2-Alkyl-4-hydroxymethylfuran-3-carboxylic acids, antibiotic production inducers discovered by Streptomyces coelicolor genome mining

Christophe Correa1, Lijiang Songa, Sean O'Rourkeb, Keith F. Chaterb, and Gregory L. Challisa

1Department of Chemistry, University of Warwick, Coventry CV4 7AL, United Kingdom; and 2Department of Molecular Microbiology, John Innes Centre, Norwich NR4 7UH, United Kingdom

Edited by Arnold L. Demain, Drew University, Madison, NJ, and approved August 28, 2008 (received for review June 6, 2008)

All of the genetic elements necessary for the production of the antibiotic methylenomycin (Mm) and its regulation are contained within the 22-kb mmy- mmf gene cluster, which is located on the 356-kb linear plasmid SCP1 of Streptomyces coelicolor A3(2). A putative operon of 3 genes within this gene cluster, mmflHLP, was proposed to direct the biosynthesis of an A-factor-like signaling molecule, which could play a role in the regulation of Mm biosynthesis. The mmflHLP operon was expressed under the control of its native promoter in S. coelicolor M512, a host lacking the SCP1 plasmid, and the ability to produce prodigine and actinorhodin antibiotics. Comparative metabolic profiling led to the identification and structure elucidation of a family of 5 new 2-alkyl-4-hydroxymethylfuran-3-carboxylic acids (AHFCAs), collectively termed Mm furans (MMFs), as the products of the mmflHLP genes. MMFs specifically induce the production of the Mm antibiotics in S. coelicolor. Comparative genomics analyses and searches of the natural product chemistry literature indicated that other streptomyces may produce AHFCAs, suggesting that they could form a general class of antibiotic biosynthesis inducers in Streptomyces species, with analogous functions to the better known γ-butyrolactone regulatory molecules.

Low-molecular-weight diffusible molecules produced by microorganisms are able to trigger such activities as secondary metabolism, morphological development, and other specialized ecological responses (1, 2). The major characterized classes of these so-called autoregulators are the acylhomoserine lactones (AHLs) in Gram-negative bacteria and the γ-butyrolactones (GBLs) in Gram-positive Streptomyces bacteria, which are an outstanding source of medically useful natural products (3, 4). GBLs are often involved in the regulation of secondary metabolism and morphological development in Streptomyces. A-factor (compound 1) (Fig. 1A), which induces production of the antibiotic streptomycin and morphological differentiation in Streptomyces griseus, was the first GBL to be discovered and has been the most investigated to date (5). Recently, the butenolidine synthase AsfA was definitively shown to be the key enzyme in A-factor biosynthesis (6).

Streptomyces coelicolor A3(2), a model streptomycete, produces at least 7 closely related GBLs (7–10). SCB1 (compound 2) (Fig. 1A), the most studied of these GBLs, appears to function mainly in the transcriptional regulation of a gene cluster proposed to direct production of an unknown polyketide metabolite (7, 11–14). The S. coelicolor chromosome contains 1 homologue of afsA called scbA, which is located adjacent to the cryptic modular polyketide synthase gene cluster (13) and is required for the biosynthesis of GBLs (15). Another afsA-like gene, mmflL, is located within the linear plasmid SCP1 adjacent to the Mm antibiotic biosynthetic gene cluster (16, 17). The mmflL gene product (353 amino acids) shares 43% similarity and 26% identity over 312 amino acids with ScbA. The putative A-factor-like molecules resulting from the catalytic activity of MmfL have not been identified. Disruption of mmflL specifically abrogates Mm production, which can be restored in an mmflL− mutant by addition of an ethyl acetate extract of the culture supernatant from an mmflL+ strain of S. coelicolor. Thus, the mmflL gene, in association with the adjacent mmflP and mmflH genes, is proposed to direct the production of diffusible molecules that activate the Mm biosynthesis operon.

Author contributions: C.C., K.F.C., and G.L.C. designed research; C.C., L.S., and S.O. performed research; C.C., L.S., and G.L.C. analyzed data; and C.C. and G.L.C. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

1To whom correspondence should be addressed: E-mail: c.corre@warwick.ac.uk.

This article contains supporting information online at www.pnas.org/cgi/content/full/0805530105/DCSupplemental.

© 2008 by The National Academy of Sciences of the USA
induce Mm production in S. coelicolor (Fig. 2). These molecules are thought to be recognized by the putative DNA-binding receptor proteins MmyR and MmfR, both of which have sequence similarity to the A-factor receptor protein ArpA and other GBL-binding proteins (18). MmyR and MmfR regulate the expression of the Mm biosynthetic genes in a manner that is responsive to the organic extract from mmy/LHP + strains. The simplest interpretation of these results, based on precedent, is that Mm production is regulated by a GBL. However, the inducers of Mm production in organic extracts were not inactivated by alkaline hydrolysis, unlike GBLs, suggesting that they do not contain a lactone. Here we report the identification and structure elucidation of a family of molecules collectively termed Mm furans (MMFs) that are assembled by the products of the mmy/LHP genes and induce Mm production in S. coelicolor.

