Sequencing and mutagenesis of genes from the erythromycin biosynthetic gene cluster of *Saccharopolyspora erythraea* that are involved in L-mycarose and D-desosamine production

Richard G. Summers,1 Stefano Donadio,1† Michael J. Staver,1 Evelyn Wendt-Pienkowski,2 C. Richard Hutchinson2 and Leonard Katz1

Author for correspondence: Leonard Katz. Tel: +1 847 937 4132. Fax: +1 847 938 3403. e-mail: leonard.katz@.abbott.com

1 Antibacterial Discovery Research Division, Abbott Laboratories, D-47P AP9A, 100 Abbott Park Road, Abbott Park, IL 60064, USA
2 School of Pharmacy, University of Wisconsin, Madison, WI 53706, USA

The nucleotide sequence on both sides of the *eryA* polyketide synthase genes of the erythromycin-producing bacterium *Saccharopolyspora erythraea* reveals the presence of ten genes that are involved in L-mycarose (*eryB*) and D-desosamine (*eryC*) biosynthesis or attachment. Mutant strains carrying targeted lesions in eight of these genes indicate that three (*eryBIV, eryBV* and *eryBVI*) act in L-mycarose biosynthesis or attachment, while the other five (*eryCII, eryCIII, eryCV, eryCV and eryCIV*) are devoted to D-desosamine biosynthesis or attachment. The remaining two genes (*eryBII* and *eryBVI*) appear to function in L-mycarose biosynthesis based on computer analysis and earlier genetic data. Three of these genes, *eryBII*, *eryCIIi* and *eryCII*, lie between the *eryAIII* and *eryG* genes on one side of the polyketide synthase genes, while the remaining seven, *eryBIV, eryBV, eryCV, eryVI, eryCV, eryCIV* and *eryBVI* lie upstream of the *eryAI* gene on the other side of the gene cluster. The deduced products of these genes show similarities to: aldohexose 4-ketoreductases (*eryBIV*), aldoketo reductases (*eryBII*), aldohexose 5-epimerases (*eryBVI*), the *dnmT* gene of the daunomycin biosynthetic pathway of *Streptomyces peucetius* (*eryBVI*), glycosyltransferases (*eryBV* and *eryCII*), the *AscC 3,4-dehydratase* from the ascyarlose biosynthetic pathway of *Yersinia pseudotuberculosis* (*eryCIV*), and mammalian *N*-methyltransferases (*eryCV*). The *eryCII* gene resembles a cytochrome P450, but lacks the conserved cysteine residue responsible for coordination of the haem iron, while the *eryCV* gene displays no meaningful similarity to other known sequences. From the predicted function of these and other known *eryB* and *eryC* genes, pathways for the biosynthesis of L-mycarose and D-desosamine have been deduced.

Keywords: desosamine, erythromycin, mycarose, sugar biosynthesis, *Saccharopolyspora*

INTRODUCTION

Erythromycin A is a clinically important macrolide antibiotic produced by the Gram-positive, filamentous soil bacterium *Saccharopolyspora erythraea*. Structurally, the antibiotic consists of three parts, a 14-membered macrolactone ring and two deoxysugars, L-cladinose and D-desosamine, that are attached to the ring. The pathway for the biosynthesis of erythromycin A has been partially elucidated by genetic studies (Paulus et al., 1990; Queener et al., 1978; Stassi et al., 1993; Weber et al., 1985, 1989, 1990, 1991) and is summarized in Fig. 1. Six phenotypic classes of mutants have been isolated; three carry lesions in either of the multistep pathways that lead to the biosynthesis of the macrolactone ring (*EryA*), L-mycarose (*EryB*) or D-desosamine (*EryC*). The remaining three phenotypic classes reflect defects in an individual enzymic step: hydroxylation of the macrolactone ring at C-6 (*EryF*) or C-12 (*EryK*), or
the O-methylation that converts mycarose to cladinose (EryG). A seventh mutational class (EryH) has been reported (Weber et al., 1990), but this may be the result of a polar mutation (Haydock et al., 1991).

The genes for the biosynthesis of erythromycin A, which occupy over 50 kb of the Saccharothrix erythraea chromosome, have been cloned (Cortes et al., 1990; Haydock et al., 1991; Stanzak et al., 1986; Tuan et al., 1990) and aspects of the organization of the gene cluster have been reported (Bevitt et al., 1992; Cortes et al., 1990; Donadio et al., 1991; Donadio & Katz, 1992; Haydock et al., 1991; Stassi et al., 1993; Weber et al., 1990, 1991). In addition, many of the genes involved in the biosynthesis of erythromycin A have been sequenced, including the eryA polyketide synthase genes (Cortes et al., 1990; Donadio et al., 1991; Donadio & Katz, 1992), the eryF (Haydock et al., 1991) and eryK (Stassi et al., 1993) P450 hydroxylase genes, the eryG O-methyltransferase gene (Haydock et al., 1991), and the erythromycin resistance gene, ermE (Uchiyama & Weisblum, 1985). The nucleotide sequences of a few of the genes required for the biosynthesis of the two deoxysugars have also been determined (Dhillon et al., 1989; Haydock et al., 1991), but the pathways for deoxysugar biosynthesis remain the least understood aspect of erythromycin biosynthesis.

