Chemistry & Biology, Vol. 11, 33–45, January, 2004, ©2004 Elsevier Science Ltd. All rights reserved. DOI 10.1016/j.chembiol.2003.12.014

Leinamycin Biosynthesis Revealing Unprecedented Architectural Complexity for a Hybrid Polyketide Synthase and Nonribosomal Peptide Synthetase

Gong-Li Tang,[1,3] Yi-Qiang Cheng,[1,3] and Ben Shen[1,2,*]
[1]Division of Pharmaceutical Sciences
[2]Department of Chemistry
University of Wisconsin
Madison, Wisconsin 53705

Summary

A 135,638 bp DNA region that encompasses the leinamycin (LNM) biosynthetic gene cluster was sequenced from *Streptomyces atroolivaceus* S-140. The boundaries of the *lnm* cluster were defined by systematic inactivation of open reading frames within the sequenced region. The *lnm* cluster spans 61.3 kb of DNA and consists of 27 genes encoding nonribosomal peptide synthetase (NRPS), polyketide synthase (PKS), hybrid NRPS-PKS, resistance, regulatory, and tailoring enzymes, as well as proteins of unknown function. A model for LNM biosynthesis is proposed, central to which is the LNM hybrid NRPS-PKS megasynthetase consisting of discrete (LnmQ and LnmP) and modular (LnmI) NRPS, acyltransferase-less PKS (LnmG, LnmI, and LnmJ), and PKS modules with unusual domain organization. These studies unveil an unprecedented architectural complexity for the LNM hybrid NRPS-PKS megasynthetase and set the stage to investigate the molecular basis for LNM biosynthesis.

Introduction

Leinamycin (LNM) is a novel antitumor antibiotic produced by several *Streptomyces atroolivaceus* species [1–3]. It is structurally characterized by an unusual 1,3-dioxo-1,2-dithiolane moiety that is spiro-fused to an 18-membered macrolactam ring of hybrid peptide-polyketide origin, a molecular architecture that has not been found to date in any other natural product (Figure 1). LNM exhibits a broad spectrum of antimicrobial activity against both gram-positive and gram-negative bacteria and shows potent antitumor activity in murine tumor models in vivo. Most significantly, it is active against tumors that are resistant to clinically important anticancer drugs, such as cisplatin, doxorubicin, mitomycin, and cyclophosphamide [2, 3]. LNM preferentially inhibits DNA synthesis and interacts directly with DNA to cause single-strand scission of DNA in the presence of thiol agents as cofactors [4–8]. The presence of the sulfoxide group in the dithiolane moiety is essential for the DNA-cleaving activity. Simple 1,3-dioxo-1,2-dithiolanes are also thio-dependent DNA cleaving agents in vitro [9, 10]. But the mechanisms for DNA cleavage by simple 1,3-dioxo-1,2-dithiolanes and LNM are distinct. Oxidative cleavage by 1,3-dioxo-1,2-dithiolanes converts molecular oxygen to DNA-cleaving oxygen radicals and is mediated by polysulfides, while alkylative cleavage by LNM is mediated by an episulfonium ion intermediate [6, 9, 10]. The latter mechanism represents an unprecedented mode of action for the thiol-dependent DNA cleavage by LNM.

LNM's unique chemical structure, potent biological activities, and novel mode of action have spurred great interest in developing LNM into a clinically useful anticancer drug [10–12]. Total synthesis of LNM has been accomplished [13, 14], and chemical modification [11, 12, 15] and functional mimics [10, 16–18] of the natural LNM have been extensively investigated. These studies generated a number of LNM analogs with improved antitumor activity, supporting the wisdom of making novel anticancer drugs based on the LNM molecular scaffold.

Complementary to making microbial metabolites and their structural analogs by chemical synthesis, "combinatorial biosynthesis," the generation of novel analogs of natural products by genetic engineering of biosynthetic pathways, offers a promising alternative that would allow the preparation of these compounds biosynthetically [19–22]. The success of this approach depends critically on (1) the development of novel methods and strategies for combinatorial manipulation of natural product biosynthetic gene clusters and (2) the continuous discovery and characterization of novel biosynthetic machinery that can be utilized for combinatorial biosynthesis. Given its unprecedented molecular architecture and potent antitumor activity, LNM offers a distinct opportunity to discover new chemistry for natural product biosynthesis and to develop new strategies and methods for combinatorial biosynthesis.

Recently, we reported the identification and localization of the LNM biosynthetic gene cluster from *S. atroolivaceus* S-140 [23]. Dissection of the biosynthetic pathway revealed a new type of polyketide synthase (PKS), which we named "acyltransferase (AT)-less PKS," in which the cognate AT domain is absent from all PKS modules and the AT activity is instead provided in *trans* by a discrete protein [24]. Here, we describe the sequence of the complete LNM biosynthetic gene cluster, determination of the cluster boundaries, and functional assignments of gene products and propose, on the basis of these results, a model for the LNM biosynthetic pathway. Our studies unveiled an unprecedented architectural complexity for the LNM hybrid NRPS-PKS megasynthetase, consisting of discrete and modular NRPSs, an AT-less modular PKS that requires a discrete, iteratively acting AT, and PKS modules with unusual domain organization. These findings set the stage to investigate the molecular basis for LNM biosynthesis and to apply combinatorial biosynthesis methods to the LNM biosynthetic machinery to generate structural diversity for anticancer drug discovery.

Results

Sequence Analysis of the *lnm* Biosynthetic Gene Cluster

Shotgun DNA sequencing of the previously identified *lnm* biosynthetic locus [23] yielded 135,638 bp of contig-

*Correspondence: bshen@pharmacy.wisc.edu
[3]These authors contributed equally to this work.

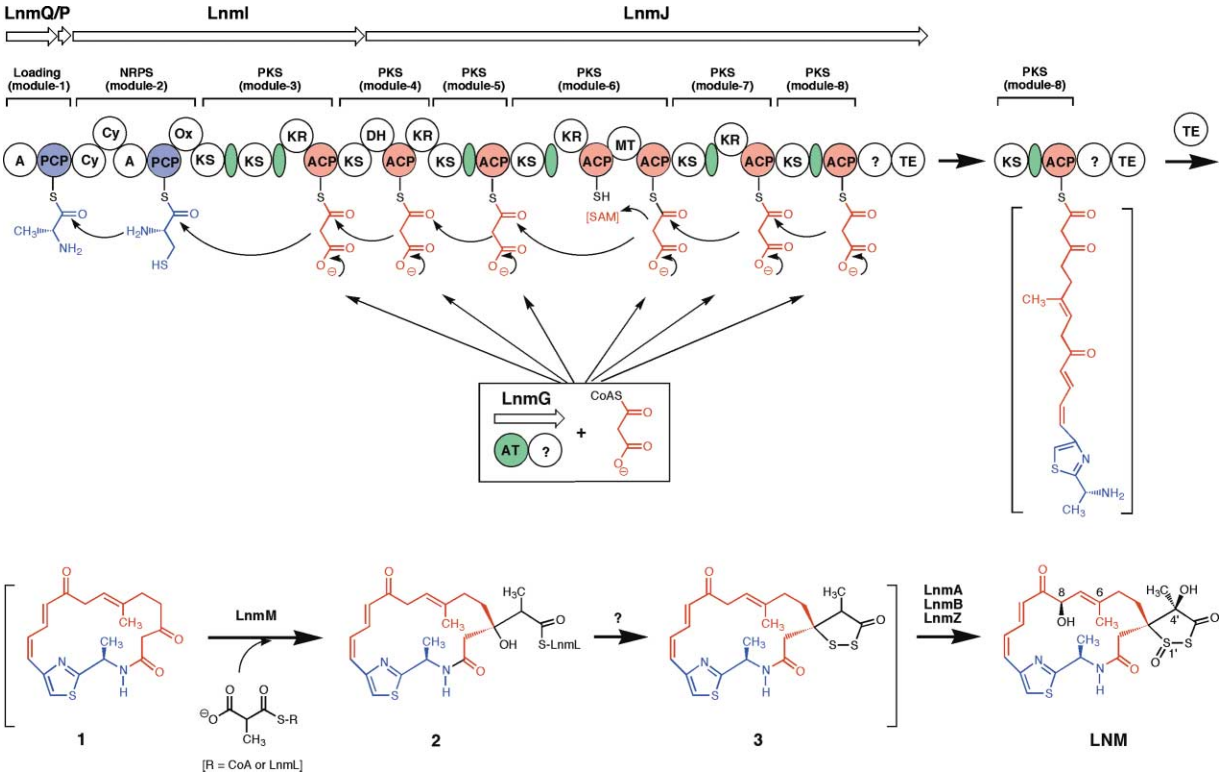


Figure 1. Proposed Model for LNM Biosynthesis and Modular Organization of the LNM Hybrid NRPS-PKS Megasynthetase with the Discrete LnmG AT Enzyme Loading the Malonyl CoA Extender Units to All Six PKS Modules

The structures in brackets are hypothetical. Color coding indicates the moiety of LNM that is of peptide (blue), polyketide (red), and other (black) origin. ACP, PCP, and AT and its proposed docking domains are shaded in red, blue, and green, respectively, to highlight the assembly-line mechanism for LNM biosynthesis. A green oval denotes an AT docking domain, and a question mark denotes a domain of unknown function. A, adenylation; ACP, acyl carrier protein; AT, acyltransferase; Cy, condensation/cyclization; DH, dehydratase; KR, ketoreductase; KS, ketosynthase; MT, methyltransferase; Ox, oxidation; TE, thioesterase.

uous DNA sequence, the overall GC content of which is 72.4%, characteristic of *Streptomyces* DNA [25]. Bioinformatic analysis of the sequenced region revealed 72 open reading frames (ORFs) (Figure 2). Functional assignments to individual ORFs were made by comparing the deduced gene products with proteins of known functions in the database and are summarized in Table 1 (only those within the *lnm* cluster). The GenBank accession number for this cluster is AF484556 (the entire sequenced region).

Determination of the *lnm* Gene Cluster Boundaries

The boundaries of the *lnm* gene cluster were accurately determined through systematic inactivation by gene replacement of ORFs within the sequenced region. Inactivation of genes (*lnmA, lnmG, lnmI, lnmJ*, and *lnmZ′*) within the *lnm* cluster abolished LNM production or resulted in the production of new metabolites, whereas inactivation of genes [*orf(−13), orf(−11), orf(−2), orf(−1), orf(+1), orf(+2), orf(+3), orf(+4)*, and *orf(+6)*] outside the *lnm* gene cluster had no effect on LNM production (Figure 3). These experiments led to the unambiguous assignment of the boundaries for the *lnm* gene cluster, which spans 61.3 kb of DNA and encompasses 27 ORFs designated *lnmA* to *lnmZ′* (Figure 2 and Table 1). Among them, 14 are biosynthetic genes encoding NRPSs (*lnmP* and *lnmQ*), PKSs (*lnmG, lnmJ, lnmL*, and *lnmM*), hybrid NRPS/PKS (*lnmI*), and other tailoring enzymes (*lnmA*, *lnmB*, *lnmD*, *lnmF*, *lnmN*, *lnmW*, and *lnmZ*), one is a regulatory gene (*lnmO*), five are resistance genes (*lnmR*, *lnmS*, *lnmT*, *lnmU*, and *lnmY*) and the remaining seven (*lnmC*, *lnmE*, *lnmH*, *lnmK*, *lnmV*, *lnmX*, and *lnmZ′*) are genes whose functions could not be predicted by sequence comparison alone.