Results

Expression of mmy/LHP in S. coelicolor M512 and Identification of the Resulting Products by Comparative Metabolic Profiling. The integrative plasmid pIJ6584, containing the mmy/LHP putative operon and the upstream native promoter region, was constructed and integrated into the chromosome of S. coelicolor M512 (which lacks SCP1) to generate S. coelicolor W74. Metabolites in organic extracts of culture supernatants of the W74 and M512 strains were compared by liquid chromatography (LC)-electrospray ionization (ESI)-MS analyses. Five compounds present in extracts of the W74 strain, but absent in extracts of the M512 strain, were identified (Fig. 3). These extracellular compounds were named MMFs 1–5.

Characterization of the MMFs by Mass Spectrometry and UV–Visible Spectroscopy. The LC-ESI-MS analyses of the MMF-containing extracts in positive-ion mode showed that MMF2 produced ions with m/z = 167 and 207, MMF1 and MMF4 both produced ions with m/z = 181 and 221, and MMF3 and MMF5 both produced ions with m/z = 195 and 235. In contrast, LC-ESI-MS analyses of the extracts in negative-ion mode showed that MMF2 produced an ion with m/z = 183, MMF1 and MMF4 both produced an ion with m/z = 197, and MMF3 and MMF5 both produced an ion with m/z = 211. The ions observed in positive-ion mode were assigned as [M+H-H2O]+ and [M+Na]+, respectively, whereas the ions observed in negative-ion mode were assigned as [M-H]-. LC-ESI-tandem MS (MS/MS) analyses showed that all 5 compounds fragmented in the same way, providing a strong indication that they share a common core structure (SI Appendix, Fig. S8).

Fractions from high-pressure LC (HPLC) separation of the organic extract of the W74 strain containing MMF1–MMF5 were collected and analyzed by ESI-time of flight (TOF)-MS in positive and negative ion modes (SI Appendix, Figs. S9–S13). From these analyses, the molecular formulae of MMF2, MMF1/ MMF4, and MMF3/MMF5 were deduced as C9H12O4, C10H14O4, and C11H16O4, respectively, further indicating that these compounds contain a common core structure that is decorated by varying alkyl chains. All 5 compounds exhibited characteristic absorbance maxima at 250 nm, suggesting the presence of a common chromophore likely associated with the common core structure.

Purification of MMF1 and MMF2 and Structure Elucidation by NMR Spectroscopy. MMF1 was isolated by HPLC from ethyl acetate extracts of the culture medium of S. coelicolor W74 grown on agar for 5 days. Approximately 1 mg of MMF1 was obtained of >90% purity. The structure of MMF1 was elucidated as 2-(2-methyl-1-propyl)-4-hydroxymethylfuran-3-carboxylic acid 3 by ESI-MS/MS and 1D and 2D NMR spectroscopy (Fig. 1B; SI Appendix, Sections 2.2 and 2.3). Key correlations from heteronuclear multiple bond correlation (HMBC) and correlated

Corre et al.
2D-NMR experiments (Sections 2.2 and 2.3). The thioester intermediates in fatty acid biosynthesis followed by tributyrin to their structural elucidation.

Biosynthetic origin of the MMFs. (A) Proposed metabolic origin of the MMFs. (B) Feeding experiments with labeled precursors of starter units in fatty acid biosynthesis to S. coelicolor. Deuterium labeling experiments with D10-leucine were specifically incorporated into MmF1, presumably via catabolism of the labeled leucine to D10-3-methylbutyryl-CoA; all 7 deuterium atoms from D7-butyrate were specifically incorporated into MmF2 and MmF5, via D7-butyryl-CoA; 7 of the 8 deuterium atoms in D8-valine were incorporated into MmF3, presumably via catabolism of the labeled valine to D7-iso-butyryl-CoA; and all of the deuterium atoms in D7-propionate were incorporated into MmF4, via D7-propionyl-CoA (Fig. 4B). These results lead us to conclude that the alkyl substituents at C-2 in MmF3, MmF4, and MmF5 are 3-methylbutyl, n-buty, and n-pentyl, respectively.

Expression of mmfl in S. coelicolor and Escherichia coli and Analysis of MMF Production by LC-MS. The integrative plasmid pBl266 contains the key mmfl gene under the control of the native T7 promoter was constructed and integrated into the chromosome of S. coelicolor M512, generating S. coelicolor W73. The role of MmF in MMF biosynthesis was investigated by analyzing organic extracts of this strain by LC-MS. These analyses indicated that S. coelicolor W73 could produce the same 5 MMFs in the same relative proportions as S. coelicolor W74 but at much lower levels. However, the absolute levels varied significantly from experiment to experiment. The mmfl gene was also cloned into pET151 (a high-copy-number plasmid that can replicate autonomously in E. coli) under the control of the IPTG-inducible T7 promoter to generate pCC001. This construct was used to transform E. coli BL21star. LC-MS analyses of culture supernatant organic extracts showed that induction of mmfl overexpression in this strain resulted in the production of MmF2 and MmF5 (Fig. 3). The structure of the major MMF produced by E. coli BL21star/pCC001 was confirmed as MmF2 by ESI-TOF-MS analyses and LC-MS/MS comparisons with MmF2 isolated from S. coelicolor W74 (SI Appendix, Figs. S6 and S7).