To better understand the biosynthesis of the erythromycin deoxysugars L-mycarose and D-desosamine, we have determined the nucleotide sequence of two segments of the erythromycin biosynthetic gene cluster that lie immediately upstream and downstream of the eryA polyketide synthase genes. Within these segments we have identified ten genes that, by targeted mutation and/or computer-assisted analysis, seem to be involved in deoxysugar biosynthesis or attachment to the erythronolide ring. From the deduced identity of these and other previously characterized mycarose and desosamine biosynthetic genes we propose biosynthetic pathways for both deoxysugars. A preliminary account of this work has been previously reported (Donadio et al., 1993).

METHODS

Strains, bacteriophage and plasmids. The bacteriophages M13mp18 and M13mp19 (Yanisch-Perron et al., 1985) and the plasmids pUC19 (Yanisch-Perron et al., 1985) and pWHM3 (Vara et al., 1989) have been described previously. The erythromycin-producing strain Saccharothrix NRRL2338 was obtained from the Northern Regional Research Laboratory, and Escherichia coli DH5α was purchased from BRL/Life Technologies.

Media and enzymes. SGGP medium was described previously by Yamamoto et al. (1986). When selection was required,
thiostrepton (Sigma) was added to 10 μg ml⁻¹. R3M plates (l⁻¹): sucrose, 103 g; K₂SO₄, 0.25 g; yeast extract, 4 g; Casamino acids, 4 g; tryptone, 4 g; agar, 22 g; H₂O, 830 ml. After sterilization, 20 ml 2:5 M MgCl₂, 20 ml 50% glucose solution, 20 ml 2:5 M CaCl₂, 12:5 ml 2 M Tris/HCl (pH 7.0) 0.2 ml of a 5000-fold concentrated trace elements solution (Hopwood et al., 1985), 2:5 ml 1 M NaOH and 0:37 ml 0:5 M KH₂PO₄ was added. When selection was required, thiostrepton was added to 25 μg ml⁻¹. Yeast extract, Casamino acids, and tryptone were purchased from DIFCO. Restriction enzymes, T4 DNA ligase, the Klenow fragment of DNA polymerase I and S1 nuclease were purchased from BRL/Life Technologies or New England Biolabs.

DNA sequence analysis. Segments of the erythromycin gene cluster that flank both sides of the eryA genes were isolated from a previously described cosmid (Stassi et al., 1993), and pUC19, M13mp18 or M13mp19 subclones were sequenced using Sequenase version 2.0 with 7-deaza-dGTP (United States Biochemical) and S⁻¹[α-³²P]dCTP (NEN Research Products). Samples were run on urea-polyacrylamide gels that were occasionally supplemented with formamide (10%) to aid in the resolution of DNA sequence compressions. ORFs were identified and analysed using GCG programs (Devereux et al., 1984) and BLAST (Altschul et al., 1990).

Construction and analysis of mutants. Mutant strains of Sac. erythraea were constructed wherein the eryBIV, eryBVI, eryCII, eryCIII, eryCV, or eryCVI genes were disabled by replacement with altered derivatives of the wild-type allele. Typically, subclones of the erythromycin gene cluster in plasmid pUC18 that carry the gene of interest along with the polar effect on expression of the surrounding sequence. The eryBIV gene was altered by replacement with altered derivatives of the wild-type allele. Typically, subclones of the erythromycin gene cluster in plasmid pUC18 that carry the gene of interest along with the polar effect on expression of the surrounding sequence. To inactivate the eryBIV gene, the DNA segment carrying the gene was digested with NcoI followed by treatment with the Klenow fragment of DNA polymerase I and religation. This introduced a frameshift mutation at codon 152 of eryBIV that should cause a translational fusion to the downstream eryBIV gene. This mutation could have a polar effect on expression of the eryBIV gene. This inhibition was inactivated by deletion of the entire reading frame. Digestion with Sphl and Nhel followed by religation of the compatible cohesi ends removes the complete eryBIV locus without altering any of the surrounding sequence. To inactivate the eryBIV gene, the DNA segment carrying the gene was digested with BamHI followed by religation. This creates an in-frame deletion of almost the entire eryBIV gene. The eryBIV gene was altered by KpnI digestion followed by religation. This creates a 301 nt deletion that removes codons 95–305 of eryBIV. The eryCII gene was altered by KpnI digestion followed by religation. This creates a 301 nt out-of-frame deletion that initiates at codon 45 of eryCII and should result in premature translation termination within the eryCII locus. To mutate the eryCII locus, the gene was digested with Sphl followed by treatment with S1 nuclease, generating a 4 nt deletion that was confirmed by DNA sequencing. This deletion creates a frameshift after codon 94 that should result in premature translation termination within the eryCII locus. The eryCIV and eryCV genes were disabled simultaneously by RsrII digestion followed by religation. This deletes the last 59 codons of eryCIV and the first 102 codons of eryCV, and should create an out-of-frame translational fusion of eryCIV to eryCV sequences. The eryCV gene was also knocked out independently by XhoI digestion followed by treatment with the Klenow fragment of DNA polymerase I and then religation. This creates a frameshift mutation after codon 192 of the eryCV gene. Finally, 666 nt internal to the 711 nt eryCVI gene were deleted by generating two PCR fragments corresponding to the flanking regions of the eryCVI locus. Ligation of these fragments creates an in-frame deletion of almost the entire eryCVI gene. To provide at least 1.5 kb of DNA on each flank of the deletion, the PCR fragments were extended with contiguous DNA from suitable subclones of the original cosmID DNA.