Genes Encoding NRPSs

The *lnmQ* and *lnmP* genes encode NRPSs with unusual architecture. In contrast to the modular NRPSs consisting of multiple domains, *lnmQ* and *lnmP* encode a discrete NRPS adenylation (A) [26] and PCP protein [27], respectively. LnmQ is highly homologous to known adenylation domains of modular NRPS, containing the highly conserved 10 motifs (Figure 4A) [26]. LnmP displays high sequence homology to both PCP domains of modular NRPSs and discrete PCPs with the signature motif of Gx(HD)S, in which the Ser residue can be post-translationally modified by the covalent attachment of the 4′-phosphopantetheine group [27, 28]. As judged by their overlapping stop and start codons, *lnmQ* and *lnmP* are most likely translationally coupled. Similar or-

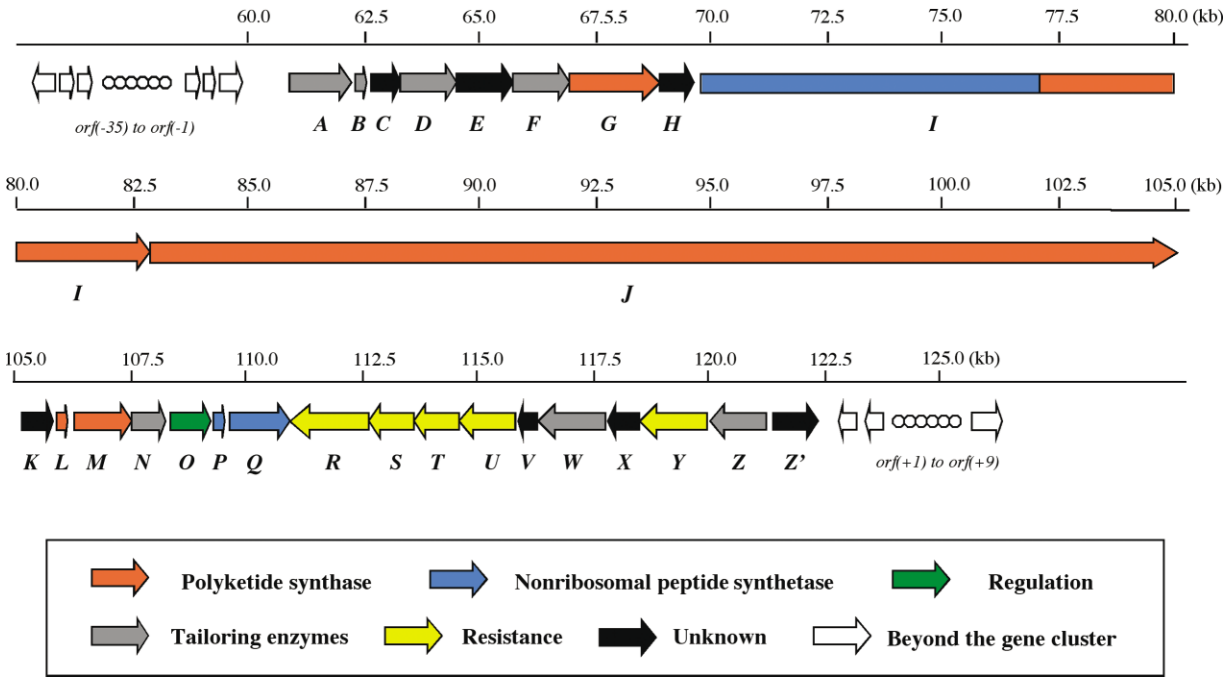


Figure 2. The Genetic Organization of the *lnm* Biosynthetic Gene Cluster

Proposed functions for individual ORFs are summarized in Table 1. The color coding legend indicates the types of genes identified with the *lnm* cluster.

ganizations for both NRPS and PKS genes are known and have been postulated to facilitate cotranslation of the coupled genes to yield equimolar amounts of proteins for optimal functional interactions, indicating that LnmQ and LnmP are likely functionally related.

To probe the role LnmQ and LnmP may play in LNM production, we inactivated *lnmQ* by replacing it with a mutant copy in which *lnmQ* was disrupted by the *aac(3)IV* apramycin-resistance gene [25]. The resultant Δ*lnmQ* mutant strain, SB3018, completely lost its ability to produce LNM. Introduction of pBS3047, in which the expression of *lnmQ* is under the control of the constitutive *ErmE** promoter [25], into SB3018 restored LNM production to approximately 40% the wild-type S-140 strain. These results unambiguously established that *lnmQ* is essential for LNM biosynthesis (Figure 4B).

Genes Encoding Hybrid NRPS/PKS

The *lnmI* gene encodes a hybrid NRPS/PKS protein containing domains characteristic of both NRPSs and PKSs. The *lnmI* gene is essential for LNM biosynthesis since inactivation of it completely abolished LNM production [23]. Residing at the N terminus of LnmI is an NRPS module (amino acid residues 1–1880) with the characteristic condensation/cyclization (Cy)-Cy-A-PCP-oxidation (Ox) domain organization [29]. We have previously demonstrated that this module activates and loads *L*-Cys to the cognate PCP and proposed that it is responsible for the biosynthesis of the thiazole moiety of LNM [23]. Following the LnmI NRPS module is a PKS module (module-3) (amino acid residues 1881–3930) with a novel β-keto synthase (KS)-KS-ketoreductase (KR)-acyl carrier protein (ACP) domain organization. To the best of our knowledge, this is the first example where tandem KS domains are identified within a PKS module. While both KSs (LnmI-KS3-1 and LnmI-KS3-2) show high homology to known KSs of PKS (Figure 5A) [30], the first KS domain contains a mutated catalytic triad of Cys-Ala-His instead of the highly conserved Cys-His-His triad known to functional KSs [31–33]. Residing at the C terminus of LnmI is a single KS domain (amino acid residues 3931–4437) that is characteristic for modular PKSs (Figure 5A). This KS presumably interacts with LnmJ in *trans* to form a complete PKS module (module 4). Among NRPSs and PKSs characterized to date, domains within a module often reside on the same protein. An NRPS or PKS module, such as module 4, that consists of domains residing on separate proteins is very rare, requiring precise protein-protein interaction between LnmI and LnmJ (Figure 1).

Genes Encoding PKS

The *lnmJ* gene encodes a giant protein of 7349 amino acid residues containing four KS, three ketoreductase (KR), one dehydratatse (DH), one methyltransferase (MT), six ACP, and one thioesterase (TE) domains characteristic to PKSs, as well as a domain unknown to PKSs (Figure 1). Inactivation of *lnmJ* completely abolished LNM production, confirming that *lnmJ* is essential for LNM biosynthesis [24]. The four KSs, like the three KSs of LnmI, are highly homologous to typical KSs from type I PKS modules (Figure 5A) and are characterized by the highly conserved catalytic triad of Cys-His-His [31–33]. The three KRs show significant homology to known KR domains with the signature NADPH binding site of GxGxxGxxxA [30, 34]. The DH domain has the

Table 1. Deduced Functions of ORFs in the *lnm* Biosynthetic Gene Cluster

Gene	Size[a]	Protein Homolog[b]	Proposed Function
orf(−35) – orf(−1)			ORFs beyond the upstream boundary
lnmA	399	RapN (T30231, 42/62)	Cytochrome P-450 hydroxylase
lnmB	78	RapM (T30230, 33/56)	Ferredoxin
lnmC	115	-	Unknown
lnmD	438	MupV (AAM12938, 23/37)	Lactone hydrolase/decarboxylase
lnmE	307	-	Unknown
lnmF	265	MupJ (AAM12923, 19/31)	Enoyl CoA hydratase
lnmG	795	MmpIII (AAM12912, 46/62)	Acyltransferase/oxidoreductase
lnmH	274	-	Unknown
lnmI	4437	PedH (AY059471, 35/53)	Hybrid NRPS/PKS
lnmJ	7349	PksM (C69679, 32/53)	PKS
lnmK	319	TaD (CAB46503, 51/64)	Unknown
lnmL	86	TaE (CAB46504, 41/57)	Discrete ACP
lnmM	416	HmgS (BAB07795, 33/51)	HMG-CoA synthase
lnmN	267	GrsT (P14686, 36/58)	Type II TE
lnmO	227	BTR (Q08530, 27/45)	Transcriptional activator
lnmP	82	AcpX (P43677, 43/56)	Discrete PCP
lnmQ	516	NRPS8 (BAB69421, 41/51)	Discrete NRPS A domain
lnmR	575	MoaD (T45539, 27/39)	ABC transporter component
lnmS	287	AgaC (T45530, 29/42)	ABC transporter component
lnmT	321	AgaB (T45531, 28/42)	ABC transporter component
lnmU	513	OphA (C98307, 26/41)	ABC transporter component
lnmV	120	-	Unknown
lnmW	516	4-CL (B87644, 24/37)	4-Coumarate-CoA ligase
lnmX	243	SCF43.15c (CAB66204, 45/56)	Unknown
lnmY	474	SCH42.31c (T35130, 36/52)	Antibiotic efflux protein
lnmZ	400	McyG (S51594, 42/58)	Cytochrome P-450 hydroxylase
lnmZ′	134	-	Unknown
orf(+1) – orf(+9)			ORFs beyond the downstream boundary

[a] Numbers are in amino acids.
[b] Given in brackets are accession numbers and percentage identity/percentage similarity.

conserved active site of HxxxGxxxxP but is only 70% of the length of typical DH domains and lacks the N-terminal 45 amino acid residues [30, 35]. The MT domain contains the three core motifs, LExGxAxA, GxxxxxxD, and LxxPxG, of *S*-adenosylmethionine (AdoMet)-dependent MTs and is highly homologous to C-MT domains of PKSs [36, 37]. The six ACPs exhibit high homology to known PKS ACPs, and all contain the signature motif around the invariable 4′-phosphopantetheine attachment site Ser residue (Figure 5B) [28, 30]. The extra domain at the C terminus of LnmJ is unprecedented in modular PKSs. It shows significant sequence homology to various tyrosine phenol-lyases (TPLs), such as the one from *Symbiobacterium thermophilum* (GenBank accession number A48380; 45% similarity and 27% identity) or *Citrobacter freundii* (accession number P31013; 41% similarity and 26% identity), but its role in LNM biosynthesis cannot be predicted on the basis of sequence analysis. The TE domain is highly homologous to other known TEs of type I PKS characterized by the conserved active site motifs of GxSxG and GxH [38].

Strikingly, the six PKS modules (modules 3 to 8) encoded by *lnmI* and *lnmJ* lack the cognate acyltransferase (AT) domain, which typically consists of approximately 320 amino acid residues and resides immediately after the KS domain within the PKS module. Close examination of LnmI and LnmJ revealed instead a short segment of 90 to 110 amino acids immediately after every KS with the exception of the C-terminal LnmI-KS (module 4), which seems to be the remnant of a functional AT domain (Figure 1). Remarkably, these segments seem to be highly conserved among all AT-less PKS modules known to date [39, 40] and share well-defined boundaries (Figure 6A). They appear to be derived by multiple deletions from a functional AT domain but cannot be functional themselves due to the lack of the highly conserved active site (GHSxG) and substrate binding [A(FS)HS] motifs [30, 35, 41] (Figure 6B). While it remains to be established if these regions play a role in LNM biosynthesis, it is tempting to propose that they could act as "docking sites" for the functional interaction between AT-less PKS enzymes and discrete ATs to constitute a functional megasynthetase. We therefore named those segments as AT "docking" domains (Figures 1 and 6).

