Döhren H (2004) *Microbiology* 150:1785–1796.
51. Dieckmann R, Graeber I, Kaesler I, Szweczyk U, von Döhren H (2005) *Appl Microbiol Biotechnol* 67:539–548.
52. Luesch H, Yoshida WY, Moore RE, Paul VJ, Corbett TH (2001) *J Am Chem Soc* 123:5418–5423.
53. Luesch H, Chandra SK, Raya RM, DeJesus PD, Orth AP, Walker JR, Belmonte JCI, Schultz PG (2006) *Nat Chem Biol* 2:158–167.
54. Gerwick WH, Proteau PJ, Nagle DG, Hamel E, Blohkin A, Slate DL (1994) *J Org Chem* 59:1243–1245.
55. Verdier-Pinard P, Sitachitta N, Rossi JV, Sackett DL, Gerwick WH, Hamel E (1999) *Arch Biochem Biophys* 370:51–58.
56. Wipf P, Reeves JT, Day BW (2004) *Curr Pharm Des* 10:1417–1437.
57. Yoo H-D, Gerwick WH (1995) *J Nat Prod* 58:1961–1965.
58. Marquez B, Verdier-Pinard P, Hamel E, Gerwick WH (1998) *Phytochemistry* 49:2387–2389.
59. Vaillancourt FH, Yeh E, Vosburg DA, O'Connor SE, Walsh CT (2005) *Nature* 436:1191–1194.
60. Chang Z, Sitachitta N, Rossi JV, Roberts MA, Flatt PM, Jia J, Sherman DH, Gerwick WH (2004) *J Nat Prod* 67:1356–1367.
61. Gu L, Jia J, Liu H, Håkansson K, Gerwick WH, Sherman DH (2006) *J Am Chem Soc* 128:9014–9015.
62. Verdier-Pinard P, Lai J-Y, Yoo H-D, Yu J, Marquez B, Nagle DG, Nambu M, White JD, Falck JR, Gerwick WH, et al. (1998) *Mol Pharmacol* 53:62–76.
63. Ming-L-J, Epperson JD (2002) *J Inorg Biochem* 91:46–58.
64. Keil A, Talaty N, Janfelt C, Noll RJ, Gao L, Ouyang Z, Cooks RG (2007) *Anal Chem* 79:7734–7739.
65. Cooks RG, Ouyang Z, Takats Z, Wiseman JM (2006) *Science* 311:1566–1570.
66. Shiea J, Huang M-Z, Hsu H-J, Lee C-Y, Yuan C-H, Beech I, Sunner J (2005) *Rapid Commun Mass Spectrom* 19:3701–3704.
67. Haefliger OP, Jeckelmann N (2007) *Rapid Commun Mass Spectrom* 21:1361–1366.
68. Sampson JS, Hawkridge AM, Muddiman DC (2006) *J Am Soc Mass Spectrom* 17:1712–1716.
69. Lim H-K, Chen J, Sensenhauser C, Cook K, Subrahmanyam V (2007) *Rapid Commun Mass Spectrom* 21:1821–1832.
70. Hu Q, Noll RJ, Li H, Makarov A, Hardman M, Cooks RG (2005) *J Mass Spectrom* 40:430–443.
71. Breitling R, Pitt AR, Barrett MP (2006) *Trends Biotechnol* 24:543–548.
72. Thompson JE, Barrow KD, Faulkner DJ (1983) *Acta Zool* 64:199–210.
73. Salomon CE, Deerinck T, Ellisman MH, Faulkner DJ (2001) *Mar Biol* 139:313–319.
74. Gillor O, Carmeli S, Rahamim Y, Fishelson Z, Ilan M (2000) *Mar Biotechnol* 2:213–223.
75. McDonnell LA, Heeren RMA (2007) *Mass Spectrom Rev* 26:606–643.
76. Signor L, Varesio E, Staack RF, Starke V, Richter WF, Hopfgartner G (2007) *J Mass Spectrom* 42:900–909.
77. Taban IM, Altelaa AFM, van der Burgt YEM, McDonnell LA, Heeren RMA, Fuchser J, Baykut G (2007) *J Am Soc Mass Spectrom* 18:145–151.
78. McLean JA, Ridenour WB, Caprioli RM (2007) *J Mass Spectrom* 42:1099–1105.
79. Cornett DS, Reyzer ML, Chaurand P, Caprioli RM (2007) *Nat Methods* 4:828–833.
80. Hsieh Y, Chen J, Korfmacher WA (2007) *J Pharmacol Toxicol Methods* 55:193–200.
81. Edwards DJ, Marquez BL, Nogle LM, McPhail K, Goeger DE, Roberts MA, Gerwick WH (2004) *Chem Biol* 11:817–833.
82. Chaurand P, Schriver KE, Caprioli RM (2007) *J Mass Spectrom* 42:476–489.
83. Verhaert PD, Conaway MCP, Pekar TM, Miller K (2007) *Int J Mass Spectrom* 260:177–184.
84. Orjala J, Gerwick WH (1996) *J Nat Prod* 59:427–430.
85. Sitachitta N, Rossi J, Roberts MA, Gerwick WH, Fletcher M, Willis CL (1998) *J Am Chem Soc* 120:7131–7132.
86. Sitachitta N, Marquez BM, Williamson RT, Rossi J, Roberts MA, Gerwick WH, Nguyen V-A (2000) *Tetrahedron* 56:9103–9113.
87. Chang Z, Flatt P, Gerwick WH, Nguyen V-A, Willis CL, Sherman DL (2002) *Gene* 296:235–247.
88. Flatt PM, O'Connell SJ, McPhail KL, Zeller G, Willis CL, Sherman DL, Gerwick WH (2006) *J Nat Prod* 69:938–944.
89. Galonic DP, Vaillancourt FH, Walsh CT (2006) *J Am Chem Soc* 128:3900–3901.
90. Faulkner DJ, Unson MD, Bewley CA (1994) *Pure Appl Chem* 66:1983–1990.
91. Flowers AE, Garson MJ, Webb RI, Dumdei EJ, Charan RD (1998) *Cell Tissue Res* 292:597–607.
92. Flatt PM, Gautschi JT, Thacker RW, Musafija-Girt M, Crews P, Gerwick WH (2005) *Mar Biol* 147:761–774.
93. Cheng Q, Xiang L, Izumikawa M, Meluzzi D, Moore BS (2007) *Nat Chem Biol* 3:557–558.
94. Dunlap WC, Battershill CN, Liptrot CH, Cobb RE, Bourne DG, Jaspers M, Long PF, Newman DJ (2007) *Methods* 42:358–376.
95. Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S, Wu D, Eisen JA, Hoffman JM, Remington K, et al. (2007) *PLoS Biol* 5:e77.
96. Lasken RS (2007) *Curr Opin Microbiol* 10:1–7.
97. McAlpine JB, Bachmann BO, Piraeae M, Tremblay S, Alarco A-M, Zazopoulos E, Farnet CM (2005) *J Nat Prod* 68:493–496.
98. Menzella HG, Reisinger SJ, Welch M, Kealey JT, Kennedy J, Reid R, Tran CQ, Santi DV (2006) *J Ind Microbiol Biotechnol* 33:22–28.
99. Kodumal SJ, Patel KG, Reid R, Menzella HG, Welch M, Santi DV (2004) *Proc Natl Acad Sci USA* 101:15573–15578.
100. Smith HO, Hutchison CA, III, Pfannkoch C, Venter JC (2003) *Proc Natl Acad Sci USA* 100:15440–15445.