Identification of a bioactive 51-membered macrolide complex by activation of a silent polyketide synthase in *Streptomyces ambofaciens*

Luisa Laureti, Lijiang Song, Sheng Huang, Christophe Corre, Pierre Leblond, Gregory L. Challis, and Bertrand Aigle

*Unité Mixte de Recherche 1128 Génétique et Microbiologie, Université Henri Poincaré, Institut Fédératif de Recherche 110 EFABA, F-54506 Vandœuvre-lès-Nancy, France;* 
*Unité Mixte de Recherche 1128 Génétique et Microbiologie, Institut National de la Recherche Agronomique, F-54506 Vandœuvre-lès-Nancy, France;* 
*Department of Chemistry, University of Warwick, CV4 7AL Coventry, United Kingdom*

Edited by Chaitan Khosla, Stanford University, Stanford, CA, and accepted by the Editorial Board February 25, 2011 (received for review December 20, 2010)

**Results and Discussion**

**Identification and in Silico Analysis of a Unique Type I PKS Gene Cluster in *S. ambofaciens* ATCC23877.** Among the cryptic biosynthetic gene clusters identified by analysis of the partial *S. ambofaciens* genome...
sequence, one cluster located at ~500 kb from the right end of the linear chromosome contains nine genes that code for putative type I modular PKSs. Sixteen further genes, encoding proteins predicted to be involved in PKS substrate supply, post- and on-PKS tailoring reactions, deoxysugar biosynthesis, regulation, and resistance flank, and are interspersed with the PKS genes (Fig. 1A and SI Appendix, Table S1). The limits of the cluster were proposed as the samR0463 and samR0487 genes on the basis of sequence comparisons (SI Appendix, Table S1). The proposed functions of the flanking genes samR0464 (predicted to encode an endoribonuclease) and samR0488 (predicted to encode a “cold-shock” DNA binding protein) are unrelated to the biosynthesis of secondary metabolites. Thus, this putative gene cluster contains 25 genes that span 150 kb, making it one of the largest polyketide biosynthetic gene clusters identified to date. Sequence analyses of the PKSs encoded by the gene cluster using SEARCHPKS (13) revealed that they contain 112 enzymatic domains organized into 25 modules (Fig. 1B). All of the enzymatic domains, except the ketoreductase (KR) domain in the last module, are predicted to harbor a functional active site. The presence of a ketosynthase (KS) domain at the N terminus of SamR0467 and a thioesterase (TE) domain at the C terminus of SamR0474 indicates that SamR0467 and SamR0474 contain the modules that initiate and terminate polyketide chain assembly, respectively (14, 15). Following the precedent of other type I modular PKS systems (16), the order in which the remaining seven PKSs act in chain assembly was assumed to follow the order of the PKS genes in the cluster (Fig. 1B). Sequence analyses of the acyltransferase (AT) domain within each module suggested that the product of the PKS is assembled from 16 molecules of malonyl-CoA, eight molecules of methylmalonyl-CoA (one of which is used to create the propionyl starter unit), and one molecule of an unknown extender unit, loaded by the AT13 domain (13, 17) (Fig. 1C and SI Appendix, Fig. S1). These analyses, combined with the putative β-carbon-processing domains in each module (Fig. 1B), predicted the planar structure of the fully processed acyl thioester intermediate attached to the acyl carrier protein (ACP) domain in the last module of the PKS, with a single ambiguity—the nature of the C-26