The *lnmG* gene encodes a protein of 795 amino acids that clearly can be divided into two distinct domains. The N-terminal domain (amino acid residues 1–317) shows high sequence homology to AT domains of known PKSs and contains the highly conserved active site of GHSxG and substrate binding motif of AFHS that is specific for malonyl-CoA [41]. The C-terminal domain (amino acid residues 318–795) exhibits low but recognizable sequence homology to a family of oxidoreductases, such as MmpIII from *Pseudomonas fluorescens* (accession number AAM12912; 59% similarity and 44% identity) or PedB from the symbiont bacterium of *Paederus fuscipes* (accession number AAL27847; 57% similarity and 39% identity). LnmG is the only AT enzyme identified within the *lnm* gene cluster, and inactivation of *lnmG* completely abolished LNM production, confirming that

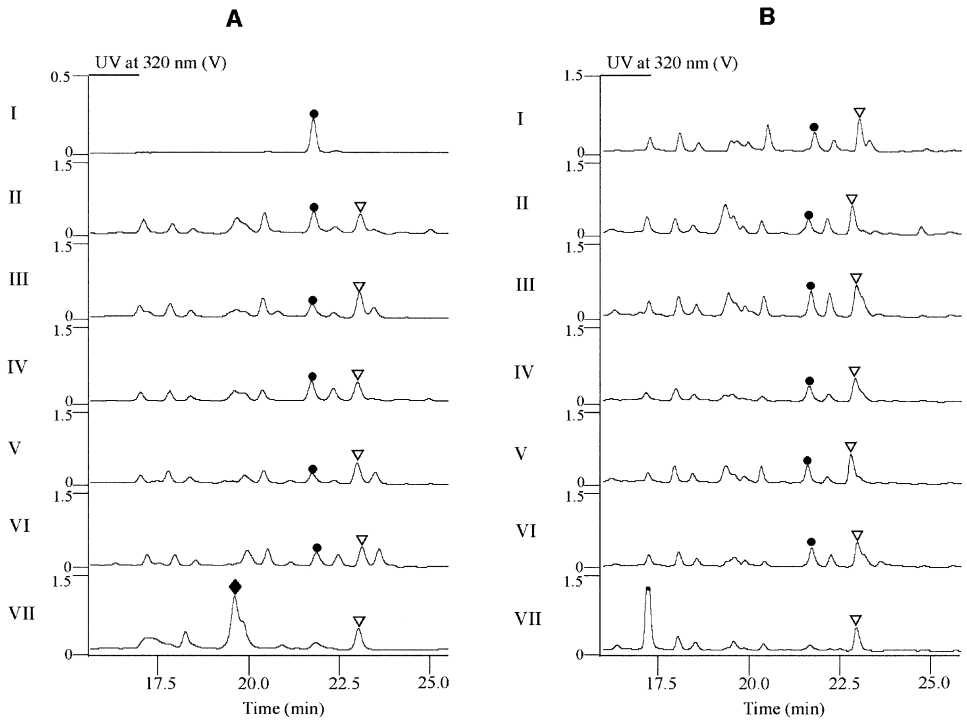


Figure 3. Determination of the *lnm* Biosynthetic Gene Cluster Boundaries and HPLC Analysis of LNM Production by *S. atroolivaceus* Wild-Type and Recombinant Strains

(A) Upstream boundary determination: I, LNM standard; II, wild-type S-140; III, SB3007 [Δ*orf(−13)*]; IV, SB3008 [Δ*orf(−11)*]; V, SB3009 [Δ*orf(−2)*]; VI, SB3010 [Δ*orf(−1)*]; VII, SB3011 (Δ*lnmA*).

(B) Downstream boundary determination: I, wild-type S-140; II, SB3012 [Δ*orf(+6)*]; III, SB3013 [Δ*orf(+4)*]; IV, SB3014 [Δ*orf(+3)*]; V, SB3015 [Δ*orf(+2)*]; VI, SB3016 [Δ*orf(+1)*]; VII, SB3017 (Δ*lnmZ′*); filled circles, LNM; open triangles, an unknown metabolite whose production is independent of LNM biosynthesis; filled diamond, a new metabolite accumulated by the Δ*lnmA* mutant.

it is essential for LNM biosynthesis [24]. We have previously characterized LnmG as a discrete AT enzyme that loads the extender units in *trans* to the LnmI and LnmJ AT-less PKS proteins for LNM biosynthesis [24].

The *lnmL* gene encodes a discrete ACP that contains the signature motif around the invariable 4′-phosphopantethein attachment site Ser residue (Figure 5B) [28, 30]. The deduced gene product of *lnmM* shows high sequence homology to both ketoacyl synthases, such as TaF from *Myxococcus xanthus* (accession number CAB46505; 76% similarity and 62% identity) or MupH from *Pseudomonas fluorescens* (accession number AAM12922; 62% similarity and 44% identity), and the 3-hydroxy-3-methylglutaryl (HMG)-CoA synthase family of enzymes, such as HmgS from *Streptomyces* sp. CL190 (accession number BAB07795; 51% similarity and 33% identity) or MvaS from *Enterococcus faecium* (accession number AAG02443; 49% similarity and 31% identity). While the molecular origin of the 1,3-dioxo-1,2-dithiolane moiety remains to be established, LnmL and LnmM could play a role in fusing this moiety to the LNM macrolactam ring (Figure 1).

Genes Encoding Tailoring Enzymes

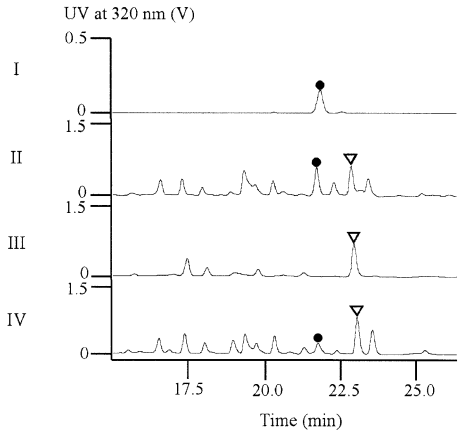
Flanking the boundaries of the *lnm* gene cluster are *lnmA* and *lnmZ*, the deduced gene products of which share significant sequence homology with each other (40% identity and 56% similarity) and are highly homologous to known cytochrome P-450 hydroxylases, such as RapN from *Streptomyces hygroscopicus* (accession number T30231; 42% identity and 62% similarity) and MycG from *Micromonospora griseorubida* (accession number S51594; 42% identity and 58% similarity). They both contain the characteristic heme binding motif of FGHGAHxCLG [42]. Translationally coupled with *lnmA* is *lnmB*, encoding a ferredoxin that would be required for the catalytic cycle of cytochrome P-450 enzymes. The fact that only one ferredoxin-encoding gene is identified within the *lnm* gene cluster suggests that the two LnmA and LnmZ cytochrome P-450 enzymes may share the same electron-transferring partner. Inspection of the LNM structure suggests at least three oxidation steps in LNM biosynthesis (hydroxylation at C-8 and C-4′ and oxidation at S-1′), and LnmA, LnmB, and LnmZ could serve as candidates for these enzyme activities (Figure 1). Indeed, preliminary analysis of the new metabolite accumulated by Δ*lnmA* mutant (Figure 3A, VII) by high-resolution mass spectrometry yielded $(M+H)^+$ and $(M+Na)^+$ ions at $m/z = 465.156$ and 487.138, consistent with the molecular formula $C_{22}H_{28}N_2O_5S_2$ [calculated 465.152 for $(M+H)^+$ and 487.134 for $(M+Na)^+$]. In contrast, high-resolution mass spectrometry analysis of the corresponding LNM degradation product isolated from the S-140 wild-type strain yielded $(M+H)^+$ and $(M+Na)^+$ ions at $m/z = 497.143$ and 519.125, consistent with the molecular formula $C_{22}H_{28}N_2O_7S_2$ [calculated 497.142 for

A

	A1	A2	A3	A4	A5
LnmQ	WTYAEA	LYAGCTAVPLN	LAYILFTSGSTGRPKG	FDLS	NLYGPTE
Ave-NRPS8	LSYGAL	LRAGATVVPLN	IAYIIFTSGSTGAPKG	FDGS	VLYGPTE
LnmI-A	MTYQEL	LKAGGVYLPME	TAYIIFTSGSTGRPKG	FDLS	SLGGATE
GrsA-A	LTYHEL	LKAGGAYVPID	LAYVIYTSGTTGNPKG	FDAS	NAYGPTE
CssA-A1	LSYSEL	LKAHLAYLPLD	LAYVIFTSGSTGKPKG	FDLS	NAYGPTE
HTS1-A3	LSYMQM	LKAGGAFMPVD	PAYLLYTSGTSGKPKG	FDLS	NSYGPTE
Consensus	ltY el	LkAggayvPld	lAYiifTSGstG PKG	FDlS	nlyGpTE

	A6	A7	A8	A9	A10
LnmQ	GELCVTGPQMFDGYL	YRTGDR	GRDDGQVKIHGYRVELSEVE	LPPYMLP	NGKTDR
Ave-NRPS8	GELYVRGPLRFPGYL	YRTGDR	GRIDHQVKIRGHRIELGEIE	LPPYMLP	NGKIDR
LnmI-A	GDLYIGGECLALGYV	YKTGDR	GRADGQVKVRGFRVELAEIE	LPEYMVP	NGKLDR
GrsA-A	GELCIGGEGLARGYW	YKTGDQ	GRIDNQVKIRGHRVELEEVE	LPTYMIP	NGKIDR
CssA-A1	GELVVSGDGLARGYT	YRTGDR	GRMDQQVKIRGHRIEPAEVE	LPSYMMP	NGKVDR
HTS1-A3	GELCIEAPSLARCYL	YRTGDL	GRKDGQIKLRGQRIELGEIE	LPHYMVP	SGKLDH
Consensus	GeLci gpglargYl	YrTGDr	GR DgQvKirGhRiEl EvE	LP YM P	nGK Dr

B



C

	235	236	239	278	299	301	322	330	331	517
LnmQ	D	L	F	T	V	C	L	I	A	K
CssA-A1	D	L	W	F	Y	I	A	V	V	K
HTS1-A3	D	L	L	F	G	I	S	V	L	K

Figure 4. Functional Analysis of *lnmQ* in LNM Biosynthesis

(A) Conserved motifs of LnmQ in comparison with other NRPS A domains. The conserved amino acids are in bold. LnmQ homologs (accession numbers in parentheses) are as follows: Ave-NRPS8 (BAB690421) from *Streptomyces avermitilis*, GrsA (AAB22346) from *Bacillus subtilis*, CssA (S41309) from *Tolypocladium niveum*, and HTS1 (Q01886) from *Cochliobolus carbonum*.

(B) HPLC analysis of LNM production by *S. atroolivaceus* wild-type and recombinant strains. I, LNM standard; II, wild-type S-140; III, SB3018 (Δ*lnmQ*); IV, SB3020 (SB3018 harboring the *lnmQ* overexpression plasmid pBS3047). Filled circles, LNM; open triangles, an unknown metabolite whose production is independent of LNM biosynthesis.

