Supporting Information

for

Sulfation and amidinohydrolysis in the biosynthesis of giant linear polyenes

Hui Hong, Markiyan Samborskyy, Katsiaryna Usachova, Katharina Schnatz, Peter F. Leadlay*

Address: Department of Biochemistry, University of Cambridge, Cambridge CB2 1GA, U.K.

Email: Peter Leadlay* - pfl10@cam.ac.uk
*Corresponding author

Experimental part and additional Figures and Schemes

1. Supplementary methods

1.1. Bacterial strains and culture conditions

*Streptomyces malaysiensis* DSM4137 was obtained from the Leibnitz Institute - DSMZ. The strain was originally deposited in the collection in connection with a Hoechst patent application, so it is not listed in the DSMZ online catalogue, but it can be obtained by specific request via the European Patent Office quoting EP 0360130 A2. The strain has been referred to in previous publications from this laboratory as *Streptomyces violaceusniger* because its 16S rRNA sequence places it within the *S. violaceusniger* clade; and on the basis of its recently-obtained whole genome sequence this has been confirmed, and the strain can now be identified as a subspecies of *Streptomyces malaysiensis* (M.S., H.H. and P.F.L., ms in preparation). *Streptomyces mediocidicus* ATCC23936 was obtained from the American Type Culture Collection through LGC Standards (Tedddington, Middlesex, U.K.). DSM4137 wild type and its mutants and *S. mediocidicus* were maintained on SFM agar (2% soya flour (Arkasoy), 2% D-mannitol, 2% agar) at 30 °C. For clethramycin, desulfoclethramycin,
mediomycin A and mediomycin B production, strains of DSM4137 and *S. mediocidicus* were cultured on fermentation liquid medium TSBY (3% TSB (Tryptic Soy Broth), 10.3% sucrose, 0.5% yeast extract) at 30 °C and 200 rpm in a rotary incubator and harvested after 2–3 days. *E. coli* strains were grown in Luria-Bertani (LB) broth (10% tryptone, 5% yeast extract, 10% NaCl) or agar (10% tryptone, 5% yeast extract, 10% NaCl, 2% agar) at 37 °C with appropriate antibiotic selection (kanamycin, at 50 µg mL⁻¹).

### 1.2. Materials, DNA isolation and manipulation

Bacterial strains, plasmids and oligonucleotides (Eurofins, Sigma) used in this work are summarised in Tables S1, S2 and S3, respectively. Restriction endonucleases were purchased from New England Biolabs (NEB). T4 DNA ligase and alkaline phosphatase were purchased from Fermentas. All chemicals were from Sigma-Aldrich. Liquid cultures for isolation of genomic DNA were grown in tryptone soya broth (Difco). DNA isolation and manipulation in *Streptomyces*, and *E. coli* were carried out using standard protocols [1,2]. PCR amplifications were carried out using Phusion® High-Fidelity DNA Polymerase (NEB). *E. coli* BL21(DE3) (Novagen) was used for protein expression.

### 1.3. Metabolite analysis and desulfoclethraramycin isolation

For small-scale analysis, DSM4137 and *S. mediocidicus* were grown in liquid TSBY medium for 2–3 days. 1 mL samples of culture broth were centrifuged at 20,000g for 15 min. The mycelia pellets were then extracted with 1 mL of methanol at 60 °C for 2 hours. The mixture was spun down and the clear methanol extract was evaporated to dryness and dissolved in 200 µL of methanol. 10 µL of the extract was analyzed by LC–UV–MS. LC–UV–MS analyses were performed on a HPLC (Agilent Technologies 1200 series) coupled to a Thermo Fisher LTQ mass spectrometer fitted with an electrospray ionization (ESI) source. The methanol extracts were loaded onto a Prodigy 5µ C18 column (4.6 × 250 mm, Phenomenex), and the samples were eluted using MQ containing 5 mM ammonium acetate (A) and acetonitrile (B) at a flow rate of 0.7 mL min⁻¹. The elution gradient for both extracts was 5% to 35% B over 10 min, 35% to 65% B over 30 min. The elution was monitored at 360 nm as well as diode array detector (DAD). The mass spectrometer was run in positive ionization mode, scanning from *m/z* 200 to 2000 in full scan mode. MS/MS analysis were performed on [M + H]⁺ ions with a normalized collision energy of 30%. High-resolution mass analysis was carried out on a Thermo Fisher Orbitrap mass spectrometer with resolution set up at 60 K.
For desulfoclethramycin production and isolation, four 1 L Erlenmeyer flasks with spirals, containing 200 ml TSBY medium, were inoculated with 5 mL 2-day TSBY seed culture of *S. malaysiensis* DSM4137 and incubated at 30 °C, 200 rpm. After 2 days, the broth was centrifuged at 9,500 rpm for 30 min. The pellet was resuspended in methanol and incubated at 60 °C for 2 h. The suspension was centrifuged at 2500g for 10 min at room temperature and the supernatant, which showed a significant yellow colour, was transferred to a round bottom flask. The solvent was evaporated and the water was removed by lyophilisation. The residue was dissolved in MeOH and purified by preparative HPLC (Agilent 1200) fitted with a Luna C18 column (100Å, 21.20 × 250 mm, Phenomenex). Compounds were eluted with 5 mM ammonium acetate (A) and MeOH (B) with a linear gradient of 5% to 60% B over 10 min, 60% B to 100% B over 20 min at a flow rate of 20 mL/min. Fractions were collected, and checked by MS analysis. Fractions containing desulfoclethramycin were combined. After removing methanol under reduced pressure, samples were lyophilized and kept at −20 °C before use.

1.4. Sulfotransferase gene knock-out in *S. malaysiensis* DSM4137

The knock-out of the sulfotransferase gene smala2697 in *S. malaysiensis* DSM4137 was performed by introducing an in-frame deletion. The construction of the deletion plasmid pYH7-smala2697 was achieved by i) PCR amplification of around 2 kbp DNA fragments upstream and downstream of smala2697, using pairs of primers smala2697-L1/L2 and smala2697-R1/R2, respectively, from genomic DNA of *S. malaysiensis* DSM4137; ii) *NdeI* restriction digestion of the cloning vector pYH7, followed by treatment with antarctic phosphatase AnP, and agarose gel purification; iii) ligation of the two fragments and the digested pYH7 plasmid by the isothermal assembly method as described previously [3], with a 50 °C for 60 min incubation step; iv) transformation of pYH7-sl/fV in *E. coli* DH10B; v) plasmid isolation, and PCR and sequencing confirmation of the inserted deletion fragment, using primers smala2697-CP1, smala2697-CP2, NdeI-L, and NdeI-R.

