Chemistry & Biology
Article

Spiroketal Polyketide Formation in *Sorangium*: Identification and Analysis of the Biosynthetic Gene Cluster for the Highly Cytotoxic Spirangienes

Bettina Frank,[1] Jens Knauber,[2] Heinrich Steinmetz,[2] Maren Scharfe,[2] Helmut Blöcker,[2] Stefan Beyer,[2] and Rolf Müller[1,2,*]
[1] Pharmaceutical Biotechnology, Saarland University, P.O. Box 151150, 66041 Saarbrücken, Germany
[2] HZI–Helmholtz Centre for Infection Research, Inhoffenstrasse 7, 38124 Braunschweig, Germany
*Correspondence: rom@mx.uni-saarland.de
DOI 10.1016/j.chembiol.2006.11.013

SUMMARY

Natural products constitute important lead structures in drug discovery. In bacteria, they are often synthesized by large, modular multienzyme complexes. Detailed analysis of the biosynthetic machinery should enable its directed engineering and production of desirable analogs. The myxobacterium *Sorangium cellulosum* So ce90 produces the cytotoxic spiroketal polyketide spirangien, for which we describe the identification and functional analysis of the biosynthetic pathway. The gene cluster spans 88 kb and encodes 7 type I polyketide synthases and additional enzymes such as a stand-alone thioesterase and 2 methyltransferases. Inactivation of two cytochrome P_{450} monooxygenase genes resulted in the production of acyclic spirangien derivatives, providing direct evidence for the involvement of these enzymes in spiroketal formation. The presence of large DNA repeats is consistent with multiple rounds of gene duplication during the evolution of the biosynthetic gene locus.

INTRODUCTION

The need for new therapeutic agents continues to be pressing. Bioactive natural products obtained from microorganisms are common sources of such medicines, but the metabolites often also serve as lead structures for drug development [1]. Prominent among these compounds are the polyketides, which are biosynthesized by polyketide synthases (PKSs) from simple acyl-CoA thioester building blocks [2]. The bacterial type I class of PKSs exhibits an assembly line architecture, in which each module within the large multienzymes (usually) catalyzes a single round of chain extension and reductive processing. Thus, the genetic organization is colinear with the sequence of biosynthetic transformations [3]. Each module contains an acyl transferase (AT) domain responsible for selection of extender units and a ketosynthase (KS) domain that incorporates the building blocks into the growing chain via a Claisen-like condensation reaction. The third essential domain within each module is the acyl carrier protein (ACP) to which the intermediates are tethered in thioester linkages via the phosphopantetheine prosthetic arm. The modules optionally incorporate domains for reductive modification of the β-keto functionalities as in fatty acid biosynthesis [2], including ketoreductase (KR), dehydratase (DH), and enoyl reductase (ER) activities. The precise complement of these elements determines the overall extent of reduction to a β-hydroxyl, an α,β double bond, or a fully reduced methylene. The completely processed polyketide chain is usually released from the enzyme complex by a C-terminal thioesterase (TE) domain through intramolecular lactonization or hydrolysis. Frequently, the enzyme-free intermediate is then modified by a series of regio- and stereospecific "tailoring" reactions such as methylation, oxidation, and glycosylation, which is usually important for the bioactivity [4].

The pharmacological and physiochemical properties of natural products can often be enhanced by modifying their structures, which may be essential for the therapeutic application. As natural products exhibit a high degree of structural and stereochemical complexity, preparation by total synthesis represents a significant challenge. A more economically viable alternative is to generate analogs by "semisynthesis," starting from existing structures [5]. A potentially powerful alternative strategy is to genetically engineer the polyketide pathways in order to rearrange the assembly lines in a targeted manner. The modular organization of type I PKSs makes them ideal candidates for approaches aimed at deleting, swapping, and mutating individual domains and modules [3]. Modification of the PKS assembly line can give rise to shortened scaffolds, altered stereochemistry, and compounds derived from new combinations of building blocks.

Tailoring enzymes have been used extensively for combinatorial biosynthesis, as they are easier to target than domains within the megasynthases, and the outcome of modifying their functions tends to be more predictable. These enzymes can be inactivated, mutated, replaced, or substituted with tailoring enzymes from other pathways, giving rise, for example, to molecules with altered oxidation states or glycolsylation patterns [3, 6]. Such

modifications may result in structural changes that are very difficult to achieve by standard synthetic chemistry, and they may also generate advanced intermediates for use in medicinal chemistry. Progress and challenges in the field recently have been described in two excellent reviews [3, 6].

Nature itself exploits the flexibility of modular systems, as demonstrated by the increasing number of PKS systems that exhibit deviations from colinearity, for example, by skipping or iterative use of modules [7, 8]. Although much has been discovered about these systems over the last decade, the analysis of newly sequenced gene clusters continues to reveal novel features of the biosynthetic pathways. However, to take full advantage of the opportunities Nature offers for combinatorial biosynthesis of altered compounds, it will be essential to decipher the molecular logic of polyketide assembly in significantly greater detail.

The genes that encode PKSs, tailoring enzymes, and other proteins involved in the biosynthesis of polyketides are usually found clustered on the chromosome, which eases their identification and analysis. With the sequencing of increasing numbers of microbial genomes, potential biosynthetic gene clusters can be identified through in silico analysis. However, a more common strategy is to screen genomic libraries for the presence of highly conserved genes within the PKSs [9].

Myxobacteria are ubiquitous in soil and are proficient producers of natural products with various biological activities. Interestingly, most of the myxobacterial secondary metabolites have been isolated from the genus *Sorangium* [10]. The complex life cycle of these bacteria and their extraordinary ability to produce secondary metabolites is mirrored by their enormous genome size. The 13.1 Mbp genome of the model strain *Sorangium cellulosum* So ce56 is the largest yet discovered in bacteria (*S. cellulosum* genome project within the BMB+F Genomik network, unpublished data and [11]). The challenges with *Sorangium sp*. are difficult handling, slow growth (generation times from 8 to 16 hr), and limited tools for genetic manipulation [12]. *S. cellulosum* So ce90 produces the epothilones, which are hybrid natural products derived from PKSs and nonribosomal peptide synthetases (NRPSs) [13, 14]. The epothilones are currently in phase III clinical trials as anticancer agents and are expected to replace taxol-derived agents once they are approved for human therapy [15].

In this article, we describe the identification and the analysis of the biosynthetic genes for the spiroketal polyketide spirangien (Figure 1) [16] from *S. cellulosum* So ce90. To analyze the function of the PKS and non-PKS enzymes involved, site-directed mutagenesis was performed to inactivate the corresponding genes, resulting in the production of a number of new, to our knowledge, spirangien derivatives. Structure elucidation revealed two sets of desmethyl derivatives and an acyclic compound, which was also produced in the glycosylated form. The spirangien PKS genes contain large stretches of sequence repeats, suggesting that the gene cluster in its current form evolved from a smaller ancestor by several gene duplication events. This may reflect the way Nature uses the flexibility of modular megasynthetases to develop compounds that offer evolutionary benefits to their producing organisms.

RESULTS AND DISCUSSION

Identification of the Spirangien Biosynthetic Gene Cluster

Among PKS domain sequences, those encoding KS domains are highly conserved. Screening of a cosmid genomic library of *S. cellulosum* So ce90 consisting of about 1600 clones with probes derived from KS sequences [17] led to the identification of the overlapping cosmids c77 and c66. Both cosmids were sequenced [17] and were found to contain 47 kb of sequence encoding the 3′ end of a modular PKS biosynthetic gene cluster (Figure 1A). Site-directed mutagenesis of the KS-encoding sequence from *spiH*, by integration of plasmid pJKB16 into the chromosome of *S. cellulosum* So ce90, resulted in loss of spirangien production (for the construction of plasmids, see Table S1 in the Supplemental Data available with this article online). Restriction analysis and end sequencing of cosmid c48, also identified during the screen, revealed a 30 kb overlap with c77. A fragment of the T7 end of cosmid c48 was used to screen a second cosmid library from *S. cellulosum* So ce90 consisting of another 2300 clones, and screening led to the identification of cosmid c4O20. This cosmid was sequenced and was found to harbor the 5′ end of the gene cluster. The remaining 3.8 kb gap was amplified from cosmid c48 by PCR and was sequenced after cloning to give pSpiL.

Sequence Analysis of the Spirangien Biosynthetic Gene Cluster

The spirangien biosynthetic gene cluster is 88.4 kb in size and shows an overall G+C content of 74.8%. Seven type I PKS-encoding genes, *spiDEFGHIJ*, are flanked by a set of three adjacent genes on each side of the gene cluster (Figure 1A; Table 1). *spiA* and *spiB* are located on the complementary strand, and the deduced proteins show sequence similarities to type II thioesterases and *O*-methyltransferases, respectively. Located directly upstream of *spiD* is *spiC*, coding for a cytochrome P_{450} monooxygenase. A second *O*-methyltransferase and another cytochrome P_{450} monooxygenase are encoded by *spiK* and *spiL*, respectively. *spiZ* is located on the complementary strand, and the encoded protein shows similarities to sigma factors.

