Chemistry & Biology
Article

Elucidating the Biosynthetic Pathway for the Polyketide-Nonribosomal Peptide Collismycin A: Mechanism for Formation of the 2,2′-bipyridyl Ring

Ignacio Garcia,[1] Natalia M. Vior,[1] Alfredo F. Braña,[1] Javier González-Sabin,[2] Jürgen Rohr,[3] Francisco Moris,[2] Carmen Méndez,[1] and José A. Salas[1,*]
[1]Departamento de Biología Funcional e Instituto Universitario de Oncología del Principado de Asturias (I.U.O.P.A), Universidad de Oviedo, 33006 Oviedo, Spain
[2]EntreChem S.L, Edificio Científico Tecnológico, Campus El Cristo, 33006 Oviedo, Spain
[3]Department of Pharmaceutical Sciences, College of Pharmacy, University of Kentucky, 789 South Limestone Street, Lexington, KY 40536-0596, USA
*Correspondence: jasalas@uniovi.es
DOI 10.1016/j.chembiol.2012.01.014

SUMMARY

The gene cluster for the bipyridyl compound collismycin was characterized from *Streptomyces* sp. CS40. Sequence analysis of a 46.7 kb DNA region revealed 27 open reading frames, 23 of which are involved in collismycin biosynthesis. Eight insertional inactivation mutants were generated in the sequenced region to prove its involvement in collismycin biosynthesis, define the boundaries of the cluster, functionally characterize some genes, and isolate two biosynthetic intermediates. A model for collismycin biosynthesis—which includes the conversion of lysine into picolinic acid, participation of a polyketide synthase-non-ribosomal peptide synthetase system, and some further modifications—is proposed. The biosynthetic pathway would include an unusual NRPS-mediated incorporation of a cysteine residue, possibly through a Michael addition and followed by the extension of the peptide chain by leucine incorporation and later removal by amidohydrolase.

INTRODUCTION

Among natural products synthesized by microorganisms, three main structural families constitute the great majority of bioactive compounds, both based on structural diversity and also in applicability: polyketides (PK), nonribosomal peptides (NRP), and hybrid PK-NRP (Newman and Cragg, 2007). Polyketides are synthesized by a sequential set of reactions catalyzed by polyketide synthases (PKS). The carbon skeleton of these compounds is biosynthesized by stepwise decarboxylative Claisen-type condensation of acyl-CoA precursors that are further modified. Three types of PKS are known to date: type I, multifunctional enzymes organized into modules, each harboring a set of distinct activities responsible for the catalysis of one cycle of polyketide chain elongation; type II, multienzyme complexes that carry out a single set of iteratively acting activities; and finally, type III, iteratively acting condensing enzymes (Wenzel and Müller, 2005; Gokhale et al., 2007; Van Lanen and Shen, 2008; Khosla, 2009). Nonribosomal peptides are small peptide molecules synthesized by nonribosomal peptide synthetases (NRPS) using proteinogenic and nonproteinogenic amino acids as monomeric building blocks, including D-amino acids and other carboxylic acids. The amino acid monomers incorporated by NRPS assembly lines are aminoacyl-AMP mixed anhydrides that follow the same chemical logic as type I PKS for chain elongation and then are modified based on the program encoded by different domains present in NRPS modules, which can include epimerization, methyltransferase, reductase, or oxidase activities (Walsh, 2004; Sieber and Marahiel, 2005; Fischbach and Walsh, 2006; Koglin and Walsh, 2009; Strieker et al., 2010). Finally, the hybrid PK-NRP compounds contain both polyketide and peptide moieties. In these compounds, usually the PK moiety is synthesized by a type I PKS (modular PKS); therefore, in the biosynthesis of PK-NRP compounds, multifunctional enzymes with an assembly line-like structure are responsible for the generation of the mixed PK-NRP chain (Du and Shen, 2001).

Within natural products synthesized by microorganisms, the bipyridyl family of compounds has not been yet studied in detail. Representatives of this family are SF2738A-F (Gomi et al., 1994), pyrisulfoxins (Tsuge et al., 1999), collismycins (Shindo et al., 1994), and caerulomycins (Funk and Divekar, 1959; McInnes et al., 1977, 1978). All these compounds contain a 2,2′-bipyridyl ring system, which is further modified with some postbipyridyl tailoring modifications (Figure 1A). Early biosynthetic studies using labeled precursors suggested that lysine-derived picolinic acid was a biosynthesis precursor (McInnes et al., 1979; Vining et al., 1988). Members of this family (including collismycin A) have been shown to exhibit antibacterial, antifungal, and cytotoxic activities (Gomi et al., 1994; Stadler et al., 2001), and they also show potentiality as anti-inflammatory agents through the binding to the dexamethasone-glucocorticoid receptor (Shindo et al., 1994) or as neuroprotectant agents by reducing oxidative stress in neurons (Martinez Gil et al., 2007).

Here we report the isolation of the gene cluster for collismycin A biosynthesis from *Streptomyces* sp. CS40, a strain isolated from the leaf-cutting ant *Acromyrmex octospinosus*. Insertional inactivation experiments provided evidence for the involvement

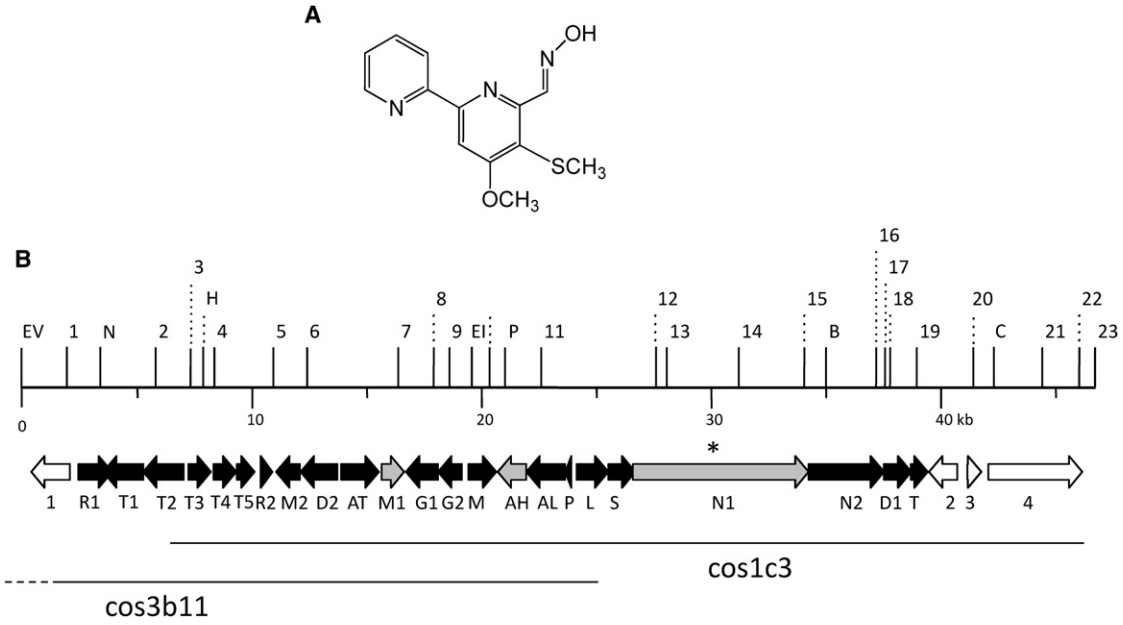

Figure 1. The Collismycin Gene Cluster

(A) Chemical structure of collismycin A and (B) genetic organization of the collismycin gene cluster in *Streptomyces* spp. CS40. White arrows indicate genes that are not involved in collismycin biosynthesis. Black and gray arrows indicate genes involved in collismycin biosynthesis. Gray arrows indicate genes inactivated in CLMM1, CLMAH, and CLM12 mutants. cos3b11 and cos1c3 represent cosmids containing the sequenced region. Asterisk indicates the BamHI fragment used for generating mutant CLM12. BamHI sites are numbered. EV, EcoRV; N, NcoI; EI, EcoRI; P, PvuI; C, ClaI, restriction sites used for cloning purposes.

of the identified genes in collismycin A biosynthesis and established the boundaries of the gene cluster. We also show experimental evidence of an interesting system for the formation of the second pyridine ring involving the incorporation and further removal of a leucine residue.

RESULTS

Isolation and Characterization of the Collismycin Gene Cluster

Work in our laboratory on the search for novel bioactive compounds prompted us to look for actinomycetes from the cuticle of leaf-cutting ants isolated from four different nests in Lambayeque, Perú. From a total of 197 strains isolated and identified as members of the actinomycetes family, we selected several strains producing interesting bioactive compounds according to their bioactive profile (unpublished data). Four of these strains were found to produce a small molecule that was identified by nuclear magnetic resonance (NMR) and mass spectrometry (MS) as collismycin A. We selected one of these strains, named *Streptomyces* sp. CS40, for the isolation and characterization of the gene cluster. Based on previous reports suggesting lysine-derived picolinic acid as a biosynthetic precursor, the possible participation of a lysine aminotransferase (LAT) in collismycin biosynthesis in this strain was evaluated. Two different LATs have been described to participate in the biosynthesis of secondary metabolites in actinomycetes. On one hand, lysine 2-aminotransferases (LAT2) involved in the biosynthesis of virginiamycin S (Namwat et al., 2002) and nikkomycin D (Bruntner and Bormann, 1998) and, on the other hand, a lysine 6-aminotransferase (LAT6) involved in the biosynthesis of cephamycin C (Coque et al., 1991). Degenerated oligonucleotides based on amino acid sequences of LAT2 and LAT6 genes were designed and used in polymerase chain reaction (PCR) amplification reactions using chromosomal DNA from *Streptomyces* sp. CS40 as template. A 565 bp PCR product was obtained with the LAT2 primers, but no amplification product was found with the LAT6 primers. Sequencing of the PCR product revealed clear similarities of the partially encoded protein with LAT2 enzymes, such as NikC from the nikkomycin pathway (53% identity in the sequenced region) and VisA from the virginiamycin pathway (47% identity in the sequenced region).

To isolate the collismycin gene cluster, we screened a library of *Streptomyces* spp. CS40 genomic DNA in cosmid pWE15 by in situ colony hybridization using the 565 bp PCR product as a probe. Three positive clones were isolated and verified by Southern hybridization. One of these cosmids, cos1C3, was submitted to partial sequencing of both ends of different BamHI fragments. Analysis of the deduced products of the different *orfs* indicated that they would code for enzymatic functions required for collismycin biosynthesis, including PKS and NRPS genes (data not shown). One of these BamHI fragments (sites 13–14, Figure 1B) was cloned into the suicide vector pOJ260 generating pOJB12, and this plasmid was used for gene disruption experiments. The resultant mutant CLM12 did not produce collismycin A when compared with the wild-type strain (Figure 2), indicating that the gene sequence within this BamHI fragment is involved in collismycin biosynthesis.