The pYH7-smala2697 construct was then introduced into *S. malaysiensis* DSM4137 by intergeneric conjugation. Freshly grown *E. coli* ET12567-pUZ8002-pYH7-smala2697 cultures at A600 ~ 0.4–0.5 were thoroughly washed, to remove antibiotics, mixed with 2–3 days old *Streptomyces* mycelium, and plated on SFM agar. Following 20–22 h of incubation at 30 °C, plates were overlaid with nalidixic acid (25 μg mL⁻¹) and apramycin (5 μg mL⁻¹). Single *Streptomyces* colonies from these plates were streaked onto SFM agar containing 50 μg mL⁻¹ apramycin, to confirm they had undergone antibiotic selection. Following further several rounds
of incubation in a non-selective TSBY medium, mutants were screened for Apr$^S$ phenotype, by patching of single colonies onto both SFM agar and SFM agar containing apramycin (50 $\mu$g mL$^{-1}$). To identify the mutants in which a double cross-over event had occurred, their genomic DNA was amplified with the smala2697-CP1/CP2 primer pair, and the resulting DNA fragments of the correct length (0.7 kb) were verified by sequencing.

1.5. Complementation of amidinohydrolases into S. malaysiensis DSM4137

The amidinohydrolase medi4948 was amplified by PCR, using as template genomic DNA of S. mediocidicus, and inserted into vector pIB139 via NdeI and EcoRV restriction sites to yield pIB139 medi4948.

The amidinohydrolase amh_A828 was amplified by PCR, using as template genomic DNA of Streptomyces olivaceus Tü4018, and inserted into vector pIB139 via NdeI and EcoRV restriction sites to yield pIB139 amh828.

The construct was then introduced by conjugation into S. malaysiensis DSM4137. The donor strain was E. coli ET12567/pUZ8002, and conjugation was carried out on 20 mL of SFM plates. After incubating at 30 °C for 20 hours, exconjugants were selected with 50 $\mu$g mL$^{-1}$ apramycin and 25 $\mu$g mL$^{-1}$ nalidixic acid. Single colonies from this plate were transferred to a SFM plate containing 50 $\mu$g mL$^{-1}$ apramycin to double check for antibiotic resistance. The patch from the confirmation plate was then inoculated into TSBY liquid culture containing 50 $\mu$g mL$^{-1}$ apramycin for production of metabolites.

1.6. Complementation of the sulfotransferase deletion mutant of S. malaysiensis DSM4137 using cloned slf genes

The in trans complementation of the S. malaysiensis DSM4137 sulfotransferase deletion mutant Δsmala2697 was done using the native smala2697, as well as sulfotransferase medi5536 from S. mediocidicus ATCC23936. Genes smala2697 and medi5536 were PCR amplified from genomic DNA, using primer pairs smala2697 _com_F/R and medi5536 _com_F/R, respectively. The cloning vector pIB139 was digested with NdeI and Eco321 and gel purified. The smala2697 and medi5536 PCR fragments were ligated by the isothermal assembly method with the digested pIB139 plasmid, to yield plasmids pIB139-smala2697 and pIB139 medi5536, respectively. The latter plasmids were used to transform E. coli DH10B, the plasmids were isolated, and their identity confirmed by PCR and sequencing using primers.
pIB-seqF and pIB-seqR. The constructs were then introduced by conjugation into the \( \Delta \text{smala2697} \) mutant. The conjugation procedure was as described in 1.4.

### 1.7. Protein expression and purification

The sulfotransferase gene \( \text{smala2697} \) was amplified by PCR, using genomic DNA of \( S. \text{malaysiensis} \) DSM4137 as template, and inserted into vector pET28a via \( NdeI \) and \( HindIII \) restriction sites to yield pET28a-smala2697.

The three amidinohydrolase genes \( \text{medi0234}, \text{medi2865} \) and \( \text{medi4948} \) were individually amplified by PCR, using genomic DNA of \( S. \text{mediocidicus} \), and inserted into vector pET28a via \( NdeI \) and \( HindIII \) restriction sites to yield pET28a-mediated0234, pET28amedi2865, and pET28a-mediated4948. The identities of the plasmids were confirmed by DNA sequencing.

The plasmids were then used to transform \( E. \text{coli} \) BL21(DE3) for protein expression. A single colony was inoculated into 10 mL of LB medium containing 50 μg ml\(^{-1}\) kanamycin and grown overnight at 37 °C, 250 rpm. An aliquot (1 mL) was retained for preparation of a glycerol stock and the remaining culture was inoculated into 1 L LB medium containing 50 μg mL\(^{-1}\) kanamycin and incubated at 37 °C, 200 rpm until \( A_{600} \) reached 0.6 before addition of 400 μL of 1 M isopropyl-\( \beta \)-D-thiogalactopyranoside (IPTG) and incubation at 22 °C overnight to induce protein expression. Cells were harvested by centrifugation at 4000g for 10 min, resuspended in lysis buffer (20 mM Tris-HCl, pH 7.8, 0.5 M NaCl, 10 mM imidazole) and lysed by sonication. The total lysate was centrifuged at 14,000 x g for 40 min, and the supernatant was loaded onto a His-Bind column (1 mL bed volume), which had been precharged with nickel ions and equilibrated with lysis buffer. The column was washed with 10 column volumes of lysis buffer. Bound proteins were then eluted with a step gradient of increasing imidazole concentration (40, 80, 100, 150, 200, 250 and 500 mM in binding buffer). The protein solutions were concentrated, and further purified by gel filtration on an ÄKTAmo 16/60 Superdex 200 Prep Grade column. The mobile phase contained 100 mM potassium phosphate, pH 7.4. Fractions containing protein of the expected size were pooled and concentrated using Amicon Ultra-4 concentrators (Millipore) fitted with either 10 kDa or 30 kDa filter. All purification steps were carried out at 4 °C. The purity of the protein was examined by 4–12% Bis-Tris Gel (Novex)
analysis and the concentration of the protein was measured by Bradford assay using bovine serum albumin as a standard.