Induction of Mm Production by MMFs. S. coelicolor W81 contains the entire mmy-mmfl gene cluster minus the mmflHLP genes. LC-MS analyses of this strain show that it does not produce MMFs or Mm antibiotics. Using a bioassay for Mm production, we showed that Mm production was restored in an agar plug with S. coelicolor W81 growing on its surface placed next to an agar plug with S. coelicolor W74 (expressing mmflHLP) growing on its surface. We hypothesized that this restoration of antibiotic production results from diffusion of MMFs from the W74 strain into the W81 strain. To test this hypothesis, we placed agar plugs containing purified MmF1 next to an agar plug with the W81 strain growing on its surface. The bioassay showed that antibiotic production was restored in the W81 strain by addition of MmF1 (Fig. 5), MmF2, MmF4, or the organic extract of E. coli BL21/pCC001 culture supernatant. Organic extraction of the plug containing the W81 strain followed by LC-MS analysis established that the produced antibiotics were the Mms.

Discussion

Genome mining has become a powerful approach for the discovery of natural products (19, 20). Here we report the discovery of the MMFs, the fifth family of previously unknown natural products of the model actinomycete S. coelicolor A3(2) to be discovered by a genome mining approach (21–24). One advantage of the genome mining approach for natural product discovery is that the genes and enzymes involved in the biosynthesis of the compounds can be simultaneously discovered, potentially leading to the rapid development of new therapeutic molecules.
thesis of the newly discovered natural product are also identified. Here we were able to identify that MmfL, MmfP, and MmfH all play a role in MMF biosynthesis and that the role played by MmfL is critical. Another advantage of the genome mining approach is that the genomic context of the identified biosynthetic genes for the newly discovered natural product often provides clues about its biological function. This was the case here; the localization of mmfLHP adjacent to genes known to be involved in Mm biosynthesis indicated that the MMFs may act as signaling molecules that control production of the Mms. We were able to prove this hypothesis by showing that purified MMF induces production of Mms in S. coelicolor strains lacking the mmfLHP genes.

Structurally related rhamnosylated 2-alkyl-4-hydroxymethylfuran-3-carboxylic acids (AHFCAs) 5 and 6 have been isolated from different Streptomyces species (Fig. 6) (25, 26). Feeding experiments showed that the aglycone of compound 6 is derived from 1 molecule of glycerol and 3 molecules of acetate (25). The results of these experiments are consistent with our proposed pathway for MMF biosynthesis (Fig. 4A). The biological function of compounds 5 and 6 and the enzymes involved in their biosynthesis have yet to be discovered. It is tempting to speculate that the biosynthesis of compounds 5 and 6 involves enzymes with significant sequence similarity to MmfL, MmfH, and MmfP and that compounds 5 and 6, or their aglycones, may function as signaling molecules in the Streptomyces species that produce them.

The observation that expression of mmfL alone in S. coelicolor M512 can lead to production of MMFs shows that MmfL plays a key role in MMF biosynthesis. MmfL shows sequence similarity to AfsA, a butenolide synthase that has recently been shown to catalyze the key steps in the biosynthesis of the GBL A-factor (6). GBLs are biosynthesized from β-ketoacyl thioesters, which are postulated to be intermediates in fatty acid biosynthesis and dihydroxyacetone phosphate (DHAP). The incorporation of labeled precursors into compound 6 (25), together with similar experiments reported here, suggest that MMFs are also derived from a β-ketothiolayest intermediate in fatty acid biosynthesis and DHAP. Further support for the involvement of fatty acid biosynthetic intermediates in MMF assembly derives from the change in the nature of the MMFs produced when MmfL is expressed in E. coli rather than S. coelicolor. In E. coli, only MMFs with linear alkyl chains are produced, whereas in S. coelicolor both straight and linear chains are produced. This finding reflects the fact that fatty acids in E. coli are derived principally from the straight chain starter acetyl-CoA, whereas in Streptomyces species fatty acids are derived from branched and linear (acetoyl-CoA, propionyl-CoA, and butyryl-CoA) chain starter units (27). Expression of mmfLHP in S. coelicolor M512 consistently leads to higher titers of MMFs (compared with expression of just mmfL), suggesting that MmfH and MmfP also play a significant role in MMF biosynthesis. Homologues of MmfH and MmfP are encoded by other genes within the S. coelicolor genome (actVA and sco3558, respectively) (7). These homologues may be able to catalyze the reaction usually catalyzed by MmfH and MmfP in their absence but with lower efficiency. Intriguingly, we identified a cluster of 3 genes that is very similar to mmfLHP in the Streptomyces avermitilis genome (28), suggesting that this streptomycete may also be able to produce AHFCAs (SI Appendix, Fig. S3).