Deciphering the Biosynthetic Logic of Spirangien Biosynthesis: Analysis of the PKS

The polyketide scaffold of spirangien is synthezised by 7 PKSs (SpiDEFGHIJ) harboring a loading module with an unusual organization, 15 additional modules for the incorporation of extender units, and a terminal TE domain (Figure 2). Thus, the modular arrangement of the PKS follows the colinear paradigm, as 16 building blocks are required

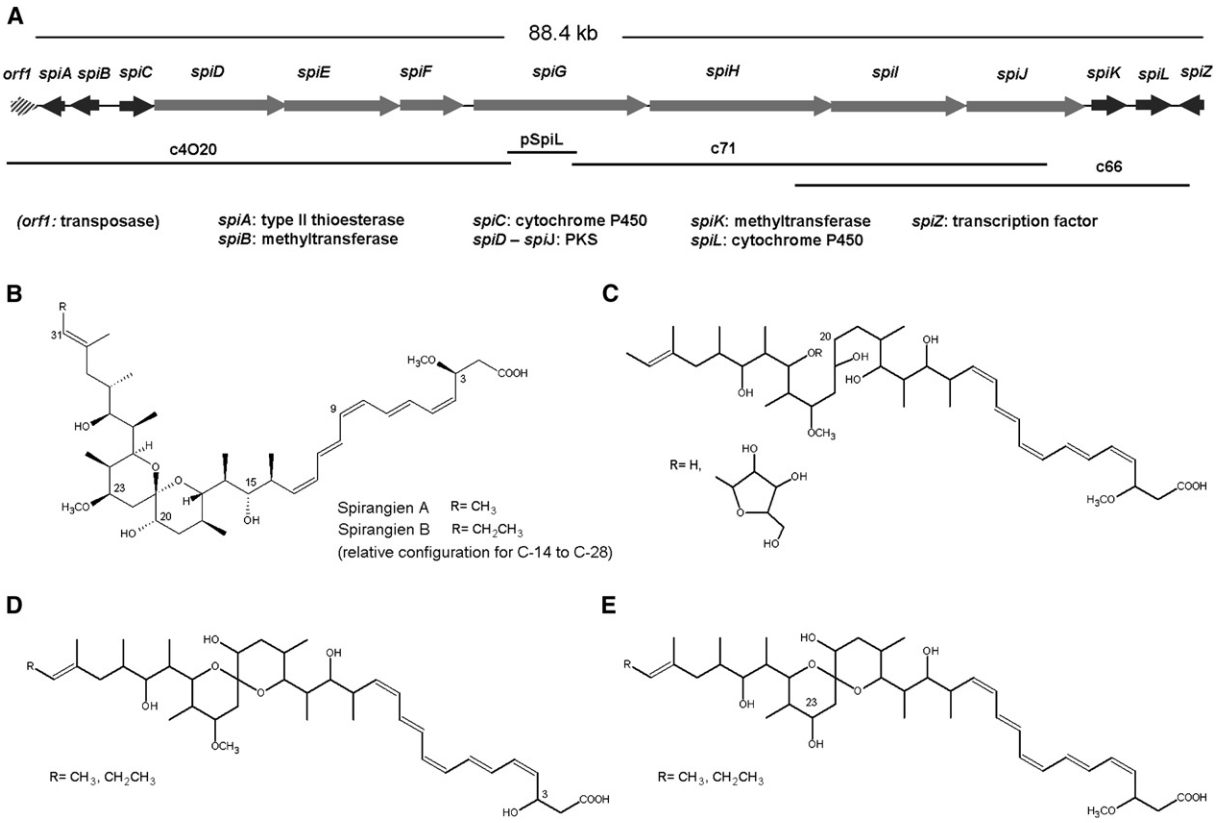

Figure 1. Organization of the Spirangien Biosynthetic Gene Cluster and Structures of Spirangien and Derivatives

(A) The inserts of cosmids c4O20, c71, and c66 and of plasmid pSpiL are indicated by lines. Seven polyketide synthase-encoding genes (light gray) are surrounded by two sets of adjacent genes (dark gray). The open reading frame *orf1* is incomplete (no stop codon). The deduced functions of the proteins encoded are listed below.

(B) Structures of spirangienes A and B.

(C) Acyclic derivatives produced by the *spiL* knockout mutant JKB19.

(D) Desmethylderivatives produced by the *spiK* knockout mutant BFSK.

(E) Desmethylderivatives produced by the *spiB* knockout mutant BFSB.

to assemble the spirangien core. The domain organization of most of the modules also correlates well with the spirangien structure, with only few exceptions. The organization of SpiD (ACP_L-KS-AT_L-AT-DH…) is unusual, but the same domain order is also present in other myxobacterial biosynthetic gene clusters [18–20]. In this module, the catalytic domains for the loading of the starter unit (ACP_L and AT_L) and the corresponding domains for the first elongation step seem to be mixed. The modules downstream of module L1 all contain the minimal set of PKS elongation domains (KS-AT-ACP) in the standard arrangement, as well as combinations of the optional reductive domains (KR, DH-KR, or the complete "reductive loop" DH-ER-KR).

The ACP_L (as all ACP domains present in the cluster) displays the highly conserved serine residue that is required for the conversion into its active *holo* form by attachment of a 4-phosphopantetheinyl arm [2]. The AT domains are responsible for selection of the acyl-CoAs and their transfer to the ACPs. All AT domains present contain the highly conserved active site GHSxG motif. Extensive sequence analysis of AT domains specific for methylmalonyl-CoA or malonyl-CoA has revealed a number of conserved residues [21] that seem to correlate with substrate specificity. A serine residue in position 200 (according to [21]) is highly conserved in the AT domains that select methylmalonyl-CoA, whereas AT domains specific for malonyl-CoA display a conserved phenylalanine residue at the equivalent position. The presence of these residues can therefore be used to predict the substrate specificity of newly sequenced AT domains. The pattern of malonyl-CoA and methylmalonyl-CoA incorporation predicted by sequence analysis of the AT domains in each module exactly matches that found in the spirangienes (the ATs of modules 5, 6, 10, 11, 12, 13, 14, and 15 are malonyl-CoA specific, while those of modules 1, 2, 3, 4, 7, 8, and 9 are methylmalonyl-CoA specific). The AT_L differs from the ATs in elongation modules at a number of conserved residues (Figure S1), consistent with its ability to select both acetyl-CoA and propionyl-CoA starter units (Figure 2).

KR domains are present in all chain-extension modules, and they exhibit a typical NADPH-binding site [22] and

Table 1. Putative Functions of Proteins Encoded in the Sequenced Region

PKS Part of the Spirangien Biosynthetic Gene Cluster

Protein (Gene)	Size (Da/bp)	Proposed Function (Protein Domains with Their Position in the Sequence)
SpiD (*spiD*)	470,885/13,602	ACP (214–411), KS (499–1776), AT (2,056–2,946), AT (3,409–4,299), DH (4,492–4,989), KR (5,974–6,510), ACP (6,847–7,020), KS (7,090–8,370), AT (8,698–9,588), DH (9,784–10,293), ER (11,335–12,258), KR (12,289–12,825), ACP (13,138–13,338)
SpiE (*spiE*)	356,285/10,281	KS (106–1,365), AT (1,684–2,577), KR (3,616–4,148), ACP (4,477–4,677), KS (4,765–6,039), AT (6,367–7,266), DH (7,465–7,977), KR (8,959–9,495), ACP (9,808–10,008)
SpiF (*spiF*)	172,992/4,953	KS (97–1,371), AT (1,699–2,604), KR (3,613–4,149), ACP (4,483–4,683)
SpiG (*spiG*)	554,784/15,996	KS (103–1,383), AT (1,720–2,634), KR (3,670–4,203), ACP (4,522–4,722), KS (4,804–6,084), AT (6,412–7,302), DH (7,498–7,995), ER (9,028–9,951), KR (9,982–10,518), ACP (10,858–11,031), KS (11,113–12,387), AT (12,718–13,611), KR (14,644–15,180), ACP (15,520–15,720)
SpiH (*spiH*)	524,296/15,018	KS (73–1,353), AT (1,672–2,562), KR (3,598–4,131), ACP (4,450–4,650), KS (4,732–6,012), AT (6,331–7,230), KR (8,257–8,793), ACP (9,133–9,333), KS (9,421–10,701), AT (11,014–11,919), DH (12,155–12,621), KR (13,699–14,235), ACP (14,554–14,754)
SpiI (*spiI*)	366,953/10,542	KS (115–1,389), AT (1,717–2,628), KR (3,658–4,194), ACP (4,522–4,722), KS (4,813–6,093), AT (6,436–7,341), DH (7,537–8,037), KR (9,154–9,690), ACP (10,009–10,209)
SpiJ (*spiJ*)	374,780/10,740	KS (106–1,377), AT (1,705–2,616), DH (2,815–3,312), KR (4,354–4,890), ACP (5,209–5,406), KS (5,482–6,755), AT (7,072–7,980), KR (8,839–9,375), ACP (9,670–9,870), TE (10,117–10,722)

Proteins Encoded Upstream of *spiC* and Downstream of *spiJ*

Protein (Gene)	Size (Da/bp)	Proposed Function of the Similar Protein	Sequence Similarity to Source	Similarity/Identity (Amino Acids)	Accession Number of the Similar Protein
SpiA (*spiA*)	29,789/792	putative thioesterase	*Nostoc* sp.	56%/36% (241)	NP_486085.1
SpiB (*spiB*)	35,748/963	possible methyltransferase	*Mycobacterium bovis*	51%/39% (220)	NP_856621.1
SpiC (*spiC*)	51,172/1,434	probable C-type cytochrome	*Pseudomonas aeruginosa*	39%/30% (229)	NP_253309.1
SpiK (*spiK*)	28,863/792	putative methyltransferase	Hypothetical symbiont bacterium of *Paederus fuscipes*	56%/42% (188)	AA547557.1
SpiL (*spiL*)	50,503/1,368	cytochrome P_{450}	*Trichodesmium erythraeum*	61%/40% (447)	ZP_00675674.1
SpiZ (*spiZ*)	23,414/639	sigma-24 factor	*Anaeromyxobacter dehalogenans*	55%/40% (158)	YP_464957.1

a conserved Lys-Ser-Tyr-Asn catalytic tetrad [23]. On the basis of amino acid motifs that seem to correlate with the stereochemistry of ketoreduction, KR domains can be classified as A or B type. The most diagnostic residue for B-type ketoreduction is an aspartate at position 95 [24], and double bonds derived from dehydration of the respective hydroxyl groups are expected to have an *E* (*trans*) configuration. Correspondingly, A-type domains lack this conserved Asp, and dehydration of an A-type hydroxyl could directly yield a *Z* (*cis*) double bond [23–25]. The spirangien PKS appears to fit with this model, as dehydration of hydroxyl intermediates generated by B-type KRs (modules 1, 11, and 13) appears to give rise to *E* double bonds, while the double bonds generated after A-type reduction (modules 10 and 12) are uniformly in *Z* configuration.