Cos1C3 and a 7,752 bp EcoRV-HindIII fragment overlapping the left-hand side of this cosmid were sequenced obtaining information of the complete gene cluster. The nucleotide sequence of

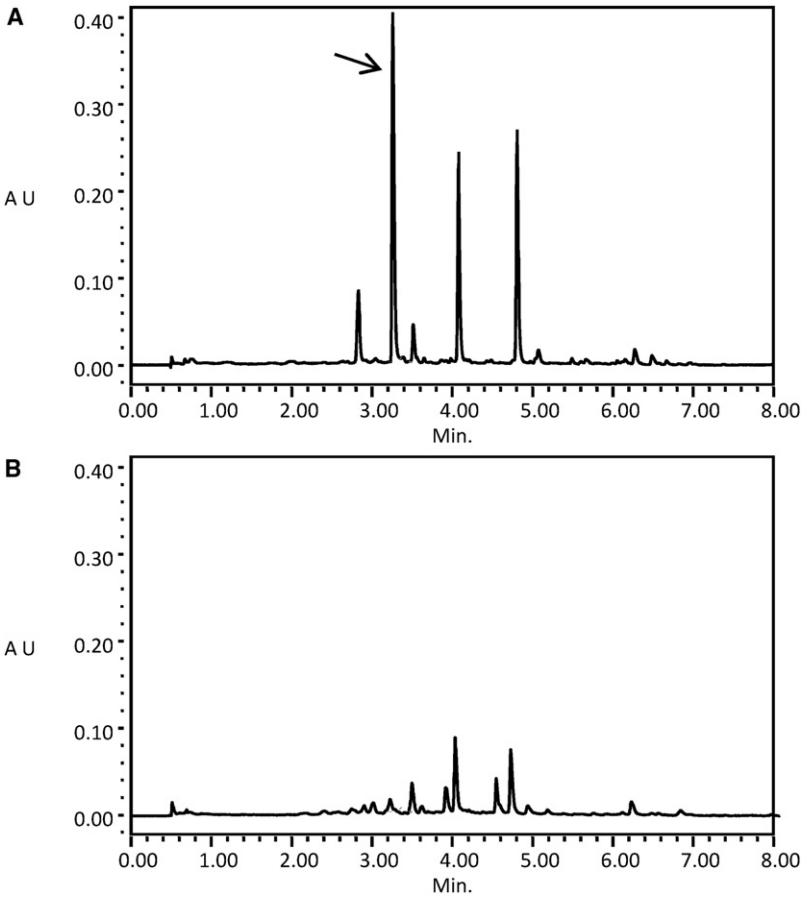

Figure 2. UPLC Analysis of Ethyl Acetate Extracts of *Streptomyces* spp. CS40and Mutant CLM12

(A) *Streptomyces* spp. CS40 and (B) mutant CLM12. The arrow indicates the UPLC mobility of collismycin A.

a 46,672 bp region was determined (EMBL or GenBank accession number HE575208). Sequence analysis and comparisons with databases revealed the presence of 27 *orfs* (Figure 1B; Table 1), all of them showing high GC content and GC bias at the third codon position that is characteristic of *Streptomyces* genes. Analysis of the deduced products of the different *orfs* showed the similarities shown in Table 1. Two TTA codons were found (one in *clmT2* and another in *clmR2*), implying an upper level of regulation during collismycin biosynthesis by some homolog to the pleiotropic regulatory gene *bldA* encoding a tRNA-Leu (Chater and Chandra, 2008).

Delimiting the Boundaries of the Collismycin Cluster

It was initially assumed that most of the 27 *orfs* could be involved in collismycin biosynthesis. This was based on the deduced functions of the gene products and on tentative roles for their products in collismycin biosynthesis. In order to determine the boundaries of the cluster, several *orfs* at both sides of the sequenced region (*orf1*, *clmR1*, *orf2*, *orf3*, and *orf4*) were knocked out by insertional inactivation. On the left-hand side, mutants in which either a gene coding for an helicase (*orf1*) or a gene coding for a TetR family regulatory protein (*clmR1*) were inactivated at the same level of the wild-type strain. From this result, it can be deduced that *clmR1* is not involved in collismycin biosynthesis. However, the *clmR1* mutant started collismycin biosynthesis earlier than did the wild-type strain, suggesting a role for this gene in collismycin biosynthesis. On the right-hand side, three genes (*orf2*, *orf3*, and *orf4*) were initially assumed to participate in collismycin biosynthesis based on their similarities to P450 monoxygenase (*orf2*), ArsR regulatory (*orf3*), and NRPS (*orf4*) genes. However, individual inactivation of these three genes did not affect collismycin biosynthesis (data not shown). According to these results we considered at the moment that all these four genes were probably not involved in collismycin biosynthesis. Moreover, in the course of these studies, we became aware of the existence of a highly similar cluster in a region of the recently sequenced genome of *Streptomyces roseosporus* NRRL15998 (Broad Institute accession number ABYB01000000). This region, which has not been annotated as a putative collismycin cluster, showed a similar genetic organization to that of the collismycin cluster with similarities at the amino acid level between gene products as high as 96%–99%. Because of this high similarity, we assumed that the *S. roseosporus* cluster was also involved in collismycin biosynthesis and therefore similar activities should be present in both clusters. *Streptomyces* sp. CS40 *orf1*, *orf2*, *orf3*, and *orf4* were not present in *S. roseosporus*; however, both strains shared *clmR1* homologs as the last gene on the left-hand side of the clusters and *clmT* homologs at the right-hand side of the clusters. Based on this information, we propose *clmR1* and *clmT* as the boundaries of the collismycin cluster, thus delimiting the *clm* cluster to a region of 36.5 kb.

Genes Involved in Picolinic Acid Biosynthesis and Its Activation to Picolinyl-CoA

Three genes in the cluster, *clmL*, *clmS*, *and clmAL* are probably involved in the biosynthesis of picolinic acid from L-lysine, its activation to picolinyl-CoA, and its loading to the PKS-NRPS hybrid protein ClmN1. The ClmL protein strongly resembles NikC of the nikkomycin pathway in *S. tendae* (Bruntner and Bormann, 1998) and VisA of the virginiamycin S pathway in *S. virginiae* (Namwat et al., 2002). These two proteins have been shown to code for a LAT2, converting L-lysine into 1-piperidine-2-carboxylic acid. The ClmS protein showed clear similarities to monomeric sarcosine oxidases, such as NikD of the nikkomycin pathway in *S. tendae* (Bruntner et al., 1999). Sarcosine oxidases catalyze the oxidation of sarcosine to formaldehyde, glycine, and H_2O_2 using FAD. ClmS exhibits an ADP-βαβ-binding fold fingerprint motif that is important for the interaction with FAD (Nishiya and Imanaka, 1996), and it also

Table 1. Proposed Functions for Collismycin Biosynthesis Genes

Gene	aa	Proposed Function in Collismycin Biosynthesis	Similar Protein, Origin, (% Identity / Similarity), Accession Number
orf1	563		DEAD/DEAH box helicase domain-containing protein, *Streptomyces viridochromogenes* DSM 40736, (85/90), ZP_07307194
clmR1	186	TetR family transcriptional regulation	TetR family transcriptional regulator *Streptomyces griseus* subsp. *griseus* NBRC 13350 (86/95), YP_001827902.1
clmT1	578	Transmembrane ABC family collismycin transport	Putative ABC transporter ATPase and permease component *Streptomyces griseus* subsp. *griseus* NBRC 13350, (90/94), YP_001827903.1
clmT2	587	Transmembrane ABC family collismycin transport	Putative ABC transporter ATPase and permease component *Streptomyces griseus* subsp. *griseus* NBRC 13350, (91/95), YP_001827904.1
clmT3	323	Transmembrane ABC family collismycin transport	Putative ABC transporter solute-binding protein *Streptomyces griseus* subsp. *griseus* NBRC 13350, (79/88), YP_001827905.1
clmT4	328	Transmembrane ABC family collismycin transport	Iron ABC transporter *Streptomyces griseus* subsp. *griseus* NBRC 13350, (86,92), YP_001827906.1
clmT5	262	Transmembrane ABC family collismycin transport	Putative ABC transporter ATP-binding protein *Streptomyces griseus* subsp. *griseus* NBRC 13350, (87/90), YP_001827907.1
clmR2	169	LuxR family transcriptional regulation	Two component transcriptional regulator, LuxR family *Thermaerobacter marianensis* DSM 12885, (56/76), YP_004102054.1
clmM2	343	Methylation	AzicL, *Kibdelosporangium* sp. MJ126-NF4, (44/59), ADB02856.1
clmD2	533	Dehydrogenation	Putative dehydrogenase and domain with unknown function *Streptomyces albaduncus* (56/69) CBH32104.1. TamL, Streptomyces sp. 307-9, (52/65), ADC79636
clmAT	545	Oxime generation	Diaminobutyrate-pyruvate aminotransferase, *Streptomyces coelicolor* A3(2), (55/70), NP_630377.1
clmM1	329	Methylation	OxyF, *Streptomyces rimosus*, (52/70), AAZ78330
clmG1	463	Dehydrogenation	GriD, *Streptomyces griseus*, (36/46), BAF36646
clmG2	344	Dehydrogenation	GriC, *Streptomyces griseus*, (42/58) BAF36645
clmM	403	Reduction	FAD-dependent oxidoreductase *Ktedonobacter racemifer* DSM 44963 (47/63) ZP_04696966.1
clmAH	414	Amidohydrolase	Amidohydrolase *Salinispora arenicola CNS-205 (53/65) YP_001536995.1*
clmAL	563	Picolinic acid activation	*NikE, Streptomyces tendae, (52/61), CAB46535*
clmP	87	PKS/NRPS loading	Acyl carrier protein *Oscillatoria* sp. PCC 6506 (39/60) ZP_07114031.1
clmL	459	Picolinic acid biosynthesis	NikC, *Streptomyces tendae*, (57/71), CAA75797
clmS	394	Picolinic acid biosynthesis	NikD, *Streptomyes tendae*, (59/69), CAB46534
clmN1	2553	PKS-NRPS	CurF, *Lyngbya majuscula* 3L, (43/59), AAT70101
clmN2	1092	NRPS	Nonribosomal peptide synthetase, *Microcystis aeruginosa* K-139, (33/53), BAH22763
clmD1	395	Dehydrogenation	Acyl-CoA dehydrogenase *Acidobacterium capsulatum* ATCC 51196, (47/62), YP_002755437.1
clmT	243	NRPS intermediate release	Putative type II thioesterase *Cyanothece* sp. ATCC 51142 (43/59) YP_001802837.1
orf2	408		Cytochrome P450 *Streptomyces sviceus* ATCC 29083 (78/86) ZP_06921933.1
orf3	202		Conserved hypothetical protein, *Streptomyces hygroscopicus* ATCC 53653, (39/48), ZP_07294271.1
orf4	1363		Nonribosomal peptide synthetase/polyketide synthase *Myxococcus xanthus* DK 1622 (43/57), YP_631961.1

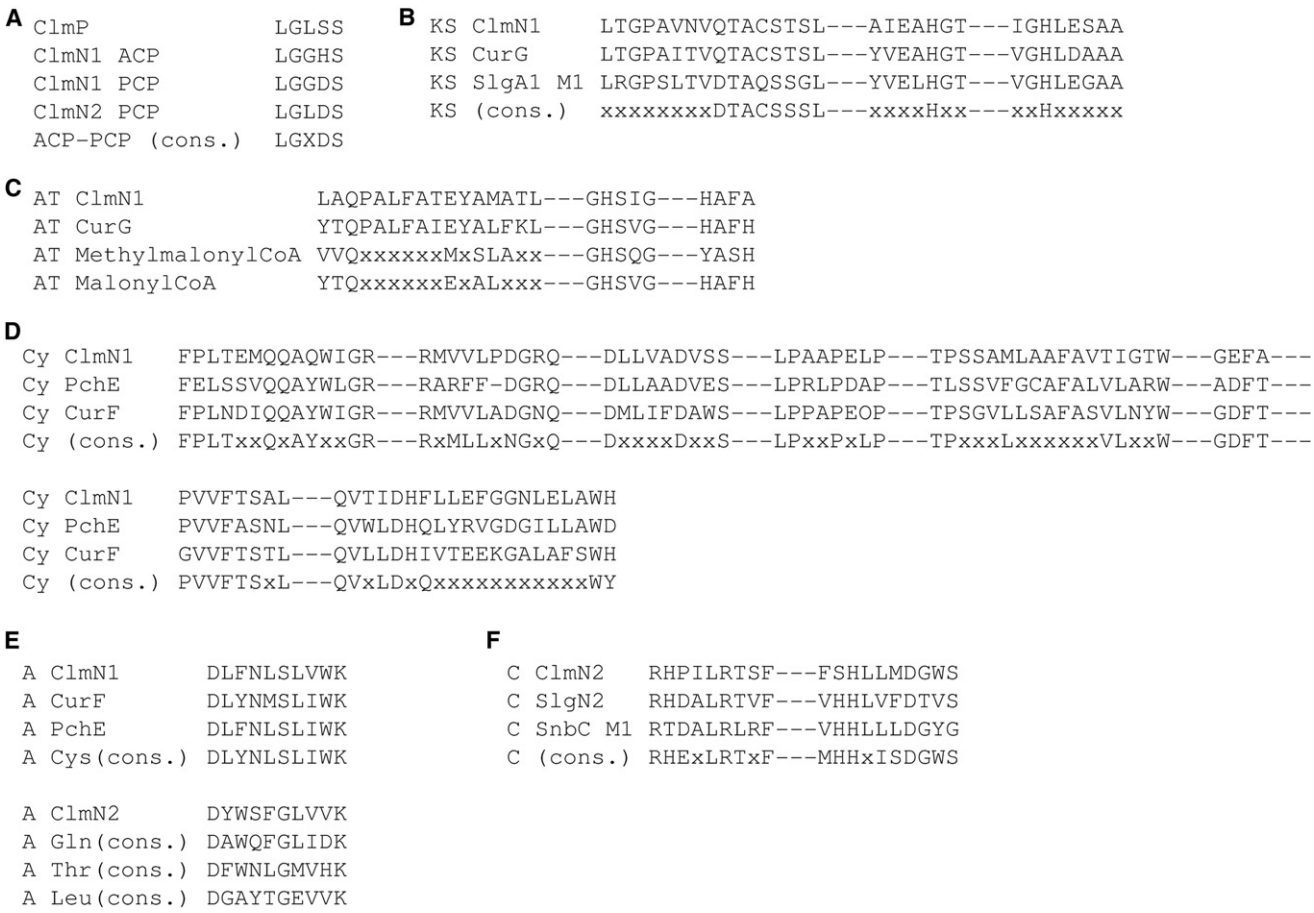

Figure 3. Sequence Alignments between Motifs of ClmN1 and ClmN2 with Conserved Motifs of Several PKS/NRPS Domains

(A) ACP-PCP domains alignment. ACP-PCP (cons.) represents the consensus sequence for acyl and peptidyl carrier proteins.