AfsA-like proteins catalyze the condensation of selected β-ketoacyl thioesters derived from fatty acid metabolism with DHAP to form phosphorylated butenolides, which undergo dephosphorylation, reduction to the corresponding butanolate, and in some cases ketoreduction, to form GBLs (6). It is conceivable that MmfL also catalyzes condensation of selected β-ketoacyl thioesters derived from fatty acid metabolism with DHAP to form phosphorylated butenolide intermediates, which undergo MmfP-catalyzed dephosphorylation. MmfH would catalyze conversion of the resulting butenolides to AHFCAs (Fig. 7). In vitro experiments with purified MmfL, MmfH, and MmfP will be required to verify these hypotheses and precisely define the roles of these proteins in AHFC biosynthesis.

Our demonstration that purified MMF1, MMF2, and MMF4 induce Mm production in S. coelicolor strains containing the Mm biosynthetic gene cluster, but lacking the mmfLHP genes shows that AHFCAs represent a class of antibiotic biosynthesis inducers distinct from GBLs. In GBL-mediated signaling, gene ex-
pression is controlled by repressor proteins that are derepressed by GBL binding. S. griseus ArpA, which binds A-factor, is the archetype of such repressor proteins (18). The mmyR and mmyR genes flanking mmyLHP both encode ArpA homologues. It is therefore tempting to speculate that MmyR and MmyR are repressors that control expression of the adjacent Mm biosynthetic genes, either directly or indirectly, by binding MMFs. Further experiments are required to fully elucidate the regulatory cascade controlling expression of the mmy cluster and to precisely define the role of MMFs in this process.

In conclusion, we have discovered a family of AHFCAs using a genome mining approach. The biological function of these AHFCAs is to induce Mm biosynthesis in S. coelicolor, analogous to the well-known GBL family of inducers of antibiotic biosynthesis in Streptomyces species. AHFCAs and GLBs appear to be derived from the same primary metabolic precursors, and the enzymatic machinery involved in the biosynthesis of these 2 classes of signaling molecules shares several similarities. It will be intriguing to discover exactly how and where the 2 biosynthetic pathways diverge. Finally, the fact that glycosylated AHFCAs have been isolated from other Streptomyces species and the fact that a putative operon containing 3 genes encoding proteins with high degrees of similarity to MmL, MmH, and Mmp, respectively, is present in X. oryzae (21), indicates that AHFCAs may be widespread in Streptomyces species where they are likely to function as a new general class of diffusible signaling molecule (SI Appendix, Section 1.2).

Materials and Methods

Strains and Plasmids. DNA fragments containing the upstream promoter region and mml or mmlHLP were subcloned into the integrative plasmid pSET152 to generate plj6566 and plj6584, respectively (SI Appendix, Section 1.1.1). These 2 plasmids were introduced into S. coelicolor M512 (SCP1, SCP2, act, red) via site-specific integration to generate S. coelicolor W73 and W74, respectively (21). These strains were analyzed by PCR to confirm the presence of the mmlH gene and the mmlHLP operon, respectively (SI Appendix, Fig. S2).

For the E. coli expression construct, the mml gene was amplified by PCR from cosmID C73 of the S. coelicolor SCP1 ordered library. A CACC sequence was introduced into the forward PCR primer before the natural start codon, to allow the directional TOPO cloning of blunt-end PCR products into pET151/D-Topo. The forward primer used was 5′-CACCATGGAACCATCACAAAC-CGCTTCTTACTG-3′. The reverse primer used was 5′-TCTATGGCCAGCCGGCTTCG-3′. The sequence of the cloned gene in the pET151/D-Topo expression vector was confirmed by DNA sequencing (Molecular Biology Service, Department of Biological Sciences, University of Warwick). One correct clone (pc001) was used to transform E. coli BL21Star(DE3) (Invitrogen) for expression of an N-terminal His6-in-frame fusion.

The mmlHLP genes of the integrative cosmid C73–787, which contains the entire Mm biosynthetic gene cluster, were replaced by the apramycin resistance gene apr [previously named aac (3)III] using PCR-targeting (21) to generate S. coelicolor MMF and MMF producing strain W81 (SI Appendix, Section 1.1.2).

