



Published in final edited form as:

J Am Chem Soc. 2024 February 14; 146(6): 4212–4220. doi:10.1021/jacs.3c13670.

Discovery and Characterization of the Fully Decorated Nocardiosis Associated Polyketide Natural Product

Shreya Kishore,

Department of Chemistry, Stanford University, Stanford, California 94305, United States

Antonio Del Rio Flores,

Department of Chemistry, Stanford University, Stanford, California 94305, United States

Stephen R. Lynch,

Department of Chemistry, Stanford University, Stanford, California 94305, United States

Kai P. Yuet,

Department of Chemistry, Stanford University, Stanford, California 94305, United States

Chaitan Khosla

Department of Chemistry, Stanford University, Stanford, California 94305, United States

Department of Chemical Engineering and Sarafan ChEM-H, Stanford University, Stanford, California 94305, United States

Abstract

The genomes of 40 strains of *Nocardia*, most of which were associated with life-threatening human infections, encode a highly conserved assembly line polyketide synthase designated as the NOCAP (NOCardiosis-Associated Polyketide) synthase, whose product structure has been previously described. Here we report the structure and inferred biosynthetic pathway of the fully decorated glycolipid natural product. Its structure reveals a fully substituted benzaldehyde headgroup harboring an unusual polyfunctional tail and an O-linked disaccharide comprising a 3- α -epimycarose and 2-*O*-methyl- α -rhamnose whose installation requires flavin monooxygenase-dependent hydroxylation of the polyketide product. Production of the fully decorated glycolipid was verified in cultures of two patient-derived *Nocardia* species. In both *E. coli* and *Nocardia* spp., the glycolipid was only detected in culture supernatants, consistent with data from genetic knockout experiments implicating roles for two dedicated proteins in installing the second sugar substituent only after the monoglycosyl intermediate is exported across the bacterial cell membrane. With the NOCAP product in hand, the stage is set for investigating the evolutionary

Corresponding Author: Chaitan Khosla – Department of Chemistry, Stanford University, Stanford, California 94305, United States; Department of Chemical Engineering and Sarafan ChEM-H, Stanford University, Stanford, California 94305, United States; khosla@stanford.edu.

The authors declare no competing financial interest.

ASSOCIATED CONTENT

Supporting Information

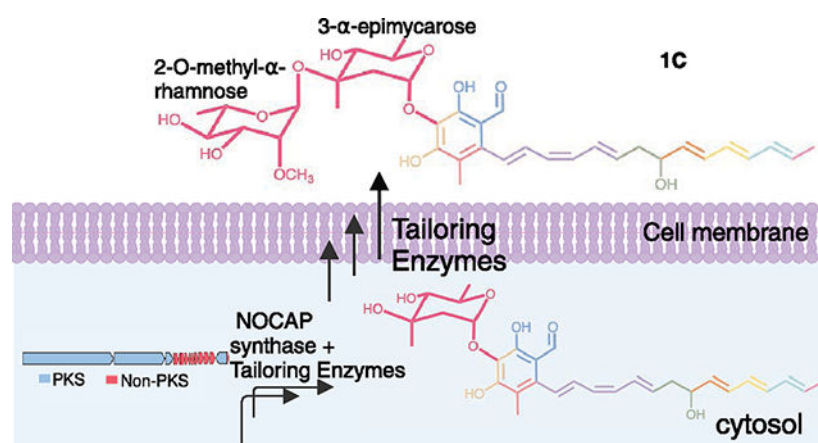
The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/jacs.3c13670>.

Detailed methods, figures, and tables (PDF)

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/jacs.3c13670>

benefit of this polyketide biosynthetic pathway for *Nocardia* strains capable of infecting human hosts.

Graphical Abstract



INTRODUCTION

Assembly line polyketide synthases (PKSs) are large multifunctional enzymes responsible for the biosynthesis of numerous structurally complex, biologically diverse, clinically used natural products.¹ Each assembly line PKS is composed of multiple modules, where individual modules encompass multiple catalytic domains that are collectively responsible for catalyzing the elongation and modification of the growing polyketide chain by one ketide unit. Of particular interest to our laboratory is an orphan PKS solely present in strains of the actinomycete *Nocardia*, termed NOCardiosis-Associated Polyketide (NOCAP) synthases. Many *Nocardia* strains are opportunistic pathogens that cause nocardiosis, especially in patients with immunosuppressive conditions.² Nocardiosis is predominantly a serious pulmonary infection, but can also be a systemic disease that disseminates through the bloodstream to other organs.²

We analyzed 337 publicly available whole genome assemblies of *Nocardia* to create a phylogenetic tree (Figure 1). The NOCAP biosynthetic gene cluster was present in 40 strains spanning 22 species (Figure 1, Table S2). Their clustering in this phylogenetic tree suggests a recent evolutionary origin of the biosynthetic pathway. It also highlights preservation of the NOCAP pathway among descendants of the last common ancestral strain, plausibly due to an added fitness provided by the natural product in their human hosts. The existence of three distinct clades of NOCAP-positive *Nocardia* is suggestive of ongoing lateral transfer of the gene cluster among *Nocardia*. Notably, 34 of the 40 NOCAP-positive strains were isolated from humans (Figure 1, pink); the remaining six (Figure 1, purple), while isolated from soil, are clustered within clades of NOCAP-positive pathogens. While there is no direct evidence as yet for the involvement of NOCAP in the pathogenesis of nocardiosis, our analysis provides a compelling argument for the biological relevance of this natural product.

The previously reported structure of the PKS product contained a substituted resorcyraldehyde headgroup linked to a 15-carbon tail with two conjugated trienes separated by a stereogenic hydroxy group (compound **1**, Figure 2).^{3,4} While our original report inferred an all-*trans* stereoconfiguration of both trienes,³ subsequent analysis of the product of a truncated NOCAP synthase⁵ raised the possibility that one of the two triene moieties had a *trans-cis-trans* configuration. Additionally, for unknown reasons, it became apparent that module 3 of the assembly line PKS was incidentally skipped, resulting in the production of a related polyketide with a 13-carbon tail (compound **2**, Figure 2).³

The NOCAP gene cluster contains 12 non-PKS genes, designated *nocapD-nocapN* (Figure 2b), most of which encode enzymes involved in sugar biosynthesis and transfer as assessed by their closest characterized homologues (Table S1).⁶ Many biologically active bacterial natural products derive their activities from atypical sugar substituents on polyketide scaffolds.^{7–9} To elucidate the structure of the fully decorated NOCAP product, we sought to refactor the complete biosynthetic gene cluster in *E. coli* BAP1. Here we report the structure of the fully decorated glycolipid product of the NOCAP biosynthetic gene cluster by isolating it from a recombinant *E. coli* strain harboring a refactored set of 4 PKS and 12 non-PKS genes, as well as from two patient-isolated *Nocardia* strains.

RESULTS AND DISCUSSION

Five plasmids [pCK-KPY222/pCK-KPY259/pCK-KPY178/pCK-KPY212/pCK-KPY215] were required to efficiently clone the NOCAP gene cluster and the 12 accessory genes into *E. coli* (Figure S1). Two new products, **1C** and **2C**, were observed in cultures of *E. coli* harboring all five plasmids. By HRMS, we identified the fully decorated polyketide **1C** with a molecular formula of C₃₇H₅₀O₁₂ (observed [M – H][–] *m/z* 685.3235, theoretical [M – H][–] *m/z* 685.3230, 0.7 ppm error) and **2C** with a molecular formula of C₃₅H₄₈O₁₂ (observed [M – H][–] *m/z* 659.3071, theoretical [M – H][–] *m/z* 659.3073, 0.3 ppm error) (Figures S2–S6). Similar to the relationship between **1** and **2** previously reported,³ **1C** and **2C** differed by a mass shift consistent with an unsaturated two-carbon (C₂H₂) moiety. **1C** and **2C** shared the same mass fragments as PKS products **1** and **2**, respectively, but with additional fragments consistent with post-PKS tailoring steps (Figures S3–S6). The relative ratio between **2C** and **1C** from heterologous expression was ~10:1, prompting us to focus our efforts on isolation of **2C** (Figure S7).