1.8. In vitro activity assays

Amidinohydrolase activity with desulfoclethramycin as substrate

Each reaction mixture (50 µL) contained 10 µM purified candidate amidinohydrolase Medi2865 (or Medi4948, or Medi0234), 1 mM purified desulfoclethramycin in 100 mM potassium phosphate buffer pH 7.5. After incubation at 37 °C for 1.5 h, 10 µL of the reaction mixture was taken out, and mixed with 50 µL of methanol. The sample was clarified by centrifugation and then analyzed by HPLC–UV–MS.

Amidinohydrolase activity was also performed after pre-incubation of protein with various metal ions. Each reaction mixture (50 µL) contained 10 µM purified Medi2865 (or Medi4948, or Medi0234), 0.5 mM MnCl₂ (or ZnCl₂, or MgCl₂) in 50 mM Tris-HCl buffer pH 8.4. After incubation at 37 °C for 30 min, purified desulfoclethramycin was added to a final concentration of 1 mM, and the reaction was allowed to continue at 37 °C for 1 to 3 h. 10 µL of the reaction mixture was taken, mixed with 50 µL methanol, and after centrifugation analyzed by HPLC–UV–MS.

Amidinohydrolase activity with L-arginine (or 4-guanidinobutyric acid, 3-guanidinopropionic acid, and 4-guanidinobutyramide) as substrate

Each reaction mixture (25 µL) contained 5 µM purified Medi2865 (or Medi4948, Medi0234), 0.5 mM MnCl₂ in 50 mM Tris-HCl buffer pH 9.0. After incubation at 37 °C for 30 min, L-arginine (or 4-guanidinobutyric acid, 3-guanidinopropionic acid, or 4guanidinobutyramide) was added to a final concentration of 1 mM, and the reaction was allowed to continue at 37 °C for 1 h. 12.5 µL of the reaction mixture was taken out, mixed with 10 µL of MQ and 3 µL of 1 M HCl, and analyzed by HPLC–MS with a Synergi 4 Polar-RP column (4.6 × 250 mm, Phenomenex). Compounds were eluted using an isocratic gradient of 5% CH₃CN with 0.1% trifluoroacetic acid (TFA) for 10 min at a flow rate of 1.0 mL min⁻¹. The mass spectrometer was set to full scan mode (from m/z 100 to 1000), and MS² on [M + H]⁺ ions at m/z 175.2, 145.2, 146.2 and 132.2 (for arginine, 4-guanidinobutyramide and 4-guanidinobutyric acid and 3-guanidinopropionic acid respectively) was performed with a normalized collision energy of 20%.
**Sulfotransferase activity with desulfoclethramycin as substrate**

Each reaction mixture (30 µL) contained 10 µM purified SMALA_2697, 1.5 mM 3’phosphoadenosine 5’-phosphosulfate (PAPS), 1 mM purified desulfoclethramycin in 100 mM potassium phosphate buffer pH 7.5. After incubation at 37 °C for 2 h, 10 µL of the reaction mixture was taken out, and mixed with 50 µL of methanol. After centrifugation, the sample was analyzed by HPLC–UV–MS.

**Sulfotransferase activity with mediomycin B as substrate**

Mediomycin B was generated in situ using desulfoclethramycin and amidinohydrolase Medi4948. Each reaction mixture (50 µL) contained 10 µM purified Medi4948, 1 mM purified desulfoclethramycin in 100 mM potassium phosphate buffer pH 7.5. After incubation at 37 °C for 1.5 h, 10 µL was taken out to check by LC–MS to make sure that desulfoclethramycin was fully converted to mediomycin B. Then purified SMALA_2697 and PAPS cofactor were added to the reaction mixture at a final concentration of 10 µM and 1.5 mM respectively, and the reaction was allowed to continue at 37 °C for 2 h. 10 µL of the reaction mixture was taken out, and mixed with 50 µL of methanol. After centrifugation, the sample was analyzed by HPLC–UV–MS.

2. Supplementary Scheme and Figures

![Scheme S1: Sulfonated natural products.](image-url)
A) desulfoclethramycin

1138.7143 $\text{C}_{63}\text{H}_{100}\text{N}_{3}\text{O}_{15}^+$

1138.7149

0.5 ppm

B) clethramycin

$\text{(C}_{63}\text{H}_{99}\text{N}_{3}\text{O}_{18}\text{NaS}^+$

1240.6537

0.3 ppm 1240.6533

e: loss of benzene via rearrangement
Figure S1: A) High-resolution MS (Orbitrap) analysis of clethramycin and desulfoclethramycin from S. malaysiensis DSM4137. B) High-resolution MS/MS analysis of desulfoclethramycin at m/z [M+H]+: 1138.8. C) UV spectrum of desulfoclethramycin.
Figure S2: A) Sequence alignment of ureohydrolases. The sequences of seventeen ureohydrolases [AMH_A828: amidinohydrolase from Streptomyces olivaceus Tü4018 (A828); AMH_A828: amidinohydrolase from Saccharomonospora azurea (caesia) DSM 43044 (A821); GbuA_PA: guanidinobutyrase from Pseudomonas aeruginosa; GpuA_PA: guanidinopropionase from Pseudomonas aeruginosa; PAH_SC: proclavaminic acid amidino hydrolase (PAH) from Streptomyces clavuligerus; Agm_BT: agmatinase from Burkholderia thailandensis; Agm_DR: agmatinase from Deinococcus radiodurans; Agm_CD: agmatinase from Clostridium difficile; Agm_TV: agmatinase from Thermoplasma. volcanium; ARG_BC: arginase from Bacillus caldovelox; ARG_TT: arginase from Thermus thermophilus; AMH_SA: amidinohydrolase from Streptomyces aizunensis; medi0234, medi2865 and medi4948: amidinohydrolases from Streptomyces mediocidicus; SMALA_0333 and SMALA_7636: amidinohydrolases from Streptomyces malaysiensis DSM4137] are aligned using MultAlign. Three well-conserved
sequences (xGGDH, DAHxD, and SxDxDxxDpxxxP) in most of the ureohydrolases are indicated by black boxes. The metal binding sites are indicated with asterisks, guanidino ligands with black triangles. **B) Phylogenetic tree of amidinohydrolases medi4948 and homologues.**

Construction of the phylogenetic tree was performed via the www.phylogeny.fr website, using default settings.