DH domains are present in modules 1, 2, 4, 7, 11, 13, and 14 and display the conserved active site motif LxxHxxxGxxxxP, which is characteristic of the enzyme family (Figure 2) [26]. However, in order to generate the conjugated double bond system in the spirangien structure, DH activity is also required after chain extension by modules 10 and 12. We assume that in these cases elimination of water is catalyzed by the DH domains of the downstream modules 11 and 13, respectively. The DH domains of modules 11 and 13 show high similarity to each other (69%), and phylogenetic analysis reveals that they

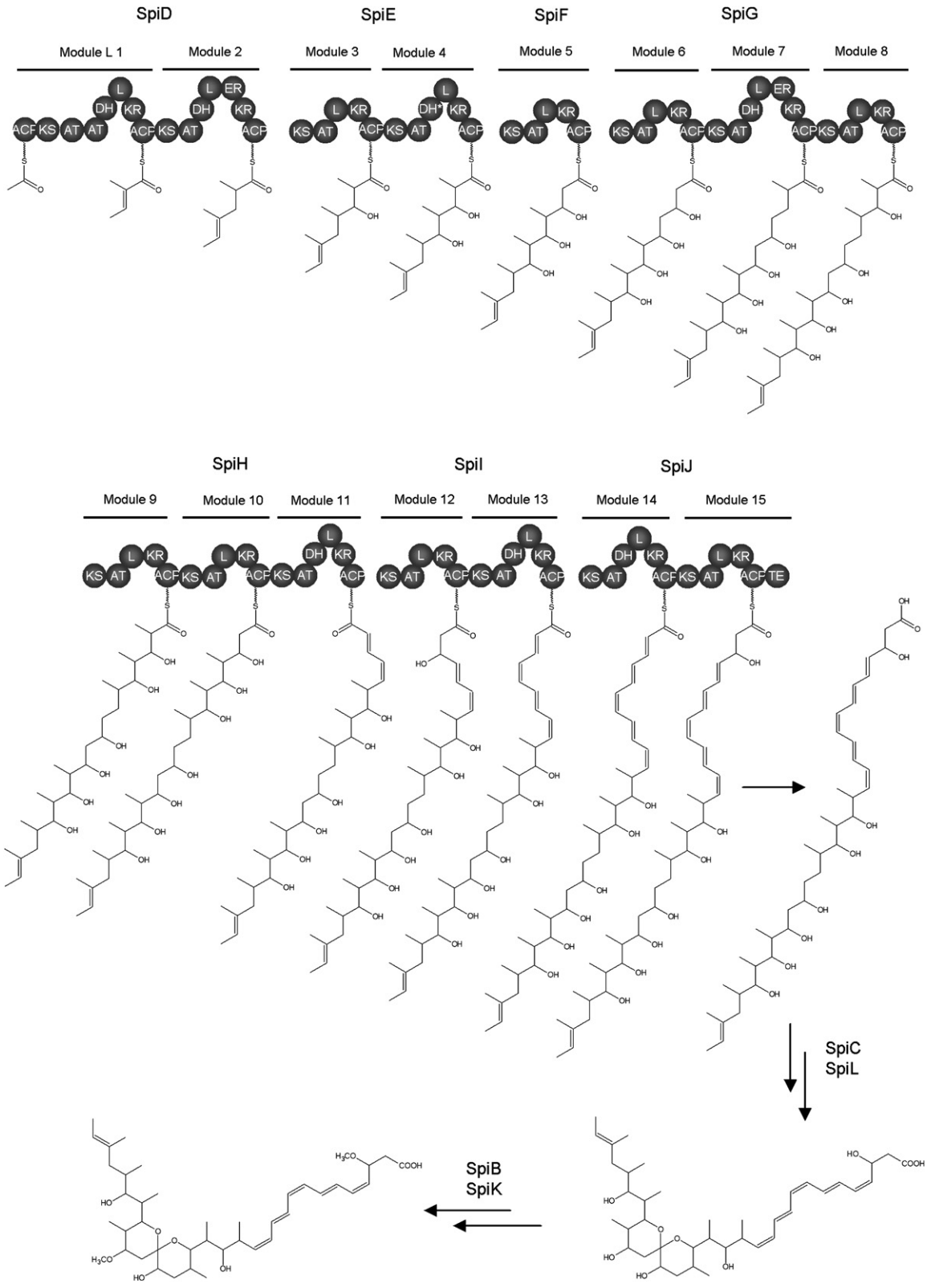

Figure 2. Model of Spirangien A Biosynthesis

Biosynthesis of spirangien B is supposed to occur analogously, but starting with ACP_L-bound propionate. The DH domain of module 4 is presumably inactive. AT-KR/DH-KR/DH-ER linker regions are inscripted "L." For discussion and further abbreviations, see the text.

differ from the other DH domains present in the PKS (Figure S2). Such a scenario (of a DH domain acting externally of the module it is located in) was also proposed for stigmatellin biosynthesis in the myxobacterium *Stigmatella aurantiaca* [18], as well as for chivosazol biosynthesis in *S. cellulosum* So ce56 [27]. In the chivosazol PKS, the modules bearing the putative iterative DHs display an unusual domain organization, which is not observed in the spirangien PKS. Experimental evidence to support an iteratively acting DH domain has been provided by Tang et al. for the DH of module 5 from the epothilone mixed PKS-NRPS pathway [28]. Although the DH domains of modules 11 and 13 appear to operate iteratively, the stereochemistry of double bonds generated in the two cycles differs (*Z* then *E*), consistent with the idea that the configuration of the hydroxyl group determines the stereochemistry of the dehydration reaction. However, there are *cis/trans* isomers of spirangienes A and B present in the culture broth (B.F. and R.M., unpublished data). Although it cannot be excluded that respective isomerizations occur during the work up procedure, this finding indicates again [27] that the classification of KR domains as A or B type cannot serve as an absolute predictor for double bond stereochemistry.

The DH domain of module 4 contains the typical conserved active site motif. However, the corresponding hydroxyl group appears to be retained in the structure. It is likely to be involved in spiroketal formation, because a knockout mutant of the P_{450}-dependent monooxygenase gene *spiL* produces acyclic spirangienes with a (partially glycosylated) hydroxyl group at the corresponding carbon (C-25) (Figure 1C). Point mutations adjacent to the core amino acids of the module 4 DH domain (Figure S2) may be responsible for this loss of function. The last module of the spirangien assembly line (module 15) terminates in a TE domain. The TE contains the conserved GXSXG motif characteristic of this family of enzymes [29] and most likely catalyzes the hydrolytic release of the fully processed polyketide chain from the multienzyme complex, yielding the free acid.

All modules of the spirangien PKS harbor long linker regions at the N terminus of the KR domains or, if present, ER domains (Figure 2). These domains span ~380 aa, share a number of conserved residues and motifs (as G^{302}LxRxxxxE and E^{352}xxxAL/VR), and can be clearly divided into two groups based on sequence comparison. One group comprises the linker regions of modules 1, 2, 4, 7, 11, 13, and 15, and the other group comprises the linker regions of modules 3, 5, 6, 8, 9, 10, 12, and 14. However, the grouping of the domains reveals no obvious biosynthetic logic. These kinds of DH-KR/DH-ER linker regions are frequently found in bacterial PKSs and are believed to contribute to intramolecular (and possibly also intermolecular) protein-protein interaction [30]. The structure of the KR domain including the linker region from the first module of the erythromycin PKS has been solved recently [30] and suggests a structural role for the linker region by stabilizing the catalytic part of the KR domain.

Tailoring Reactions in Spirangien Biosynthesis

The biosynthesis of the polyketide scaffold by the PKS assembly line is followed by a series of tailoring reactions. Spirangien displays four functional groups that are unlikely to derive from PKS reactions: two methoxy groups in positions 3 and 23, a hydroxyl group in position 20 (C-20 originates from C-2 of the incorporated acetate unit), and the spiroketal function in the center of the molecule (Figure 1). Four enzymes encoded in regions that flank the PKS genes could be responsible for the introduction of these modifications: SpiB and SpiK show homology to *O*-methyltransferases and display the typical *S*-adenosylmethionine (SAM)-binding motif [31], whereas SpiC and SpiL resemble cytochrome P_{450} monooxygenases. To assign the exact function of these proteins in spirangien biosynthesis, the corresponding genes were inactivated separately by site-directed mutagenesis. Manipulation of *Sorangium* strains is difficult to achieve, as only limited genetic tools are available [12]. The most commonly used selection markers are not applicable to *S. cellulosum* So ce90 because of its broad resistance spectrum. In addition, to date no self-replicating units have been identified for myxobacteria. To achieve gene inactivation, pSUPHyg-based constructs harboring a fragment for homologous recombination and a hygromycin-resistance gene [32, 33] were introduced into *S. cellulosum* So ce90 via biparental mating. Correct integration into the chromosome was verified by Southern blot analysis, and the mutants were evaluated for secondary metabolite production after cultivation in the presence of XAD adsorber resin. A summary of the mutants generated in this work and their corresponding phenotypes is given in Table 2.