![Fig. 1. Stambomycin biosynthetic gene cluster. (A) Organization of the stambomycin biosynthetic gene cluster. Limits of the cluster have been assigned using sequence comparisons. (B) Module and domain organization of the PKS encoded by the cluster, assigned using the SEARCHPKS program (13). The ketoreductase domain in module 24 is predicted to be nonfunctional because the tyrosine and asparagine residues of the catalytic triad are absent (42). ACP, acyl carrier protein; AT, acyltransferase; DH, dehydratase; ER, enoyl reductase predicted to generate a 2R acyl thioester intermediate; ERS, enoyl reductase predicted to generate a 2S acyl thioester intermediate; KR*, ketoreductase predicted to generate a 2R, 3R-acyl thioester intermediate; KR+, ketoreductase predicted to generate a 2R, 3S-acyl thioester intermediate; KR#, ketoreductase predicted to generate a 2S, 3S-acyl thioester intermediate; KR^, ketoreductase predicted to generate a 2S, 3R-acyl thioester intermediate; KS, β-ketoacyl synthase; KSQ, β-ketoacyl synthase in which the active site cysteine residue is replaced by glutamine; TE, thioesterase. (C) Predicted structure of the fully assembled polyketide chain attached to the ACP domain in the last module of the PKS. The structure of the side chain at C-26 could not be predicted because the nature of the substrate of the AT domain in module 12 could not be inferred from sequence comparisons. Atoms in the polyketide chain are colored according to the precursors from which they derive: blue, malonyl-CoA; red, methylmalonyl-CoA; purple, malonyl-CoA derivative of unknown structure.](https://www.pnas.org/content/108/15/6259)

Laureti et al. PNAS | April 12, 2011 | vol. 108 | no. 15 | 6259
substituted by the AT13 domain (Fig. 1C). The stereochemistry of each stereocenter in this intermediate was predicted by sequence analyses of the KR domains, which control the stereochemistry of the α and β carbons, and the enoylreductase (ER) domains, which control only the α carbon stereochemistry in the intermediates attached to each module of the PKS (Fig. 1 B and C and SI Appendix, Figs. S2 and S3) (18, 19). Substructure and similarity searches of the Chemical Abstracts database using the predicted structure of the final PKS-bound intermediate indicated that the product of the cryptic gene cluster was likely to be novel.

Five genes in the cluster were found to be homologous to genes within the spiramycin biosynthetic gene cluster that are responsible for the biosynthesis of mycaminosone, one of three deoxysugars within the spiramycin structure (7) (SI Appendix, Fig. S4). Thus, these five genes are predicted to encode enzymes catalyzing the conversion of α-d-glucose-1-phosphate to nucleotide diphosphate (NDP)-d-mycaminose (SI Appendix, Table S1 and Fig. S4). In addition, analysis with SEARCHGTr (20) indicated that the samR0481 gene encodes a putative glycosyl transferase (GT) capable of transferring a mycaminosyl residue from NDP to one or more of the hydroxyl groups in the product of the PKS. Taken together, these data strongly suggested that the final product of the cluster is a glycosylated lactone. It has been observed that GTs often require an auxiliary protein to glycosylate their acceptors, especially when the donor substrate is an α-d sugar like mycaminose (21). No gene encoding such a protein has been identified within the cryptic biosynthetic gene cluster. Nguyen et al. (22) recently characterized an auxiliary protein, encoded by the smr6 gene within the spiramycin biosynthetic gene cluster, which interacts with the mycaminosyltransferase Srm5, responsible for catalyzing transfer of the mycaminosyl residue from NDP to the spiramycin aglycone. They also showed that another auxiliary protein, Srm28, is flexible, and that in addition to the forosaminyltransferase Smr29, Srm28 can partner Srm5. Thus, it is possible that SamR0481 requires either Smr6 or Srm28, or both, for activity.

Three other genes are postulated to be involved in the biosynthesis of the predicted unique glycosylated polyketide. The samR0478 and samR0479 genes encode putative cytochromes P450, which usually catalyze post-PKS hydroxylation reactions (23). The predicted product of samR0485 is a type II thioesterase. Such enzymes typically catalyze hydrolysis of aberrant acyl groups attached to the ACP domains of the PKS that block chain elongation (24).

Activation of the Silent Cryptic Gene Cluster. Poor or no expression of the gene cluster could explain why its metabolic product has never been identified. To test this hypothesis, we examined the expression of selected PKS genes within the cluster under two growth conditions known to elicit production of the macrolide spiramycin (25). RT-PCR analyses indicated that the PKS genes were expressed very poorly or not at all throughout growth (Fig. 2), implying that the product of the gene cluster either is not produced or is produced in quantities likely to be below the detection limit of most analytical tools.