(B) KS domain alignment. KS (cons.) represents consensus sequence for ketosynthase domain.

(C) AT domains alignment, where AT methylmalonylCoA indicates consensus sequence for propionate incorporation, and AT malonylCoA indicates consensus sequence for acetate incorporation.

(D) Condensation/cyclization domain alignment. Cy (cons.) represents consensus sequences for condensation/cyclization domain.

(E) Adenylation domain alignment. A Cys (cons.), A Gln (cons.), A Thr (cons.), and A (Leu (cons.) represents consensus Stachelhaus sequences for cysteine, glutamine, threonine, and leucine incorporation, respectively.

(F) Condensation domain alignment. C (cons.) represents consensus sequence for condensation domain.

possesses a cysteine residue (Cys323), which is likely to form part of the active site of the enzyme (Nishiya et al., 1995). The ClmAL protein shows the presence of an AMP-binding domain, characteristic of the acetyl-CoA synthetase-like superfamily. Proteins with an AMP-binding domain contain a consensus sequence involved in ATP binding and amino acid adenylation. ClmAL matches with all consensus core sequences described for this domain (de Crécy-Lagard et al., 1997), and the amino acid sequence between positions 213 and 222 (LSGGTTGPPK) contains the signature characteristic of active AMP-binding domains (LSGGxTxxxK). According to this information, the consecutive action of ClmL, ClmS, and ClmAL would convert L-lysine into activated picolinyl-CoA.

Genes Involved in the Formation of the Bipyridyl Ring

Four PKS-NRPS or related genes have been identified in the cluster: *clmP*, *clmN1*, *clmN2*, and *clmT*. The deduced sequence of *clmP* shows high similarity (above 60%) and identity (above 39%) with predicted acyl carrier proteins (ACP) from different organisms. ClmP also contains the core motif (LGLSS) present in ACPs (LGXDS) with the active serine residue (Marahiel et al., 1997; Figure 3A). ACPs act as "swinging arms" for the attachment of activated fatty acids, allowing its elongation and processing by PKS proteins. These domains are usually encoded within the PKS, but sometimes ACPs act as stand-alone domains, like AcmD in actinomycin biosynthesis (Pfennig et al., 1999). We propose ClmP as a free-standing ACP charging picolinic acid from the acyl-CoA forming enzyme ClmAL to ClmN1.

Genes *clmN1* and *clmN2* would encode a hybrid PKS-NRPS and NRPS proteins, respectively. Analysis of the deduced amino acid sequence of ClmN1 reveals the presence of seven domains organized into two modules: a PKS extender module and a NRPS elongation module (Figure 4). The PKS extender module contains ketosynthase (KS), acyltransferase (AT), and ACP

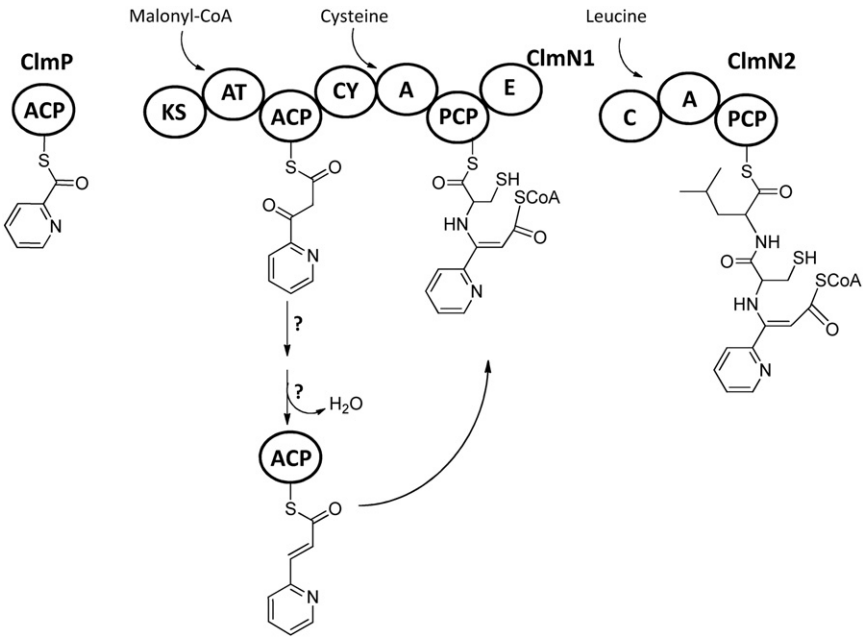

Figure 4. Scheme Showing the Multi-Enzymatic Hybrid PKS-NRPS Complex Formed by ClmP, ClmN1, and ClmN2 and the Hypothetical Reactions that It Catalyzes

The question marks indicate that the enzymes catalyzing the reduction steps following malonyl-CoA incorporation have not been yet identified.

domains, with no loading domain. The KS domain matches the conserved residues present in consensus KS domains (Figure 3B; Aparicio et al., 1996; Chang et al., 2004). The AT domain contains the active site motif (GHSxG; Reeves et al., 2001) and the substrate specificity signature for malonyl-CoA (Figure 3C; Reeves et al., 2001). The NRPS module in ClmN1 contains condensation/cyclization (Cy), adenylation (A), peptidyl carrier protein (PCP), and epimerizarion (E) domains. The ClmN1 NRPS module shows a high similarity with the CurF NRPS module (44% identity and 60% similarity), both having a similar organization, except that the E domain present in ClmN1 was absent in CurF. CurF is a hybrid PKS-NRPS involved in curacin A biosynthesis in the cyanobacterium *Lyngbya majuscule*. The NRPS module in CurF elongates the growing chain by condensing a cysteine residue to generate a thiazoline ring (Chang et al., 2004). A domain contains a ten amino acids signature determining substrate specificity, known as Stachelhaus sequence, and it is possible to predict the possible amino acid loaded by comparing this sequence with other domains (Stachelhaus et al., 1999). Stachelhaus analysis of the A domain of ClmN1 (DLFNLSLVWK) suggests that this module incorporates a cysteine residue (Figure 3E).

ClmN1 NRPS module contains a cyclization domain (Cy). This type of domain is usually involved in the generation of heterocyclic derivatives of cysteine, threonine, or serine and their condensation into the nascent nonribosomal peptide. ClmN1 Cy domain contains a sequence (DLLVADVSS) described as the core signature motif of the Cy domain (DxxxxDxxS; Di Lorenzo et al., 2008). Additionally, the ClmN1 Cy domain sequence contains other specific Cy consensus sequences also present in the CurF Cy domain (Figure 3D; Chang et al., 2004).

Analysis of the ClmN1 NRPS module predicts the presence of an E domain. These domains are usually associated with the modification of the incorporated amino acid from an L-amino acid into a D-amino acid (Marahiel et al., 1997). Although the presence of the E domain in ClmN1 is suggested based on sequence analysis, this domain lacks the active site motif present in functional E domains (HHxxxDxVSW; de Crécy-Lagard et al., 1997). We therefore propose that the ClmN1 E domain is not active or is not acting as a canonical E domain.

The *clmN2* gene would code for a monomodular NRPS. It contains three domains: C, A, and PCP. The C domain contains the highly conserved core motifs C2 (RHExLRTxF) and C3 (HxxxDG) with the active histidine residue (Figure 3F; Marahiel et al., 1997). Analysis of the A domain does not match any of the motifs present in databases and thus makes it difficult to predict the amino acid used as substrate. However, the fact that the C domain contains an active site suggested that a second amino acid should be incorporated by the ClmN2 protein. ClmN2 also contains a putative PCP domain containing the conserved core sequence for ACP/PCP domains (Figure 3A).

The gene *clmT* encodes a type II thioesterase (TE) domain. Stand-alone type II TEs are often encoded within NRPS gene clusters and are proposed to act by removing incorrectly loaded amino acids from PCP domains of NRPS proteins or removing acetyl groups from ACP domains to regenerate free thiol groups (Schwarzer et al., 2002). However, recent studies reveal that stand-alone type II TEs could also act by releasing an intermediate product from ACP for subsequent modifications (Whicher et al., 2011). Since ClmN1 and ClmN2 lack the TE domain, we propose that ClmT could be involved in releasing the growing PK-NRP chain from ClmN2.

Three other genes (*clmM*, *clmD1*, and *clmD2*) in the cluster could participate in the heterocyclization of the second ring and thus form the 2,2′-bipyridyl structure. Blast analysis reveals that ClmM shows high similarity (above 40% identity and 60% similarity) with putative FAD-dependent oxidoreductases and monoxygenases. In addition, ClmM contains an UbiH domain present in 2-polyprenyl-6-methoxyphenol hydroxylases and related FAD-binding oxidoreductases. This suggests that *clmM* could code for a putative monoxygenase. The ClmD1 predicted sequence shows high homology (above 45% identity and 60% similarity) with predicted acyl-CoA dehydrogenases from different organisms. The predicted sequence contains an acyl-CoA dehydrogenase domain. This domain is typically found in FAD-dependent oxidases reducing CoA-bound substrates. The ClmD2 sequence shows high homology with TamL (52% identity and 65% similarity), a protein probably involved in oxidative tailoring modifications during tirandamycin biosynthesis

(Carlson et al., 2010) and with SpnJ (50% identity and 62% similarity), a FAD-dependent secondary alcohol oxidase involved in spinosin biosynthesis (Kim et al., 2007). It shows the presence of a FAD-binding domain (residues 43 to 123), which is involved in a variety of oxidation reactions (Mattevi et al., 1997) and is usually present in enzymes using FAD as cofactor, mainly oxidoreductases. The exact roles of ClmM, ClmD1, and ClmD2 in collismycin biosynthesis are not clear at the moment, although at least two oxidations and one dehydration steps are necessary for heterocyclization to generate the 2,2′-bipyridyl ring. These enzymes could participate in these reactions.

Genes Involved in Further Modifications

Upon formation of the bipyridyl ring, several additional genes would introduce further chemical modifications to generate the final product collismycin A. Two genes in the cluster (*clmM1* and *clmM2*) would code for methyltransferases. Analysis of the amino acid sequence of ClmM1 and ClmM2 shows that both would code for *S*-adenosylmethionine-dependent methyltransferases. However, ClmM1 and ClmM2 are significantly different, only sharing 37% identity. ClmM1 is homolog to OxyF (52% identity and 70% similarity), which is probably involved in *C*-methylation during oxytetracycline biosynthesis (Zhang et al., 2006). On the other hand, the closest homolog to ClmM2 is AzicL (44% identity and 59% similarity), responsible for an *O*-methylation during azicemicins biosynthesis (Ogasawara and Liu, 2009). Based on this information, we propose that ClmM2 could be involved in the generation of the methoxy group present in position 4 in collismycin and ClmM1 could be responsible for catalyzing a methylation event necessary for intermediate rearrangement during earlier steps of the biosynthesis. This role for ClmM1 was confirmed by insertional inactivation (see subsection Unusual Incorporation of a Leucine Residue and Its Further Removal during Biosynthesis).

