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 Hutchinson[2] and Leonard Katz[1]

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*, *eryCIV*, *eryCV* and *eryCVI*) are devoted to D-desosamine biosynthesis or attachment. The remaining two genes (*eryBII* and *eryBVII*) 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*, *eryCVI*, *eryBVI*, *eryCIV*, *eryCV* and *eryBVII* 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 (*eryBVII*), the *dnmT* gene of the daunomycin biosynthetic pathway of *Streptomyces peucetius* (*eryBVI*), glycosyltransferases (*eryBV* and *eryCIII*), the AscC 3,4-dehydratase from the ascarylose biosynthetic pathway of *Yersinia pseudotuberculosis* (*eryCIV*), and mammalian *N*-methyltransferases (*eryCVI*). 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
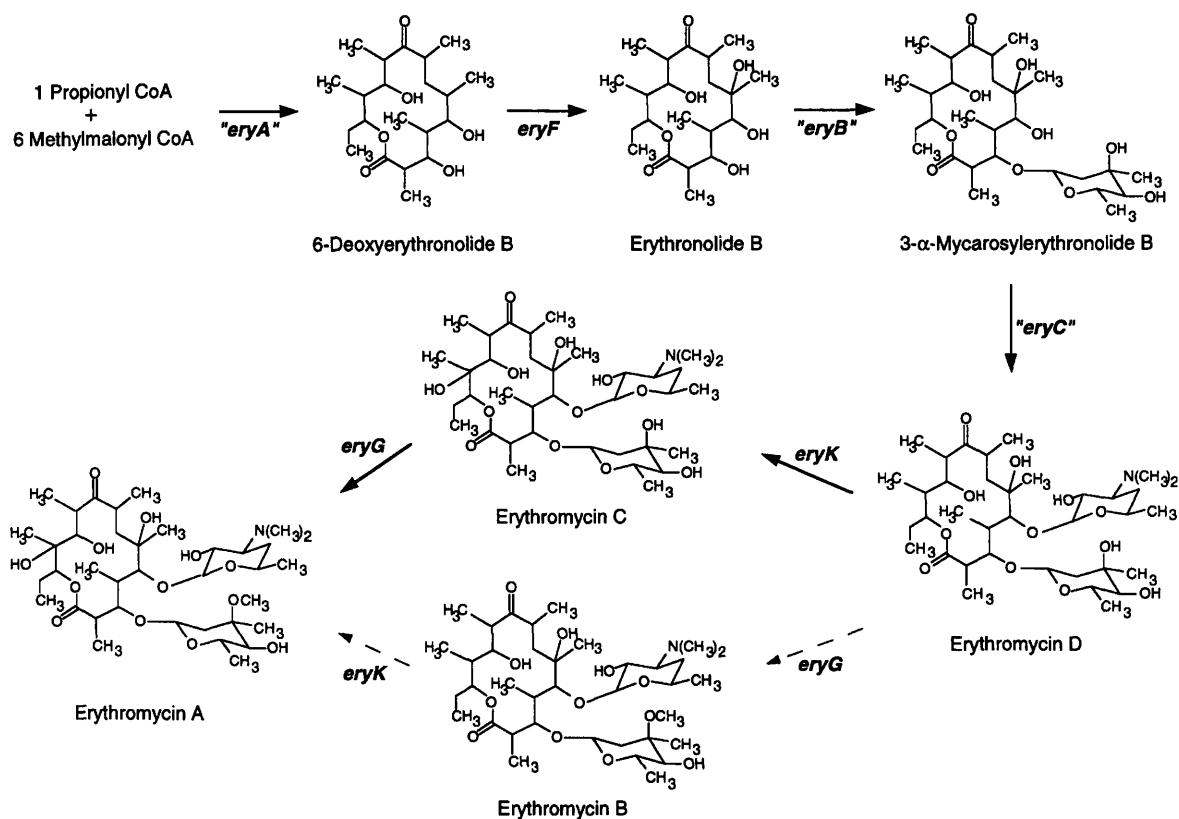
Fig. 1. Biosynthetic scheme for erythromycin A. The principal steps in erythromycin A biosynthesis, as determined by genetic analysis, are shown. *eryA, eryB* and *eryC* mutants are defective in the multistep pathways leading to the biosynthesis of 6-deoxyerythronolide B, L-mycarose and D-desosamine, respectively. *eryF, eryK* and *eryG* mutations affect the single enzymic steps for C-6 and C-12 hydroxylation and the conversion of L-mycarose to L-cladinose, respectively.