### A)

| Sequences | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 | 110 | 120 |
|-----------|----|----|----|----|----|----|----|----|----|-----|-----|-----|
| SMALA_2697|    |    |    |    |    |    |    |    |    |     |     |     |
| Sautol_AQA11921.1|    |    |    |    |    |    |    |    |    |     |     |     |
| Smelan_SEB92269.1|    |    |    |    |    |    |    |    |    |     |     |     |
| Smed_medi15536|    |    |    |    |    |    |    |    |    |     |     |     |
| Kmed9733_WP_035796292.1|    |    |    |    |    |    |    |    |    |     |     |     |
| Sblast_BAW35627.1|    |    |    |    |    |    |    |    |    |     |     |     |
| S.Rtd22_NE_CP015726.1|    |    |    |    |    |    |    |    |    |     |     |     |
| S.RK95-74_BAW35600.1|    |    |    |    |    |    |    |    |    |     |     |     |
| S.PRh5_EXU69913.1|    |    |    |    |    |    |    |    |    |     |     |     |
| Sviol_AEM87304.1|    |    |    |    |    |    |    |    |    |     |     |     |
| Srapa_AGP57770.1|    |    |    |    |    |    |    |    |    |     |     |     |
| Siran_CDR09769.1|    |    |    |    |    |    |    |    |    |     |     |     |
| Shygr_AQW50862.1|    |    |    |    |    |    |    |    |    |     |     |     |
| S.DSM7348_GRF0413|    |    |    |    |    |    |    |    |    |     |     |     |
| SMALA_0226|    |    |    |    |    |    |    |    |    |     |     |     |
| Smed_medi1571|    |    |    |    |    |    |    |    |    |     |     |     |

### B)

| Sequences | 70 | 80 | 90 | 100 | 110 | 120 |
|-----------|----|----|----|-----|-----|-----|
| SMALA_2697|    |    |    |     |     |     |
| Sautol_AQA11921.1|    |    |    |     |     |     |
| Smelan_SEB92269.1|    |    |    |     |     |     |
| Smed_medi15536|    |    |    |     |     |     |
| Kmed9733_WP_035796292.1|    |    |    |     |     |     |
| Sblast_BAW35627.1|    |    |    |     |     |     |
| S.Rtd22_NE_CP015726.1|    |    |    |     |     |     |
| S.RK95-74_BAW35600.1|    |    |    |     |     |     |
| S.PRh5_EXU69913.1|    |    |    |     |     |     |
| Sviol_AEM87304.1|    |    |    |     |     |     |
| Srapa_AGP57770.1|    |    |    |     |     |     |
| Siran_CDR09769.1|    |    |    |     |     |     |
| Shygr_AQW50862.1|    |    |    |     |     |     |
| S.DSM7348_GRF0413|    |    |    |     |     |     |
| SMALA_0226|    |    |    |     |     |     |
| Smed_medi1571|    |    |    |     |     |     |