(C) The substrate selectivity-conferring code of LnmQ in comparison with codes for the *D*-Ala-specific A domains of CssA-A1 and HTS1-A3. The same and similar residues are highlighted in bold.

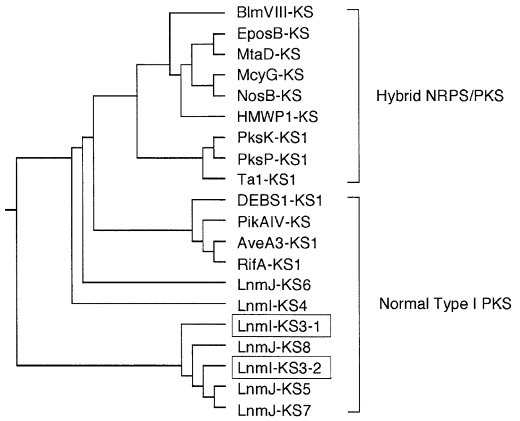
$(M+H)^+$ and 519.124 for $(M+Na)^+$] [60]. The molecular weight difference of 32 between these two compounds (i.e., 2 H atoms versus 2 OH groups) agrees perfectly with the predicted function of LnmA as a cytochrome P-450 hydroxylase and suggests that inactivation of *lnmA* apparently prevents hydroxylation of the corresponding LNM biosynthetic intermediate at both C-8 and C-4′ positions.

The *lnmD* gene encodes a protein exhibiting low sequence homology to several 3-oxoadipate enol-lactone hydrolase/4-carboxymuconolactone decarboxylases, such as MupV from *Pseudomonas fluorescens* (accession number AAM12938; 23% identity and 37% similarity). The *lnmF* gene product resembles known enoyl CoA hydratases, such as MupJ from *Pseudomonas fluorescens* (accession number AAM12923; 19% identity and 31% similarity). The deduced gene product of *lnmW* shows high sequence homology to a family of 4-coumarate-CoA ligases, such as the one from *Caulobacter crescentus* (accession number B87644; 24% identity and 37% similarity). It is not clear what role these proteins could play in LNM biosynthesis.

The deduced gene product of *lnmN* is highly homologous to other known type II TEs, such as GrsT from *Bacillus subtilis* (accession number P14686; 36% identity and 58% similarity) or McyT from *Planktothrix agardhii* (accession number CAD29792; 41% identity and 58% similarity), and is characterized by the conserved active site motifs of GxSxG and GxH [38]. Type II TEs have been identified from both polyketide and nonribosomal peptide biosynthetic gene clusters. It has been generally accepted that type II TEs play an "editing" role by removing misprimed ACPs (for PKS) or PCPs (for NRPS) to ensure the catalytic efficiency and fidelity in polyketide [43, 44] or peptide biosynthesis [45], respectively. The identification of the LnmN type II TE from the *lnm* gene cluster provides an excellent opportunity to investigate if LnmN plays a similar editing role in a mechanistic analogy to type II TEs for PKSs or NRPSs, but with a relaxed substrate specificity by removing misprimed acyl intermediates from both the ACP and PCP domains of the LNM hybrid NRPS-PKS megasynthetase.

Regulatory, Resistance, and Other Genes Encoding Proteins of Unknown Function

The *lnmO* gene is the only apparent regulatory gene identified within the *lnm* cluster. The deduced gene product of *lnmO* is a 25.7 kDa protein that belongs to a family of transcriptional activators, such as DNR from *Alcaligenes faecalis* (accession number BAA90776; 27%



B

LnmJ-ACP6-1	FGVDSLVSL
LnmJ-ACP6-2	FGVDSLVSL
LnmJ-ACP4	LGLDSGHML
LnmJ-ACP7	LGFDSITLT
LnmJ-ACP8	LGLDSIYRM
LnmI-ACP3	LGVESVLLG
LnmJ-ACP5	MGVDSVGAV
LnmL	LGADSVDRV
Consensus	lGvdS

Figure 5. Sequence Analysis of the LnmIJ PKSs

(A) Phylogenetic analysis of the LnmI and LnmJ KSs and their homologs from other polyketide and hybrid peptide-polyketide biosynthetic gene clusters. KS-containing homologs (natural product/accession numbers in parentheses) are as follows: BlmVIII (bleomycin/AF210249), EposB (epothilone/AF217189), MtaD (myxothiazole/AF188287), McyC (microcystin/AF183408), NosB (nostopeptolide/AF204805), HMWP1 (yersiniabactin/af091251), PksK and PksP (unknown compound/AL009126), TA1 (TA antibiotic/AJ006977), DEBS1 (erythromycin/Q03131), PikAIV (pikromycin/AF079138), AveA3 (avermectin/AB032367), and RifA (rifamycin/AF040570).

(B) Compilation of the core sequences of LnmI and LnmJ ACP domains as well as the discrete LnmL ACP. The invariant Ser and other conserved residues are in bold.

identity and 47% similarity) or BTR from *Bordetella pertussis* (accession number Q08530; 27% identity and 45% similarity). Inactivation of *lnmO* by gene replacement completely abolished LNM production, confirming its involvement in LNM production (G.-L.T., B. Yun, Y.-Q.C., and B.S., unpublished data).

At least five genes, *lnmR*, *lnmS*, *lnmT*, *lnmU*, and *lnmY*, could be identified within the *lnm* gene cluster whose deduced gene products could confer LNM resistance to *S. atroolivaceus* S-140. LnmR shows high sequence homology to various ATP hydrolases, such as MoaD from *Agrobacterium tumefaciens* (accession number T45539; 27% identity and 39% similarity) or YliA from *E. coli* (accession number P75769; 27% identity and 39% similarity). Both LnmS and LnmT show significant sequence homology to transmembrane proteins, such as AgaC from *Agrobacterium tumefaciens* (accession number T45530; 29% identity and 42% similarity) and DppC from *Brucella melitensis* (accession number AF3306; 28% identity and 40% similarity) or AgaB from *Agrobacterium tumefaciens* (accession number T45531; 28% identity and 42% similarity) and DppB from *Brucella melitensis* (accession number AF3535; 27% identity and 41% similarity), respectively, but share no sequence homology with each other. LnmU exhibits significant sequence homology to a family of periplasmic oligopeptide binding proteins, such as OphA from *Agrobacterium tumefaciens* (accession number C98307; 26% identity and 41% similarity) or DppA from *Brucella melitensis* (accession number P23847; 22% identity and 37% similarity). Together, LnmR, LnmS, LnmT, and LnmU could be envisaged to constitute an ATP binding cassette transporter complex [46] for active transport of LNM out of producing cells. LnmY displays significant sequence homology to various antibiotic efflux proteins, such as SCH42.31c from *Streptomyces coelicolor* (3)2 (accession number T35130; 36% identity and 52% similarity) or McT from *Streptomyces lavendulae* (accession number AAD32747; 32% identity and 52% similarity), which could potentially provide an alternative resistance mechanism to *S. atroolivaceus* S-140 by passively dispersing LNM across the cell membrane.

The remaining genes, *lnmC*, *lnmE*, *lnmH*, *lnmK*, *lnmV*, *lnmX*, and *lnmZ′*, identified within the *lnm* gene cluster encode proteins that either show no significant sequence homology to any proteins in the databases or resemble proteins with unknown functions. While it remains to be established what role these proteins could play in LNM biosynthesis, inactivation of each of them by gene replacement completely abolished LNM production, confirming unambiguously that they are essential for LNM biosynthesis (G.-L.T., B. Yun, Y.-Q.C., and B.S., unpublished data).

Discussion

The LNM Biosynthetic Gene Cluster Consisting of 27 Genes

Given the unprecedented structure of LNM, we reasoned that knowing the precise boundaries of the *lnm* gene cluster should greatly facilitate our effort to postulate a model for LNM biosynthesis and to carry out functional analysis of the *lnm* gene cluster. Two rounds of gene inactivation were carried out, identifying the upstream boundary to be between *orf(−1)* and *lnmA*. On the basis of sequence analysis, we initially predicted that the upstream boundary of the *lnm* cluster would start from *orf(−2)* that encodes an NRPS module. To test this hypothesis, we chose to inactivate *orf(−2)* as a target within the *lnm* cluster and *orf(−11)* and *orf (−13)*, which encode an NRPS module and a discrete NRPS condensation (C) enzyme, respectively, as targets beyond the predicted upstream boundary. They were replaced, respectively, with a mutant copy in which the target ORF was substituted with the apramycin-resistance gene, *aac(3)IV* [27]. To our surprise, the resultant *S. atroolivaceus* SB3007 [Δ*orf(−13)*], SB3008 [Δ*orf(−11)*], and SB3009 [Δ*orf(−2)*] mutant strains exhibited the same phenotype as the wild-type S-140 strain and produced LNM at a level comparable to the S-140 strain from HPLC analysis (Figure 3A, II versus III, IV, or V). These results suggested that *orf(−2)*, *orf(−11)*, and *orf(−13)* are all outside of the *lnm* gene cluster. We next inactivated *orf(−1)*, which also encodes a discrete NRPS C enzyme, and *lnmA*, which encodes a cytochrome

A

```
LnmJ-AT-dock5     LADIAHTLRV GRSPLAVRLA VVCGEPEELR R---RLAAFL DGDEPGEGVF TGRADDDKEP VRLERA----
PedH-AT-dock11    LADIAFTLQT GRKAMDFRLA VVVEGVEARL RAVESLRAYL RNETPGPTVF VDNVLEDHSR VREQLVGSAG
LnmI-AT-dock3-2   LADVAHTLQS GREPLRERVA LVAYDVAGLC RALDLFAS-- GDTG--AWVH GRTPGGALPD GPK-------
LnmJ-AT-dock7     LADLAHTSRV GRRELPERLA VTAASHAQLA ARLREFAA-- TGVAGEGVST GTARKGGAGS GLG-------
LnmJ-AT-dock6     PADLAFTLGV GRAHLPVRAA VIARNVPELR RRLRLLQS-- GAQAPGCFRT GQGAAAGDLD EQTRAELAGR
MmpD-AT-dock2     VVDMSYSLLT GRQRFEQRFH VVVADRAELI AALRRGMP-- ADTADLAEAH QRRLQG---- -LSARATGQS
PedF-AT-dock5     LADLAYTLQT GREAMEQRVA LLVGDLAGLL EALSALRE-- ERPCPVSVWS GRVEPGPSRG AETVNADQPA
LnmI-AT-dock3-1   LDDVAWTLQT GRASLGHRLT LSADGLDGVR AGLTAFLD-- GRACPG---- ---------- -LATAAADPA
LnmJ-AT-dock8     LARVAYTLQT GRTGHRHRFA VRVRDRDELI GALEAFAA-- GELPDHAATG ---------- --TARRDAPS
PedF-AT-dock1     LRDIAYTLQV GRDAFEHRLA LVVDSQQQLI EGLECYLE-E RQPSQGEG-- ---------- -AVYQGQVAS
PedH-AT-dock8     IVDIAYTLQV GREALSQRLA LIVTDLVDLK TRLRSLLE-G GEEPSGVYR- GDTKANKAAL QEIDDDDRSL
PedF-AT-dock6     LHDIAYTLQV GREAMPRRLA LAVTSLAQLA DRLQTWLE-Q PTQTEGVQQG LVTQEAEEQF DTVLGDEDRA
PedF-AT-dock4     LFDIAYTLQV GREALDERLG LVAVSLQELS RQLAAFLG-E EAEQPLLYRG -RVQRNKDAL QALANDEEFQ
PedH-AT-dock9     LADLAYTLQV GRDAMAERLA MTADSMEELE HKLRAFVEGR SGEVKDLYQG -SVKQNKRIL SAFAGDEEMQ
MmpA-AT-dock5     LRDIAYTLQV GRQAMNARLS CIATSTADLM DALRRYCA-- -GEAHPGVQS -TTLKDADRL SLFGQDESAL
MmpB-AT-dock      LHDLAYTLQV GREALDARAA FTAQSVQVLK ERLVALAD-- GAQHPDVLIG ----QALKPV RLRAG---ET
MmpD-AT-dock3     LHDLAYTLQT GREALNARLG FLAHSIDDVQ ACLREYLQ-- GALTSGRVQV GSARQDENPL VRLLGEDDLS
PedH-AT-dock10    LRNVAYTLQV GREAMQHRLA FSARSIEDAR RILEAFAQ-- GREVARLYRG ------YVKT ARDSRSGRRD
MmpA-AT-dock6     LQALAYTLQV GREAWEWRVA LLVTSGDELV RELRAFID-G ALEGPSWWSG ---CLPEAHS LATRPSEQAC
PedF-AT-dock3     LQDLAYTLQV GRQAMDWRVA FLVKDLHDLS EKLERFLQ-G DSLVQDCFQG ---RVATS-- -----VMDAA
MmpD-AT-dock4     MAAMAYTLMA GRKHHEWRLA MVTHDAGQLQ QSLEGWLQ-G RDEATVHSGH WDVRRFVEQQ EVLQAARECL
Consensus         l d aytlqv GR     Rla         l    l

LnmJ-AT-dock5     ----AELFR- --LGRLSELA RAWADGAAVE WDDCRAGDG- --VRPRRVPL PAHPLD---E RSYW
PedH-AT-dock11    QAILQRALM- --EPDLRALA GWWVKGIKLP WHQLHAG--- --WKRKRVVL PTYPFE---R KSYW
LnmI-AT-dock3-2   AVLDAAADR- --DAELLRLG RHWTGGGTVD WPGLHP---- --VRRRLVSL PSYPFA---E DRHW
LnmJ-AT-dock7     AQELTAALA- --GRRWADAV EHWTLGGRVD WRTADAG--- --RLVRKVAF PTYPFN---R SRHW
LnmJ-AT-dock6     ARSGSPAER- --VAALERLA AAYAAGQDLD WQSLSYG--- --DRPRRVPL PTYPFG---G DRHW
MmpD-AT-dock2     LYYLQ----- ----QLAVFA DAWAAGQAVD FAPLFDA--- --LDCQRMPL PGYPF----AR ERYW
PedF-AT-dock5     AELLQRIPQW LAEGALDELA QAWVAGAPID WCQLRRR--- --RPPRRVHL PSYPF----AR ERYW
LnmI-AT-dock3-1   LAGVPAG--- --AQDLARAA GIWLRGHAVD FARLWS---- --APARRVPL PVQDFTVLAQ ERHW
LnmJ-AT-dock8     VQSDE----- ----DPALLR KSWCEGADVP WHTWWPK--- --TPG-RVPL PTAPF----AR TRHW
PedF-AT-dock1     ESQSLPFT-- --EDDLAAVA RQWVAGGAVL WP-----VPV GPKKPRRVRL PAYPF----DK RAYW
PedH-AT-dock8     EKLIAYFS-- --QDDVHKLA KIWTQGVEVD WPSLYARMPF AGRSPRRVAL PTYPF----AR QRHW
PedF-AT-dock6     AAVERWVE-- --KGQYEKLL DAWTRGWAID WN----VLYC TDTRPRRIGL PTYPF----AR RRYW
PedF-AT-dock4     ETVDKWLA-- --RRKYSKLL KHWVTGLSVD WTRLYS---- D-VLPRRIRL PVYPF----VR QRYW
PedH-AT-dock9     EALDKWIQ-- --RGKLAKLL EIWVAGLNID WQQLYGSDRS G-TPPHRISA PGYPF----AE QRHW
MmpA-AT-dock5     TLLDQWYA-- --ESKWAQLA QIWRAGEVID WARLVAPG--- ---TARRISL PGYSF----DR EPIP
MmpB-AT-dock      AVPVQSPA-- --MDDNQAVM RHWVSGGQVE WAQLYQGF-- ---QPARISL PLYPF----VR ERCW
MmpD-AT-dock3     AMVAQWAA-- --TGQQDKLL AIWVSGGDID WQALLPGR-- ---WGRIISL PGYPF----AR ERFW
PedH-AT-dock10    ESVAEPIR-- --GKDHDAVL AIWVKGVDVN WQELYAAES- --DLPYRISL PTYPF----AR ERYW
MmpA-AT-dock6     AAIRQMFD-- --QADLGGIL RHWVQGEAVD WSPLYHSH-- ---RPARLGL PTYPF----AR QRYW
PedF-AT-dock3     ATPLPVAQ-- --DREQAAIA KAWVTGRLVD WKELPRRG-- ---TPHRISL PTYPF----AE ERYW
MmpD-AT-dock4     IRLMGNRDQS TRGEALTSLA QAFVGGYELD YSSLFMAD-- ---ERRRVPL PTYPFL---R QSYW
Consensus                               w G   d w  l           rrv l Ptypf    r  ryw
```

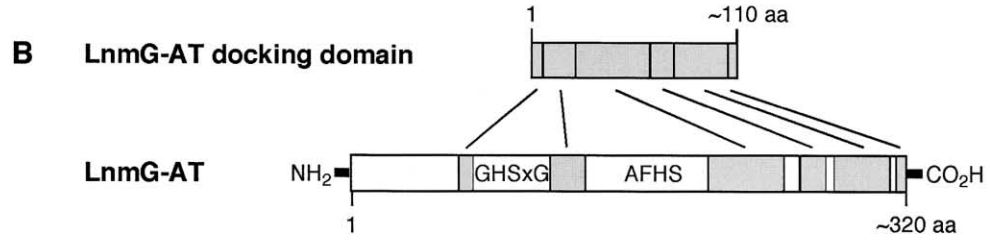


Figure 6. Sequence Analysis of AT-less PKSs

(A) Compilation of deduced amino acid sequences of AT docking domains from LnmI and LnmJ and other AT-less PKSs. PedF and PedH are AT-less PKSs from the pederin biosynthetic gene cluster (accession number AY059471), and MmpA, MmpB, and MmpD are AT-less PKSs from the mupirocin biosynthetic gene cluster (accession number AF318063). The conserved amino acid residues are highlighted in bold.
(B) Diagram of homologous regions between the AT docking domain and the functional LnmG-AT domain. Regions containing the conserved active site of GHSxG and substrate binding motif of AFHS for a functional AT domain are absent in the AT docking domain.

P-450 hydroxylase, leading to the isolation of the *S. atroolivaceus* SB3010 [Δ*orf(−1)*] and SB3011 (Δ*lnmA*) mutant strains, respectively. The SB3010 strain produced LNM in a level comparable to the wild-type S-140 strain, indicative that inactivation of *orf(−1)* has no apparent effect on LNM production. In contrast, the SB3011 strain completely lost its ability to produce LNM and instead accumulated new metabolites that were not detected in the S-140 strain, suggesting that *lnmA* is essential for LNM biosynthesis (Figure 3A, II versus VI or VII; the abolishment of LNM production in VII was confirmed by LC-mass spectrometric analysis). Taken together, these results allowed the assignment of the upstream boundary of the *lnm* gene cluster to be between *orf(−1)* and *lnmA* (Figure 2).

The downstream boundary of the *lnm* gene cluster was defined by the same strategy to be between *lnmZ′* and *orf(+1)*. Initial sequence analysis failed to predict a

putative downstream boundary due to its complex genetic organization. Within the sequenced downstream region, the genes appear to be organized as multiple transcriptional units, as evidenced by the frequent change of their transcriptional directions, and many of the ORFs lack homology to proteins of known function in the databases, further complicating sequence-based functional assignment. Therefore, a series of gene replacement experiments were carried out to inactivate *lnmZ′*, *orf(+1)*, *orf(+2)*, *orf(+3)*, *orf(+4)*, and *orf(+6)*, which were predicted to encode proteins of unknown function [LnmZ′, ORF(+1), ORF(+3)], a type II thioesterase [ORF(+2)], a TetR-family transcriptional regulator [ORF(+4)], and a putative hydrolase/lactonase [ORF(+6)], respectively. The Δ*lnmZ′* mutant strain, SB3017, essentially lost its ability to produce LNM. It instead accumulated new metabolites that were not detected from the S-140 strain, suggesting that *lnmZ′* is critical for LNM biosynthesis (Figure 3B, I versus VII; the trace amount of LNM produced in VII was confirmed by LC-mass spectrometric analysis). In contrast, the other mutant strains, SB3016 [Δ*orf(+1)*], SB3015 [Δ*orf(+2)*], SB3014 [Δ*orf(+3)*], SB3013 [Δ*orf(+4)*], and SB3012 [Δ*orf(+6)*], exhibited the same phenotype as and produced LNM in a level comparable to the S-140 strain (Figure 3B), indicating that they are all outside of the *lnm* gene cluster (Figure 3B, I versus II, III, IV, V, or VI). Taken together, these results showed the downstream boundary of the *lnm* gene cluster to be at the end of *lnmZ′* (Figure 2).

Model for LNM Biosynthesis

Precise determination of the *lnm* gene cluster boundaries allowed us to propose a model for LNM biosynthesis on the basis of the genes within the cluster and their deduced functions (Figure 1). According to the hybrid NRPS-PKS assembly-line enzymology [20, 29, 47], LNM biosynthesis could be envisaged to begin at the loading NRPS module consisting of LnmQ and LnmP. LnmQ selects, activates, and loads a *D*-Ala to LnmP to initiate LNM biosynthesis. The LnmI NRPS module then selects, activates, and loads a *L*-Cys to its cognate PCP and catalyzes the condensation between the aligned *D*-Ala-*S*-PCP and *L*-Cys-*S*-PCP. Subsequent cyclization and oxidation yield the thiazonyl-S-PCP intermediate. At this point, the growing peptidyl-*S*-PCP intermediate is switched from the NRPS to the PKS assembly-line biosynthetic machinery. The discrete LnmG AT provides the missing AT activity in *trans* to LnmI and LnmJ and loads the malonyl CoA extender units to all ACP domains of the six LnmI and LnmJ PKS modules. The docking domains within the AT-less PKS modules could facilitate the interaction between LnmG and LnmI and LnmJ. Sequential elongations of the thiazonyl-*S*-PCP intermediate by the LnmI and LnmJ PKS modules complete the biosynthesis of the LNM hybrid peptide-polyketide carbon backbone. The full-length acyl-*S*-ACP intermediate is then released and cyclized by the TE domain of LnmJ to yield a macrolactam intermediate such as **1** (Figure 1).

The unprecedented structure of the 1,3-dioxo-1,2-dithiolane moiety of LNM has precluded us from predicting its biosynthetic pathway a priori. However, the identification of LnmL (ACP) and LnmM (HMG-CoA synthase) within the *lnm* gene cluster provided a clue that allowed us to speculate about how the 1,3-dioxo-1,2-dithiolane moiety could be fused to **1** in LNM biosynthesis. We propose that LnmM, in a mechanistic analogy to HMG-CoA synthase, could catalyze the condensation of methylmalonyl CoA at the β-keto group of **1** (or its linear acyl-*S*-ACP precursor before cyclization) to afford intermediate **2**. The Similar reactions have been proposed for the biosynthesis of several polyketides with alkyl side chains [40, 48]. Methylmalonyl CoA could also be presented to LnmM in its ACP-activated form of methylmalonyl-*S*-LnmL, the formation of which could be catalyzed by LnmG. Tethering metabolites of primary metabolism to carrier proteins such as ACP and PCP as a general strategy to sequester and thus divert them to secondary metabolism has been proposed previously [49, 50]. Further modifications of **2**, via **3** as a possible intermediate, by other tailoring enzymes could finally afford LNM. LnmA (P-450 hydroxylase), LnmB (ferredoxin), and LnmZ (P-450 hydroxylase) could serve as excellent candidates to catalyze the C-8 and C-4′ hydroxylation and S-1′ oxidation, respectively, thus converting **3** to LNM (Figure 1). While the order of many of these steps has to be determined experimentally, the proposed model for LNM biosynthesis is consistent with the genes and the functions of their deduced gene products identified within the *lnm* cluster.

Unprecedented Architectural Complexity of the LNM Hybrid NRPS/PKS Megasynthetase

Initiation and Peptide Biosynthesis by NRPS

According to the LNM structure, the current paradigm for hybrid peptide-polyketide biosynthesis would predict two NRPS modules for the biosynthesis of the peptide moiety of LNM [20, 29, 47]. An NRPS loading module, presumably having an A-PCP-epimerase (E) domain organization, would select and activate the readily available *L*-Ala. Upon loading of *L*-Ala to the cognate PCP to form an *L*-Ala-*S*-PCP intermediate, the E domain would convert *L*-Ala into *D*-Ala to set the stage for chain elongation. Alternatively, the loading module could be of A-PCP organization, indicative of direct activation and loading of *D*-Ala to the cognate PCP to initiate LNM biosynthesis. Sequence analysis of the *lnm* gene cluster, however, failed to identify any NRPS with either A-PCP-E or A-PCP organization. Instead, we only identified LnmQ and LnmP, a pair of discrete NRPS A and PCP proteins, within the *lnm* cluster (Figure 2).

The proposal that LnmQ and LnmP constitute the loading module to initiate LNM biosynthesis is consistent with the fact that *lnmQ* and *lnmP* are translationally coupled, ensuring that LnmQ and LnmP are produced in equimolar amounts for optimal interactions. This proposal also agrees well with the *D*-Ala specificity of LnmQ predicted according to the amino acid specificity-conferring codes of the A domain [51, 52]. Although a *D*-Ala-specific A domain of bacterial origin was not known prior to this work, two *D*-Ala-specific A domains of fungal origin, CssA-A1 for cyclosporin biosynthesis from *Tolypocladium niveum* [53] and HTS1-A3 for HC-toxin biosynthesis from *Cochliobolus carbonum* [54], have been characterized. Moderate conservation of the amino acid

specificity-conferring codes between LnmQ and the two known *D*-Ala-specific A domains is apparent (Figure 4C), supporting the functional assignment of LnmQ and LnmP to constitute the loading module that directly activates and incorporates *D*-Ala into LNM biosynthesis. LnmQ and LnmP therefore represent a novel architecture for an NRPS loading module consisting of discrete A and PCP proteins (Figure 1, module 1).

Following the priming of LnmP with *D*-Ala by LnmQ, the second NRPS module (module 2), residing at the N terminus of LnmI with a Cy-Cy-A-PCP-Ox domain organization, elongates the *D*-Ala-*S*-PCP with the cognate *L*-Cys-*S*-PCP and cyclizes and oxidizes the resultant *D*-alaninyl-*L*-cysteinyl-*S*-PCP to yield the thiazonyl-*S*-PCP intermediate. The LnmI NRPS module is characterized by tandem Cy domains, which are rare and have only been previously identified for the vibriobactin gene cluster from *Vibrio cholerae* [55]. In a mechanistic analogy to VibF, we propose that the second Cy domain is responsible for condensation, yielding the *D*-alaninyl-*L*-cysteinyl-*S*-PCP intermediate that is subsequently cyclized by the first Cy domain to afford the *D*-alaninylthiazolinyl-*S*-PCP intermediate. Oxidation of the latter by the Ox domain then yields the thiazonyl-*S*-PCP intermediate, setting the stage for the elongation steps to switch from the NRPS to the PKS assembly-line machinery (Figure 1, module 2).

NRPS-to-PKS Transition

A fundamental question for hybrid peptide-polyketide biosynthesis is how a hybrid NRPS-PKS controls the transition between the NRPS and PKS biosynthetic machinery [20, 47]. We have previously shown that the KS domains from PKS modules at the hybrid NRPS/PKS interface are unique in comparison with those from normal PKS modules. While all KSs contain the highly conserved Cys-His-His catalytic triad, those from PKS modules at the hybrid NRPS/PKS interface fall into a distinct subfamily in a phylogenetic analysis. The latter observation led to the speculation that PKS modules at the hybrid NRPS/PKS interface might have evolved a unique KS domain to facilitate the transition from peptide to polyketide biosynthesis [20]. Surprisingly, the LnmI PKS module at the hybrid NRPS/PKS interface (Figure 1, module 2/module 3) is characterized by an unprecedented tandem KS architecture, but neither of the KS domains fall into the hybrid NRPS/PKS subfamily of KSs (Figure 5A). Instead, they are more closely related to KSs from normal PKS modules, with the exception that the first KS contains a mutated catalytic triad of Cys-Ala-His. Since the His-His residues are essential for malonyl-*S*-ACP decarboxylation to generate the corresponding carbon anion, and the Cys residue catalyzes condensation between the resultant carbon anion and the acyl-*S*-KS to form a C-C bond [30–33], the first KS domain alone cannot be sufficient to catalyze the entire chain elongation step. We propose that the first KS domain catalyzes the transfer of the growing peptide intermediate of peptidyl-*S*-PCP from the upstream NRPS module (module 2) to its Cys residue, and the second KS domain catalyzes the decarboxylative condensation between the resulting peptidyl-*S*-KS and the cognate malonyl-*S*-ACP (module 3) to complete the elongation step. The LnmI hybrid NRPS/PKS protein with tandem KS domains in the PKS module therefore might represent a novel mechanism to facilitate the transition from peptide to polyketide biosynthesis (Figure 1, module 3).

Elongation and Termination by PKS

The LnmI PKS-bound growing intermediate continues to be elongated by the five PKS modules (modules 4 to 8) on LnmJ, furnishing the full-length LNM peptide-polyketide backbone (Figure 1). While the deduced LnmI and LnmJ PKS functions are consistent with what would be required for the biosynthesis of the polyketide moiety of LNM from the acyl CoA precursors, several notable features of the LnmI and LnmJ PKS are unprecedented. Most strikingly, none of the six PKS modules of LnmI and LnmJ contain the cognate AT domain. We have previously named LnmI and LnmJ as AT-less PKSs and demonstrated that the missing AT activity is provided by the discrete LnmG AT enzyme that acts iteratively in *trans* to load the malonyl CoA extender units to ACP domains of the six PKS modules for LNM biosynthesis [24]. Intriguingly, the short segments sandwiched between KS and the downstream domains within an AT-less PKS module appear to be conserved among all AT-less PKSs known to date (Figure 6). We now propose that they may serve as indigenous docking sites and facilitate protein-protein interaction between AT-less PKSs and discrete ATs to constitute a functional megasynthetase (Figure 1, module 3 to module 8).

Other novel features of the LnmG, LnmI, and LnmJ PKS megasynthetase include domain redundancy, domain misposition, having a domain missing, or having an extra domain. As examples, (1) PKS modules 3 and 6 each lack a DH domain, and PKS module 7 lacks both DH and ER domains. (2) PKS module 6 has two ACP domains flanking an MT domain, although the presence of the MT domain in PKS module 6 is consistent with the malonyl-CoA-specificity of LnmG, indicating that the methyl group at C-6 is of AdoMet origin. (3) Extra domains unprecedented in PKS are found between PKS module 8 and the TE and at the C terminus of LnmG, respectively (Figure 1). It should be pointed out that the definition of these features as atypical is very subjective, based on the so-called "colinearity rule" for modular PKSs [20, 29, 30, 47]. However, numerous exceptions to the latter have been observed recently, and PKSs with atypical domain and modular organizations may be much more common than currently thought [56]. While the details of how the LNM hybrid NRPS-PKS megasynthetase catalyzes LNM biosynthesis from the amino acid and short carboxylic acid precursors have to wait for future in vivo and in vitro experimentation, the current analysis provides a working model that can be used to formulate research hypotheses and to design experiments to further these investigations. Regardless of these unusual features, LNM biosynthesis appears to be terminated by a mechanism common to both polyketide and peptide biosynthesis [29, 30, 38]. The LnmJ TE offloads the full-length hybrid peptide-polyketide intermediate from the LNM hybrid NRPS-PKS megasynthetase and cyclizes it into the macrolactam intermediate **1**. Post-NRPS and -PKS modifications by the tailoring enzymes finally afford LNM (Figure 1).

Significance

A 135,638 bp DNA region that encompasses the biosynthetic gene cluster for the antitumor antibiotic LNM

was sequenced from *S. atroolivaceus* S-140. Systematic inactivation of ORFs within this region resulted in the precise determination of the *lnm* gene cluster boundaries. Bioinformatic and genetic analysis of the *lnm* cluster allowed us to propose a model for LNM biosynthesis. The assembly of the hybrid peptide-polyketide backbone of LNM from the amino acid and carboxylic acid precursors is proposed to be catalyzed by the LNM hybrid NRPS-PKS megasynthetase with an unprecedented architectural complexity, consisting of discrete and modular NRPSs, AT-less PKSs, and PKS modules with unusual domain organizations. Modifications of the nascent hybrid peptide-polyketide intermediate to form LNM are proposed to involve tailoring enzymes that catalyze novel chemistry for the introduction of an alkyl branch into the polyketide backbone and the formation of the 1,3-dioxo-1,2-dithiolane moiety. These findings set the stage to investigate the molecular basis of LNM biosynthesis and to apply combinatorial biosynthesis methods to the LNM biosynthetic machinery to increase structural diversity for anticancer drug discovery.

Studies on peptide, polyketide, and hybrid peptide-polyketide biosynthesis in the past decade have benefited greatly from the "colinearity rule" for most of the modular PKSs, NRPSs, and hybrid NRPS-PKSs known to date. Innovations in methodologies for cloning biosynthetic gene clusters and advances in technologies for DNA sequencing and bioinformatic analysis have facilitated the unveiling of NRPSs, PKSs, and hybrid NRPS-PKSs with novel mechanisms and structures. The LNM hybrid NRPS-PKS megasynthetase is exceptional in this regard for its unprecedented architectural complexity. These findings underscore once again nature's versatility in evolving complex pathways for natural product biosynthesis. They also provide new opportunities to study the fundamental enzymology of and to develop new combinatorial biosynthesis methods for the NRPS, PKS, or hybrid NRPS-PKS biosynthetic machinery.

Experimental Procedures

Strains, Plasmids, Chemicals, Biochemicals, and LNM Production and Analysis

The wild-type *S. atroolivaceus* S-140 strain [23, 24], *E. coli* S17-1 [57], plasmids pSET151 [57] and pBS3031 [24], the *aac(3)IV* apramycin-resistance gene cassette [25], and the *ErmE** promoter [25] were described previously. Common chemicals and biochemicals were from commercial sources. LNM production and isolation from both the wild-type and recombinant *S. atroolivaceus* strains and LNM analysis by HPLC and mass spectrometry were carried out as previously reported [23, 24].

DNA Sequencing, Analysis, and Manipulation

The *lnm* gene cluster was previously localized to five overlapping cosmids, pBS3004, pBS3005, pBS3006, pBS3007, and pBS3008 [24]. DNA sequencing by a shotgun method of the first four cosmids yielded a 135,638 bp contiguous DNA sequence. Bioinformatic analyses of DNA and protein sequence were carried out with the Genetics Computer Group (GCG) program (Madison, WI) [58] or the Clustal W program [59]. Functional assignments were made by utilization of the BLAST server at the National Center for Biotechnology Information (Bethesda, MD) and comparison of the deduced gene products with proteins of known functions in the database. General procedures for genetic manipulations in *E. coli* and in *S. atroolivaceus*, the conjugation between *E. coli* S17-1 and *S. atroolivaceus*, selection of exconjugants, and confirmation of homologous recombination by Southern hybridization were performed as described previously [23, 24].

Inactivation by Gene Replacement

To inactivate *orf(−13)*, XbaI and HindIII sites were introduced into *orf(−13)* by PCR using the following two pairs of primers: 5′-TAA TACGACTCACTATAGGGCGA-3′/5′-GCTCTAGACTCCTTCGACCT GTTCGACC-3′ (the XbaI site is underlined) and 5′-GGAGAAGCTT GACGAAGAAGCCGATGAGC-3′/5′-GACAGCGATGACCGAGAC-3′ (the HindIII site is underlined). A 1.5 kb HindIII-XbaI fragment containing the *aac(3)IV* apramycin-resistance gene was inserted into the engineered XbaI and HindIII sites, resulting in the replacement of a 326 bp internal fragment of *orf(−13)* with *aac(3)IV*. The mutated *orf(−13)* was then moved as a 3.7 kb Eco*R*I-PstI fragment into the same sites of pSET151 to yield pBS3034. To inactivate *orf(−11)*, an internal 12 bp NcoI-NotI fragment was replaced with a 1.5 kb NcoI-NotI fragment containing *aac(3)IV*, and the mutated *orf(−11)* was cloned as a 6.0 kb XbaI-HindIII fragment into the same sites of pSET151 to yield pBS3035. To inactivate *orf(−2)*, an internal 695 bp BglII-BglII fragment was replaced with a 1.5 kb BglII-BamHI fragment containing *aac(3)IV*, and the mutated *orf(−2)* was transferred as a 9.3 kb EcoRI fragment into the same site of pSET151 to afford pBS3036. To inactivate *orf(−1)*, an internal 1263 bp NcoI-NotI fragment was replaced with a 1.5 kb NcoI-NotI fragment containing *aac(3)IV*, and the mutated *orf(−1)* was cloned as a 4.7 kb SphI-EcoRI fragment into the same sites of pSET151 to furnish pBS3037. To inactivate *lnmA*, an internal 428 bp XhoI-PvuII fragment was replaced with a 1.5 kb XhoI-PvuII fragment containing *aac(3)IV*, and the mutated *lnmA* was moved as a 2.8 kb *Hind*III-EcoRI fragment into the same sites of pSET151 to yield pBS3038.

To inactivate *orf(+6)*, a BglII site was introduced into *orf(+6)* by PCR using the following pairs of primers: 5′-GGCTACGCATGCTAT CTCGCCGAGAAGG-3′/5′-GGTGAGATCTTGCGAGTTGGTGCGTG-3′ and 5′-TCGCAAGATCTCACCCATGCTTCAGTCC-3′/5′-GAAGAA TTCTGATGAAAAGACCCTGTG-3′ (the BglII site is underlined). A 1.5 kb BglII-BamHI fragment containing *acc(3)IV* was then inserted into the engineered BglII site, and the mutated *orf(+6)* was cloned as a 3.4 kb SphI-EcoRI fragment into the same sites of pSET151 to afford pBS3039. To inactivate *orf(+4)*, an internal 243 bp NotI-AatII fragment was replaced with a 1.5 kb NotI-AatII fragment containing *acc(3)IV*, and the mutated *orf(+4)* was moved as a 3.0 kb SphI-EcoRI fragment into the same sites of pSET151 to yield pBS3040. To inactivate *orf(+3)*, a 1.5 kb KpnI-BamHI fragment containing *acc(3)IV* was inserted into the internal KpnI site, and the mutated *orf(+3)* was cloned as a 3.2 kb SphI fragment into the same site of pSET151 to afford pBS3041. To inactivate *orf(+2)*, an internal 84 bp NcoI-KpnI fragment was replaced with a 1.5 kb BamHI-KpnI fragment containing *acc(3)IV*, and the mutated *orf(+2)* was transferred as a 3.4 kb HindIII-XbaI fragment into the same sites of SET151 to furnish pBS3042. To inactivate *orf(+1)*, a 1.5 kb NcoI fragment containing *acc(3)IV* was inserted into the internal NcoI site, and the mutated *orf(+1)* was cloned as a 3.8 kb PstI-EcoRI fragment into the same sites of pSET151 to give pBS3043. To inactivate *lnmZ′*, an internal 84 bp NotI-BamHI fragment was replaced with a 1.5 kb NotI-BamHI fragment containing *acc(3)IV*, and the mutated *lnmZ′* was transferred as a 4.2 kb SphI-EcoRI fragment into the same sites of pSET151 to form pBS3044.

To inactivate *lnmQ*, a 1.5 kb KpnI-XbaI fragment containg *aac(3)IV* was inserted into the internal KpnI-XbaI sites, and the mutated *lnmQ* was cloned as a 4.8 kb BamHI-HindIII fragment into the same sites of pSET151 to yield pBS3045.

Constructs pBS3034 through pBS3045 were each introduced into *S. atroolivaceous* S-140 by conjugation. The desired double-crossover homologous recombinant events were selected for using the apramycin-resistant and thiostrepton-sensitive phenotype, leading to the isolation of mutant strains SB3007 [Δ*orf(−13)*], SB3008 [Δ*orf(−11)*], SB3009 [Δ*orf(−2)*], SB3010 [Δ*orf(−1)*], SB3011 (Δ*lnmA*), SB3012 [Δ*orf(+6)*], SB3013 [Δ*orf(+4)*], SB3014 [Δ*orf(+3)*], SB3015 [Δ*orf(+2)*], SB3016 [Δ*orf(+1)*], SB3017 (Δ*lnmZ′*), and SB3018 (Δ*lnmQ*), respectively. The genotypes of these mutants were confirmed by Southern analysis.

Complementation of the Δ*lnmQ* mutant

To construct the *lnmQ* expression plasmid, a 450 bp EcoRI-BamHI fragment harboring the *ErmE** promoter and a 2.7 kb BamHI-SphI

fragment containing the intact *lnmQ* were cloned into the EcoRI-SphI sites of pBS3031 to yield pBS3047. The expression of *lnmQ* in pBS3047 is under the control of the *ErmE** promoter. pBS3047 was introduced into the Δ*lnmQ* mutant strain of *S. atroolivaceus* SB3018 by conjugation. Selection for apramycin and thiostrepton resistance resulted in the isolation of the recombinant strain SB3020 that harbors pBS3047. The *S. atroolivaceus* SB3020 strain was cultured and analyzed by HPLC and mass spectrometry for LNM production with the S-140 wild-type strain as a control.

Acknowledgments

We thank Kyowa Hakko Kogyo Co. Ltd., Tokyo, Japan, for an authentic sample of leinamycin, the wild-type *S. atroolivaceus* S-140 strain, and assistance in sequencing the *lnm* gene cluster. This work is supported in part by University of California BioSTAR Program Grant Bio99-10045 and Kosan Biosciences, Inc. (Hayward, CA). B.S. is a recipient of National Science Foundation CAREER Award MCB9733938 and National Institutes of Health Independent Scientist Award AI51689.

Received: August 5, 2003
Revised: October 14, 2003
Accepted: October 16, 2003
Published: January 23, 2004