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 *Sac. 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 *Sac. erythraea* 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, thio-strepton 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 5'-[α-³²P] or 5'-[α-³³P]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, eryBV, eryBVI, eryCII, eryCIII, eryCIV, 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 1·5–3 kb contiguous DNA on each flank were altered in vitro using standard recombinant DNA techniques (Sambrook et al., 1989). The construction of each altered gene is summarized below. To disable eryBIV, 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 eryBV gene. This mutation could have a polar effect on expression of eryBV. The eryBV gene was inactivated by deletion of the entire reading frame. Digestion with SpeI and NheI followed by religation of the compatible cohesive ends removes the complete eryBV locus without altering any of the surrounding sequence. To inactivate eryBVI, the DNA segment carrying the gene was digested with BamHI followed by religation. This creates an in-frame 630 nt deletion that removes codons 95–305 of eryBVI. 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 eryCIII, the gene was digested with SphI followed by treatment with S1 nuclease, generating a 4 nt deletion that was confirmed by DNA sequencing. This deletion causes a frameshift after codon 94 that should result in premature translation ter-mination within the eryCIII gene. The eryCIV and eryCV genes were disabled simultaneously by RsrII digestion fol-lowed 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 inte-grated 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 analyis (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 erythro-mycin 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 erythronolide B, 3-α-L-mycarosyl 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-α-L-mycarosyl 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)⁻¹ 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
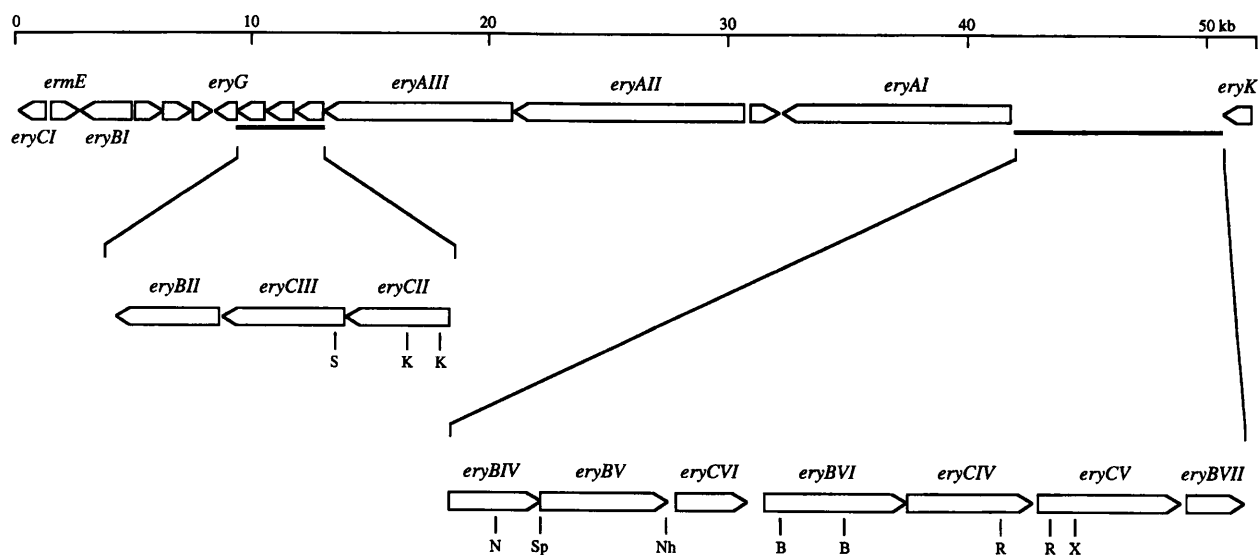
Fig. 2. Organization of the erythromycin biosynthetic gene cluster. The regions described in this work have been underlined and are shown in greater detail. The approximate locations of the restriction sites used to generate the mutations described herein are indicated below each gene. Abbreviations: B, *Bam*HI; K, *Kpn*I; N, *Nco*I; Nh, *Nhe*I; R, *Rsr*II; S, *Sph*I; Sp, *Spe*I; X, *Xho*I. The small ORF between *eryAI* and *eryAII* is insertion sequence IS*1136* (Donadio & Staver, 1993). Although not shown, a similar IS-element lies just beyond *eryK* (Pereda *et al.*, 1997).

al., 1991). These ORFs would be transcribed in the same direction as the upstream *eryAIII* gene and the downstream *eryG* gene, and all three ORFs, the *eryAIII* 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 *eryAIII* 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 *eryAIII*. 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 *eryAIII* 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 *eryAIII*. 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).

