Identification of the Putative Bryostatin Polyketide Synthase Gene Cluster from "*Candidatus* Endobugula sertula", the Uncultivated Microbial Symbiont of the Marine Bryozoan *Bugula neritina*

Sebastian Sudek,[†] Nicole B. Lopanik,[‡,§] Laura E. Waggoner,[†,§] Mark Hildebrand,[†] Christine Anderson,[†] Haibin Liu,[‡] Amrish Patel,[†] David H. Sherman,[‡] and Margo G. Haygood*[,†]

Scripps Institution of Oceanography, Marine Biology Research Division, and Center for Marine Biotechnology and Biomedicine, University of California San Diego, La Jolla, California 92093-0202, and Department of Medicinal Chemistry, Life Sciences Institute, University of Michigan, Ann Arbor, Michigan 48109-2216

Received July 21, 2006

The bryostatins are protein kinase C modulators with unique structural features and potential anticancer and neurological activities. These complex polyketides were isolated from the marine bryozoan *Bugula neritina*, but recent studies indicate that they are produced by the uncultured symbiotic bacterium "*Candidatus* Endobugula sertula" ("*E. sertula*"). Here we present the putative biosynthetic genes: five modular polyketide synthase (PKS) genes, a discrete acyltransferase, a β-ketosynthase, a hydroxy-methyl-glutaryl CoA synthase (HMG-CS), and a methyltransferase. The cluster was sequenced in two closely related "*E. sertula*" strains from different host species. In one strain the gene cluster is contiguous, while in the other strain it is split into two loci, with one locus containing the PKS genes and the other containing the accessory genes. Here, we propose a hypothesis for the biosynthesis of the bryostatins. Thirteen PKS modules form the core macrolactone ring, and the pendent methyl ester groups are added by the HMG-CS gene cassette. The resulting hypothetical compound bryostatin 0 is the common basis for the 20 known bryostatins. As "*E. sertula*" is to date uncultured, heterologous expression of this biosynthetic gene cluster has the potential of producing the bioactive bryostatins in large enough quantities for development into a pharmaceutical.

Marine invertebrates, particularly sessile ones lacking physical defenses, are rich sources of bioactive compounds. The number of compounds characterized from marine sources has increased significantly in recent years.[1] Often symbiotic microorganisms have been proposed as the true producers due to structural similarities with known microbial compounds, and in a few cases, convincing arguments for marine compounds produced by symbiotic bacteria have been made.[2−5] Many bioactive compounds have strong cytotoxic, anti-inflammatory, antimicrobial, or antiviral activity, but the development into usable drugs is typically hampered by the scarcity of source material and/or the exceedingly low concentration of the compound in the source organism. Large-scale isolation of compounds from natural sources is often environmentally detrimental and cost-prohibitive. Perhaps in part due to the scarcity of these compounds in nature, to our knowledge, only one marine natural product is in routine clinical use at this time [Prialt (Elan Pharmaceuticals, Inc.), the synthetic equivalent of a naturally occurring conopeptide]. There are four possible alternative routes to provide sufficient amount of compounds: (i) aquaculture of the source animal, (ii) direct culturing of the symbiont, (iii) total chemical synthesis, and (iv) heterologous expression of biosynthetic genes in a suitable host. All four methods have different potential advantages and drawbacks, and no one approach will likely be the solution for all compounds.

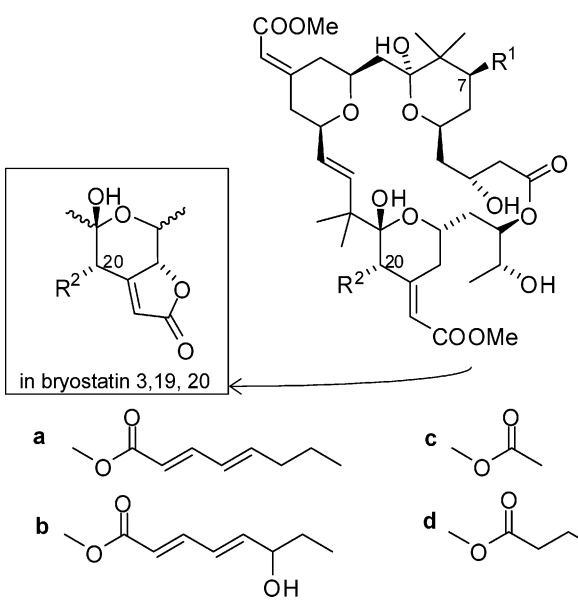
The bryostatins are cyclic polyketides isolated from the temperate marine bryozoan *Bugula neritina*.[6,7] Their common feature is a 25-membered macrolactone ring with three component pyran rings (Figure 1). Bryostatin 1 has activity against a variety of cancer cell lines (pancreatic and renal cancer, leukemia, non-hodgkin's lymphoma, melanoma, and others) and is in a number of phase II/III clinical trials (www.clinicaltrials.gov) alone and in combination with other drugs.[8−10] The effects of bryostatin 1, which modulates the activity of protein kinase C, is markedly different in different cell lines (reviewed in ref 8). More recently bryostatin 1 has shown beneficial effects in countering depression and dementia in a rat model system,[11] as well as enhancing long-term memory in a mollusk species.[12] One of the ecological roles of the bryostatins is to protect the host larvae from predators.[13]

There are three recognized sibling species of *Bugula neritina*, distinguishable by their mitochondrial cytochrome oxidase I gene sequence: deep (found on the West Coast of the United States), shallow (found on both East and West Coasts), and Northern Atlantic (found on the East Coast). The deep and shallow species contain strains of a bacterial symbiont "*Candidatus* Endobugula sertula" ("*E. sertula*") that differ by <0.6% in 16S rRNA sequence.[14,15] In contrast, the North Atlantic sibling species does not appear to have any "*E. sertula*" strains, as all bacterial 16S rRNA sequences isolated were 10−11% different from the deep and shallow strains.[14] The sibling *B. neritina* species vary in their bryostatin composition. Of the 20 described bryostatins, most diversity is found in the C-7 and C-20 position (Figure 1). The deep species is normally found below ~10 m and contains bryostatins with an octa-2,4-dienoate substituent at C-20 (among them, the clinically relevant bryostatin 1); the shallow species, found above ~10 m, contains bryostatins without this substituent. The third species, Northern Atlantic, which does not have "*E. sertula*", also does not appear to have bryostatins (20, N. Lopanik unpub. data).

Complex polyketides such as the bryostatins are synthesized by modular polyketide synthases (PKS). Modular PKSs are typically only found in bacteria, making a bacterial symbiont of *B. neritina* a prime candidate for the true biosynthetic source of the bryostatins. "*E. sertula*" is consistently associated with *B. neritina*, has not been found elsewhere in the surrounding water column, and is transferred vertically between generations.[16] Elimination of "*E. sertula*" with antibiotics leads to *B. neritina* larvae in the next generation without bryostatin.[4,13] Messenger RNA corresponding to a PKS gene fragment from the cluster described here has been identified and localized to "*E. sertula*" cells.[4] More recently, an entire PKS gene, *bryA*, has been cloned from the "*E. sertula*"/*B. neritina* association; *bryA* has four PKS modules consistent with the early steps of

* Corresponding author. Tel: (503)748-1993. Fax: (503)748-1464. E-mail: haygoodm@ebs.ogi.edu.
[†] Scripps Institution of Oceanography.
[‡] University of Michigan.
[§] These authors contributed equally.