Compound **2C** was isolated as an acid-sensitive (Figure S8), faint yellow solid from ~80 L of *E. coli* BAP1 supernatant (yield ~12.5 μg/L) (Figure S9). The structure of **2C** was fully elucidated by extensive 1D and 2D NMR experiments and analyses, including ¹H 1D, ¹³C 1D, COSY, TOCSY, HSQC, HMBC, and ROESY experiments (Figures 3, S10–S33, Table 1). Consistent with the previously elucidated PKS products,³ COSY, TOCSY, and HMBC experiments confirmed the same connectivity of **2C**.

To elucidate the structure of the sugars, two anomeric proton–carbon sets at H22/C22 (δ_H 5.24, δ_C 102.49 ppm) and H29/C29 (δ_H 5.31, δ_C 90.15 ppm) were identified. COSY, TOCSY, HSQC, and HMBC experiments established carbon–carbon connectivity from C29 to C34 (17.55 ppm) on one sugar. Additionally, we observed a singlet *O*-methyl substituent

at C30 (81.53 ppm), two hydroxy substituents at C31 (71.37 ppm) and C32 (73.86 ppm), and a methyl substituent at C33 (69 ppm), suggesting a 6-deoxy sugar moiety. COSY, TOCSY, HSQC, and HMBC experiments established carbon–carbon connectivity from C22 to C27 (18.55 ppm) on the other sugar moiety. We observed a hydroxy substituent at C25 (77.61 ppm), a tertiary carbon, C₂₄ (78.2 ppm) between C23 and C25, a methyl substituent at C26 (69.26 ppm), and two diastereotopic protons at C23 (41.8 ppm), implying a dideoxy sugar moiety. HMBC correlations indicated that the tertiary carbon C24 was attached to a methyl substituent as well as to the other sugar via an O-linkage, resulting in a disaccharide (Figures 3, S22–S26). An HMBC correlation between H22 (5.24 ppm) and C4 (132.04 ppm) established the linkage of the disaccharide at the C4 position on the aromatic headgroup of the PKS products (Figures S25).

The predicted exact mass of the NMR-deduced structure of compound **2C** (m/z 659.3073 [$M - H$][−]) is consistent with the observed HRMS value (Figure S3). We can be assured that the glycosyl groups attached to both compounds **1** and **2** are identical because (i) the m/z difference between **2** and **2C** is the same as the m/z difference between **1** and **1C**; (ii) **1C** and **2C** differ by a mass shift consistent with an unsaturated two-carbon (C₂H₂) moiety; and (iii) **1C** and **2C** have consistent fragmentation patterns wherein the fragments for **1C** are shifted by the m/z of an unsaturated two-carbon (C₂H₂) moiety (Figures S3 and S5), as previously reported for compounds **1** and **2**.³

To assign the relative configuration of **2C**, we analyzed ROESY cross peaks between substituents attached to nonadjacent carbon atoms. A ROESY cross peak between H28–H26 (1.66 ppm, 4.08 ppm) and another between H25–H23a (4.08 ppm, 2.04 ppm) established this sugar as a 3-epimycarose because these sets of protons would interact with each other only if they were axial. H22 has a 4.4 Hz J -coupling, suggesting an equatorial orientation and an α -anomer sugar (Figure S30). Despite the *syn* penalty for the 1–3 diaxial connection, the sugar gains added stability due to the anomeric effect, potentially counteracting the diaxial strain.¹¹ For the other sugar, we observed ROESY cross peaks between H31–H33 (3.73 ppm, 3.86) and H35–H32 (3.52 ppm, 3.42 ppm), establishing this sugar as a 2-*O*-methyl-rhamnose sugar because these sets of protons would only interact with each other if they were axial. H29 has a 1.7 Hz J -coupling (Figure S30), indicating again an equatorial orientation and an α -anomer sugar: 2-*O*-methyl- α -rhamnose.

In the alkyl chain, C18–C19 (J = 13.8 Hz), C16–C17 (J = 15.2 Hz), and C12–C13 (J = 14.9 Hz) are *trans* double bonds (Table 1). Two ROESY correlations, between H8/H9–H1 and H8/H9–H21, confirm the C8–C9 *trans* bond (Figures S27 and S29). Only the H8/H9–H21 correlation would exist in the *cis* configuration. Two lines of evidence suggest a *cis* C10–11 bond, dictating a *trans*–*cis*–*trans* configuration of **2C**: (i) While H17–H19 and H16–H18 show representative *trans* ROESY correlations, H11–H13 displays a notably more intense correlation, reflecting a closer proximity in *cis* configuration (Figure S32); (ii) a ROESY cross-peak between H12–H9 (Figure S33) would only be predicted if the C10–C11 bond were *cis*.

Although most polyketides are all *trans*, some *cis* polyketides have been observed.¹² In our case, a tailoring enzyme doublet of doublet of quartets encoded within the NOCAP

cluster is unlikely to promote isomerization because the truncated PKS product synthesized by modules 1–5, with no tailoring enzymes, possesses the analogous bond in the *cis* configuration (Figure S34).⁵ The *cis* bond may be a consequence of PKS stereospecificity or an endogenous *E. coli* enzyme. For instance, FabA from the fatty acid synthase of *E. coli* can reversibly isomerize *trans*-2-decenoyl-ACP to *cis*-3-decenoyl-ACP.¹³ Since *Nocardia* also contains FabA homologues, it is possible that the same *trans* to *cis* isomerization of the C10–C11 occurs in the native strain.

We also sought to identify compounds **1C** and/or **2C** in two patient-derived, NOCAP-positive *Nocardia* strains. Compound **1C** was detected in the spent culture media of *Nocardia abscessus* (DSM: 44432) and *Nocardia pneumoniae* (DSM 44730) (Figure 4). The retention time and UV spectrum of the *Nocardia*-derived metabolite are identical to those of *E. coli*-derived **1C** based on HRMS (Figures 4, S35), suggesting that **1C** made by *Nocardia* is also in the *trans*–*cis*–*trans* configuration. Production of **1C** was correspondingly abolished in a Δ KS₆DH₆ *Nocardia abscessus* mutant (Figures 4, S36). Derivatization of **1C** from both sources with Girard's reagent T¹⁴ confirmed the presence of an aldehyde (Figure S37). Incidental skipping of module 3 to produce compound **2C** seems to be an artifact of heterologous expression in *E. coli* since **2C** was not detected in *Nocardia* cultures.

To decode the biosynthetic pathway from PKS products (**1** and **2**) to **1C** and **2C**, we constructed single gene deletion plasmids of each of the accessory genes in the NOCAP gene cluster. Co-transformation of each of these plasmids with the remaining four plasmids required to produce **1C** and **2C** in *E. coli* BAP1 led to the identification of the full set of essential biosynthetic genes. Four genes, *nocapE*, *nocapI*, *nocapN*, and *nocapP*, were found to be nonessential for **1C** and **2C** biosynthesis (Figure S38 and Table S1). Because *E. coli* BAP1 lacks NcapE, NcapN, and NcapP homologues, it is unlikely that they were substituted by an endogenous enzyme. In contrast, NcapI has 29%/46% identity and similarity to *E. coli* GDP-mannose 4,6-dehydratase and could be dispensable in this heterologous host.

Based on the structures of **1C** and **2C**, we anticipated that the substituted resorcyaldehyde headgroup would need to be hydroxylated at C4 before disaccharide incorporation. NcapM, a putative flavin-dependent monooxygenase, was identified as the candidate enzyme for this transformation.¹⁵ To test this hypothesis, we engineered plasmid pCK-KPY285 to include NcapM and introduced it into *E. coli* BAP1 along with plasmids pCK-KPY222 and pCK-KPY259, required for production of the PKS product(s). HRMS analysis revealed that the resulting strain produced two compounds consistent with structures **1B** (molecular formula C₂₃H₂₅O₅; observed [M – H][–] *m/z* 381.1710, theoretical [M – H][–] *m/z* 381.1707, 0.8 ppm error) and **2B** (molecular formula C₂₁H₂₃O₅; observed [M – H][–] *m/z* 355.1544, theoretical [M – H][–] *m/z* 355.1551, 1.9 ppm error) (Figures S39–S44). As with compounds **1** and **2**, **1B** and **2B** differed by a mass shift consistent with a C₂H₂ moiety. The mass shift from **1** to **1B** and **2** to **2B** was also consistent with the addition of a hydroxy group. All the expected HRMS fragments of compounds **1** and **2**³ were shifted by +16 in the fragmentation patterns of **1B** and **2B** (Figures S41, S43), confirming that NcapM is responsible for catalyzing the installation of a hydroxy moiety at C4 of **1** and **2** (Figure S45).