The inactivation of both putative methyltransferase genes, *spiB* and *spiK*, resulted in the loss of production of spirangienes A and B. However, new derivatives were produced with molecular weights that were 14 mass units below that of the wild-type metabolites (Figure 3), indicating that each methyltransferase is responsible for the generation of one of the methoxy groups at positions 3 and 23. The level of production of spirangienes in the wild-type and derivatives in the mutants was comparable. In order to assign each methyltransferase to a specific methylation event, feeding experiments were performed with L-[*methyl*-^{13}C]methionine. Spirangien-enriched fractions obtained from these experiments were analyzed by ^{13}C NMR spectroscopy and were compared to the corresponding wild-type fractions. As both acetyl-CoA and propionyl-CoA serve as starter units in spirangien biosynthesis and additional isomers with variations in double bond configurations are present (B.F. and R.M., unpublished data), signals corresponding to the 3-OMe group were present between 56.7 and 57.0 ppm in the ^{13}C NMR spectrum of the wild-type spirangienes (Figure S4). As expected, signals for the 23-OMe and 3-OMe groups were observed in the wild-type sample from the feeding experiment, while the spectra of fractions derived from the mutants each lacked one of the signals. In the case of the *spiB* knockout mutant, the 23-OMe-derived signal was missing, whereas the sample from the *spiK* knockout

Table 2. Phenoytype of Mutants after Gene Inactivation

Mutant	Plasmid for Inactivation	Inactivated Gene (Deduced Function)	Phenotype
BFSA	pBFS_A	*spiA* (type II thioesterase)	spirangien production
BFSB	pBFS_B	*spiB* (methyltransferase)	production of 23-desmethyl spirangienes
BFSC	pBFS_C	*spiC* (cytochrome P_{450} monooxygenase)	no spirangien production
BFSCP	pBFS_CP	see text and Figure 4	spirangien production
JKB18	pJKB16	*spiH* (PKS: KS module 10)	no spirangien production
BFSK	pBFS_K	*spiK* (methyltransferase)	production of 3-desmethyl spirangienes
JKB19	pJKB29	*spiL* (cytochrome P_{450} monooxygenase)	production of acyclic spirangien derivatives

This work is the reference for all mutants.

mutant did not exhibit the 3-OMe-derived signal (Figure S4). Therefore, it can be concluded that SpiB is responsible for methylation of the C-23 OH group and that SpiK catalyzes methyl group transfer to the C-3 hydroxyl in spirangienes (Figures 1D and 1E). In addition to the sequence-based analysis, incorporation of the ^{13}C methyl group derived from methionine also provides direct evidence that both methyltransferases act in a SAM-dependent manner.

The timing of formation of the spiroketal in spirangien biosynthesis is currently unknown. The two cytochrome P_{450} monooxygenases SpiC and SpiL are most likely involved in the introduction of the hydroxyl group at carbon 20 and the formation of the neighboring spiroketal function, with C-21 as the central carbon atom. To decipher the function of SpiC and SpiL in spirangien biosynthesis and to elucidate the mechanism of spiroketal formation, site-directed mutagenesis was performed to generate knockout mutants of both SpiC and SpiL. The knockout mutant of SpiL (JKB19) failed to produce spirangienes A and B, but two new derivatives, produced at the wild-type level, were detected. Structure elucidation by NMR analysis after fermentation (*4* l) revealed new acyclic derivatives missing the hydroxyl group at carbon 20. One of the *seco* derivatives is glycosylated with a pentose at the 25-OH position, which is normally part of the spiroketal functionality in spirangienes A and B (Figures 1C and 5). According to NMR reference data from the literature, the sugar moiety is most likely a methylated α-D-arabinofuranoside [34]. To our knowledge, this compound represents the first example of a glycosylated metabolite from *S. cellulosum* So ce90. However, the same sugar moiety is also part of the metabolite icumazol B, produced by another *S. cellulosum* strain [35]. After inactivation of *spiC*, neither spirangien nor cyclic or acyclic derivatives could be detected.

SpiC is encoded 13 bp upstream of the PKS SpiD, indicating that the genes are transcriptionally linked (Figure 4A). To ensure transcription of the PKS-encoding genes in the knockout mutant, the *aphII* promoter [33] was joined to the fragment used for homologous recombination (Figure 4B). To exclude the possibility that

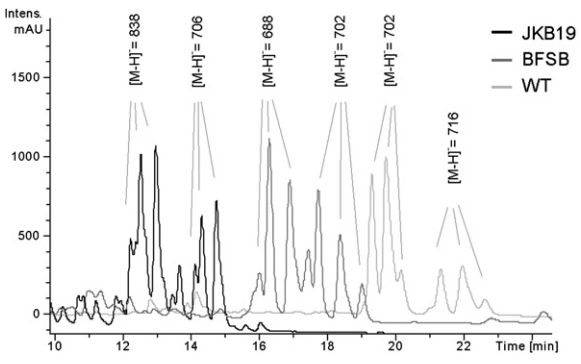

Figure 3. HPLC Chromatograms of Culture Extracts of Wild-Type *S. cellulosum* So ce90, the *spiL* Knockout Mutant JKB19, and the *spiB* Knockout Mutant BFSB11

Spirangiens A and B and derivatives were detected by UV absorption at 334 nm and mass analysis. The multiple peaks that correspond to spirangienes A and B and their derivatives originate from different *cis*/*trans* isomers (data not shown).

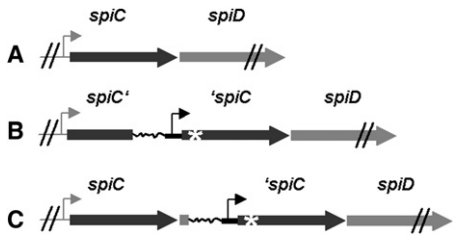

Figure 4. Organization of *spiC* and *spiD* in Wild-Type and Mutants of *S. cellulosum* So ce90

(A) Wild-type. The first PKS-encoding gene, *spiD*, starts 13 bp downstream of *spiC* and seems to be controlled by the same promoter (gray).

(B) The *spiC* kockout mutant BFSC. To ensure transcription of the PKS-encoding genes, the *aphII* promotor (black) was cloned 5′ of the homologous fragment for recombination in the *spiC* knockout mutant BFSC. The asterisk indicates an introduced stop codon, resulting in two inactive copies of *spiC*. The vector backbone sequence is indicated by a wiggly line.

(C) The complementation mutant BFSCP. The mutant BFSCP harbors an intact copy of *spiC* under the control of the natural promoter, and the PKS-encoding genes are under the control of the introduced *aphII* promoter.

transcription of the PKS genes from this artificial promotor was inefficient, a second construct for homologous recombination, which harbors additional 399 base pairs resembling the missing 3′ end of *spiC* (pBFS_CP), was designed. Integration of pBFS_CP results in a genotype with one intact copy of *spiC* controlled by the natural promoter, and a second inactive copy of *spiC* that is connected to *spiD* and controlled by P_{aphII} (Figure 4C). In the corresponding mutant BFSCP, spirangien production was restored, indicating that P_{aphII} promotes sufficient transcription of the PKS-encoding genes (an overview of plasmids, mutants, and phenotypes is given in Table 2). Thus, it can be concluded that SpiC is essential for spirangien biosynthesis.

In principle, cytochrome P_{450}-catalyzed oxidation and spiroketal formation could occur during chain assembly on the PKS, or alternatively after the intermediate is released from the multienzyme. Inactivation of SpiL leads to the production of an acyclic spirangien derivative that resembles a fully processed polyketide chain that has been released by the terminal TE domain of the PKS (and further decorated with both methoxy groups). The glycosylated derivative presumably arises because of the promiscuous activity of a glycosyl transferase belonging to some other biochemical pathway. Production of such acyclic intermediates by the SpiL knockout mutant supports the hypothesis that oxidative modification by SpiL and SpiC occurs after hydrolytic chain release from the multienzyme complex.

Spiroketal moieties are present in some other polyketide-derived natural products such as the avermectins and related compounds produced by *Streptomyces avermitilis* [36]. In these cases, the spiroketal function is thought to derive from nucleophilic attack of two hydroxyl groups on a central carbonyl carbon, resulting in the formation of a hemi-acetal that is then converted to the final acetal moiety. A different mechanism has been postulated for the formation of the spiroketal moieties in the polyketide monensin, produced by *Streptomyces cinnamonensis.* Recent findings indicate that after epoxidation of PKS-derived double bonds, two epoxide hydrolase-like enzymes, MonBI and MonBII, catalyze nucleophilic attack of a hemiacetal-derived hydroxy group on an epoxide carbon. This attack results in the formation of an acetal and concomitant ring opening of the epoxide to generate a free hydroxyl group, which can subsequently attack a second epoxide, resulting in the typical polyether structure found in monensin [37].