Culture Conditions. The M512, W73, and W74 strains were grown on AlphMM agar medium [30 mm L-alanine, 5 mm K-HPO4, 5 mm MgSO4·hydrate, 10 g/L glycerol, and 15 g/L agar in tap water, adjusted to pH 5.0]. After incubation at 30 °C for 5 days, the agar, together with the mycelia, was transferred to a 250-ml flask containing 50 mL of ethyl acetate. The mixture was agitated by swirling and allowed to stand overnight. The ethyl acetate was decanted. The culture supernatant was adjusted to pH 3 with concentrated HCl and was extracted twice with 50 mL of ethyl acetate. The combined extracts were evaporated to dryness and redissolved in 1 mL of 1:1 H2O/MeOH. A 50-μL aliquot of the resulting solution was analyzed by LC-MS as described below.

Analytical Methods. LC-MS analyses were carried out by using a reverse-phase column (Agilent C18, 150 × 4.6 mm, 5 μm) connected to an Agilent 1100 HPLC instrument. The outflow was routed to a Bruker HCT+ ion trap mass spectrometer fitted with an electrospray source operating in positive- or negative-ion mode. A 5 min isocratic elution (70:30 with H2O containing 0.1% HCOOH and MeCN (formic acid was replaced by NH3 when detecting in negative-ion mode) followed by a gradient reaching 50:50 after 25 min permitted the separation of the MMFs. Direct injection M5 (Bruker HCT+) and ESI-TOF-MS (Bruker MicroTOF) experiments were carried on fractions collected from the LC-MS analyses.

Purification of MMF1. Forty Petri plates, each containing 25-ml AlphMM agar medium, were inoculated with spores of S. coelicolor W74. After 5 days at 30 °C, the medium was removed from the plates, combined, and extracted with 500-mL ETOAc. The organic extract was separated from the agar by filtration and dried over MgSO4. The solvent was removed under vacuum, and the residue was redissolved in 2 ml of 50:50 H2O/MeOH. MMF1 was purified further by HPLC (condition: HPLC column Eclipse X2 150 × 4.6 mm, 5 μm) attached to an Agilent 1100 HPLC instrument fitted with a quaternary pump and variable wavelength detector, monitoring absorbance at 265 nm. Four injections of 0.5 mL were performed with H2O containing 0.1% HCOOH (solvent A)/MeCN (solvent B) as the eluent. The elution profile was as follows: 95:5 solvent A/solvent B for 5 min, then 100 solvent A/solvent B over 25 min. The flow rate was 5 mL/min. The collected fractions were analyzed by ESI-MS, and those containing the compound with m/z = 181 (retention time, 15.5–16 min) were combined and the solvent was removed under vacuum. The residue was resuspended in 2 ml of 50:50 H2O/MeOH and further purified by HPLC with a different elution profile (75:25 solvent A/solvent B for 5 min then 100 solvent A/solvent B over 25 min). The collected fractions were analyzed by ESI-MS, and those containing the compound with m/z = 181 (retention time, 10.5 min) were combined and the solvent was removed using a rotary evaporator. The residue was dried under vacuum and analyzed by 1H, 13C, and 15N NMR experiments (CDCl3, Bruker AV700 spectrometer equipped with a TCI cryoprobe). MMF1 was then quantified by 1H-NMR spectroscopy by adding 5 μmol of ChCl3 as an internal standard. A 1 mg/ml solution of MMF1 in methanol was then prepared and used to assess the induction of Mm production.

Incorporation Experiments with Labeled Precursors. Fifty milliliters of AlphMM liquid medium was inoculated with spores of S. coelicolor W74 and incubated at 30 °C at 180 rpm. After 24 h, a solution of the appropriate quantity of labeled precursor (D9-Leucine, D3-Butyric acid, D3-Valine, D5-Proionic acid, or 13C6-Isocitrate) (≥99%) was added to the perfusion medium (with 500-ml EtOAc). The organic extract was separated from the agar by washing, filter-sterilized, and then fed to the culture. Thereafter, this feeding procedure was repeated 4 more times at 12-h intervals. The culture was harvested after 5 days of total incubation time and extracted as described above. The extracts were analyzed by LC-MS (SI Appendix, Figs. S14–S18).

Mm Biosynthesis Induction Assay. S. coelicolor M145 is an SCP1-deficient strain and is therefore sensitive to Mm. M145 was used in a bioassay for Mm production. An AlaphMM (pH 5.0, allowing optimal diffusion of Mm) plate was inoculated with S. coelicolor M512, W74, or W81 and incubated at 30 °C for 48 h. A plug from the plates containing confluent lawns of M512 or W74 cells was then transferred to a plug-containing confluent lawn of Mm onto an AlaphMM (pH 5.0) agar plate inoculated with a spore suspension of the Mm-sensitive strain M145. Zones of growth inhibition around the plugs containing the M81 strains, after 96–120 h of additional incubation at 30 °C, indicated Mm was being produced. A variation of this assay, solutions (20 μL) containing HPLC-purified MMF1 were dropped onto a plug of AlaphMM (pH 5.0) agar placed next to an agar plug containing confluent lawns of W81 cells on an AlaphMM (pH 5.0) agar plate inoculated with a spore suspension of S. coelicolor M145. E. coli growth was detected around the plugs containing the W81 strain, as described above.