- **A)** Phylogenetic tree of amidinohydrolases medi4948 and homologues.
- **B)** Phylogenetic tree of amidinohydrolases medi4948 and homologues.
| PROTEIN  | PRIMARY SEQUENCE                                                                 |
|---------|----------------------------------------------------------------------------------|
| Actin_AGS77324.1 | -DDVFAEQR ALSYETLLEE PEKELILRLE F-----GVEP -HRTWLDASI AHDL------ |
| S.CDSM7348.ORF0413 | -SRPDADFVSLSSYEGGSL PERETLRAH H-----GVEP -LPEWLAGR ALLL------ |
| Smed_medi1571     | -AEVPAPQRT ALAYEYLLDL PEEELSLRAL F-----GVEP -LPEWLAGR ALLL------ |
| MtubStf1_PDB_2ZQ5_A | RARYDSAQFY DDDYHDILAD PGTVIADYHR HFG---LTLSD- EEARQMTT VHAE--- |
| MtubStf3_CCP45048.1 | REVLDFPRFE ERVEYDLGGD PEQRLALIQ HLG---LGFDE CYLPIRQVL ADNH--- |
| MavStf9_PDB_226V_A  | RDRDFNADVL TVREYILRAD PAFVRIEVE FLG------LGQDRG QVA---AV ANCTLRRM |
| Sargent_AGU42411.1 | -AGLPFLQL HLSDYAVVAG PVPLTRFRGR F------VGLAE- PQRVWAVR GQVD- |
| Telg2_PDB_3MGC_A    | SDFRPNADVL TVREYILKGD PAFVRIEVE FLG------LGQDRG QVA---AV ANCTLRRM |
| Telg3_AJC60996.1   | RDRDFNADVL TVREYILKGD PAFVRIEVE FLG------LGQDRG QVA---AV ANCTLRRM |
| Telg4_PDB_3NIB_A    | RDRDFNADVL TVREYILKGD PAFVRIEVE FLG------LGQDRG QVA---AV ANCTLRRM |
| Stoyoc_PDB_20V8     | HESFFNAAVL PEGERAVKE FDLE------LGQDRG QVA---AV ANCTLRRM |
| Actin_AGS77324.1    | RGSFDPDDVL TVREYLRKGD PAGLRTVEFLG----LGQDRG QVA---AV ANCTLRRM |
| S.DSM7348.ORF0413 | -DDVFAEQR ALSYETLLEE PEKELILRLE F-----GVEP -HRTWLDASI AHDL------ |
| S.CDSM7348.ORF0413 | -SRPDADFVSLSSYEGGSL PERETLRAH H-----GVEP -LPEWLAGR ALLL------ |
| Smed_medi1571     | -AEVPAPQRT ALAYEYLLDL PEEELSLRAL F-----GVEP -LPEWLAGR ALLL------ |
| MtubStf1_PDB_2ZQ5_A | RARYDSAQFY DDDYHDILAD PGTVIADYHR HFG---LTLSD- EEARQMTT VHAE--- |
| MtubStf3_CCP45048.1 | REVLDFPRFE ERVEYDLGGD PEQRLALIQ HLG---LGFDE CYLPIRQVL ADNH--- |
| MavStf9_PDB_226V_A  | RDRDFNADVL TVREYILRAD PAFVRIEVE FLG------LGQDRG QVA---AV ANCTLRRM |
| Sargent_AGU42411.1 | -AGLPFLQL HLSDYAVVAG PVPLTRFRGR F------VGLAE- PQRVWAVR GQVD- |
| Telg2_PDB_3MGC_A    | SDFRPNADVL TVREYILKGD PAFVRIEVE FLG------LGQDRG QVA---AV ANCTLRRM |
| Telg3_AJC60996.1   | RDRDFNADVL TVREYILKGD PAFVRIEVE FLG------LGQDRG QVA---AV ANCTLRRM |
| Telg4_PDB_3NIB_A    | RDRDFNADVL TVREYILKGD PAFVRIEVE FLG------LGQDRG QVA---AV ANCTLRRM |
| Stoyoc_PDB_20V8     | HESFFNAAVL PEGERAVKE FDLE------LGQDRG QVA---AV ANCTLRRM |
| Actin_AGS77324.1    | RGSFDPDDVL TVREYLRKGD PAGLRTVEFLG----LGQDRG QVA---AV ANCTLRRM |
| S.DSM7348.ORF0413 | -DDVFAEQR ALSYETLLEE PEKELILRLE F-----GVEP -HRTWLDASI AHDL------ |
| S.CDSM7348.ORF0413 | -SRPDADFVSLSSYEGGSL PERETLRAH H-----GVEP -LPEWLAGR ALLL------ |
| Smed_medi1571     | -AEVPAPQRT ALAYEYLLDL PEEELSLRAL F-----GVEP -LPEWLAGR ALLL------ |
| MtubStf1_PDB_2ZQ5_A | RARYDSAQFY DDDYHDILAD PGTVIADYHR HFG---LTLSD- EEARQMTT VHAE--- |
| MtubStf3_CCP45048.1 | REVLDFPRFE ERVEYDLGGD PEQRLALIQ HLG---LGFDE CYLPIRQVL ADNH--- |
| MavStf9_PDB_226V_A  | RDRDFNADVL TVREYILRAD PAFVRIEVE FLG------LGQDRG QVA---AV ANCTLRRM |
| Sargent_AGU42411.1 | -AGLPFLQL HLSDYAVVAG PVPLTRFRGR F------VGLAE- PQRVWAVR GQVD- |
| Telg2_PDB_3MGC_A    | SDFRPNADVL TVREYILKGD PAFVRIEVE FLG------LGQDRG QVA---AV ANCTLRRM |
| Telg3_AJC60996.1   | RDRDFNADVL TVREYILKGD PAFVRIEVE FLG------LGQDRG QVA---AV ANCTLRRM |
| Telg4_PDB_3NIB_A    | RDRDFNADVL TVREYILKGD PAFVRIEVE FLG------LGQDRG QVA---AV ANCTLRRM |
| Stoyoc_PDB_20V8     | HESFFNAAVL PEGERAVKE FDLE------LGQDRG QVA---AV ANCTLRRM |
| Actin_AGS77324.1    | RGSFDPDDVL TVREYLRKGD PAGLRTVEFLG----LGQDRG QVA---AV ANCTLRRM |
Figure S3: Phylogenetic analysis of sulfotransferase genes in actinobacteria. SMALA_2697 from *S. malaysiensis* DSM4137, Sautol from *S. autolyticus* CGMCC0516, Smelan from *S. melanosphorafaciens* DSM40318, Smed_med5536 from *S. mediocidicus* ATCC 23936, Kmed from *Kitasatospora mediocidica* KCTC9733, Sblast from *S. blastmyceticus* NBRC12747, S.RTd22 from *Streptomyces* sp. RTd22, S.PRh5 from *Streptomyces* sp. PRh5 CCTCC2013487, Sviol from *S. violaceusniger* Tü 4113, Srapa from
S. rapamycinicus NRRL5491, Siran from S. iranensis HM 35, Shygr from S. hygroscopicus XM201, S.DSM7348 from Streptomyces sp. DSM7348 – are sulfotransferases (Slf) associated with a clethramycin/mediomycin cluster. S.RK95-74 from Streptomyces sp. RK9574 has been reported to contain a neomediomycin cluster [4]. SMALA_0226 and Smed_medi1571 are additional Slf genes in S. malaysiensis DSM4137 and S. mediocidicus ATCC 23936, respectively. Sargent is an Slf associated with the production of carbapenem MM4550 in S. argenteolus ATCC11009. MtubStf1 is an Slf involved in the production of sulfated trehalose glycolipids in Mycobacterium tuberculosis H37rv. MtubStf3 is an Slf involved in the production of sulfomenaquinone S881 in M. tuberculosis H37rv. MavStf9 is a Mycobacterium avium Slf of unknown function. Stoyoc is the StaL sulfotransferase in the teicoplanin A47934 biosynthetic pathway in S. toyocaensis. Actino is a tailoring Slf within the UK-68,597 glycopeptide biosynthetic cluster in Actinoplanes sp. ATCC. Teg12, Teg13, and Teg 14 are tailoring Slfs within the TEG cluster isolated from metagenomic DNA and predicted to produce a polysulfated teicoplanin-like glycopeptide. GenBank or PDB accession numbers are shown for each Slf sequence. A) **Sequence alignment of sulfotransferases.** The multiple alignment of amino acid sequences was performed using NCBI COBALT. Three well-conserved sequences of the 5′-PAP- and 3′-PAPS- binding motifs are indicated by a solid-line and dot-line boxes, respectively. B) **Phylogenetic tree of sulfotransferases.** Maximum-likelihood inference was performed in MEGA 6.06 on the WAG+I+G4+F substitution model with a bootstrap resampling of 1000 replicates. Nodes of the tree are labeled with bootstrap values, %. Slf genes associated with giant linear polyene biosynthetic gene clusters are shown shaded green. The Slf genes labelled with an asterisk denote known clethramycin, mediomycin or neomediomycin gene clusters.
Figure S4: In-frame deletion of sulfotransferase gene smala2697 in Streptomyces malaysiensis DSM4137. Lane 1: marker; Lane 2 and 3: PCR product from WT (1,642 bp) and Δsmala2697 respectively.