10.1021/np060361d CCC: $37.00
Published on Web 12/22/2006



bryostatin	R^1	R^2
1	c	a
2	OH	a
3	c	a
4	e	d
5	e	c
6	d	c
7	c	c
8	d	d
9	c	d
10	e	H
11	c	H
12	d	a
13	d	H
14	e	OH
15	c	b
16	e	H
17	e	H
18	e	H
19	e	d
20	e	H

Figure 1. Structure of bryostatins, modified from ref 4. Bryostatin 1 is currently in clinical trials for a variety of cancer treatments.

bryostatin biosynthesis.[17] On the basis of the available structures of the different bryostatins, we hypothesize that a common precursor to all bryostatins is produced by a PKS in "*E. sertula*".

PKSs are large multimodular enzymes that elongate and modify the nascent polyketide in an assembly line fashion. Each module performs one specific elongation step by adding an acyl-CoA precursor and then transferring the polyketide chain to the next module. Each module has several catalytic domains. A typical type I PKS core module consists of a β-keto acyl synthase (KS) domain, an acyltransferase (AT) domain, and an acyl carrier protein (ACP) domain. The AT domain selects and loads the extender unit onto the ACP domain. The KS condenses the nascent polyketide received from the upstream ACP onto the ACP-bound acyl extender unit in a decarboxylative reaction, effectively adding two carbons to the polyketide. In most characterized PKSs, the AT functionality is integrated into each PKS module. However, in several systems, there are no AT domains within the PKS genes, and the AT domain resides on a discrete and separate gene in close proximity to the PKS genes.[18−21] In the leinamycin system, the AT domain was shown to be able to load the predicted extender unit (malonyl-CoA) onto ACP domains excised from the PKS genes, demonstrating that the discrete AT has the capability of functioning and interacting with the PKS.[18] Each module can contain different accessory domains. After extension, the β-keto group can be reduced successively to a hydroxyl group by a ketoreductase (KR) domain, to a *trans*-double bond by a dehydratase (DH) domain, and to the fully saturated β-carbon by an enoyl reductase (ER) domain.[22,23] A number of other domains are occasionally found embedded in PKS modules; for instance, methyl groups are added to the polyketide carbon backbone of yersiniabactin and epothilone by methyltransferase (MT) domains.[24,25] In most cases, there is collinearity between the order of modules on the gene and the order of the biosynthetic reactions. However, there are several examples where domains or entire modules are skipped or used more than once, suggesting that exceptions to the rule of collinearity are not uncommon.[19,26−29]

Here we present the entire ~80 kbp gene cluster from both the shallow and deep forms of *B. neritina* proposed to code for the biosynthetic machinery to make a common precursor of the bryostatins.

Results and Discussion

The *bry* Gene Cluster: Overview. The construction and screening of "*E. sertula*"-enriched *B. neritina* DNA libraries has previously been reported.[17] Briefly, a gene fragment from "*E. sertula*" amplified using degenerate primers for KS domains was used to identify four overlapping cosmids of a *B. neritina* DNA library. Due to instability upon propagation, these cosmids could not be characterized. Subcloned fragments of the cosmid inserts were used as probes to repeatedly screen two libraries under high-stringency conditions: a λ phage library of California deep, "*E. sertula*"-enriched *B. neritina* and a North Carolina shallow *B. neritina* fosmid library. The clones identified were sequenced, assembled, and annotated. The identified genes cluster with the modular PKS gene *bryA* we reported previously[17] and together constitute the entire putative bryostatin biosynthesis cluster. With one notable exception (see below) the arrangement of the genes is identical between deep and shallow strains of "*E. sertula*". The sequence identity is high (98.8%). A draft sequence was also obtained from California shallow *B. neritina* and proved to be virtually identical to North Carolina shallow (data not shown).

The *bry* gene cluster has two parts: five genes (*bryA-D* and *X*) totaling 71 kb coding for modular PKSs and four genes (*bryP-S*) totaling 6 kb with accessory functions, two acyltransferase domains on a single ORF, a discrete β-ketoacyl synthase, a β-hydroxy-β-methyl-glutaryl CoA synthase, and a methyltransferase (Figure 2 and Table 1, function proposed from BLAST searches). Both parts of the cluster have opposing direction of transcription. The PKS genes have very small intergenic regions (0−6 bp) except for *bryD* to *bryA* (134 bp). In CA/NC shallow "*E. sertula*" the accessory genes are directly upstream (202 bp) of the large PKS genes but in opposite orientation. In CA deep "*E. sertula*" a transposase lies directly upstream of the PKS, and although the accessory genes are present, their position relative to the PKS is unknown. They are not within ~30 kb upstream or ~5 kb downstream of the PKS genes as determined by draft quality sequencing of these regions (data not shown). There is a second transposase upstream of *bryP-R* in deep "*E. sertula*". It is interesting to note that the region directly upstream of *bryB* and *bryP* is extremely AT-rich (~75% over 200 nucleotides). The transposase genes found directly upstream of the two gene sets *bryABCDX* and *bryPQRS* in deep "*E. sertula*" strongly suggest transposons as the cause of rearrangement. Since the two transposases upstream of *bryB* and *bryP* are not identical, multiple transposition events must have taken place. The acyltransferase activity of *bryP* is integral for PKS function, so we assume the gene arrangement found in shallow "*E. sertula*" with all *bry* genes in one locus is ancestral. The fragmentation of a PKS pathway

