

Analysis of an 8.1-kb DNA Fragment Contiguous with the Erythromycin Gene Cluster of *Saccharopolyspora erythraea* in the *eryCI*-Flanking Region

Andrew R. Reeves,¹ Gerhard Weber,² William H. Cernota,¹ and J. Mark Weber^{1*}
Fermalogic, Inc., Chicago, Illinois 60612,¹ and Biochemie, GmbH, R&D Microbiology, Kundl, Austria²

Received 20 May 2002/Returned for modification 18 July 2002/Accepted 15 August 2002

An 8.1-kb region of the *Saccharopolyspora erythraea* genome, significant for its contiguity to the known genes of the erythromycin biosynthetic gene cluster, was mutationally analyzed and its DNA sequence was determined. The region lies immediately adjacent to *eryCI*. The newly characterized region is notable for a large, 3.0-kb segment, predicted not to be translated, followed by four probable genes: an acetyltransferase gene, a protease inhibitor gene, a methyltransferase gene, and a transposase gene. Because the probable functions of the genes in this region are not required for erythromycin biosynthesis or resistance and because a deletion of a 6.0-kb portion of this region had no effect on erythromycin biosynthesis, this region marks the outside boundary of the erythromycin gene cluster. Therefore, *eryCI* represents the end of the cluster. These results complete the analysis of the erythromycin gene cluster and eliminate the possibility that additional sought-after pathway-specific structural or regulatory genes might be found within or adjacent to the cluster.

Microorganisms isolated from soil have been used for the production of bioactive molecules such as antibiotics for over a half century. The mycelial gram-positive actinomycetes are an exceptionally rich source of these compounds. Much effort has been devoted to genetic analysis of the antibiotic biosynthesis process (14), which has been greatly aided by the clustering of antibiotic biosynthesis and resistance genes in the chromosomes of these organisms. A related area of interest in this field concerns the need of the producing organism to control the level of production of these toxic molecules in its growth environment. In some antibiotic gene clusters this function is performed by a pathway-specific regulatory gene located within the cluster (2, 4, 10, 11, 20, 26, 35).

Erythromycin is a clinically important macrolide antibiotic; it is also the parent molecule for several commercially successful semisynthetic macrolide and ketolide derivatives (6). As such, it would be of both great biological and commercial interest to understand the regulatory mechanisms operating in *Saccharopolyspora erythraea* that control the level of production of this important antibiotic.

The erythromycin gene cluster was cloned prior to the advent of automated DNA sequencing; consequently, it was analyzed in fragments by several different laboratories. This could explain how the cluster had come to be referred to as “entirely known” (8, 9) when, in fact, significant information regarding a 6-kb region immediately downstream of *eryCI* was lacking. This oversight persisted, despite reports in the earlier literature that suggested that this region contained a sought-after pathway-specific regulatory gene(s) (12, 30). Furthermore, this region might have been predicted to contain additional early-stage deoxysugar biosynthesis genes, for example, a dTDP-glucose synthase or a dTDP-glucose dehydratase, that

had been found in the tylosin biosynthetic pathway gene cluster but not yet in the erythromycin gene cluster (19, 28).

In this study, then, we sought to determine the probable identities of the genes contained in this region contiguous with the erythromycin gene cluster and whether they are required for erythromycin biosynthesis. In so doing we completed the analysis of the erythromycin gene cluster, a project that had its beginnings 20 years ago with the cloning of the erythromycin resistance gene (29).

MATERIALS AND METHODS

Strains and plasmids. FL1347, a red variant derivative of *S. erythraea* ATCC 11635, and BCl1, a commercial derivative of *S. erythraea* ATCC 11635, were used for this study. A list of the strains and plasmids constructed or used in this study is given in Table 1.

Chemicals, biochemicals, and growth media. Erythromycin A, thiostrepton, kanamycin sulfate, ampicillin, and tetrazolium chloride were purchased from Sigma (St. Louis, Mo.). The sporulation agar used for *S. erythraea* BCl1 was E20A agar medium. E20A agar medium per 1 liter of aqueous solution is 5 g of Bacto Soytone (Difco), 5 g of soluble starch, 3 g of CaCO₃, 2.1 g of morpholinepropanesulfonic acid buffer, and 20 g of Bacto Agar. R2T2 medium (R2T medium [34] without the peptone) was used for protoplast regeneration for all strains of *S. erythraea* and as a sporulation medium for the FL1347 (red variant) strain of *S. erythraea*.

Plasmids constructed for DNA sequencing. The *eryCI*-flanking region was originally cloned in plasmid pMW3 during an earlier study (32). A 6.7-kb *HindIII*-*BglII* fragment from pMW3 was subcloned into *Bam*HI-*HindIII*-digested pUC19 to make pFL2000.

The remaining portion of the *eryCI*-flanking region was cloned from pMW3 by PCR to produce a fragment covering the region from the unique *BglII* site at the 3' end of open reading frame (ORF) ORFL2 (Fig. 1, site 11) to the 3' end of ORFL4. The reaction was performed according to the protocol of the manufacturer with the Expand Long Template PCR kit (Roche Molecular Biochemicals, Indianapolis, Ind.) by using kit buffer 3. The amplification reaction involved a heating step at 98°C for 4 min, followed by 30 cycles at 98°C for 30 s, 55°C for 1 min, and 68°C for 10 min. The PCR fragment was cloned into pCRtopo (Invitrogen, Carlsbad, Calif.) to form pFB46.

Plasmid pFL2070 was constructed to obtain additional nucleotide (nt) sequence information on ORFL3. The insert in pFL2070 was produced by using primers K16-3 (5'-CGGAATTCCTTGTGAGGCGCTGGGACGG-3' [the *EcoRI* site is indicated in boldface]) and K17-3 (5'-CAAAGCTTCATGCCCGCGGTAGTCCC-3' [the *HindIII* site is indicated in boldface]). The resulting

* Corresponding author. Mailing address: Fermalogic, Inc., 2201 West Campbell Park Dr., Chicago, IL 60612. Phone: (312) 738-0050. Fax: (312) 738-0963. E-mail: mark.weber@fermalogic.com.