Figure S5: 4 - 12% Bis-Tris SDS-PAGE analysis of A) amidinohydrolases. Lane 1, protein standards; Lane 2, Medi4948; Lane 3, Medi2865; Lane 4, Medi0234. B) Sulfotransferase SMALA_2697. Lane 1, protein standards; Lane 2, SMALA_2697.
Figure S6: HPLC–UV–MS analysis of polyenes.  A) LC–UV (360 nm) trace of methanol extract from mycelium of DSM4137 wild type, showing the production of clethramycin (1a) and desulfoclethramycin (1b) at m/z 1240.8 ([M + Na]^+) and 1138.8 ([M + H]^+) , respectively. B) LC–UV (360 nm) trace of methanol extract from mycelium of DSM4137 complemented with the amidinohydrolase-encoding gene amh_A828 from the marginolactone desertomycin biosynthesis from *S. olivaceus* Tü4018 [5], showing that this enzyme was not able to convert 1a and 1b to their amino forms.

3. Supplementary Tables

| Strain          | Genotype/Characteristics                                                                 | Reference |
|-----------------|-----------------------------------------------------------------------------------------|-----------|
| *E. coli*       |                                                                                         |           |
| DH10B           | F* mcrA Δ(mrr-hsdRMS - mcrBC), □80lacZΔM15, Δ lacX74 recA1 endA1 araD139 Δ (ara leu)7697 galU galK rpsL nupG λ– | Invitrogen|
|                 | host for general cloning                                                                 |           |
| BL21(DE3)       | F* ompT hsdS8 (rB–, mB–) gal dcm (λDE3 lysogen)                                         | Invitrogen|
| ET12567 (pUZ8002) | (F dam-13::Tn9 dcm-6 hsdM hsdR recF143 zjj202::Tn10 galK2 gatT22 ara14 pacY1 xyl-5 leuB6 thi-1) Donor strain for conjugation between *E. coli* and *Streptomyces* | [6]       |
S. *malaysiensis*  
**DSM4137**

- WT-DSM4137: wild type strain, producing azalomycin, desulfo clethramycin, and clethramycin [7]
- □smala2697: smala2697 in-frame deletion mutant from DSM4137
- □smala2697::smala2697: smala2697 mutant complemented with plasmid pIB139-slFV
- □smala2697::medi5536: smala2697 mutant complemented with plasmid pIB139-slFM
- DSM4137::medi4948: DSM4137 complemented with plasmid pIB139medi4948
- DSM4137::amh-A828: DSM4137 complemented with plasmid pIB139amh828

*S. mediocidicus*  
**ATCC23936**

mediomycin- and clethramycin-producing strain [8]

---

**Table S2: Plasmids used in this work.**

| Plasmid            | Genotype/Characteristics                                      | Reference |
|--------------------|--------------------------------------------------------------|-----------|
| pYH7               | *E. coli*-Streptomyces shuttle vector                        | [9]       |
| pYH7-slFV          | *smala2697* gene disruption construct in which a 948 bp internal fragment of *smala2697* was deleted inframe | this work |
| pIB139             | *E. coli*-Streptomyces shuttle vector, *attP* (ΦC31), *int*, *PermE* |           |
| pIB139-medi4948    | Amidinohydrolase *medi4948* complementation plasmid          | This work |
| pIB139-amh828      | Amidinohydrolase *amh-A828* complementation plasmid          | This work |
| pIB139-smala2697   | Sulfotransferase *smala2697* complementation plasmid         | This work |
| pIB139-medi5536    | Sulfotransferase *medi5536* complementation plasmid          | This work |
| pET28a(+)          | *E. coli* protein expression vector                          | Invitrogen|
| pET28a-smala2697   | Sulfotransferase *SMALA_2697* protein expression construct with N-terminal His-tag based on pET28a(+) | this work |
| pET28a-medi0234    | Medi0234 protein expression construct with N-terminal His-tag based on pET28a(+) | this work |
| pET28a-medi2865    | Medi2865 protein expression construct with N-terminal His-tag based on pET28a(+) | this work |
| pET28a-medi4948    | Medi4948 protein expression construct with N-terminal His-tag based on pET28a(+) | this work |
Table S3: Oligonucleotide primers used in this work.