Table 1. Mutations, phenotype and deduced function of *eryB* and *eryC*

Locus	Previous designation	Mutation	Phenotype	Predicted molecular mass (Da)	Predicted function
eryBII	ORF7*	†	EryB	35537	3-Ketoreductase
eryBIV	ORF13‡	4 nt insertion	EryB§	33975	4-Ketoreductase
eryBV	ORF14‡	1248 nt deletion	EryB	45531	Mycarosyl transferase
eryBVI	ORF16‡	630 nt deletion	EryB	57657	2,3-Dehydratase (?)
eryBVII	ORF19‡	None	ND	21291	5-Epimerase
eryCII	ORF9*	301 nt deletion	EryC	38420	Unknown
eryCIII	ORF8*	4 nt deletion	EryC	45922	Desosaminyl transferase
eryCIV	ORF17‡	524 nt deletion‖	EryC	43347	3,4-Dehydratase
eryCV	ORF18‡	4 nt insertion	EryC	53898	Unknown
eryCVI	ORF15‡	654 nt deletion	EryC	25956	3-Amino dimethyltransferase

ND, Not determined.

* Haydock *et al.* (1991).

† Weber *et al.* (1990).

‡ Donadio *et al.* (1993) and Katz & Donadio (1993).

§ Strain also produces a small amount of two bioactive species.

‖ Deletion extends into the *eryCV* locus (see text).

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, however, although a TDP-D-glucose 4,6-dehydratase has been purified from *Sac. erythraea* (Vara & Hutchinson, 1988) and its gene has been cloned (Linton *et al.*, 1995). 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-L-mycarose. These include 2,3-dehydration and reduction to form the 2-deoxysugar, 5-epimerization to form the L-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 ascarylose, 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 *eryAIII* 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*
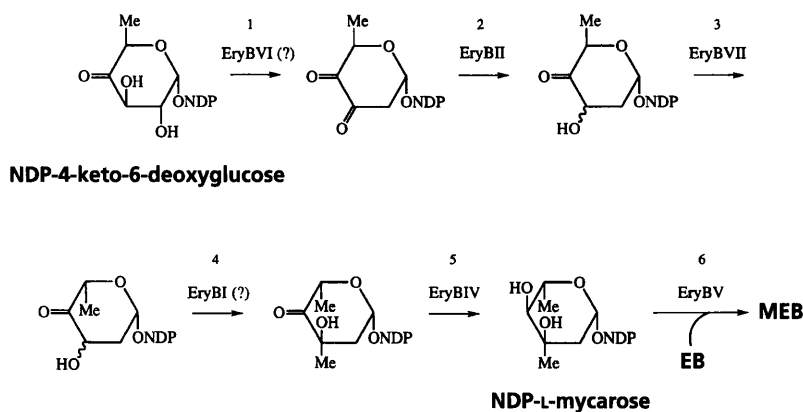
Fig. 3. Scheme for the biosynthesis of L-mycarose. The putative gene products believed to be associated with the various steps are indicated at the respective point in the pathway.

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 aryl-alcohol dehydrogenase of the fungus *Phanerochaete chrysosporium* (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-α-L-mycarosylerythronolide B to erythromycin A, typical of the EryB phenotype. Yet, the mutant also produces small amounts of two bioactive compounds that migrate slightly more rapidly than erythromycins A and C, respectively, during TLC. Based on the likely function of *eryBIV* (see below), these compounds may be the 4-keto-L-mycarosyl derivatives of erythromycins A and C.

The deduced product of *eryBIV* displays similarity to a large family of sugar oxidoreductases that includes the mechanistically well-characterized UDPgalactose 4-epimerase encoded by the *galE* gene of *E. coli* (Bauer *et al.*, 1992; Lemaire & Müller-Hill, 1986). Also in this family are *ascF* from the ascarylose pathway of *Yersinia pseudotuberculosis* (Thorson *et al.*, 1994), *rfbJ* from the

D-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 RfbJ and 20% identical to GalE), all members have a rigorously conserved GxxGxxG motif within the first twenty residues of their N-termini 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.

eryBV. The *eryBV* 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 *eryBV* gene extends for 1247 nt and encodes a protein of 45·5 kDa. Strains carrying a chromosomal deletion of the entire *eryBV* gene have an EryB phenotype.