After alteration, the gene and its surrounding DNA were transferred to plasmid pWHM3, an E. coli–Streptomyces shuttle plasmid that replicates poorly in Sac. erythraea. Homology-based, integrative transformation of Sac. erythraea protoplasts with the pWHM3 derivatives was then carried out as described by Weber & Losick (1988), except that R3M plates were used for growth on solid media. From several of the resulting thiostrepton-resistant transformants, the integrated plasmid was evicted by at least two rounds of non-selective growth in SGGP medium followed by protoplasting and plating for single colonies. The single colonies were then screened for thiostrepton-sensitivity, and chromosomal DNA from the sensitive strains was examined by Southern analysis (Sambrook et al., 1989) to determine if, following integration and excision, the mutated allele had replaced the wild-type copy. For each of the mutant strains analysed here, the predicted restriction map was confirmed by Southern analysis (data not shown).

Mutants were analysed for their ability to synthesize erythromycin A or its biosynthetic intermediates by TLC analysis of ethyl acetate extracts of basic (pH 9) culture broths, essentially as described by Weber et al. (1990). After removing the cells from 1–10 ml culture grown in SGGP media, 6 μl NH₄OH ml⁻¹ was added and the basic supernatant was extracted once with ethyl acetate. Thin-layer chromatograms of concentrated extracts were then developed with isopropyl ether: methanol: NH₄OH (70:35:2, by vol.) versus authentic samples of erythromelin B, 3-azetinopyrrolyl erythronolide B, and erythromycins A, B, C and D.

The mutants were also analysed for their ability to bioconvert erythromycin intermediates to erythromycin A. The EryB mutants were grown in SGGP media for 2–3 d at 30 °C and were then fed 3-azetinopyrrolyl erythronolide B (20 μg ml⁻¹ of culture) from a stock solution in DMSO (50 mg ml⁻¹). After overnight growth, the culture supernatant was analysed as above for the presence of erythromycin A. The EryC mutants were grown and analysed similarly, but these strains were fed 20 μg erythromycin D (ml culture)³ from a DMSO stock solution (50 mg ml⁻¹).

RESULTS AND DISCUSSION

Nucleotide sequence of the regions flanking the eryA genes

Taking advantage of the fact that antibiotic biosynthesis genes are almost always clustered on the chromosome or on plasmids of producing bacteria, we determined the nucleotide sequence of the chromosomal regions that flank the eryA genes, expecting to find more genes involved in erythromycin biosynthesis (Fig. 2). These sequences extend more than 3 kb downstream of eryAIII, filling the gap between eryAIII and eryG, and over 8 kb upstream of eryAI, covering the entire segment between eryAI and eryK.

Three ORFs whose existence had been described earlier were confirmed between eryAIII and eryG (Haydock et
These ORFs would be transcribed in the same direction as the upstream *eryAll* gene and the downstream *eryG* gene, and all three ORFs, the *eryAll* gene and the *eryG* gene are closely spaced, with the largest gap (30 nt) falling between the third ORF (*eryBII*) and *eryG*. Interestingly, the translational stop codon for *eryAll* overlaps the putative ATG start codon of the first ORF (*eryCII*), suggesting that the three ORFs and *eryG* might be transcribed as part of the same message as *eryAll*. There is some evidence, however, that *eryG* has an independent promoter (Weber et al., 1989).

On the other side of the *eryA* genes, in the gap between *eryAI* and *eryK* we identified seven ORFs that also appear to be organized in an operon-like fashion. Transcription of these ORFs, however, would diverge from the *eryAI* gene and converge on *eryK*. In general, the seven ORFs are fairly closely spaced, with the largest gap (153 nt) falling between the third (*eryCVI*) and fourth (*eryBVI) ORFs.

**Functional analysis of the erythromycin biosynthetic ORFs**

To determine whether the ORFs flanking the *eryA* genes encode components involved in erythromycin A biosynthesis, several strains were constructed in which the chromosomal copies of the wild-type alleles were replaced with mutated variants. The ability of the mutant strains to synthesize erythromycin A or its biosynthetic intermediates was then determined by TLC analysis of culture broth extracts. To ensure that the mutant phenotypes were not due to a polar effect generated by the intended lesion, the mutant strains were also assessed for their ability to bioconvert advanced intermediates in the erythromycin pathway. Two of the three ORFs between *eryAll* and *eryG* and six of the seven ORFs between *eryAI* and *eryK* were characterized in this way (Table 1).