The deduced product of *clmAH* shows high similarity with a putative amidohydrolase from *Salinispora arenicola* (52% identity and 64% similarity; US DOE Joint Genome Institute, YP_001536995.1). The amidohydrolase family includes metallo-dependent enzymes involved in the hydrolysis of C-N bonds (both peptidic and nonpeptidic) by the nucleophilic attack of a deprotonated molecule of water. ClmAH shares with several members of this family a conserved metal binding site, involving four histidines and one aspartic acid residue, which are not near each other in the amino acid sequence but are very close in the tridimensional structure of the protein (Holm and Sander, 1997). The role of ClmAH in collismycin biosynthesis was initially not very clear. It was difficult to anticipate which amide bond would be hydrolyzed by this protein, but it was later on established by insertional inactivation (see subsection Unusual Incorporation of a Leucine Residue and Its Further Removal during Biosynthesis).

ClmG1 shows homology with NAD(P)-dependent aldehyde dehydrogenases, such as GriD from *S. griseus* subsp. *griseus* (55% identity 67% similarity). Next to *clmG1* and in the same orientation, *clmG2* would code for a protein homologous to different AMP-binding proteins, such as GriC also from *S. griseus* subsp. *griseus* (49% identity and 63% similarity). GriC and GriD have been reported to act together as a carboxylic acid reductase involved in the conversion of 3-amino-4-hydroxybenzoic acid (3,4-AHBA) to 3-amino-4-hydroxybenzaldehyde (3,4-AHBAL) during the biosynthesis of grixazone in *S. griseus*. In that reaction, GriC was proposed to convert 3,4-AHBA to an acyl-AMP intermediate, which is later reduced to 3,4-AHBAL by GriD with the use of NAD(P)H (Suzuki et al., 2007). Their homology to ClmG1 and ClmG2 suggest that these enzymes could catalyze a similar reaction during collismycin biosynthesis, reducing a 2,2′-bipyridine intermediate with a carboxylic group to an aldehyde and thus preparing that position for the introduction of the oxime group. Finally, ClmAT predicted amino acid sequence shows the presence of an acetyl ornithine aminotransferase domain. This domain belongs to aminotransferase class-III family and enzymes containing this domain are involved in transamination or decarboxylation of basic amino acids and their derivatives in pyridoxal phosphate-dependent reactions. ClmAT shows high identity with putative diaminobutyrate-pyruvate aminotransferases (above 50% identity and above 70% similarity) and with BatP (28% identity and 47% similarity), aminotransferase involved in kalimantacin/batumin-related polyketide biosynthesis in *Pseudomonas fluorescens* (Mattheus et al., 2010). We propose ClmAT could participate in the formation of the oxime group in collismycin A biosynthesis.

Regulation and Secretion

Two possible regulatory genes, *clmR1* and *clmR2*, and five putative transport genes, *clmT1* to *clmT5*, are located on the left-hand side of the collismycin cluster. *clmR1* would code for a transcriptional regulator of the TetR family. It contains the usual N-terminal helix-turn-helix (HTH) DNA binding domain that is characteristic of this family, involving three α helices and part of the fourth one. TetR regulators act mainly as transcriptional repressors, and they control drug resistance and antibiotic production, as well as other metabolic processes (Ramos et al., 2005). Inactivation of *clmR1* does not abolish collismycin biosynthesis but rather anticipates the starting time for collismycin biosynthesis (data not shown). The second regulatory gene, *clmR2*, shows high homology with LuxR-family transcriptional regulators. They all have a C-terminal HTH DNA binding domain that forms a four-helical structure, whereas the N-terminal region of the protein usually contains an auto-inducer binding domain or a response regulator domain (Aravind et al., 2005). Unlike them, ClmR2 does not show any homology or identifiable domain in its N-terminal region, making this protein more similar to GerE, a small transcriptional regulator from *Bacillus subtilis*, which represents the simplest member of this family, since it lacks the usual "receiver" domain and is only composed of the HTH DNA binding domain (Ducros et al., 2001).

Genes *clmT1* and *clmT2* code for proteins homologous to ABC transporters, in which the hydrophobic transmembrane domain and the ATP-binding domain are fused in a single polypeptide. The hydrophobic domain is located in the N-terminal region of the protein and would encode six transmembrane helices, and the nucleotide binding domain is in the C-terminal region and contains the Walker A and Walker B motifs characteristic of the ATP-binding cassette family (Méndez and Salas, 2001). These genes are located together and transcribed in the same direction, opposite to that of the surrounding genes. The products of these two genes could form a transport system with two transmembrane domains and two nucleotide binding domains forming a heterodimer (Kos and Ford, 2009; Méndez

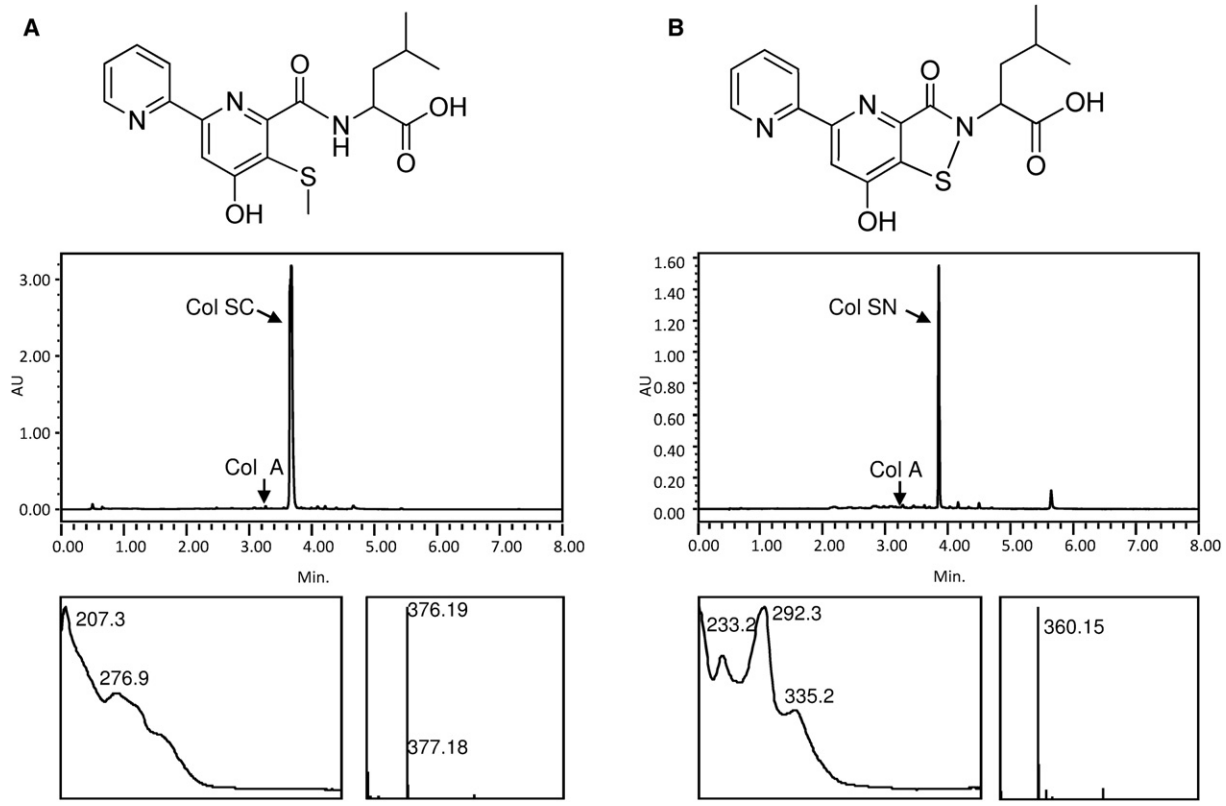

Figure 5. Isolation of Collismycin SC and Collismycin SN

(A) UPLC chromatogram of extracts of mutant CLMAH and chemical structure of collismycin SC (col SC) and its absorption spectrum and mass analysis. For comparison, UPLC retention of collismycin A (col A) is indicated.

(B) UPLC chromatogram of extracts of mutant CLMM1 and chemical structure of collismycin SN (col SN) and its absorption spectrum and mass analysis of collismycin SN (col SN). For comparison, UPLC retention of collismycin A (col A) is indicated.

See also Tables S1 and S2.

and Salas, 2001). The deduced sequence of *clmT3* reveals high similarity to ABC transporter solute-binding proteins (SBPs), such as FepB-like proteins. Structurally, ClmT3 would be similar to cluster A SBPs involved in the binding and uptake of ferric siderophores and several metal ions (Berntsson et al., 2010). The gene *clmT4* would code for the permease component of the transport system, similar to FepD-like proteins. It contains nine predicted transmembrane helices. The ClmT5 predicted sequence shows high homology with the ATPase subunit of ABC transporters involved in ferric iron uptake, such as FepC-type proteins. It contains the characteristic nucleotide-binding domain with one Walker A and one Walker B motif (Méndez and Salas, 2001). The fact that these three genes (*clmT3*, *clmT4*, and *clmT5*) are similar to different components of the Fep transport system involved in the uptake of chelated ferric iron-siderophore complexes (Stephens et al., 1995) and also to three consecutive genes from *Frankia* sp. Ccl3 (YP_483314.1, YP_483316.1, and YP_483315.1, with levels of identity/homology of 31/47%, 53/70%, and 51/65%, respectively) suggests that they could conform a single transport system. Furthermore, the hypothetic permease component of this system (ClmT4), contains putative interaction residues with another transmembrane component, a solute-binding protein (probably ClmT3) and an ATPase domain (possibly ClmT5), which reinforces the idea of a system integrated by three independently coded proteins that interact with each other in order to form a complete functional ABC system.

Unusual Incorporation of a Leucine Residue and Its Further Removal during Biosynthesis

As mentioned previously, a clear role for the ClmAH amidohydrolase was not initially anticipated based on the structure of collismycin and on tentative models for its biosynthesis. We therefore decided to generate a mutant in *clmAH* by gene replacement. The resultant mutant CLMAH did not produce collismycin, but it accumulated a compound that was not detected in the wild-type strain (Figure 5A). MS analysis of this compound revealed a surprisingly high molecular weight: 375 mass units versus 275 mass units for collismycin A. This compound was purified, and its structure was elucidated by using one-dimensional (1D) ^{1}H, two-dimensional (2D) ^{1}H correlated spectroscopy (COSY) ^{1}H, ^{13}C heteronuclear single quantum coherence (HSQC)-edited, and heteronuclear multiple bond correlation (HMBC) nuclear magnetic resonance (NMR) experiments. A first analysis allowed us to identify a 2,2′-bipyridyl fragment identical to that observed in the reference compound collismycin A. Thus, the

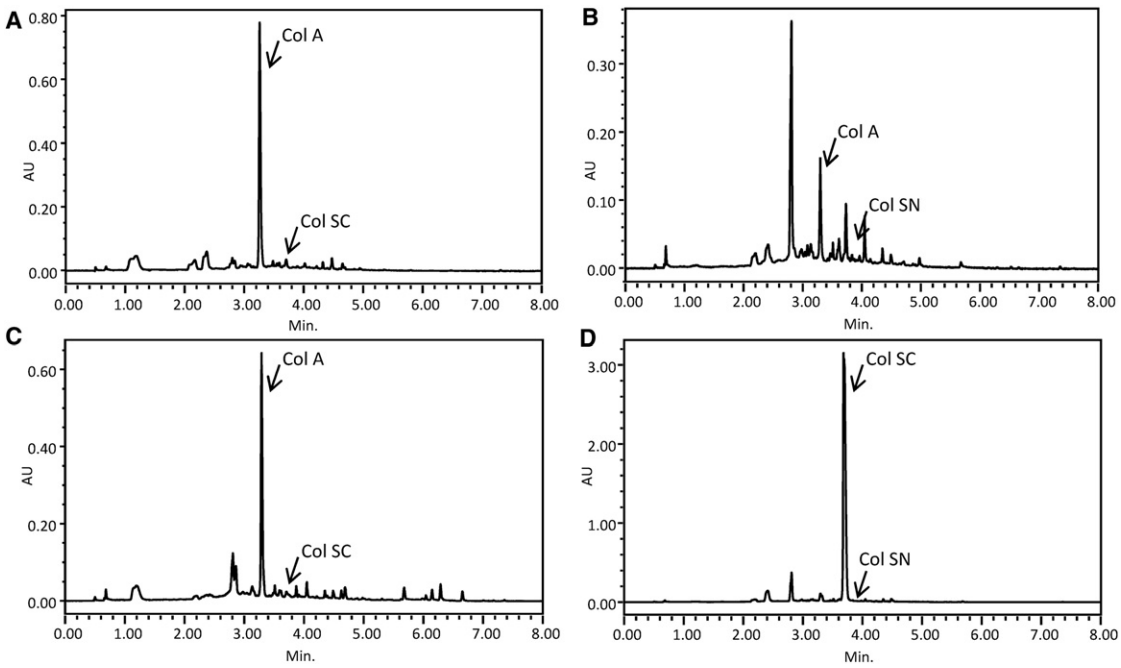

Figure 6. Biotransformation Experiments with Collismycin SC and collismycin SN

(A) Mutant CLM12 grown in the presence of collismycin SC (col SC).