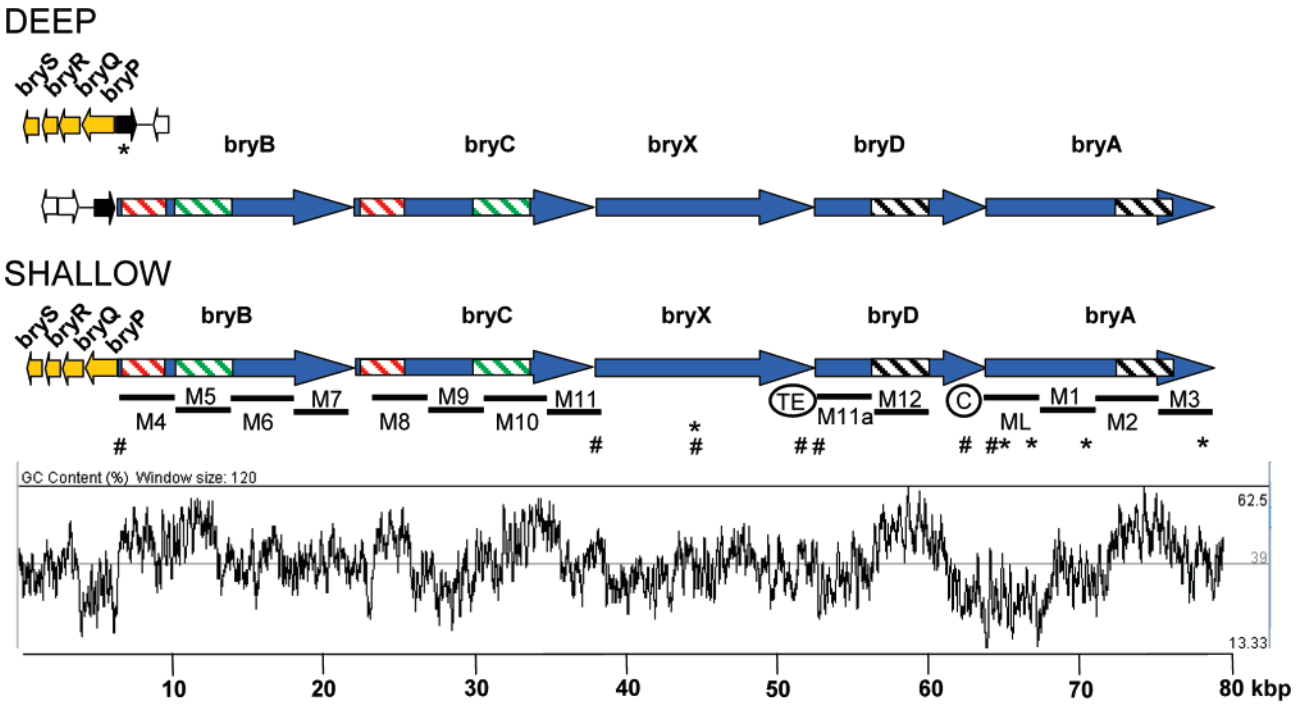


Figure 2. The *bry* cluster in deep and shallow "*E. sertula*". Modular polyketide synthase genes in blue, genes with accessory function in yellow, and transposases in black. Genes with primary metabolism homologues in open arrows. The red-, green-, and black-striped regions are perfect repeats. Black bars are PKS modules proposed to be involved in bryostatin biosynthesis. The two possible off-loading domains, a thioesterase in *bryX* and a NRPS condensation domain in *bryD*, are circled (not to scale). Transcripts were detected by RT-PCR (∗) or RPAs (#) in native *B. neritina* RNA samples. The %GC plot shows higher values in repeats. Average, maximum, and minimum %GC (for a 120 bp window) are shown on the right.

Table 1. Genes and PKS Domains in the *bry* Cluster[a]

gene	size (amino acids)	predicted catalytic domains and modules	homologues (by blastX)	expectation value
bryA	4888	L: DHh-KRh-FkbH-ACP 1: KS-KR-ACP 2: KS-KR-ACP 3: KS-(*O-MT*)-ACP		
bryB	5511	4: KS-*(DH)*-MT-ACP 5: KS-DH-KR-ACP 6: KS-DH[b]-KR-ACP 7: KS-(*O-MT*)-ACP	*B. subtilis* PKS *Paederus fuscipes* symbiont *ped* cluster	∼1e−160 (for KS domains)
bryC	5381	8: KS-DH-PS-KR-ACP 9: KS-*(DH)*-MT-ACP 10: KS-KR-ACP 11: KS-ACP		
bryX	4396	(MT-ACP) (KS-DH-ACP) (KS-ACP) TE-KS		
bryD	3476	11a: KR-ACP 12: KS-DH[b]-KR-ACP KS-ACP-NRPS		
bryP	632	AT-AT	*B. subtilis* pksC	2e−75
bryQ	405	KS	*B. subtilis* pksF	2e−124
bryR	419	HMG-CS	*B. subtilis* pksG	1e−163
brySS	364	MT	*Vibrio vulnificus* SAM-dependent methyltransferase	3e−103

[a] KS = *β*-ketosynthase, KR = ketoreductase, DH = dehydrogenase, ACP = acyl carrier protein, MT = methyltransferase, TE = thioesterase, AT = acyltransferase, HMG-CS = *β*-hydroxy-*β*-methyl-glutaryl CoA synthase, PS = pyran synthase. The loading module (L) has some unique domains; DHh and KRh are DH and KR-like sequences; together with the FkbH domain they are proposed to catalyze the formation of the unusual D-lactate starter unit as described previously.[17] Module 8 contains a unique domain, named pyran synthase (PS), proposed to form the tetrahydropyran ring from C11−C15. [b] Active site mutated, proposed to be inactive (*italics*), role in bryostatin biosynthesis unclear; see discussion for details

is not unprecedented. The *ped* genes are found in three loci in the *Paederus fuscipes* symbiont genome,[2,30] while the closely related *onn* genes[2] are clustered. The *ped* loci are also bordered by transposase-like sequences. A symbiotic lifestyle generally leads to genome degradation.[31] It is possible that the strictly symbiotic lifestyle of "*E. sertula*" has also led to the fragmentation of the putative bryostatin biosynthesis pathway.

A peculiar feature of the PKS genes is the presence of three large perfect repeats ranging in size from 2.3 to 3.5 kbp. The first repeat spans the first quarter of *bryB* and *C*, and the second repeat

is directly adjacent to the first in *bryB*, but about 6 kb apart in *bryC*. The third repeat is found in the latter half of *bryD* and *A* (Figure 2). None of the repeat regions span ORFs. While the GC% of the cluster is generally low at ~40%, it is notably higher in the repeat regions (~49%). The arrangement of genes and the position of the repeats have been confirmed by a series of Southern blots and long-range PCRs (data not shown).

Downstream of the PKS cluster we found ~3 kb of truncated ORFs and hypothetical protein genes followed by a glutathione reductase gene, defining one end of the *bry* cluster. The gene density is notably lower in this area compared to the *bry* cluster. In the deep CA strain of "*E. sertula*", there is a transposase directly upstream of the five large PKS genes *bryA*−*D* + *X* (Figure 2). Upstream of the transposase there are a number of potential pseudogenes, primary metabolism genes and relatively long intergenic regions, but no homologies to *bryP*−*S* or the sequences surrounding these genes in the shallow "*E. sertula*" strains. In both strains of "*E. sertula*", *bryS* is followed by a series of oxidoreductase genes of unknown function (Figure 2).

Domain Analysis. The best sequence matches for the entire *bry* PKS cluster are the *Bacillus subtilis pksM* genes[32] and the *Paederus fuscipes* symbiont *ped* genes.[21] BlastP[33] analysis on the NCBI server (www.ncbi.nlm.nih.gov/blast) allowed the definition of catalytic domains and modules within the "*E. sertula*" PKS genes. *bryA* contains four modules thought to initiate bryostatin biosynthesis and has been described previously.[17] *bryB* and *bryC*, the first two of the five large PKS genes, also contain four modules each (M4−11). *bryX* and *D* have two modules each as well as a number of single domains (see Table 1). We propose that the first KR-ACP (M11a) and the first complete modules of *bryD* (M12), as well as either the NRPS condensation domain in *bryD* or the TE domain in *bryX*, are used in bryostatin biosynthesis (Figure 2), while the other domains of *bryX* have unknown functions.

The large repeats are unprecedented in PKS genes to our knowledge. The first repeat in *bryBC* (Figure 2) contains only a single KS domain. The second repeat contains a DH, MT, ACP, and KS. The third repeat in *bryDA* contains the C-terminal region of a KS, DH, KR, ACP and the *N*-terminal region of another KS. Repeats are generally thought of as regions of increased genetic recombination, and this domain content could allow "re-shuffling" of the PKS modules. For example, if the second repeat was deleted, the preceding KS of module 4 would be in the right position to interact with the now orphaned reductive domains of module 5. On the other hand, if this repeat was inserted at the right position between a KS and an ACP in a hypothetical module, the newly fused gene would then contain two fully fledged modules, where the KS of the original module would need to interact with the repeat's ACP, and the repeat's KS would interact with the original ACP. The boundary of the third repeat is within a KS domain. Judging by the position of catalytic residues relative to the repeat's boundary, deletion or insertion events would not lead to new functional KS domains.

One interesting aspect of the repeats found in the *bry* cluster is that KS domains are generally thought to be relatively specific to the chain length of their substrate.[34−36] We propose that KSs identical in sequence perform in module 4/8 and module 5/10, interacting with radically different substrates in both cases. It is possible that the formation of the tetrahydropyran rings (see below) can mask the true chain length and allow identical KSs to catalyze reactions on such different substrates.

The *bry* PKS genes lack integrated AT domains in each module. BryP could function as a discrete acyltransferase, bringing a malonyl CoA extender unit to each of the 12 modules. This type of AT mechanism has been proposed in the closely related *ped* cluster and demonstrated biochemically in leinamycin biosynthesis.[18] BryP contains two distinct AT domains: the first domain has the complete active site residues GHS...R; in the second domain R is mutated to Q. The specificity motif defined for integrated ATs (YASH for methylmalonyl-CoA extension and HAFH for malonyl-CoA extension[37]) is only partially present, but closer to the expected extender unit malonyl-CoA (see Supporting Information).