References

1. Hara, M., Asano, K., Kawamoto, I., Takiguchi, T., Katsumata, S., Takahashi, K., and Nakano, H. (1989). Leinamycin, a new antitumor antibiotic from *Streptomyces*: producing organism, fermentation and isolation. J. Antibiot. (Tokyo) *42*, 1768–1774.
2. Hara, M., Takahashi, I., Yoshida, M., Asano, K., Kawamoto, I., Morimoto, M., and Nakano, H. (1989). DC *107*, a novel antitumor antibiotic produced by a *Streptomyces sp*. J. Antibiot. (Tokyo) *42*, 333–335.
3. Nakano, H., and Tamaoki, T. (1992). Mechanism based screens for natural products leads as sources for antitumor drugs. In Harnessing Biotechnology for the 21st Century, Proc. 9th Int. Biotechnol. Symp. Expo., M.R. Ladisch and A. Bose, eds. (Washington, D.C.: American Chemical Society), pp. 72–75.
4. Hara, M., Saitoh, Y., and Nakano, H. (1990). DNA strand scission by the novel antitumor antibiotic leinamycin. Biochemistry *29*, 5676–5681.
5. Asai, A., Hara, M., Kakita, S., Kanda, Y., Yoshida, M., Saito, H., and Saitoh, Y. (1996). Thiol-mediated DNA alkylation by the novel antitumor antibiotic leinamycin. J. Am. Chem. Soc. *118*, 6802–6803.
6. Gates, K.S. (2000). Mechanisms of DNA damage by leinamycin. Chem. Res. Toxicol. *13*, 953–956.
7. Breydo, L., Zang, H., Mitra, K., and Gates, K.S. (2001). Thiol-independent DNA alkylation by leinamycin. J. Am. Chem. Soc. *123*, 2060–2061.
8. Breydo, L., and Gates, K.S. (2002). Activation of leinamycin by thiols: a theoretical study. J. Org. Chem. *67*, 9054–9060.
9. Mitra, K., Kim, W., Daniels, J.S., and Gates, K.S. (1997). Oxidative DNA cleavage by the antitumor antibiotic leinamycin and simple 1,2-dithiolan-3-one 1-oxides: Evidence for thiol-depedent conversion of molecular oxygen to DNA-cleaving radicals mediated by polysulfides. J. Am. Chem. Soc. *119*, 11691–11692.
10. Chatterji, T., Kizil, M., Keerthi, K., Chowdhury, G., Pospisil, T., and Gates, K.S. (2003). Small molecules that mimic the thiol-triggered alkylating properties seen in the natural product leinamycin. J. Am. Chem. Soc. *125*, 4996–4997.
11. Kanda, Y., Ashizawa, T., Kakita, S., Saito, H., Gomi, K., and Okabe, M. (1998). Synthesis and antitumor activity of leinamycin derivatives: Modifications of C-8 hydroxy and C-9 keto groups. Bioorg. Med. Chem. Lett. *8*, 909–912.
12. Kanda, Y., Ashizawa, T., Kakita, S., Takahashi, Y., Kono, M., Yoshida, M., Saitoh, Y., and Okabe, M. (1999). Synthesis and antitumor activity of novel thioester derivatives of leinamycin. J. Med. Chem. *42*, 1330–1332.
13. Fukuyama, T., and Kanda, Y. (1994). Total synthesis of (+)-leinamycin. J. Synth. Org. Chem. Jpn. *52*, 888–899.
14. Kanda, Y., and Fukuyama, T. (1993). Total synthesis of (+)-leinamycin. J. Am. Chem. Soc. *115*, 8451–8452.
15. Kanda, Y., Ashizawa, T., Kawashima, K., Ikeda, S.-I., and Tamaoki, T. (2003). Synthesis and antitumor activity of novel C-8 ester derivatives of leinamycin. Bioorg. Med. Chem. Lett. *13*, 455–458.
16. Behroozi, S.I., Kim, W., and Gates, K.S. (1995). Reaction of N-propanethiol with 3H–1,2-benzodithiol-3-one 1-oxide and 5,5-dimrthyl-1,2-dithiolane-3-one 1-oxide studies related to the reaction of antitumor antibiotic leinamycin with DNA. J. Org. Chem. *60*, 3964–3966.
17. Behroozi, S.I., Kim, W., Dannaldson, J., and Gates, K.S. (1996). 1,2-Dithiolane-3-one 1-oxides-a class of thiol-activated DNA-cleaving agents that are structurally related to the natural product leinamycin. Biochemistry *35*, 1768–1774.
18. Lee, A.H.F., Chan, A.S.C., and Li, T. (2003). Synthesis of 5-(7-hydroxyhept-3-enyl)-1,2-dithiolan-3-one 1-oxide, a core functionality of antibiotic leinamycin. Tetrahedron *59*, 833–839.
19. Strohl, W.R. (2001). Biochemical engineering of natural product biosynthesis pathways. Metab. Eng. *3*, 4–14.
20. Du, L., Sánchez, C., and Shen, B. (2001). Hybrid peptide-polyketide natural products: biosynthesis and prospects toward engineering novel molecules. Metab. Eng. *3*, 78–95.
21. Du, L., and Shen, B. (2001). Biosynthesis of hybrid peptide-polyketide natural products. Curr. Opin. Drug Discov. Dev. *4*, 215–228.
22. Walsh, C.T. (2002). Combinatorial biosynthesis of antibiotics: challenges and opportunities. Chembiochem *3*, 124–134.
23. Cheng, Y.-Q., Tang, G.-L., and Shen, B. (2002). Identification and localization of the antitumor macrolactam leinamycin biosynthesis gene cluster from *Streptomyces atroolivaceus* S-140. J. Bacteriol. *184*, 7013–7024.
24. Cheng, Y.Q., Tang, G.L., and Shen, B. (2003). Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis. Proc. Natl. Acad. Sci. USA *100*, 3149–3154.
25. Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F., and Hopwood, D.A. (2000). Practical *Streptomyces* Genetics (Norwich, United Kingdom: The John Innes Foundation).
26. Conti, E., Stachelhaus, T., Marahiel, M.A., and Brick, P. (1997). Structural basis for the activation of phenylalanine in the nonribosomal biosynthesis of gramicidin. EMBO J. *16*, 4147–4183.
27. Weber, T., Baumgartner, R., Renner, C., Marahiel, M.A., and Holak, T.A. (2000). Solution structure of PCP, a prototype for the peptidyl carrier domains of modular peptide synthetases. Structure Fold. Des. *8*, 407–418.
28. Walsh, C.T., Gehring, A.M., Weinreb, P.H., Quadri, L.E.N., and Flugel, R.S. (1997). Post-translational modification of polyketide and nonribosomal peptide synthases. Curr. Opin. Chem. Biol. *1*, 309–315.
29. Mootz, H.D., Schwarzer, D., and Marahiel, M.A. (2002). Ways of assembling complex natural products on modular nonribosomal peptide synthetases. Chembiochem *3*, 490–504.
30. Staunton, J., and Weissman, K.J. (2001). Polyketide biosynthesis: a millennium review. Nat. Prod. Rep. *18*, 380–416.
31. Olsen, J.G., Kadziola, A., von Wettetein-Knowles, P., Siggaard-Andersen, M., Lindquist, Y., and Larsen, S. (1999). The X-ray crystal structure of β-ketoacy [acyl carrier protein] synthase I. FEBS Lett. *460*, 46–52.
32. He, M., Varoglu, M., and Sherman, D.H. (2000). Structural modeling and site-directed mutagenesis of the actinorhodin β-ketoacyl-acyl carrier protein synthase. J. Bacteriol. *182*, 2619–2623.
33. Kwon, H.-J., Smith, W.C., Scharon, A.J., Hwang, S.H., Kurth, M.J., and Shen, B. (2002). C—O bond formation by polyketide synthases. Science *297*, 1327–1330.
34. Reid, R., Piagentini, M., Rodriguez, E., Ashley, G., Viswanathan, N., Carney, J., Santi, D.V., Hutchinson, C.R., and McDaniel, R.M. (2003). A model of structure and catalysis for ketoreductase domains in modular polyketide synthases. Biochemistry *42*, 72–79.
35. Aparicio, J.F., Molnar, I., Schwecke, T., Konig, A., Haydock, S.F., Khaw, L.E., Staunton, J., and Leadlay, P.F. (1996). Organi-

zation of the biosynthetic gene cluster for rapamycin in *Streptomyces hygroscopicus*: analysis of the enzymatic domains in the modular polyketide synthase. Gene *169*, 9–16.
36. Walsh, C.T., Chen, H., Keating, T.K., Hubbard, B.K., Losey, H.C., Luo, L., Marshall, C.G., Miller, D.A., and Patel, H.M. (2001). Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines. Curr. Opin. Chem. Biol. *5*, 525–534.
37. Kagan, R.N., and Clarke, S. (1994). Widespread occurrence of three sequence motifs in diverse S-adenosylmethionine-dependent methyltransferases suggests a common structure for these enzymes. Arch. Biochem. Biophys. *310*, 417–427.
38. Tsai, S.-C., Miercke, L.J., Krucinski, J., Gokhale, R., Chen, J.C., Foster, P.G., Cane, D.E., Khosla, C., and Stroud, R.M. (2001). Crystal structure of the macrocycle-forming thioesterase domain of the erythromycin polyketide synthase: Versatility from a unique substrate channel. Proc. Natl. Acad. Sci. USA *98*, 14808–14813.
39. Piel, J. (2002). A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of *Paederus* beetles. Proc. Natl. Acad. Sci. USA *99*, 14002–14007.
40. El-Sayed, K.A., Hothersall, J., Cooper, S.M., Stephens, E., Simpson, T.J., and Thomas, C.M. (2003). Characterization of the mupirocin biosynthesis gene cluster from *Pseudomonas fluorescens* NCIMB 10586. Chem. Biol. *10*, 419–430.
41. Reeves, C.D., Murli, S., Ashley, G.W., Piagentini, M., Hutchinson, C.R., and McDaniel, R. (2001). Alteration of the substrate specificity of a modular polyketide synthase acyltransferase domain through site-directed mutations. Biochemistry *40*, 15464–15470.
42. Poulos, T.L. (1995). Cytochrome P450. Curr. Opin. Struct. Biol. *5*, 767–774.
43. Heathcote, M.L., Staunton, J., and Leadlay, P.F. (2001). Role of type II thioesterases: evidence for removal of short acyl chains produced by aberrant decarboxylation of chain extender units. Chem. Biol. *8*, 207–220.
44. Kim, B.S., Cropp, T.A., Beck, B.J., Sherman, D.H., and Reynolds, K.A. (2002). Biochemical evidence for an editing role of thioesterase II in the biosynthesis of the polyketide pikromycin. J. Biol. Chem. *277*, 48028–48034.
45. Schwarzer, D., Mootz, H.D., Linne, U., and Marahiel, M.A. (2002). Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases. Proc. Natl. Acad. Sci. USA *99*, 14083–14088.
46. Locher, K.P., Lee, A.T., and Rees, D.C. (2002). The *E. coli* BtuCD structure: a framework for ABC transporter architecture and mechanism. Science *296*, 1091–1098.
47. Cane, D.E., and Walsh, C.T. (1999). The parallel and convergent universes of polyketide synthases and non-ribosomal peptide synthetases. Chem. Biol. *6*, R319–R325.
48. Kingston, D.G.I., Kolpak, M.X., LeFevere, J.W., and Borup-Grochtmann, I. (1983). Biosynthesis of antibiotics of the virginiamycin family. 3. biosynthesis of virginiamycin M_1. J. Am. Chem. Soc. *105*, 5106–5110.
49. Thomas, M.G., Burkart, M.D., and Walsh, C.T. (2002). Conversion of L-proline to pyrrolyl-2-carboxyl-*S*-PCP during undecylprodigiosin and pyoluteorin biosynthesis. Chem. Biol. *9*, 171–184.
50. Yu, T.-W., Bai, L., Clade, D., Hoffmann, D., Toelzer, S., Trinh, K.Q., Xu, J., Moss, S.J., Leistner, E., and Floss, H.G. (2002). The biosynthetic gene cluster of the maytansinoid antitumor agent ansamitocin from *Actinosynnema pretiosum*. Proc. Natl. Acad. Sci. USA *99*, 7968–7973.
51. Stachelhaus, T., Mootz, H.D., and Marahiel, M.A. (1999). The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. Chem. Biol. *6*, 493–505.
52. Challis, G.L., Ravel, J., and Townsend, C.A. (2000). Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. Chem. Biol. *7*, 211–224.
53. Schmidt, B., Riesner, D., Lawen, A., and Kleinkauf, H. (1992). Cyclosporin synthetase is a 1.4 MDa multienzyme polypeptide. Re-evaluation of the molecular mass of various peptide synthetases. FEBS Lett. *307*, 355–360.
54. Scott-Craig, J.S., Panaccione, D.G., Pocard, J.-A., and Walton, J.D. (1992). The cyclic peptide synthetase catalyzing HC-toxin production in the filamentous fungus *Cochliobolus carbonum* is encoded by a 15.7-kilobase open reading frame. J. Biol. Chem. *267*, 26044–26049.
55. Marshall, C.G., Hillson, N.J., and Walsh, C.T. (2002). Catalytic mapping of the vibriobactin biosynthetic enzyme VibF. Biochemistry *41*, 244–250.
56. Shen, B. (2003). Polyketide biosynthesis beyond the type I, II, and III polyketide synthase paradigms. Curr. Opin. Chem. Biol. *7*, 285–295.
57. Bierman, M., Logan, R., O'Brien, K., Seno, E.T., Nagaraja Rao, R., and Schoner, B.E. (1992). Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces* spp. Gene *116*, 43–49.
58. Devereux, J., Haeberli, P., and Smithies, O. (1984). A comprehensive set of sequence analysis programs for VAX. Nucleic Acids Res. *12*, 387–395.
59. Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res. *22*, 4673–4680.
60. Asai, A., Saito, H., and Saitoh, Y. (1997). Thiol-independent DNA cleavage by a leinamycin degradation product. Bioorg. Med. Chem. *5*, 723–729.

Accession Numbers

The sequence reported in this paper has been deposited in GenBank under accession number AF484556.