Each of the mutants analysed fell into one of the two previously described phenotypic classes, *EryB* and *EryC*, related to deoxysugar biosynthesis (Weber et al., 1985). The *EryB* mutants accumulate the aglycone erythronolide B which is assayed for by TLC analysis as described in Methods, yet these mutants are fully capable of converting 3-α-L-mycarosyl-erythronolide B (or any later intermediate) to erythromycin A (also described in Methods). These mutants are blocked in their ability to synthesize mycarose and/or attach it to the erythronolide ring. Strains mutated in three of the seven ORFs lying between *eryAI* and *eryK* had an *EryB* phenotype. The *EryC* mutants, on the other hand, accumulate 3-α-L-mycarosyl-erythronolide B, and convert erythromycin D (or later intermediates) to erythromycin A. These mutants cannot synthesize desosamine and/or are unable to attach it the macrolactone ring. Mutation in three of the ORFs lying between *eryAI* and *eryK* resulted in an *EryC* phenotype, as did mutation in two of the ORFs downstream of *eryAll*. The roles of the remaining two ORFs (the final ORFs in each of the two collinear groups) were deduced from computer searches and earlier genetic data; each is most likely involved in mycarose biosynthesis. It is thus apparent that the mycarose and desosamine biosynthetic genes are interspersed among one another on both sides of the *eryA* polyketide synthase genes (Fig. 2). This is consistent with the lack of functional grouping that had been alluded to by previous studies (Dhillon et al., 1989; Haydock et al., 1991; Weber et al., 1990).
Early steps in mycarose and desosamine biosynthesis

Based on analogy to the biosynthesis of other bacterial 6-deoxysugars (Liu & Thorson, 1994), the biosynthesis of both mycarose and desosamine should proceed through a nucleotide-linked intermediate NDP-4-keto-6-deoxy-D-glucose. While the identity of the nucleotide bound to glucose is often central to the regulation of deoxysugar biosynthesis in some organisms (Liu & Thorson, 1994), it is not currently known what nucleotide is employed by *Sac. erythraea* in the biosynthesis of mycarose and desosamine. TDP-oleandrose is the nucleotide sugar precursor of the oleandrose units of the avermectins produced by *Streptomyces avermitilis* (Schulman et al., 1990), and in the tylosin producer, *Streptomyces fradiae*, TDP-4-keto-6-deoxy-D-glucose is thought to be the precursor for the three tylosin sugars. Moreover, in *Str. fradiae* the two genes responsible for the production of TDP-4-keto-6-deoxy-D-glucose have been localized to the tylosin gene cluster (Merson-Davies & Cundliffe, 1994). Homologues of these early biosynthetic genes have not been discovered in the erythromycin gene cluster (Vara & Hutchinson, 1988) and its gene has been cloned (Linton et al., 1993). Based on current understanding, it is possible that erythromycin production relies upon general cellular pools of TDP-4-keto-6-deoxy-D-glucose for the precursor for mycarose and desosamine biosynthesis, rather than having dedicated enzymes to produce this key intermediate.

From nucleotide-linked 4-keto-6-deoxy-D-glucose, the biosynthetic pathways for mycarose and desosamine should diverge. Schemes for the formation of each sugar are outlined in Figs 3 and 6, respectively.

Mycarose biosynthesis and genetics

As shown in Fig. 3, five steps should be required to convert NDP-4-keto-6-deoxy-D-glucose to NDP-1-mycoarose. These include 2,3-dehydration and reduction to form the 2-deoxysugar, 5-epimerization to form the 1-sugar, 3-C-methylation and 4-ketoreduction. While the order of these transformations has not been established, the chemistry involved places certain constraints on the pathway. Thus, ketoreduction at C-4 is most likely the last step since the presence of the ketone group would facilitate most of the reactions at the adjacent carbons. The 2,3-dehydration, on the other hand, probably occurs early since it requires the proton at C-3 and must therefore precede 3-C-methylation. Epimerization at C-5 could conceivably occur at any point prior to ketoreduction, yet, during the biosynthesis of ascaryllose, a 3,6-dideoxysugar, epimerization follows the dehydration and reduction steps (Thorson et al., 1994).

Of the ten ORFs analysed in this work, five appear to be involved in mycarose biosynthesis or its attachment to the erythronolide ring. One of these ORFs lies immediately upstream of the *eryG* gene on the *eryALL* side of the PKS genes; this ORF has been named *eryBII* in accordance with earlier studies of this region (Weber et al., 1990; Haydock et al., 1991). The remaining four ORFs lie between *eryAI* and *eryK*, and these have been named *eryBIV*, *eryBV*, *eryBVI* and *eryBVII*, respectively (Fig. 2). The characteristics and probable function of each of these genes, based largely on computer-assisted analysis, are summarized below.