TABLE 1. Plasmids constructed or used in this study and strains used in this study

Plasmid or strain	Parent plasmid	Plasmid insert or strain description	Reference
Plasmids			
pMW3	pIJ702	Original clone of DNA from <i>eryCI</i> -flanking region of NRRL 2338 red variant strain	32
pFB46	pCRtopo (Invitrogen)	3.5-kb PCR fragment from pMW3 blunt end cloned into pCRtopo for DNA sequencing	This study
pFL2000	pUC19	<i>HindIII</i> - <i>BglII</i> subclone from pMW3	This study
pFL2063	pFL8	Contains two PCR fragments from ends of <i>eryCI</i> -flanking region; precursor plasmid to pFL2067	This study
pFL2067	pFL8	The Km ^r gene from pUC4K, originally from Tn903 (Amersham Pharmacia Biotech), cloned into pFL2063 between clone 1 and clone 2 for deletion of <i>eryCI</i> -flanking region	This study
pFL2070	pUC19	PCR product for resequencing of ORFL3	This study
Strains			
<i>S. erythraea</i> FL1347		Red variant of ATCC 11635	This study
<i>S. erythraea</i> BC1		Commercial derivative of ATCC 11635	This study
<i>S. erythraea</i> FB50		Commercial BC1 strain carrying a 6-kb deletion in <i>eryCI</i> -flanking region generated by gene replacement by using plasmid pFL2067	This study
<i>S. erythraea</i> FL2067		FL1347 red variant strain carrying a 6-kb deletion in <i>eryCI</i> -flanking region generated by gene replacement by using plasmid pFL2067	This study
<i>E. coli</i> DH5 α -e		Host strain for transformations	Gibco BRL, Baltimore, Md.

PCR product was purified from an agarose gel and digested with *HindIII* and *EcoRI* and then ligated to pUC19, which had also been cut with *HindIII* and *EcoRI*.

Both strands of the inserts from plasmids pFL2000 and pFB46 were sequenced by Microsynth GmbH (Balgach, Switzerland). The insert from plasmid pFL2070 was sequenced at ACGT, Inc. (Northbrook, Ill.).

Plasmids for deletion of *eryCI*-flanking region. Plasmid pFL8, which facilitates integrative transformation into *S. erythraea* and autonomous replication in *Escherichia coli* and *Streptomyces lividans* (which allows screening for clones in *E. coli* according to the blue or white color obtained with 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside), was constructed through the ligation of *MseI*-digested pIJ487 (31) to *NdeI*-digested pBS+ (Stratagene, Inc., La Jolla, Calif.). *MseI* and *NdeI* have compatible overlapping ends. Two smaller *MseI* fragments were eliminated from pIJ487 during this step, with a portion of the unneeded fd terminator and a portion of the unneeded *melC2* gene removed. The ligation mixture was transformed into *E. coli* DH5 α -e, selecting for ampicillin resistance (Ap^r), and plasmid DNA was prepared from a collection of primary transformants. This mixture of plasmid DNA was transformed into *S. lividans* TK21, selecting for thiostrepton resistance (Th^r). DNA from the Th^r *Streptomyces lividans* transformants was transformed back into *E. coli*, and the resulting plasmid was designated pFL6. To create pFL8, plasmid pFL6 was digested with *BglII* and then religated; this removed the polylinker that was originally carried on pIJ487.

Plasmid pFL2067 was constructed to delete a 6.0-kb portion of the *eryCI*-flanking region from the chromosome. pFL2063 was first constructed from one 1.8-kb *EcoRI*-*BglII* DNA fragment (clone 1; Fig. 2A and C) carrying ORFL4 and part of ORFL3 generated by PCR with primer pair K16 (5'-CGGAATTCCAC CGAGTAGTCGTTGGC-3' [the *EcoRI* site is indicated in boldface]) and K17 (5'-GAAGATCTGCCACGATGAAGTACGT-3' [the *BglII* site is indicated in boldface]). Another *HindIII*-*BglII* DNA fragment (clone 2; Fig. 2A and C) carried DNA from the entire *eryCI* gene and a part of *ermE* and was generated by PCR with primer pair K18 (5'-GTAGATCTCCGCGGACCGTATCGA-3' [the *BglII* site is indicated in boldface]) and K19 (5'-GAAAGCTTCTCCCGCA ACGACTTCG-3' [the *HindIII* site is indicated in boldface]) (Fig. 2).

For PCR amplification the cycling conditions were a heating step at 95°C for 1 min, followed by 25 cycles at 95°C for 1 min, 65°C for 30 s, and 72°C for 2 min, with a final extension step at 72°C for 7 min. The reaction mixture contained 5 U of *pfuTurbo* polymerase in a final volume of 50 μ l. Mg²⁺ was added to a final concentration of 2 mM.

After the amplification steps, the PCR products were run out on a 1% agarose gel and purified with the GeneClean II kit (Bio 101). The purified DNA was digested with *EcoRI*-*BglII* (from primer pair K16-K17) and *HindIII*-*BglII* (from primer pair K18-K19) in double digestions. After another purification step, the two PCR products were ligated in a three-component ligation reaction to

EcoRI-*HindIII*-digested pFL8 Δ G (pFL8 with the unique *BglII* site removed) to form pFL2063. In a subsequent step, the 1.2-kb kanamycin resistance (Km^r) gene from pUC4K (Amersham Pharmacia Biotech, Piscataway, N.J.) was ligated as a *BamHI* fragment into the *BglII* site between the two cloned PCR fragments in pFL2063. The ligation mixture was transformed into *E. coli* DH5 α -e, and the cells were spread on Luria-Bertani plates containing ampicillin (100 μ g/ml) and kanamycin (40 μ g/ml). Km^r and Ap^r colonies were screened for the presence of the cloned PCR products by *ENase* digestion. The plasmids were digested with *EcoRI*, *BglII*, and *HindIII* in single and double digestions to verify the presence of the correct inserts. Several recombinant plasmids were obtained and were designated pFL2067 (Fig. 2A).