The protein encoded by *eryBV* 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

```
         1                                                                                                    100
EryBIV   vngisdsprq litLlGaSGF VGsavlreLr dh.pvrlrav sRggapavpp gaaevedLr. .......... .adllepgra aaaiedadvi VhlVaHaAgg
AscF     .......mk. .llItGvSGY IGshlvnyLa nlggyeiygi sRneildqdi nq....lLln ikifqldrds lpdilkrvrp .........d V..ViHlAsc
RfbJ     ....mtflke yviVsGaSGF IGkhlleaLk k.sgisvvai tRdvi..knn sn....aLan vrwcswdnie llveelsids .........a LigIiHlAte
StrL     ...mspyprp rwlVtGaSGm lGreltplLd rr.gaavtal gRghl.ditd gaa....... .......... .......vrs avaehrpavv VncaawtAvd
GalE     ......... vlVtGgSGY IGsh...... ..tcvqllqn ghdviildnl cnskrsvLpv ierlggkhpt fvegdirnea lmteilhdha IdtViHfAgl
Consensus ---------- ---V-G-SGF IG------L-- ---------- -R-------- -------L-- ---------- ---------- ---------- V--V-H-A--

         101                                                                                                   200
EryBIV   stwrsatsdp e.aervNvgl mhdLvgAlhd rrrstppvlL yasTaqaanp saasr..... ........Ya qqKteaEril rkatdegrvr gvilrlpav.
AscF     flsqhsykni keiiksNvef pteLleAmnd ...vgvkkiI ..nTgtsWqc fnsdtY.... ...npvnlYa asKqafEdil kfyinae.gf sainlklfdt
RfbJ     y..ghktssl iniedaNvik plkLldlaik ...yradifL ..nTd.sFfa kkdfnY.... ...qhmrpYi itKrhfDeig hyyanmh.di sfvnmrlehv
StrL     ....eaesep alamavNgeg prhLaqA... .cravgavlL qlsTdyvFpg sggrpYredh ptg.prtvYg ctKragErav levlpdtgyi vrtawl....
GalE     kavgesvqkp leyydnNdng tlrLisAmra ...anvknfI fsssatvYgd qpkipYvesf ptgtpqspYg ksKlmvEqil tdlqkaqpdw siallryfnp
Consensus ---------- ------N--- ---L--A--- ----------L ---T---F-- -----Y---- --------Y- --K---E--- ---------- ----------

         201                                                                                                   300
EryBIV   YGqsgp.... ..sgpmgrgv vaamirrala gepltmwhdg gvrrdl.... ...lhVeDva tAFaaAlEhh dalag...gt WalGadrsep lgDifravsg
AscF     YG........ ..gvdkrrkl is.ll..... ..ddiaknnk qldmspgeql ldlvhInDvc rAFkiAiDkl celpseyvvs Ygvsnkyrvt lkElvsiYe.
RfbJ     YG........ ..pgdgenkf ipyii..... ..dclnkkqs cvkcttgeqi rdfifVdDvv nAYltilEnr kevps..yte YqvGtgagvs lkDflvyLqn
StrL     YGaggp.... ........nf vakmirlead edtvlvvddq hgqptwt... ...adLaDrl aALgaAalag tapagiyhat ntgGttwnal apEtfrlLga
GalE     vGahpsgdmg edpqgipnnl mpyiaqvavg rrdslaifgn dyptedgtgv rdyihVmDla dghvvAmEkl ankpgvhm.. YnlGagvgns vlDvvnaF..
Consensus YG-------- ---------- ---------- ---------- ---------- -----V-D-- -AF--A-E-- ---------- Y--G------ --D----L--

         301                                                          383
EryBIV   sva..rqtgs pavdvvtvpa pehaEandfr sddidstefr srtgWrprvs LtDgidrtva altpteeh.. ....
AscF     .....rvnnv klninfgtre yrnrEvmvpc tniqnl.... ..pdWevvip Lsqglky... .......... ....
RfbJ     t....mmpgs ssifefgaie qrdnEimfsv annknlkamg wkpnFdykkg IeEllkrl.. .......... ....
StrL     dparvrptts lalarpavr. prysvldqsr wkaagleplr hw.....raa LtEsf....p alcgragrpv pgpr
GalE     s....kacgk pvnyhfapr. ...rEgafrp tgrtpakptv .nlnWrvtrt LdEmaqdtwh w......... ....
Consensus ---------- ---------- ----E----- ---------- ----W----- L-E------- ---------- ----
```

Fig. 4. Alignment of the deduced EryBIV product with other known and putative sugar 4-ketoreductases. AscF, L-ascarylose biosynthesis (Thorson *et al.*, 1994); RfbJ, D-abequose biosynthesis (Wyk & Reeves, 1989); StrL, L-dihydrostreptose biosynthesis (Pissowotzki *et al.*, 1991); GalE, UDPgalactose 4-epimerase (Lemaire & Müller-Hill, 1986). The GxxGxxG and YxxxKxxxE motifs are shown in bold.

to encode the enzyme responsible for transfer of L-daunosamine to the aromatic polyketide ε-rhodomycinone 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 high degree of relatedness (43 % identity) between the product of *eryBV* and 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 catalyse these transfers. Finally, although the degree of identity is lower, EryBV also resembles other glycosyltransferases such as the rhamnosyltransferase encoded by *rhlB* (21 % identity) of *Pseudomonas aeruginosa* (Ochsner *et al.*, 1994).