By consolidating experimental data from the single gene deletions and bioinformatic analyses of the eight essential accessory genes in the NOCAP gene cluster (Figure S38 and Table S1), we propose a biosynthetic pathway for fully decorated compounds **1C** and **2C** (Figures 5, S46). Initial steps in the biosynthesis of both TDP-3- α -epimycarose and TDP-2-*O*-methyl- α -rhamnose are predicted to utilize housekeeping enzymes involved in rhamnose biosynthesis (RmlA-D) that are encoded in the genomes of all *Nocardia* strains as well as *E. coli* BAP1.⁷ This is analogous to spinosyn biosynthesis, where a single set of enzymes supplies rhamnose for cell wall and antibiotic biosynthesis.¹⁶ Subsequent steps in the biosynthesis of the two TDP-sugars require the activities of NocapD, NocapJ, NocapG, and NocapH (Figure S46).

We propose that NocapK, a homologue of the glycosyltransferase in elloramycin biosynthesis,¹⁷ transfers TDP-3-epimycarose onto the 4-phenol substituent of **1B** and **2B**. The second glycosylation is likely catalyzed by the tandem activities of two integral membrane proteins, NocapF and NocapL, whose closest characterized homologues are found in mycobacteria (which are close relatives of *Nocardia*¹⁸). NocapF is homologous to mycobacterial polyprenol monophosphomannose (PPM) synthase that converts intracellular GDP-mannose into PPM,^{19,20} whereas NocapL is a homologue of a transmembrane mannosyltransferase that utilizes PPM to mannosylate its glycosyl-acceptor substrate at the external face of the cell membrane.²¹ Together, these two mycobacterial proteins can flip PPM across the cell membrane, although the underlying mechanism is not well understood. Based on that precedent,²² a pathway is proposed for converting polyketide products **1** and **2** into fully decorated compounds **1C** and **2C** with concomitant release outside the cell (Figure 5). This elaborate biosynthetic pathway, involving specific transmembrane transporters to glycosylate and transport the natural product out of the cell, underscores the biological potential of compounds **1C** and **2C** in *Nocardia*. Our proposed pathway is supported by detection of PKS products compounds **1** and **2** and hydroxylated **1B** and **2B** within the *E. coli* cell pellet, while fully decorated compounds **1C** and **2C** were only detected in *E. coli* and *Nocardia* supernatants (Figures S47 and S48).

Whereas the absolute configuration of the disaccharide was not experimentally determined, it may be inferred from the biosynthetic pathway. The common precursor to both saccharide units is presumably D-glucose-1-phosphate (Figure S46). RmlC, putatively involved in biosynthesizing the 2-*O*-methyl-rhamnose sugar (Figure S46), is known to catalyze a 3,5 double epimerization reaction to eventually form dTDP-L-rhamnose.²³ It is therefore likely that this sugar is L-2-*O*-methyl-rhamnose. Because RmlC or any other epimerase is not involved in the biosynthesis of dTDP-3-epimycarose (Figure S46), it is likely that this sugar is D-3-epimycarose.

Phylogenetic analysis of the NOCAP biosynthetic gene cluster revealed that virtually all of the *Nocardia* strains harboring its encoded biosynthetic pathway are human pathogens. Motivated by this observation, we have decoded via extensive NMR and spectrometric analyses the structure of its unusual, highly functionalized glycolipid product. NOCAP comprises a fully substituted benzaldehyde headgroup with an unusual polyfunctional alkyl tail and an *O*-linked disaccharide unit derived from 3- α -epimycarose and 2-*O*-methyl- α -rhamnose sugars. Genetic analysis has established that the NOCAP biosynthetic pathway

involves successive action of a 3 MDa assembly line polyketide synthase, a flavin-dependent monooxygenase, a cytosolic glycosyltransferase, and an extracellular glycosyltransferase. With the structure of this secreted glycolipid in hand, we set the stage to interrogate its role in the pathogenicity of *Nocardia*.

EXPERIMENTAL METHODS

General HPLC Solvents.

HPLC A: water; HPLC B: acetonitrile. For low-resolution MS, an Ultivo Triple Quadrupole LC/MS was used, and for high-resolution an MS 6545XT AdvanceBio LC/Q-TOF was used. For NMR, a Bruker Neo-500 NMR and Inova 600 NMR were used. For HPLC, an Agilent C18 column was used.

Description of Plasmids Used.

Plasmid pCK-KPY178 encodes the malonyl-CoA synthetase MatB from *Streptomyces coelicolor*,²⁴ the *Rhizobium leguminosarum* malonate transporter MatC,²⁵ the loading module (module L), and trans-AT-TEII (tAT-TEII) of the NOCAP synthase. Plasmid pCK-KPY222 encodes modules 1 and 2 from *N. pneumoniae* as one bimodular protein and module 3 from *N. araoensis*. (As described previously,³ combining PKS subunits from different patient-derived *Nocardia* strains was necessary because certain subunits could not be expressed as functional proteins in *E. coli*.) Plasmid pCK-KPY259 encodes module 4-KS5 and intact NocapB from *N. araoensis*. Plasmid pCK-KPY212 encodes the 12 non-PKS genes found in the NOCAP gene cluster in *N. araoensis*. Finally, plasmid pCK-KPY215 encodes Sfp from *Bacillus subtilis*²⁶ as well as GroL1, GroL2, and GroS from *Streptomyces coelicolor*, which together constitute the GroEL–GroES molecular chaperone system that forms a nanocage to allow encapsulation of non-native substrate proteins and provides a physical environment optimized to promote and accelerate protein folding.²⁷

Heterologous Production of 1, 2, 1B, 2B, 1C, and 2C in *E. coli*.

Growth media used in this protocol had the following compositions: (a) LB broth: 25 g/L LB broth Miller granulated powder, supplemented with 100 mg/L carbenicillin disodium, 50 mg/L kanamycin monosulfate, 50 mg/L streptomycin sulfate, 25 mg/L gentamicin sulfate, and 25 mg/L chloramphenicol. (b) Terrific Broth: 47.6 g/L Terrific Broth (modified) powder, 10 g/L glycerol, supplemented with 100 mg/L carbenicillin disodium, 50 mg/L kanamycin monosulfate, 50 mg/L streptomycin sulfate, 25 mg/L gentamicin sulfate, and 25 mg/L chloramphenicol.

For 1 and 2: *E. coli* BAP1 [pCK-KPY178/pCK-KPY222/pCK-KPY259]

For 1B and 2B: *E. coli* BAP1 [pCK-KPY285/pCK-KPY222/pCK-KPY259]

For 1C and 2C: *E. coli* BAP1 [pCK-KPY178/pCK-KPY222/pCK-KPY259/pCK-KPY212/pCK-KPY215]

A single colony of the appropriate *E. coli* BAP1 strain was used to inoculate an overnight seed culture of LB broth (15 mL) grown at 30 °C in a 50 mL Falcon tube. The seed

culture (10 mL) was used to inoculate 1 L of Terrific Broth in 2.5 L Pyrex Erlenmeyer flasks and agitated at 30 °C (200 rpm) until the OD₆₀₀ reached ~0.4. After addition of a 10 mL/L solution of 500 mM sodium malonate, pH 7.4, a 1 mL/L solution of 50 mM calcium pantothenate, and a 0.25 mL/L solution of 200 mM isopropyl β -D-1-thiogalactopyranoside (IPTG), the cultures were agitated (200 rpm) at 20 °C for an additional 48 h. The cultures were then transferred to centrifuge bottles and centrifuged at 5000g for 25 min at 20 °C.

Isolation and Purification of 1, 2, 1B, and 2B.

As previously described in Yuet et al.,³ for analytical verification of biosynthesis of **1**, **2**, **1B**, and **2B**, small samples of 250 mg of frozen cells were resuspended in 0.3 mL of *n*-hexane (Thermo Fisher Scientific) and 0.2 mL of isopropanol (Thermo Fisher Scientific) and vortexed with glass beads for 30 min. After centrifugation at 15000g for 10 min, the upper organic phases were collected, dried in vacuo, redissolved in methanol, and analyzed by LC-MS (Agilent Ultivo QQQ LC-MS system or Agilent 6545 Q-TOF LC-MS). Samples were protected from light at all times.

For isolation and purification:³

HPLC A: 99.9% (v/v) water, 0.1% (v/v) formic acid (Thermo Fisher Scientific)

HPLC B: 99.9% (v/v) acetonitrile, 0.1% (v/v) formic acid (Thermo Fisher Scientific)

Guided by Matyash's lipid extraction method,²⁸ the cell pellet was resuspended in 65 mL/L culture water, 150 mL/L culture methanol (Thermo Fisher Scientific), and 500 mL/L culture methyl *tert*-butyl ether (Thermo Fisher Scientific) and sonicated in a Branson Ultrasonics M1800 ultrasonic cleaning bath for 1 h. Phase separation was induced by the addition of a 65 mL/L culture of brine. The upper organic phases were collected, pooled, dried in vacuo, and redissolved in 20 mL of 25% (v/v) methanol in HPLC B. This mixture was diluted to a final volume of 90 mL with HPLC A and equally loaded onto six pre-equilibrated C18 solid-phase extraction cartridges (3M Empore 7 mm/3 mL). The cartridges were each washed with 1 mL of HPLC A, 1 mL of 25/75 HPLC B/HPLC A, 1 mL of 50/50 HPLC B/HPLC A, 1 mL of 75/25 HPLC B/HPLC A, and 1 mL of HPLC B. Fractions identified by LC-MS (a Waters SQ detector 2 LC-MS system and/or Agilent 6545 Q-TOF LC-MS system) to contain **1**, **2**, **1B**, or **2B** (50/50 HPLC B/HPLC A, 75/25 HPLC B/HPLC A and HPLC B) were pooled, dried in vacuo, redissolved in 35/65 HPLC B/HPLC A, and filtered through a 0.45 μ m PFTE membrane (VWR). This mixture was separated with a gradient elution method (35/65 HPLC B/HPLC A to 80/20 HPLC B/HPLC A over 1 h at 2 mL/min) on an Agilent 1260 Infinity LC system equipped with an Agilent Eclipse XDB-C8 column (5 μ m, 250 $\times$ 9.4 mm).

Isolation and Purification of 2C and 1C.

Supernatant and pellets from *E. coli* cultures were separated and tested for the presence of **1C** and **2C**. As judged by LC-MS, both **1C** and **2C** were detected predominantly in the supernatant. For analytical verification of the biosynthesis of **1C** and **2C**, 25 mL of supernatant was extracted twice with ethyl acetate (25 mL each time), followed by drying

with magnesium sulfate. Ethyl acetate was then removed by rotary evaporation, and the extract was resuspended in methanol for LC-MS analysis.

For Isolation and Purification.

The supernatant was transferred to fresh 2.5 L Pyrex Erlenmeyer flasks, 20 g/L Amberlite IRA-900 chloride form resin was added, and the flasks were agitated (200 rpm) at 20 °C for 2 h. The resin was filtered out using vacuum filtration, airdried, transferred to a new Erlenmeyer flask along with 125 mL of ethyl acetate/20 g of resin, and agitated (200 rpm) at 20 °C for an additional 2 h to allow the metabolites to be extracted by ethyl acetate from the resin. The ethyl acetate was then removed using a rotary evaporator. The crude extract was then redissolved in 2 mL of ethyl acetate and purified using silica column chromatography. The mobile phase started with 100% ethyl acetate (300 mL) followed by 2% methanol, 98% ethyl acetate (300 mL), and finally 5% methanol, 98% ethyl acetate (400 mL). The fractions containing compounds **1C** and **2C** were identified using LC-MS, pooled, and dried under vacuum. The sample was then redissolved in 2 mL 60:40 HPLC A:HPLC B and filtered through a 0.45 μ m PTFE membrane. This mixture was separated with a gradient elution method (isocratic 40% B for 20 min $\rightarrow$ 50% B over 20 min $\rightarrow$ 70% B over 10 min at a flow rate of 7 mL/min) on an Agilent C18 column. Fractions identified by LC-MS to contain **1C** and **2C** were separately pooled and lyophilized. Yields for **2C** were ~1 mg from 60 to 80 L of culture, and yields for **1C** were significantly lower.

NMR of Compound 2C.

Samples were prepared in 0.5 mL of CDCl₃ in standard borosilicate glass NMR tubes (5 mm $\times$ 8 in.). NMR spectra of **2C** were acquired on a 600 MHz Inova spectrometer. Experiments acquired included 1D ¹H NMR, COSY, HSQC, HMBC, and ROESY (600 MHz Inova spectrometer). 1D ¹³C NMR spectra were acquired on a Bruker Avance Neo 500 MHz NMR spectrometer equipped with a 5 mm liquid nitrogen cooled broadband Prodigy probe and a Varian Inova 600 MHz NMR equipped with a 5 mm room-temperature HCN probe. Automatic baseline corrections, phasing, and referencing were used to correct for offsets in the acquired data.

The two-dimensional NMR experiments were acquired on a Varian Inova 600 MHz NMR instrument. The COSY was acquired with 4 scans of 4096 complex points in t₂ and 300 real points in t₁ over 8400 Hz in both dimensions for an acquisition time of 0.24 s with a recycle delay of 1.8 s. The HSQC was acquired with sensitivity enhancement with 8 scans of 2048 complex points in t₂ and 300 complex points in t₁ over 8400 Hz in the H1 dimension and 25,641 Hz in the C13 dimension for an acquisition time of 0.12 s with a recycle delay of 2.2 s. The HMBC was acquired with 44 scans of 4096 complex points in t₂ and 300 real points in t₁ over 8400 Hz in the H1 dimension and 35,199 Hz in the C13 dimension for an acquisition time of 0.24 s with a recycle delay of 2.2 s. The TOCSY was acquired with a mixing time of 80 ms with 4 scans of 4096 complex points in t₂ and 512 complex points in t₁ over 8400 Hz in both dimensions for an acquisition time of 0.24 s with a recycle delay of 1.8 s. The ROESY was initially acquired with a mixing time of 200 ms with 8 scans of 4096 complex points in t₂ and 512 complex points in t₁ over 8400 Hz in both dimensions for an acquisition time of 0.24 s with a recycle delay of 2.3 s. A higher resolution ROESY

was acquired with a mixing time of 300 ms with 8 scans of 16,384 complex points in t2 and 2048 complex points in t1 over 8400 Hz in both dimensions for an acquisition time of 1.02 s with a recycle delay of 2.5 s. Data were processed with either a VNMRJ 4.2 (Agilent) or MestreNova 14.

Girard T Derivatization of Compound 1C.

To convert the aldehyde-containing **1C** to the hydrazone-containing compound, 0.1 mL extracts containing **1C** were incubated for 1 h at room temperature after the addition of 0.1 mL of 2% (w/v) Girard's reagent T (MilliporeSigma) in methanol and 20 μ L of glacial acetic acid (ThermoFisher Scientific).

Nocardia Strains and Growth Conditions.

Nocardia abscessus DSM 44432 and *Nocardia pneumoniae* DSM 44730 were acquired from the Leibniz Institute DSMZ. The growth media was adapted from Chen et al.²⁹ as follows. Both strains were cultured in 5294 media agar plates (2.5 g/L potato starch, 0.5 g/L yeast extract, 2.5 g/L glucose, 2.5 g/L glycerol, 0.75 g/L corn steep liquor, 0.5 g/L peptone, 0.25 g/L sodium chloride, 0.75 g/L calcium carbonate, 15 g/L agar, pH 7.2) for 3 days at 30 °C to allow for sporulation. Individual spore colonies were inoculated into 100 mL of seed media (10 g/L potato starch, 5 g/L glucose, 3 g/L NZ-amine A, 2 g/L yeast extract, 5 g/L tryptone, 1 g/L potassium phosphate dibasic, 0.5 g/L magnesium sulfate, 3 g/L calcium carbonate, pH 7.0) in a 500 mL baffled flask at 37 °C for 5 days. After the 5-day period, 1.5 mL of the seed culture was inoculated into 150 mL of production media (20 g/L potato starch, 5 g/L glucose, 3 g/L yeast extract, 20 g/L glycerol, 15 g/L cottonseed flour, pH 7.0) in a 500 mL baffled flask at 30 °C for 5 days.

Preparation of Electrocompetent *Nocardia abscessus*.

The general procedure to prepare electrocompetent *Nocardia* was adapted from a previous protocol.³⁰ A single colony of *N. abscessus* maintained in 5294 media agar was used to inoculate 0.5 mL of BHI media, which was grown overnight at 37 °C. The seed culture was inoculated into 50 mL of BHI supplemented with 1% (w/v) glycine and 0.5% (v/v) Tween 80 in a 500 mL baffled flask and was cultured at 37 °C for 2 days. Thereafter, the culture was transferred to a 50 mL conical tube and incubated at 4 °C for 1 h. The chilled cells were then pelleted using a Beckman Coulter Avanti J-15R centrifuge (3200g, 10 min) to separate the spent media and cell pellet. The spent media was disposed of, and the cell pellet was resuspended in 20 mL of sterile water. After a 30 min incubation at 4 °C, the cells were centrifuged again, and the water was carefully decanted. This process was repeated twice more to remove residual secondary metabolites and ionic compounds from the cells. The cells were then resuspended in 25 mL of ice-cold 10% glycerol, centrifuged, and decanted twice more. Lastly, the cells were resuspended in 1 mL of ice-cold 10% glycerol and aliquoted into 100 μ L portions for electroporation.

KS6DH6 Gene Disruption in *Nocardia abscessus*.

Genomic DNA from *N. abscessus* was purified using a Quick-DNA fungal/bacterial kit (Zymo Research). pAD3 was constructed by amplifying 1.5 kb regions upstream and

downstream of KS6DH6 and integrated into the backbone of pNV18. pNV18 is an *E. coli*-*Nocardia* shuttle vector that can be used for gene disruption and delivery in *Nocardia* species.³¹ Between the flanking sequences from the NOCAP synthase, a gentamycin resistance cassette was integrated in pAD3. To facilitate DNA transfer into *N. abscessus*, pAD3 was linearized with XbaI and gel-purified. A 30 ng amount of linearized pAD3 was added to 100 μ L of electrocompetent *N. abscessus* and subsequently incubated for 30 min at 4 °C. Cells were then transferred to a 1 mm electroporation cuvette and electroporated at 25 μ F, 100 Ω , and 1.25 kV. The cells were recovered with addition of 600 μ L of BHI and incubated at 37 °C for 2 h. After 3 days of plating onto 5294 agar media plates supplemented with 150 μ g/mL of gentamycin, resulting colonies were restreaked onto 5294 agar media plates containing either gentamycin or kanamycin. Double-crossover mutants displaying kanamycin sensitivity and gentamycin resistance were cultured, and the presence of the gentamycin gene disruption cassette was confirmed by PCR. Furthermore, primers were designed outside of the targeted deletion region to confirm the KS6DH6 disruption.

LC-MS Analysis of *Nocardia* Extracts.

After a 5-day fermentation period, production cultures were pelleted using a Beckman Coulter Avanti J-15R centrifuge (3200g, 30 min) to separate the spent media and cell pellet. The spent media was extracted 1:1 by volume with ethyl acetate (3 $\times$). The ethyl acetate fractions were pooled and dried under rotary evaporation. The extract was resuspended in 2 mL of methanol for subsequent liquid chromatography mass spectrometry (LC-MS) analyses. The cell pellet was frozen at -80 °C for 1 h and extracted with 20 mL of methanol (3 $\times$). The methanol extracts were pooled and dried under rotary evaporation. The resulting cell pellet extract was redissolved in 2 mL of methanol for LC-MS analyses.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank Stanford ChEM-H Metabolomics Knowledge Center for providing instrumentation (Agilent 6545 Q-TOF mass spectrometer with Agilent 1290 Infinity II UHPLC) and technical support. We thank Khosla lab members Tom Privalsky and Harrison Anthony Besser for helpful discussions and suggestions. We also thank Amy Hsu and Michael Abers for critical feedback on the manuscript and figures.

Funding

The work was supported by National Institutes of Health (NIH) Grant R35 GM141799 (to C.K.), Stanford Graduate Fellowship (to S.K.), Stanford Science Fellowship (to A.D.R.F), and NIH Grant F32 GM123637 (to K.P.Y.). The work utilized the Bruker Neo 500 MHz NMR spectrometer in the Stanford Medicinal Chemistry Knowledge Center funded by the NIH High End Instrumentation grant (1 S10 OD028697-01).

REFERENCES

- (1). Walsh CT Polyketide and Nonribosomal Peptide Antibiotics: Modularity and Versatility. Science 2004, 303 (5665), 1805–1810. [PubMed: 15031493]
- (2). Wilson JW Nocardiosis: Updates and Clinical Overview. Mayo Clin Proc. 2012, 87 (4), 403–407. [PubMed: 22469352]

- (3). Yuet KP; Liu CW; Lynch SR; Kuo J; Michaels W; Lee RB; McShane AE; Zhong BL; Fischer CR; Khosla C Complete Reconstitution and Deorphanization of the 3 MDa Nocardiosis-Associated Polyketide Synthase. *J. Am. Chem. Soc.* 2020, 142 (13), 5952–5957. [PubMed: 32182063]
- (4). Kuo J; Lynch SR; Liu CW; Xiao X; Khosla C Partial In Vitro Reconstitution of an Orphan Polyketide Synthase Associated with Clinical Cases of Nocardiosis. *ACS Chem. Biol.* 2016, 11 (9), 2636–2641. [PubMed: 27384917]
- (5). Guzman KM; Yuet KP; Lynch SR; Liu CW; Khosla C Properties of a “Split-and-Stuttering” Module of an Assembly Line Polyketide Synthase. *J. Org. Chem.* 2021, 86 (16), 11100–11106. [PubMed: 33755455]
- (6). UniProt: a worldwide hub of protein knowledge | Nucleic Acids Research | Oxford Academic. <https://academic.oup.com/nar/article/47/D1/D506/5160987> (accessed 2023-05-17).
- (7). Thibodeaux CJ; Melanç CE III; Liu H Natural-Product Sugar Biosynthesis and Enzymatic Glycodiversification. *Angew. Chem., Int. Ed.* 2008, 47 (51), 9814–9859.
- (8). Schnaitman CA; Klena JD Genetics of Lipopolysaccharide Biosynthesis in Enteric Bacteria. *Microbiological Reviews* 1993, 57 (3), 655–682. [PubMed: 7504166]
- (9). Thorson JS; Hosted TJ Jr; Jiang J; Biggins JB; Ahlert J Natures Carbohydrate Chemists The Enzymatic Glycosylation of Bioactive Bacterial Metabolites. *Curr. Org. Chem.* 2001, 5 (2), 139–167.
- (10). Letunic I; Bork P Interactive Tree Of Life (iTOL) v5: An Online Tool for Phylogenetic Tree Display and Annotation. *Nucleic Acids Res.* 2021, 49 (W1), W293–W296. [PubMed: 33885785]
- (11). Juaristi E; Cuevas G Recent Studies of the Anomeric Effect. *Tetrahedron* 1992, 48 (24), 5019–5087.
- (12). Yin Z; Dickschat J Cis Double Bond Formation in Polyketide Biosynthesis. *Natural Product Reports* 2021, 38 (8), 1445–1468. [PubMed: 33475122]
- (13). Leesong M; Henderson BS; Gillig JR; Schwab JM; Smith JL Structure of a Dehydratase-Isomerase from the Bacterial Pathway for Biosynthesis of Unsaturated Fatty Acids: Two Catalytic Activities in One Active Site. *Structure* 1996, 4 (3), 253–264. [PubMed: 8805534]
- (14). Girard A; Sandulesco G Sur Une Nouvelle Série de Réactifs Du Groupe Carbonyle, Leur Utilisation à l'extraction Des Substances Cétoniques et à La Caractérisation Microchimique Des Aldéhydes et Cétones. *Helv. Chim. Acta* 1936, 19 (1), 1095–1107.
- (15). Eswaramoorthy S; Bonanno JB; Burley SK; Swaminathan S Mechanism of Action of a Flavin-Containing Monooxygenase. *Proc. Natl. Acad. Sci. U. S. A.* 2006, 103 (26), 9832–9837. [PubMed: 16777962]
- (16). Madduri K; Waldron C; Merlo DJ Rhamnose Biosynthesis Pathway Supplies Precursors for Primary and Secondary Metabolism in *Saccharopolyspora Spinosa*. *J. Bacteriol.* 2001, 183 (19), 5632–5638. [PubMed: 11544225]
- (17). Patallo EP; Blanco G; Fischer C; Braña AF; Rohr J; Méndez C; Salas JA Deoxysugar Methylation during Biosynthesis of the Antitumor Polyketide Elloramycin by *Streptomyces Olivaceus*: Characterization of Three Methyltransferase Genes. *J. Biol. Chem.* 2001, 276 (22), 18765–18774. [PubMed: 11376004]
- (18). Ridell M Serological Study of Nocardiae and Mycobacteria by Using “Mycobacterium” Pellegrino and Nocardia Corallina Precipitation Reference Systems. *International Journal of Systematic and Evolutionary Microbiology* 1974, 24 (1), 64–72.
- (19). Gurcha SS; Baulard AR; Kremer L; Loch C; Moody DB; Muhlecker W; Costello CE; Crick DC; Brennan PJ; Besra GS Ppm1, a Novel Polyprenol Monophosphomannose Synthase from *Mycobacterium Tuberculosis*. *Biochem. J.* 2002, 365 (2), 441–450. [PubMed: 11931640]
- (20). Baulard AR; Gurcha SS; Engohang-Ndong J; Gouffi K; Loch C; Besra GS In Vivo Interaction between the Polyprenol Phosphate Mannose Synthase Ppm1 and the Integral Membrane Protein Ppm2 from *Mycobacterium Smegmatis* Revealed by a Bacterial Two-Hybrid System. *J. Biol. Chem.* 2003, 278 (4), 2242–2248. [PubMed: 12427759]
- (21). Morita YS; Sena CBC; Waller RF; Kurokawa K; Sernee MF; Nakatani F; Haites RE; Billman-Jacobe H; McConville MJ; Maeda Y; Kinoshita T PimE Is a Polyprenol-Phosphate-Mannose-Dependent Mannosyltransferase That Transfers the Fifth Mannose of Phosphatidylinositol

- Mannoside in Mycobacteria *. J. Biol. Chem. 2006, 281 (35), 25143–25155. [PubMed: 16803893]
- (22). Mishra AK; Driessen NN; Appelmek BJ; Besra GS Lipoarabinomannan and Related Glycoconjugates: Structure, Biogenesis and Role in Mycobacterium Tuberculosis Physiology and Host-Pathogen Interaction. FEMS Microbiol Rev. 2011, 35 (6), 1126–1157. [PubMed: 21521247]
- (23). Mistou M-Y; Sutcliffe IC; van Sorge NM Bacterial Glycobiology: Rhamnose-Containing Cell Wall Polysaccharides in Gram-Positive Bacteria. FEMS Microbiol Rev. 2016, 40 (4), 464–479. [PubMed: 26975195]
- (24). Hughes AJ; Keatinge-Clay A Enzymatic Extender Unit Generation for In Vitro Polyketide Synthase Reactions: Structural and Functional Showcasing of Streptomyces Coelicolor MatB. Chemistry & Biology 2011, 18 (2), 165–176. [PubMed: 21338915]
- (25). Lombó F; Pfeifer B; Leaf T; Ou S; Kim YS; Cane DE; Licari P; Khosla C Enhancing the Atom Economy of Polyketide Biosynthetic Processes through Metabolic Engineering. Biotechnol. Prog. 2001, 17 (4), 612–617. [PubMed: 11485419]
- (26). Quadri LE; Weinreb PH; Lei M; Nakano MM; Zuber P; Walsh CT Characterization of Sfp, a Bacillus Subtilis Phosphopantetheinyl Transferase for Peptidyl Carrier Protein Domains in Peptide Synthetases. Biochemistry 1998, 37 (6), 1585–1595. [PubMed: 9484229]
- (27). Duchêne A-M; Kieser HM; Hopwood DA; Thompson CJ; Mazodier P Characterization of Two groEL Genes in Streptomyces Coelicolor A3 (2). Gene 1994, 144 (1), 97–101. [PubMed: 7913076]
- (28). Matyash V; Liebisch G; Kurzchalia TV; Shevchenko A; Schwudke D Lipid Extraction by Methyl-Tert-Butyl Ether for High-Throughput Lipidomics. J. Lipid Res. 2008, 49 (5), 1137–1146. [PubMed: 18281723]
- (29). Chen J; Frediansyah A; Männle D; Straetener J; Brötz-Oesterhelt H; Ziemert N; Kayser L; Gross H New Nocobactin Derivatives with Antimuscarinic Activity, Terpenibactins A-C, Revealed by Genome Mining of Nocardia Terpenica IFM 0406. ChemBioChem. 2020, 21 (15), 2205–2213. [PubMed: 32196864]
- (30). Dhakal D; Kumar Jha A; Pokhrel A; Shrestha A; Sohng JK Genetic Manipulation of Nocardia Species. Current Protocols in Microbiology 2016, 40 (1), 10F.2.1–10F.2.18.
- (31). Chiba K; Hoshino Y; Ishino K; Kogure T; Mikami Y; Uehara Y; Ishikawa J Construction of a Pair of Practical Nocardia-Escherichia Coli Shuttle Vectors. Jpn. J. Infect Dis 2007, 60 (1), 45–47. [PubMed: 17314425]

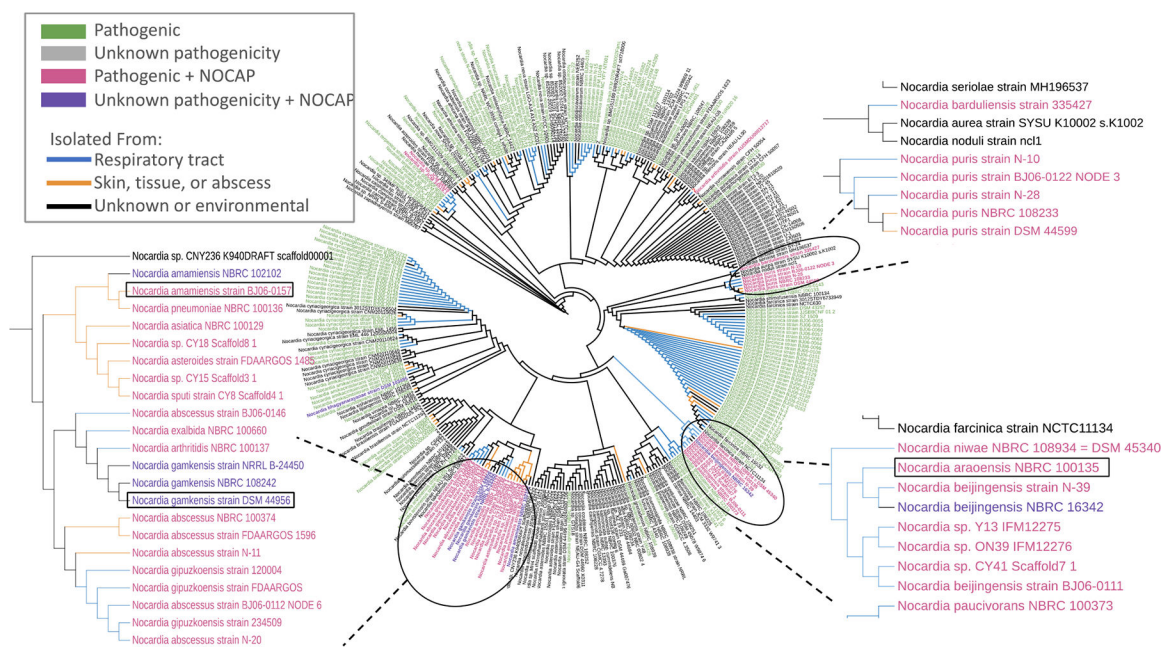
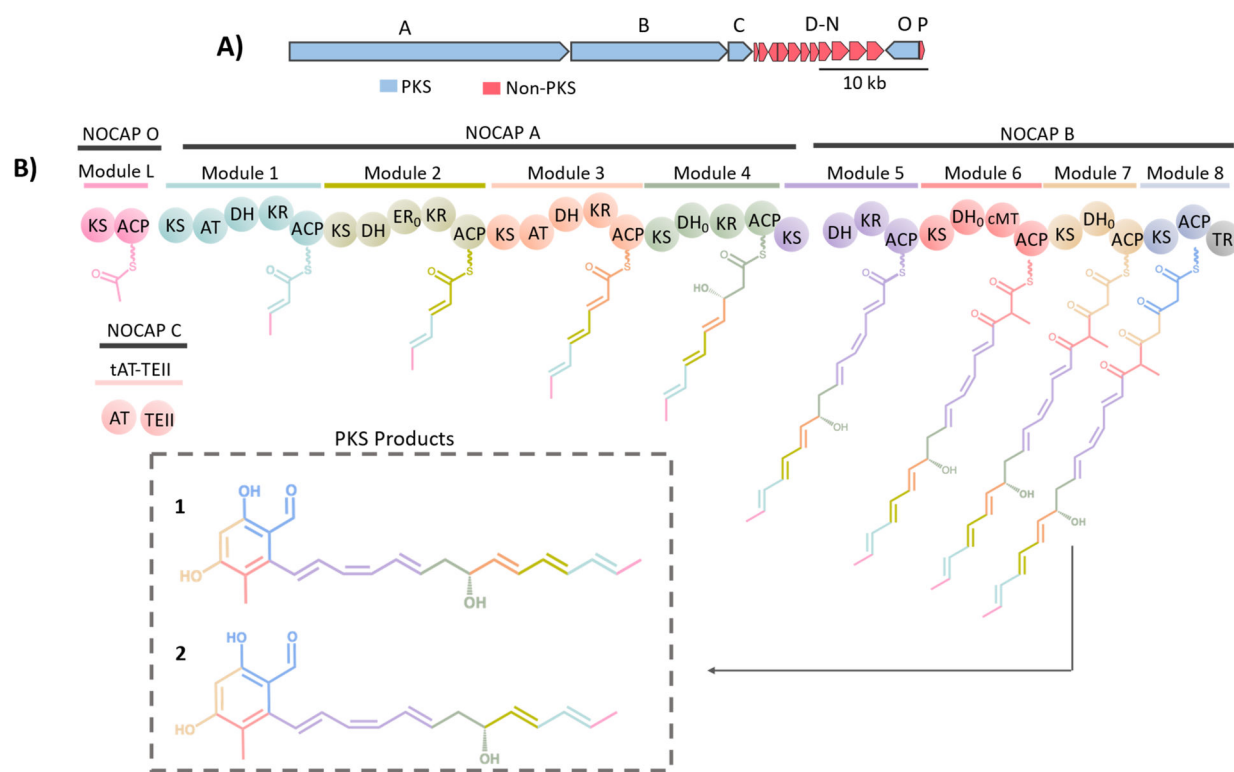


Figure 1. Phylogenetic tree of 337 *Nocardia* genomes based on ClustalW alignment of the *gyrB* housekeeping gene made in iTOL.¹⁰ Black boxes are used to highlight the strains used in the study. NOCAP-positive strains are highlighted in pink or purple.

**Figure 2.**

(A) Genomic architecture of the NOCAP synthase. See Table S1 for closest characterized homologues of *nocapD-N*. (B) Prototypical NOCAP synthase from *N. pneumoniae*. Biosynthesis of **1** was performed by the NOCAP synthase. Key: KS, ketosynthase; AT, acyltransferase; DH, dehydratase; KR, ketoreductase; ER, enoylreductase; cMT, C-methyltransferase; ACP, acyl carrier protein; TR, thioester reductase; and TEII, thioesterase II.

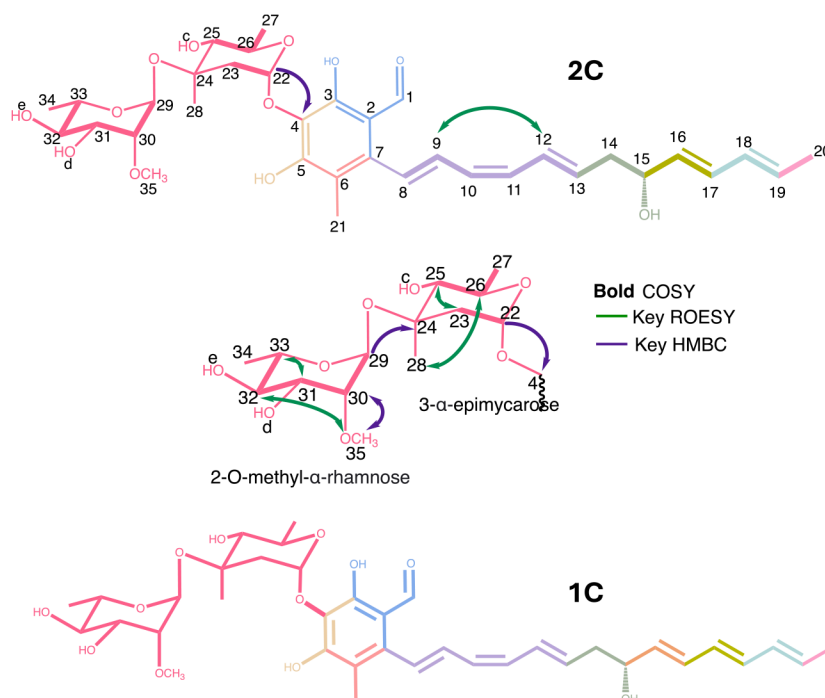


Figure 3. Structure of **2C** based on 2D NMR data. The middle panel shows the disaccharide with its relative configuration assigned based on ROESY data. Green (ROESY) and purple (HMBC) arrows only indicate key correlations; for additional ROESY and HMBC correlations, see Figures S22–S33. The corresponding structure of **1C** is shown in the lower panel.

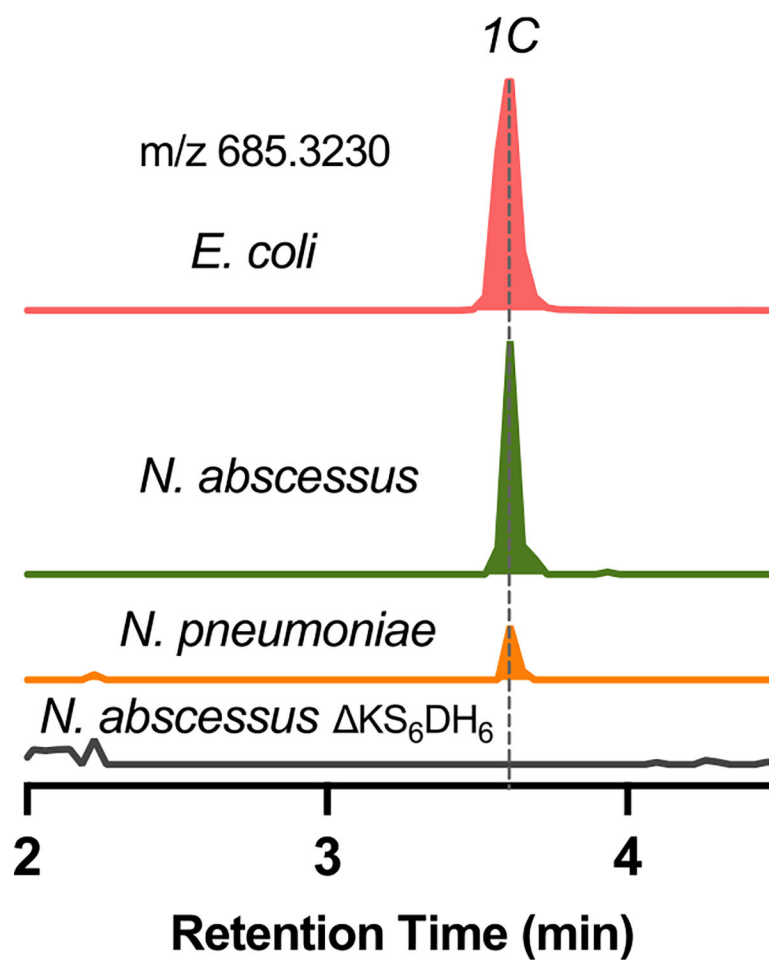


Figure 4.

Extracted ion chromatogram of compound **1C** from supernatants of *E. coli* BAP1 harboring the complete NOCAP biosynthetic gene cluster (pink), *Nocardia abscessus* (green), *Nocardia pneumoniae* (orange), and *Nocardia abscessus* Δ KS₆DH₆ (gray). The spectrum was acquired on an Agilent 6545 Q-TOF LC-MS system. A 10 ppm mass error tolerance was used for each trace.