(B) Mutant CLM12 grown in the presence of collismycin SN (col SN).

(C) Mutant CLMM1 grown in the presence of collismycin SC (col SC).

(D) Mutant CLMAH grown in the presence of collismycin SN (col SN).

^{1}H-^{1}H COSY spectrum showed a sequence of 3′-H (δ 8.43), 4′-H (δ 8.01), 5′-H (δ 7.52), and 6′-H (δ 8.72), typical of a mono-substituted pyridine unit, which is connected with a second tetra-substituted pyridine moiety (HMBC spectrum). These four substituents were established as the 2-pyridyl unit, a hydroxyl group ($\delta_{C\text{-}OH}$ 166.6), a methylthio group (δ_H 2.34, δ_C 17.7), and an aminocarbonyl group (δ_H 8.70, δ_C 167.0). Moreover, the ^{1}H-^{1}H coupling of the amidic proton (NH) allowed us to identify a leucine residue attached to this position, which was responsible for the unexpected high molecular weight. This compound was 6-[N-(1-Carboxy-3-methylbutyl)carbamoyl]-4-hydroxy-5-methylthio[2,2′-bipyridine] and was named as collismycin SC.

The presence of this leucine residue in collismycin SC (probably incorporated by the ClmN2 NRPS module) and its absence in collismycin A suggests that a possible role for the amidohydrolase ClmAH could be the removal of the leucine residue before the formation of the oxime group. Interestingly, collismycin SC only contains one of the two methyl groups in collismycin; the methyl group at the sulfur atom is present but the methyl group at the hydroxyl group is absent. We therefore assumed that it was possible that one of the two methyltransferase genes could be responsible for the introduction of the methyl group. Based on the similarities with methyltransferases in databases, ClmM1 was a more likely candidate. Inactivation of ClmM1 by gene replacement rendered the mutant CLMM1, which did not produce collismycin but accumulated a compound, showing also higher mass than collismycin A (359 mass units versus 275 mass units for collismycin A; Figure 5B). We found in this compound analogous chemical shifts and atom connectivity to the NMR data of collismycin SC but without resonances from the N-H amidic proton and the methylthio group. After careful analysis, we established unambiguously that the pyridine unit of this compound is fused with an N-substituted isothiazole ring. This compound was 2-(1-Carboxy-3-methylbutyl)-3-oxo-5-(2-pyridyl)-2,3-dihydro-1,2-thiazolo[4,5-b]pyridine and was named collismycin SN. From here we deduced that ClmM1 is responsible for the *S*-methylation, event causing the breakage of the N-S bond and subsequent isothiazole ring opening.

The unexpected structures of collismycin SC and collismycin SN prompted us to discard the possibility that they could be shunt products and not real intermediates in collismycin biosynthesis. With this aim, we carried out some biotransformation experiments in which either collismycin SC or SN was fed to the CLM12 (in which the *clmN1* NRPS gene was inactivated), the CLMAH, and the CLMM1 mutants. Both collismycin SC and SN were efficiently converted into collismycin A by the CLM12 mutant (Figures 6A and 6B). In addition, collismycin SC was converted into collismycin A by the CLMM1 mutant (Figure 6C), whereas collismycin SN was converted into collismycin SC by the CLMAH mutant (Figure 6D). These experiments demonstrate that (1) collismycin SN and SC are biosynthetic intermediates; (2) the amidohydrolase ClmAH acts after the methyltransferase ClmM1; and (3) collismycin SN is the substrate of ClmM1, which introduces the methyl group at the sulfur atom, and consequently, ClmM2 should be responsible for the methylation of the hydroxyl group at position 4 of collismycin in later steps of the biosynthesis.

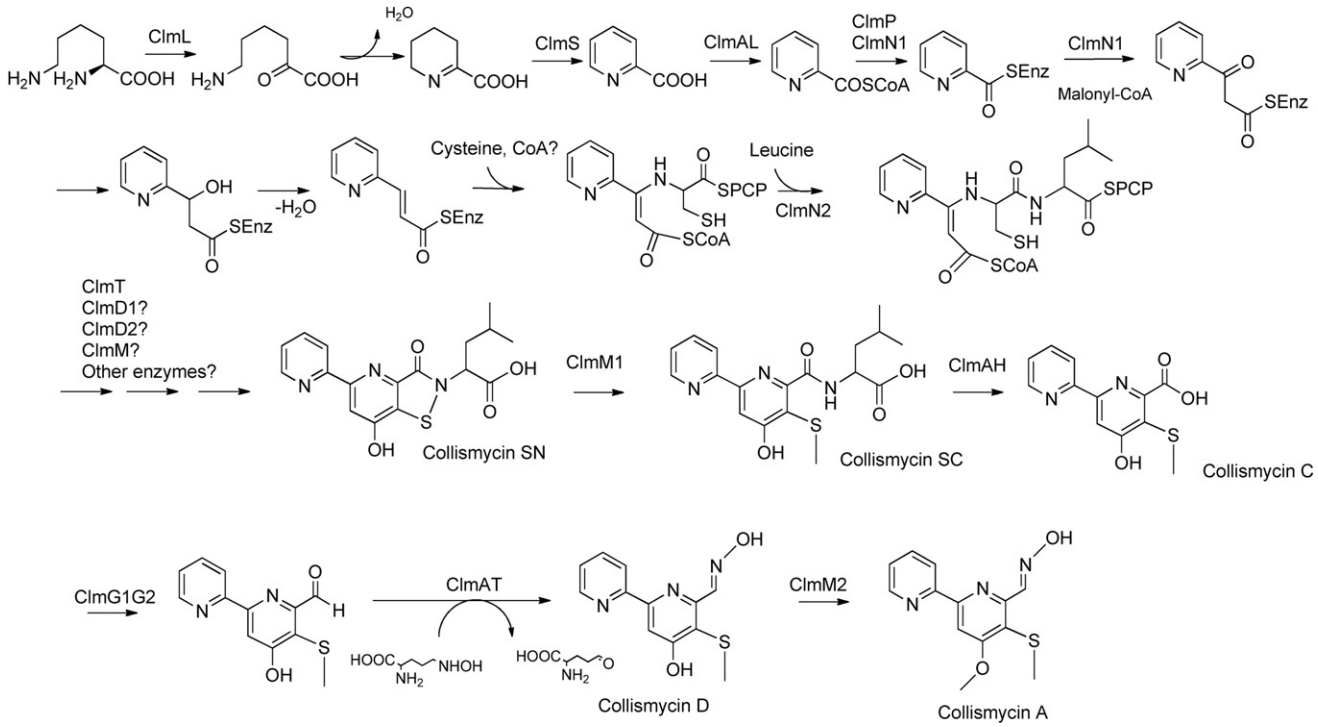

Figure 7. Proposed Pathway for the Biosynthesis of Collismycin A

The different steps in collismycin A biosynthesis with putative assignations for the different enzymes are indicated.

DISCUSSION

The gene cluster for the bipiridyl collismycin A has been isolated and characterized and found to extend along a chromosomal region of approximately 36 kb. For the isolation of the cluster, we took advantage of the use of a specific gene probe, the lysine 2-aminotransferase gene, which commits lysine into the biosynthetic pathway through the formation of picolinic acid. The cluster consists of 23 genes: 16 coding for structural proteins; two coding for pathway regulatory proteins; and five coding for collismycin secretion. The involvement of the isolated cluster in collismycin A biosynthesis was confirmed by insertional inactivation of a hybrid PKS-NRPS gene, and the limits of the cluster were identified by generating mutants in flanking genes. Interestingly, database comparisons of the genes of this cluster allowed the identification of a similar cluster in *S. roseosporus*, which had not been annotated as such. Both clusters were identical at the level of genetic organization and with a very high degree of identity/similarity (above 93%), suggesting a common evolutionary origin.

Based on the deduced products of the genes involved in collismycin A biosynthesis and the chemical structures of the two intermediates isolated from mutants CLMAH and CLMM1, we propose a pathway for collismycin A biosynthesis (Figure 7) through three different phases: (1) biosynthesis and activation of picolinic acid from L-lysine; (2) formation of the 2,2′-bipiridyl ring; and (3) modifications of the 2,2′-bipyridyl structure.

In the first phase, picolinic acid would be synthesized from L-lysine with the participation of two enzymes, ClmL (lysine 2-aminotransferase) and ClmS (oxidase), homologs of which have been found in other biosynthesis pathways also utilizing picolinic acid as biosynthesis precursor, such as nikkomycin and virginiamycin pathways. Picolinic acid would be finally activated by ClmAL (acyl CoA ligase) to picolinyl-CoA, which would be then loaded to the hybrid PKS-NRPS enzyme by ClmP, acting as a loading enzyme. As a result of the action of these enzymes, the first pyridine ring of collismycin is formed.

Formation of the second pyridine ring would occur in the second phase and would require the participation of a PKS-NRPS system, involving two genes (*clmN1* and *clmN2*) for the incorporation of a malonyl-CoA extender unit, a cysteine and a leucine. The incorporation of a malonate unit and a cysteine would be catalyzed by the ClmN1 hybrid PKS-NRPS. The enzymes catalyzing the proposed reduction steps following malonyl-CoA incorporation are not yet identified. Interestingly, the incorporation of the cysteine residue likely occurs by Michael addition (1,4-addition) of the cysteine rather than the usual direct 1,2-attack at the thioester carbonyl, something very unusual for NRPS but not for PKS (Mihali et al., 2008). From the chemical reactivity, a 1,4-Michael addition is clearly preferred over an 1,2-addition, and enzymes usually follow chemical reactivity logic. Moreover, proposing a Michael addition here also facilitates the overall biosynthetic process, by avoiding a chemically and enzymatically unprecedented and unlikely rearrangement after the fact. Note that a Michael addition was also proposed for the first NRPS module reaction in bleomycin biosynthesis (Shen et al., 2002). Along with the cysteine incorporation, the switch from ACP- to the PCP-tethering occurs. Simultaneously with this, we also propose CoA-activation of the former

ACP-linked malonyl-carbonyl that is necessary for the ring cyclization reaction that happens later. This carbonyl needs to be activated sooner or later and not necessarily as a CoA ester; it also could form an anhydride with phosphoric acid (which would occur through phosphorolysis, like in the well-known glyceraldehyde-3-phosphate dehydrogenase reaction of glycolysis). That activation likely happens along with the cysteine incorporation but may happen also later; therefore, the question mark appears in Figure 7.