eryBVI. 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 55·1 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 fradiae* (58 % identity over the first 98 residues; Merson-Davies & Cundliffe, 1994), ORF3 from the daunorubicin biosynthetic gene cluster

Fig. 5. Comparison of the structures of L-mycarose, L-daunosamine, D-desosamine and D-mycaminose.

of *Streptomyces* sp. strain C5 (41 % identity; Dickens *et al.*, 1996), and *dnmT* from the daunorubicin pathway of *Str. 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 aldohexose 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*, *dnmT* 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, *eryBI*, has been described, although its function is not known (Haydock *et al.*, 1991). In light of the present data, it is possible that *eryBI* might encode the C-methyltransferase. Finally, there is some evidence for an another *eryB* locus, ORF3, immediately upstream of

eryBI (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-desosamine should comprise six steps (Fig. 6), and based on the genetic data described below, this would include a pyridoxamine-catalysed dehydration and reduction that is similar to a reaction first described for the biosynthesis of ascarylose (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 *eryAIII*. Three more desosamine biosynthetic genes, *eryCVI*, *eryCIV* 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-α-L-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 *eryAIII* polyketide synthase gene. While several alternative ATG start codons fall further downstream from *eryAIII*, 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 38·4 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.
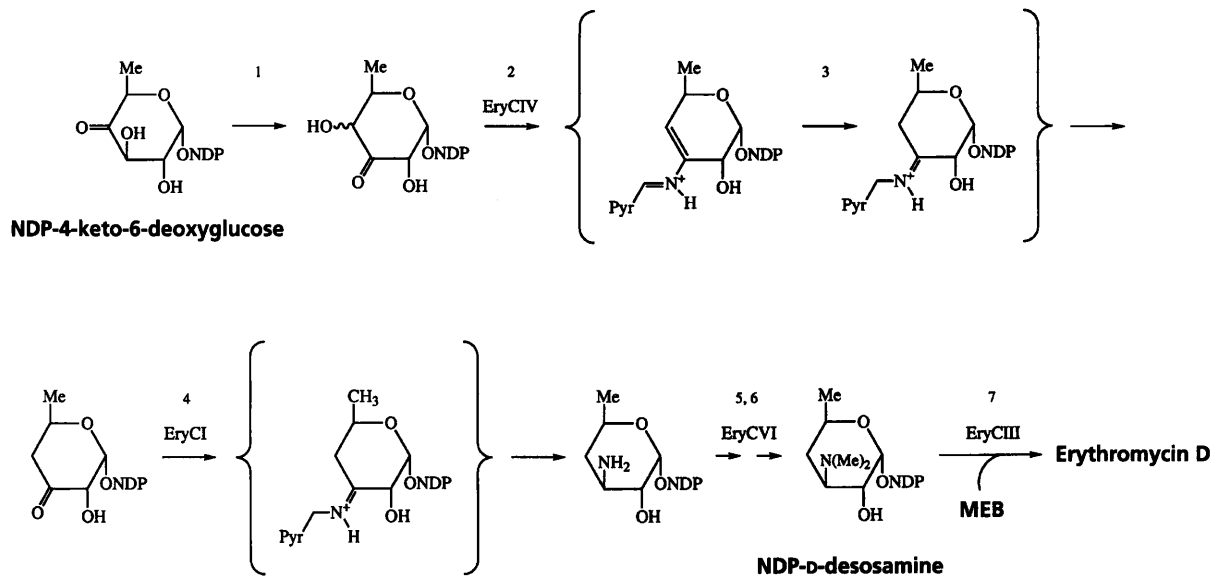
Fig. 6. Scheme for the biosynthesis of D-desosamine. The putative involvement of the pyridoxamine cofactor in the 3,4-dehydration and reduction and the 3-amination is shown in brackets. The putative gene products believed to be associated with the various steps are indicated at the respective point in the pathway.

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-D-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 spiromycin 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 deoxyamino sugar biosynthesis including *tylB* 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 ascarylose 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 D-desosamine biosynthesis requires both deoxygenation and amino 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 TylB protein (61% identity) contrasted with the relationship between EryCIV and TylB (31% identity). TylB participates in the biosynthesis of D-mycaminose, 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 53·9 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 *eryCII*, 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.

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* 215, 403–410.

Bauer, A. J., Rayment, I., Frey, P. A. & Holden, H. M. (1992). The molecular structure of UDP-galactose 4-epimerase from *Escherichia coli* determined at 2·5 Å resolution. *Proteins* 12, 372–381.

Bevitt, D. J., Cortes, J., Haydock, S. F. & Leadlay, P. F. (1992). 6-Deoxyerythronolide B synthase 2 from *Saccharopolyspora erythraea*. Cloning of the structural gene, sequence analysis and inferred domain structure of the multifunctional enzyme. *Eur J Biochem* 204, 39–49.