The in silico analysis of the gene cluster identified three genes coding for potential pathway-specific regulators. The samR0468 and samR0469 genes encode a putative two-component system, and samR0484 encodes a putative large ATP-binding regulator of the LuxR family (LAL) protein (SI Appendix, Table S1). LAL regulators contain an ATP-binding domain at the N terminus (Walker A and B motifs) and a helix-turn-helix domain at the C terminus, most likely responsible for DNA binding (26). Members of the LAL family have been previously described as activators of polyketide biosynthesis, such as PikD and RapH, which activate pikromycin and rapamycin production, respectively (27, 28). Thus, we decided to examine whether constitutive overexpression of the samR0484 gene triggers expression of the PKS genes within the cryptic cluster. The samR0484 gene was cloned into the conjugative and integrative vector pIB139 (29), placing it under the control of the strong constitutive ermE* promoter. The resulting plasmid was integrated into the attB site of the chromosome of S. ambofaciens to create the strain ATCC/OE484, and the empty pIB139 vector was separately integrated into the chromosome of S. ambofaciens as a control to create the strain ATCC/pIB139. Comparative transcriptional analyses by RT-PCR showed that the PKS genes within the cryptic cluster were expressed in the ATCC/OE484 strain throughout growth, thus suggesting that the LAL regulator is a pathway-specific activator of the gene cluster (Fig. 2). In contrast, little or no expression of the PKS genes was detected in the control strain. Constitutive expression of the genes encoding the two-component system resulted in no induction of expression of the PKS genes relative to the control (SI Appendix, SI Materials and Methods and Fig. S5).

Isolation and Structure Elucidation of the Metabolic Products of the Cryptic PKS Gene Cluster. To identify the product(s) of the gene cluster, the supernatant and the mycelium extracts of the strains ATCC/pIB139 and ATCC/OE484 were analyzed by comparative metabolic profiling using liquid chromatography-electrospray ionization-time of flight-mass spectrometry (LC-ESI-TOF-MS). Thus, two major species with molecular formulas C73H133NO22 (calculated for C73H132NO22; 1362.9232; found: 1362.9236; calculated for C73H133NO22; 1362.9232; found: 1362.9232) were identified in the methanolic mycelial extracts of the ATCC/OE484 strain that were absent in extracts of the ATCC/ pIB139 strain (Fig. 3 and SI Appendix, Figs. S6 and S7). To confirm that these compounds were the products of the cryptic gene cluster, we replaced the samR0467 gene coding for the first PKS in the assembly line in the ATCC/OE484 strain with a kanamycin resistance cassette. LC-ESI-TOF-MS analysis of mycelial extracts of this strain confirmed that production of the compounds was abrogated (Fig. 3). Similarly, inactivation of the samR0481 gene, encoding the glycosyl transferase, also abolished production of the compounds. LC-ESI-TOF-MS analysis indicated that the corresponding glycosylated products were produced instead (calculated for C65H119O19; 1203.8340; found: 1203.8344; calculated for C65H119O19; 1203.8340; found: 1203.8344) (SI Appendix, Figs. S8 and S9.).

A search of the Chemical Abstracts database did not retrieve any compounds with molecular formulas corresponding to the
identified metabolic products of the cryptic gene cluster. Thus, we purified these unique compounds from mycelial extracts of the ATCC/OE484 strain using semipreparative HPLC. 1H NMR analysis of the higher-molecular-weight species indicated that it consists of two almost identical compounds, which we named stambomycins A and B, in a ratio of ∼5:1. COSY, TOCSY, HMBC, HSQC, NOESY, and PENDANT NMR experiments showed that stambomycins A and B consist of a common 51-membered glycosylated macrolide core containing a tetrahydropyran resulting from the addition of the C-7 hydroxyl group to a 12-membered intermediate attached to the C-5 keto group (Fig. 4 and SI Appendix, Figs. S10–S22 and Table S2). The NMR experiments confirmed the close relationship between the planar structure of the macrocyclic aglycone and the predicted structure of the final intermediate attached to the ACP domain in the last module of the PKS (Figs. 1C and 4). 1H coupling constants of ~15 Hz for H-12/H-13 and H-48/H-49 and NOESY correlations between the C-10 methyl group and C-12, and between the C-24 methyl group and C-26, indicate that the C-12/C-13, C-48/C-49, C-10/C-11, and C-24/C-25 double bonds all have E configuration (Fig. 4). COSY and HMBC correlations combined with comparisons of 1H NMR data with the literature (30, 31) established that β-mycaminose is attached to the aglycone (Fig. 4 and SI Appendix, Fig. S14 and Table S2), as predicted by sequence analysis of the deoxysugar biosynthesis genes within the cluster. NMR spectroscopic analysis of the aglycones isolated from the samR0481 mutant of the ATCC/OE484 strain confirmed that they lacked the β-mycaminose residue at C-5 (SI Appendix, SI Materials and Methods, Figs. S31–S40, and Tables S4 and S5).

