Electronic Supplementary Information

Biosynthesis of the uridine-derived nucleoside antibiotic A-94964: identification and characterization of the biosynthetic gene cluster provide insight into the biosynthetic pathway

Taro Shiraishi, Makoto Nishiyama, and Tomohisa Kuzuyama

Biotechnology Research Center and Collaborative Research Institute for Innovative Microbiology, The University of Tokyo, 1-1-1, Yayoi, Bunkyo-ku, Tokyo, 113-8567, Japan

Address correspondence to: Tomohisa Kuzuyama. Fax, +81-3-5841-8030; E-mail, utkuz@mail.ecc.u-tokyo.ac.jp

Contents

Materials and methods ........................................... S1

Table S1. Deduced functions of ORFs in cosT49 ............... S4
Table S2. Oligonucleotides used in this study .................. S5

Figure S1. Schematic representation for gene deletion by λRed and Cre/loxP recombination systems ............. S6
Figure S2. Metabolite analysis of S. albus::cosT49Δorf19 .... S7
Figure S3. Domain architecture of Anb6-9 ....................... S8
Figure S4. Metabolite analysis of S. albus::cosT26Anb3 and S. albus::cosT49Anb8-9 .................. S9
Figure S5. Metabolite analysis of S. albus::cosT49Anb11 .... S10
Figure S6. Metabolite analysis of S. albus::cosT49Anb10::pKU1021anb10 ............... S11

References .......................................................... S12
Materials and methods

Bacterial strains, plasmids, and culture conditions. *Escherichia coli* DH5α was used for routine cloning, *E. coli* XL1BLUE MRF’ was used for the construction of a cosmid library, *E. coli* BW25141/pKD46 was used for the gene deletion in a cosmid by λRed recombination system, *E. coli* DH5α/pTH18cs::cre was used for the removal of the deletion cassette by Cre/loxp recombination system. *Streptomyces* sp. SANK 60404 is a gift from Daiichi Sankyo (Tokyo, Japan), which had previously been isolated from a soil sample collected in Okinawa, Japan. A pOJ446 vector was used to construct the cosmid library of *Streptomyces* sp. SANK 60404. *Streptomyces albus* G153 was used as a heterologous host for A-94964 production. *Streptomyces* sp. SANK 60404 or each *S. albus* transformant was inoculated into 10 mL TSB medium (30 g L⁻¹ tryptic soya broth) and incubated with shaking (300 rpm) at 30°C for 2 day. Two mL of the preculture was inoculated into 100 mL of the A-94964 production medium and incubated with rotating (180 rpm) at 27°C for 7 day. A-94964 standard is also a gift from Daiichi Sankyo.

Genomic DNA isolation, DNA sequencing and de novo assembly. *Streptomyces* sp. SANK 60404 mycelium was inoculated in TSB medium. After 2 days at 30°C, genomic DNA was isolated by phenol chloroform extraction. The isolated genomic DNA was then subjected to Illumina DNA sequencing. A 10-kb mate pair library sequencing was performed with an Illumina Genome Analyzer IIx. The Edena de novo short-reads assembler (Genomic Research Laboratory, Geneva, Switzerland) was used for de novo assembly. An assembly of the sequence reads yielded 1,565 contigs with 7,706,959 total base pairs.

Cosmid library construction, screening, and sequencing. Total DNA from *Streptomyces* sp. SANK 60404 was prepared and partially digested with Sau3AI. DNA fragments larger than 20 kb were ligated with a BamHI- and phosphatase-treated pOJ446, packaged with a LAMBDA INN packaging kit (Nippon Gene, Saitama, Japan), and introduced into *E. coli* XL1Blue MRF’ cells according to the manufacturer’s instructions. The resulting cosmid library of SANK 60404 was screened by colony hybridization with a DNA fragment containing the orf17 as a probe (ECL Direct; GE Healthcare Japan, Tokyo, Japan). A positive cosmid cosT26 was sequenced by the shotgun method (Genotech, Inc., Daejeon, Korea) and annotated with the FramePlot 4.0beta (http://nocardia.nih.go.jp/fp4/) and protein BLAST program (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Introduction of the cosmid library into *S. albus* G153 cell was performed with polyethylene glycol-mediated protoplast transformation. Protoplasts were prepared using the standard protocol.

Gene deletion by λRed recombination system. The schematic scheme is represented for gene
deletion by λRed recombination system (Fig. S1). \(^7\) Using λRed recombination system, each target gene was deleted by replacement with the aph gene that confers kanamycin-resistance. The plasmid pKU479 was used as a template for amplification of the aph gene inserted between two mut-loxP sequences. Each mut-loxP sequences-containing aph gene cassette with approximately 40-nt homology arms corresponding to the flanking regions of a target gene was amplified by PCR with a set of primers listed in Table S2. The amplified gene cassettes were individually introduced into E. coli BW25141/pKD46/cosT26 or E. coli BW25141/pKD46/cosT49. Kanamycin-resistant clones were then selected and each cosmide containing the mut-loxP sequences-containing aph gene cassette was prepared from the selected clones. The prepared cosmide was introduced into E. coli DH5α/pTH18cs::cre to remove the aph gene cassette by Cre/loxP recombination system. Kanamycin-sensitive clones were then selected and each cosmide, where the target gene is deleted, was prepared from the selected clones. The prepared cosmide was introduced into S. albus, and the resultant transformant was used for the following metabolite analysis.