Bibb, M. J. & Cohen, S. N. (1982). Gene expression in *Streptomyces*: construction and application of promoter-probe plasmid vectors in *Streptomyces lividans*. *Mol Gen Genet* 187, 265–277.

Cortes, J., Haydock, S. F., Roberts, G. A., Bevitt, D. J. & Leadlay, P. F. (1990). An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of *Saccharopolyspora erythraea*. *Nature* 348, 176–178.

Denoya, C. & Dubnau, D. (1989). Mono- and dimethylating activities and kinetic studies of the *ermC* 23 S rRNA methyltransferase. *J Biol Chem* 264, 2615–2624.

Devereux, J., Haeberli, P. & Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res* 12, 387–395.

Dhillon, N., Hale, R. S., Cortes, J. & Leadlay, P. F. (1989). Molecular characterization of a gene from *Saccharopolyspora erythraea* (*Streptomyces erythraeus*) which is involved in erythromycin biosynthesis. *Mol Microbiol* 3, 1405–1414.

Dickens, M. L., Ye, J. & Strohl, W. R. (1996). Cloning, sequencing, and analysis of aklaviketone reductase from *Streptomyces* sp. strain C5. *J Bacteriol* 178, 3384–3388.

Donadio, S. & Katz, L. (1992). Organization of the enzymatic domains in the multifunctional polyketide synthase involved in erythromycin formation in *Saccharopolyspora erythraea*. *Gene* 111, 551–560.

Donadio, S. & Staver, M. J. (1993). IS*1136*, an insertion element in the erythromycin gene cluster of *Saccharopolyspora erythraea*. *Gene* **126**, 147–151.

Donadio, S., Staver, M. J., McAlpine, J. B., Swanson, S. J. & Katz, L. (1991). Modular organization of genes required for complex polyketide biosynthesis. *Science* **252**, 675–679.

Donadio, S., Stassi, D., McAlpine, J. B. & 9 other authors (1993). Recent developments in the genetics of erythromycin formation. In *Industrial Microorganisms: Basic and Applied Molecular Genetics*, pp. 257–265. Edited by R. H. Baltz, G. D. Hegeman & P. L. Skatrud. Washington, DC: American Society for Microbiology.

Ellis, E. M., Judah, D. J., Neal, G. E. & Hayes, J. D. (1993). An ethoxyquin-inducible aldehyde reductase from rat liver that metabolizes aflatoxin B_1 defines a subfamily of aldo-keto reductases. *Proc Natl Acad Sci USA* **90**, 10350–10354.

Geistlich, M., Losick, R., Turner, J. R. & Rao, R. N. (1992). Characterization of a novel regulatory gene governing the expression of a polyketide synthase gene in *Streptomyces ambofaciens*. *Mol Microbiol* **6**, 2019–2029.

Haydock, S. F., Dowson, J. A., Dhillon, N., Roberts, G. A., Cortes, J. & Leadlay, P. F. (1991). Cloning and sequence analysis of genes involved in erythromycin biosynthesis in *Saccharopolyspora erythraea*: sequence similarities between EryG and a family of S-adenosylmethionine-dependent methyltransferases. *Mol Gen Genet* **230**, 120–128.

Hopwood, D. A., Bibb, M. J., Chater, K. F., Kieser, T., Bruton, C. J., Kieser, H. M., Lydiate, D. J., Smith, C. P., Ward, J. M. & Schrempf, H. (1985). *Genetic Manipulation of Streptomyces: a Laboratory Manual*. Norwich: John Innes Foundation.

Jiang, X.-M., Neal, B., Santiago, F., Lee, S. J., Romana, L. K. & Reeves, P. R. (1991). Structure and sequence of the *rfb* (O antigen) gene cluster of *Salmonella* serovar typhimurium (strain LT2). *Mol Microbiol* **5**, 695–713.

Katz, L. & Donadio, S. (1993). Polyketide synthesis: prospects for hybrid antibiotics. *Annu Rev Microbiol* **47**, 875–912.

Kim, K.-S. & Farrand, S. K. (1996). Ti plasmid-encoded genes responsible for catabolism of the crown gall opine mannopine by *Agrobacterium tumefaciens* are homologs of the T-region genes responsible for synthesis of this opine by the plant tumour. *J Bacteriol* **178**, 3275–3284.

Krügel, H., Schumann, G., Hänel, F. & Fiedler, G. (1993). Nucleotide sequence analysis of five putative *Streptomyces griseus* genes, one of which complements an early function in daunorubicin biosynthesis that is linked to a putative gene cluster involved in TDP-daunosamine formation. *Mol Gen Genet* **241**, 193–202.