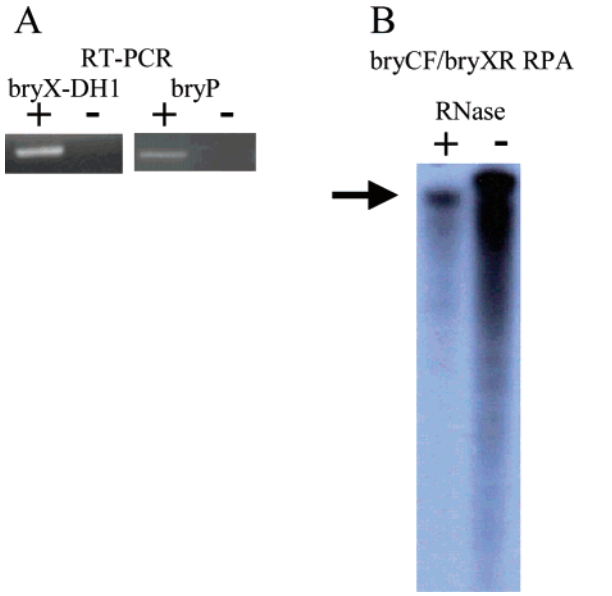


Figure 3. Representative data for detection of transcription summarized in Figure 2. (A) RT-PCR products from *bryX* (shallow "*E. sertula*"/*B. neritina* RNA) and *bryP* (deep "*E. sertula*"/*B. neritina* RNA) in which reverse transcriptase was included (+), showing the presence of transcript, or omitted (−), demonstrating the absence of DNA contamination, in the RT-PCR protocol. (B) BryCF/BryXR RPA products in which RNase A/T1 was included (+), demonstrating the presence of transcript, or omitted (−), showing the full length of the labeled RNA probe.

***bry* Cluster Transcription.** Two very sensitive techniques, reverse transcription PCR (RT-PCR) and ribonuclease protection assays (RPAs), were used to detect transcripts from the *bry* cluster in RNA preparations from *B. neritina*. The presence of *bryX* transcripts in shallow *B. neritina* and *bryP* transcripts in deep *B. neritina* was demonstrated with RT-PCR (Figure 3A). *bryB*, *bryC*, *bryX*, *bryD*, and *bryA* transcripts were detected in shallow-type *B. neritina* RNA with RPAs (*bryC* and *bryX* in Figure 3B as examples). These results are consistent with transcription of the entire *bry* cluster in *B. neritina*.

Bryostatin Biosynthesis. We have shown previously that the *bry* PKS cluster is the only large PKS cluster present in either the shallow or the deep strains of "*E. sertula*".[17] The presence of such large open reading frames in an otherwise degenerate genome (as judged from ~35 kb of DNA sequence surrounding the PKS cluster) and the demonstration of transcription throughout the region suggest that these genes are functional. We propose the following model for the biosynthesis of a bryostatin precursor (Figure 4): Biosynthesis starts with *bryA* as described previously.[17] The next four modules (M4−7) come from *bryB*, the first PKS gene in the cluster extending the chain from 9 to 17 carbons. A methyltransferase domain in M4 introduces the first characteristic *gem*-dimethyl group. The successive use of a single MT domain has also been suggested in pederin[21] and yersiniabactin biosynthesis;[24] in the latter it is supported by biochemical evidence. In addition to the MT, M4 also contains a DH, which lacks the HxxxGxxxxP signature motif and, therefore, is proposed to be inactive. M9, where the second *gem*-dimethyl group is introduced (see below), has a similarly modified DH also linked to an MT, while M3 and M7, which both also have MT domains, do not.

bryC also contains four modules (M8−11), which extend the chain to 25 carbons. M8 contains KS-DH-PS-KR-ACP; the KR and DH introduce a second double bond. There is another domain

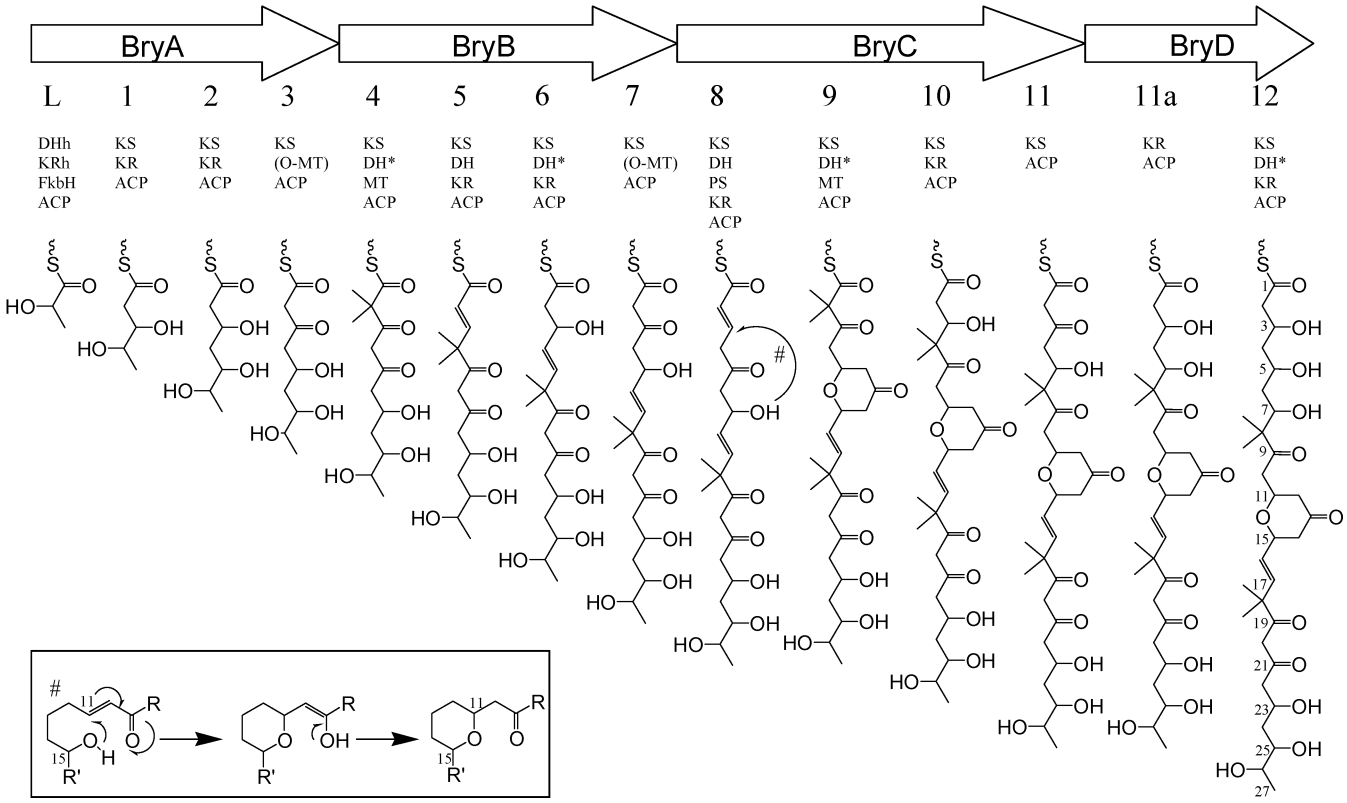


Figure 4. Proposed pathway for bryostatin biosynthesis. Functional arrangement of domains and the intermediate polyketides are shown. Arrows denote proteins, PKS modules are designated by L and 1−12, and their domains are listed. Box details the formation of the pyran ring via Michael reaction. See Table 1 for abbreviations. *****Active site mutated, proposed to be inactive. (O-MT) in MT3 and MT7 may be involved in HMG-CS mechanism; see discussion for details.

