Nucleotide Sequence Analysis of the Carbomycin Biosynthetic Genes Including the 3-O-Acyltransferase Gene from *Streptomyces thermotolerans*

Akira Arisawa,* Hiroshi Tsubekawa, Kazuhiko Okamura, and Rokuro Okamoto

Mercian Corporation, Central Research Laboratories, 4–9–1 Johnan, Fujisawa-shi, Kanagawa 251, Japan

Received June 10, 1994

A 3.2-kb DNA fragment of the carbomycin biosynthetic region including the 3-O-acyltransferase gene (acyA) from *Streptomyces thermotolerans* was sequenced, and four ORFs were found in the fragment. The second ORF, designated ORF-A, was transcribed in the opposite direction to the other three ORFs. The first ORF was identified as *carA*, a gene for carbomycin resistance. The amino acid sequence of ORF-A was homologous to proteins of the cytochrome P-450 family. *Streptomyces lividans* transformed with pCB20, in which ORF-A was subcloned, epoxidized carbomycin B at its C-12, 13 positions, thus producing carbomycin A. The third ORF, the amino acid sequence of which showed a homology to macrolide antibiotics O-acyltransferases was identified as *acyA*. The last ORF (ORF-B), which starts just 3 bp downstream from the TGA termination codon of *acyA*, was thought to be a carbomycin 4-O-methyltransferase gene, because the amino acid sequence deduced from ORF-B showed high homology to a putative midemycin 4-O-methyltransferase encoded on *mdmC*.

Carbomycin, an acyl 16-membered macrolide antibiotic, is produced by *Streptomyces thermotolerans* ATCC 11416 (Fig. 1). As usual for antibiotics at large, the genes for biosynthesis of carbomycin are thought to be clustered in the chromosomal DNA of *S. thermotolerans*.4-7 Characterization and organization of the gene cluster would contribute not only to better understanding of basic mechanism for production of the macrolide antibiotics but also, to construction of genetically-engineered strains.3-4 A number of studies on biosynthetic genes of other macrolide antibiotics such as erythromycin,5-10 tylosin,11,12 and spiramycin13 have been done so far along this line.

In earlier studies on genes involved in carbomycin biosynthesis, two genes for resistance to carbomycin (*carA* and *carB*)14 and later, a mycarsyl 4'-O-seryl-CoA transferase gene (*carE*)15 were cloned by Epp et al. Independent of their work, we cloned two genes named *acyB1* and *acyB2*. *acyB1* was identical to the *carE* gene while *acyB2* was thought to be a regulatory gene that would enhance *acyB1* expression.15 We have also cloned the macrolide 3-O-acyltransferase gene (*acyA*) from a region distant from *acyB1* by chromosome walking.4

We report here the nucleotide sequences of a 3.2-kb DNA fragment of the carbomycin biosynthetic region that includes *acyA*, and three other genes. How these three genes in addition to *acyA* function will be discussed on the basis of the results of a homology search of the deduced amino acid sequence and the result of gene-expression in *S. lividans*.

**Materials and Methods**

*Bacterial strains and plasmids.* *S. thermotolerans* TH475, a mutant strain of *S. thermotolerans* ATCC11416 that had high acyltransferase activity but no activity for carbomycin-lactone formation was used as a donor strain for the carbomycin biosynthetic region. The isolation of original *acyA* plasmid p53A and subcloning of the *acyA* region into pMAA25 are described in our previous report.11 pUC118 (Takara Shuzo Co., Ltd., Kyoto, Japan) was used for the sequencing study. *S. lividans* TK2416 provided from D.A. Hopwood (John Innes Institute, Norwich, UK) was used as a host strain for subcloning and gene-expression of ORF-A. pLI35017 (from Hopwood) and pUC18 (Nippon Gene, Toyama, Japan) were used to construct pCB20, details of which will be shown later.