**Gene complementation of anb10 in S. albus::cosT49Δanb10.** The anb10 gene was amplified by PCR from cosT49, using the primers listed in Table S2. The primers were designed to amplify the anb10 gene from the upstream region that includes its ribosomal binding site. A DNA fragment containing the anb10 gene was cloned downstream of the rpsJ promoter in pKU1021\(^7\) to give pKU1021anb10. S. albus::cosT49Δanb10 was transformed with pKU1021anb10 according to a previously reported protocol.\(^4\)

**Analysis of metabolites.** After cultivation of Streptomyces sp. SANK 60404 or each S. albus transformant, two-times volume of acetone was added to the culture broth. After the extraction by acetone, the acetone was evaporated in vacuo, the remaining residue was dissolved in 50% acetonitrile. The resultant solution was analyzed on an HPLC system (Jasco, Tokyo, Japan) equipped with a Capcell Pak C18 UG120 column (4.6 φ × 250 mm; Shiseido, Tokyo, Japan) under the following condition: mobile phase 50% acetonitrile + 50% acetonitrile with 5.0 mM triethylamine phosphate (pH 3.0) at flow rate of 1.0 mL/min. The resultant solution was also analyzed on a high-resolution Triple TOF 5600 MS instrument (SCIEX, Tokyo, Japan) equipped with a UFLC Nexera system (Shimadzu, Kyoto, Japan). In the MS and MS/MS analysis, a Capcell Pak C18 IF2 column (2.0 φ × 50 mm; Shiseido, Tokyo, Japan) or an ACQUITY UPLC BEH Amide column (2.1 φ× 50 mm; Waters, Tokyo) was used. LC condition was as follows: (i) for CAPCELL PAK C18 IF column, mobile phase A, 10% acetonitrile + 10 mM ammonium formate (pH 3.0); mobile phase B, 90% acetonitrile + 10 mM ammonium formate (pH 3.0); 2–98% B over 5 min, 98% B for 5 min, and then 2% A for 5 min, at a flow rate of 0.4 ml/min; (ii) for ACQUITY UPLC BEH Amide column, mobile phase A, 50%
acetonitrile + 10 mM ammonium acetate (pH 9.0); mobile phase B, 95% acetonitrile + 10 mM ammonium acetate (pH 9.0); 100–0% B over 5 min, 0% B for 2.5 min, and then 100% B for 2.5 min, at a flow rate of 0.4 ml/min. MS and MS/MS analyses were simultaneously performed using electrospray ionization in positive mode.
<table>
<thead>
<tr>
<th>Gene Bank ID</th>
<th>BLAST Hit Protein (CUGB)</th>
<th>ORF</th>
<th>Proposed function of each ORF encoded in cosT49, Accession number: LC41326.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>Table S1</strong>. Proposed functions of each ORF encoded in cosT49. Accession number: LC41326.</td>
</tr>
</tbody>
</table>
### Table S2. Oligonucleotides used in this study. Start codon and stop codon are underlined except for anb10comp-fw and anb10comp-rv.