ACKNOWLEDGMENTS. We thank Dr. Nadia Kadi (University of Warwick) for assistance with cloning and overexpression of mmlH in E. coli. This work was performed in the United Kingdom through the Biological Sciences Research Council Grant BB/E008003/1 and European Union Integrated Project ActinoGEN Contract 005224.


15. Hsiao N-H, et al. (2007) ScbA from Streptomyces coelicolor A3(2) has homology to fatty acid synthases and is able to synthesize γ-butyrolactones. Microbiology 153:1394–1404.


Supporting information

2-Alkyl-4-hydroxymethylfuran-3-carboxylic acids, antibiotic production inducers discovered by *Streptomyces coelicolor* genome mining

Christophe Corre, Lijiang Song, Sean O'Rourke, Keith F. Chater and Gregory L. Challis

Contents:

1 Strains and plasmids
   1.1 Construction of plasmids and strains
       1.1.1 Generation of plasmids pIJ6566 and pIJ6584 and introduction into *S. coelicolor* M512 to generate *S. coelicolor* W73 and W74, respectively
       1.1.2 Generation of plasmid pCC002 and introduction into M145 to generate *S. coelicolor* W81
       1.2 Sequence comparison of *S. coelicolor* and *S. avermitilis* AHFCA gene clusters
       1.3 Inducing activity contained in *E. coli* BL21 / pCC001 organic extract

2 Analytical data
   2.1 UV/Vis Spectroscopy
   2.2 Mass Spectrometry data
   2.3 NMR Spectroscopy data

3 References
1 Strains and plasmids

The strains and plasmids used in this study are summarized in Table S1.

**Table S1**: Strains and plasmids used in this study

<table>
<thead>
<tr>
<th>Strains / Plasmids</th>
<th>Relevant properties</th>
<th>Source / Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strains</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M145</td>
<td>SCP1(^{-}), SCP2(^{-}) (Methylenomycin-sensitive)</td>
<td>Kieser T, et al., 2000 (1)</td>
</tr>
<tr>
<td>M512</td>
<td>M145 with ΔactII-ORF4 ΔredD (MMF minus, Methylenomycin minus)</td>
<td>Floriano and Bibb, 1996 (2)</td>
</tr>
<tr>
<td>W73</td>
<td>M512 with pIJ6566::attB (MMF producer)</td>
<td>This study</td>
</tr>
<tr>
<td>W74</td>
<td>M512 with pIJ6584::attB (MMF producer)</td>
<td>This study</td>
</tr>
<tr>
<td>W81</td>
<td>M145 with pCC002::attB (MMF minus, Mm minus)</td>
<td>This study</td>
</tr>
<tr>
<td><strong>Plasmids</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pIJ6566</td>
<td>pSET152 containing mmfL gene and promoter region</td>
<td>This study</td>
</tr>
<tr>
<td>pIJ6584</td>
<td>pSET152 containing mmfLHP genes and promoter region</td>
<td>This study</td>
</tr>
<tr>
<td>C73_787</td>
<td>Integrative cosmid containing the full methylenomycin cluster including the MMFs biosynthetic genes</td>
<td>Gust B, et al., 2004 (3)</td>
</tr>
<tr>
<td>pCC002</td>
<td>C73_787 with mmfL→P::apr</td>
<td>This study</td>
</tr>
<tr>
<td>pCC001</td>
<td>pET151 containing mmfL</td>
<td>This study</td>
</tr>
</tbody>
</table>

1.1 Construction of plasmids and strains

1.1.1 Generation of plasmids pIJ6566 and pIJ6584 and introduction into S. coelicolor M512 to generate S. coelicolor W73 and W74, respectively

A 5,188 bp SexA1, EcoRI DNA fragment from the C73 cosmid\(^{1}\) containing the mmfLHP genes as well as other genes was introduced into the integrative plasmid pSET152 to generate pIJ6582. A 1.7 kb XhoI, BglII DNA fragment of the cosmid containing the mmfL gene was introduced into pSET152 to generate pIJ6566. Digestion of pIJ6582 with Scal and XbaI generated a 3,134 bp fragment that was ligated with pIJ6566 (digested with Scal and XbaI) to generate pIJ6584 containing just the mmfLHP putative operon and the upstream native promoter (Fig. S1).
Fig. S1. Procedure for construction of the integrative plasmids pIJ6566 and pIJ6584 containing mmfL and mmfLHP, respectively.
S. coelicolor M512 does not carry the plasmid SCP1 and consequently cannot produce the methylenomycins. The plasmids pJ6566 and pJ6584 were introduced into S. coelicolor M512 via conjugation from E. coli ET12567/pUZ8002 to generate S. coelicolor W73 and W74, respectively, following the procedure described by Gust et al.2