that has 26% identity over ~500 amino acids to a region in module 5 of *ped*F annotated as a dehydratase domain, which we have termed the pyran synthase (PS). We propose that this region catalyzes the formation of the tetrahydropyran ring that pederin and bryostatin have in common, as the dehydratase domain in module 5 of *ped*F and the PS region in module 8 of *bryC* are both in the correct position to form this group. Analogous to Piel,[21] we propose the bryostatin pyran ring from C-11 to C-15 is formed by a Michael-type reaction catalyzed by the newly proposed pyran synthase domain. The nucleophilic oxygen on C-15 attacks the double bond between C-11 and C-12, leading to an intermediate state with a shifted double bond and a hydroxy group on C-9, which can then rearrange to form the functionality found in bryostatin (see box Figure 4). The second *gem*-dimethyl group could be added onto the nascent polyketide by the MT domain embedded in M9.

As the function of the next gene, *bryX*, is unclear, we hypothesize that it is skipped (see below). After module 11, the first incomplete module of *bryD* consisting of KR and ACP (M11a) would reduce the β-keto group left untouched by module 11 proper. While this "division of labor" with one module doing the extension and the next doing the reductive step is very unusual, it could be achieved if there is an equilibrium between the ACP in modules 11 and 11a that allows the transfer of the polyketide directly from one ACP to the next without a ketosynthase step. Alternatively, since the KR of M11a is on the N-terminal end of BryD, it could reduce the keto group without the need to transfer the polyketide to the ACP of M11a. The transfer to the next regular module (M12) could still be achieved by an appropriate spatial arrangement of the catalytic domains. PKS genes starting with KR-ACP are also found in the *pks*2 cluster of *Bacillus amyloliquefaciens*.[19] The next complete module on *bryD* is proposed to do the final extension step. We propose that the NRPS condensation domain in *bryD* together with an adjacent KS (as in cyclosporin reviewed in ref 38) or the TE domain in *bryX* then cyclizes the molecule from C-1 to C-25. Two five-carbon pyran rings form by the condensation of a keto and a hydroxyl group from C-19 to C-23 and C-5 to C-9.

One interesting feature of the bryostatins is the presence of two exocyclic olefinic bonds on C-13 and C-21. These functionalities are proposed to be synthesized by the products of *bryR* and possibly *bryS*. It is unclear whether this happens as the polyketide chain is elongating or after cyclization and release. Because these groups are present on all bryostatins [although it is further modified in bryostatins 3, 19, and 20 (Figure 1)], it seems likely that this occurs on the growing polyketide chain. In addition, these modifications occur at the last module of BryA and BryB, where the polyketide chain may be physically more accessible for modification by the HMG-CoA synthase, BryR. We hypothesize that BryR condenses acetyl-CoA onto the β-keto groups at C-13 and 21. A similar function for HMG-CoA synthase homologues has been proposed in jamaicamide[39] and mupirocin[20] biosynthesis. After dehydration, the carboxylic acid group would be *O*-methylated either by the embedded methyltransferases in M3 and M7 or by BryS (Figure 5). In a phylogenetic analysis, the embedded MTs in M3 and M7 group with other PKS-embedded *O*-MTs as opposed to *C*-MTs (see Supporting Information). The function of the stand-alone KS, BryQ, is unclear. Recent research suggests that these stand-alone KS groups decarboxylate a malonyl-ACP/PCP (peptidyl carrier proteins analogous to ACP in nonribosomal peptide synthases) domain into an acetyl-ACP/PCP domain, which is then used by the HMG-CS as a substrate to acylate the β-keto group.[40] However, there is no apparent stand-alone ACP/PCP domain present in this gene cluster, and BryR, the HMG-CoA synthase, may be able to directly use acyl-CoA substrates. Whether BryQ could decarboxylate an acyl-CoA is unclear. The resulting compound bryostatin 0 (Figure 5) is hypothetical, but is a plausible common basis for the 20 known bryostatins. In all bryostatins (with the exception of bryostatin 3, 19, and 20, which undergo an additional lactonization), further modifications are restricted to C-7 and C-20 (Figure 1). The

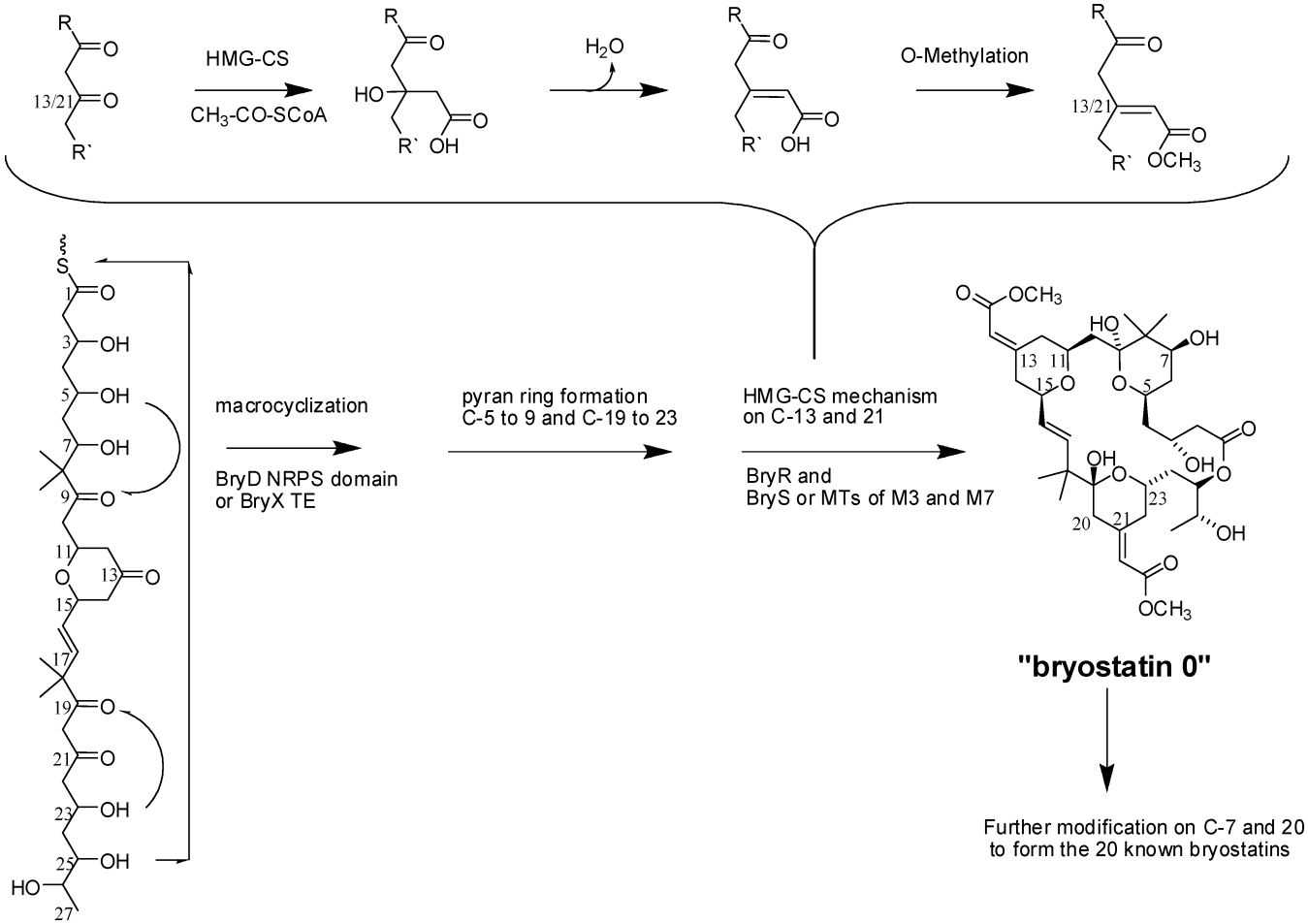


Figure 5. Maturation steps. The order of maturation steps is unknown; they may also take place during polyketide synthesis. Top row: HMG-CS pathway to form exocyclic olefinic bonds.

mechanism (or even the physical location) of further maturation into these compounds remains to be investigated and could be mediated by either "*E. sertula*" or its host.