In the linear precursor to monensin and avermectin, a keto group, which ultimately gives rise to the spiroketal functionality, is present at the carbon center. However, based on sequence analysis, there is no indication that the KR domain of module 6 in the spirangien PKS is inactive. Thus, it is unlikely that a keto group is retained at C-21 during chain assembly; therefore, a similar mechanism of spiroketal formation to the avermectin hypothesis would have to involve the reintroduction of the keto group by a tailoring enzyme [36]. Although it cannot be excluded that spirangien biosynthesis involves an epoxidation

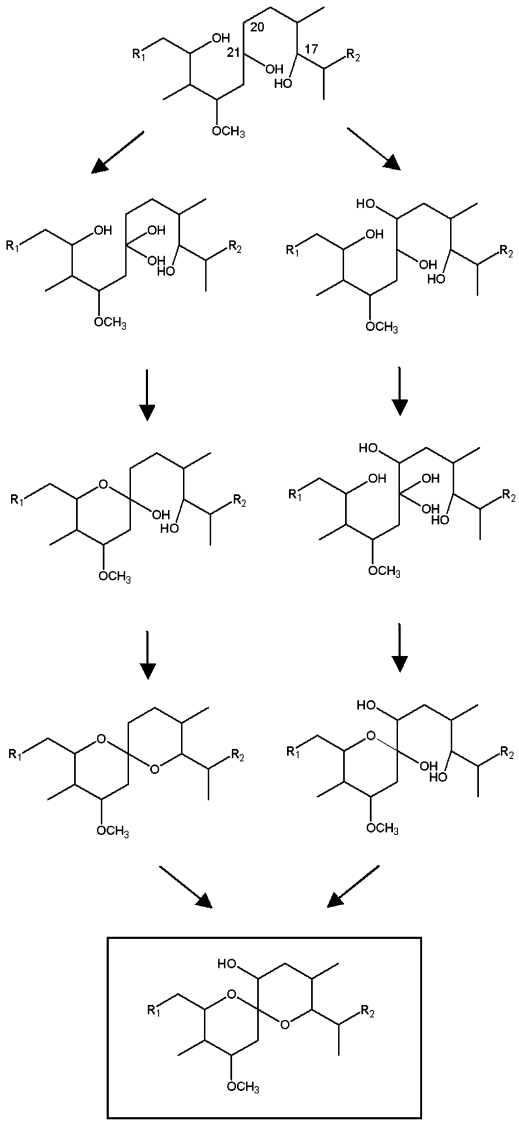

Figure 5. Hypothetical Pathways of Spiroketal Formation in Spirangien Biosynthesis, Starting with the Acyclic Derivative Produced by the *spiL* Mutant

The single reactions may be catalyzed by SpiL or SpiC or may occur spontaneously.

step comparable to monensin biosynthesis, SpiL and SpiC show no significant similarities to epoxidases and epoxide hydrolases, and no such enzymes are encoded near the spirangien biosynthetic gene locus.

Further investigation will be required to fully explain the formation of the spiroketal moiety, but based on the available evidence, we propose the following two alternative mechanisms. SpiL may act first by directly or indirectly reoxidizing the alcohol at C-21 to a keto group, followed by spontaneous or enzyme-catalyzed cyclization, giving rise to the spiroketal moiety (hypothetical reaction schemes are depicted in Figure 5, the ketogroup introduced at C-21 is shown in a hydrated form). In this

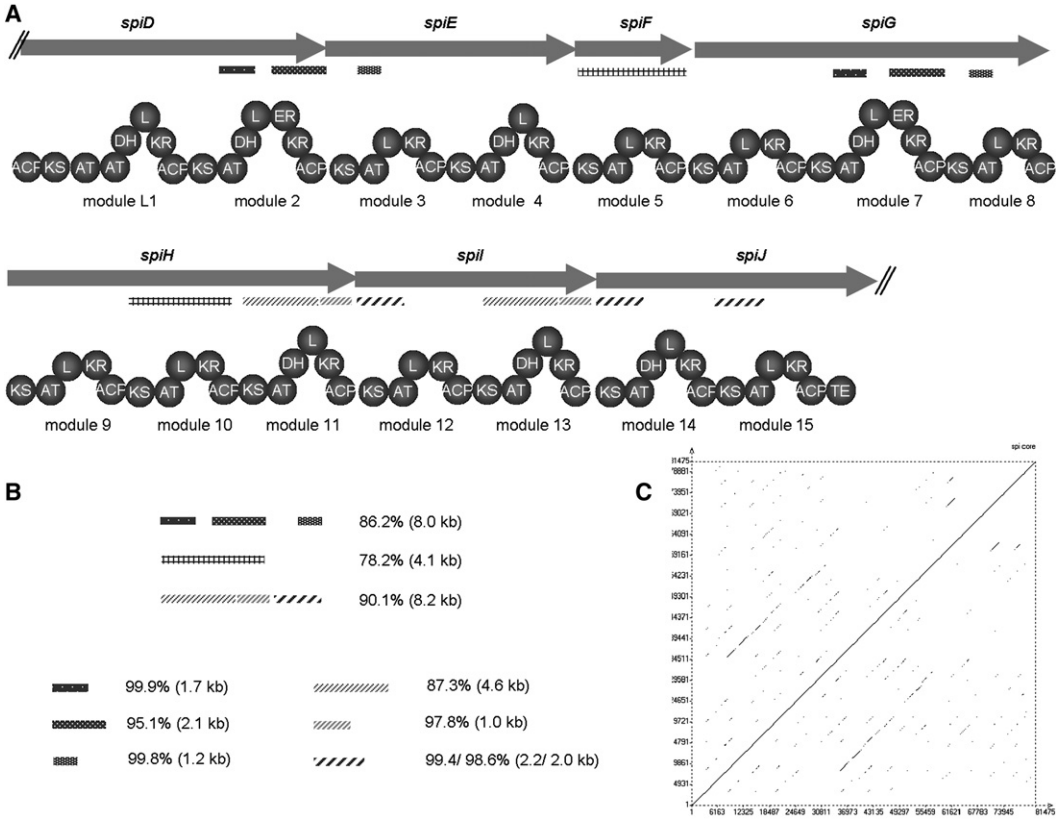

Figure 6. Sequence Repeats within the Spirangien Biosynthetic Gene Cluster

(A) Location of the repetitive DNA regions within the cluster in relation to the encoded domains.

(B) Degree of identity and length of the single repeats and stretches of repeats.

(C) Sequence repeats were identified by dot plot analysis of the PKS core cluster: the DNA sequence is compared with itself, and regions with a certain degree of similarity (>85%/30 bp) are pictured as stretches. The diagonal line shows the matching of every residue with itself and is therefore not of interest for the analysis of repeats.

scenario, SpiC would not be capable of introducing a second hydroxyl group to C-21; thus *spiL* inactivation would lead to the production of the acyclic spirangienes. Alternatively, SpiL may hydroxylate C-20 before SpiC performs oxidation/cyclization reactions. Here, inactivation of *spiL* would also lead to a substrate that cannot be processed by SpiC. It is also conceivable that SpiC acts first in spiroketal formation by either mechanism. Unfortunately, no product can be found in the *spiC* mutant that might clarify the type of modification introduced by SpiL or SpiC. The intermediate may be unstable and may break up into products that are readily degraded in the cells or after excretion into the medium. In fact, spirangienes themselves are already unstable compounds.

SpiA Is Not Essential for Spirangien Biosynthesis

The gene product of *spiA* shows homology to type II thioesterases, which are often found in PKS and NRPS biosynthetic gene clusters [38]. In contrast to type I thioesterases, which catalyze the release of the fully processed polyketide chain from the multienzyme complex, type II TEs are not thought to be directly involved in biosynthesis. Inactivation of these enzymes can lead to decreased product yields as in tylosin biosynthesis [39]; however, in other cases, no effect on product formation has been observed (e.g., for *pikAV* from the pikromycin cluster [40]). Inactivation of *spiA* also leads to no significant effect on spirangien production. Type II thioesterases have been proposed to play an editing role in polyketide pathways by removing aberrantly decarboxylated extender units from the ACPs that would otherwise block the biosynthesis [38], or by removing certain alternative starter units from the ACP of the loading module [41]. Another possible function of type II TEs is to prevent the accumulation of intracellular acyl-CoA [42]. Type II TEs involved in nonribosomal peptide biosynthesis have been characterized in more detail [43], and they were shown to regenerate misacylated NRPSs [44]. The exact function of SpiA, and type II TEs in polyketide biosynthesis more generally, remains unclear.

Sequence Repeats: Indications for the Origin of Spirangien Biosynthesis?

Dot plot analysis of the PKS portion of the spirangien biosynthetic gene cluster reveals the presence of seven large stretches of sequence that are repeated at least once in

the gene cluster (Figure 6). The twin sequences show up to 99.9% identity to each other on the nucleic acid level, and they vary in size from 1.0 to 4.6 kb. The different repeat regions are clustered in three stretches that are 8.0, 4.1, and 8.2 kb in size; the mutual sequence identity between the corresponding regions is 86.2%, 78.2%, and 90.1%, respectively (Figure 6B). The first 8.0 kb stretch of repeats is present at the junctions between both modules 2 and 3 and between modules 7 and 8, and it spans from AT to AT (Figure 6A). The second repeat region is spanning *spiF* completely plus the part of *spiH* coding for the module 10. The third stretch of repeats forms the junctions between modules 11 and 12 and between modules 13 and 14. The 2 kb of the 3′ region encode for a KS-AT didomain and is additionally found in module 15; the mutual sequence identity is 99.4% for modules 12 and 14 and 98.6% for modules 12 and 15.