One-step integration and eviction of pFL2067. Plasmid pFL2067 was chromosomally integrated into *S. erythraea* by standard procedures (34), except that primary transformants were selected by using kanamycin (10 μ g/ml) instead of thiostrepton. Primary transformants appearing on R2T2 plates with kanamycin were patched onto a series of three E20A plates containing thiostrepton at 15 μ g/ml, kanamycin at 10 μ g/ml, and no added antibiotic, respectively. Of 96 colonies tested, all 96 grew on the kanamycin-supplemented plates and the plates containing no drug. Ninety-one colonies grew on the thiostrepton-supplemented plates. The five putative Km^r thiostrepton-susceptible (Th^s) strains with gene replacements were designated FB50a to FB50e.

Southern blotting verification of deletion in FB50 strains. Chromosomal DNA from strains FB50a to FB50e was prepared (13) and digested with *NorI*. The Southern blotting procedure (Roche Molecular Biochemicals) was performed, with the prehybridization and hybridization steps carried out at 60°C (Fig. 2). Hybridizations were performed in the presence of an *eryCI* probe (for the size and the location of the probe, see Fig. 2C). The 4.4-kb hybridizing band, band A, which corresponds to the native *eryCI*-flanking region and which extends into the *ermE* gene, appears only in the parent strain, BC1. In FB50 deletion candidates, band A is changed to a 5.1-kb band, band B.

PCR verification of deletion in FB50 strains. The approach described above produces a unique PCR fragment that would be created only if the Km^r gene was substituted for the deleted region. PCR primer pair K68-K69 (sequences below) was designed to amplify a 1,656-bp fragment that comprised the 1,268-bp Km^r gene and 388 bp of DNA located on both sides of the resistance gene (Fig. 2). Primer K68 (5'-GGAATTCTGG CTCAACTGCGCAGACAA-3' [where the boldface indicates *EcoRI*]) anneals in the *eryCI*-flanking region 175 bp upstream of the Km^r gene. Primer K69 (5'-CAGAGCTCTTGATCTGCGTTCCGGCTC-3' [where the boldface indicates *SacI*]) anneals in the *eryCI* gene 213 bp downstream of the Km^r gene. Both primer sequences are present in the strain BC1 chromosome, but no amplified product is expected in the parent strain because the 2-min extension reaction used would not amplify a PCR fragment that large. The estimated size of each of the fragments in all three lanes is 1,692 bp