Several structural features of the stambomycins were not predictable from sequence analyses of the biosynthetic enzymes, including the nature of the C-26 side chain and the sites of glycosylation, hydroxylation, and macrocycle formation. A 4-methyl-n-hexyl side chain is proposed to be attached to C-26 in stambomycin A on the basis of COSY and HMBC correlations (Fig. 4 and SI Appendix, Figs. S10, S12, S13, S17, and S20). Similar analyses suggested that stambomycin B contains a 5-methyl-n-hexyl side chain at C-26. Several HMBC and NOESY correlations indicate that the deoxysugar residue is attached to the C-5 hydroxyl group of the aglycone (Fig. 4). COSY and HMBC correlations suggest that non–PKS-derived hydroxyl groups are attached to C-28 and C-50, and that the macrolactone results from formation of an ester bond between C-1 and the C-50 hydroxyl group (Fig. 4).

1H, COSY, HMBC, HSQC, NOESY, and PENDANT NMR experiments indicated that the lower-molecular-weight species consists of two compounds, which we named stambomycins C and D, in a ratio of ~3:1, that differ from stambomycins A and B only in the structure of the alkyl chain attached to C-26 (SI Appendix, Table S3 and Figs. S23–S30). Stambomycin C is proposed to have a 4-methyl-n-pentyl group at this position, whereas stambomycin D is proposed to bear an n-hexyl group (Fig. 4 and SI Appendix, Figs. S10, S12, S13, S26, and S29).

The experimentally elucidated structures of the stambomycins suggest the assembly of these remarkable macrolides involves unprecedented biosynthetic chemistry. The various hexyl/heptyl substituents at C-26 indicate that the AT domain of PKS module 12 selectively recognizes and loads hexyl/heptylmalonyl-CoA extender units onto the ACP domain of this module (SI Appendix, Fig. S41). Malonyl-CoA, methylmalonyl-CoA, ethylmalonyl-CoA, (2-propenyl)malonyl-CoA, and (2-chloroethyl)malonyl-CoA are all known or hypothesized to be used as extender units by type I modular PKSs (32). Crotonyl-CoA reductase/carboxylase-like enzymes catalyze reduction and carboxylation of α, β-unsaturated thioester precursors in the biosynthesis of ethylmalonyl-CoA, (2-propenyl)malonyl-CoA, and (2-chloroethyl)malonyl-CoA (32–34). No enzymes with sequence similarity to crotonyl-CoA reductase/carboxylases are encoded by genes within the stambomycin biosynthetic gene cluster, suggesting that a different mechanism is used to biosynthesize the hexyl/heptylmalonyl-CoA extender units in S. ambobaciens.

The C-1/O-50 ester linkage in the stambomycins suggests that hydroxylation of C-50, probably catalyzed by one of the two cytochrome P450s encoded within the cluster, occurs during polyketide chain assembly on one or more ACP-bound intermediates attached to the PKS (SI Appendix, Fig. S41). The C-50 hydroxyl group is not present in the 25(methyl)malonyl-CoA starter unit predicted to be recognized and decarboxylated by the loading module of the PKS, and is required to offload the fully assembled polyketide chain from the active site Ser residue of the TE domain by macrocyclization. Further experiments will be required to test this hypothesis.