*Chemicals and enzymes.* Carbomycin B was purified from a culture broth of *S. thermotolerans* ATCC11416. A mixture of carbomycin A and B was obtained from Pfizer Inc. (Gorton, Conn., U.S.A.) and was used as a reference standard. T4 DNA ligase and restriction endonucleases except *AalII* were purchased from Takara Shuzo Co., Ltd., (Kyoto, Japan). *AalII* and T4 DNA polymerase were from Toyobo Co., Ltd., (Osaka, Japan).

*Sequencing and analysis of DNA.* A 2.1-kb *Sph*Ⅰfragment and a 1.05-kb *Sph*Ⅰ-*Sal*Ⅰfragment of p53A were subcloned separately into pUC118 and

---

* Corresponding author.

**Abbreviations:** kb, kilobase pair; ORF, open reading frame.
sequenced by the primer-walking method with a Sequenase version 2.0 kit using 7-deaza-dGTP in place of dGTP (United States Biological Corporation, Cleveland, Ohio, U.S.A.). Primers were synthesized with a DNA synthesizer model 381A (Applied Biosystems Inc., Foster City, Calif., U.S.A.) [25P]dCTP (30 TBq/mmol, Amersham, Buckinghamshire, UK) was used for labeling DNA. A short DNA region (SalI–SplH, 124 bp) was sequenced directly using denatured p33A DNA as a template by the method of Hattori and Sakaki.18 The sequence data was analyzed with GENETYX-Mac software programs (Software Development Co., Ltd., Tokyo, Japan) for localization of streptomycete genes, prediction of amino acid sequences, and homology search.

Construction of pCB20. A BglII linker (5'AGATCT3') was synthesized with the DNA synthesizer and inserted into the single HincII site of pUC18. The resulting plasmid (pUC-Bgl) was digested with BglII, blunted with T4 DNA polymerase, and ligated to PstI-digested pJ350, cohesive ends of which were also blunted to the polymerase, to construct an Escherichia coli-Streptomyces shuttle vector pSAN-lac. A 2.0-kb PstI–SplH fragment containing the ORF-A region from p33A was isolated from a 0.8% agarose gel and was purified with Gene Clean II (Bio 101, La Jolla, Calif., U.S.A.). The DNA fragment was subcloned into PstI–SplH site of pSAN-lac, to construct pCB20.

Expression of ORF-A in S. lividans. A transformant of S. lividans TK24 with pCB20 was cultivated in 25 ml of Tryptic Soy Broth (Difco), with 5 µg/ml thioperin at 28°C for 3 days. The mycelia were harvested from the culture broth and washed with 2 ml of 250 mM TE buffer (N-tris(hydroxymethyl)methyl-2-aminoethanesulfonic acid, pH 7.2). A 0.25-g portion as a wet weight of the mycelia was suspended in 1 ml of 250 mM TE buffer (pH 7.2) containing 0.2% glucose, 0.2% (w/v) trace elements solution,140 and 1.0 µg/ml of carbomycin B. The suspension was incubated at 28°C for 4 or 24 h with shaking at 200 rpm. After addition of 0.25 ml of 50% K2HPO4 to the suspension, carbomycin was extracted with 0.5 ml of ethylacetate. A three-µl portion of the extract was spotted on a TLC plate (Merck art. 5715) and developed with a solvent system of ethylacetate–diethylamine–H2O–methanol = (100:2:2:1) to detect epoxidized products of carbomycin B. For coloration, the TLC plate was immersed in 10% H2SO4 and baked at 120°C for 5 min. The TLC spots were printed by Quick Copy (Fuji Film, Tokyo, Japan).
Results

Sequence analysis of the acyA region

A 3.2-kb DNA fragment containing acyA from p53A was sequenced (Fig. 2). The 3267-bp nucleotide sequence was analyzed to search for some ORFs using a software program of frame analysis (Fig. 3). We found two complete and two truncated ORFs having the codon usage pattern specific for Streptomyces genes proposed by Bibb et al.10