The presence of the mmtL and mmtLHP genes in the S. coelicolor W73 and W74 strains, respectively, was confirmed by PCR using genomic DNA isolated from the strains (Fig. S2). The primers used were as follows:

A: mmtL_sc_for 5'-TGTTCGACGGCTGCCCTCC-3'
B: mmtL_sc_rev 5'-AGGCCACCGCCTTCGTCTG-3'
C: mmtL_sc_for_b 5'-GAAAGGCGTGGGTGTTG -3'
D: mmpP_sc_rev 5'-TACCCTCGGGAAGGTATG -3'

![Fig. S2. PCR analysis of the S. coelicolor M512, W73, W74 and W81 strains.](image)

1.1.2 Generation of plasmid pCC002 and introduction into M145 to generate S. coelicolor W81

The mmtLHP genes within the C73-787 cosmid (derived from the S. coelicolor SCP1 ordered cosmid library (3)) were replaced with the apr apramycin resistance gene (previously known as aac(3)IV) using a PCR-targeting based gene replacement method to generate pCC002.(4) The apramycin resistance gene was amplified from pW60 using the forward primer:

5' TCCCCGTACGAGGACACACCGGAGATGTAGCGCCCCCTATGATTCCGGGGATCCGTCGACC-3',

and the reverse primer:

5' AAAAGTGGAGGCAGCACGGCTACCGCGCGAGGATTCATGTAGGCTGGAGCTGCTTC-3'.

Underlined sequences correspond to the 20 nucleotides P1 and 19 nt P2
sequences described in the procedure developed by Gust et al. (4) Correct replacement of the \textit{mmfLHP} genes with \textit{apr} on pCC002 was confirmed by PCR using the primers C and D (section 1.1.1) which are complimentary to regions upstream and downstream, respectively, of the region of the C73-787 cosmid replaced.

The pW60 plasmid was constructed by amplifying the \textit{apr} gene from a gel-purified \textit{HindIII} digestion fragment of the plasmid pSPM83 (kindly provided by Dr Jean-Luc Pernodet, CNRS, Universite Paris-Sud 11, Orsay, France) using the primers C and D (section 1.1.1) which are complimentary to regions upstream and downstream, respectively, of the region of the C73-787 cosmid replaced. The pW60 plasmid was constructed by amplifying the \textit{apr} gene from a gel-purified \textit{HindIII} digestion fragment of the plasmid pSPM83 (kindly provided by Dr Jean-Luc Pernodet, CNRS, Universite Paris-Sud 11, Orsay, France) using the primers: 5'- TGTACGGCCACAGAGATGATGTGCAC-3' and 5'- ATTCGGGATCCGTCGAC

The six underlined nucleotides in bold correspond to \textit{HindIII} restriction sites. The underlined sequences (not bold) correspond to the 20 nt P1 and 19 nt P2 sequences described above. Bold sequences are complementary to sequences of pSPM83 and the nucleotides that are not in bold or underlined correspond to flip recombinase target (FRT) sequences. The gel-purified PCR product was cloned into pGEM-T-Easy (Promega) to generate pW60. The pCC002 plasmid was introduced into \textit{S. coelicolor} M145 via conjugation from \textit{E. coli} ET12567/pUZ8002 to generate \textit{S. coelicolor} W81 following the procedure described by Gust et al. (4).

1.2 Sequence comparison of \textit{S. coelicolor} and \textit{S. avermitilis} AHFCA gene clusters

\begin{center}
\begin{tabular}{lcl}
Protein & Proposed function & Homologue (% identity / % similarity) \\
MmfR & DNA binding receptor protein & SAV2270 (47 / 64) \\
MmfL & Butenolide synthase & AvaA (47 / 64) \\
MmfH & Flavin dependent dehydrogenase & SAV2267 (44 / 58) \\
MmfP & Phosphatase & SAV2266 (36 / 51) \\
MmyR & DNA binding receptor protein & SAV2268 (38 / 54) \\
\end{tabular}
\end{center}

\textbf{Fig. S3.} Comparison of the genetic organization of the known \textit{S. coelicolor} and putative \textit{S. avermitilis} AHFCA gene clusters. Sequence identities / similarities between each of the encoded proteins are tabulated.
1.3 Inducing activity contained in *E. coli* BL21 / pCC001 organic extract

**Fig. S4.** Induction of methylenomycin production by organic extract of *E. coli* BL21 / pCC001 culture supernatant. Right panel: methylenomycin production was induced in the MMF and methylenomycin non-producing W81 strain grown on the plug by addition of an organic extract corresponding to 15 mL of culture supernatant. Left panel: control experiment in which extract of *E. coli* BL21 culture supernatant was added.
2. Analytical data

2.1 UV/Vis Spectroscopy

The UV/Vis spectroscopic properties of MMFs 1-5 were obtained from DAD data (Fig. S5).