*eryBII*. The *eryBII* gene lies immediately upstream of *eryG* gene in a region that was previously determined to harbour an *eryB* locus (Weber et al., 1990). The *eryBII*
product, a 333 amino acid polypeptide, shows the greatest similarity (36% identity) to the putative product of the moca gene of the mannityl opine utilization operon of Agrobacterium tumefaciens (Kim & Farrand, 1996). Although the role of moca is not known, the deduced product appears to belong to a family of alcohol dehydrogenases and/or aldoketo reductases that includes such diverse members as the aryI-alcohol dehydrogenase of the fungus Phanerochaete chrysorion (Reiser et al., 1994), the L-fucose dehydrogenase from Pseudomonas (Yamamoto-Otake et al., 1994), and the aflatoxin B1 aldehyde reductase from the rat (Ellis et al., 1993). By its inclusion in this family, it is likely that eryBII encodes one of the two reductases that should be required for mycarose biosynthesis (Fig. 3, steps 2 and 5). As described below, eryBIV most likely encodes the reductase that catalyses the last step in mycarose biosynthesis. Consequently, eryBII probably encodes the reductase that acts during the dehydration/reduction sequence that leads to the deoxygenation of C-2. Oddly, the deduced product of eryBII also displays 28% identity to the β-subunit of the voltage-gated K⁺ channel from the rat (Rettig et al., 1994). The significance of this relationship, however, is not clear.

**eryBIV.** The eryBIV gene is situated 225 nt upstream of eryAI and is the first of seven genes that would be divergently transcribed from eryAI. This gene spans 969 nt and should yield a protein of 34 kDa. A frameshift mutation in eryBIV causes accumulation of erythronolide B while allowing biotransformation of 3-deoxy-o-o-abequose pathway of Salmonella enterica (Wyk & Reeves, 1989), strL from the streptomycin biosynthetic pathway of Streptomyces griseus (Pissowotzki et al., 1991), and kde from an unidentified pathway in Sac. erythraea (Linton et al., 1995). While this enzyme family is considerably diverged at the level of primary structure (for example, EryBIV is only 19% identical to Rfb) and 20% identical to GalE), all members have a rigorously conserved GxxGxxG motif within the first twenty residues of their N-terminals that resembles the Rossmann fold motif characteristic of a nucleotide-binding site (Fig. 4) (Wierenga & Hol, 1983). Additionally, at about 100 residues downstream of the GxxGxxG sequence, these proteins contain a conserved YxxxKxxxD/E motif, the tyrosine and lysine residues of which have also been found to be involved in cofactor binding (Bauer et al., 1992). [It should be noted that similar motifs also characterize NDPsugar 4,6-dehydratases, such as the rfbB gene of Salmonella enterica (Jiang et al., 1991). In the dehydratases, however, the spacing between the two motifs is consistently greater, approaching 130 residues.] Mechanistically, GalE must reduce a ketone at C-4 during the interconversion of glucose and galactose, and, where the sugar substrate is known, the other members of this family are also thought to act upon a C-4 ketone. This strongly suggests that eryBIV encodes the C-4 ketoreductase involved in mycarose biosynthesis.

**eryBIV.** The eryBIV gene appears to start immediately downstream of eryBIV at a GTG codon that overlaps the TAG stop codon of eryBIV. Translation initiation at this site would yield a protein very similar in size to several closely related sequences, and the GTG codon is preceded by a likely RBS based on complementarity to the 3′ end of Streptomyces lividans (Bibb & Cohen, 1982) and the recently isolated Sac. erythraea 16S rRNAs (M. J. Staver & R. G. Summers, unpublished). The eryBIV gene extends for 1247 nt and encodes a protein of 45.5 kDa. Strains carrying a chromosomal deletion of the entire eryBIV gene have an EryB phenotype.

The protein encoded by eryBIV is most closely related to the deduced product of dnrS (46% identity) from the daunorubicin biosynthetic pathway of Streptomyces peucetius (Otten et al., 1995). The dnrS gene is thought
to encode the enzyme responsible for transfer of L-daunosamine to the aromatic polyketide ɛ-rodhozymycinone during daunorubicin biosynthesis. The strong similarity between EryBV and DnRS suggests that EryBV is most likely the mycarosyl transferase. Additional support for this notion comes from the degree of identity (43% identity) between the product of eryCIII, a desosamine biosynthesis or attachment gene (described later). Starting from NDP-4-keto-6-deoxyglucose, the biosynthetic pathways for mycarose and desosamine have no common steps. Both sugars must be transferred to the erythronolide ring, however, and related enzymes in each pathway might be expected to catalyze these transfers. Finally, although the degree of identity is lower, EryBV also resembles other glycosyltransferases such as the rhamnosyltransferase encoded by eryBVI (21% identity) of Pseudomonas aeruginosa (Ochsen et al., 1994).

epyBVI. This gene could start at a GTG codon 84 nt downstream from the eryCVI gene. This GTG codon is not preceded by a likely RBS, however, and a more likely start point is at an ATG codon 69 nt further downstream. From this codon, the gene would span 1463 nt and should encode a protein of 55 kDa. A strain that carries a 630 nt in-frame deletion within the eryBVI gene has an EryB phenotype, indicating that this gene is involved in mycarose biosynthesis.

The deduced product of eryBVI is most closely related to three sequences, a partial sequence from the tylosin pathway in Streptomyces fradiæ (58% identity over the first 98 residues; Merson-Davies & Cundliffe, 1994), ORF3 from the daunorubicin biosynthetic gene cluster of Streptomyces sp. strain C5 (41% identity; Dickers et al., 1996), and dnmT from the daunorubicin pathway of Streptomyces peucetius (41% identity; Scotti & Hutchinson, 1996). Like erythromycin, tylosin contains mycarose, and an eryBVI analogue in that gene cluster is not surprising. The dnmT gene, however, has been shown to be involved in the biosynthesis of L-daunosamine, a 2,3,6-trideoxy-3-amino aldoheoxose that shares just a few features with mycarose (Fig. 5). Common steps in the biosynthesis of daunosamine and mycarose should
include 2,3-dehydration, 5-epimerization and 4-keto-reduction. [The reduction that follows 2,3-dehydration during mycarose biosynthesis is not thought to be required for daunosamine formation (Otten et al., 1995).] Since the epimerization and ketoreduction steps in mycarose production are most likely catalysed by the products of eryBVII and eryBIV, respectively, this comparison suggests that eryBVI encodes the 2,3-dehydratase.

Interestingly, the deduced products of eryBVI, dnnnT and ORF3 appear to comprise two duplicated halves. For EryBVI, N-terminal residues 60–256 are 28% identical to C-terminal residues 302–500. Consequently, these enzymes may have two active sites.

**eryBVII.** The eryBVII gene most likely begins at a GTG codon 28 nt downstream from eryCV yielding a protein of 193 amino acids (=21290 Da). No mutations have been isolated in eryBVII, but the deduced product is related to a large family of sequences that have been postulated to be sugar C-5 or C-3, C-5 epimerases. While direct experimental evidence for this postulate is lacking, these genes are ubiquitous to gene clusters devoted to L-sugar biosynthesis, for example rfbC from the L-rhamnose pathway of *Salmonella* (Jiang et al., 1991; Xiang et al., 1994), ascE from the L-ascarylose pathway of *Yersinia pseudotuberculosis* (Thorson et al., 1994), strM from the L-dihydrostreptose pathway of *Streptomyces griseus* (Pissowotzki et al., 1991), and ORF4, which is probably involved in L-daunosamine biosynthesis, from the daunorubicin pathway of *Streptomyces griseus* (Krügel et al., 1993). Moreover, these genes have not been found in gene clusters devoted to D-sugar biosynthesis.

The putative EryBVII enzyme is most closely related to the deduced product of ORF4 from *Str. griseus* (61% identity). Because mycarose and daunosamine are both 2,6-dideoxy sugars, the similarity between their epimerases could be due to similarities in their substrates (i.e. 2,6-dideoxy intermediates). This would suggest that epimerization follows deoxygenation (as depicted in Fig. 3), which would be analogous to the order observed during L-ascarylose biosynthesis (Thorson et al., 1994).

The eryBVII gene seems to be the last deoxysugar-related biosynthetic gene on the right hand side of the erythromycin gene cluster as shown in Fig. 2. Downstream from eryBVII and partially overlapping its 3'-end is the previously characterized and convergently transcribed eryK gene (Stassi et al., 1993). Beyond this there do not appear to be any more erythromycin biosynthetic genes (Pereda et al., 1997).

The five eryB genes described above account for all of the steps required for mycarose biosynthesis except for C-methylation at C-3 (Fig. 3). An additional locus, eryBl, has been described, although its function is not known (Haydock et al., 1991). In light of the present data, it is possible that eryBl might encode the C-methyltransferase. Finally, there is some evidence for an another eryB locus, ORF3, immediately upstream of eryBl (Haydock et al., 1991; Weber et al., 1990). It is not clear, however, what role an additional eryB gene might play in mycarose biosynthesis unless multiple proteins are required for some steps in the pathway. Moreover, it has recently been shown that an analogue of ORF3 in the daunorubicin pathway of *Str. peucetius* is not required for daunosamine biosynthesis, suggesting that this is not a necessary sugar biosynthetic gene (C. R. Hutchinson, unpublished).

**Desosamine biosynthesis and genetics**

The biosynthetic pathway leading from NDP-4-keto-6-deoxy-D-glucose to NDP-D-daunosamine should comprise six steps (Fig. 6), and based on the genetic data described below, this would include a pyridoxamine-catalysed dehydrogenation and reduction that is similar to a reaction first described for the biosynthesis of ascarlose (Weigel et al., 1992). The remaining four steps would be an initial 4,3-keto isomerization to prepare the sugar for deoxygenation, 3-amination (also utilizing pyridoxamine), and finally, two methyl transfers to the newly added amino group. The order of the dehydration/reduction and amination steps is not known. If amination precedes the dehydration/reduction (opposite to the order depicted in Fig. 6), then the latter reactions would most likely employ a pyridoxal (versus pyridoxamine) cofactor, but the mechanism would remain essentially similar.

Like the eryB genes, the desosamine biosynthetic genes are distributed on both sides of the eryA polyketide synthase genes. Two desosamine biosynthetic or attachment loci, eryCII and eryCIII, lie immediately downstream of eryAI11. Three more desosamine biosynthetic genes, eryCIV, eryCV, and eryCv, are interspersed among the four eryB genes upstream of eryAI (Fig. 2). The characteristics of the eryC genes are summarized below.

**eryCII.** A mutant strain carrying a 301 nt deletion in the eryCII gene displays a typical EryC phenotype, i.e. it accumulates 3-a-~-mycarosyl erythronolide B, but can bioconvert erythromycin D to erythromycin A. This gene most likely starts at an ATG codon that overlaps the TGA stop codon of the eryAI11 polyketide synthase gene. While several alternative ATG start codons fall further downstream from eryAI11, only the overlapping ATG is preceded by a canonical purine-rich putative RBS. Assuming that translation initiates at this first ATG, the eryCII locus would span 1083 nt and encode a protein of 384 kDa.

The deduced product of eryCII is most closely related to the predicted product of dnrQ from the daunorubicin pathway of *Str. peucetius* (36% identity). The dnrQ gene is involved in L-daunosamine biosynthesis, although its exact function is not known (Otten et al., 1995). Interestingly, both the putative EryCII and DnrQ proteins display similarity to a number of P450 enzymes, but both proteins lack the rigorously conserved cysteine residue that normally coordinates the haem iron.
A comparison of the proposed pathways for desosamine biosynthesis (Fig. 6) and daunosamine biosynthesis (Otten et al., 1995) fails to reveal where these two proteins might act. Based on the proposed function of the eryC genes described below, however, it is possible to speculate that eryCII encodes either the 4-keto-6-deoxyglucose isomerase (step 1) or the reductase that completes deoxygenation at C-4 (step 3) since these steps cannot be accounted for by other gene products. However, neither of these steps is thought to be required for daunosamine biosynthesis (Otten et al., 1995).

**eryCIII.** The eryCIII gene most likely begins 10 nt downstream of eryCII at an ATG codon that is preceded by a possible RBS. The gene extends for 1266 nt and should encode a protein of 45920 Da. A frameshift mutation that would cause premature translation termination within eryCIII causes the mutant strain to accumulate 3-α-l-mycarosylerythronolide B as expected for an EryC mutant. Earlier studies of this region of the chromosome had suggested that this locus might be involved in mycarose biosynthesis (it was originally included within the eryBII locus; Weber et al., 1990; Haydock et al., 1991); however, these earlier results are most easily explained by a polar effect of the original disruption mutations over the downstream eryBII gene.

As described above, the deduced products of eryCIII and eryBV are closely related to each other (43% identity) and to the presumed daunosamine transferase encoded by dnrS (Otten et al., 1995), suggesting that eryCIII encodes the NDP-α-desosamine:3-α-l-mycarosyl erythronolide B transferase. The EryCIII protein is also 28% identical to the rhamnosyltransferase encoded by the rhlB gene of *P. aeruginosa* (Ochsner et al., 1994).

**eryCVI.** The eryCVI gene probably begins at an ATG codon 53 nucleotides downstream of eryBV. This locus spans 713 nt and should encode a protein of 26 kDa. An in-frame deletion of almost the entire locus yields an EryC phenotype.

The deduced EryCVI enzyme is most closely related to the putative product of rdmD from *Streptomyces purpurascens* (Niemi & Mäntsälä, 1995) (55% identity), and it also strongly resembles the predicted product of srmX (52% identity) from the spiramycin biosynthetic gene cluster of *Streptomyces ambofaciens* (Geistlich et al., 1992). An alignment of the deduced sequences of EryCVI, RdmD and SrmX reveals a strongly conserved nonapeptide motif, LLDVACGTG, that runs from residues 42 to 50 of the three proteins. When this sequence was used to search the GenBank/EMBL database, strong homology to several mammalian N-methyltransferases was discovered. Indeed, the predicted EryCVI product is 26% identical to the glycine N-methyltransferase of the rat (Ogawa et al., 1987). Since desosamine is an N-dimethylamino sugar, these findings suggest that eryCVI encodes the desosaminyl N-dimethyltransferase. Dimethylation by a single enzyme is not uncommon, as exemplified by the ErmC N-methyltransferase of *Staphylococcus aureus* (Denoya & Dubnau, 1989). By analogy, rdmD and srmX also probably encode the sugar N-dimethyltransferases that are required for the biosynthesis of the rhodosamine moiety of rhodomycin and the mycaminose moiety of spiramycin, respectively.

**eryCIV.** The eryCIV locus most likely initiates at an ATG codon that overlaps the TGA stop codon of the eryBVI gene and ends at a TGA codon 1202 nt downstream, yielding a predicted protein of 43.3 kDa. Insertional
inactivation of eryCIV and a deletion mutation that also encompasses part of the downstream eryCV gene cause an EryC phenotype. While it is formally possible that this phenotype is due solely to a polar effect of the insertion or loss of the eryCV gene in the deletion mutant, this possibility seems unlikely. The putative EryCIV enzyme is related to the predicted products of a number of genes involved in deoxy and deoxynamino sugar biosynthesis including tyIB from the tylosin biosynthetic gene cluster of Str. fradiae (Merson-Davies & Cundliffe, 1994), dnrJ from the daunomycin biosynthetic gene cluster of Str. peucetius (Stutzman-Engwall et al., 1992), and, interestingly, the previously characterized eryCI gene (Dhillon et al., 1989). The predicted EryCIV enzyme is also 25% identical to the mechanistically well-characterized AscC dehydratase from the asparagine pathway of Y. pseudotuberculosis (Thorson et al., 1994). AscC is a novel, pyridoxamine-dependent dehydratase (Weigel et al., 1992). It has been proposed, however, that many members of this family are more likely to be aminotransferases based on their ubiquitous occurrence in amino sugar biosynthetic pathways and the fact that pyridoxamine is the cofactor commonly employed for biological transaminations (Thorson et al., 1993). The finding of two AscC-like gene products (EryCI and EryCIV) devoted to desosamine biosynthesis provides some support for this idea since a desosamine biosynthesis requires both deoxygenation and amin transfer.

Comparisons of the putative products of eryCIV and eryCI to their homologues from other amino sugar pathways suggests that eryCI is the gene for the transaminase, leaving eryCIV to encode the dehydratase. Most telling is the strong relationship between the EryCI enzyme and the deduced TyIB protein (61% identity) contrasted with the relationship between EryCIV and TyIB (31% identity). TyIB participates in the biosynthesis of D-mycarosamine, an analogue of D-desosamine that is not deoxygenated at C-4 (Fig. 5).

**eryCV.** The eryCV gene possibly starts at an ATG codon 38 nt downstream from eryCIV. Assuming translation initiation at this site, the eryCV gene would span 1469 nt and should encode a protein of 539 kDa. While a frameshift mutation at this locus yields an EryC phenotype, the deduced product of eryCV displays no meaningful similarity to current entries in the GenBank/EMBL database. As was the case for eryCI, based on the likely function of the other eryC genes, it is possible that eryCV is involved in one of the unassigned steps in Fig. 6, encoding either the initial isomerase or the reductase.

**Conclusion**

The ten genes analysed here, plus the previously described eryBI (Haydock et al., 1991) and eryCI (Dhillon et al., 1989) genes probably account for all of the genetic information required to convert NDP-4-keto-6-deoxyglucose into L-mycarose and D-desosamine, and to attach them to the erythronolide ring.

Unlike the situation in the tylosin-producer, Str. fradiae (Merson-Davies & Cundliffe, 1994), the genes for production of the early intermediate NDP-4-keto-6-deoxyglucose do not appear to be associated with the erythromycin biosynthetic gene cluster. It is possible that this precursor is obtained from general cellular pathways.

While our interpretations are based, in part, on TLC data which cannot unambiguously establish the chemical identity of erythromycin derivatives, in one case, TLC data suggest that novel derivatives may have been produced by a strain mutated in eryBIV, most likely the 4-keto-mycarosyl derivatives of erythromycins A and C. If this can be confirmed by rigorous chemical analysis, it offers the hope that additional erythromycin derivatives might be created by rational manipulation of the mycarose and desosamine biosynthetic pathways.

**ACKNOWLEDGEMENTS**

We would like to thank T. Kavanaugh, L. Zulawinski and S. Nannapaneni for synthesis of oligonucleotides.

**REFERENCES**


expression of genes from tylosin biosynthetic genes from the galE and genes, one of which complements an early function in daunorubicin biosynthesis of novel deoxysugars by bacteria.

The erythromycin-producing Agrobacterium tumefaciens thymidine diphospho-4-keto-6-deoxyglucose 3,5-epimerase from Arabidopsis thaliana catalyzes the hydroxylation step in erythromycin biosynthesis.

Cloning and characterization of the EryR and EryO, two proteins responsible for synthesis of this opine by the plant tumour.


An ethoxyquin-inducible aldehyde reductase from rat liver that metabolizes aflatoxin B1 defines a subfamily of aldo-keto reductases. Proc Natl Acad Sci USA 90, 10350-10354.


Cloning and sequence analysis of genes encoding a daunorubicin biosynthesis enzyme and a glycosyl transferase involved in daunorubicin biosynthesis.

Cloning and characterization of the Streptomyces peucetius dnrQS genes encoding a daunorubicin biosynthesis enzyme and a glycosyl transferase involved in daunorubicin biosynthesis.

Cloning and characterization of the Streptomyces peucetius dnrQS genes encoding a daunorubicin biosynthesis enzyme and a glycosyl transferase involved in daunorubicin biosynthesis.

Cloning and characterization of the Streptomyces peucetius dnrQS genes encoding a daunorubicin biosynthesis enzyme and a glycosyl transferase involved in daunorubicin biosynthesis.

Cloning and characterization of the Streptomyces peucetius dnrQS genes encoding a daunorubicin biosynthesis enzyme and a glycosyl transferase involved in daunorubicin biosynthesis.

Cloning and characterization of the Streptomyces peucetius dnrQS genes encoding a daunorubicin biosynthesis enzyme and a glycosyl transferase involved in daunorubicin biosynthesis.

Cloning and characterization of the Streptomyces peucetius dnrQS genes encoding a daunorubicin biosynthesis enzyme and a glycosyl transferase involved in daunorubicin biosynthesis.


Received 10 June 1997; accepted 26 June 1997.