| Primer                  | Nucleotide sequence (5' to 3')                  | Restriction site(s) |
|-------------------------|-------------------------------------------------|---------------------|
| **primers for protein expression** |                                                 |                     |
| medi0234-fwd            | ttt CATATG ACGATCCCAGCCACGCCCAGGG              | NdeI               |
| medi0234-rev            | agctga AAGCTT TCAGCAGCGCACCTCTCCGT             | HindIII             |
| medi2865-fwd            | ttt CATATG AGCACCAACCCCCCCGGCCCCCG            | NdeI               |
| medi2865-rev            | agctga AAGCTT TCAGTCGCGAGCGGCGCGCGGA           | HindIII             |
| medi4948-fwd            | ttt CATATG ACGTTCCCAACGCACAAGAC              | NdeI               |
| medi4948-rev            | AAGCTT TCAGGGTTGTCCAGTGG                    | HindIII             |
| smala2697-fwd           | TTTCATATG GTCACCGAGAAGTTGACAT                | HindIII             |
| smala2697-rev           | AGCTGAAAGCTT CTACTGGTGAGCGGCCGGCAG            |                     |
| **primers for smala2697 gene in-frame deletion** |                                                 |                     |
| smala2697-L1            | TGATCAAGCCGCAATACCTCATATG                    |                     |
|                        | TGCTGTAAACCGGTCCGACATCTCTGTA                 |                     |
| smala2697-L2            | CAGCGAGTGAG GTTGCAGCCAGCAAAATGTCACTTT         |                     |
| smala2697-R1            | GTCGTCGCGCAC CTCACCTCGCTGGTGACAGTTG          |                     |
| smala2697-R2            | CCGCGCGGTCCGGATCCCCCGCATATG                  |                     |
|                        | TGACCGTCTTCACTCGGCGAGAAGCG                  |                     |
| **primers for PCR screening of deletion mutants** |                                                 |                     |
| smala2697-CP1           | ATGACCGTCTCGCTACAGGAG                      |                     |
| smala2697-CP2           | GCACACCAGATGATAGGCA                             |                     |
| NdeI-L                  | GCTCAGGGCGACAGTAC                             |                     |
| NdeI-R                  | CTGACCGGCAATCACCAAC                           |                     |
| **primers for smala2697 gene complementation** |                                                 |                     |
| smala2697_com_F         | AATCGTGCCGGTTGGATAGGATCCACATATGGTG            |                     |
|                        | AACCGAAGATGGAC                                 |                     |
|                        | ACAGGAAACACGCTATGACATGATACGAAATTGTA          |                     |
| smala2697_com_R         | TCCTACTGGTGAGCGGCCAG                           |                     |
| **primers for medi5536 gene complementation** |                                                 |                     |
| medi5536_com_F          | GTGCCGGTTGGATGATCCACATATGCACACGGAG            |                     |
|                        | AGTTGACCTTTTG                                 |                     |
|                        | TGACATGATTACGAATTCGATATCTCTACAGCCCTG         |                     |
| medi5536_com_R          | GCCCCGCCAGTG                                    |                     |
| **Primers for PCR screening and sequencing of complementation mutants** |                                                 |                     |
| pIB-seqF                | GATCTTGACGGCTGGCGGAG                               |                     |
| pIB-seqR                | CACTCATTAGGCACCCCGGAG                             |                     |
| **primers for medi4948 gene complementation** |                                                 |                     |
| medi4948_com_F          | ttt CATATG ACGTCCCCCAACGACAAGAC              | NdeI               |
|                         | agctga CATATG TCAGGGCTTGCCT                   | EcoRV              |
| **primers for amh_A828 gene complementation** |                                                 |                     |
| amh828_com_F            | ttt CATATG AGCGAGACACCCCCGAGTGCGA             | NdeI               |
|                         | agctga CATATG TCACTTGAGCGGAGAAGCGCA          | EcoRV              |

S22
Table S4a: Properties of genes within the clethramycin biosynthetic gene cluster of *Streptomyces* sp. DSM4137

| ORF     | Product            | % Identity/similarity | Species/FUNCTION          | Putative       | Database entry size (aa) |
|---------|--------------------|-----------------------|---------------------------|----------------|--------------------------|
| smala2696R | 311           | 96/98                | *Streptomyces iranensis* | LysR regulator | WP_044575084             |
| smala2697R | 347           | 84/90                | *Streptomyces rapamycinicus* | sulfotransferase AGP57770.1 |
| smala2698 | 882            | 95/97                | *Streptomyces violaceusniger* | LuxR regulator | AEM87305.1               |
| smala2699 | 552            | 93/95                | *Streptomyces rapamycinicus* | arginine oxidase | AGP57768.1               |
| smala2700R | 469           | 96/97                | *Streptomyces violaceusniger* | acyl-CoA ligase | AEM87307.1               |
| smala2701R | 207           | 93/97                | *Streptomyces violaceusniger* | TEII           | AEM87308.1               |
| smala2702 | 326            | 96/98                | *Streptomyces violaceusniger* | ABC transporter | AEM87309.1               |
| smala2703 | 477            | 84/88                | *Streptomyces violaceusniger* | ABC transporter | AEM87310.1               |
| smala2704 | 312            | 96/97                | *Streptomyces rapamycinicus* | ACP:malonyl transferase | AGP57763.1               |
| smala2705R | 199           | 94/96                | *Streptomyces rapamycinicus* | TetR regulator | AGP61306.1               |
| smala2706 | 304            | 95/97                | *Streptomyces rapamycinicus* | △,△-hydrolase | AGP61305.1               |
| smala2707R | 907           | 90/92                | *Streptomyces rapamycinicus* | TetR regulator | AGP58152.1               |
| smala2708 | 248            | 96/98                | *Streptomyces sp.* | short chain dehydrogenase | AGP58153.1               |
| smala2709R | 144           | 93/97                | *Streptomyces sp.* | glycosyltransferase | WP_030771662             |
| smala2710R | 167           | 99/100               | *Streptomyces scabiei* | bacteriocin biosynthesis protein | KFG10609.1               |
| smala2711R | 353           | 65/74                | *Streptomyces sclerotiorum* | lanthionine synthetase | WP_030569306             |
| smala2712R | 1012          | 64/75                | *Streptomyces scabiei* | lantibiotic dehydratase | KFF98219.1               |
| smala2713R | 54            | —                    | *Streptomyces sp.* | putative lantibiotic precursor | —                        |
| smala2714R | 400           | 74/85                | *Streptomyces sp.* | protein-Lisoaspartate Omethyltransferase | WP_045558093             |
| smala2715R | 287 | 57/69 | Streptomyces viridochromogenes taurine dioxygenase | AFV30253.1 |
| smala2716 | 155 | 83/92 | Streptomyces iakyrus NUDIX hydrolase | WP_033313602 |
| smala2717 | 389 | 73/83 | Streptomyces viridochromogenes Xre regulator | ELS55765.1 |
| smala2718R | 69 | — | — | — |
| smala2719R | 219 | 81/90 | Streptomyces sp. NTK937 DNA binding protein | KDQ67008. |
| smala2720 | 106 | — | — | — |
| smala2721 | 71 | — | — | — |
| smala2722R | 3902 | 91/93 | Streptomyces iranensis PKS CleA9 | CDR09758.1 |
| smala2723R | 5783 | 92/95 | Streptomyces violaceusniger PKS CleA8 | AEM87318.1 |
| smala2724R | 3206 | 93/96 | Streptomyces sp. PRh5 PKS CleA7 | EXU62495.1 |
| smala2725R | 7345 | 87/90 | Streptomyces violaceusniger PKS CleA6 | AEM87320.1 |
| smala2726R | 5250 | 91/94 | Streptomyces violaceusniger PKS CleA5 | AEM87321.1 |
| smala2727R | 1664 | 93/95 | Streptomyces sp. PRh5 PKS CleA4 | EXU62661.1 |
| smala2728R | 8599 | 91/93 | Streptomyces violaceusniger PKS CleA3 | AEM87323.1 |
| smala2729R | 3444 | 90/93 | Streptomyces iranensis PKS CleA2 | CDR09746.1 |
| smala2730R | 8207 | 90/93 | Streptomyces sp. PRh5 PKS CleA1 | EXU66032.1 |
| smala2731R | 514 | 91/96 | Streptomyces rapamycinicus membrane protein | AGP57745.1 |
| smala2732R | 166 | 96/98 | Streptomyces rapamycinicus membrane protein | AGP57744.1 |
| smala2733R | 185 | 68/82 | Streptomyces sp. 769 membrane protein | AJC60945.1 |
| smala2734 | 414 | 93/96 | Streptomyces rapamycinicus sensor kinase | AGP57742.1 |
| smala2735 | 186 | 96/99 | Streptomyces rapamycinicus LuxR regulator | AGP57741.1 |
| smala2736R | 851 | 93/95 | Streptomyces rapamycinicus LuxR regulator | AGP57740.1 |
| smala2737R | 253 | 93/94 | Streptomyces rapamycinicus TEII thioesterase | AGP57739.1 |