Lemaire, H.-G. & Müller-Hill, B. (1986). Nucleotide sequences of the *galE* and *galT* gene of *E. coli*. *Nucleic Acids Res* **14**, 7705–7711.

Linton, K. J., Jarvis, B. W. & Hutchinson, C. R. (1995). Cloning the genes encoding thymidine diphosphoglucose 4,6-dehydratase and thymidine diphospho-4-keto-6-deoxyglucose 3,5-epimerase from the erythromycin-producing *Saccharopolyspora erythraea*. *Gene* **153**, 33–40.

Liu, H. & Thorson, J. S. (1994). Pathways and mechanisms in the biogenesis of novel deoxysugars by bacteria. *Annu Rev Microbiol* **48**, 223–256.

Merson-Davies, L. A. & Cundliffe, E. (1994). Analysis of five tylosin biosynthetic genes from the *tylIBA* region of the *Streptomyces fradiae* genome. *Mol Microbiol* **13**, 349–355.

Niemi, J. & Mäntsälä, P. (1995). Nucleotide sequences and expression of genes from *Streptomyces purpurascens* that cause the production of new anthracyclines in *Streptomyces galilaeus*. *J Bacteriol* **177**, 2942–2945.

Ochsner, U. A., Fiechter, A. & Reiser, J. (1994). Isolation, characterization, and expression in *Escherichia coli* of the *Pseudomonas aeruginosa rhlAB* genes encoding a rhamnosyltransferase involved in rhamnolipid biosurfactant synthesis. *J Biol Chem* **269**, 19787–19795.

Ogawa, H., Konishi, K., Takata, Y., Nakashima, H. & Fujioka, M. (1987). Rat glycine methyltransferase: complete amino acid sequence deduced from a cDNA clone and characterization of the genomic DNA. *Eur J Biochem* **168**, 141–151.

Otten, S. L., Liu, X., Ferguson, J. & Hutchinson, C. R. (1995). Cloning and characterization of the *Streptomyces peucetius dnrQS* genes encoding a daunosamine biosynthesis enzyme and a glycosyl transferase involved in daunorubicin biosynthesis. *J Bacteriol* **177**, 6688–6692.

Paulus, T. J., Tuan, J. S., Luebke, V. I., Maine, G. T., DeWitt, J. P. & Katz, L. (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.

Pereda, A., Summers, R. G. & Katz, L. (1997). Nucleotide sequence of the *ermE* distal flank of the erythromycin biosynthesis cluster in *Saccharopolyspora erythraea*. *Gene* **193**, 65–71.

Pissowotzki, K., Mansouri, K. & Piepersberg, W. (1991). Genetics of streptomycin production in *Streptomyces griseus*: molecular structure and putative function of genes *strELMB2N*. *Mol Gen Genet* **231**, 113–123.

Queener, S. W., Sebek, O. K. & Vézina, C. (1978). Mutants blocked in antibiotic synthesis. *Annu Rev Microbiol* **32**, 593–636.

Reiser, J., Muheim, A., Hardegger, M., Frank, G. & Fiechter, A. (1994). Aryl-alcohol dehydrogenase from the white-rot fungus *Phanerochaete chrysosporium*. *J Biol Chem* **269**, 28152–28159.

Rettig, J., Heinemann, S. H., Wunder, F., Lorra, C., Parcej, D. N., Dolly, J. O. & Pongs, O. (1994). Inactivation properties of voltage-gated K^+ channels altered by presence of β-subunit. *Nature* **369**, 289–294.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Schulman, M. D., Acton, S. L., Valentino, D. L. & Arison, B. H. (1990). Purification and identification of dTDP-oleandrose, the precursor of the oleandrose units of the avermectins. *J Biol Chem* **265**, 16965–16970.

Scotti, C. & Hutchinson, C. R. (1996). Enhanced antibiotic production by manipulation of the *Streptomyces peucetius dnrH* and *dnmT* genes involved in doxorubicin (adriamycin) biosynthesis. *J Bacteriol* **178**, 7316–7321.

Stanzak, R., Matsushima, P., Baltz, R. H. & Rao, R. N. (1986). Cloning and expression in *Streptomyces lividans* of clustered erythromycin biosynthesis genes from *Streptomyces erythraeus*. *Bio/Technology* **4**, 229–232.

Stassi, D., Donadio, S., Staver, M. J. & Katz, L. (1993). Identification of a *Saccharopolyspora erythraea* gene required for the final hydroxylation step in erythromycin biosynthesis. *J Bacteriol* **175**, 182–189.

Stutzman-Engwall, K. J., Otten, S. L. & Hutchinson, C. R. (1992). Regulation of secondary metabolism in *Streptomyces* spp. and overproduction of daunorubicin in *Streptomyces peucetius*. *J Bacteriol* **174**, 144–154.