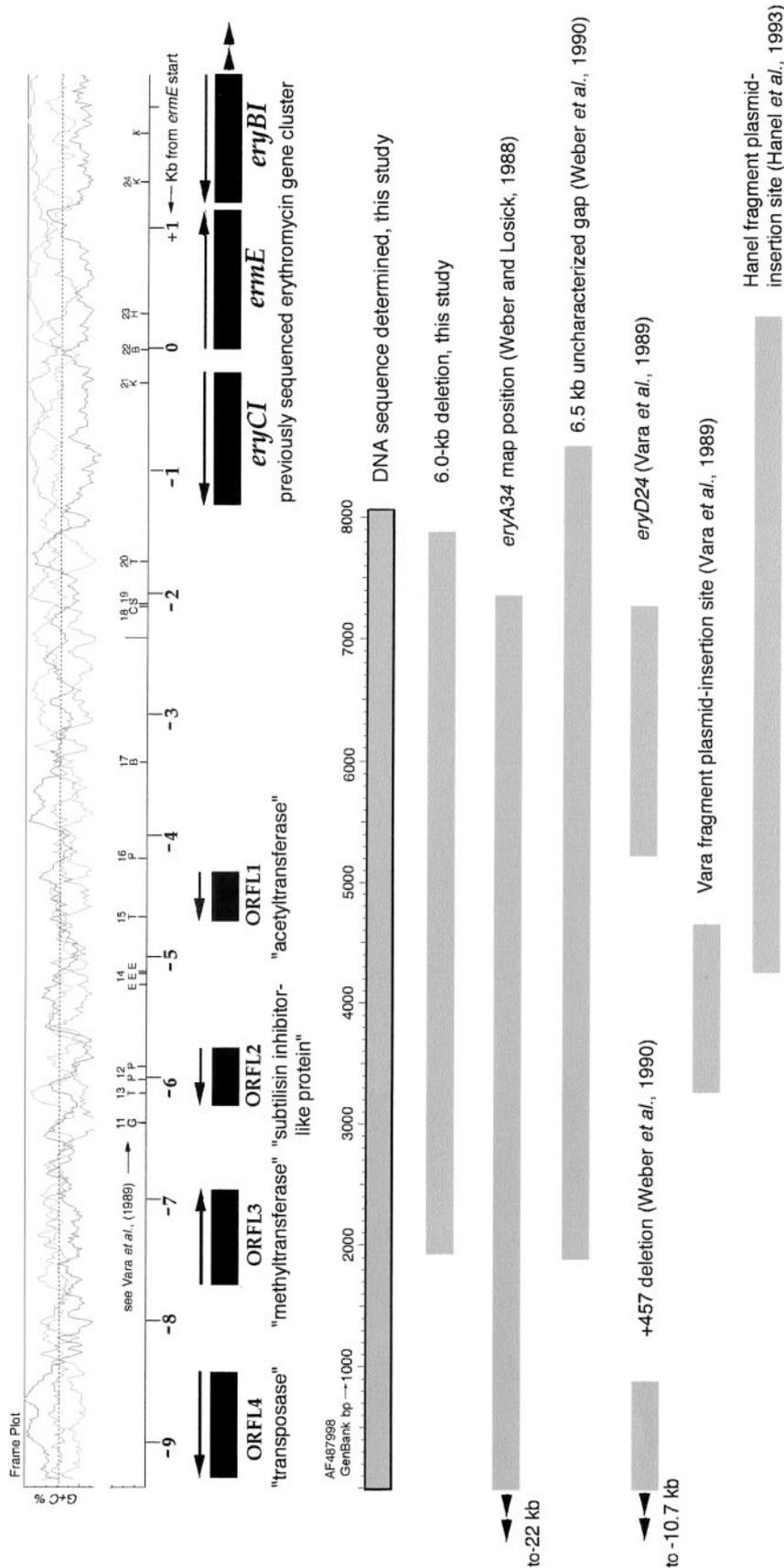


FIG. 1. Schematic map of the *eryCI*-flanking region of the *S. erythraea* erythromycin biosynthetic gene cluster. The ORFs are shown as black boxes, with the direction of transcription indicated by arrows above the black boxes. The map positions (in kilobases) starting from the first nucleotide of the *ermE* gene are given above the arrows. Above the map position scale are the locations of the restriction enzyme sites originally identified by Vara et al. (30) and refined in this study by DNA sequence analysis; compare this figure to Fig. 3 of Vara et al. (30). Restriction site abbreviations: G, *Bgl*II; B, *Bam*HI; C, *Cla*I; E, *Eco*RI; H, *Hind*III; K, *Kpn*I; P, *Pst*I; S, *Sph*I; T, *Syl*I. The frame plot (percent G+C content) for this region of the chromosome is shown at the top (3, 17). Below the ORFs are segments of the *eryCI*-flanking region that are referred to in the text and are represented by eight grey boxes. The uppermost grey box, with a black border, represents the extent of the region subjected to DNA sequence analysis in this study. The scale (in base pairs) at the top of this grey box correlates to the numbering of the sequence with GenBank accession no. AF487998.

TABLE 2. Characteristics of *eryCI*-flanking region

Feature	Size (bp)	Position (nt) in GenBank sequence	No. of amino acids	pI	Identity	Mol wt
Noncoding region	3,037	5061–8097	NA ^a	NA	Noncoding region	
ORFL1	387	4674–5060	129	9.68	Acetyltransferase	13,553
<i>att</i> site	34	4470–4503	NA	NA	Attachment site	
ORFL2 precursor protein	444	3144–3587	148	9.44	SSLI precursor protein	15,869
ORFL2 mature protein	336	3144–3479	112	6.77	SSLI mature protein	11,841
ORFL3	813	1654–2466	271	7.19	Methyltransferase	29,247
ORFL4	879	92–970	293	10.35	Transposase	33,026

^a NA, not applicable.

(extrapolated from a standard curve of the 1-kb ladder), which is very close to the known size of 1,656 bp. No amplified product of any size is observed in the lane containing strain BC1 DNA (lane 4). The sizes and the presence of the amplification products in lanes 1 to 3 confirm that the deletion has occurred in strains FB50a to FB50c. PCR conditions were as follows: total chromosomal DNA was isolated from three deletion candidate strains (strains FB50a to FB50c) and BC1 as described previously. The final Mg²⁺ concentration was 1.5 mM. All other reaction components were added as described above. Five units of *Taq* polymerase (MBI Fermentas, Vilnius, Lithuania) was used in all reactions. The cycling program consisted of a heating step without *Taq* polymerase for 5 min at 95°C, followed by 25 cycles of 1 min at 95°C, 1 min at 60°C, and 2 min at 72°C. A final extension cycle at 72°C for 7 min was performed to fill in protruding ends.

Shake-flask fermentation for production of erythromycin. Fermentations were performed on a Lab-line model 3525 incubator-shaker. Working stock spore suspensions were maintained at –80°C in 20% glycerol. Samples (25 µl) of spores from working stocks were transferred aseptically to capped test tubes (16 by 125 mm) containing 4 ml of sterile SCM broth (18). Test tube (seed) cultures were grown for 2 days at 32°C, with shaking at a 10° angle at 375 rpm (calibrated with a strobe light) on a 1-in. rotary displacement. Soluble complete medium (SCM) cultures (25 ml) containing 1.3 ml of soybean oil were started from 2.5 ml of the seed cultures and were allowed to grow for 5 days at 32°C and 375 rpm with a 1-in. rotary displacement. After a 5-day fermentation, the flasks were reweighed, and their weights were adjusted to the original weights by the addition of water to compensate for evaporation. The cultures were streaked onto agar plates to check for contamination.

Bioassay for erythromycin. A double-agar layer system on a large plate (245-mm² bioassay dish; catalog no. 431111; Corning Costar, Cambridge, Mass.) was used for the bioassay for erythromycin. The bottom agar layer consisted of 100 ml of tryptic soy broth (TSB) agar. Once the agar solidified (after it was allowed to sit for 30 min to 1 h at room temperature), a top agar layer was poured. Top agar consisted of 100 ml of TSB agar containing 250 µl of 1% triphenyl-tetrazolium red and a sufficient quantity of Th' *Bacillus subtilis* spores to produce a confluent lawn of growth. The upper layer was allowed to solidify at room temperature for 1 h. Broth samples (15 µl) from red variant strains FL1347 and FL2067 were spotted directly onto 1/4-in. bioassay discs (Schleicher & Schuell, Keene, N.H.) and allowed to dry for 30 min. Broth samples from strains producing larger amounts (strains BC1 and FB50) were first diluted 1:10 before being bioassayed to keep the erythromycin concentrations within the values on the standard curves. Stock erythromycin A solution (100 mg/ml) was prepared in 95% ethanol. Standard erythromycin A solutions to be used for the bioassay were prepared at 25, 50, 100, and 250 µg/ml in Tris-EDTA (pH 8.0) buffer. The bioassay plate was incubated overnight at 37°C. After incubation, the diameters of the inhibition zones were measured and converted to concentrations by using the standard curve produced for each plate.

Thin-layer chromatography. Thin-layer chromatography was performed by a previously described method (34), except that only a 1-ml sample of broth, with the appropriate scaling down of the amounts of the reagents used, was assayed.