Sequence repeats with nearly 100% identity on the nucleotide level have also been found in other biosynthetic gene clusters, but they are mostly restricted to single domains [45, 46]. Examples from myxobacteria include the myxothiazol and the epothilon biosynthetic gene clusters [13, 14, 19]. A remarkable and so far unique exception is the mycolactone biosynthetic gene cluster sequenced from *Mycobacterium ulcerans* [47], in which two giant PKSs, MLSA and MLSB, harbor nine and eight modules, respectively. All 17 modules represent 6 groups sharing >98% homology. The arrangement of the modules with respect to these groups appears to be almost random in the gene cluster, which poses the question of the substrate specificity of domains within this megasynthetase.

The strong conservation of PKS function is naturally mirrored in a certain degree of sequence homology. However, the extent of sequence repetition in the spirangien biosynthetic gene cluster is striking. Based on the location and size of these DNA repeats, one can speculate that the spirangien biosynthetic gene cluster has evolved through three major duplication events. An ancestral gene locus may have comprised 11 modules instead of the 16 modules present today. It can be assumed that sequence repeats with lower overall homology arose from earlier duplications, while stretches with very high similarity derive from more recent events. Strongly conserved regions within the repeat stretches are interrupted by less conserved regions that may be more tolerant to the accumulation of point mutations over generations. A first duplication event involving module 2 and portions of module 3 may have resulted in the corresponding regions of modules 7 and 8, while duplication of module 5 may have resulted in module 10. The most recent evolutionary duplication of the module 11/12 junction may have given rise to module 13 and the N terminus of module 14, and then subsequently to the N-terminal portion of module 15 (Figure 6). Recombination events involving deletions of whole modules have been described for the polyene polyketide nystatin from *Streptomyces sp*. and have been postulated for the cyclic pentapeptide nodularin from cyanobacteria species [48, 49]. In these cases, the evolutionarily related biosynthetic gene clusters were identified, and their products were characterized. Spirangienes have been detected in a number of *S. cellulosum* strains [10], and resequencing of the corresponding biosynthetic gene clusters as well as screening for putative ancestral spirangienes of shorter lengths should give further insight into the development of the present biosynthetic gene cluster.

These examples highlight how rearrangements within the modular megasynthetase systems can lead to the development of new secondary metabolites that may offer evolutionary benefits for the producer. A second way to expand the biosynthetic potential of a species is the horizontal gene transfer of secondary metabolic genes among related but also divergent taxa [50, 51]. An indication for horizontal gene transfer is the presence of transposase-like sequences adjacent to bacterial biosynthetic gene loci [50]. Remnants of such a transposase-encoding gene are also located in the 5′ region of the spirangien biosynthetic gene locus (Figure 1A). In the case of spirangien, Nature has apparently used both horizontal transfer and gene duplication to evolve the present biosynthetic cluster.

SIGNIFICANCE

The spirangienes produced by *S. cellulosum* So ce90 are highly cytotoxic and antifungal natural products. The analysis of the biosynthetic gene cluster provides important insights into the corresponding and, to our knowledge, novel biosynthetic pathway by which these metabolites are formed. Direct evidence is given that two cytochrome P_{450} monooxygenases are involved in the formation of the central spiroketal moiety. Genetic engineering of the biosynthetic gene locus resulted in the production of novel, to our knowledge, acyclic analogs as well as two different desmethyl derivatives. Large stretches of DNA sequence repeats have been identified within the gene cluster. Taken together with the presence of remnants of a transposase-encoding gene, this finding provides some explanation for the way the spirangien biosynthetic gene cluster has developed during evolution. The detailed analysis of this biosynthetic gene locus also has relevance for understanding more general aspects of PKS mechanism and evolution. Such analyses still continue to reveal new biosynthetic aspects, despite the great progress that has been made in the last two decades. Combinatorial biosynthesis has been practiced by Nature ever since, making use of deletions, inactivations, insertions, and duplications for the generation of new and potent compounds. Understanding in detail how this is achieved during the course of evolution will hopefully allow us to better harness combinatorial biosynthesis in the laboratory for use in drug discovery.

EXPERIMENTAL PROCEDURES

Strains and Culture Conditions

S. cellulosum So ce90 was grown at 30°C with 170 rpm in liquid G51t medium [52]. After conjugation on P agar plates [11], cells were plated

on PM agar plates. PM agar plates consist of 0.05% caseine peptone (Marcor), 0.15% $MgSO_4 \times 7\ H_20$, 50 mM HEPES (Serva), 1.5% agar (Difco) (pH adjusted to 7.4 with KOH). After autoclaving, 10 ml/l sterile solutions of 5% NH_4SO_4, 10% $CaCl_2 \times 2\ H_2O$, 0.08% Fe-EDTA (Fluka), 0.625% K_2HPO_4, glucose (35%), 1.04% Na-dithionite (Merck), and 10% autoclaved liquid culture of *S. cellulosum* So ce90 were added. For the analysis of secondary metabolites, *S. cellulosum* So ce90 was cultivated in liquid E medium [53]. If required, the media were supplemented with 150 μg/ml hygromycin B and 120 μg/ml tobramycin.

E. coli DH10B, *E. coli* SURE, and *E. coli* ET12567/pUZ8002 were grown in LB medium at 37°C. *E. coli* strains containing cosmids were grown at 30°C. Antibiotics were used at the following concentrations: 100 μg/ml hygromycin B, 50 μg/ml kanamycin sulfate, 50 μg/ml ampicillin.

DNA Isolation, Manipulation, Analysis, and PCR

All restriction enzymes were purchased from MBI Fermentas. Genomic DNA of *S. cellulosum* So ce90 was isolated by using the Puregene Genomic DNA Purification Kit (Gentra) according to the manufacturer's protocol. Southern Blot analysis of genomic DNA was performed by following the standard protocol of the DIG DNA labeling and detection kit (Roche Diagnostics). Plasmid DNA purification was performed by using the NucleoSpin Plasmid Kit (Macherey-Nagel). Polymerase chain reaction (PCR) was carried out by using *Taq* DNA-Polymerase (MBI Fermentas). DMSO and glycerol were added to the reaction mixture to final concentrations of 5% and 2.5%, respectively. Conditions for amplification with an Eppendorf Mastercycler gradient thermal cycler were as follows: denaturation, 30 s at 95°C; annealing, 30 s at 55–63°C; extension, 50–90 s at 72°C for 30 cycles, and a final extension for 10 min at 72°C. PCR products were purified with the NucleoSpin Plasmid Kit (Macherey-Nagel).

PCR products generated for the construction of inactivation plasmids were verified by sequencing. The cosmid library was made by using partially digested (Sau3AI) genomic DNA of *S. cellulosum* So ce90, which was size fractionated with 3% agarose gels and cloned into the predigested (BamHI-XbaI) SuperCos vector (Stratagene). Ligations were packaged by using the Gigapack III Gold Packaging extract (Stratagene) and were transfected into *E. coli* SURE (Stratagene) according to the manufacturer's protocol.

All other DNA manipulations were performed according to standard protocols [54]. DNA and amino acid sequence analysis was carried out by using the VectorNTI software package (Invitrogen) and the Lasergene software package (DNASTAR, Inc.). Comparison with GenBank data was performed with BLAST.

For detailed information about the construction of plasmids, see Table S1.

Screening of the Cosmid Library

Screening of the cosmid library was performed as previously described [17].

Gene Inactivation in *Sorangium cellulosum* So ce90

Constructs for gene inactivation were based on pSUPHyg containing a homologous fragment for recombination (see Supplemental Data) and were conjugated into *S. cellulosum* So ce90 as described previously [52]. Transformants were selected on PM agar plates containing hygromycin B and tobramycin, and single colonies became visible after 10–14 days. Mutants were analyzed for the correct integration of the plasmids by Southern blot hybridization, after digestion with appropriate restriction endonucleases.

Analysis of Secondary Metabolite Production

S. cellulosum So ce90 and mutants were grown in E medium with 1% XAD adsorber resin (Rohmer and Haas) for at least 14 days. XAD beads and cells were harvested by centrifugation and extracted successively with acetone and methanol. The combined extracts were evaporated and redissolved in methanol, resulting in a 100-fold concentration of the original culture volume. Extracts were analyzed by using HPLC-MS. Separation was carried out with an Agilent 1100 series system equipped with a photodiode array detector and coupled to a Bruker HCT plus mass spectrometer operating in negative ionization mode at a scan range from m/z = 100–1100. A 125 × 2 mm Nucleodur C18/3 μm RP column (Macherey-Nagel) was used for separation with a solvent system consisting of H_2O (A) and acetonitrile (B), each containing 0.1% formic acid. The following gradient was applied: 0–2 min 35% B, 2–15 min linear from 35% B to 75% B, 15–25 min isocratic at 75% B, 25–30 min linear from 75% B to 95% B, 30–33 min isocratic at 95% B. Spirangienes were identified by MS analysis (spirangien A and *cis/trans* isomers: R_t 19.5–20.4 min, $[M-H]^-$ = 701.5; spirangien B and *cis/trans* isomers: R_t 21.6–22.8 min, $[M-H]^-$ = 715.5).