The first ORF, from nucleotide position 1–192 encoding 63 amino acids, corresponded exactly to the C-terminal region of carA, which is reported as a carbomycin resistance gene.14,20 An inverted repeat found downstream from carA

Fig. 3. Frame Analysis of the acyA Region.
The average G+C contents of the first (−−−−), second (-----), and third (-----) positions in each triplet within a sliding 50 codons are plotted. The possible open reading frames in each direction (N1>, N2>, N1<, N2<, and N3<) are shown overhead, where N1> to N3> correspond to three shifts of reading frame for one strand of DNA, while N1< to N3< correspond to the other strand. Symbols (—) and (—) show, respectively, the location of ATG initiation codons and that of termination codons. The localization and orientation of four genes (carA, ORF-A, acyA and ORF-B) are shown together with some cleavage sites of restriction endonucleases, where B: BamHI, Bg: BglII, H: HindIII, P: PstI, S: SalI, Sp: SalII, X: XhoI. Lines at the bottom represent regions cloned in indicated plasmids.

Fig. 4. Harr Plot Comparisons13 of ORF-A Protein with EryF (A), and ORF-B Protein with MdmC (B).
A: Deduced product of ORF-A (this study) versus β-deoxyerythronolide B hydroxylase from Saccharopolyspora erythraea, EryF.8,102 B: Deduced product of ORF-B (this study) versus putative micamycin 4-O-methyltransferase from Streptomyces incurvatus, MdmC.21 In each case, the window was 5 and the stringency was 2. Numbers along axes refer to the residue numbers of each protein.
Fig. 5. Alignment of Amino Acid Sequences of the Macrolide-O-Acyltransferases.
Amino acid sequences encoded in acyA (this study), mdmB\(^{11}\) and acyBI (carE)\(^{1,11}\) are aligned with GENETYX-MAC, multi-alignment application program that uses Needleman–Wunsch algorithm.\(^{29}\) Numbers at the right-end of the sequences indicate the residue numbers of each sequence. Identities across all three sequences are indicated by asterisks.

(nucleotide positions 211–241) was supposed to function as a transcriptional terminator.

The second ORF (nucleotide positions 1483–248) was designated ORF-A, encoded 411 amino acids (\(M_r, 45,273\)) and was transcribed in the reverse direction to the other three ORFs. The amino acid sequence deduced from the nucleotide sequence was homologous to proteins of the cytomegalovirus family (data not shown). Maximum matching (37.4% identity) was given to 6-deoxyerythronolide B hydroxylase encoded on eryF from \(Saccharopolyspora erythraea\)^{6,10} (Fig. 4A). Especially, a heme-binding domain conserved in P-450 proteins was found in the amino acid sequence encoded 427–398, suggesting that the product from ORF-A might be an oxidase involved in the oxidation process of carbamylic biosynthesis.

The third ORF (nucleotide positions 1666–2835) consisting of 389 amino acids (\(M_r, 42,879\)) was confirmed to be acyA, because the deduced amino acid sequence of the gene showed a high homology to the \(mdmB\) product, midecamycin 3-O-acyltransferase from \(Streptomyces mycarfaciens\)^{21} (66.3% identity) and acyBI (carE) product\(^{7}\) (45.2% identity) (Fig. 5). In our previous study, \(S. lividans\) and \(S. fradiae\) transformed with pMAA25 (Fig. 3) containing this ORF expressed 3-O-acyltransferase activity.\(^{41}\)

A C-portion truncated gene, the coding sequence of which starts just 3 bp downstream from the TGA termination

codon of acyA (nucleotide position 2839) was designated to ORF-B. The amino acid sequence of ORF-B had a high homology to the \(mdmC\) product, which is a putative 4-O-methyltransferase of midecamycin from \(S. mycaraefaciens\)^{21} (79.0% identity) (Fig. 4B). This suggests that ORF-B is
carbomycin 4-O-methyltransferase gene.