HPLC. For determination of erythromycin A and by-products, a method based on high-performance liquid chromatography (HPLC)–mass spectrometry (MS) was used. Measurements were carried out on an Agilent Technologies 1100 series LC/MSD system that comprised a capillary pump, a micro-vacuum degasser, an autosampler, a thermostatted column compartment, and an MSD detector. The LC-MSD system was used with an electrospray ionization source. Complete system control and data evaluation were done on an Agilent Technologies ChemStation for LC-MS. The HPLC separation was carried out on an Asahipak ODP-50, 5-µm column (125 by 2.0 mm; Agilent Technologies) with an ammonium formate buffer-acetonitrile mobile phase mixture. Electrospray ionization-MS in single-ion monitoring mode was used for detection.

For calibration, an erythromycin A solution of known concentration was used.

Database searches and sequence alignments. The DNA sequence of the *eryCI*-flanking region was subjected to searches by use of the BLASTN, BLASTP, and BLASTX programs (1). Searches based on conserved protein family motifs were also performed.

Nucleotide sequence accession number. The nucleotide sequence of the *eryCI*-flanking region has been submitted to the GenBank database and has been given accession number AF487998.

RESULTS

Cloning of *eryCI*-flanking region. The *eryCI*-flanking region described in this report is 8.1 kb and extends from the 3' end of *eryCI* to the 3' end of ORFL4. By using the traditional *ery* gene cluster mapping index, the region extends from –1.25 kb to –9.4-kb (Fig. 1). For GenBank deposit AF487998, the nucleotide numbering proceeds from left to right, from ORFL4 to *eryCI* (Fig. 1). The DNA used for sequencing of the *eryCI*-flanking region was subcloned from plasmid pMW3 (32).

DNA sequence analysis of *eryCI*-flanking region. The *eryCI*-flanking region (Table 2) begins with a 3,037-bp region next to *eryCI* for which the BLASTX and BLASTN programs found no significant sequence similarities. FramePlot analysis (3, 17) of this region does not depict clearly defined ORFs, in contrast to the region to the right, within the *ery* cluster, where ORFs are densely packed and can be clearly visualized (Fig. 1). One potentially significant hairpin-loop structure (12-bp stem, two mismatches) was found between nt 7670 and 7703, and a 12-bp perfect direct repeat was between nt 7789 and 7825; both of these were close to the 3' end of the *eryCI* gene.

ORFL1, the reverse complement of nt 4674 to 5060 (GenBank accession no. AF487998), shows an AGAACG ribosome binding site 7 bp upstream from a GTG start codon, followed by 129 amino acid (aa) residues (molecular weight, 13,553; pI 9.58) and a TGA stop codon. Analysis of the deduced amino acid sequence of ORFL1 with the BLASTP program revealed that it has a large central domain (aa residues 27 to 107) characteristic of the GNAT family of acetyltransferases. This family of acetyltransferases is known to play an important role in antibiotic resistance, particularly to aminoglycoside antibiotics (7). Two aminoglycoside antibiotic resistance genes from *Acinetobacter* showed the highest degree of homology [aminoglycoside 6'-*N*-acetyltransferase from *Acinetobacter haemolyticus* (Protein Information Resource [PIR] accession no. I39502) and the aminoglycoside 6'-*N*-acetyltransferase-Ik protein from an *Acinetobacter* sp. (PIR accession no. I39505)]. The analysis with the BLASTP program also revealed 34% identity over a 45-amino-acid region to the sequence of a predicted protein from the kasugamycin biosynthesis gene cluster (an-

other aminoglycoside antibiotic) of *Streptomyces kasugaensis* M338-M1 (16).

The GNAT family also includes the histone acetyltransferases of eukaryotes. These enzymes catalyze the acetylation of lysine residues in histones, leading to transcriptional activation; however, the analysis with the BLASTP program did not identify this specific type of protein as homologous to ORFL1. Considering the close proximity of ORFL1 to the erythromycin gene cluster, it might have been thought to be a tailoring enzyme, but to the best of our knowledge, no acetylated tailored products of erythromycin have been reported, and our results show that ORFL1 plays no role in the biosynthesis of erythromycin A. ORFL1 is followed by a gap of 169 bp of noncoding DNA until an *att* site is reached at nt 4470 to 4503, and then another 881 bp separates the *att* site from the start codon for ORFL2.

ORFL2, a precursor protein of 148 aa residues with a molecular weight of 15,869 and a pI of 9.48, showed significant sequence similarity to subtilisin inhibitor-like (SSIL) precursor proteins (proteinaceous inhibitors of proteolytic enzymes) from a wide variety of actinomycetes. The predicted mature protein of ORFL2 would have 112 aa, a molecular weight of 11,841, and a pI of 6.77. The strongest match for ORFL2 (40% identity over 242 aa) was to another SSIL protein previously identified in *S. erythraea* by Pereda et al. (24) as ORF4 (referred to in this report as ORFR4) in the *eryK*-flanking region of the *ery* gene cluster. Analysis with the BLAST program revealed that ORFL2 has significant homologies to many other SSIL precursor proteins from other actinomycetes. SSIL precursor proteins are widely distributed in nature; however, little is known about their biological role or significance. One SSIL protein-coding sequence from *Streptomyces venezuelae* has been used commercially to provide the signal sequence for the large-scale production of secreted tumor necrosis factor alpha (25).

ORFL2 (precursor form) is the reverse complement of nt 3144 to 3587. It begins with an ATG start codon and is followed by a 36-aa signal peptide, as determined by SignalP computer analysis (21). SignalP analysis of ORFR4, the subtilisin-like inhibitor from the study of Pereda et al. (24), shows that it also has a classic signal sequence. No ATG or GTG start codon was found close to the ORFR4 signal sequence, indicating that the gene is either not translated or not translated by the usual mechanisms. We also determined the nucleotide sequence for the 5' flanking region of ORFR4 and found it to be in agreement with the results of Pereda et al. (24). Our interpretation of ORFR4, if it is translated, indicates that an additional 6 aa would exist on the 5' end of the mature protein described by Pereda et al. (24), which would then be preceded by a 37-aa signal peptide. Another similarity between these two genes is the stem-loop structures following both of these ORFs. The stem-loop structure (nt 2543 to 2573) is farther away from the end of ORFL2 than it is from the end of ORFR4. A large gap of 673 bp follows the end of ORFL2 before the gap reaches the 3' end of ORFL3.

The predicted amino acid translation for ORFL3 contains 271 aa and has a molecular weight of 29,247 and a pI of 7.19. From the analysis with the BLASTP program, the sequences with the two highest degrees of homology to ORFL3 (nt 1654 to 2466) were for uncharacterized proteins: the putative *S*-

adenosyl-L-methionine methyltransferase from *Sinorhizobium meliloti*, (38% identity over 231 aa; the protein with GenBank accession no. CAC49637) and the probable *O*-methyltransferase from *Mesorhizobium loti* (34% identity over 199 aa; the protein with GenBank accession no. q98g94). The analysis with the BLASTP program showed a domain in the predicted amino acid sequence of ORFL3 that is common to the UbiE and COQ5 family *C*-methyltransferases and a domain common to the PCMT family of *O*-methyltransferases.

O-Methyltransferases are known to play a role in antibiotic biosynthesis, either through the methylation of deoxysugars or through the polyketide portion of the structure (see reference 22 and references therein). One of the high-scoring ORFL3 homologs found by analysis with the BLASTP program was to a *C*-5-*O*-methyltransferase, the product of the *aveD* gene of the avermectin biosynthetic pathway in *Streptomyces avermitilis* (38% identity over 283 aa; GenBank accession no. JC6531 and PIR accession no. JC6531) (15). The protein encoded by *aveD*, in turn, shows significant homology to that encoded by *eryG*, the *O*-methyltransferase in the *ery* gene cluster (23, 33). Only one methyltransferase, the product of the *eryG* gene, has been found to be required for erythromycin A biosynthesis. Although ORFL3 is close to the *ery* gene cluster, it is not required for erythromycin A biosynthesis (see "Mutational analysis of *eryCI*-flanking region" below).

The deduced product of the ORFL4 gene, a protein of 293 aa residues with a molecular weight of 33,026 and the reverse complement of nt 92 to 970, shows a very high degree of sequence similarity to a probable transposase for IS1533 from *Mycobacterium tuberculosis* (57% identity over 264 aa; PIR accession no. D70668). ORFL4 has a homolog, ORFR6, in the *eryK*-flanking region. Interestingly, both transposases are located one ORF downstream of a subtilisin-like inhibitor and are transcribed in the same direction. The sequences of the ORFL4 and ORFR6 transposases, however, do not share a significant degree of homology to one another, and only the sequence of ORFR6 shows similarity to the transposase in the region between *eryAI* and *eryAII* (5).

Mutational analysis of *eryCI*-flanking region. Plasmid pFL2067 (Fig. 2) was constructed so that a 6.0-kb portion of the *eryCI*-flanking region encompassing the area between -1.5 and -7.5 kb (between nt 1925 and 7888) could be deleted. The deletion removed the 3' half of ORFL3, all of ORFL2 and ORFL1, and the 3-kb noncoding region next to *eryCI* (Fig. 1; see the grey bar labeled 6.0-kb deletion). Deletion strains FB50 and FL2067, for which the deletions were confirmed by PCR and Southern blot analysis (Fig. 2), were grown in shake flasks for erythromycin production (see Materials and Methods). The results of the agar plate bioassay (see Materials and Methods) revealed that the ability of the strains to make antibiotic was not noticeably affected, indicating that ORFs 1, 2, and 3 are not required for erythromycin production. Further analysis of fermentation broth extracts by thin-layer chromatography and HPLC showed that the bioactive material produced by the deletion mutants migrated and eluted identically to erythromycin A (data not shown).

In this study we also determined that ORFL4, the presumptive transposon homolog, would have been deleted in an earlier study (deletion +457; see Fig. 1 in reference 34). Since the +457 deletion was reported to have no effect on erythromycin

production, we can now conclude that ORFL4 is also not required for erythromycin production.

DISCUSSION

The erythromycin gene cluster had come to be referred to in the literature as “entirely known” (8, 9), although a significant segment of DNA contiguous with the cluster adjacent to *eryCI* had been left essentially uncharacterized (34).

The only information available regarding the *eryCI*-flanking region was presented in two reports that mapped erythromycin biosynthetic genes to this region by complementation analysis (30) and suggested the presence of pathway-specific regulatory genes through plasmid insertion experiments (12, 30). Since publication of those reports, no mutational analyses concerning this region have been reported and no DNA sequence information had been published or deposited.

The earlier results were not further developed, although any gene in this region that played a regulatory or structural role in erythromycin biosynthesis would have been considered part of the cluster. Also, there was significant potential commercial value in knowledge of this region. The potential was heightened because two genes predicted to play a role in the early steps of erythromycin deoxysugar biosynthesis were still unidentified (27, 28) and because no pathway-specific regulatory genes had yet been found for Em.

Ironically, we found that the erythromycin gene cluster does in fact end with *eryCI* and that there are no additional erythromycin biosynthetic or regulatory genes in the 8.1-kb *eryCI*-flanking region. Hence, references to *eryCI* as the boundary gene stand without need for correction. In this study it was shown that ORFs in this region code for enzymes that are not known to be required for erythromycin biosynthesis or resistance. Furthermore, deletion analysis showed that this region is not required for production of erythromycin.

Interestingly, pathway-specific regulatory genes have recently been reported for two other macrolide antibiotics; pikromycin has one regulator (35), and tylosin has five regulators (2). In each case it was reported that these regulatory genes are contiguous with or contained within their respective clusters. Therefore, the results reported here show that there are further differences between macrolide antibiotic gene clusters with respect to the number of regulatory genes that they contain.

In an earlier study dealing with this region, the *eryD24* locus, which defined a putative regulatory gene controlling deoxysugar biosynthesis, was mapped by complementation analysis (30) immediately downstream of *eryCI* (Fig. 1). The evidence presented in this report shows clearly that a deoxysugar biosynthesis or regulatory gene is not located in or near this area.

Two other previous results are relevant to this report. They concern reports of significant strain improvement effects caused by plasmid insertions in this area of the chromosome (12, 30). When the first plasmid, plasmid pWHM14 (30), is inserted into the *eryCI*-flanking region, it was reported to cause significant overproduction of erythromycin (Fig. 1). The 1-kb cloned fragment (the Vara fragment) responsible for the insertion overlapped portions of ORFL1 and ORFL2 and the region between them. The second plasmid, plasmid pWHM55, contained a 5.3-kb fragment (the Hanel fragment) that was

also reported to cause a significant increase in the level of erythromycin production. The Hanel fragment overlapped the complete *eryCI* gene, the 3-kb noncoding region next to *eryCI*, and all of ORFL1 (Fig. 1). The DNA sequence information reported here does not provide any further insights into the strain improvement effects reported earlier, for example, how the effects might be rationally explained or further exploited.

Some of the results from earlier studies that also dealt with the *eryCI*-flanking region were not further clarified by this study. For example, the *eryB* locus that Vara et al. (30) mapped downstream of *eryCI* was beyond the region analyzed in this study. If the mapping for the *eryB* locus done by Vara et al. (30) is correct, it will represent a rare case in which an essential antibiotic biosynthesis gene falls well outside the bounds of the antibiotic biosynthesis gene cluster. Additional DNA sequence analysis and mutational analysis (up to -20 kb) will be needed to answer this question.

Finally, the *eryA34* mutation, originally genetically mapped downstream of *eryCI* (32), was not found within the 8.1-kb region analyzed in this study. It could conceivably fall within the uncharacterized region farther to the left, up to -22 kb, but since the only known structural *eryA* locus is now known to be on the other (right) side of *ermE*, this earlier conclusion, based on less definitive genetic mapping data, is likely to be incorrect.

ACKNOWLEDGMENTS

We thank Roy Wesley and Keqian Yang for technical support and helpful comments.

REFERENCES

- Altshul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**:403–410.
- Bate, N., A. R. Butler, A. R. Gandecha, and E. Cundliffe. 1999. Multiple regulatory genes in the tylosin biosynthetic cluster of *Streptomyces fradiae*. *Chem. Biol.* **6**:617–624.
- Bibb, M. J., P. R. Findlay, and M. W. Johnson. 1984. The relationship between base composition and codon usage in bacterial genes and its use for the simple and reliable identification of protein-coding sequences. *Gene* **30**:157–166.
- Distler, J., A. Ebert, K. Mansouri, K. Pissowotzki, M. Stockmann, and W. Piepersberg. 1987. Gene cluster for streptomycin biosynthesis in *Streptomyces griseus*: nucleotide sequence of three genes and analysis of transcriptional activity. *Nucleic Acids Res.* **15**:8041–8056.
- Donadio, S., and M. J. Staver. 1993. IS1136, an insertion element in the erythromycin gene cluster of *Saccharopolyspora erythraea*. *Gene* **126**:147–151.
- Douthwaite, S., and W. S. Champney. 2001. Structures of ketolides and macrolides determine their mode of interaction with the ribosomal target site. *J. Antimicrob. Chemother.* **48**(Suppl. T 1):1–8.
- Dyda, F., D. C. Klein, and A. B. Hickman. 2000. GCN5-related *N*-acetyltransferases: a structural overview. *Annu. Rev. Biophys. Biomol. Struct.* **29**:81–103.
- Gaisser, S., G. A. Bohm, M. Doumith, M. C. Raynal, N. Dhillon, J. Cortes, and P. F. Leadlay. 1998. Analysis of *eryBI*, *eryBIII* and *eryBVII* from the erythromycin biosynthetic gene cluster in *Saccharopolyspora erythraea*. *Mol. Gen. Genet.* **258**:78–88.
- Gaisser, S., J. Reather, G. Wirtz, L. Kellenberger, J. Staunton, and P. F. Leadlay. 2000. A defined system for hybrid macrolide biosynthesis in *Saccharopolyspora erythraea*. *Mol. Microbiol.* **36**:391–401.
- Geistlich, M., R. Losick, J. R. Turner, and R. N. Rao. 1992. Characterization of a novel regulatory gene governing the expression of a polyketide synthase gene in *Streptomyces ambofaciens*. *Mol. Microbiol.* **6**:2019–2029.
- Gramajo, H. C., E. Takano, and M. J. Bibb. 1993. Stationary-phase production of the antibiotic actinorhodin in *Streptomyces coelicolor* A3(2) is transcriptionally regulated. *Mol. Microbiol.* **7**:837–845.
- Hanel, F., G. Schumann, G. Fiedler, and H. Krugel. 1993. Stimulation of erythromycin A yield by integration of a chromosomal DNA fragment including the *eryCI* gene into the chromosome of *Saccharopolyspora erythraea*. *Bio/Technol. Lett.* **15**:105–110.
- Hopwood, D. A., M. J. Bibb, K. F. Chater, T. Kieser, C. Bruton, H. M. Kieser,

- D. J. Lydiate, C. P. Smith, J. M. Ward, and H. Schremp. 1985. Genetic manipulation of *Streptomyces*: a laboratory manual. John Innes Foundation, Norwich, United Kingdom.
14. Hopwood, D. A. 1999. Forty years of genetics with *Streptomyces*: from in vivo through in vitro to in silico. *Microbiology* **145**:2183–2202.
 15. Ikeda, H., T. Nonomiya, M. Usami, T. Ohta, and S. Omura. 1999. Organization of the biosynthetic gene cluster for the polyketide anthelmintic macrolide avermectin in *Streptomyces avermitilis*. *Proc. Natl. Acad. Sci. USA* **96**:9509–9514.
 16. Ikeno, S., T. Tsuji, K. Higashide, N. Kinoshita, M. Hamada, and M. Hori. 1998. A 7.6kb DNA region from *Streptomyces kasugaensis* M338-M1 includes some genes responsible for kasugamycin biosynthesis. *J. Antibiot.* **51**:341–352.
 17. Ishikawa, J., and K. Hotta. 1999. FramePlot: a new implementation of the frame analysis for predicting protein-coding regions in bacterial DNA with a high G + C content. *FEMS Microbiol. Lett.* **174**:251–253.
 18. McAlpine, J. B., J. S. Tuan, D. P. Brown, K. D. Grebner, D. N. Whittern, A. Buko, and L. Katz. 1987. New antibiotics from genetically engineered Actinomycetes. I. 2-Norethromycins, isolation and structural determinations. *J. Antibiot.* **40**:1115–1122.
 19. Merson-Davies, L. A., and E. Cundliffe. 1994. Analysis of five tylosin biosynthetic genes from the *tylBA* region of the *Streptomyces fradiae* genome. *Mol. Microbiol.* **13**:349–355.
 20. Narva, K. E., and J. S. Feitelson. 1990. Nucleotide sequence and transcriptional analysis of the *redD* locus of *Streptomyces coelicolor* A3(2). *J. Bacteriol.* **172**:326–333.
 21. Nielsen, H., J. Engelbrecht, S. Brunak, and G. von Heijne. 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **10**:1–6.
 22. Patallo, E. P., G. Blanco, C. Fischer, A. F. Brana, J. Rohr, C. Mendez, and J. A. Salas. 2001. Deoxysugar methylation during biosynthesis of the antitumor polyketide elloramycin by *Streptomyces olivaceus*. Characterization of three methyltransferase genes. *J. Biol. Chem.* **276**:18765–18774.
 23. Paulus, T. J., J. S. Tuan, V. E. Lueke, G. T. Maine, J. P. DeWitt, and L. Katz. 1990. Mutation and cloning of *eryG*, the structural gene for erythromycin O-methyltransferase from *Saccharopolyspora erythraea*, and expression of *eryG* in *Escherichia coli*. *J. Bacteriol.* **172**:2541–2546.
 24. Pereda, A., R. Summers, and L. Katz. 1997. Nucleotide sequence of the *ermE* distal flank of the erythromycin biosynthesis cluster in *Saccharopolyspora erythraea*. *Gene* **193**:65–71.
 25. Pozidis, C., E. Lammertyn, A. S. Politou, J. Anne, A. S. Tsiftoglou, G. Sianidis, and A. Economou. 2001. Protein secretion biotechnology using *Streptomyces lividans*: large-scale production of functional trimeric tumor necrosis factor alpha. *Biotechnol. Bioeng.* **72**:611–619.
 26. Raibaud, A., M. Zalacain, T. G. Holt, R. Tizard, and C. J. Thompson. 1991. Nucleotide sequence analysis reveals linked n-acetyl hydrolase, thioesterase, transport, and regulatory genes encoded by the bialaphos biosynthetic gene cluster of *Streptomyces hygroscopicus*. *J. Bacteriol.* **173**:4454–4463.
 27. Salah-Bey, K., M. Doumith, J.-M. Michel, S. Haydock, J. Cortés, P. F. Leadlay, and M.-C. Raynal. 1998. Targeted gene inactivation for the elucidation of deoxysugar biosynthesis in the erythromycin producer *Saccharopolyspora erythraea*. *Mol. Gen. Genet.* **257**:542–553.
 28. Summers, R. G., S. Donadio, M. J. Staver, E. Wendt-Pienkowski, C. R. Hutchinson, and L. Katz. 1997. Sequencing and mutagenesis of genes from the erythromycin biosynthetic gene cluster of *Saccharopolyspora erythraea* that are involved in L-mycarose and D-desosamine production. *Microbiology* **143**:3251–3262.
 29. Thompson, C. J., T. Kieser, J. M. Ward, and D. A. Hopwood. 1982. Physical analysis of antibiotic-resistance genes from *Streptomyces* and their use in vector construction. *Gene* **20**:51–62.
 30. Vara, J., M. Lewandowska-Skarbek, Y. G. Wang, S. Donadio, and C. R. Hutchinson. 1989. Cloning of genes governing the deoxysugar portion of the erythromycin biosynthesis pathway in *Saccharopolyspora erythraea* (*Streptomyces erythraeus*). *J. Bacteriol.* **171**:5872–5881.
 31. Ward, J. M., G. R. Janssen, T. Kieser, M. J. Bibb, M. J. Buttner, and M. J. Bibb. 1986. Construction and characterisation of a series of multi-copy promoter-probe plasmid vectors for *Streptomyces* using the aminoglycoside phosphotransferase gene from Tn5 as indicator. *Mol. Gen. Genet.* **203**:468–478.
 32. Weber, J. M., and R. Losick. 1988. The use of a chromosome integration vector to map erythromycin resistance and production genes in *Saccharopolyspora erythraea* (*Streptomyces erythraeus*). *Gene* **68**:173–180.
 33. Weber, J. M., B. Schoner, and R. Losick. 1989. Identification of a gene required for the terminal step in erythromycin A biosynthesis in *Saccharopolyspora erythraea* (*Streptomyces erythraeus*). *Gene* **75**:235–241.
 34. Weber, J. M., J. O. Leung, G. T. Maine, R. H. Potenz, T. J. Paulus, and J. P. DeWitt. 1990. Organization of a cluster of erythromycin genes in *Saccharopolyspora erythraea*. *J. Bacteriol.* **172**:2372–2383.
 35. Wilson, D. J., Y. Xue, K. A. Reynolds, and D. H. Sherman. 2001. Characterization and analysis of the PikD regulatory factor in the pikromycin biosynthetic pathway of *Streptomyces venezuelae*. *J. Bacteriol.* **183**:3468–3475.