The incorporation of an extra amino acid, a leucine, would be catalyzed by the ClmN2 NRPS. The presence of malonate and cysteine had been already anticipated based on the structure of collismycin, on predictions for a possible biosynthesis pathway, and on amino acid signatures present in the PKS and NRPS domains of *clmN1*. However, the incorporation of a leucine residue was completely unexpected. Insertional inactivation of the *clmAH* amidohydrolase and of the *clmM1* methyltransferase genes gave rise to the isolation of two compounds, collismycin SC and collismycin SN, with masses clearly higher than that of collismycin A and both containing an extra leucine residue elongating the cysteine. Both compounds are biosynthetic intermediates (and not shunt products), since they were efficiently converted to collismycin A when fed to a deletion mutant in the PKS-NRPS ClmN1. This temporarily bound leucine residue might help to fit the proposed intermediate into the enzyme responsible for formation of the second pyridine ring. The effect of its isobutyl side chain might also contribute to an enhanced nucleophilicity of ring methylene group next to the sulfur atom that is necessary for the ring closure. Breakage of the S-N bond would be catalyzed by the ClmM1 methyltransferase during the methylation event. This was demonstrated by the conversion of collismycin SN into collismycin SC when collismycin SN was added to cultures of the CLMAH mutant. The presence of a third ring in collismycin SN was surprising and induced us to propose a mechanism in which the formation of this third ring is necessary for the generation of the 2,2′-bipyridyl structure followed by the breakage of this third ring when the ClmM1 methyltransferase methylates the sulfur atom. Finally, for the formation of the 2,2′-bipyridyl ring, two oxidations and one dehydration step would be necessary and catalyzed by enzymes not yet well defined in the gene cluster.

In the final phase, the leucine residue would be removed by the ClmAH amidohydrolase followed by the reduction of the carboxyl group to an aldehyde group (ClmG1 and ClmG2 proteins) and the formation of the oxime by the ClmAT aminotransferase and finally the methylation of the hydroxyl group at position 4 by the ClmM2 methyltransferase. The presence of an oxime function in collismycin A is not very frequent in bacterial natural products. In plants, the formation of oximes is usually mediated by members of the cytochrome P450 family (Sibbesen et al., 1995; Wittstock and Halkier, 2000), and the participation of the cytochrome P450 NocL in the conversion of a primary amine to an oxime in nocardicin A biosynthesis has been also reported (Kelly and Townsend, 2002). In the collismycin A cluster no gene coding for a cytochrome P450 has been found. We therefore propose a role for the ClmAT transaminase in the introduction of the oxime group using either delta-(N)-OH ornithine or epsilon-(N)-hydroxy-lysine as a donor. Experiments now in progress in our laboratory are focused on the generation of individual knock-outs in most of the genes of the cluster in order to clearly establish a definitive function for the different genes and their products and the isolation of collismycin A derivatives.

SIGNIFICANCE

Collismycin A is synthesized through a hybrid polyketide synthase-nonribosomal peptide synthetase (PKS-NRPS) system, leading to the formation of the 2,2′-bipyridyl ring structure, characteristic of this compound. This biosynthesis process has some particular features not frequently found in other systems. One is the incorporation of a cysteine residue, possibly by a Michael addition, which is unusual but the most logical way for the design of the bipyridyl system omitting any major rearrangements. Another feature is the incorporation by a NRPS of a leucine residue that no longer remains in the final structure but rather is removed by an amidohydrolase in later steps of the biosynthesis. The presence of this leucine might facilitate the formation of the second pyridine ring, as discussed above. The isolation and characterization of the collismycin A gene cluster opens up the way for future studies aimed at engineering through combinatorial biosynthesis to generate analogs of this compound with potentially improved therapeutic activities.

EXPERIMENTAL PROCEDURES

Strains, Culture Conditions, and Plasmids

Bacterial strains used in this work were *Streptomyces* spp. CS40, collismycin producer, and *Escherichia coli* strains DH10B (Invitrogen, Carlsbad, CA, USA) and ET12567 (pUB307; Kieser et al., 2000) used for subcloning and strain LE392MP (Epicenter, Madison, WI, USA) used to generate and propagate the cosmid library. Growth media for *Streptomyces* strains were tryptone soya broth (TSB) for propagation growth, a medium for sporulation (Fernández et al., 1998), and R5A as production medium (Fernández et al., 1998). Conjugation between *E. coli* and *Streptomyces* was performed following standard procedures (Kieser et al., 2000). *E. coli* media were those described by Sambrook et al. (1989). When clones containing antibiotic resistance gene markers were grown, the medium was supplemented with the appropriate antibiotics: 100 μg/ml ampicillin, 25 μg/ml apramycin, 25 μg/ml chloramphenicol, 50 μg/ml nalidixic acid, 10 μg/ml tetracycline, 50 μg/ml thiostrepton, and 20 μg/ml tobramycin. Plasmids used in this work were pBluescript II SK+ (Fermentas, Glen Burnie, MD, USA) for routine cloning, pCR-Blunt (Invitrogen) and pGEM-T (Promega, Madison, WI, USA) for subcloning of PCR products, and pHZ1358 (Sun et al., 2009) and pOJ260 (Bierman et al., 1992) for gene replacement and gene disruption experiments in *Streptomyces*, respectively. pUO9090 (Prado et al., 1999) was the source of the apramycin resistance gene *aac3(IV)* and pWE15 (Stratagene) was used for constructing the cosmid library.

DNA Manipulation

DNA manipulations were performed in accordance with standard procedures for *E. coli* (Sambrook et al., 1989) and *Streptomyces* (Kieser et al., 2000). A cosmid library of *Streptomyces* spp. CS40 genomic DNA was constructed. DNA fragments obtained from a partial Sau3AI digestion were ligated to BamHI-digested cosmid pWE15, and the resultant recombinant DNA was packaged using MaxPlax Lamda Packaging extracts (Epicenter) in accordance with the manufacturer's handbook. After infection of *E. coli* LE392MP, the resultant *E. coli* transductants were selected with ampicillin, picked, transferred to a 96-well micro-titer plates containg Luria broth medium and ampicillin to construct the *Streptomyces* spp. CS40 gene library, and kept at −20°C until use. Clones from the library were replica plated onto Luria agar plates containing ampicillin, and after overnight growth at 37°C, colonies were transferred to nylon membrane filters for in situ colony hybridization analysis as described (Sambrook et al., 1989) and screened using labeled probe

generated using DIG DNA labeling and detection kit (Roche, Madison, WI, USA).

DNA Sequencing and Analysis of the Collismycin Gene Cluster

The collismycin gene cluster was located in a DNA region delimited by cosmid cos1c3 and a 7,748 bp EcoRV/HindIII fragment from cosmid cos3B11 (sites E–H, Figure 1), which partially overlaps cos1c3. An EcoRV/HindIII fragment from cosmid cos3B11 was cloned into pBluescript II SK+ to generate pBSKB8. Sequencing was performed on double-stranded DNA templates with the dideoxynucleotide chain-reaction method (Sanger et al., 1977) and the Cy5 Autocycle Sequencing Kit (Pharmacia Biotech, Piscataway, NJ, USA). An Alf-express automatic DNA sequencer (Pharmacia Biotech) was used. Computer-aided database searching and sequence analysis were carried out with the University of Wisconsin Genetics Computer Group software (Devereux et al., 1984) and the Blast program (Altschul et al., 1990). Analysis of PKS and NRPS predicted proteins was carried out using programs ASMPKS (Tae et al., 2007) and the NRPS predictor (Rausch et al., 2005). Analysis of transmembrane regions in putative membrane proteins were carried out using program TMHMM (version 2.0; Krogh et al., 2001).

Construction of Plasmids for Gene Inactivation

Involvement of hybrid PKS-NRPS gene in collismycin biosynthesis was determined by gene disruption. For this purpose pOJB12 was generated in the *Streptomyces* suicide vector pOJ260 (Bierman et al., 1992). pOJB12 was obtained by subcloning a 3.1 kb BamHI fragment from cos1c3 into pOJ260. This construct was used for generation of mutant strain CLM12.

The boundaries of the gene cluster were determined by inactivation of *orf1*, *orf2*, *orf3*, *orf4*, and *clmR1* as follows.

(1) *orf1*. For inactivation of *orf1* a 2.1 kb EcoRV/BamHI fragment (sites E–1, Figure 1) was obtained from pBSKB8 and subcloned into EcoRV site of pUO9090 to generate pUO2K. Then, a 3.6 kb BamHI (sites 1–2, Figure 1) fragment obtained from pBSKB8 was subcloned into the blunt-ended HindIII site of pUO2K to generate pUO2K3K. From this construct, a 6.5 kb SpeI fragment was rescued (containing *orf1* disrupted by the insertion of the *aac3(IV)* cassette) and cloned into BamHI blunt-ended pHZ1358 to generate pHZ2K3K that was finally used for the gene replacement experiment.

(2) *orf2*. For inactivation of orf2 the 5′- and 3′-end regions of orf2 were amplified by PCR. A 1,526 bp fragment from the 5′-end was amplified from cos1c3 using oligoprimers orf22 del1 (5′-GAATTCCGGTGAAT CGTCGCGG-3′, EcoRI underlined) and orf22 del2 (5′-AAGCTTCGG CACGCCTCGCGT-3′, HindIII underlined). A 1,500 bp fragment from the 3′-end was amplified from cos1c3 using oligoprimers orf22 del3 (5′-GATATCGCGTAGCCCGTCCACGG-3′, EcoRV underlined and orf22 del4 (5′-GATATCGCTATCTGCTGACCGGCC-3′, EcoRV underlined). PCR conditions used were 97°C, 5 min; 30 cycles of 95°C, 30 s, 55°C and 68°C, 1 min and a final extension at 68°C 10 min. Pfx DNA polymerase (Invitrogen) and 1x PCR Enhancer Solution (Invitrogen) were used for all amplifications. The PCR products were then subcloned into pCR-Blunt and sequenced. The 5′-end fragment was recovered by EcoRI/HindIII digestion and subcloned into pUO9090 to generate pU022A. The 3′-end was recovered by EcoRV digestion and subcloned into pUO22A to generate pUO22AB. From this construct, a 4.5 kb SpeI fragment was rescued (containing most of orf2 deleted and substituted by the *aac3(IV)* cassette) and cloned into BamHI blunt-ended pHZ1358 to generate pHZ22AB that was finally used for the gene replacement experiment.

(3) *orf3*. For inactivation of orf3 the 5′- and 3′-end regions of orf3 were amplified by PCR. A 1,589 bp fragment from the 5′-end was amplified from cos1c3 using oligoprimers orf23 del1 (5′-GAATTCGCGATCTCGA GCTCGGCCTT-3′, EcoRI underlined) and orf23 del2 5′-AAGCTTTTGT GATCAGTTGAGGCCGGG-3, HindIII underlined). A 1,653 bp fragment from the 3′-end was amplified from cos1c3 using oligoprimers orf23 del3 (5′-GGATCCGCATCCTCGACGTGAACCCC-3′, BamHI underlined) and orf23 del4 (5′-GATATCTACTCCAGGGTGCCGTCGGG-3′, EcoRV underlined). PCR conditions were the same as described above. The PCR products were then subcloned into pCR-Blunt and sequenced. The 5′-end fragment was recovered by EcoRI/HindIII digestion and subcloned into pUO9090 to generate pU023A. The 3′-end was recovered by EcoRV/BamHI digestion and subcloned into pUO23A to generate pUO23AB. From this construct, a 4.7 kb SpeI fragment was rescued (containing most of orf3 deleted and substituted by the *aac3(IV)* cassette) and cloned into BamHI blunt-ended pHZ1358 to generate pHZ23AB that was finally used for the gene replacement experiment.

(4) *orf4*. For inactivation of *orf4,* a 2,9 kb BamHI fragment (sites 20–21, Figure 1) from cos1c3 was cloned into pBluescript SK + to generate pBSKB23. From this construct, a 2.2 kb ClaI/BamHI fragment (sites C–21, Figure 1) was obtained and cloned into BamHI blunt-ended pOJ260 to generate pOJB23P. This plasmid was used for the gene disruption experiment.

(5) *clmR1*. For inactivation of *clmR1*, a 3.6 kb BamHI fragment (sites 1–2, Figure 1) from pBSKB8 was cloned into the BamHI site of pBluescript II SK + to generate pBSK3K. Then, the *aac3(IV)* cassette from pUO9090 was rescued with HindIII/EcoRV, blunt-ended and cloned into the blunt-ended NcoI site (site N, Figure 1) of the previous construct pBSK3K generating pBSK3KA. From this construct, a 5.1 kb NotI/EcoRI fragment was recovered, blunt-ended and cloned into BamHI blunt-ended pHZ1358 to generate pHZ3KA. This plasmid was used for the gene replacement experiment.

Two additional mutants were generated in the *clmAH* and *clmM1* genes as follows.