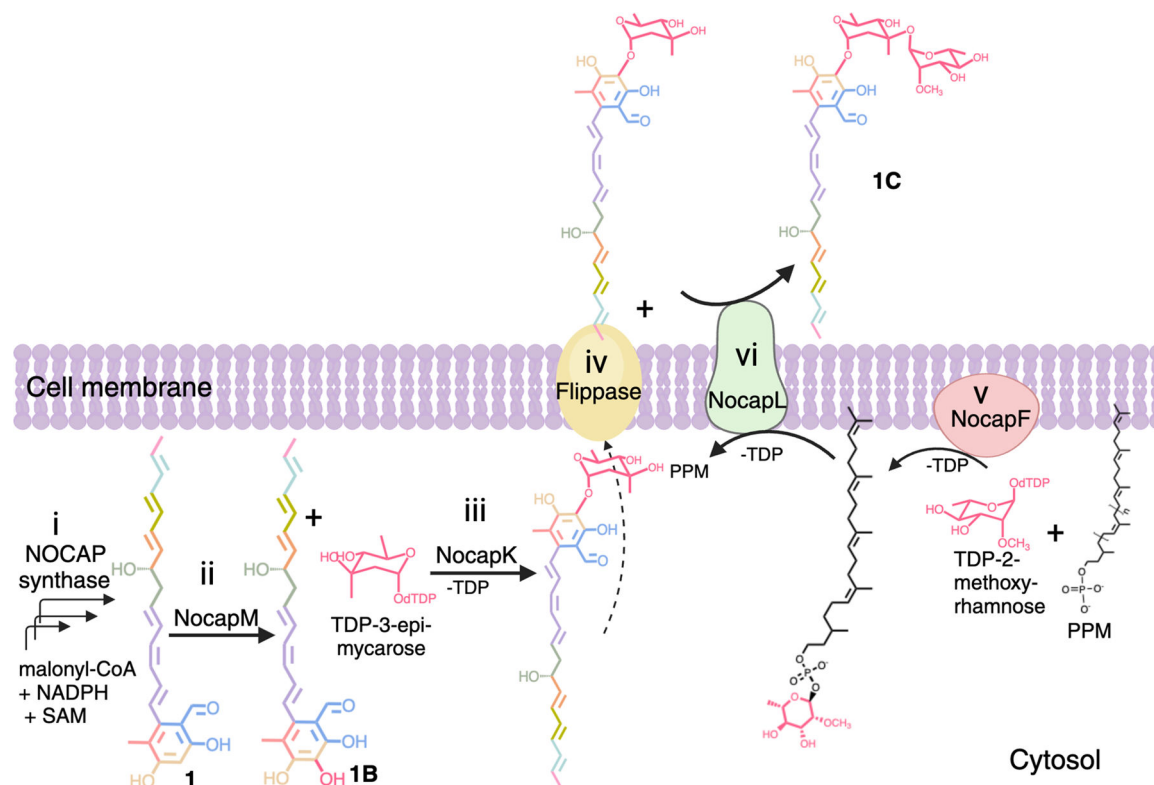


Figure 5.

Proposed NOCAP biosynthetic pathway. In the cytosol (i) compound **1** is synthesized by the NOCAP synthase and (ii) hydroxylated by NocapM to yield **1B**, (iii) which is glycosylated by NocapK using TDP-3-epi-mycarose as the cosubstrate (whose biosynthesis is shown in Figure S46). (iv) The resulting monoglycosyl intermediate is exported by an unidentified, presumably housekeeping, flippase. (v) Meanwhile NocapF synthesizes polyprenol 2-*O*-methyl- α -rhamnosyl phosphate from the corresponding TDP-sugar (whose biosynthesis is shown in Figure S46) and polyprenol monophosphomannose (PPM); finally (vi) NocapL transfers 2-*O*-methyl- α -rhamnose onto the monoglycosyl intermediate, yielding fully decorated **1C**, which is released into the environment. The figure was created with [BioRender.com](https://www.biorender.com) with permission number SA26AEHCOW.

Table 1.

NMR Assignments for 2C^a

atom number	δ_H	δ_C	HMBC	COSY	splitting pattern and <i>J</i> couplings
1	9.9	195.71	C ₂ , C ₃ , C ₄	n/a	s
2	n/a	113.61	n/a	n/a	n/a
3	n/a	155.46	n/a	n/a	n/a
4	n/a	132.04	n/a	n/a	n/a
5	n/a	154.66	n/a	n/a	n/a
6	n/a	115.98	n/a	n/a	n/a
7	n/a	139.65	n/a	n/a	n/a
8	6.62	126.55	C ₉	n/a	m
9	6.62	134.44	C ₈	H ₁₀	m
10	6.14	127.17	C ₉ , C ₈	H ₉	m
11	6.14	131.83	C ₁₂ , C ₁₃	H ₁₂	m
12	6.56	128.49	n/a	H ₁₁ , H ₁₃	m
13	5.82	132.91	C ₁₁ , C ₁₄	H ₁₂ , H ₁₄	dt, <i>J</i> = 14.9, 7.4 Hz
14	2.39	41.21	C ₁₅ , C ₁₃ , C ₁₂	H ₁₃ , H ₁₅ , H ₁₂	m
15	4.21	72.12	C ₁₆	H ₁₄ , H ₁₆	q, <i>J</i> = 6.4 Hz
16	5.57	132.33	C ₁₇ , C ₁₅ , C ₁₈	H ₁₅ , H ₁₇	dd, <i>J</i> = 15.2, 6.7 Hz
17	6.2	131.38	C ₁₈ , C ₁₅	H ₁₆ , H ₁₈	dd, <i>J</i> = 15.3, 10.4 Hz
18	6.04	130.79	C ₁₇	H ₁₇ , H ₁₉ , H ₂₀	ddq, <i>J</i> = 15.3, 10.5, 1.5 Hz
19	5.72	130.54	C ₁₇ , C ₂₀	H ₁₈ , H ₂₀	dq, <i>J</i> = 13.8, 6.7 Hz
20	1.75	18.27	C ₁₉	H ₁₉	dd, <i>J</i> = 6.7, 1.6 Hz
21	2.13	11.89	C ₅ , C ₆ , C ₇	n/a	s
22	5.24	102.49	C ₂₅ , C ₂₆ , C ₂₄ , C ₂₈ , C ₄	H _{23a} , H _{23b}	d, <i>J</i> = 4.4 Hz
23a	2.04	41.8	C ₂₂ , C ₂₄	H _{23b} , H ₂₂ , H ₂₈	dd, <i>J</i> = 14.2, 4.6 Hz
23b	2.54	41.8	C ₂₂ , C ₂₄ , C ₂₈	H _{23a} , H ₂₂	m
24	n/a	78.34	n/a	n/a	n/a
25	3.33	77.61	C ₂₆ , C ₂₇ , C ₂₄	H ₂₆	d, <i>J</i> = 9.5 Hz
26	4.08	69.26	n/a	H ₂₇ , H ₂₅	dq, <i>J</i> = 9.4, 6.1 Hz
27	1.41	18.55	C ₂₅ , C ₂₆	H ₂₆	d, <i>J</i> = 6.1 Hz
28	1.65	16.93	C ₂₅ , C ₂₄ , C ₂₃	H _{23a}	s
29	5.31	90.15	C ₃₀ , C ₃₁ , C ₃₃ , C ₂₄	H ₂₈ , H ₃₀	d, <i>J</i> = 1.7 Hz
30	3.41	81.53	C ₃₅ , C ₃₁ , C ₃₂	H ₃₁ , H ₂₉	m
31	3.73	71.37	C ₃₂	H ₃₁ , H ₃₂ , H ₃₀	dd, <i>J</i> = 9.5, 3.7 Hz
32	3.42	73.86	C ₃₄ , C ₃₃	H ₃₁ , H ₃₃	m
33	3.86	69	n/a	H ₃₄ , H ₃₂	dq, <i>J</i> = 9.5, 6.3 Hz
34	1.34	17.55	C ₃₃ , C ₃₂	H ₃₃	d, <i>J</i> = 6.3 Hz
35	3.51	59.22	C ₃₀	n/a	s
a	12.38	OH at C3	C ₂ , C ₃ , C ₄	n/a	s

atom number	δ_H	δ_C	HMBC	COSY	splitting pattern and <i>J</i> couplings
b	7.53	OH at C5	n/a	n/a	s
c	4.13	OH at C25	n/a	H ₂₅	s
d	2.27	OH at C31	n/a	H ₃₁	s
e	2.32	OH at C32	n/a	n/a	s
g	7.44	OH at C15	n/a	n/a	s

^adq: doublet of quartets, s: singlet, m: multiplet (cannot resolve), dd: doublet of doublets, d: doublet, q: quartet, dq: doublet of quartets, ddq: doublet of doublet of quartets.