**Bioactivity of the Stambomycins.** The broad range of biological activities exhibited by macrolides, including antibacterial (35), antifungal (36), and antiproliferative (37) activities, prompted us to investigate the biological activity of the stambomycins. They showed moderate activity against Gram-positive bacteria (SI Appendix, Table S6) but no activity against Gram-negative bac-

---

**Fig. 3.** Detection of the stambomycins. Base peak chromatograms from LC-ESI-TOF-MS analyses of methanolic mycelial extracts of ATCC/pIB139 (Top), ATCC/OE484 (Middle), and ATCC/OE484/ΔA467 (Bottom) strains. Peaks corresponding to stambomycins A/B and C/D in the ATCC/OE484 strain are labeled.

**Fig. 4.** Structure elucidation of the stambomycins. Structures of stambomycins A–D proposed on the basis of HRMS and NMR spectroscopic analyses. Key COSY, HMBC, and NOESY correlations for structural elements that were not predicted by sequence analysis of the biosynthetic enzymes are indicated by bold lines, single-headed arrows, and double-headed arrows, respectively. J coupling constants observed in 1H NMR spectra for H-12/H-13 and H-49/H-50 are indicated.
teria, or fungi. More interestingly, they inhibited proliferation of human adenocarcinoma (HT29) cell lines with a similar IC_{50} (1.77 ± 0.04 μM for stambomycins A/B and 1.74 ± 0.04 μM for stambomycins C/D) to the clinical anticancer agent doxorubicin (1.32 ± 0.08 μM), but were fourfold less toxic toward adult Chinese hamster ovary sable cells (CHO-K1; 8.47 ± 0.67 μM for stambomycins A/B, 8.46 ± 0.52 μM for stambomycins C/D) than doxorubicin (1.99 ± 0.25 μM). Significant antiproliferative activities were also observed against the human breast (MCF7), lung (H460), and prostate (PC3) cancer cell lines (SI Appendix, Table S6). These data indicate that stambomycins may represent promising new leads for anticancer drug discovery.

Conclusions

In summary, we have used constitutive overexpression of a putative pathway-specific LAL regulator to induce expression of a silent type I modular PKS gene cluster in S. ambifaciens. This enabled us to identify a unique structural class of polyketides with promising antitumor activity that represents the largest macrolides ever to be isolated from an actinomyces. In addition, structure elucidation of these molecules disclosed unique and interesting chemical features; in particular, the incorporation of hexyl- and heptylmalonyl-CoA extender units and a mechanism of macrolactonization involving trans hydroxylation of PKS-bound intermediate(s) to release the fully assembled polyketide chain from the PKS. Manipulation of a putative pathway-specific regulatory gene has been used to uncover a unique natural product by activating expression of a silent hybrid PKS-NRPS gene cluster in a filamentous fungus (38). A similar tactic has hitherto not been used for the identification of natural metabolic products of cryptic gene clusters in actinomycetes, which are one of the main producers of microbial antibiotics and other bioactive natural products. Database searches indicate that LAL regulators are present in numerous gene clusters that direct the production of known secondary metabolites (SI Appendix, Table S7), as well as several cryptic biosynthetic gene clusters for which the metabolic product(s) remain to be discovered (SI Appendix, Table S8). The findings reported here suggest that our approach is likely to represent a powerful general strategy for realizing the drug-discovery potential offered by the numerous cryptic biosynthetic pathways identified in actinomycete genome sequences.

Materials and Methods

Bacterial strains, plasmids and BACs are described in SI Appendix, Table S9, and their manipulation is described in SI Appendix, SI Materials and Methods.

Bioinformatics Analysis.

The programs SEARCHPKS (13) and SEARCHGTr (20) were used to analyze the PKSs and the glycosyltransferase encoded by genes within the cluster. Functional analysis of each gene product and sequence comparisons were carried out using Blast2 and Blast 2.2.27 (http://blast.ncbi.nlm.nih.gov/blast.cgi).

Overexpression of the LAL Regulator.