<table>
<thead>
<tr>
<th>Oligonucleotide</th>
<th>Sequence</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>orf9p-fw</td>
<td>5'-gggcgggatgatcaccgggccgaggaa-3'</td>
<td>screening of cosmid library</td>
</tr>
<tr>
<td>orf9p-rv</td>
<td>5'-gggtacgcgtgcactgcctgaccacc-3'</td>
<td>screening of cosmid library</td>
</tr>
<tr>
<td>orf17p-fw</td>
<td>5'-gggtcgctcgactcgactcgcc-3'</td>
<td>screening of cosmid library</td>
</tr>
<tr>
<td>orf17p-rv</td>
<td>5'-gggtcgctcgactcgactcgcc-3'</td>
<td>screening of cosmid library</td>
</tr>
<tr>
<td>dorf5-fw</td>
<td>5'-AGCCGATGCACATACCTGTCGATCACCTGAGAAATCTCATGCCAGTGAATTCGAGCGACTCGAGT-3'</td>
<td>gene deletion</td>
</tr>
<tr>
<td>dorf5-rv</td>
<td>5'-GAGTGCCACCGTGCCTCCCGTTCACCGGACCGAATATCCGGGTACCGAGCGAACGCGTT-3'</td>
<td>gene deletion</td>
</tr>
<tr>
<td>dorf22-fw</td>
<td>5'-CCCGTGACGGTCACTGGTGCGAACAGCGGGTCAGAGCCATGCGCCTGGCCGCGTGTTGTGGTC-3'</td>
<td>gene deletion</td>
</tr>
<tr>
<td>dorf22-rv</td>
<td>5'-TGTCCGGCAGGCACCTACAGMAGAACCAGGTGGCCGCGTCGAGCGAACCGCGCTGTT-3'</td>
<td>gene deletion</td>
</tr>
<tr>
<td>dorf21-fw</td>
<td>5'-CCCTCCTCCCGGCCTCGCGGAGGTCCGCGTGGTCTCGCCGCGGAAGAGCCCGCCTGACCGC-3'</td>
<td>gene deletion</td>
</tr>
<tr>
<td>dorf21-rv</td>
<td>5'-GCCCTCAGGCGCCTCGCGGAGGTCCGCGTGGTCTCGCCGCGGAAGAGCCCGCCTGACCGC-3'</td>
<td>gene deletion</td>
</tr>
<tr>
<td>dorf20-fw</td>
<td>5'-GTACCCGACGGCCGCACCGCAGACCGAATATCCGGGTACCGAGCGAACGCGTT-3'</td>
<td>gene deletion</td>
</tr>
<tr>
<td>dorf20-rv</td>
<td>5'-GACCCGTCACGGGGTCCGGCCACTGGTGCGAACGGTGTG-3'</td>
<td>gene deletion</td>
</tr>
<tr>
<td>dorf19-fw</td>
<td>5'-ACGCCTGTCACCGGGCCACCGCAGACCGAATATCCGGGTACCGAGCGAACGCGTT-3'</td>
<td>gene deletion</td>
</tr>
<tr>
<td>dorf19-rv</td>
<td>5'-GCTCCTCCGCGGCCTCGCGGAGGTCCGCGTGGTCTCGCCGCGGAAGAGCCCGCCTGACCGC-3'</td>
<td>gene deletion</td>
</tr>
<tr>
<td>danb8-9-fw</td>
<td>5'-OTACCCGACGGCCGCACCGCAGACCGAATATCCGGGTACCGAGCGAACGCGTT-3'</td>
<td>gene deletion</td>
</tr>
<tr>
<td>danb8-9-rv</td>
<td>5'-ACGCCTGTCACCGGGCCACCGCAGACCGAATATCCGGGTACCGAGCGAACGCGTT-3'</td>
<td>gene deletion</td>
</tr>
<tr>
<td>danb3-rv</td>
<td>5'-ACGCCTGTCACCGGGCCACCGCAGACCGAATATCCGGGTACCGAGCGAACGCGTT-3'</td>
<td>gene deletion</td>
</tr>
<tr>
<td>danb11-rv</td>
<td>5'-ACGCCTGTCACCGGGCCACCGCAGACCGAATATCCGGGTACCGAGCGAACGCGTT-3'</td>
<td>gene deletion</td>
</tr>
<tr>
<td>anb10comp-fw</td>
<td>5'-GGGCTAGAAGGGGGCACGACGACCGAATATCCGGGTACCGAGCGAACGCGTT-3'(XbaI site underlined)</td>
<td>gene complementation</td>
</tr>
<tr>
<td>anb10comp-rv</td>
<td>5'-GGGCAACTCTCCAGACCGAATATCCGGGTACCGAGCGAACGCGTT-3'(HindIII site underlined)</td>
<td>gene complementation</td>
</tr>
</tbody>
</table>
Figure S1. Schematic representation for gene deletion by λRed and Cre/loxP recombination systems.
Figure S2. Metabolite analysis of *S. albus*:cosT49Δorf19. a. Extracted ion count chromatograms for a possible intermediate of A-94964 that lacks one sugar moiety (1) (m/z 907.3948 [M+H]+). b. MS spectrum of 1. c. MS/MS spectrum of 1. The predicted fragmentation patterns of 1 are also shown. d. MS/MS spectrum of A-94964. The predicted fragmentation patterns of A-94964 are also shown.
**Figure S3.** Domain architecture of Anb6-9. ACP, acyl carrier protein; ER, enoyl reductase; KS, ketosynthase; AT, acyltransferase; DH, dehydratase; KR, ketoreductase.
Figure S4. Metabolite analysis of *S. albus*:cosT26Δanb3 and *S. albus*:cosT49Δanb8-9. a. Extracted ion count chromatograms for 2 (m/z 631.1465 [M+H]+). b. MS spectrum of 2. c. MS/MS spectrum of 2. The predicted fragmentation patterns of 2 are also shown.
Figure S5. Metabolite analysis of *S. albus::cosT49Δanb11*. a. Extracted ion count chromatograms for 3 (m/z 333.09286 [M+H]+). b. MS spectrum of 3. c. MS/MS spectrum of 3. The predicted fragmentation patterns of 2 are also shown.
Figure S6. Metabolite analysis of \emph{S. albus::cosT49\Delta anb10}. Extracted ion count chromatograms for A-94964 (\textit{m/z} 1069.4476 \textit{[M+H]^+}) obtained via LC-MS analysis of the culture extracts from each transformant. The metabolite marked with asterisk (*) corresponds to an isomer of A-94964.
Reference