Thorson, J. S., Lo, S. F., Liu, H. & Hutchinson, C. R. (1993). Biosynthesis of 3,6-dideoxyhexoses: new mechanistic reflections upon 2,6-dideoxy, 4,6-dideoxy, and amino sugar construction. *J Am Chem Soc* 115, 6993–6994.

Thorson, J. S., Lo, S. F., Ploux, O., He, X. & Liu, H. (1994). Studies of the biosynthesis of 3,6-dideoxyhexoses: molecular cloning and characterization of the *asc* (ascarylose) region from *Yersinia pseudotuberculosis* serogroup VA. *J Bacteriol* 176, 5483–5493.

Tuan, J. S., Weber, J. M., Staver, M. J., Leung, J. O., Donadio, S. & Katz, L. (1990). Cloning of genes involved in erythromycin biosynthesis from *Saccharopolyspora erythraea* using a novel actinomycete–*Escherichia coli* cosmid. *Gene* 90, 21–29.

Uchiyama, H. & Weisblum, B. (1985). N-Methyl transferase of *Streptomyces erythraeus* that confers resistance to the macrolide-lincosamide-streptogramin B antibiotics: amino acid sequence and its homology to cognate R-factor enzymes from pathogenic bacilli and cocci. *Gene* 38, 103–110.

Vara, J. A. & Hutchinson, C. R. (1988). Purification of thymidine-diphospho-D-glucose 4,6-dehydratase from an erythromycin-producing strain of *Saccharopolyspora erythraea* by high resolution liquid chromatography. *J Biol Chem* 263, 14992–14995.

Vara, J., Lewandowska-Skarbek, M., Wang, Y.-G., Donadio, S. & Hutchinson, C. R. (1989). Cloning of genes governing the deoxysugar portion of the erythromycin biosynthesis pathway in *Saccharopolyspora erythraea* (*Streptomyces erythraeus*). *J Bacteriol* 171, 5872–5881.

Weber, J. M. & Losick, R. (1988). The use of a chromosome integration vector to map erythromycin resistance and production genes in *Saccharopolyspora erythraea* (*Streptomyces erythraeus*). *Gene* 68, 173–180.

Weber, J. M., Wierman, C. K. & Hutchinson, C. R. (1985). Genetic analysis of erythromycin production in *Streptomyces erythraeus*. *J Bacteriol* 164, 425–433.

Weber, J. M., Schoner, B. & Losick, R. (1989). Identification of a gene required for the terminal step in erythromycin A biosynthesis in *Saccharopolyspora erythraea* (*Streptomyces erythraeus*). *Gene* 75, 235–241.

Weber, J. M., Leung, J. O., Maine, G. T., Potenz, R. H. B., Paulus, T. J. & Dewitt, J. P. (1990). Organization of a cluster of erythromycin genes in *Saccharopolyspora erythraea*. *J Bacteriol* 172, 2372–2383.

Weber, J. M., Leung, J. O., Swanson, S. J., Idler, K. B. & McAlpine, J. B. (1991). An erythromycin derivative produced by targeted gene disruption in *Saccharopolyspora erythraea*. *Science* 252, 114–117.

Weigel, T. M., Liu, L. & Liu, H. (1992). Mechanistic studies of the biosynthesis of 3,6-dideoxyhexoses in *Yersinia pseudotuberculosis*: purification and characterization of CDP-4-keto-6-deoxy-D-glucose-3-dehydrase. *Biochemistry* 31, 2129–2139.

Wierenga, R. K. & Hol, W. G. J. (1983). Predicted nucleotide-binding properties of p21 protein and its cancer-associated variant. *Nature* 302, 842–844.

Wyk, P. & Reeves, P. (1989). Identification and sequence of the gene for abequose synthase, which confers antigenic specificity on group B salmonellae: homology with galactose epimerase. *J Bacteriol* 171, 5687–5693.

Xiang, S.-H., Hobbs, M. & Reeves, P. R. (1994). Molecular analysis of the *rfb* gene cluster of a group D2 *Salmonella enterica* strain: evidence for its origin from an insertion sequence-mediated recombination event between group E and D1 strains. *J Bacteriol* 176, 4357–4365.

Yamamoto, H., Maurer, K. H. & Hutchinson, C. R. (1986). Transformation of *Streptomyces erythraeus*. *J Antibiotics* 39, 1304–1313.

Yamamoto-Otake, H., Nakano, E. & Koyama, Y. (1994). Cloning and sequencing of the L-fucose dehydrogenase gene from *Pseudomonas* sp. no.1143. *Biosci Biotech Biochem* 58, 2281–2282.

Yanisch-Perron, C., Vieira, J. & Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* 33, 103–119.