The model described here implies that the rules for collinearity between genes and their proteins in polyketide biosynthesis are not absolute. Examples of noncollinear polyketide synthesis include avermectin, an anthelmintic compound produced by *Streptomyces avermitilis*,[41] and rapamycin, an immunosuppressant produced by *S. hygroscopicus*.[42] The proposed order of biosynthesis for bryostatin is different from the gene order: *bryA*, the last gene in the cluster, is the first gene in the biosynthetic scheme, and *bryX* is skipped. The presence of an uninterrupted ORF of this size in both strains of "*E. sertula*" as well as our RT-PCR data show that this gene is important. However, the domain content of *bryX* is enigmatic, since it contains PKS domains, which are not organized as conventional modules. A few functions required to produce complete bryostatins are not apparent in the cluster, including the formation of C7/C20 substituents and phosphopantetheinylation of the ACP domains, but the domain structure of *bryX* does not appear to be appropriate for these functions (except for the TE domain as a potential longer-chain acyltransferase adding the carboxylic acids found ester-linked to C-7/C-20 in some bryostatins). Also, *bryX* is identical in both deep and shallow "*E. sertula*", which consistently have different suites of bryostatins.

Functional heterologous expression of the entire *bry* cluster would provide the hypothetical compound bryostatin 0 (Figure 5), differing from bryostatin 1 in C-7 and C-20 substituents. This compound contains all proposed pharmacophore elements.[43−45] C-7 has been shown to be irrelevant for pharmaceutical activity. C-20 substituents have recently been proposed as "tunable elements" to modify bryostatin potency.[46] A chemical synthesis pathway for C-20 modification of bryostatin 0 would presumably be difficult due to steric congestion, leading to undesired side products.[46] It may be possible to simply overpower this problem with an inexpensive and basically unlimited supply of bryostatin 0 followed by separation techniques to isolate products with the correct bryostatin 1 substitution at C-20. Alternatively, identification and functional expression of the additional biosynthetic genes from either host or symbiont would allow the direct production of authentic bryostatin 1.

The sequence data presented are the first step toward solving the bryostatin supply problem using a biosynthetic route. There are substantial technical hurdles to overcome: the *bry* PKS cluster is enormous in size, which makes cloning challenging and functional expression of the entire cluster daunting. Also as discussed above, some genes in bryostatin biosynthesis have not been identified. In spite of these potential difficulties, expressing the biosynthetic genes in a heterologous host would be very cost-effective to provide a virtually unlimited supply of authentic bryostatin. The *bry* genes also have the potential to be used in a combinatorial biosynthesis approach. By deleting or modifying modules or domains variations of bryostatin could be produced and complement the chemically synthesized "bryologues" currently being developed. This compares favorably with the other potential routes to bryostatin: while technically feasible, aquaculture of *B. neritina* to produce bryostatins has not proven to be cost-effective.[47] Direct culturing of obligate-symbiotic bacteria is notoriously difficult because typically it is challenging for the bacteria to adapt to conditions outside of the narrow range provided by the host. No exception to this rule, "*E. sertula*" has proven unculturable so far. While total chemical

Table 2. Samples Examined in This Study[a]

sample location	*"E. sertula"* strain	continuous *bry* cluster	*bry* gene accession number	identity of *bry* genes
Torrey Pines Artificial Reef II, La Jolla, CA	deep	no	DQ889941 and DQ889942	100%
Mission Beach, San Diego CA	?	?	N.A.	(~98%)
Scripps Pier, La Jolla, CA	shallow	yes	N.A.	(~98%)
Radio Island Jetty, Morehead City, NC	shallow	yes	EF032014	98%

[a] The identity percentage is relative to the Torrey Pines sample. The Mission Beach and Scripps Pier samples were only partially sequenced, so the identity percentage does not apply to the entire gene cluster.

synthesis of bryostatin is possible,[48] its practical use is limited because of the >70 steps needed. In our opinion, the synthesis of simpler analogues based on the pharmacologically active part of the molecule[49] is the other promising route. It has already led to a number of compounds with interesting activities.[50]

Conclusions. The *bry* cluster described in this paper is an example of a PKS cluster found in a marine invertebrate that is implicated in the production of a bioactive compound originally isolated from the host animal. While the organization of the genes and lack of appropriately sized PKS genes elsewhere in the genome make it likely that the *bry* cluster is responsible for bryostatin production, there is no direct evidence. This would require either culturing and mutational studies of "*E. sertula*" or heterologous expression of all (or at least significant parts) of the *bry* genes. As described above, both of these tasks are formidable. The *bry* genes share a surprising number of features with the pederin/onamide clusters: domains from these clusters are usually the best hits in blastP searches after the *B. subtilis pks* genes; both clusters have methyltransferases, discrete ATs, and, in the case of onnamide, HMG-CoA synthases. They also share a domain that presumably is responsible for the formation of a tetrahydropyran ring via a Michael-type reaction mechanism. There are three long perfect repeats in the *bry* cluster. While their borders do not exactly coincide with PKS modules, the recombination events commonly associated with repeated stretches of DNA hint at a mechanism how modular PKS genes can quickly evolve. In conclusion, the *bry* cluster has some unique and interesting features and we provide some testable hypotheses for its function, opening a rich field of further study.

Experimental Section

Sample Collection and Genotype Determination. Colonies of *Bugula neritina* were collected by scuba diving from three sites along the coast of San Diego, CA: Mission Bay (water depth ~5 m, sample not genotyped, but previous collections determined to be shallow), Scripps Pier (~ 7 m, CA shallow), Torrey Pines Artificial Reef II (~15 m, CA deep). A fourth sample was collected from Radio Island Jetty in Morehead City, NC (~7 m, NC shallow). The genotype of collected specimens was determined as described previously via a sequence polymorphism in the "*E. sertula*" 16S rRNA gene.[15] The samples are listed in Table 2.

DNA Isolation, Cloning, and Sequencing. All procedures to enrich *B. neritina* DNA preparations for "*E. sertula*" as well as the construction of the SuperCos I cosmid, λ-ZAP, and λ-DASHII (Stratagene) libraries were described previously.[17,51] Briefly, since DNA preparations from *B. neritina* contain only minor amounts of "*E. sertula*" DNA, different (ultra)centrifugation techniques were employed to enrich the symbiont DNA to approximately 10% of the total DNA. All libraries were constructed from partially digested genomic DNA according to manufacturer's protocols. From the Mission Bay-derived library four SuperCos clones spanned most of the *bry* cluster, leaving two gaps. However, due to stability issues, the entire cluster was recloned into λ-DASHII (insert size ~15 kb) for CA shallow and λ-ZAP (insert size ~5 kb) for CA deep and screened using cosmid fragments as probes. Thirty-two λ-DASHII and >100 λ-ZAP clones were mapped to the *bry* cluster by end sequencing, and overlapping clones were picked and fully sequenced by primer walking. Sequencing was performed on ABI3100 sequencers at the UCSD Center for AIDS research and through SeqXcel Inc., San Diego. The CA shallow cluster was only partially sequenced (~75% coverage); the CA deep cluster was sequenced to 2−4-fold coverage. Reads were assembled with Sequencher (Gene Codes Corp). Three long perfect repeats made assembly challenging. Each repeat was covered by two overlapping clones, which contained part of the adjacent nonrepeat region.