Feeding of Labeled Precursors, Purification, and NMR Analysis of Desmethyl Derivatives

The mutants BFSB, BFSK, and the wild-type were grown in 500 ml E medium with 1% XAD supplemented with L-[*methyl*-^{13}C]methionine (1 mM final concentration), added in three portions at 4, 7, and 9 days. XAD beads and cells were harvested by centrifugation and extracted with methanol. After evaporation of the organic solvent, the remaining aqueous phase was extracted with ethyl acetate. The organic phase was evaporated, and the remaining oily residue was redissolved in methanol and extracted with heptane. The methanolic phase was concentrated and chromatographed with methanol on Sephadex LH 20 (Amersham Biosciences). Spirangien-containing fractions were combined, evaporated, and analyzed by NMR. NMR spectra were recorded at 125.7 MHz on a Bruker Advance 500 with CD_3OD as solvent and internal standard (see Figure S4). ^{13}C enrichments of 3-OMe and 23-OMe after feeding of L-[*methyl*-^{13}C]methionine were calculated in comparison to natural abundance. The enrichments of 3-OMe and 23-OMe were found to be 5.5-fold and 8.4-fold, respectively.

Isolation and NMR Analysis of *Seco* Spirangienes

A shaking flask fermentation (4 l) of JKB19 cultivated in the presence of 100 ml adsorber resin XAD-16 (Rohm and Haas) was passed through a sieve to recover the resin. The resin was washed with water, transferred to a glass column, and eluted with 800 ml methanol. The extract was evaporated to dryness to yield 3.6 g of a crude product. The residue was diluted with water and extracted three times with ethyl acetate. The combined organic layers were dried with sodium sulfate and evaporated in vacuo to yield 2.4 g of an oily residue. The resulting extract was partitioned between methanol and *n*-heptane to give an extract (1.59 g) that was separated by chromatography on Sephadex LH-20 (Fluka) with dichloromethane:methanol at a ratio of 8:2 as eluent (column, 7 × 71 cm; flow, 3.5 ml/min; UV detection, 313 nm). The spirangien-containing fraction (0.66 g) was further separated by MPLC (ODS AQ C18 16 μm, 3 × 48 cm, Kronlab) (flow, 13 ml/min; methanol:50 mM ammonium acetate buffer [pH 6.5], 75:25; UV detection: 313 nm). A further fine separation (C18 7 μm, 250 × 21 mm, Nucleosil) (flow, 12 ml/min; acetonitrile:50 mM ammonium acetate buffer [pH 5.0], 58:42; UV detection, 313 nm) gave 3.1 mg *seco* spirangien A and 6.0 mg *seco* spirangien glycoside.

UV, Shimadzu UV2102 PC UV/VIS scanning spectrometer; solvent, methanol (Merck); ESI-HPLC-MS, PE Siex API 2000 LC-MS; mass spectrometer, Micromass ESI QTOF; NMR, Bruker ARX 600 spectrometer (^{1}H: 600 MHz and ^{13}C: 150 MHz).

Structural Data on *Seco* Spirangienes

Seco Spirangien A

R_t = 18.2 min (analytical HPLC). UV (MeOH): λ_{max} (lgε) = 308 (4.21), 320 (4.50), 336 (4.67), 354 (4.62). ^{1}H and ^{13}C NMR: see Table S2. ESI MS: m/z (%): M+H$^+$ 707 (100%), M−H$^+$ 705 (100%). HRMS (ESI) for $C_{41}H_{70}O_9$ $[M+Na]^+$ calc'd. 729.4917; found 729.4935.

Seco 25 O-Spirangien A Furanoside

R_t = 17.0 min (analytical HPLC). UV data are identical to those of *seco* spirangien A. ^{1}H and ^{13}C NMR (only data differing from those for *seco* spirangien A are given): ^{1}H NMR: δ = 1.60–1.57a, 1.60–1.57b (22-H_2); 3.59 (23-H); 2.11 (24 H); 3.74 (25 H); 1.94 (26-H) ddq 7.2, 6.8, 13.9;

5.10 (1′-H) d 1.9; 4.05 (2′-H) m; 3.92 (3′-H) dd 5.7, 5.7; 4.07 (4′-H) m; 3.75 (5′-Ha), m; 3.66 (5′-Hb) dd 5.3, 11.7 ppm. ^{13}C NMR: δ = 38.1 (C-22), 79.4 (C-23), 38.0 (C-24), 82.5 (C-25), 36.9 (C-26), 110.0 (C-1′), 82.2 (C-2′), 77.1 (C-3′), 85.4 (C-4′), 62.0 (C-5′). ESI MS m/z (%): $M+NH_4^+$ 856.2 (100%), $M-H^+$ 836.9 (100%). HRMS (ESI) for $C_{46}H_{78}O_{13}$ $[M+Na]^+$: calc'd. 861.5340; found 861.5387.

Supplemental Data

Supplemental Data include a phylogenetic tree of AT domains (Figure S1), an alignment of the DH domain core regions (Figure S2), a phylogenetic tree of DH domains (Figure S3), ^{13}C NMR signals for methoxy groups in the spirangienes (Figure S4), details on the construction of plasmids (Table S1), and ^{13}C and ^{1}H NMR data of seco spirangien (Table S2) and are available at http://www.chembiol.com/cgi/content/full/14/2/221/DC1/.

ACKNOWLEDGMENTS

We thank K. Gerth for help in cultivating *S. cellulosum* So ce90 and HPLC analysis and S.C. Wenzel, H.B. Bode, and K.J. Weissman for helpful comments regarding this manuscript. Research in R.M.'s laboratory was supported by grants from the German Bundesministerium für Bildung und Forschung and the Deutsche Forschungsgemeinschaft.

Received: October 18, 2006
Revised: November 20, 2006
Accepted: November 27, 2006
Published: February 23, 2007

REFERENCES

1. Newman, D., Cragg, G., and Snader, K. (2003). Natural products as sources of new drugs over the period 1981–2002. J. Nat. Prod. *66*, 1022–1037.
2. Staunton, J., and Weissman, K.J. (2001). Polyketide biosynthesis: a millennium review. Nat. Prod. Rep. *18*, 380–416.
3. Weissman, K.J., and Leadlay, P.F. (2005). Combinatorial biosynthesis of reduced polyketides. Nat. Rev. Microbiol. *3*, 925–936.
4. Rix, U., Fischer, C., Remsing, L.L., and Rohr, J. (2002). Modification of post-PKS tailoring steps through combinatorial biosynthesis. Nat. Prod. Rep. *19*, 542–580.
5. Khosla, C., and Tang, Y. (2005). A new route to designer antibiotics. Science *308*, 367–368.
6. Clardy, J., and Walsh, C. (2004). Lessons from natural molecules. Nature *432*, 829–837.
7. Moss, S.J., Martin, C.J., and Wilkinson, B. (2004). Loss of co-linearity by modular polyketide synthases: a mechanism for the evolution of chemical diversity. Nat. Prod. Rep. *21*, 575–593.
8. Wenzel, S.C., and Müller, R. (2005). Formation of novel secondary metabolites by bacterial multimodular assembly lines: deviations from text book biosynthetic logic. Curr. Opin. Chem. Biol. *9*, 447–458.
9. Bode, H.B., and Müller, R. (2005). The impact of bacterial genomics on natural product research. Angew. Chem. Int. Ed. *44*, 6828–6846.
10. Gerth, K., Pradella, S., Perlova, O., Beyer, S., and Müller, R. (2003). Myxobacteria: proficient producers of novel natural products with various biological activities—past and future biotechnological aspects with the focus on the genus *Sorangium*. J. Biotechnol. *106*, 233–253.
11. Pradella, S., Hans, A., Sproer, C., Reichenbach, H., Gerth, K., and Beyer, S. (2002). Characterisation, genome size and genetic manipulation of the myxobacterium *Sorangium cellulosum* So ce56. Arch. Microbiol. *178*, 484–492.
12. Kopp, M., Irschik, H., Gross, F., Perlova, O., Sandmann, A., Gerth, K., and Müller, R. (2004). Critical variations of conjugational DNA transfer into secondary metabolite multiproducing *Sorangium cellulosum* strains So ce12 and So ce56: development of a mariner-based transposon mutagenesis system. J. Biotechnol. *107*, 29–40.
13. Molnar, I., Schupp, T., Ono, M., Zirkle, R., Milnamow, M., Nowak-Thompson, B., Engel, N., Toupet, C., Stratmann, A., Cyr, D.D., et al. (2000). The biosynthetic gene cluster for the microtubule-stabilizing agents epothilones A and B from *Sorangium cellulosum* So ce90. Chem. Biol. *7*, 97–109.
14. Julien, B., Shah, S., Ziermann, R., Goldman, R., Katz, L., and Khosla, C. (2000). Isolation and characterization of the epothilone biosynthetic gene cluster from *Sorangium cellulosum*. Gene *249*, 153–160.
15. Höfle, G., and Reichenbach, H. (2005). Epothilone, a myxobacterial metabolite with promising antitumor activity. In Anticancer Agents from Natural Products, G.M. Cragg, D.G. Kingston, and D.J. Newman, eds. (Boca Raton, FL: Taylor & Francis), pp. 413–450.
16. Niggemann, J., Bedorf, N., Flörke, U., Steinmetz, H., Gerth, K., Reichenbach, H., and Höfle, G. (2005). Spirangien A and B, highly cytotoxic and antifungal spiroketals from the myxobacterium *Sorangium cellulosum*: isolation, structure elucidation and chemical modifications. Eur. J. Org. Chem. *2005*, 5013–5018.
17. Beyer, S., Kunze, B., Silakowski, B., and Müller, R. (1999). Metabolic diversity in myxobacteria: identification of the myxalamid and the stigmatellin biosynthetic gene cluster of *Stigmatella aurantiaca* Sg a15 and a combined polyketide-(poly)peptide gene cluster from the epothilone producing strain *Sorangium cellulosum* So ce90. Biochim. Biophys. Acta *1445*, 185–195.
18. Gaitatzis, N., Silakowski, B., Kunze, B., Nordsiek, G., Blöcker, H., Höfle, G., and Müller, R. (2002). The biosynthesis of the aromatic myxobacterial electron transport inhibitor stigmatellin is directed by a novel type of modular polyketide synthase. J. Biol. Chem. *277*, 13082–13090.
19. Silakowski, B., Schairer, H.U., Ehret, H., Kunze, B., Weinig, S., Nordsiek, G., Brandt, P., Blocker, H., Höfle, G., Beyer, S., and Müller, R. (1999). New lessons for combinatorial biosynthesis from myxobacteria: the myxothiazol biosynthetic gene cluster of *Stigmatella aurantiaca* DW4/3–1. J. Biol. Chem. *274*, 37391–37399.
20. Ligon, J., Hill, S., Beck, J., Zirkle, R., Monar, I., Zawodny, J., Money, S., and Schupp, T. (2002). Characterization of the biosynthetic gene cluster for the antifungal polyketide soraphen A from *Sorangium cellulosum* So ce26. Gene *285*, 257–267.
21. Yadav, G., Gokhale, R.S., and Mohanty, D. (2003). Computational approach for prediction of domain organization and substrate specificity of modular polyketide synthases. J. Mol. Biol. *328*, 335–363.
22. Tang, L., Yoon, Y.J., Choi, C.Y., and Hutchinson, C.R. (1998). Characterization of the enzymatic domains in the modular polyketide synthase involved in rifamycin B biosynthesis by *Amycolatopsis mediterranei*. Gene *216*, 255–265.
23. Reid, R., Piagentini, M., Rodriguez, E., Ashley, G., Viswanathan, N., Carney, J., Santi, D.V., Hutchinson, C.R., and McDaniel, R. (2003). A model of structure and catalysis for ketoreductase domains in modular polyketide synthases. Biochemistry *42*, 72–79.
24. Caffrey, P. (2003). Conserved amino acid residues correlating with ketoreductase stereospecificity in modular polyketide synthases. ChemBioChem *4*, 654–657.
25. Baerga-Ortiz, A., Popovic, B., Siskos, A.P., O'Hare, H.M., Spiteller, D., Williams, M.G., Campillo, N., Spencer, J.B., and Leadlay, P.F. (2006). Directed mutagenesis alters the stereochemistry of catalysis by isolated ketoreductase domains from the erythromycin polyketide synthase. Chem. Biol. *13*, 277–285.