**Fig. S5.** HPLC-DAD analysis of MMFs in organic extracts of *S. coelicolor* W74. Top panel: UV-Vis absorbance spectrum of MMF2 showing a $\lambda_{\text{max}}$ at 250 nm resulting from the common furan core in the MMFs. Middle panel: 2-D UV/Vis chromatogram showing that the five MMFs have identical UV-Vis absorbance spectra. Bottom panel: 1-D UV/Vis chromatogram showing that all five MMFs absorb at 250 nm.
2.2 Mass Spectrometry data

LC-MS/MS data for MMF2 from *S. coelicolor* W74 and *E. coli* BL21Star(DE3) / pCC001 are shown in Figs. S6 and S7.

**Fig. S6.** LC-MS/MS data of MMF2 from *S. coelicolor* W74 (positive ion mode detection)

**Fig. S7.** LC-MS/MS data of MMF2 from *E. coli* BL21 star / pCC001 (positive ion mode detection)
Comparison of MS$^3$ data for MMF2, MMFs 1 and 4, MMFs 3 and 5 is shown in Fig. S8.

Fig. S8. Tandem MS data in negative ion mode detection of MMF2, MMF1 and 4 (which have a common tandem MS pattern), MMF3 and 5 (which have a common tandem MS pattern)
High-resolution MS data for MMFs 1-5 is shown in Figs. S9-S13

Fig. S9. ESI-TOF-MS spectrum of MMF1 from W74 in negative ion mode

Fig. S10. ESI-TOF-MS spectrum of MMF2 from W74 in negative ion mode
Fig. S11. ESI-TOF-MS spectrum of MMF3 from W74 in negative ion mode

Calculated m/z for $\text{C}_{11}\text{H}_{15}\text{O}_{4}$ = 211.0976 (4.25 ppm error)

Fig. S12. ESI-TOF-MS spectrum of MMF4 from W74 in negative ion mode

Calculated m/z for $\text{C}_{10}\text{H}_{13}\text{O}_{4}$ = 197.0819 (1.98 ppm error)
Fig. S13. ESI-TOF-MS spectrum of MMF5 from W74 in negative ion mode

Calculated m/z for C$_{11}$H$_{15}$O$_4$ = 211.0976 (0.64 ppm error)
LC-MS data for incorporation of labeled precursors into MMFs 1-5 is shown in Figs. S14-S18.

**Fig. S14.** LC-MS data for the incorporation of d10-leucine into MMF1 (positive ion mode detection) and structural interpretation
**Fig. S15.** LC-MS data for the incorporation of d₇-butyric acid into MMF2 (positive ion mode detection) and structural interpretation
Fig. S16. LC-MS data for the incorporation of d₈-valine into MMF3 (positive ion mode detection) and structural interpretation
**Fig. S17.** LC-MS data for the incorporation of d5-propionic acid into MMF4 (positive ion mode detection) and structural interpretation.
**Fig. S18.** LC-MS data for the incorporation of d$_7$-butyric acid into MMF5 (positive ion mode detection) and structural interpretation

### 2.3 NMR Spectroscopy data

$^1$H, COSY, HSQC, HMBC and $^{13}$C NMR spectra for MMF1 (3) are shown in Figs. S19-S23.
Fig. S19. $^1$H-NMR spectrum of 3
Fig. S20. COSY spectrum of 3
Fig. S21. HSQC spectrum of 3
Fig. S22. HMBC spectrum of 3
Fig. S23. $^{13}$C-NMR spectrum of 3
NMR data for MMF2 are shown in Figs. S24-S26.

![Chemical structure](image)

<table>
<thead>
<tr>
<th></th>
<th>$^1H$[^a]</th>
</tr>
</thead>
<tbody>
<tr>
<td>H-5</td>
<td>7.25 (1H, s)</td>
</tr>
<tr>
<td>COOH</td>
<td>-</td>
</tr>
<tr>
<td>H-7</td>
<td>4.53 (2H, s)</td>
</tr>
<tr>
<td>CH$_2$OH</td>
<td>3.52 (1H, br m)</td>
</tr>
<tr>
<td>H-1’</td>
<td>2.94 (2H, d, 7.3 Hz)</td>
</tr>
<tr>
<td>H-2’</td>
<td>1.50 (2H, m)</td>
</tr>
<tr>
<td>H-3’</td>
<td>0.92 (3H, t, 7.4 Hz)</td>
</tr>
</tbody>
</table>

[^a] $^1$H NMR (700 MHz, CD$_2$Cl$_2$) data for 4 referenced to residual CH$_2$Cl$_2$ as internal standard.

**Fig. S24.** Chemical structure and $^1$H-NMR assignments for MMF2 isolated from *S. coelicolor* W74. Selected correlations observed in the COSY spectrum are indicated by bold lines. [a] $^1$H NMR (700 MHz, CD$_2$Cl$_2$) data for 4 referenced to residual CH$_2$Cl$_2$ as internal standard.
Fig. S25. $^1$H-NMR spectrum for 4
Fig. S26. COSY spectrum for 4
3. References:


