Biosynthetic Pathway and Gene Cluster Analysis of Curacin A, an Antitubulin Natural Product from the Tropical Marine Cyanobacterium *Lyngbya majuscula*[†]

Zunxue Chang,[‡,||] Namthip Sitachitta,[§,||] James V. Rossi,[⊥] Mary Ann Roberts,[§] Patricia M. Flatt,[§] Junyong Jia,[‡] David H. Sherman,*[,‡] and William H. Gerwick*[,§]

Department of Medicinal Chemistry, College of Pharmacy, Department of Chemistry, and Department of Microbiology and Immunology, University of Michigan, Ann Arbor, Michigan 48109, College of Pharmacy, Oregon State University, Corvallis, Oregon 97331, and Department of Biochemistry and Biophysics, Oregon State University, Corvallis, Oregon 97331

Curacin A (**1**) is a potent cancer cell toxin obtained from strains of the tropical marine cyanobacterium *Lyngbya majuscula* found in Curaçao. Its structure is unique in that it contains the sequential positioning of a thiazoline and cyclopropyl ring, and it exerts its potent cell toxicity through interaction with the colchicine drug binding site on microtubules. A series of stable isotope-labeled precursors were fed to cultures of curacin A-producing strains and, following NMR analysis, allowed determination of the metabolic origin of all atoms in the natural product (one cysteine, 10 acetate units, two S-adenosyl methionine-derived methyl groups) as well as several unique mechanistic insights. Moreover, these incorporation experiments facilitated an effective gene cloning strategy that allowed identification and sequencing of the approximately 64 kb putative curacin A gene cluster. The metabolic system is comprised of a nonribosomal peptide synthetase (NRPS) and multiple polyketide synthases (PKSs) and shows a very high level of collinearity between genes in the cluster and the predicted biochemical steps required for curacin biosynthesis. Unique features of the cluster include (1) all but one of the PKSs are monomodular multifunctional proteins, (2) a unique gene cassette that contains an HMG-CoA synthase likely responsible for formation of the cyclopropyl ring, and (3) a terminating motif that is predicted to function in both product release and terminal dehydrative decarboxylation.

Marine cyanobacteria have yielded an amazing assortment of chemically diverse and biologically relevant natural products.[1,2] Several promising clinical candidates coming from the marine environment can be traced to metabolic processes of cyanobacteria, including dolastatin 10,[3] dolastatin 15,[4] and scytonemin.[5] In this regard, our previous investigations of a Curaçao collection of *Lyngbya majuscula* resulted in a highly active antiproliferative and cytotoxic compound, curacin A (**1**).[6] Curacin A has been shown to block cell cycle progression by interacting with the colchicine binding site on tubulin and inhibiting microtubule polymerization.[7] Simultaneously, there has been considerable interest to produce curacin A by total synthesis[8,9] and to produce synthetic as well as semisynthetic analogues to explore structure–activity relationships in this new potential drug class.[10] An elegant combinatorial synthesis approach has been applied that has resulted in a chemically stabilized curacin A derivative with improved water solubility characteristics, yet it retains curacin A's remarkable biological effect on microtubules.[11]

Analysis of the more than 200 diverse natural products reported to date from marine cyanobacteria results in structural patterns that reflect intriguing metabolic trends in this group. Marine cyanobacteria frequently generate secondary metabolites that are of mixed polyketide/nonribosomal peptide origin. This type of hybrid biosynthesis requires the integrated activity of polyketide synthases (PKS) and nonribosomal peptide synthetases (NRPS) and

1 R = CH₃, Curacin A
3 R = H, Curacin D

2 Barbamide
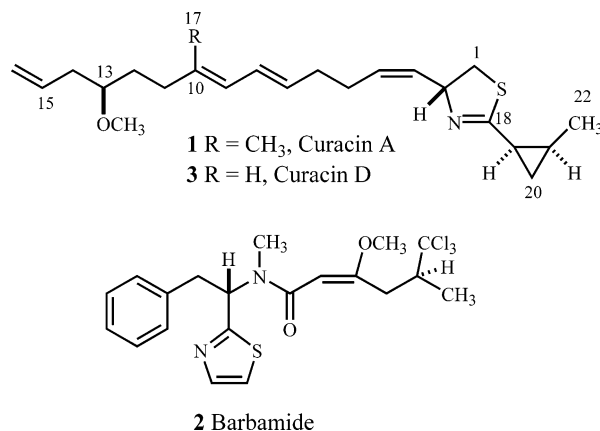
is common only in a few groups of microorganisms, most notably, the cyanobacteria and myxobacteria.[1,12] In cyanobacteria, ketide units are connected to peptide units through two distinct types of linkages: (1) the activated carboxyl functionality of amino acids is extended with acetate-derived units to produce "ketide extended amino acids", and (2) amino acids or the corresponding hydroxy acids are found in amide or ester linkage with the terminating carboxyl function of a polyketide.[1] Additionally, cyanobacteria appear to have a multitude of unusual tailoring enzymes that catalyze oxidation, methylation, and perhaps most interestingly, halogenation.[1] The functional group additions result in natural products that possess vinyl halides (chlorine, bromine, iodine), alkyl halides including dichloromethylene and trichloromethyl groups, and even an acetylenic bromide functionality.[1]

Despite the great interest in the unique structures and biological activities of marine cyanobacterial metabolites, only a few biosynthetic studies have appeared. This is a consequence of problems with current laboratory culture
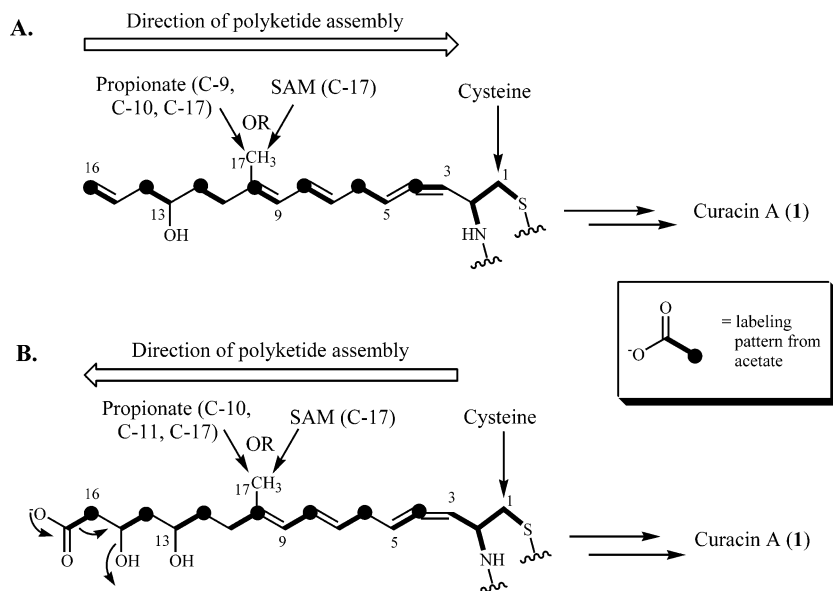
Figure 1. Two hypothetical alternative directions of assembly of the polyketide section of curacin A (**1**). (A) A 14-carbon lipid chain is proposed to be constructed beginning at C-16 that condenses with a cysteine residue with concomitant loss of the carboxyl group of the amino acid. (B) Assembly could occur by initiation with cysteine (or a modified cysteine) followed by seven cycles of polyketide extension, including an unusual terminal decarboxylation to form the C-15−C-16 olefin. Experimental results described in the text are consistent with (B).
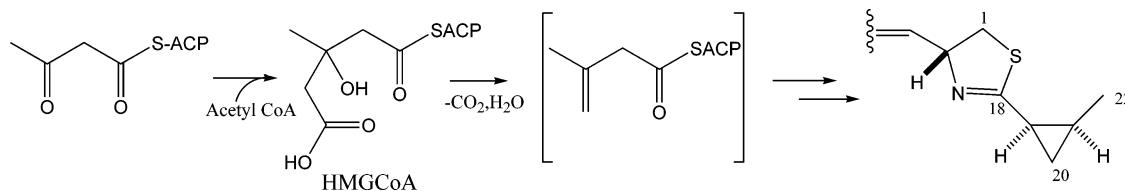

Figure 2. Proposed metabolic source of the C-18 to C-22 section of curacin A (**1**). Experimental results discussed in the text support the assembly of a branched triketide with the possible intermediacy of an HMGCoA residue.

methods, as cyanobacteria are notoriously difficult to purify from contaminating bacteria, grow slowly or not at all, have poorly defined nutritional requirements, and are still at a very formative stage of biochemical characterization.[13,14] Our previous work with the biosynthesis of the halogen-containing metabolite barbamide (**2**) represents an exception.[15−18] The Curaçao strain that produces barbamide has been maintained in culture for over 10 years and has been used extensively in biosynthetic precursor incorporation experiments. Despite its slow growth rate, this strain readily incorporates exogenously supplied isotope-labeled precursors. Precursor incorporation studies demonstrated that the biosynthetic subunits of barbamide (**2**) include leucine, phenylalanine, and cysteine combined with one acetate unit and two methionine-derived methyl groups.[15−17] Moreover, these experiments were instrumental in demonstrating the intermediacy of trichloroleucine in this biosynthetic pathway and establishing the chiral preference of the putative halogenase.[17] This subsequently led to the isolation and sequencing of the barbamide biosynthetic gene cluster and the partial biochemical characterization of several key features associated with its assembly.[18] This combined approach of pathway studies employing isotope incorporation and molecular genetics has been extraordinarily powerful in revealing novel features of the biosynthesis of barbamide (**2**) and providing insight into a number of biosynthetic processes in cyanobacteria as well as other classes of organisms. Here, we have applied a similar approach to study the biosynthesis of a co-occurring metabolite in this *L. majuscula* 19L strain, the potent antitubulin natural product curacin A (**1**).

Curacin A (**1**) is a unique natural product structure in the relationship of its two lipid chains and sequential thiazoline and cyclopropyl rings. Two hypotheses for the construction of this carbon skeleton were considered (Figure 1). In the first case (Figure 1A), we proposed that a 14-carbon lipid chain could be constructed beginning at C-16 and, similar to sphingolipid biosynthesis, that this could condense with a cysteine residue with concomitant loss of the carboxyl group of the amino acid. Alternatively, assembly of this section of curacin A (**1**) could occur by initiation with cysteine (or a modified cysteine) followed by seven cycles of polyketide extension, including an unusual terminal decarboxylation to form the C-15−C-16 olefin (Figure 1B).

In either scenario, the C-17 methyl branch is located at a C-2 carbon and thus could be derived from a propionate extender unit or from SAM-dependent methylation. The five carbons forming the cyclopropyl ring were predicted to link to the amino and sulfhydryl groups of cysteine and were envisioned to arise from four possible hypotheses: (1) acetate making a mevalonic acid-type intermediate, (2) a polyketide-type assembly of acetate with the branching carbon (C-20) coming from C-2 of acetate or an unknown C-1 source (Figure 2), (3) xylose through the non-mevalonate pathway to form isopentenyl pyrophosphate, or (4) a modification of leucine.

Results and Discussion

Pathway Investigation by Precursor Incorporation Experiments. To distinguish between the two proposed biosynthetic routes for the C-1 to C-16 section of curacin A (**1**) (Figure 1), isotope-labeled precursors ([1,2-^{13}C$_2$]-acetate, [1-^{13}C,^{18}O]acetate, [2-^{13}C]acetate, [1-^{13}C,^{2}H$_3$]acetate, [^{2}H$_3$,2-^{13}C]acetate, [2-^{13}C]-D,L-mevalonolactone, [2-^{13}C,^{15}N]-glycine, [1-^{13}C]glycine, and [methyl-^{13}C]-L-methionine) were

Table 1. NMR Data for Curacin A (**1**) Produced with Supplementation from Sodium [1,2-13C₂]-, [1-13C,18O₂]-, [2-13C]-, [²H₃,1-13C]-, or [²H₃,2-13C]Acetate (all spectra referenced to the centerline of the solvent (C₆D₆) at δ 128.39)

	[1,2-13C₂]acetate[a]			[1-13C,18O₂]acetate[a]		[2-13C]acetate[a]	[²H₃,1-13C]acetate[e]	[²H₃,2-13C]acetate[f]
carbon	δ_C	J_{CC} (Hz)	% relative incorporation[b]	Δδ (α-isotope shift)	% relative incorporation[d]	% relative incorporation[d]	Δδ (β-isotope shifts)	Δδ (α-isotope shift)
1	40.35				107.7	104.4		
2	74.73				93.6	93.2		
3	131.71		c		114.5	100.3		
4	131.25	42.0	c		97.4	187.5		0.351
5	28.54	42.0	129.9		163.6	99.7	0.107	
6	33.52	43.7	129.6		99.3	175.3		0.368
7	131.75	43.7	c		146.0	103.0		
8	128.28		c		c	c		
9	125.92	57.0	126.7		174.2	108.3	0.055	
10	136.83	43.0	125.6		83.0	214.7		
11	36.18	43.0	126.9		164.9	104.9		
12	32.56	39.5	131.0		108.7	183.0		0.385
13	80.33	39.5	127.5	0.02	176.0	111.2	0.055	
14	38.43	42.4	132.4		110.6	182.7		0.736
15	135.73	42.4	127.1		157.5	102.3	0.078	
16	117.18				94.4	190.4		0.546
17	16.97				92.7	101.9		
18	168.80	67.2	117.7		163.3	108.9		
19	20.51	67.2, 14.4	124.3		101.9	179.8		
20	14.61	14.4	141.8		111.9	188.0		0.295
21	16.37	44.2, 14.4	127.2		181.6	104.7	0.138	
22	12.72	44.2	134.5		96.9	199.6		0.558
OMe	57.70				95.2	100.9		

[a] 13C NMR spectra recorded on a Bruker AM400 spectrometer operating at 100.61 MHz. [b] Percent relative incorporation = A/B where A = total integrated NMR signal at that position and B = integrated NMR signal of the uncoupled resonance at that position. [c] Not determined due to overlapping NMR signals. [d] Relative incorporation = C/D where C = total integrated NMR signal at that position and D = integrated NMR signal at that position normalized to C-1, C-2, C-3, C-17 and the OCH₃ signal. [e] 13C NMR spectrum recorded on a Bruker DRX600 spectrometer operating at 150.90 MHz. [f] ²H-decoupled 13C NMR spectrum recorded on a General Electric GN Omega 500 spectrometer operating at 125.76 MHz

added to a laboratory-cultured strain of *L. majuscula* 19L, known to produce curacin A (**1**) and barbamide (**2**). Results of these precursor incorporation experiments are described below.

The carbon skeleton of curacin A (**1**) from C-3 to C-16 is predicted to derive from the incorporation of intact acetate units. However, the pattern of incorporation differs in the two proposed pathways and can be used to determine the correct orientation of acetate assembly (Figure 1). In Figure 1A, carbon positions 3–4, 5–6, 7–8, 9–10, 11–12, 13–14, and 15–16 are predicted to derive from intact acetate units, whereas in Figure 1B intact labeling patterns should be observed at carbon positions 4–5, 6–7, 8–9, 10–11, 12–13, and 14–15. Feeding experiments using [1,2-13C₂]acetate showed that approximately 30% of the carbon at each site came from the 13C-labeled precursor and gave rise to distinct sets of coupled signals with $^1J_{CC}$ values between 39 and 57 Hz (the magnitude depends on carbon hybridization) (Table 1). Careful measurement of these values demonstrated the intact incorporation of acetate units as predicted in Figure 1B, indicating that cysteine is incorporated into curacin A prior to extension of the C-3 to C-16 ketide chain. Additional coupling observed between C-18/C-19 ($^1J_{CC}$ = 67.2 Hz) and C-21/C-22 ($^1J_{CC}$ = 43.9 Hz) showed that four of the five atoms of the cyclopropyl ring-containing side chain were also derived from intact acetate units, thus supporting either a mevalonate (hypothesis #1) or a mixed acetate/C-1 origin for these five carbons (hypothesis #2, Figure 2).

Additional experiments with [1-13C,18O]acetate or [2-13C]-acetate were used to determine which carbon atoms originate from the C-1 or C-2 position of acetate. These results identified two carbon atoms that come from C-2 of a fragmented acetate unit and defined the origin of the oxygen atom at C-13. Following [1-13C,18O]acetate feedings, a 46–81% enhancement in 13C was measured at each of

the eight predicted C-1-derived carbon atoms (C-5, C-7, C-9, C-11, C-13, C-15, C-18, and C-21) (Table 1). Additionally, the 13C-enriched peak for C-13 was shifted slightly upfield (0.02 ppm) by a primary 18O isotope-induced shift, thus indicating an acetate-derived oxygen at this position (Supporting Information). Since C-1 of acetate was incorporated at C-18 and C-21, we concluded that the C-18 to C-22 fragment originated from either a classical mevalonate or a branched triketide pathway (hypotheses #1 and #2, respectively; Figure 2). Thus, we ruled out potential derivation from the non-mevalonate pathway (=methyl erythritol phosphate pathway, hypothesis #3) that would require processing of acetate via the glyoxylate and tricarboxylic acid cycles and result in labeling only at either C-20 or C-21,[19,20] and a potential leucine origin for this five-carbon fragment (hypothesis #4).

Precursor incorporation studies using [2-13C]acetate showed 75–115% 13C-enrichment at the expected carbon atoms (C-4, C-6, C-8, C-10, C-12, C-14, C-16, C-19, and C-22). Moreover, 13C-enrichment was also observed at C-20 (88% over natural abundance), a finding consistent with the origin of the C-18 to C-20 from either the classical mevalonic acid pathway or a branched triketide pathway. Interestingly, the 13C NMR of this sample also showed a $^1J_{CC}$ = 12 Hz in the 13C-enriched signals for both C-19 and C-20, indicating a substantial enrichment of the acetate pool during this incorporation experiment (Supporting Information).

The 13C-enriched acetate incorporation studies confirmed the polyketide nature of the C-4 to C-16 portion of curacin A (**1**) and implicated a novel terminal decarboxylation reaction that leads to the fragmentation of the final acetate unit. Two mechanistic possibilities were proposed for the terminal decarboxylation, as depicted in Figure 3. On one hand (Figure 3A), a ketone at C-13 and unsaturation between C-14 and C-15 could provide a bridge for electron
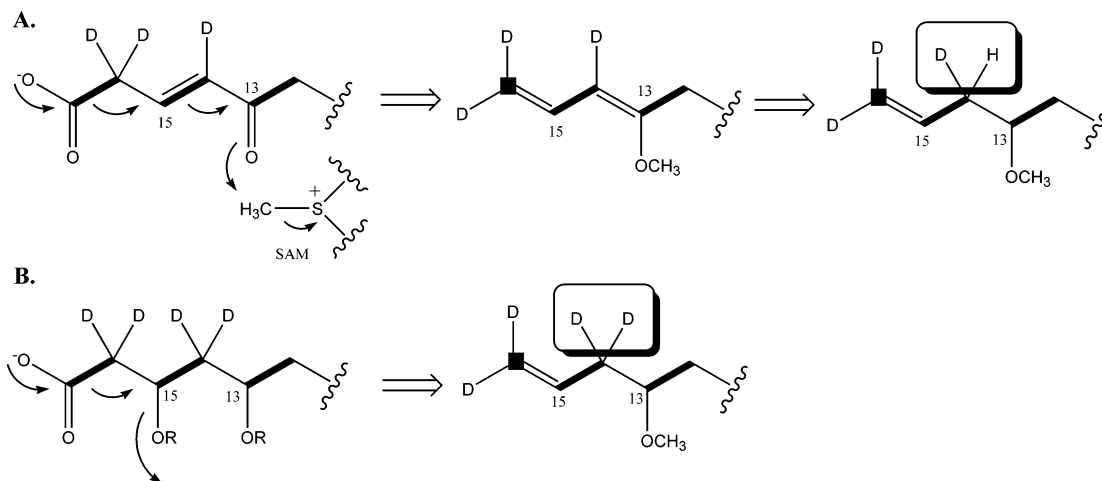
Figure 3. Two hypothetical routes by which terminal decarboxylation could occur during the formation of curacin A (**1**). In pathway A, the intermediacy of a C-14/C-15 double bond leads to retention of a single deuterium atom at C-14, whereas in pathway B two deuterium atoms are retained at this position. Experimental results described in the text are consistent with pathway B.

flow from the carbon−carbon bond being broken to the addition of a methyl group to the C-13 oxygen atom and neutralization of charge on sulfur in S-adenosyl methionine (SAM). Alternatively (Figure 3B), reduction of the C-15 intermediate ketone to the hydroxyl level would provide an alternate exit point (leaving group) for electrons generated during decarboxylation.

These two hypotheses imply a different level of intermediate oxidation at C-14, a feature that could be probed with ^{2}H-labeled acetate. The first of these experiments utilized [1-^{13}C,^{2}H$_3$]acetate (187 mg), in which the ^{13}C label acts as a reporter nuclei for ^{2}H incorporation (β-isotope-induced ^{2}H shift).[21] Labeled curacin A displayed an upfield shifted signal for C-15 ($\Delta\delta = 0.078$), consistent with the presence of two ^{2}H atoms at C-14. This labeling pattern supports the mechanism for terminal decarboxylation presented in Figure 3B. The reduction of the intermediate C-13 ketone most likely occurs via a normal ketoreductase (KR) mechanism within the modular PKS assembly process.[22,23] Additional β-isotope-induced shifts were observed for C-5, C-9, C-13, and C-21, the magnitude of which were consistent with a single ^{2}H atom at C-4, C-8, and C-12 and two ^{2}H atoms at C-22. The retention of only a single ^{2}H atom at C-12 and none at C-19 (one expected) is anomalous. Moreover, due to the near overlap of the C-3 and C-7 ^{13}C NMR resonance lines, the deuterium content at C-6 could not be measured using the β-isotope-induced shift approach.

As shown by the [2-^{13}C]acetate incorporation experiment, two acetate units in curacin A (**1**) are fragmented via decarboxylation of C-1 positions during the biosynthetic process, leading to isolated carbon atoms deriving from C-2 of acetate at C-16 and C-20. In these cases, the C-1 atom is lost, and another approach was required to follow the fate of ^{2}H atoms at these positions. Hence, [^{2}H$_3$,2-^{13}C]-acetate was provided to cultures of *L. majuscula* strain 19L, the curacin A isolated, and the ^{2}H α-isotope shift in the ^{13}C NMR spectrum measured (Table 1). A single ^{2}H atom was retained at C-4, C-6, C-12, and C-20, while two ^{2}H atoms were present at C-14, C-16, and C-22. From the two ^{2}H-acetate incorporation experiments, the following conclusions can be made: (1) the retention of two ^{2}H atoms at C-14, observed in both experiments, is consistent with the mechanism presented in Figure 3B for decarboxylation following the final ketide extension, (2) the retention of ^{2}H at most positions is as predicted from standard polyketide assembly, and (3) anomalies in the ^{2}H content are observed

at C-6 (one observed, two expected), C-12 (one observed, two expected), and C-19 (one expected, none observed).

Results from biosynthetic incorporation of labeled acetate indicated that the C-18 to C-22 section of curacin A (**1**) derived from either a terpenoid precursor (e.g., DMAPP, IPP) or a branching triketide precursor (Figure 2). To investigate the possibility of a terpenoid precursor, feeding experiments using [2-^{13}C]-D,L-mevalonolactone were conducted. The resulting curacin A (**1**) displayed no significant ^{13}C enrichments. Indeed, in light of a recent study of the pathway producing terpenes in cyanobacteria, which appears to be different from either the mevalonate or non-mevalonate pathways, incorporation from mevalonolactone into curacin A would have been surprising.[24] Additionally, analysis of ^{13}C-acetate enrichment in sections C-18 to C-22 and C-4 to C-16 supported the direct incorporation of acetate, rather than indirect incorporation through a mevalonate-type intermediate. A mevalonate-type intermediate would likely have reduced the levels of ^{13}C-enrichment in the C-18 to C-22 section (Table 1). The combination of comparable levels of ^{13}C-enrichment to this section, and the lack of incorporation from mevalonolactone, indicated that a branched triketide was the likely origin of C-18 to C-22 (Figure 2), a hypothesis supported by analysis of the biosynthetic gene cluster (vide infra).

At this point, the emerging picture of curacin A (**1**) biosynthesis suggested that a branched triketide condensed with cysteine and seven subsequent rounds of chain extension with malonate. To provide support for the involvement of cysteine as a biosynthetic precursor, trial experiments were conducted with amino acids lacking an isotope label. Unfortunately, both cysteine and its direct precursor, serine, were found to be toxic to the cyanobacterium when supplied at greater than 20 mg/L. Hence, we employed glycine on the basis of its reported ability to be better tolerated by microorganisms.[25] Because it was expected that a portion of the glycine provided to *L. majuscula* would lose C-1 via decarboxylation, we utilized [2-^{13}C,^{15}N]glycine with the goal of observing an intact ^{13}C−^{15}N bond in the resulting curacin A (C-2 to N). Significant enrichment was observed at C-2 of curacin A (**1**), appearing as an upfield isotope shifted signal with a $^1J_{NC} = 2.6$ Hz (Table 2 and Supporting Information). Additionally, the C-2 signal showed significant coupling to C-1 (both appearing as a triplet with $^1J_{CC} = 29.9$ Hz), which must arise from C-2 of glycine contributing to the C-1 pool (e.g., N^5,N^{10}-methylene tetrahydrofolate) and subsequent recombination

Table 2. ^{13}C NMR Data of Curacin A (**1**) Derived from [1-^{13}C]- and [2-^{13}C,^{15}N]Glycine and [*methyl*-^{13}C]-L-Methionine[a]

C no.	[1-^{13}C]glycine % relative incorporation[b]	[2-^{13}C,^{15}N]glycine % relative incorporation[c]	[*methyl*-^{13}C]- L-methionine % relative incorporation[d]
1	100.0	253.4	98.2
2	100.7	199.9	98.9
3	132.7	100.0	103.1
4	101.5	132.0	106.0
5	111.3	122.5	101.9
6	107.9	115.9	102.6
7	104.3	119.7	109.4
8	*e*	*e*	*e*
9	118.9	130.4	104.1
10	103.0	193.2	113.5
11	111.9	117.8	117.1
12	110.6	121.9	104.7
13	109.8	130.5	105.1
14	109.6	128.8	106.9
15	104.3	146.4	106.0
16	110.7	130.9	112.8
17	101.8	413.9	201.3
18	88.5	109.9	95.3
19	106.6	120.1	92.4
20	109.2	121.0	98.4
21	106.8	109.7	100.1
22	123.5	135.9	103.2
O-CH$_3$	97.6	309.7	187.8

[a] All ^{13}C NMR spectra were recorded on a Bruker AM400 spetrometer operating at 100.61 MHz. ^{13}C NMR spectra were referenced to the center line of the solvent (C$_6$D$_6$) at 128.3. [b] Relative incorporation = A/B, where A = total integrated NMR signal at that center and B = normalized integrated NMR signal at that center when compared to C-1, C-2, C-17, and O-CH$_3$ of **1**. [c] Relative incorporation = A/C, where A = total integrated NMR signal at that center and C = normalized integrated NMR signal at that center when compared to C-3 of **1**. Normalization to multiple carbon centers was not applicable to this feeding experiment because C-2 of glycine will be converted into C-2 of pyruvate and subsequently into C-1 of acetate via its catabolic pathway. [d] Relative incorporation = A/B, where A = total integrated NMR signal at that center and B = normalized integrated NMR signal at that center when compared to C-1, C-2, and C-3. [e] Not determined due to the overlapping of the NMR signal with the solvent signal.

with [2-^{13}C,^{15}N]glycine. Indeed, enrichment of the C-1 pool from [2-^{13}C,^{15}N]glycine was confirmed by observing strong labeling of the C-13 *O*-methyl signal as well as the pendant *C*-methyl group (C-17) (Table 2). In a second experiment using [1-^{13}C]glycine, curacin A showed a modest enrichment at C-3, confirming that processing of glycine to cysteine occurs as expected with C-1 of glycine becoming C-1 of cysteine. Decreased incorporation supports the metabolic lability of glycine in *L. majuscula*.

The origin of the two pendant methyl groups in curacin A (**1**), the C-13 *O*-methyl and the C-10 *C*-methyl (=C-17), were predicted to derive from SAM. Feeding experiments using [methyl-^{13}C]-L-methionine showed significant ^{13}C-enrichment in both pendent methyl groups (*O*-CH$_3$ = 88%, *C*-CH$_3$ = 101% over natural abundance (Table 2)). Interestingly, we previously isolated and described a curacin A analogue (curacin D, **3**) that lacks the C-17 methyl group.[26] We speculated that structural variation in curacin biosynthesis was likely introduced late in the assembly process as a "tailoring" reaction, and this inferred a methionine rather than propionate origin. While this predicted origin of C-17 was shown to be correct on the basis of precursor incorporation studies, analysis of the biosynthetic gene cluster suggests that the C-17 methyl group is introduced during polyketide elongation (see below).

Probing the *L. majuscula* Genomic Library and Characterization of the Curacin A Biosynthetic Gene Cluster. The ability to make structurally diverse metabolites of mixed polyketide/polypeptide origin from the same strain, including curacin A (**1**),[6] barbamide (**2**),[27] carmabin,[28] antillatoxin,[29] and several malyngamides,[30] presented a significant challenge for the efficient cloning of a target metabolic gene cluster. As previously described for analysis of the barbamide (**2**) gene cluster,[18] the cosmid library derived from *L. majuscula* genomic DNA was screened using several NRPS and PKS probes. Cosmid clones isolated from this initial screening strategy were then further characterized using specific structural motifs to identify the pathway of interest.

The *L. majuscula* library was screened by colony hybridization using a general PKS probe amplified by PCR with primers designed from conserved sequence of KS domains of several PKS gene clusters.[18] Sixty positive recombinant clones were further grouped by Southern hybridization using general PKS and NRPS probes, resulting in the isolation of three families of overlapping cosmids: A, B, and C.[18] Group A cosmid inserts were shown to encode the biosynthetic gene cluster for barbamide (**2**) biosynthesis.[18] Restriction mapping and Southern analysis using the general PKS and NRPS probes were conducted to evaluate groups B and C as candidates for the curacin A gene cluster.[18] On the basis of the structural features of curacin A (**1**) and the precursor incorporation studies, one NRPS module for cysteine and nine PKS modules were expected in the curacin A gene cluster. Group C cosmids were eliminated because they contained multiple NRPS modules and an insufficient number of PKS modules.[18] Group B cosmids contained robust signals corresponding to several PKS regions on a large fragment of DNA (~15 kb) consistent with the predicted number of PKS modules for curacin biosynthesis.[18] However, a general NRPS probe failed to hybridize to the group B cosmids in Southern analysis. To determine if group B cosmids contained a cysteine-specific NRPS module, a pair of PCR primers (PS720F and PS770R) were designed to amplify the adenylation domain of NRPS modules specific for cysteine incorporation and thiazoline/thiazole formation.[18] Sequence analysis of the PCR products indicated that group B cosmids contain an NRPS module specific for cysteine, consistent with precursor incorporation results for curacin A (**1**). The overlapping cosmids of group B were aligned with restriction mapping analysis, and three overlapping cosmids, pLM54, pLM9, and pLM17, were selected for DNA sequencing (Figure 4).

Nucleotide Sequencing and Analysis of the Curacin A (1) Biosynthetic Gene Cluster. The DNA sequences of the three overlapping cosmids were determined as described previously[18] and assembled into an 87-kb contig. Nucleotide sequence analysis revealed 31 putative open reading frames (ORFs) with a relatively low (45.5%) overall %G + C content, consistent with the *L. majuscula* genome. Sequence comparisons with the NCBI database indicated that the 14 ORFs spanning 63.7 kb of DNA at the center of the sequenced area might be responsible for curacin A biosynthesis, and they were designated *curA*–*curN* (Figure 4). No putative genes related to secondary metabolites were identified in the region preceding *curA*, and therefore it likely marks the 5′-end boundary of the *cur* gene cluster. The genetic architecture of the cluster includes the coincident direction of transcription and small intergenic regions (0–90 bp) between ORFs. The postulated *cur* gene products align well with specific families of
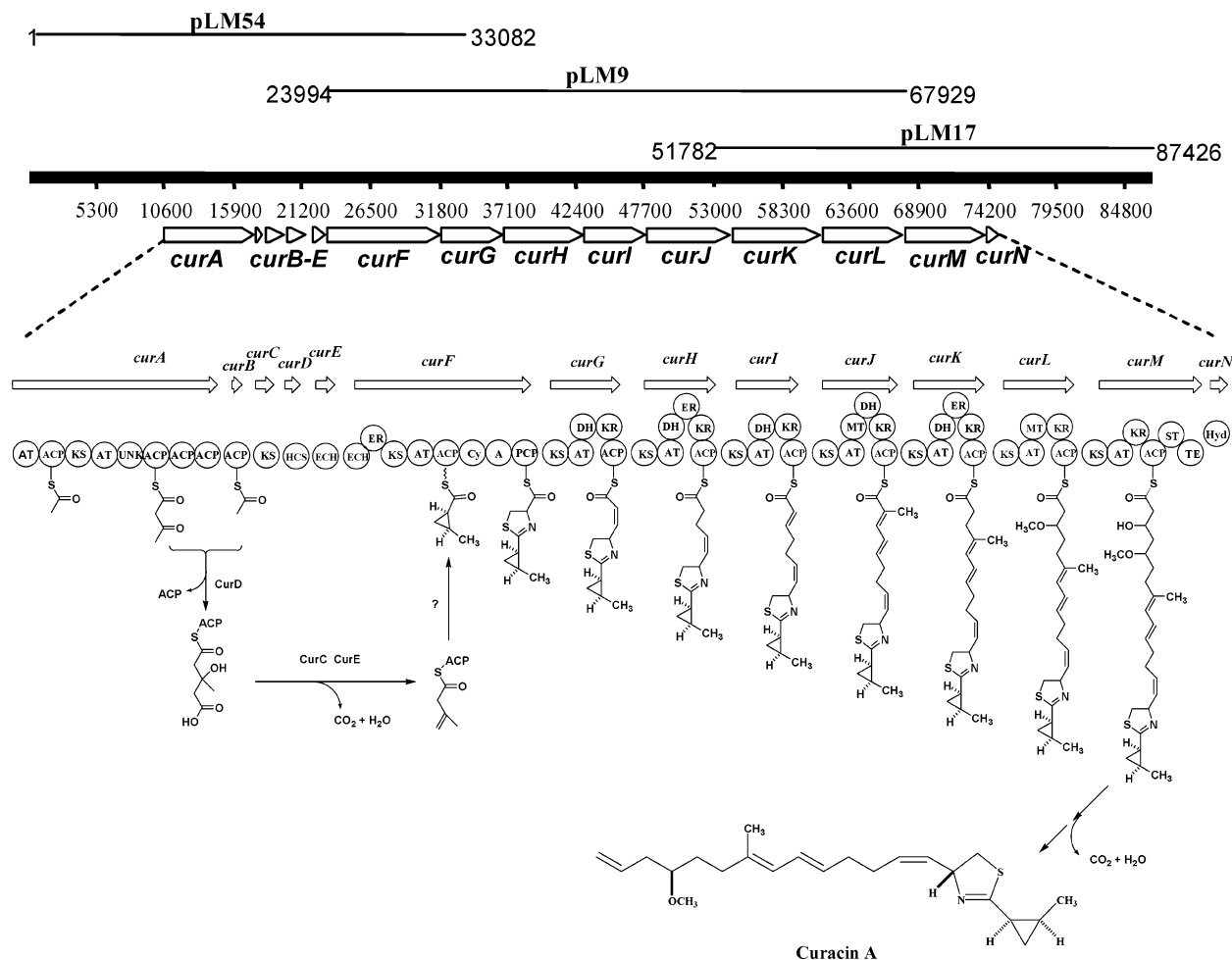
Figure 4. (a). Distribution of sequenced cosmids and ORFs on the sequenced area and the domain organization of the *cur* gene cluster. (b) Proposed pathway of curacin A (**1**) biosynthesis in *L. majuscula*. ACP, acyl carrier protein; KS, β-ketoacyl-ACP synthase; KR, β-ketoacyl-ACP reductase; AT, acyl transferase; HCS, 3-hydroxyl-3-methylglutaryl-CoA (HMG-CoA) synthase; ECH, enoyl-CoA hydratase/isomerase; DH, β-hydroxy-acyl-ACP dehydratase; O-MT, *O*-methyl transferase; ER, enoyl reductase; PCP, peptidyl carrier protein; Cy, condensaton/cyclization domain; A, adenylation domain; ST, sulfotransferase; TE, thioesterase; Hyd, α/β hydrolase; UNK, putative oxidase.

Table 3. Characterization of Protein Sequences Encoded by the *curA* Biosynthetic Gene Cluster

protein	amino acid	domains[a]	proposed function	sequence similarity	ident/sim	accession
CurA	2311	AT, ACP, KS, AT, ACP, ACP, ACP	PKS	StiC (*Stigmatella auratiaca*)	50%/67%	CAD19087
CurB	79	ACP	acyl carrier protein	ACP (*Bacillus subtilis*)	53%/72%	NP570904
CurC	408	KS	β-ketoacyl synthase	PksF (*Bacillus subtilis*)	40%/56%	NP389594
CurD	442	HCS	HMG-CoA synthase	FAB-H-like condensation	44%/63%	AAN85526
CurE	254	ECH	enoyl-CoA hydratase/ isomerase	enoyl-CoA hydratase/carnitine racemase (*Nostoc punctiforme*)	35%/53%	A48956
CurF	3195	ECH, ER, KS, AT, UNK, ACP, C, A, PCP	mixed PKS/NRPS	C-term = BarG (*L. majuscula*) N-term = PKS (*T. erythraeum*)	56%/71% 49%/68%	AAN32981 ZP00074378
CurG	1583	KS, AT KR, DH, ACP	PKS	NosB (*Nostoc* sp. GSV224)	51%/68%	AAF15892
CurH	2199	KS, AT, ER, DH, KR, ACP	PKS	StiF (*Stigmatella auratiaca*)	40%/56%	CAD19090
CurI	1656	KS, AT, DH, KR, ACP	PKS	StiB (*Stigmatella auratiaca*)	44%/60%	CAD19086
CurJ	2298	KS, AT, MT, DH, KR, ACP	PKS	StiC (*Stigmatella auratiaca*)	46%/64%	CAD19087
CurK	2232	KS, AT, ER, DH, KR, ACP	PKS	StiF (*Stigmatella auratiaca*)	40%/58%	CAD19090
CurL	1956	KS, AT, MT, KR, ACP	PKS	StiE (*Stigmatella auratiaca*)	38%/56%	CAD19089
CurM	2147	KS, AT, KR, ACP, ST, TE	PKS/TE	StiB (*Stigmatella auratiaca*)	43%/60%	CAD19086
CurN	341		hydrolase	haloalkane dehalogenase hydrolase	50%/64% 39%/58%	P59336 NP820220

[a] AT = acyl transferase, ACP = acyl carrier protein, KS = β-ketoacyl synthase; HCS = HMG-CoA synthase, ECH = enoyl CoA hydratase, ER = enoyl reductase, C = condensation, A = adenylation, PCP = peptidyl carrier protein, KR = keto reductase, DH = dehydratase, MT = methyl transferase, ST = sulfotransferase, UNK = domain of unknown function.

proteins in the GenBank database relating to secondary metabolism and have enabled reasonable prediction of functions for each of the corresponding ORFs (Table 3).

Nine PKS modules and one NRPS module were identified in the gene cluster that match the predictions based on curacin A (**1**) precursor incorporation experiments. The monomodular structure encoded by the PKS genes is a remarkable feature of the *cur* metabolic system. With the exception of *curF*, which is a hybrid PKS/NRPS bimodule, all other PKS genes encode a single type I module. The
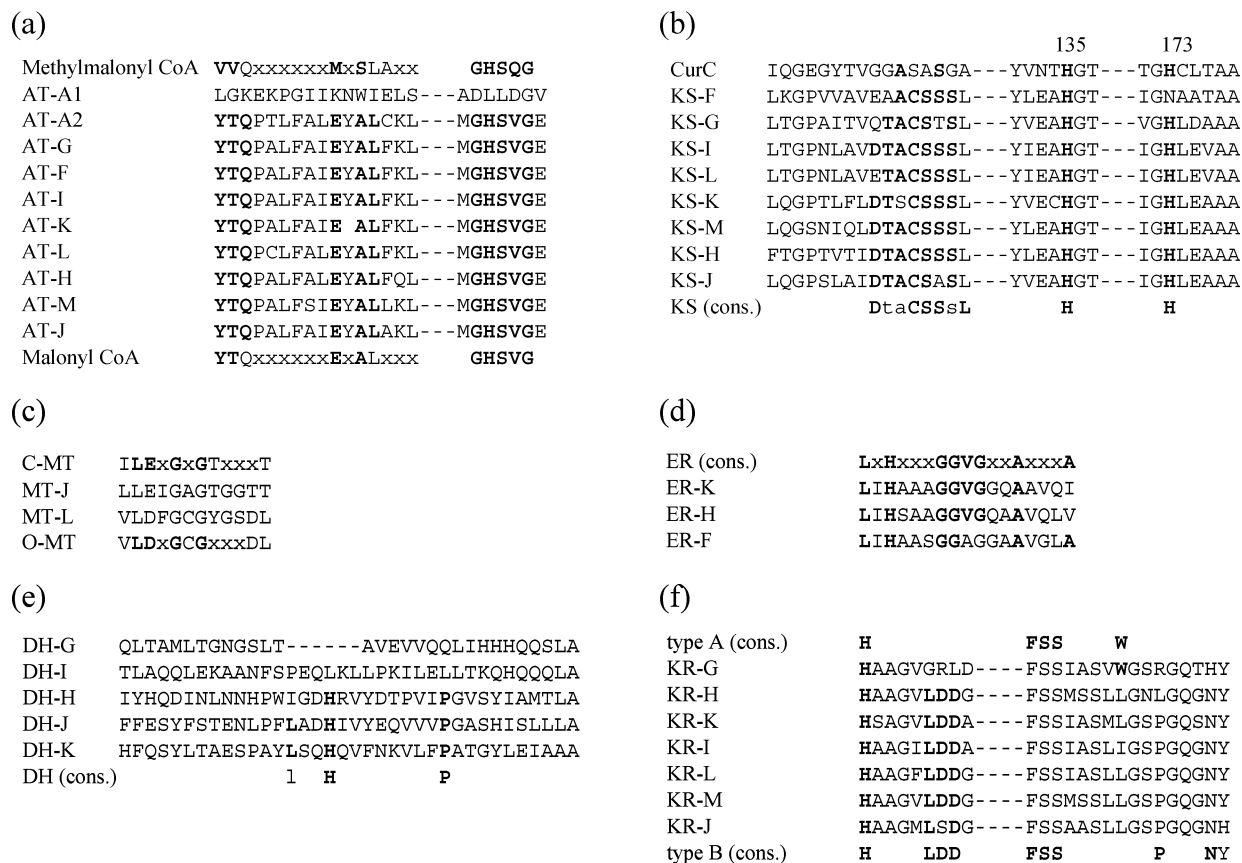
(a)

```
Methylmalonyl CoA  VVQxxxxxxxMxSLAxx      GHSQG
AT-A1              LGKEKPGIIKNWIELS---ADLLDGV
AT-A2              YTQPTLFALEYALCKL---MGHSVGE
AT-G               YTQPALFAIEYALFKL---MGHSVGE
AT-F               YTQPALFAIEYALFKL---MGHSVGE
AT-I               YTQPALFAIEYALFKL---MGHSVGE
AT-K               YTQPALFAIE ALFKL---MGHSVGE
AT-L               YTQPCLFALEYALFKL---MGHSVGE
AT-H               YTQPALFALEYALFQL---MGHSVGE
AT-M               YTQPALFSIEYALLKL---MGHSVGE
AT-J               YTQPALFAIEYALAKL---MGHSVGE
Malonyl CoA        YTQxxxxxxExALxxx      GHSVG
```

(b)

```
                                    135        173
CurC      IQGEGYTVGGASASGA---YVNTHGT---TGHCLTAA
KS-F      LKGPVVAVEAACSSSL---YLEAHGT---IGNAATAA
KS-G      LTGPAITVQTACSTSL---YVEAHGT---VGHLDAAA
KS-I      LTGPNLAVDTACSSSL---YIEAHGT---IGHLEVAA
KS-L      LTGPNLAVETACSSSL---YIEAHGT---IGHLEVAA
KS-K      LQGPTLFLDTSCSSSL---YVECHGT---IGHLEAAA
KS-M      LQGSNIQLDTACSSSL---YLEAHGT---IGHLEAAA
KS-H      FTGPTVTIDTACSSSL---YLEAHGT---IGHLEAAA
KS-J      LQGPSLAIDTACSASL---YVEAHGT---IGHLEAAA
KS (cons.)          DtaCSSsL       H         H
```

(c)

```
C-MT    ILExGxGTxxxT
MT-J    LLEIGAGTGGTT
MT-L    VLDFGCGYGSDL
O-MT    VLDxGCGxxxDL
```

(d)

```
ER (cons.)    LxHxxxGGVGxxAxxxA
ER-K          LIHAAAGGVGGQAAVQI
ER-H          LIHSAAGGVGQAAVQLV
ER-F          LIHAASGGAGGAAVGLA
```

(e)

```
DH-G      QLTAMLTGNGSLT------AVEVVQQLIHHHQQSLA
DH-I      TLAQQLEKAANFSPEQLKLLPKILELLTKQHQQQLA
DH-H      IYHQDINLNNHPWIGDHRVYDTPVIPGVSYIAMTLA
DH-J      FFESYFSTENLPFLADHIVYEQVVVPGASHISLLLA
DH-K      HFQSYLTAESPAYLSQHQVFNKVLFPATGYLEIAAA
DH (cons.)              l  H          P
```

(f)

```
type A (cons.)    H              FSS     W
KR-G              HAAGVGRLD----FSSIASVWGSRGQTHY
KR-H              HAAGVLDDG----FSSMSSLLGNLGQGNY
KR-K              HSAGVLDDA----FSSIASMLGSPGQSNY
KR-I              HAAGILDDA----FSSIASLIGSPGQGNY
KR-L              HAAGFLDDG----FSSIASLLGSPGQGNY
KR-M              HAAGVLDDG----FSSMSSLLGSPGQGNY
KR-J              HAAGMLSDG----FSSAASLLGSPGQGNH
type B (cons.)    H    LDD      FSS       P   NY
```

Figure 5. Sequence alignments between the key motifs of deduced *cur* gene products with the conserved motifs of several PKS domains: (a) AT domain alignment, where methylmalonyl CoA recognition indicates propionate incorporation and malonyl CoA recognition indicates acetate incorporation; (b) KS domain alignment; (c) MT domain alignment; (d) ER domain alignment; (e) DH domain alignment; and (f) KR domain alignment.

putative HMG-CoA synthetase encoded by *curD* supports hypothesis #2 for cyclopropane ring formation (Figure 2). The location of embedded methyltransferase domains in the deduced CurJ and CurL monomodular PKSs also supports the SAM-derived origin of the two pendant methyl groups in curacin A (**1**), the C-13 *O*-methyl and the C-10 *C*-methyl (=C-17). The overall architecture and catalytic domain organization for the *cur* cluster are collinear and completely consistent with the curacin A structure, as well as the precursor incorporation results.

Sequence Analysis of the Deduced *cur* Gene Products and Elucidation of the Mechanism of Curacin Biosynthesis. CurA encodes a 2311 amino acid protein with (AT_{A1}) ACP_{A1}, KS_A, AT_{A2}, UNK, ACP_{A2}, ACP_{A3}, and ACP_{A4} domains showing high similarity to other modular PKSs of various microorganisms. A unique feature of the protein composition of the biosynthetic system is a series of three tandem ACP domains (ACP_{A2}, ACP_{A3}, and ACP_{A4}) located at the C-terminus of the polypeptide. Another ACP domain (ACP_{A1}) precedes the KS_A domain. A Pfam HMM search of CurA revealed the presence of an AT_{A1} domain at the N-terminus of CurA. Further BLAST analysis indicated that this domain shares homology with the *N*-acetyltransferase (GNAT) domain of PedI in the putative pederin gene cluster. The GNAT domain of PedI was believed to load the starting acetyl unit in pederin biosynthesis.[31] However, the AT_{A1} domain found in CurA is presumed to be inactive on the basis of the absence of the conserved active site GHSxG motif and the substrate-specific motif (Figure 5a). Therefore, the malonyl-CoA starter unit might be loaded onto ACP_{A1} by the AT_{A2} domain that lies downstream of the KS_A domain. The AT_{A2} may also be responsible for loading the first extending malonyl-CoA unit onto ACP_{A2} as is found in the soraphen

gene cluster of *Sorangium cellulosum* So ce26 and the myxothiazol gene cluster of *Stigmatella aurantiaca* DW4/3-1.[32,33] Two tandem AT domains, located in the first PKS module of SorA and MtaB, were believed to be responsible for loading two different substrates. In myxothiazol biosynthesis, the first AT domain was believed to load the benzoyl-CoA starter unit that subsequently shifts backward to ACP1 preceding the KS domain; the other AT catalyzes loading of the first extension unit forward to ACP2.[34] In curacin A (**1**) biosynthesis, malonyl-CoA is believed to be the starter unit and the first extender unit. Therefore, one AT domain might be sufficient to load both positions. The CurA KS_A domain is predicted to catalyze the first condensation to form the diketide intermediate (Figure 4).

The tandem ACP domains found at the end of CurA, the four small ORFs encoding CurB, CurC, CurD, and CurE, and the first two domains of CurF are predicted to be involved in the formation of the unique cyclopropyl ring structure. Since the cyclopropyl ring is a critical component for the biological activity of curacin A (**1**), the biochemical mechanisms involved in its formation are of particular interest.[10,11,35] This *curB−F* cassette is highly related to a corresponding set of genes in the jamaicamide gene cluster (*jam*) isolated from a strain variant of *L. majuscula* (Figure 6).[36] Furthermore, both of these pathways share homology with the "pksX" pathway from *Bacillus subtilis*, possibly involved in the biosynthesis of difficidin, the mupirocin pathway (Mup) from *Pseudomonas fluorescens*, and the antibiotic TA gene cluster from *Myxococcus xanthus*.[37−39] Each of these putative pathways require tailoring enzymes that introduce and modify a pendant carbon atom to the nacent polyketide chain during assembly (Figure 6). The addition of a pendant carbon atom in the curacin pathway results in the formation of the cyclopropyl ring structure,
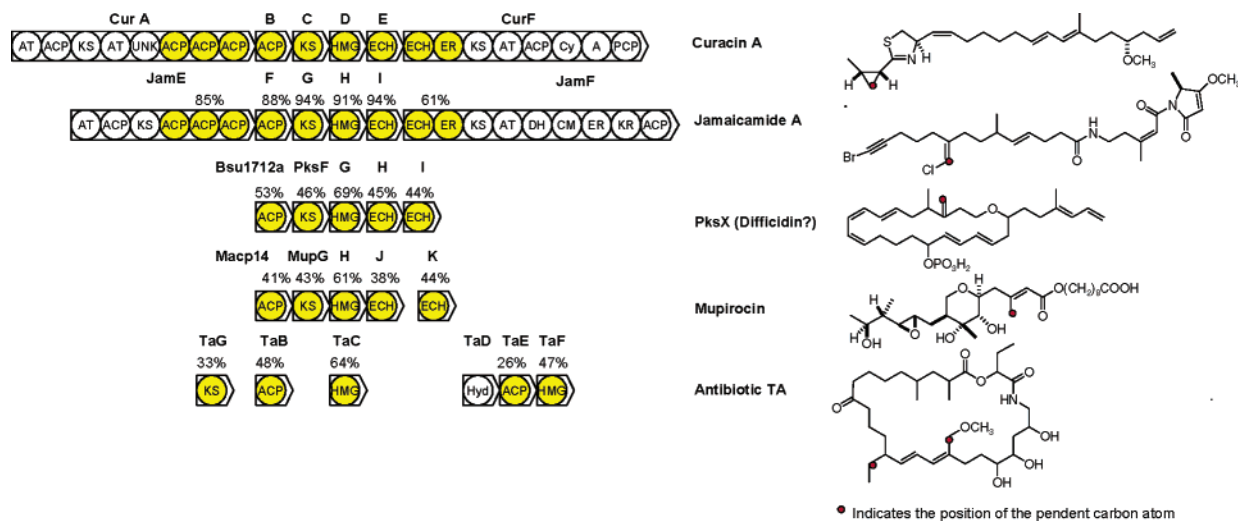
Figure 6. Putative gene cassette required for the addition of a pendent carbon atom from acetate to a nacent polyketide is conserved in several biosynthetic pathways. Alignment of the putative *curA* gene cassette required for the formation of the cyclopropyl ring functionality with biosynthetic gene cassettes from the jamaicamide, pksX (putatively coding for difficidin), mupirocin, and antibiotic TA biosynthetic gene clusters. Percent identity with the corresponding Cur homologue is indicated for each encoded protein. For the antibiotic TA, TaE was aligned with CurB, TaG is aligned with CurC, and TaF was aligned with CurD.

whereas in jamaicamide biosynthesis a vinyl chloride functional group is generated (Figure 6). ACP, KS, HMG-CoA synthase, and the enoyl-CoA hydratase domains have also been found in the pksX, mupirocin, and antibiotic TA pathways and correspond with the addition of pendant carbon atoms (Figure 6). In fact, two HMG-CoA synthetase domains are present in the antibiotic TA gene cluster and have been shown by gene disruption to be essential for antibiotic biosynthesis[39] (Figure 6).

Although rigorous biochemical analysis will be required to dissect the earliest steps of curacin A biosynthesis, we presume that the *curD*-encoded HMG-CoA synthase catalyzes condensation of an acetate unit with acetoacetyl-ACP to form (*S*)-3-hydroxy-3-methylglutaryl-ACP (HMG-ACP). The resulting intermediate would likely be modified by decarboxylation and dehydration to form an isopentenyl-ACP intermediate that is later isomerized to form the cyclopropyl ring. Sequence analysis of the CurC KS domain revealed that the invariant Cys residue in the active site is replaced by Ser, indicating the absence of the condensation function of the enzyme (Figure 5b). As with KS^Q domains found in numerous modular PKS systems,[40] CurC might function as a decarboxylase and is thus predicted to mediate the decarboxylation of the HMG-ACP intermediate. The ECH encoded by *curE*, together with the ECH domain in CurF, might catalyze the dehydration and isomerization to form the cyclopropyl ring intermediate (Figure 6). Alternatively a novel unknown domain in the CurA protein (UNK) that shares homology with phytanoyl-CoA hydroxylase (PhyH) might contribute to formation of the cyclopropyl ring (Figure 4). The proposed reactions are consistent with precursor incorporation studies and the biosynthetic mechanism described in Figure 2.

Due to the presence of the ECH and ER domains within CurF, we predict that cyclopropyl ring formation will be completed while the intermediate product is docked as an acyl-ACP intermediate to CurF. Further investigations of the KS active site within CurF reveal that while the active site cysteine residue in the DtaCSSsL motif is found (Figure 5b), the replacement of the conserved His-173 by Asn will likely render the enzyme ineffective for ketide extension and supports the hypothesis that it functions in the formation of the cyclopropyl ring.[41]

Following the PKS module on CurF is the single Cur pathway NRPS module. Sequence comparison with the GenBank database indicated a high similarity with NRPSs specific for the incorporation of cysteine and subsequent heterocyclization. The amino acid residues in the substrate binding pocket in the A domain of CurF are consistent with the conserved amino acids in the substrate binding pockets for cysteine recognition (Figure 7a). The core sequences in the C domain of CurF match well with the conserved core sequences of heterocyclization domains for thiazole/thiazoline formation (Figure 7b) and were therefore assigned Cy. The architecture of the NRPS module is thus consistent with the structure of curacin A (**1**) as well as the results of precursor incorporation experiments.

The remaining seven monomodular PKS modules in the curacin A (**1**) biosynthetic system are encoded by seven *cur* genes following *curF*, designated as *curG* to *curM*, that are responsible for seven rounds of condensation with malonyl-CoA extender units. The domain organization of the PKS modules corresponds closely to the structure of curacin A and the precursor incorporation results. All AT domains in these monomodules contain the amino acid motif specific for malonyl-CoA recognition (Figure 5a).[42,43] Embedded methyltransferase domains are located in the CurJ and CurL PKS modules. The sequence LLEIGAGTGGTT in MT$_J$ matches exactly the motif I(L)LExGxGTxxxT of *C*-methylation domains,[33] which is presumed to catalyze transfer of the C-17 methyl group to C-10 of curacin A (**1**). Sequence analysis and collinear placement of the CurL MT domain support its role in catalyzing *O*-methylation (Figure 5c). ER domains are located in each of the CurH and CurK modules, and the conserved LxHxxxGGVGxxAxxxA active site motif of functional ER domains was found in both (Figure 5d).[44] DH domains were expected in each of the CurG, CurH, CurI, CurJ, and CurK multifunctional proteins; however, only three functional DH domains were evident. Specifically, CurH, CurJ, and CurK contain the conserved His residue at the active sites of typical PKS DH domains, whereas this residue is missing in the CurG and CurI modules (Figure 5e). It is likely that the DH domains in CurH, CurJ, and CurK are essential for the subsequent function of the ER and *C*-methylation domains in these modules. On the other hand, no further processing domains follow the DH domains in CurG and CurI. Thus,

(a)

Position	235	236	239	278	299	301	322	330
Consensus(Cys)	D	L	Y	N	LM	SV	LMG	I
CurA	D	L	Y	N	L	S	L	I
BarG2	D	L	Y	N	L	S	L	I

(b)

	Z1	Z2	C3	Z3
Consensus	**FPLTXXQXAYXXGR**	**RXMLLXNGXQ**	**DXXXXDXXS**	**LPXXPXLP**
Cy-CurF	FPLNDIQQAYWIGR	RMVVLADGNQ	DMLIFDAWS	LPPAPEIP
Cy-BarG2	FPLTEIQQAYWLGR	RMVILPNGEQ	DALIADAWS	LPPAPELP

	Z4	Z5	Z6	Z7
Consensus	**TPXXXLXXXXXXVLXXW**	**GDFT**	**PVVFTSXL**	**QVXLDXQXXXXXXXXXXXXWY**
Cy-CurF	TPSGVLLSAFASVLNYW	GDFT	GVVFTSTL	QVLLDHIVTEEKGALAFSWH
Cy-BarG2	TPSGVLLAAFADFLAYW	GDFT	GVVFTSTL	QVWLDNSVAEQNGALLLIWN

Figure 7. (a) Comparison of the residues in the amino acid binding pockets of the adenylation domains of *curF* with the consensus sequence. The numbers represent the positions of the amino acids of GrsA.[52] (b) Alignment of the core sequences in the heterocyclization domain of CurF and BarG2 with the consensus core sequence of heterocyclization domains.[53]

dehydration might be catalyzed by other alternative dehydratases present in the gene cluster. The enoyl-CoA hydratase/isomerase product of *curE* is a potential candidate for this task.

The *cis* C-3/C-4 double bond and the terminal double bond of curacin A (**1**) represent additional interesting features of the molecule. Sequence analysis of the modules in CurG and CurM gave useful information on the potential mechanism of their formation. Two types of ketoreductases (KR) were found that are presumed to play a significant role in setting double-bond configuration in polyketide biosynthesis.[45] Specifically, *cis* or *trans* double-bond formation may be determined by the stereochemistry of the hydroxyl group that undergoes subsequent dehydration. Sequence alignment analysis indicated that the KR of CurG is different from all other KR domains in the cluster; the former belong to type A KRs and all others belong to type B KRs (Figure 5f).[45] The former is, therefore, predicted to result in a hydroxyl group with *S* stereochemistry, which is favored to form a *cis* double bond following dehydration. The later can form alcohols with *R* stereochemistry, resulting in the formation of a *trans* double bond by dehydration.

The precursor incorporation experiments indicated that the terminal double bond of curacin A (**1**) was formed by decarboxylation and ensuing dehydration (Figure 3B). The presence of a KR domain but not a DH or ER domain in CurM is consistent with these observations. The enzymes involved in the HMG-ACP modification might also be involved in catalyzing these two steps.

Finally, the TE domain is located at the C-terminus of the final PKS monomodule. An apparent sulfotransferase domain (ST) is located between the last ACP domain and TE domain, although its function is unclear. One small ORF, designated *curN*, is located immediately downstream of *curM*. The amino acid sequence of CurN shows significant similarity to α/β-hydrolase or acyltransferase. It is possible that CurN may be involved in modification of the final curacin A (**1**) chain elongation intermediate leading to the terminal alkene functionality.

Conclusions. The biosynthesis of curacin A (**1**) has been explored using precursor incorporation studies and molecular genetics to clone, sequence, and analyze the corresponding *cur* biosynthetic gene cluster. A particularly notable feature of the pathway involves a unique biochemical mechanism of cyclopropyl ring formation. In this process, HMGCoA-synthase-mediated reactions add an acetate unit to the β-carbonyl function of acetoacetyl-ACP substrate. The product is subsequently decarboxylated and rearranged to yield the methyl cyclopropyl ring system of the curacins. An NRPS module encoded by *curF* is responsible for linking this five-carbon fragment to cysteine as well as the subsequent heterocyclization and dehydration to the thiazoline moiety. Hence, the precise steps involved in generating these uniquely disposed cyclopropyl and thiazoline rings are now positioned to be fully explored biochemically. Several rounds of PKS extension utilizing malonyl-CoA, along with one *O*-methyl and one *C*-methyl transferase reaction, yield the remaining atoms of the curacin A molecule. In addition, a final decarboxylative dehydration presumably terminates the biosynthetic sequence to yield the characteristic C-15/C-16 olefin. A structural analysis of other marine cyanobacterial metabolites suggests that this terminating decarboxylation motif is relatively common (e.g., kalkitoxin,[46] somocystinamide[47]), thus representing yet another molecular signature for cyanobacterial secondary metabolism. The largely monomodular nature of the curacin A biosynthetic gene cluster is quite unusual and suggests it may be possible to manipulate this gene cluster in a number of intriguing ways (e.g., engineering domain and module exchanges). These several unique features of the curacin A (**1**) biosynthetic pathway and gene cluster will provide new opportunities for metabolic engineering of secondary metabolites as well as provide opportunities developing gene probes of greater specificity for identification and analysis of additional pathways that share some of these distinctive features.

Experimental Section

General Experimental Procedures. Most NMR spectra were recorded on a Bruker AM400 spectrometer operating at 400.13 MHz for ^{1}H NMR and at 100.61 MHz for ^{13}C NMR. The NMR spectra of **1** isolated from the [^{2}H$_3$,1-^{13}C]acetate feeding experiment were recorded on a Bruker DRX600 spectrometer (Bruker Biospin Corp., Billerica, MA) operating at a proton frequency of 600.01 MHz and a carbon frequency of 150.90 MHz. The NMR spectra of **1** isolated from the [^{2}H$_3$,2-^{13}C]acetate feeding experiment were recorded on a GN Omega 500 spectrometer (General Electric Global Research, Niskayuna, NY) operating at 500.11 Hz for the ^{1}H NMR and at 125.76 MHz for ^{13}C NMR. Chemical shifts are adjusted to the center line of solvent at 128.39 ppm (C$_6$D$_6$). High-performance liquid chromatography (HPLC) utilized a M6000A pump (Waters Corp., Milford, MA), a Rheodyne 7125 injector, and a Lambda-Max 480 LC spectrophotometer (Waters Corp.). Aluminum-backed thin-layer chromatography (TLC) sheets

Table 4. Calculation of [1-^{13}C,^{18}O$_2$]Acetate Incorporation into Curacin A (**1**)

carbon #	total integral of natural abundance curacin A	normalization factor compared to:					total integral of enriched curacin A	normalized integration value of enriched curacin A compared to:					% enrichment compared to:					average relative enrichment (%)
		C-1	C-2	C-3	C-17	OCH$_3$		C-1	C-2	C-3	C-17	OCH$_3$	C-1	C-2	C-3	C-17	OCH$_3$	
1	4.3	1.0	1.0	0.9	1.5	1.4	2.9	2.9	2.5	3.1	2.5	2.6	100.0	115.0	94.0	116.1	113.1	107.7
2	4.2	1.0	1.0	0.9	1.5	1.4	2.5	2.9	2.5	3.1	2.5	2.6	87.0	100.0	81.8	101.0	98.4	93.6
3	4.6	1.1	1.1	1.0	1.6	1.5	3.3	3.1	2.7	3.3	2.7	2.8	106.3	122.2	100.0	123.5	120.3	114.5
4	3.9	0.9	0.9	0.8	1.4	1.2	2.4	2.6	2.3	2.8	2.3	2.3	90.4	104.0	85.0	105.0	102.3	97.4
5	4.4	1.0	1.0	1.0	1.5	1.4	4.6	3.0	2.6	3.2	2.6	2.7	152.0	174.7	142.9	176.5	171.9	163.6
6	4.5	1.1	1.1	1.0	1.6	1.4	2.8	3.1	2.7	3.3	2.6	2.7	92.2	106.0	86.7	107.1	104.3	99.3
7	3.9	0.9	0.9	0.8	1.3	1.2	3.6	2.6	2.3	2.8	2.3	2.3	135.6	155.9	127.5	157.5	153.4	146.0
9	4.0	0.9	0.9	0.9	1.4	1.3	4.4	2.7	2.4	2.9	2.4	2.4	161.8	186.0	152.2	187.9	183.0	174.2
10	1.5	0.3	0.3	0.3	0.5	0.5	0.8	1.0	0.9	1.1	0.9	0.9	77.1	88.7	72.5	89.6	87.3	83.0
11	4.4	1.0	1.0	1.0	1.5	1.4	4.6	3.0	2.6	3.2	2.6	2.7	153.2	176.1	144.1	177.9	173.3	164.9
12	4.3	1.0	1.0	0.9	1.5	1.4	3.0	2.9	2.5	3.1	2.5	2.6	101.0	116.1	95.0	117.3	114.3	108.7
13	3.8	0.9	0.9	0.8	1.3	1.2	4.2	2.6	2.2	2.7	2.2	2.3	163.5	187.9	153.8	189.9	184.9	176.0
14	4.3	1.0	1.0	0.9	1.5	1.4	3.0	2.9	2.5	3.1	2.5	2.6	102.7	118.1	96.6	119.3	116.2	110.6
15	3.0	0.7	0.7	0.6	1.0	0.9	2.9	2.0	1.7	2.1	1.7	1.8	146.3	168.1	137.6	169.9	165.5	157.5
16	3.9	0.9	0.9	0.8	1.3	1.2	2.3	2.6	2.3	2.8	2.3	2.3	87.7	100.8	82.5	101.9	99.2	94.4
17	2.9	0.7	0.7	0.6	1.0	0.9	1.7	2.0	1.7	2.1	1.7	1.7	86.1	99.0	81.0	100.0	97.4	92.7
18	1.0	0.2	0.2	0.2	0.3	0.3	1.0	0.7	0.6	0.7	0.6	0.6	151.7	174.3	142.6	176.1	171.6	163.3
19	3.9	0.9	0.9	0.8	1.4	1.2	2.5	2.6	2.3	2.8	2.3	2.3	94.7	108.8	89.0	110.0	107.1	101.9
20	4.5	1.1	1.1	1.0	1.6	1.4	3.2	3.1	2.7	3.3	2.7	2.7	103.9	119.5	97.7	120.7	117.6	111.9
21	4.3	1.0	1.0	0.9	1.5	1.4	4.9	2.9	2.5	3.1	2.5	2.6	168.7	193.9	158.6	195.9	190.8	181.6
22	3.3	0.6	0.8	0.7	1.2	1.1	2.0	2.3	2.0	2.4	2.0	2.0	90.0	103.4	84.6	104.5	101.8	96.9
OCH$_3$	3.1	0.7	0.7	0.7	1.1	1.0	1.9	2.1	1.9	2.3	1.8	1.9	88.4	101.6	83.1	102.7	100.0	95.2

(silica gel 60 F$_{254}$; Merck & Co., Inc., Whitehouse Station, NJ) were used for TLC. Vacuum liquid chromatography (VLC) was performed with silica gel G for TLC (Merck & Co., Inc.). [2-^{13}C]-D,L-mevalonolactone was purchased from Isotec, Inc. (Miamisburg, OH), and all of the other stable isotope precursors were purchased from Cambridge Isotope Laboratories (Andover, MA).

General Culture Conditions and Isolation Procedure. Approximately 3 g of *L. majuscula* strain 19L was inoculated into a 2.8 L Fernbach flask containing 1 L of SWBG11 medium. The culture was grown at 28 °C under uniform illumination (4.67 μmol photon S^{-1} m^{-2}), aerated, and equilibrated for 3 days prior to addition of isotope-labeled precursors. Cultures of *L. majuscula* were harvested on day 9 or 10 after the inoculation, blotted dry, and extracted with 2:1 CH$_2$Cl$_2$/ MeOH. The extract was gravity filtered through glass wool, dried in vacuo, weighed, and applied to a silica gel column (1.5 cm i.d. × 15 cm) in 5% EtOAc/hexanes and eluted with a stepped gradient elution of 5% EtOAc to 100% EtOAc. A fraction containing curacin A (**1**) (eluted with 5% EtOAc/ hexanes) was subjected to a final purification by NP-HPLC [Verapack silica 10 μm, 4.1 mm × 30 mm, 4% EtOAc/hexanes, and UV detection at 254 nm, t_R = 22 min] to give pure curacin A (**1**, 11% of the extractable lipid; typically 5−15 mg recovered, or 2−5 mg per liter of culture media).

Calculation of ^{13}C Incorporation into Curacin A (1). The calculation of ^{13}C incorporation into curacin A (**1**) from [1-^{13}C,^{18}O$_2$]acetate supplementation is presented as an example (Table 4). Column A shows all 23 carbons of **1**. Columns B and H show the integrated values of the ^{13}C signals of natural abundance curacin A and enriched curacin A, respectively. Next, the normalization factors in comparison to C-1 (column C) were calculated by individually dividing the integration values of C-1 to C-22 and the OCH$_3$ group (table positions B1 to B23) by the integration value of C-1 (table position B1). The normalization factors compared to C2, C3, C17, and OCH$_3$ are shown in columns D, E, F, and G, respectively. Multiplication of the normalization factors (column C) by the integration value for the C-1 resonance of enriched **1** (table position H1) provides the normalized integration value compared to C-1 (column I). The normalization values compared to C-2, C-3, C-17, and the OCH$_3$ group are shown in columns J, K, L, and M, respectively. The percent ^{13}C enrichment of each carbon in **1** in comparison to C-1 (column N) was calculated by dividing the integration values of enriched **1** (column H) by the normalized integration values (column J) and multiplying

Table 5. Calculation of [1,2-^{13}C$_2$]Acetate Incorporation into Curacin A (**1**)[a]

carbon #	total integral of enriched curacin A	integration value for uncoupled signal	% relative enrichment from intact incorporation
5	59.5	45.8	129.9
6	60.0	46.3	129.6
9	49.8	39.3	126.7
10	18.9	15.1	125.6
11	58.7	46.3	126.9
12	59.1	45.1	131.0
13	53.0	41.6	127.5
14	59.1	44.6	132.4
15	35.6	28.0	127.1
16	54.9	49.7	110.5
18	11.4	9.7	117.7
19	45.7	36.8	124.3
20	47.7	33.6	141.8
21	53.7	42.2	127.2
22	47.7	35.4	134.5

[a] Carbon signals for C-4, C-7, and C-8 were obscured due to overlap. Carbon signals for C-1, C-2, C-3, C-17 and the O-Me signal were not enriched from this precursor source.

by 100. The percent ^{13}C enrichment of each carbon in comparison to C-2, C-3, C-17, and the OCH$_3$ group are shown in columns O, P, Q, and R, respectively. The average percent enrichment for each carbon atom (e.g., the average of numbers in columns N, O, P, Q, and R) is displayed in column S.

The relative ^{13}C enrichments in curacin A (**1**) from the [1,2-^{13}C$_2$]acetate feeding experiment were calculated as follows (Table 5). Column B shows the total integration value of the ^{13}C NMR signal for each resonance of **1** from this feeding experiment. The integration value of the ^{13}C signal appearing in the center of each cluster, representing those ^{13}C atoms not coupled to adjacent ^{13}C atoms for each resonance of **1**, is shown in column C. The percentage relative enrichment from intact [1,2-^{13}C$_2$]acetate incorporation appears in column D and was calculated by dividing the total integration values for **1** (column B) by the integration values of the "uncoupled" ^{13}C signals (column C) and multiplying by 100.

Sodium [1-^{13}C,^{18}O$_2$]Acetate Feeding Experiment. Sodium [1-^{13}C,^{18}O$_2$]acetate (375 mg) was mixed with unlabeled sodium acetate (622 mg) and administered to 3 × 1 L cultures on days 3, 6, and 8, and then all three 1 L cultures were harvested on day 10. A total of 8.4 mg of labeled curacin A (**1**)

was isolated from the crude organic extract. The ^{13}C NMR spectral data and percent relative enrichment in curacin A (**1**) are summarized in Table 1.

Sodium [2-^{13}C]Acetate Feeding Experiment. Sodium [2-^{13}C]acetate (187 mg) was mixed with unlabeled sodium acetate (311 mg) and fed to 3 × 1 L cultures on days 3, 6, and 8, and then all three 1 L cultures were harvested on day 10. A total of 6.1 mg of labeled curacin A (**1**) was isolated from the crude organic extract. The ^{13}C NMR spectral data and percent relative enrichment in **1** are summarized in Table 1.

Sodium [1,2-^{13}C$_2$]Acetate Feeding Experiment. Sodium [1,2-^{13}C$_2$]acetate (208 mg) was mixed with unlabeled sodium acetate (415 mg) and provided to 3 × 1 L cultures on days 3, 6, and 8, and then all three 1 L cultures were harvested on day 10. A total of 10.0 mg of labeled curacin A (**1**) was isolated from the crude organic extract. The ^{13}C NMR spectral data and percent relative enrichment in **1** are summarized in Table 1.

Sodium [1-^{13}C,2-^{2}H$_3$]Acetate Feeding Experiment. Sodium [1-^{13}C,2-^{2}H$_3$]acetate (187 mg) was mixed with unlabeled sodium acetate (311 mg) and provided to 3 × 1 L cultures on days 3, 6, and 8, and then all three 1 L cultures were harvested on day 10. A total of 6.0 mg of labeled curacin A (**1**) was isolated from the crude organic extract. The ^{13}C spectral data and the measured β-isotope induced shifts of **1** are summarized in Table 1.

Sodium [2-^{13}C,2-^{2}H$_3$]Acetate Feeding Experiment. Sodium [2-^{13}C,2-^{2}H$_3$]acetate (187 mg) was mixed with unlabeled sodium acetate (311 mg) and supplied to 3 × 1 L cultures on days 3, 6, and 8, and then all three 1 L cultures were harvested on day 10. A total of 3.9 mg of labeled curacin A (**1**) was isolated from the crude organic extract. The ^{2}H-decoupled ^{13}C NMR spectral data and the α-isotope induced shifts of **1** are summarized in Table 1.

[2-^{13}C]-D,L-Mevalonolactone Feeding Experiment. [2-^{13}C]-D,L-Mevalonolactone (100 mg) was provided to 3 × 1 L cultures on days 3, 6, and 8, and then all three 1 L cultures were harvested on day 10. A total of 7.4 mg of **1** was isolated from the crude organic extract. The ^{13}C NMR spectrum of curacin A (**1**) from this feeding experiment showed no enhancement of any carbon.

[1-^{13}C]Glycine Feeding Experiment. [1-^{13}C]Glycine (300 mg) was supplied to 2 × 1 L cultures on days 3, 6, and 8, and both cultures were then harvested on day 10. A total of 10.8 mg of labeled curacin A (**1**) was isolated from the crude organic extract. The ^{13}C NMR spectrum of **1** from this feeding experiment showed enhancement of the C-3 resonance at δ 131.71 (132%, Table 2).

[2-^{13}C,^{15}N]Glycine Feeding Experiment. [2-^{13}C,^{15}N]-Glycine (225 mg) was supplied to 2 × 1 L cultures on days 3, 6, and 8, and both cultures were then harvested on day 10. A total of 10.2 mg of labeled curacin A (**1**) was isolated from the crude organic extract. The ^{13}C NMR spectrum of **1** isolated from this feeding experiment showed enhancement of signals at δ 16.97 (C-17, s; 413%), 40.35 (C-1, d; $^1J_{CC}$ = 29.9 Hz, 253%), 57.70 (O*C*H$_3$, s, 309%), and 74.73 (C-2, d, $^1J_{CN}$ = 2.6 Hz; and dd, $^1J_{CC}$ = 29.9 Hz and $^1J_{CN}$ = 2.6 Hz; 199%) (Table 2).

L-[Methyl-^{13}C]Methionine Feeding Experiment. L-[*methyl*-^{13}C]Methionine (60 mg) was fed to 2 × 1 L cultures on days 3, 6, and 8, and both cultures were then harvested on day 10. A total of 11.2 mg of labeled curacin A (**1**) was isolated from the crude organic extract. The ^{13}C NMR spectrum of **1** isolated from this feeding experiment showed enhancement of signals at δ 16.97 (C-17, 201%) and 57.7 (O*C*H$_3$, 187%) (Table 2).

Bacterial Strains and Culture Conditions. *E. coli* strain DH5α was routinely used in this study as a host for DNA cloning and sequencing. *E. coli* was cultured overnight at 37 °C in Luria-Bertani (LB) broth or LB agar plates. Ampicillin and apramycin were added to cultures containing bacterial strains with plasmids or cosmids at final concentrations of 100 and 50 μg/mL, respectively.

DNA Manipulation. Plasmid DNA isolation, *E. coli* strain transformation, and colony and Southern hybridization were performed according to standard methods.[48] High molecular

Table 6. Primer Sequences Used for the Isolation and Characterization of the *curA* Biosynthetic Gene Cluster (bold letters show the restriction site, modified sequences are in italics)

primer	comments	sequence
KS1Up	PKS forward	5′-MGIGARGCIHWISMIATGGAYCCICAR-CAIMG-3′
KSD1	PKS reverse	5′-GGRTCICCIARISWIGTICCIGTICCITG-3′
PSF	NRPS forward	5′-AA***GAATTC***GCIGGIGGIGCITAIGTNCC-3′
PSR	NRPS reverse	5′-AA***GAATTC***CCCNCGIATITTIACITG-3′
PS720F	NRPS for Cys forward	5′-AT***GAATTC***GAYYTIWSIGTITAYGAY-ITITTYGG-3′
PS770R	NRPS for Cys reverse	5′-A***TGAATTC***ARIGCIGGIACIWSRTTCCA-3′

weight genomic DNA was prepared from *L. majuscula* 19L as described previously.[18] Briefly, the fresh filaments of *L. majuscula* were frozen in liquid nitrogen, ground with a mortar and a pestle for 10 min, and extracted with phenol/chloroform. The aqueous solution was transferred to a new tube, and the DNA was precipitated by ethanol. The genomic DNA was then redissolved in TE buffer and purified by a Genomic DNA purification kit (Qiagen, Valencia, CA).

Pfu Turbo DNA polymerase (Stratagene, La Jolla, CA) was used in PCR to amplify DNA fragments from *L. majuscula* genomic DNA with the following thermocycler conditions: denature at 95 °C for 45 s; anneal at 55 °C for 45 s; and extension at 72 °C for 1 min for 30 cycles. PCR primers (Table 6) were obtained from Life Technology, Inc. PCR products were purified using the PCR purification kit (Qiagen). The PCR conditions and degenerate primers KS1Up and KSD1 used for amplifying KS fragments from *L. majuscula* genomic DNA were as described by Beyer.[49] Primers PSF and PSR (Table 6) were used to amplify general NRPS fragments of *L. majascula* (bold letters show the restriction site, modified sequences are in italics). For amplification of NRPS fragments specifically for cysteine and thiazole/thiazoline formation, primers PS720F and PS770R were used.

Construction and Screening of the Cosmid Library. High molecular genomic DNA of *L. majuscula* was partially digested with *Sau*3AI and fractionated by sucrose gradient centrifugation. The DNA fragments of 30−45 kb were collected and cloned into *Bam*HI/*Hpa*I digested cosmid vector pOJ446.[50] In vitro packaging was performed with Gigapack III Gold (Stratagene) according to the manufacturer's protocol. About 2000 colonies (~20-fold coverage of the genome) were screened using the general PKS probe. The DIG High Prime Labeling kit (Roche Diagnostics Corp., Indianapolis, IN) was used to label the probes for Southern blot and colony hybridizations. The hybridization reaction was at 65 °C overnight and washed first with 1× SSC and then with 0.2× SSC at 60 °C, both for 30 min.

DNA Sequencing and Sequence Analysis. Cosmids pLM54, pLM9, and pLM17 were prepared for DNA sequencing by a shotgun cloning approach. The cosmid DNA was subjected to random fragmentation by nebulization to yield an average fragment size of 2−3 kb. Plasmid DNA (500 μg) was dissolved in 2 mL of TE buffer, transferred into a nebulizer (IPI Medical Products Inc, Chicago, IL), and nebulized with N$_2$ gas at 10−15 psi for 150 s. The DNA fragments were precipitated and resuspended in 500 μL of TE buffer and then gel-purified using a Gelase Extraction kit (Epicentre, Madison, WI). DNA fragments of 2−3 kb were blunt-ended by S1 nuclease, phosphorylated by T4 DNA kinase, and cloned in *Sma*I/BAP-treated pUC18 (Amersham Biosciences Corp. Piscataway, NJ). The random subclones were prepared for sequencing using a Miniprep 96 kit (Qiagen) and sequenced using an ABI Prism 3700 DNA analyzer (Perkin-Elmer, Boston, MA). The chromatogram sequencing data were assembled and edited using the Seqman program (DNASTAR, Inc., Madison, WI). DNA and protein sequences were routinely analyzed using Mac-Vector 6.5 (Accelrys Corp., San Diego, CA). NCBI BLAST searches were performed for sequence homology analysis of the *cur* genes and putative polypeptides.[51]

References and Notes

(1) Gerwick, W. H.; Tan, L.; Sitachitta, N. *The Alkaloids*; Academic Press: San Diego, 2001; pp 75–184.
(2) Faulkner, D. J. *Nat. Prod. Rep.* **2002**, *19*, 1–48.
(3) Bai, R.; Pettit, G. R.; Hamel, E. *Biochem. Pharmacol.* **1990**, *39*, 1941–1949.
(4) Bai, R.; Friedman, S. J.; Pettit, G. R.; Hamel, E. *Biochem. Pharmacol.* **1992**, *43*, 2637–2645.
(5) Proteau, P. J.; Gerwick, W. H.; Garcia-Pichel, F.; Castenholz, R. *Experientia* **1993**, *49*, 825–829.
(6) Gerwick, W. H.; Proteau, P. J.; Nagle, D.; Hamel, E.; Blokhin, A.; Slate, D. L. *J. Org. Chem.* **1994**, *59*, 1243–1245.
(7) Blokhin, A. V.; Yoo, H. D.; Geralds, R. S.; Nagle, D. G.; Gerwick, W. H.; Hamel, E. *Mol. Pharmacol.* **1995**, *48*, 523–531.
(8) Wipf, P.; Xu, W. *J. Org. Chem.* **1996**, *61*, 6556–6562.
(9) Muir, J. C.; Pattenden, G.; Ye, T. *J. Chem. Soc., Perkin Trans.* **2002**, 2243–2250, and references therein.
(10) Verdier-Pinard, P.; Lai, J. Y.; Yoo, H. D.; Yu, J.; Marquez, B.; Nagle, D. G.; Nambu, M.; White, J. D.; Falck, J. R.; Gerwick, W. H.; Day, B. W.; Hamel, E. *Mol. Pharmacol.* **1998**, *53*, 62–76.
(11) Wipf, P.; Reeves, J. T.; Balachandran, R.; Day, B. W. *J. Med. Chem.* **2002**, *45*, 1901–1917.
(12) Silakowski, B.; Kunze, B.; Muller, R. *Gene* **2001**, *275*, 233–240.
(13) Rorrer, G. L.; Modrell, J.; Zhi, C.; Yoo, H.-D.; Nagle, D.; Gerwick, W. H. *J. Appl. Phycol.* **1995**, *7*, 187–195.
(14) Burja, A. M.; Abou-Mansour, E.; Banaigs, B.; Payri, C.; Burgess, J. G.; Wright, P. C. *J. Microbiol. Methods* **2002**, *48*, 207–219.
(15) Sitachitta, N.; Rossi, J.; Roberts, M. A.; Gerwick, W. H.; Fletcher, M. D.; Willis, C. L. *J. Am. Chem. Soc.* **1998**, *120*, 7131–7132.
(16) Williamson, R. T.; Sitachitta, N.; Gerwick, W. H. *Tetrahedron* **1999**, *40*, 5175–5178.
(17) Sitachitta, N.; Marquez, B. L.; Williamson, R. T.; Rossi, J.; Roberts, M. A.; Gerwick, W. H.; Nguyen, V. A.; Willis, C. L. *Tetrahedron* **2000**, *46*, 9103–9114.
(18) Chang, Z.; Flatt, P.; Gerwick, W.; Nguyen, V.; Willis, C.; Sherman, D. *Gene* **2002**, *296*, 235–247.
(19) Flesch, G.; M., R. *Eur. J. Biochem.* **1988**, *175*, 405–411.
(20) Rohmer, M.; Knani, M.; Simonin, P.; Sutter, B.; Sahm, H. *Biochem. J.* **1993**, *295*, 517–524.
(21) Vederas, J. C. *Nat. Prod. Rep.* **1987**, *4*, 277–337.
(22) Weissman, K. J.; Timoney, M.; Bycroft M, G., P.; Hanefeld, U.; Staunton, J.; Leadlay, P. F. *Biochemistry* **1997**, *38*, 13849–13855.
(23) Lau, J.; Fu, H.; Cane, D. E.; Khosla, C. *Biochemistry* **1999**, *38*, 1643–1651.
(24) Ershov, Y. V.; Gantt, R. R.; Cunningham, F. X. J.; Gantt, E. *J. Bacteriol.* **2002**, *184*, 5045–5051.
(25) Abell, C.; Doddrell, D. M.; Garson, M. J.; Laue, E. D.; Staunton, J. *J. Chem. Soc., Chem. Commun.* **1983**, 694–696.
(26) Marquez, B.; Verdier-Pinard, P.; Hamel, E.; Gerwick, W. H. *Phytochemistry* **1998**, *49*, 2387–2389.
(27) Orjala, J.; Gerwick, W. H. *J. Nat. Prod.* **1996**, *59*, 427–430.
(28) Hooper, G. J.; Orjala, J.; Schatzman, R. C.; Gerwick, W. H. *J. Nat. Prod.* **1998**, *61*, 529–533.
(29) Orjala, J.; Nagle, D. G.; Hsu, V.; Gerwick, W. H. *J. Am. Chem. Soc.* **1995**, *117*, 8281–8282.
(30) Orjala, J.; Nagle, D.; Gerwick, W. H. *J. Nat. Prod.* **1995**, *58*, 764–768.
(31) Piel, J.; Wen, G.; Platzer, M.; Hui, D. *ChemBioChem* **2004**, *5*, 93–8.
(32) Ligon, J.; Hill, S.; Beck, J.; Zirkle, R.; Molnar, I.; Zawodny, J.; Money, S.; Schupp, T. *Gene* **2002**, *285*, 257–267.
(33) Wilkinson, C. J.; Frost, E. J.; Staunton, J.; Leadlay, P. F. *Chem. Biol.* **2001**, *8*, 1197–1208.
(34) Silakowski, B.; Schairer, H. U.; Ehret, H.; Kunze, B.; Weinig, S.; Nordsiek, G.; Brandt, P.; Blocker, H.; Hofle, G.; Beyer, S.; Muller, R. *J. Biol. Chem.* **1999**, *274*, 37391–37399.
(35) Blokhin, A. V.; Yoo, H. D.; Geralds, R. S.; Nagle, D. G.; Gerwick, W. H.; Hamel, E. *Mol. Pharmacol.* **1995**, *48*, 523–531.
(36) Edwards, D. J.; Marquez, B.; Nogle, L. M.; McPhail, K.; Goeger, D.; Roberts, M. A.; Gerwick, W. H. *Chem. Biol.* **2004**, in press.
(37) Albertini, A. M.; Caramori, T.; Scoffone, F.; Scotti, C.; Galizzi, A. *Microbiology* **1995**, *141*, 299–309.
(38) El-Sayed, A. K.; Hothersall, J.; Cooper, S. M.; Stephens, E.; Simpson, T. J.; Thomas, C. M. *Chem. Biol.* **2003**, *10*, 419–430.
(39) Paitan, Y.; Orr, E.; Ron, E. Z.; Rosenberg, E. *Microbiology* **1999**, *145*, 3059–3067.
(40) Bisang, C.; Long, P. F.; Cortes, J.; Westcott, J.; Crosby, J.; Matharu, A. L.; Cox, R. J.; Simpson, T. J.; Staunton, J.; Leadlay, P. F. *Nature* **1999**, *401*, 502–505.
(41) Tang, L.; Yoon, Y. J.; Choi, C. Y.; Hutchinson, C. R. *Gene* **1998**, *216*, 255–265.
(42) Haydock, S. F.; Aparicio, J. F.; Molnar, I.; Schwecke, T.; Khaw, L. E.; Konig, A.; Marsden, A. F.; Galloway, I. S.; Staunton, J.; Leadlay, P. F. *FEBS Lett.* **1995**, *374*, 246–248.
(43) Silakowski, B.; Nordsiek, G.; Kunze, B.; Blocker, H.; Muller, R. *Chem. Biol.* **2001**, *8*, 59–69.
(44) Kakavas, S. J.; Katz, L.; Stassi, D. *J. Bacteriol.* **1997**, *179*, 7515–7522.
(45) Caffrey, P. *ChemBioChem* **2003**, *4*, 654–657. Lett. **2002**, *4*, 1095–1098.
(46) Wu, M.; Okino, T.; Nogle, L. M.; Marquez, B. L.; Williamson, R. T.; Sitachitta, N.; Berman, F. W.; Murray, T. F.; McGough, K.; Jacobs, R.; Colsen, K.; Asano, T.; Yokokawa, F.; Shioiri, T.; Gerwick, W. H. *J. Am. Chem. Soc.* **2000**, *122*, 12041–12042.
(47) Nogle, L. M.; Gerwick, W. H. *Org. Lett.* **2002**, 4, 1095–1098.
(48) Sambrook, J.; Fritsch, E. F.; Maniatis, T. *Molecular Cloning; A Laboratory Manual*; Cold Springs Harbor Laboratory Press: New York, 1989.
(49) Beyer, S.; Kunze, B.; Silakowski, B.; Muller, R. *Biochim. Biophys. Acta* **1999**, *1445*, 185–195.
(50) Bierman, M.; Logan, R.; O'Brien, K.; Seno, E. T.; Rao, R. N.; Schoner, B. E. *Gene* **1992**, *116*, 43–49.
(51) Altschul, S. F.; Madden, T. L.; Schaffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. *Nucl. Acids Res.* **1997**, *25*, 3389–3402.
(52) Challis, G. L.; Ravel, J.; Townsend, C. A. *Chem. Biol.* **2000**, *7*, 211–224.
(53) Konz, D.; Marahiel, M. A. *Chem. Biol.* **1999**, *6*, R39–R48.