A fosmid library from shallow NC populations of *B. neritina*/"*E. sertula*" was generated using CopyControl Fosmid Library Preparation kit (Epicentre). The library was probed with portions of KS genes that had been isolated by PCR from genomic DNA using primers based on CA shallow "*E. sertula*" sequence. Positive fosmids (2) were subcloned into pSMART-LC Kan sequencing vector (Lucigen Corp.) and end-sequenced. Because the whole cluster was not contained on these two fosmids, probes were generated from the 5′ and 3′ ends of the cluster and used to reprobe the library. The additional positive fosmids (2) were subcloned and end-sequenced. Sequences were assembled using SeqMan (Lasergene). The fosmids were sequenced to 2−5-fold coverage.

Restriction mapping and Southern blotting were performed according to established protocols as described previously.[17]

RNA Isolation and Reverse Transcription PCR (RT-PCR). Enriched bacterial fraction RNA was isolated from Scripps Pier (shallow) and Torrey Pines Artificial Reef (deep) adult *B. neritina* as described previously.[17] RNA used for RT-PCR was purified on an RNeasy mini kit spin column (QIAGEN Inc.), treated with DNase I (QIAGEN Inc.) as described by the manufacturer, and eluted with 30 μL of RNase-free water (QIAGEN Inc).

RT-PCR was performed as previously described.[17] Gene-specific primers BryX-DH1R (GGC GTT GCC CAG GCA ATA TGT TGC) or BryPR (ACG TGA ATG AAA GGC AGC GC) were used to generate cDNA. PCR was performed on the cDNA with the corresponding primer pair, BryX-DH1F (GCT TTA CCC TGC TAT CCT TTT GCC) and BryX-DH1R or BryPF (GTG GGC AGG GTT CAC AGC AC) and BryPR. To confirm that RT-PCR products reflected RNA content rather than DNA contamination, control RT-PCR reactions without reverse transcriptase were conducted for each primer pair.

Ribonuclease Protection Assay (RPA). In order to make transcription templates for radio-labeled RNA probes, PCR was performed on shallow *B. neritina* DNA with the following *bry* cluster primer pairs: bryBF (GGT GAT GCC AAG ATG ACC ACC GC) and bryBR (CAC GAT CAT GAT TTA AAC GCT G), bryCF (CTA GAT GAT GAC GAT GAG ATT G) and bryXR (CAG ACT GCA ACA TGC GTA AGG C), bryX-DH1F and bryX-DH1R, bryX-KS3F (CCA GCC TTT TGA TCA GTT CAA GTC) and bryX-KS3R (GGG ATC TCC TAG TTT AGT CCC AGT), bryDF (TTC ATT TAT GTG CAG GAC ACA TAC) and bryDR (CCC AGA AGC CGA TTC CAG ATG CCC), and bryDendF (CAG ATA AAC CTA TAG AAG AGA TTG) and bryAR (GTT TTT TTC GGT ATT GTC GAA TGC). *bry* cluster PCR products were purified from agarose gels with a Quiaex II kit (QIAGEN Inc.) and cloned into pCR4-TOPO plasmids (Invitrogen) using a TOPO TA cloning kit for sequencing (Invitrogen). Following transformation into TOP10 One Shot Chemically Competent *E. coli* (Invitrogen), growth of transformants and isolation of plasmid DNA with a QIAprep spin miniprep kit (QIAGEN Inc.), plasmids were digested with *SpeI* or *NotI* (New England Bio Labs), depending on insert orientation, so that transcription of the plasmid would terminate shortly after the insert

region. The digested plasmid was then purified by standard phenol/chloroform extraction and ethanol precipitation. ^{32}P-labeled RNA probes complementary to *bry* cluster mRNA were transcribed with T3 RNA polymerase or T7 RNA polymerase using a MAXIscript in vitro transcription kit (Ambion) and [^{32}P]UTP (Amersham Pharmacia Biotech). In order to maximize the portion of full-length transcripts in the RPA probe, the product of the transcription reaction was run on a denaturing polyacrylamide gel (6% acrylamide, 8 M urea, 1X TBE), cut out of the gel, and incubated in probe elution buffer (Ambion) overnight at 37 °C. Then 5 μg of enriched bacterial fraction RNA from shallow *B. neritina* and 8×10^{-4} cpm of labeled RNA probe were used with a Hybspeed RPA kit (Ambion) for each RPA. In the RPA, the RNA probe hybridized with its complement from the genomic RNA. Upon the addition of RNase A/T1, the single-stranded RNA was degraded, leaving only the double-stranded probe−target hybrid, which is protected. The products were run on a denaturing polyacrylamide gel (6% acrylamide, 8 M urea, 1X TBE) and exposed to X-ray film at −80 °C.

Genbank Accession Numbers. "*E. sertula*" deep *bryABCDX* and *bryPQRS* were deposited in Genbank as DQ889941 and DQ889942, respectively. The complete North Carolina "*E. sertula*" shallow *bry* cluster was deposited as EF032014.

Acknowledgment. N.B.L. was supported by a NIH NRSA postdoctoral fellowship (5F32CA110636). L.E.W. was supported by a Department of Defense postdoctoral traineeship award (DAMD17-00-1-0181). Research support was from the National Institutes of Health (5R01CA079678-03) and the Department of Defense (DAMD17-00-1-0183). We are especially grateful to C. Sheehan and E. Miller for technical assistance and E. Schmidt for helpful discussions. We would also like to thank Y. Mao, D. Wilson, and Y. Xue for assistance with building and probing the NC shallow fosmid library and N. Lindquist for facilitating collection of NC shallow *B. neritina*.