Putative functions of the encoded proteins were deduced from analyses with the BlastP program (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The % identity/similarity for the protein in the database with the highest end-to-end similarity is indicated. R designates a gene lying on the opposite strand. The entire genome sequence of *S. malaysiensis* DSM4137 has been deposited in GenBank where it can be accessed as *S. malaysiensis* Bioproject PRJNA396489, Biosample SAMN07427119.
| ORF      | Product identity/similarity | Species Function | Putative          | Database entry size (aa) |
|----------|-----------------------------|------------------|-------------------|--------------------------|
| orf5537  | 210                         | 54/66            | *Nocardiopsis* sp. | WP_053619846             |
| medi5536 | 349                         | 96/98            | *Streptomyces* blastmyceticus | BAM21064.1 |
| medi5535R| 942                         | 95/97            | *Streptomyces* blastmyceticus | BAM21065.1 |
| medi5534R| 553                         | 97/98            | *Streptomyces* blastmyceticus | BAM21066.1 |
| medi5533  | 468                         | 83/90            | *Streptomyces* violaceusniger | AEM87307.1 |
| medi5532  | 213                         | 80/86            | *Streptomyces* hygroscopicus | WP_030843507 |
| medi5531R | 314                         | 80/87            | *Streptomyces* hygroscopicus | WP_030843518 |
| medi5530  | 3833                        | 77/84            | *Streptomyces* violaceusniger | AEM87317.1 |
| medi5529  | 5702                        | 79/86            | *Streptomyces* rapamycinicus | AGP57754.1 |
| medi5528  | 3212                        | 82/89            | *Streptomyces* himastatinicus | EFL26042.1 |
| medi5527  | 7131                        | 79/85            | *Streptomyces* violaceusniger | AEM87320.1 |
| medi5526  | 5163                        | 78/85            | *Streptomyces* violaceusniger | AEM87321.1 |
| medi5525  | 1651                        | 79/87            | *Streptomyces* sp. PRh5 | EXU62661.1 |
| medi5524  | 8399                        | 77/83            | *Streptomyces* violaceusniger | AEM87320.1 |
| medi5523  | 3377                        | 76/83            | *Streptomyces* sp. PRh5 | EXU66033.1 |
| medi5522  | 8123                        | 76/83            | *Streptomyces* sp. PRh5 | EXU66032.1 |
| medi5521  | 518                         | 70/81            | *Streptomyces* rapamycinicus | AGP57745.1 |
| medi5520  | 151                         | 85/91            | *Streptomyces* iranensis | CDR09741.1 |
| medi5519  | 177                         | 60/77            | *Streptomyces* aizunensis | AAX98180.1 |
| medi5518R | 417                         | 73/83            | *Kitasatospora mediocidica* | WP_035796319 |
Putative functions of the encoded proteins were deduced from analyses with the BlastP program (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The % identity/similarity for the protein in the database with the highest end-to-end similarity is indicated. The entire genome sequence of *S. mediocidicus* ATCC23936 has been deposited in GenBank where it has been classified as *S. blastmyceticus* and can be accessed as *S. blastmyceticus* Bioproject PRJNA411827, Biosample SAAMN07688521.

### 4. Supplementary References

[1] Sambrook, J. Russell, D W. Molecular Cloning: A Laboratory Manual, 3rd ed. Cold Spring Harbor Laboratory Press, New York, 2001.

[2] Kieser, T, Bibb, M, Buttner, M, Chater, K F, Hopwood, D A, Practical *Streptomyces Genetics*. The John Innes Foundation, Norwich, 2001.

[3] Gibson, D. G.; Young, L.; Chuang, R. Y.; Venter, J. C.; Hutchison, C. A. III.; Smith, H. O. *Nat. Methods* 2009, 6, 343–345.

[4] Zhang, L.; Hashimoto, T.; Qin, B.; Hashimoto, J.; Kozone, I.; Kawahara, T.; Okada, M.; Awakawa, T.; Ito, T.; Asakawa, Y.; Ueki, M.; Takahashi, S.; Osada, H.; Wakimoto, T.; Ikeda, H.; Shin-Ya, K.; Abe, I. *Angew. Chem. Int. Ed.* 2017, 56, 1740–1745.

[5] Hong, H.; Samborskyy, M.; Lindner, F.; Leadlay, P F. *Angew. Chem. Int. Ed.* 2016, 55, 1118–1123.

[6] MacNeil, D. J.; Gewain, K. M.; Ruby, C. L.; Dezeny, G.; Gibbons, P. H.; MacNeil, T. *Gene* 1992, 111, 61–68.

[7] Hong, H.; Fill, T.; Leadlay, P. F. *Angew. Chem. Int. Ed.* 2013, 52, 13096–13099.

[8] Cai, P.; Kong, F.; Fink, P.; Ruppen, M. E.; Williamson, R. T.; Keiko, T. *J. Nat. Prod.* 2007, 70, 215–219.

[9] Sun, Y.; Hahn, F.; Demydchuk, Y.; Chettle, M.; Tosin, M.; Osada, H.; Leadlay, P. F. *Nat. Chem. Biol.* 2010, 6, 99–101.