(1) *clmAH*. A 1,504 bp fragment upstream *clmAH* was amplified by PCR using oligoprimers orf13 del1alt (5′-CTGCACACGGCCGTCAC-3′) and orf13 del2 (5′-GGTGCTACGCCTCCACG-3′), using the PCR conditions described above. The PCR product was subcloned into pCR-blunt and sequenced. This construct was then digested with EcoRI and a 1.5 kb fragment subcloned into pUO9090 to generate pUO13A. Then, a 1.9 kb EcoRI/PvuI fragment from cos1c3 (sites EI–P, Figure 1) was blunt-ended and cloned into pCR-blunt vector. This construct was verified by sequencing, and the 1.9 kb EcoRI fragment recovered and cloned into EcoRV-digested pUO13A to generate pUO13AB. From this construct, a 4.9 kb SpeI fragment was recovered, blunt-ended and cloned into BamHI blunt-ended pHZ1358 to generate pHZ13AB. This plasmid was used for gene replacement experiment.

(2) *clmM1*. For inactivation of clmM1 the 5′- and 3′-end regions of clmM1 were amplified by PCR. A 1,597 bp fragment from the 5′-end was amplified from cos1c3 using oligoprimers orf9 del1 (5-GAATTCTGGA CGAGGACGGTCAGGAGG-3′, EcoRI underlined) and orf9 del2 (5′-AA GCTTATGCTCCATCAGTCCCGACACGA-3′, HindIII underlined). A 1,542 bp fragment from the 3′-end was amplified from cos1c3 using oligoprimers orf9 del3 (5′-TTGTACCGGGGTAGCCGTCGC-3′) orf9 del4 (5′-CGAGTTCAAGCCGGTACGGGAGT-3′). PCR conditions were the same as described above. The PCR products were then subcloned into pCR-Blunt and sequenced. The 5′-end fragment was recovered by EcoRI/HindIII digestion and subcloned into pUO9090 to generate pU09A. The 3′-end was recovered by EcoRI digestion and subcloned into pUO9A to generate pUO9AB. From this construct, a 4.6 kb SpeI fragment was rescued (containing most of clmM1 deleted and substituted by the *aac3(IV)* cassette) and cloned into BamHI blunt-ended pHZ1358 to generate pHZ9AB that was finally used for the gene replacement experiment.

Generation of *Streptomyces spp*. CS40 Mutant Strains

Constructs pOJB23P and pOJB12 were introduced into *Streptomyces* spp. CS40 by intergeneric conjugation from *E. coli* ET12567 (pUB307). Transconjugants were selected for resistance to apramycin and integration into the *Streptomyces* spp. CS40 chromosome verified by Southern hybridization.

Constructs pHZ2K3K, pHZ22AB, pHZ23AB, pHZ3KA, pHZ9AB, and pHZ13AB were introduced into *Streptomyces* spp. CS40 by intergeneric conjugation from *E. coli* ET12567 (pUB307). Transconjugants, in which the double-crossover event took place, were identified by their resistance to apramycin and sensitivity to thiostrepton and verified by Southern hybridization.

Analysis of Collismycin Production by UPLC and LC-MS

Routinely, collismycin production was assessed by cultivating *Streptomyces* spp. CS40 or mutants on R5A solid production medium. After six days at 30°C, agar plugs containing 1.5 ml agar media were extracted with 1.5 ml ethyl acetate/1% formic acid, and the presence of collismycin or related compounds in the extracts were analyzed by ultra-performance liquid chromatography (UPLC). In liquid cultures, strains were grown as a seed culture in TSB (30 ml in a 250 ml Erlenmeyer flask). After two days incubation in a rotary incubator (30°C, 250 rpm), 2.5% v/v of the cultures were used to inoculate 30 ml R5A liquid medium. After seven days of incubation, 1 ml aliquots of fermentation broths were removed, and 1 ml ethyl acetate/1% formic acid was added to the sample and mixed vigorously for 30 min. The phases were separated by centrifugation and then the ethyl acetate removed by evaporation using a Speed-Vac. Residues were resuspended in methanol and clarified by centrifugation. Analysis of collismycin production was performed by reversed-phase chromatography on Acquity UPLC equipment with a BEH C18 column (2.1 × 100 mm; Waters, Milford, MA, USA) with acetonitrile and 0.1% trifluoroacetic acid (TFA) as eluent. Samples were eluted with 10% acetonitrile for 1 min, followed by a linear gradient from 10% to 80% acetonitrile over 7 min at a flow rate of 0.5 ml/min and a column temperature of 30°C. For high-performance liquid chromatography-mass spectrometry, an Alliance chromatographic system coupled to a ZQ4000 mass spectrometer and a Symmetry C18 column (2.1 × 250mm; Waters) were used. Solvents were the same as listed previously, and elution was performed with an initial isocratic hold with 10% acetonitrile for 4 min followed by a linear gradient form 10% to 88% acetonitrile over 26 min at 0.25 ml/min. MS analyses were done by electrospray ionization in the positive mode, with a capillarity voltage of 3 kV and a cone voltage of 20 V. Detection and spectral characterization of peaks was performed in both cases by photodiode array detection and Empower software (Waters), extracting bidimensional chromatograms at 332 nm.

Isolation of Collismycins SN and SC

Collismycins SN and SC were purified from 2.4 L cultures of *Streptomyces* spp. mutants CLMM1 and CLMAH, respectively, following culture conditions described previously. The extracts were dissolved in 5 ml of a mixture of dimethyl sulfoxide (DMSO) and methanol (1:1) and centrifuged, and the resulting upper lipid layer was removed. The first purification step was done by chromatography in an XTerra PrepRP18 column (19 × 300 mm; Waters) with acetonitrile and 0.05% TFA in water as solvents. A linear gradient from 30% to 100% acetonitrile in 7 min followed by a 30 min isocratic hold with 100% acetonitrile was used, at a flow rate of 15 ml/min. Peak of interest was collected in 0.1 M phosphate buffer (pH 7.0). The solution obtained was partially evaporated in rotavapor to reduce the acetonitrile concentration and then applied to a solid-phase extraction cartridge (Sec-Pak C18; Waters), washed with water to remove salts and eluted with methanol. Further purifications were performed in isocratic conditions with a Symmetry C18 column (7.8 × 300 mm; Waters), using mixtures of acetonitrile and 0.05% TFA in water, at flow rate of 7 ml/min. As described above peaks were always collected on buffer, desalted by solid-phase extraction and finally, to remove the TFA bound to collismycins, the cartridge was washed with 0.1% ammonia in water, and after further washing with water, the retained compounds were eluted with methanol and again lyophilized.

Biotransformations

Biotransformation experiments were carried out in 24 square, deepwell plates (Enzyscreen, Leiden, The Netherlands). *Streptomyces* spp. CS40 wild-type strain and mutants CLMM12, CLMAH, and CLMM1 were grown as seed cultures in 2 ml of TSB. After 24 hr in an orbital shaker at 30°C and 250 rpm, 0.1 ml of each culture was used to inoculate 3 ml of R5A or R5A supplemented with 0.1 mg of collismycin SN or collismycin SC. After six days of further incubation (30°C/250 rpm), 2 ml of cultures were extracted with 2 ml ethyl acetate/1% formic acid and were analyzed by UPLC.

Structure Elucidation and Characterization of Collismycin Derivatives

The structures of the collismycin derivatives were elucidated by NMR spectroscopy and mass spectrometry. NMR data were recorded in DMSO-d_6 at 298 K using a Bruker Avance Ultrashield Plus 600 spectrometer (600MHz for ^{1}H, 150MHz for ^{13}C experiments) and subsequently processed using the program Topspin (version 1.3; Bruker GmbH, Karlsruhe, Germany). The chemical shifts are given in delta (δ) values and the coupling constants (*J*) in Hertz (Hz). All NMR assignments are based on H,H-COSY, HSQC, HMBC, and NOESY spectra, allowing an unambiguous assignment of all NMR signals.

SUPPLEMENTAL INFORMATION

Supplemental Information includes two tables and can be found with this article online at doi:10.1016/j.chembiol.2012.01.014.

ACKNOWLEDGMENTS

We wish to thank Carlos Olano and Jesús Cortes for a critical reading of our manuscript. This research was supported by a Spanish Ministry of Science and Innovation grant (BIO2009-07643) and the Red Temática de Investigación Cooperativa de Centros de Cáncer (Ministry of Health, ISCIII-RETIC RD06/0020/0026 to J.A.S). We thank the Spanish Ministry of Science and Innovation for the PhD student fellowship (F.P.U.) to Natalia M. Vior.

Received: September 12, 2011
Revised: January 13, 2012
Accepted: January 17, 2012
Published: March 22, 2012

REFERENCES

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

Aparicio, J.F., Molnár, I., Schwecke, T., König, A., Haydock, S.F., Khaw, L.E., Staunton, J., and Leadlay, P.F. (1996). Organization of the biosynthetic gene cluster for rapamycin in *Streptomyces hygroscopicus*: analysis of the enzymatic domains in the modular polyketide synthase. Gene *169*, 9–16.

Aravind, L., Anantharaman, V., Balaji, S., Babu, M.M., and Iyer, L.M. (2005). The many faces of the helix-turn-helix domain: transcription regulation and beyond. FEMS Microbiol. Rev. *29*, 231–262.

Berntsson, R.P., Smits, S.H., Schmitt, L., Slotboom, D.J., and Poolman, B. (2010). A structural classification of substrate-binding proteins. FEBS Lett. *584*, 2606–2617.

Bierman, M., Logan, R., O'Brien, K., Seno, E.T., Rao, R.N., and Schoner, B.E. (1992). Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces* spp. Gene *116*, 43–49.

Bruntner, C., and Bormann, C. (1998). The *Streptomyces tendae* Tü901 L-lysine 2-aminotransferase catalyzes the initial reaction in nikkomycin D biosynthesis. Eur. J. Biochem. *254*, 347–355.

Bruntner, C., Lauer, B., Schwarz, W., Möhrle, V., and Bormann, C. (1999). Molecular characterization of co-transcribed genes from *Streptomyces tendae* Tü901 involved in the biosynthesis of the peptidyl moiety of the peptidyl nucleoside antibiotic nikkomycin. Mol. Gen. Genet. *262*, 102–114.

Carlson, J.C., Fortman, J.L., Anzai, Y., Li, S., Burr, D.A., and Sherman, D.H. (2010). Identification of the tirandamycin biosynthetic gene cluster from *Streptomyces sp*. 307-9. ChemBioChem *11*, 564–572.

Chang, Z., Sitachitta, N., Rossi, J.V., Roberts, M.A., Flatt, P.M., Jia, J., Sherman, D.H., and Gerwick, W.H. (2004). Biosynthetic pathway and gene cluster analysis of curacin A, an antitubulin natural product from the tropical marine cyanobacterium *Lyngbya majuscula*. J. Nat. Prod. *67*, 1356–1367.

Chater, K.F., and Chandra, G. (2008). The use of the rare UUA codon to define "expression space" for genes involved in secondary metabolism, development and environmental adaptation in *streptomyces*. J. Microbiol. *46*, 1–11.

Coque, J.J., Liras, P., Laiz, L., and Martín, J.F. (1991). A gene encoding lysine 6-aminotransferase, which forms the beta-lactam precursor alpha-aminoadipic acid, is located in the cluster of cephamycin biosynthetic genes in *Nocardia lactamdurans*. J. Bacteriol. *173*, 6258–6264.

de Crécy-Lagard, V., Blanc, V., Gil, P., Naudin, L., Lorenzon, S., Famechon, A., Bamas-Jacques, N., Crouzet, J., and Thibaut, D. (1997). Pristinamycin I

biosynthesis in *Streptomyces pristinaespiralis*: molecular characterization of the first two structural peptide synthetase genes. J. Bacteriol. *179*, 705–713.

Devereux, J., Haeberli, P., and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. *12*, 387–395.

Di Lorenzo, M., Stork, M., Naka, H., Tolmasky, M.E., and Crosa, J.H. (2008). Tandem heterocyclization domains in a nonribosomal peptide synthetase essential for siderophore biosynthesis in *Vibrio anguillarum*. Biometals *21*, 635–648.

Du, L., and Shen, B. (2001). Biosynthesis of hybrid peptide-polyketide natural products. Curr. Opin. Drug Discov. Devel. *4*, 215–228.