The SamR0484-encoding sequence was amplified by PCR from S. ambifaciens ATCC23877 genomic DNA with the primers OE484-F and OE484-R (sequences in SI Appendix, Table S10). The PCR conditions, using Takara polymerase (Fermentas), were as follows: 2 min at 94 °C, 30 cycles of 10 s at 94 °C, 30 s at 60 °C, and 30 s at 72 °C; and a final extension of 5 min at 72 °C. To check possible contamination of genomic DNA, the same PCR program (35 cycles instead of 28) was applied to RNA samples, after DNase treatment, using as a control primers designed to amplify the hrdB gene, which encodes the major sigma factor and is considered to be constitutively expressed.

Construction of S. ambifaciens Mutant Strains.

The REDIRECT system (40) was used to make an in-frame deletion of samr0467 in S. ambifaciens ATCC/OE484 as described previously (41). The neo + oriT mutation cassette, derived from the plasmid pH776 (40), was used as template in the PCR with the primers D467-F and D467-R. E. coli BW25113/pKD20 was transformed, first with the BAC of interest (BBB), and then with the PCR product (~1400 bp) to replace the biosynthetic gene samr0467 by homologous recombination. The chloramphenicol resistance gene of the vector pBluBAC11 was replaced by a spectinomycin resistance gene, using the same strategy. E. coli ET1025/pJZ8002 was transformed with the mutated BAC (BBBspec/aJ467::neo+oriT) for conjugation with the ATCC/OE484 strain. The kanamycin-resistant and spectinomycin-sensitive clones were picked and analyzed by PCR, Southern blotting, and pulsed-field gel electrophoresis.

Identification, Isolation, and Structural Elucidation of Stambomycins A/B and C/D.

The PKS products were identified by LC-MS comparative metabolic profiling of S. ambifaciens ATCC23877, ATCC/OE484, and ATCC/OE484/A467 strains grown in liquid MP5 medium (details in SI Appendix). To purify stambomycins, solid MP5 medium was used. A 50-mL quantity of MP5 molten solid medium was poured into a 10 × 10 cm plate. The cultures were grown for 10 days. The mycelia were harvested, washed with water, and then extracted with 40 mL methanol. The mixture was sonicated for 10 min, 80% A/20% B; 20 min, 100% B; 35 min, 100% B. Absorbance was recorded at 240 nm. Fractions containing stambomycins were concentrated methanol extract using semipreparative HPLC on a reverse-phase column (250 mm × 21 mm, 25 μL). The mobile phases were 60% A/40% B; 0 min, 100% A; 10 min, 100% B; 20 min, 100% B; 35 min 100% B. Absorbance was monitored at a wavelength of 240 nm. Fractions containing stambomycins were identified by ESI-MS and combined. The combined fractions were evaporated under reduced pressure and resuspended in a small volume of 50% aqueous methanol. Stambomycins were purified by the combined fractions by semi-preparative HPLC on the same column using the following mobile phase: 0 min, 60% A/40% B; 15 min, 5% A/95% B; 20 min, 100%; 25 min, 100% B. Fractions containing a mixture of stambomycins A and B were collected, combined, and lyophilised, as were separate fractions containing a mixture of stambomycins C and D. An 7.8-mg quantity of stambomycins A/B and 4.3 mg stambomycins C/D were obtained from 10 plates. NMR spectra were recorded in d6-MeOH and, in the case of HMBC spectra, for stambomycins A/B also d6-DMSO, on Bruker Avance 700 MHz or 500 MHz spectrometers equipped with cryoprobes.

Acknowledgments.

We thank Peter Leadlay (University of Cambridge, Cambridge, United Kingdom) for the kind gift of pI3819. We likewise thank Dr. Peter Grice (University of Cambridge, Cambridge, United Kingdom) with assistance in the recording of NMR spectra. The Bruker MaxXs mass spectrometer used in this research was obtained, through Birmingham Science City: Innovative Uses for Advanced Materials in the Modern World (West Midlands Centre for Advanced Materials Project 2), with support from Advantage West Midlands and was partly funded by the European Regional Development Fund. This research was supported by the European Commission under the 6th framework program (integrated project actinogen; FP6-522).


3. Challis GL (2008) Mining microbial genomes for new natural products and bio-
421–431.
37. Olano C, Méndez C, Sambucetti L (2009) Antitumor compounds from marine acti-