26. Donadio, S., and Katz, L. (1992). Organization of the enzymatic domains in the multifunctional polyketide synthase involved in erythromycin formation in *Saccharopolyspora erythraea*. Gene *111*, 51–60.
27. Perlova, O., Gerth, K., Hans, A., Kaiser, O., and Müller, R. (2006). Identification and analysis of the chivosazol biosynthetic gene cluster from the myxobacterial model strain *Sorangium cellulosum* So ce56. J. Biotechnol. *121*, 174–191.
28. Tang, L., Ward, S., Chung, L., Carney, J.R., Li, Y., Reid, R., and Katz, L. (2004). Elucidating the mechanism of *cis* double bond formation in epothilone biosynthesis. J. Am. Chem. Soc. *126*, 46–47.
29. Konz, D., and Marahiel, M.A. (1999). How do peptide synthetases generate structural diversity? Chem. Biol. *6*, R39–R48.
30. Keatinge-Clay, A.T., and Stroud, R.M. (2006). The structure of a ketoreductase determines the organization of the β-carbon processing enzymes of modular polyketide synthases. Structure *14*, 737–748.
31. Kagan, R.M., and Clarke, S. (1994). Widespread occurrence of three sequence motifs in diverse *S*-adenosylmethionine-dependent methyltransferases suggest a common structure for these enzymes. Arch. Biochem. Biophys. *310*, 417–427.
32. Simon, R., O'Connell, M., Labes, M., and Pühler, A. (1986). Plasmid vectors for the genetic analysis and manipulation of *Rhizobia* and other Gram negative bacteria. Methods Enzymol. *118*, 643–659.
33. Blondelet-Rouault, M.H., Weiser, J., Lebrihi, A., Branny, P., and Pernodet, J.L. (1997). Antibiotic resistance gene cassettes derived from the omega interposon for use in *E. coli* and Streptomyces. Gene *190*, 315–317.
34. Breitmaier, E., and Voelter, W. (1987). Carbon-13 NMR Spectroscopy, Third Edition (Weinheim, Germany: VCH Verlag).
35. Reichenbach, H., Höfle, G., Böhlendorf, B., and Irschik, H. German patent DE4305486. August 1994.
36. Ikeda, H., and Omura, S. (1997). Avermectin biosynthesis. Chem. Rev. *97*, 2591–2609.
37. Gallimore, A.R., Stark, C.B., Bhatt, A., Harvey, B.M., Demydchuk, Y., Bolanos-Garcia, V., Fowler, D.J., Staunton, J., Leadlay, P.F., and Spencer, J.B. (2006). Evidence for the role of the *monB* genes in polyether ring formation during monensin biosynthesis. Chem. Biol. *13*, 453–460.
38. 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.
39. Butler, A.R., Bate, N., and Cundliffe, E. (1999). Impact of thioesterase activity on tylosin biosynthesis in *Streptomyces fradiae*. Chem. Biol. *6*, 287–292.
40. Chen, S., Roberts, J.B., Xue, Y., Sherman, D.H., and Reynolds, K.A. (2001). The *Streptomyces venezuelae pikAV* gene contains a transcription unit essential for expression of enzymes involved in glycosylation of narbonolide and 10-deoxymethynolide. Gene *263*, 255–264.
41. Hu, Z., Pfeifer, B.A., Chao, E., Murli, S., Kealey, J., Carney, J.R., Ashley, G., Khosla, C., and Hutchinson, C.R. (2003). A specific role of the *Saccharopolyspora erythraea* thioesterase II gene in the function of modular polyketide synthases. Microbiology *149*, 2213–2225.
42. Zheng, Z., Gong, Q., Liu, T., Deng, Y., Chen, J.C., and Chen, G.Q. (2004). Thioesterase II of *Escherichia coli* plays an important role in 3-hydroxydecanoic acid production. Appl. Environ. Microbiol. *70*, 3807–3813.
43. Linne, U., Schwarzer, D., Schroeder, G.N., and Marahiel, M.A. (2004). Mutational analysis of a type II thioesterase associated with nonribosomal peptide synthesis. Eur. J. Biochem. *271*, 1536–1545.
44. 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.
45. Aparicio, J.F., Molnar, I., Schwecke, T., Konig, 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.
46. Brautaset, T., Sekurova, O.N., Sletta, H., Ellingsen, T.E., Strm, A.R., Valla, S., and Zotchev, S.B. (2000). Biosynthesis of the polyene antifungal antibiotic nystatin in *Streptomyces noursei* ATCC 11455: analysis of the gene cluster and deduction of the biosynthetic pathway. Chem. Biol. *7*, 395–403.
47. Stinear, T.P., Mve-Obiang, A., Small, P.L., Frigui, W., Pryor, M.J., Brosch, R., Jenkin, G.A., Johnson, P.D., Davies, J.K., Lee, R.E., et al. (2004). Giant plasmid-encoded polyketide synthases produce the macrolide toxin of *Mycobacterium ulcerans*. Proc. Natl. Acad. Sci. USA *101*, 1345–1349.
48. Brautaset, T., Bruheim, P., Sletta, H., Hagen, L., Ellingsen, T.E., Strom, A.R., Valla, S., and Zotchev, S.B. (2002). Hexaene derivatives of nystatin produced as a result of an induced rearrangement within the *nysC* polyketide synthase gene in *S. noursei* ATCC 11455. Chem. Biol. *9*, 367–373.
49. Moffitt, M.C., and Neilan, B.A. (2004). Characterization of the nodularin synthetase gene cluster and proposed theory of the evolution of cyanobacterial hepatotoxins. Appl. Environ. Microbiol. *70*, 6353–6362.
50. Lopez, J.V. (2003). Naturally mosaic operons for secondary metabolite biosynthesis: variability and putative horizontal transfer of discrete catalytic domains of the epothilone polyketide synthase locus. Mol. Genet. Genomics *270*, 420–431.
51. Jenke-Kodama, H., Sandmann, A., Müller, R., and Dittmann, E. (2005). Evolutionary implications of bacterial polyketide synthases. Mol. Biol. Evol. *22*, 2027–2039.
52. Jaoua, S., Neff, S., and Schupp, T. (1992). Transfer of mobilizable plasmids to *Sorangium cellulosum* and evidence for their integration into the chromosome. Plasmid *28*, 157–165.
53. Gerth, K., Steinmetz, H., Höfle, G., and Reichenbach, H. (2001). Studies on the biosynthesis of epothilones: the PKS and epothilone C/D monooxygenase. J. Antibiot. (Tokyo) *54*, 144–148.
54. Sambrook, J., and Russell, D.W. (2001). Molecular Cloning: A Laboratory Manual (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Accession Numbers

The sequence of the *Sorangium cellulosum* So ce90 spirangien biosynthetic gene cluster reported in this paper has been deposited in the EMBL database with accession code AM407731.