Ducros, V.M., Lewis, R.J., Verma, C.S., Dodson, E.J., Leonard, G., Turkenburg, J.P., Murshudov, G.N., Wilkinson, A.J., and Brannigan, J.A. (2001). Crystal structure of GerE, the ultimate transcriptional regulator of spore formation in *Bacillus subtilis*. J. Mol. Biol. *306*, 759–771.

Fernández, E., Weissbach, U., Sánchez Reillo, C., Braña, A.F., Méndez, C., Rohr, J., and Salas, J.A. (1998). Identification of two genes from *Streptomyces argillaceus* encoding glycosyltransferases involved in transfer of a disaccharide during biosynthesis of the antitumor drug mithramycin. J. Bacteriol. *180*, 4929–4937.

Fischbach, M.A., and Walsh, C.T. (2006). Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms. Chem. Rev. *106*, 3468–3496.

Funk, A., and Divekar, P.V. (1959). Caerulomycin, a new antibiotic from *Streptomyces caeruleus* Baldacci. I. Production, isolation, assay, and biological properties. Can. J. Microbiol. *5*, 317–321.

Gokhale, R.S., Sankaranarayanan, R., and Mohanty, D. (2007). Versatility of polyketide synthases in generating metabolic diversity. Curr. Opin. Struct. Biol. *17*, 736–743.

Gomi, S., Amano, S., Sato, E., Miyadoh, S., and Kodama, Y. (1994). Novel antibiotics SF2738A, B and C, and their analogs produced by *Streptomyces* sp. J. Antibiot. *47*, 1385–1394.

Holm, L., and Sander, C. (1997). An evolutionary treasure: unification of a broad set of amidohydrolases related to urease. Proteins *28*, 72–82.

Kelly, W.L., and Townsend, C.A. (2002). Role of the cytochrome P450 NocL in nocardicin A biosynthesis. J. Am. Chem. Soc. *124*, 8186–8187.

Khosla, C. (2009). Structures and mechanisms of polyketide synthases. J. Org. Chem. *74*, 6416–6420.

Kim, H.J., Pongdee, R., Wu, Q., Hong, L., and Liu, H.W. (2007). The biosynthesis of spinosyn in *Saccharopolyspora spinosa*: synthesis of the cross-bridging precursor and identification of the function of SpnJ. J. Am. Chem. Soc. *129*, 14582–14584.

Kieser, T., Bibb, M., Chater, K., Butter, M., and Hopwood, D. (2000). Practical *Streptomyces* Genetics: A Laboratory manual (Norwich, UK: John Innes Foundation).

Koglin, A., and Walsh, C.T. (2009). Structural insights into nonribosomal peptide enzymatic assembly lines. Nat. Prod. Rep. *26*, 987–1000.

Kos, V., and Ford, R.C. (2009). The ATP-binding cassette family: a structural perspective. Cell. Mol. Life Sci. *66*, 3111–3126.

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. *305*, 567–580.

Marahiel, M.A., Stachelhaus, T., and Mootz, H.D. (1997). Modular peptide synthetases involved in nonribosomal peptide synthesis. Chem. Rev. *97*, 2651–2674.

Martinez Gil, A., Egea, P.U., Medina Padilla, M., García Palomero, E., Perez Baz, J., Fernandez Chimeno, R.I., Medarde Fernandez, A., Canedo Hernandez, L.M., Romero Millan, F., Castro Morera, A., et al. (2007). Use of collismycin and derivatives thereof as oxidative stress inhibitors. Patent WO/2007/017146.

Mattevi, A., Fraaije, M.W., Mozzarelli, A., Olivi, L., Coda, A., and van Berkel, W.J. (1997). Crystal structures and inhibitor binding in the octameric flavoenzyme vanillyl-alcohol oxidase: the shape of the active-site cavity controls substrate specificity. Structure *5*, 907–920.

Mattheus, W., Gao, L.J., Herdewijn, P., Landuyt, B., Verhaegen, J., Masschelein, J., Volckaert, G., and Lavigne, R. (2010). Isolation and purification of a new kalimantacin/batumin-related polyketide antibiotic and elucidation of its biosynthesis gene cluster. Chem. Biol. *17*, 149–159.

Méndez, C., and Salas, J.A. (2001). The role of ABC transporters in antibiotic-producing organisms: drug secretion and resistance mechanisms. Res. Microbiol. *152*, 341–350.

McInnes, A., Smith, D., Wright, J., and Vining, L. (1977). Caerulomycins B and C, new 2, 2 -dipyridyl derivatives from *Streptomyces caeruleus*. Can. J. Chem. *55*, 4159–4165.

McInnes, A., Smith, D., Walter, J., Wright, J., Vining, L., and Arsenault, G. (1978). Caerulomycin D, a novel glycosidic derivative of 3, 4-dihydroxy-2, 2 -dipyridyl 6-aldoxime from *Streptomyces caeruleus*. Can. J. Chem. *56*, 1836–1842.

McInnes, A., Smith, D.G., Walter, J.A., Vining, L.C., and Wright, J.L.C. (1979). The biosynthesis of caerulomycin A in *Streptomyces caeruleus*. Incorporation of 14C- and 13C-labeled precursors and analyses of labeling patterns by ^{13}C NMR. Can. J. Chem. *57*, 3200–3204.

Mihali, T.K., Kellmann, R., Muenchhoff, J., Barrow, K.D., and Neilan, B.A. (2008). Characterization of the gene cluster responsible for cylindrospermopsin biosynthesis. Appl. Environ. Microbiol. *74*, 716–722.

Namwat, W., Kinoshita, H., and Nihira, T. (2002). Identification by heterologous expression and gene disruption of VisA as L-lysine 2-aminotransferase essential for virginiamycin S biosynthesis in *Streptomyces virginiae*. J. Bacteriol. *184*, 4811–4818.

Newman, D.J., and Cragg, G.M. (2007). Natural products as sources of new drugs over the last 25 years. J. Nat. Prod. *70*, 461–477.

Nishiya, Y., Zuihara, S., and Imanaka, T. (1995). Active site analysis and stabilization of sarcosine oxidase by the substitution of cysteine residues. Appl. Environ. Microbiol. *61*, 367–370.

Nishiya, Y., and Imanaka, T. (1996). Analysis of interaction between the *Arthrobacter* sarcosine oxidase and the coenzyme flavin adenine dinucleotide by site-directed mutagenesis. Appl. Environ. Microbiol. *62*, 2405–2410.

Ogasawara, Y., and Liu, H.W. (2009). Biosynthetic studies of aziridine formation in azicemicins. J. Am. Chem. Soc. *131*, 18066–18068.

Pfennig, F., Schauwecker, F., and Keller, U. (1999). Molecular characterization of the genes of actinomycin synthetase I and of a 4-methyl-3-hydroxyanthranilic acid carrier protein involved in the assembly of the acylpeptide chain of actinomycin in *Streptomyces*. J. Biol. Chem. *274*, 12508–12516.

Prado, L., Lombó, F., Braña, A.F., Méndez, C., Rohr, J., and Salas, J.A. (1999). Analysis of two chromosomal regions adjacent to genes for a type II polyketide synthase involved in the biosynthesis of the antitumor polyketide mithramycin in *Streptomyces argillaceus*. Mol. Gen. Genet. *261*, 216–225.

Ramos, J.L., Martínez-Bueno, M., Molina-Henares, A.J., Terán, W., Watanabe, K., Zhang, X., Gallegos, M.T., Brennan, R., and Tobes, R. (2005). The TetR family of transcriptional repressors. Microbiol. Mol. Biol. Rev. *69*, 326–356.

Rausch, C., Weber, T., Kohlbacher, O., Wohlleben, W., and Huson, D.H. (2005). Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs). Nucleic Acids Res. *33*, 5799–5808.

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-specific mutations. Biochemistry *40*, 15464–15470.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA *74*, 5463–5467.

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.

Shen, B., Du, L., Sanchez, C., Edwards, D.J., Chen, M., and Murrell, J.M. (2002). Cloning and characterization of the bleomycin biosynthetic gene cluster from *Streptomyces verticillus* ATCC15003. J. Nat. Prod. *65*, 422–431.

Shindo, K., Yamagishi, Y., Okada, Y., and Kawai, H. (1994). Collismycins A and B, novel non-steroidal inhibitors of dexamethasone-glucocorticoid receptor binding. J. Antibiot. *47*, 1072–1074.

Sibbesen, O., Koch, B., Halkier, B.A., and Møller, B.L. (1995). Cytochrome P-450TYR is a multifunctional heme-thiolate enzyme catalyzing the conversion of L-tyrosine to p-hydroxyphenylacetaldehyde oxime in the biosynthesis of the cyanogenic glucoside dhurrin in *Sorghum bicolor* (L.) Moench. J. Biol. Chem. *270*, 3506–3511.

Sieber, S.A., and Marahiel, M.A. (2005). Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics. Chem. Rev. *105*, 715–738.

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.

Stadler, M., Bauch, F., Henkel, T., Mühlbauer, A., Müller, H., Spaltmann, F., and Weber, K. (2001). Antifungal actinomycete metabolites discovered in a differential cell-based screening using a recombinant TOPO1 deletion mutant strain. Arch. Pharm. (Weinheim) *334*, 143–147.

Stephens, D.L., Choe, M.D., and Earhart, C.F. (1995). *Escherichia coli* periplasmic protein FepB binds ferrienterobactin. Microbiology *141*, 1647–1654.

Strieker, M., Tanović, A., and Marahiel, M.A. (2010). Nonribosomal peptide synthetases: structures and dynamics. Curr. Opin. Struct. Biol. *20*, 234–240.

Sun, Y., He, X., Liang, J., Zhou, X., and Deng, Z. (2009). Analysis of functions in plasmid pHZ1358 influencing its genetic and structural stability in *Streptomyces lividans* 1326. Appl. Microbiol. Biotechnol. *82*, 303–310.

Suzuki, H., Ohnishi, Y., and Horinouchi, S. (2007). GriC and GriD constitute a carboxylic acid reductase involved in grixazone biosynthesis in *Streptomyces griseus*. J. Antibiot. *60*, 380–387.

Tae, H., Kong, E.B., and Park, K. (2007). ASMPKS: an analysis system for modular polyketide synthases. BMC Bioinformatics *8*, 327.

Tsuge, N., Furihata, K., Shin-Ya, K., Hayakawa, Y., and Seto, H. (1999). Novel antibiotics pyrisulfoxin A and B produced by *Streptomyces californicus*. J. Antibiot. *52*, 505–507.

Van Lanen, S.G., and Shen, B. (2008). Advances in polyketide synthase structure and function. Curr. Opin. Drug Discov. Devel. *11*, 186–195.

Vining, L.C., McInnes, A.G., McCulloch, A.W., Smith, D.G., and Walter, J.A. (1988). The biosynthesis of caerulomycins in *Streptomyces caeruleus*. Isolation of a new caerulomycin and incorporation of picolinic acid and glycerol into caerulomycin A. Can. J. Chem. *66*, 191–194.

Walsh, C.T. (2004). Polyketide and nonribosomal peptide antibiotics: modularity and versatility. Science *303*, 1805–1810.

Wenzel, S.C., and Müller, R. (2005). Formation of novel secondary metabolites by bacterial multimodular assembly lines: deviations from textbook biosynthetic logic. Curr. Opin. Chem. Biol. *9*, 447–458.

Whicher, J.R., Florova, G., Sydor, P.K., Singh, R., Alhamadsheh, M., Challis, G.L., Reynolds, K.A., and Smith, J.L. (2011). Structure and function of the RedJ protein, a thioesterase from the prodiginine biosynthetic pathway in Streptomyces coelicolor. J. Biol. Chem. *286*, 22558–22569.

Wittstock, U., and Halkier, B.A. (2000). Cytochrome P450 CYP79A2 from *Arabidopsis thaliana* L. Catalyzes the conversion of L-phenylalanine to phenylacetaldoxime in the biosynthesis of benzylglucosinolate. J. Biol. Chem. *275*, 14659–14666.

Zhang, W., Ames, B.D., Tsai, S.C., and Tang, Y. (2006). Engineered biosynthesis of a novel amidated polyketide, using the malonamyl-specific initiation module from the oxytetracycline polyketide synthase. Appl. Environ. Microbiol. *72*, 2573–2580.