Supporting Information Available: Partial alignment of acyltransferase genes and phylogenetic analysis of methyltransferase domains. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Faulkner, D. J. *Nat. Prod. Rep.* **2002**, *19*, 1−49.
(2) Piel, J.; Hui, D.; Wen, G.; Butzke, D.; Platzer, M.; Fusetani, N.; Matsunaga, S. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 16222−16227.
(3) Schmidt, E. W.; Nelson, J. T.; Rasko, D. A.; Sudek, S.; Eisen, J. A.; Haygood, M. G.; Ravel, J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 7315−7320.
(4) Davidson, S.; Allen, S. W.; Lim, G. E.; Anderson, C. M.; Haygood, M. G. *Appl. Env. Microbiol.* **2001**, *67*, 4531−4537.
(5) Piel, J. *Nat. Prod. Rep.* **2004**, *21*, 519−538.
(6) Pettit, G. R.; Herald, C. L.; Doubek, D. L.; Herald, D. L. *J. Am. Chem. Soc.* **1982**, *104*, 6846−6847.
(7) Pettit, G. R. *Prog. Chem. Org. Nat. Prod.* **1991**, *57*, 153−195.
(8) Mutter, R.; Wills, M. *Bioorg. Med. Chem.* **2000**, *8*, 1841−1860.
(9) Blackhall, F. H.; Ranson, M.; Radford, J. A.; Hancock, B. W.; Soukop, M.; McGown, A. T.; Robbins, A.; Halbert, G.; Jayson, G. C. *Br. J. Cancer* **2001**, *84*, 465−469.
(10) Mohammad, R. M.; Adsay, N. V.; Philip, P. A.; Pettit, G. R.; Vaitkevicius, V. K.; Sarkar, F. H. *Anti-Cancer Drugs* **2001**, *12*, 735−740.
(11) Sun, M. K.; Alkon, D. L. *Eur. J. Pharmacol.* **2005**, *512*, 43−51.
(12) Alkon, D. L.; Epstein, H.; Kuzirian, A.; Bennett, M. C.; Nelson, T. J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 16432−16437.
(13) Lopanik, N.; Lindquist, N.; Targett, N. *Oecologia* **2004**, *139*, 131−139.
(14) McGovern, T.; Hellberg, M. *Mol. Ecol.* **2003**, *12*, 1207−1215.
(15) Davidson, S. K.; Haygood, M. G. *Biol. Bull.* **1999**, *196*, 273−280.
(16) Haygood, M.; Davidson, S. *Appl. Environ. Microbiol.* **1997**, *63*, 4612−4616.
(17) Hildebrand, M.; Waggoner, L. E.; Liu, H.; Sudek, S.; Allen, S.; Anderson, C.; Sherman, D. H.; Haygood, M. *Chem. Biol.* **2004**, *11*, 1543−1552.
(18) Cheng, Y.-Q.; Tang, G.-L.; Shen, B. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3149−3154.
(19) Chen, X. H.; Vater, J.; Piel, J.; Franke, P.; Scholz, R.; Schneider, K.; Koumoutsi, A.; Hitzeroth, G.; Grammel, N.; Strittmatter, A. W.; Gottschalk, G.; Sussmuth, R. D.; Borriss, R. *J. Bacteriol.* **2006**, *188*, 4024−4036.
(20) El-Sayed, A. K.; Hothersall, J.; Cooper, S. M.; Stephens, E.; Simpson, T. J.; Thomas, C. M. *Chem. Biol.* **2003**, *10*, 419−430.
(21) Piel, J. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 14002−14007.
(22) Donadio, S.; Staver, M. J.; McAlpine, J. B.; Swanson, S. J.; Katz, L. *Science* **1991**, *759*, 675−679.
(23) Hopwood, D. A.; Sherman, D. H. *Annu. Rev. Genet.* **1990**, *24*, 37−66.
(24) Miller, D. A.; Luo, L.; Hillson, N.; Keating, T. A.; Walsh, C. T. *Chem. Biol.* **2002**, *9*, 333−344.
(25) Molnar, I.; Schupp, T.; Ono, M.; Zirkle, R. E.; Milnamow, M.; Nowak-Thompson, B.; Engel, N.; Toupet, C.; Stratmann, A.; Cyr, D. D.; Gorlach, J.; Mayo, J. M.; Hu, A.; Goff, S.; Schmid, J.; Ligon, J. M. *Chem. Biol.* **2000**, *7*, 97−109.
(26) Xue, Y.; Wilson, D.; Sherman, D. H. *Gene* **2000**, *245*, 203−211.
(27) Gaitatzis, N.; Silakowski, B.; Kunze, B.; Nordsiek, G.; Blocker, H.; Hofle, G.; Muller, R. *J. Biol. Chem.* **2002**, *277*, 13082−13090.
(28) Wenzel, S. C.; Muller, R. *Curr. Opin. Chem. Biol.* **2005**, *9*, 447−458.
(29) Moss, S. J.; Martin, C. J.; Wilkinson, B. *Nat. Prod. Rep.* **2004**, *21*, 575−593.
(30) Piel, J.; Wen, G.; Platzer, M.; Hui, D. *ChemBioChem* **2004**, *5*, 93−98.
(31) Moran, N. A. *Cell* **2002**, *108*, 583−586.
(32) Hofemeister, J.; Conrad, B.; Adler, B.; Hofemeister, B.; Feesche, J.; Kucheryava, N.; Steinborn, G.; Franke, P.; Grammel, N.; Zwintscher, A.; Leenders, F.; Hitzeroth, G.; Vater, J. *Mol. Genet. Genomics* **2004**, *272*, 363−378.
(33) Altschul, S. F.; Gish, W.; Miller, W.; Wyers, E. W.; Lipman, D. J. *J. Mol. Biol.* **1990**, *215*, 403−410.
(34) Scarsdale, J. N.; Kazanina, G.; He, X.; Reynolds, K. A.; Wright, H. T. *J. Biol. Chem.* **2001**, *276*, 20516−20522.
(35) Holzbaur, I. E.; Ranganathan, A.; Thomas, I. P.; Kearney, D. J. A.; Reather, J. A.; Rudd, B. A. M.; Staunton, J.; Leadlay, P. F. *Chem. Biol.* **2001**, *8*, 329−340.
(36) He, X.; Reynolds, K. A. *Antimicrob. Agents Chemother.* **2002**, *46*, 1310−1318.
(37) Del Vecchio, F.; Petkovic, H.; Kendrew, S. G.; Low, L.; Wilkinson, B.; Lill, R.; Cortes, J.; Rudd, B. A.; Staunton, J.; Leadlay, P. F. *J. Ind. Microbiol. Biotechnol.* **2003**, *30*, 489−494.
(38) Keating, T. A.; Ehrmann, D. E.; Kohli, R. M.; Marshall, C. G.; Trauger, J. W.; Walsh, C. T. *ChemBioChem* **2001**, *2*, 99−107.
(39) Edwards, D. J.; Marquez, B. L.; Nogle, L. M.; McPhail, K.; Goeger, D. E.; Roberts, M. A.; Gerwick, W. H. *Chem. Biol.* **2004**, *11*, 817−833.
(40) Calderone, C. T.; Kowtoniuk, W. E.; Kelleher, N. L.; Walsh, C. T.; Dorrestein, P. C. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 8977−8982.
(41) Ikeda, H.; Nonomiya, T.; Usami, M.; Ohta, T.; Omura, S. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9509−9514.
(42) Schwecke, T.; Aparicio, J.; Molnar, I.; Konig, A.; Khaw, L.; Haycock, S.; Oliynyk, M.; Caffrey, P.; Cortes, J.; Lester, J.; Bohm, G.; Staunton, J.; Leadlay, P. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 7839−7843.
(43) Rando, R. R.; Kishi, Y. *Biochemistry* **1992**, *31*, 2211−2218.
(44) Wender, P. A.; Cribbs, C. M.; Koehler, K. F.; Sharkey, N. A.; Herald, C. L.; Kamano, Y.; Pettit, G. R.; Blumberg, P. M. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 7197−7201.
(45) Wender, P. A.; DeBrabander, J.; Harran, P. G.; Jimenez, J. M.; Koehler, M. F.; Lippa, B.; Park, C. M.; Siedenbiedel, C.; Pettit, G. R. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 6624−6629.
(46) Wender, P. A.; Baryza, J. L. *Org. Lett.* **2005**, *7*, 1177−1180.
(47) Mendola, D. *Biomol. Eng.* **2003**, *20*, 441−458.
(48) Hale, K. J.; Hummersone, M. G.; Manaviazar, S.; Frigerio, M. *Nat. Prod. Rep.* **2002**, *19*, 413−453.
(49) Wender, P. A.; Hinkle, K. W.; Koehler, M. F.; Lippa, B. *Med. Res. Rev.* **1999**, *19*, 388−407.
(50) Wender, P. A.; Clarke, M. O.; Horan, J. C. *Org. Lett.* **2005**, *7*, 1995−1998.
(51) Hildebrand, M.; Waggoner, L. E.; Lim, G. E.; Sharp, K. H.; Ridley, C. P.; Haygood, M. G. *Nat. Prod. Rep.* **2004**, *21*, 122−142.

NP060361D