Expression of ORF-A in S. lividans

To study the possible function of ORF-A, we subcloned a 2.0-kb PstI-SphI fragment including ORF-A in PC20 (Fig. 3). A bioconversion test of carbomycin B was done using the transformant of S. lividans with PC20 as described in Materials and Methods. Figure 6 shows that the transformant had an activity to epoxidize carbomycin B at its C-12, 13 positions, producing carbomycin A. S. lividans harboring pSAN-lac, in which the ORF-A fragment was not inserted, showed no activity for the epoxidation. From the results of the bioconversion experiment and the homology search described in the previous section, we concluded that ORF-A encoded a cytochrome P-450 epoxygenase that catalyzed, at C-12, 13, epoxidation of carbomycin B.

Discussion

When the carbomycin resistance genes carA and carB were cloned originally by Epp et al.,14 there was no information such as how far the two genes are distant from each other in the chromosomal DNA of S. thermotolerans. Later studies showed, however, that carE (acyB1) and gene(s) involved in the carbomycin-lactone formation, carG2,22 linked to carB.

By chromosome walking, we have cloned the 3.2-kb DNA fragment from S. thermotolerans containing the acyA gene and found that this gene was located in a region about 40–80 kb distant from acyB1 (carE).41 In this report, we showed from the nucleotide sequence analysis of the DNA fragment that the carbomycin-resistance gene carA, cytochrome P-450 epoxygenase gene (ORF-A) and a putative carbomycin 4-O-methyltransferase gene (ORF-B) were clustered around acyA. Therefore, all the carbomycin biosynthetic genes cloned so far are located within a region of about 80 kb that contains of two parts, i.e., carE (acyB1)-carB-carG and carA-ORF-A-acyA-ORF-B. Since all genes except carA in the latter cluster are thought to be responsible for modification of carbomycin-lactone, genes for 18-hydroxylation and 18-dehydrogenation of the lactone, which are not cloned yet, could be found near the region of acyA.

The nucleotide sequence presented in Figure 2 shows that ORF-A is transcribed as a monocistronic transcript, because it has a putative promoter sequence at its 5′-region and has an inverted repeat specific for a termination of transcription at its 3′-region. On the other hand, there is no promoter-like sequence upstream region of ORF-B, the protein-coding sequence of which starts immediately from the 3′-terminus of acyA. Hence, ORF-B would be transcribed polycistronically with acyA. The positional and probably, transcriptional relation between acyA and ORF-B in S. thermotolerans would correspond to that between mdmB and mdmC in S. mycophycinus.21

Cytochrome P-450 proteins are structurally-related oxidases found in a wide range of organisms including animals, plants, and bacteria. Some P-450 proteins in actinomycetes are involved in biosynthesis of antibiotics and other secondary metabolites.23–27 We found from the bioconversion experiment of carbomycin B that ORF-A encoded a cytochrome P-450 epoxygenase that catalyzed C-12, 13 epoxidation of carbomycin B. The amino acid sequence deduced from ORF-A showed the highest homology to 6-deoxyerythronolide B hydroxylase (37.4% identity) from Saccharopolyspora erythraea (Fig. 4A) and a little less homology to mycaminic IV hydroxylase (34.3% identity, data not shown) from Micromonospora griseorubida.27 These three P-450 proteins must be closely related to each other because they are all specific for the oxidation of macrolactone. Since it is suggested that mycaminic IV hydroxylase has both activities of 12, 13-epoxidation and 14-hydroxylation,27 it would be of significance to know which domains of the three P-450 proteins decide the specificities for hydroxylation and epoxidation of the macrolactone.

The deduced gene product of acyB1 (45.2% identity) than that of mdmB (66% identity), because both acyA and mdmB catalyze 3-O-acetylation, while acyB1 catalyses 4′-O-acetylation of macrolides. A comparison of these sequences may provide a basis for better understanding of what are determinants to specify the catalytic site of the macrolide antibiotics.

Acknowledgements. We thank H. Asou-Takahashi for excellent technical assistance.

References