Modular Organization of Genes Required for Complex Polyketide Biosynthesis

STEFANO DONADIO, MICHAEL J. STAVER, JAMES B. MCALPINE, SUSAN J. SWANSON, LEONARD KATZ

In *Saccharopolyspora erythraea*, the genes that govern synthesis of the polyketide portion of the macrolide antibiotic erythromycin are organized in six repeated units that encode fatty acid synthase (FAS)–like activities. Each repeated unit is designated a module, and two modules are contained in a single open reading frame. A model for the synthesis of this complex polyketide is proposed, where each module encodes a functional synthase unit and each synthase unit participates specifically in one of the six FAS-like elongation steps required for formation of the polyketide. In addition, genetic organization and biochemical order of events appear to be colinear. Evidence for the model is provided by construction of a selected mutant and by isolation of a polyketide of predicted structure.

POLYKETIDES ARE A LARGE AND DIVERSE CLASS OF NATURAL products that includes antibiotics, pigments, and immunosuppressants and have applications in medicine, agriculture, and industry (for an example, see Fig. 1). Biosynthesis of polyketides is believed to occur by a series of condensations of carbon units in a manner similar to that of long chain fatty acids (LCFA) (*1*). The LCFAs are formed by fatty acid synthase (FAS) through a process whereby a starter unit (commonly acetate) is condensed to the extender unit (malonate). The resulting β-keto group is then fully processed (reduced), and the cycle resumes with the condensation of a new extender unit (*2*). Most polyketides, however, contain structural complexities that can be accounted for by the use of different extender units at various steps and by variations in the extent of processing of the β-carbon (β-ketoreduction, dehydration, enoylreduction). Although this complexity exists, in living organisms there is believed to be a polyketide synthase (PKS) that produces only one or a few related molecular structures (*3*). Thus, an understanding of the biosynthesis of complex polyketides must include a description of the mechanism by which the PKS both selects the correct substrate and decides the fate of the β carbon at each step.

The polyketide portion of macrolide antibiotics is synthesized through the condensation of short chain carbon units; for example, seven propionates in the case of erythromycin (*4*). β-Hydroxy-acyl thioesters, which mimic hypothetical intermediates in the synthesis but not the corresponding β-keto derivatives, are incorporated in vivo into the macrolide rings of erythromycin and tylosin, providing support for a FAS-like origin of these molecules (*5*). Consistent with this idea is the detection of branched-chain fatty acids in fermentation broths of tylosin and mycinamicin producers (*6*). A FAS-like mechanism for the synthesis of the erythromycin aglycone 6-deoxyerythronolide B (6dEB) requires that (i) six methylmalonyl–coenzyme A (mmCoA) units, three of each enantiomer, are successively condensed to a propionyl-CoA starter unit; (ii) that β-ketoreduction occurs after each condensation step except step three so that a keto group is left at C-9; and (iii) that dehydration and enoyl reduction take place only after the fourth condensation to introduce a methylene at C-7 (Fig. 1). Consequently, a full set of FAS activities (*2*) is required for 6dEB synthesis: acyltransferase (AT), β-ketoacyl carrier protein synthase (KS), and acyl carrier protein (ACP) for chain elongation; β-ketoreductase (KR), dehydratase (DH), and enoyl reductase (ER) for processing of the β carbon; and thioesterase (TE) for release and lactonization of the full-length chain. In addition, the hypothetical 6dEB PKS must also be programmed at each step to select the correct enantiomeric extender unit and to process the β carbon to the appropriate degree.

Six modules with FAS-like domains in 6dEB synthesis. The segment of the chromosome required for the formation of 6dEB in *S. erythraea* has been designated *eryA*. A 5-kb DNA fragment of *eryA* had been identified by its ability to restore erythromycin production when introduced into a mutant blocked in the synthesis of 6dEB. Hybridization of chromosomal DNA with this segment

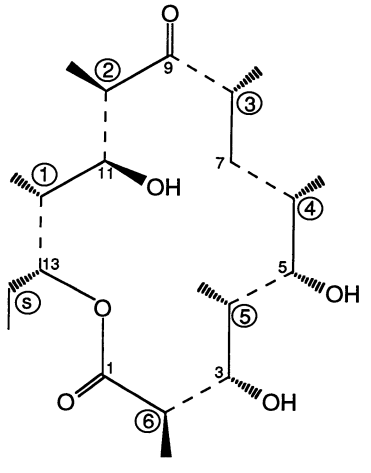

Fig. 1. Structure of 6-deoxyerythronolide B. The dotted lines represent C-C bonds formed during synthesis steps 1 through 6 by condensation between the starter unit propionyl-CoA (encircled S) and the extender units methylmalonyl-CoA encircled 1 to 6. The stereochemistry of extender units 1, 3, and 4 should be compared to that of units 2, 5, and 6. Lactonization of the acyl chain between C-1 and C-13 results in the formation of 6-deoxyerythronolide B. Hydroxylation of the lactone ring at C-6, followed in order by mycarose and desosamine attachment at hydroxyls at C-3 and C-5, respectively, results in the formation of the first bioactive compound, erythromycin D (*4*).

S. Donadio, M. J. Staver, and L. Katz are in the Department of Corporate Molecular Biology, Abbott Laboratories, Abbott Park, IL 60064. J. B. McAlpine and S. J. Swanson are in the Bioactive Microbial Metabolite Project, Abbott Laboratories, Abbott Park, IL 60064.

has indicated the existence of several closely linked homologous loci (*7*). Gene disruption experiments (*7, 8*) showed that *eryA* encompasses about 10 kb of DNA and identified the end of *eryA* proximal to *ermE*, the erythromycin resistance gene (*9*). Similar experiments were performed in our study to establish the other boundary for *eryA*. Contiguous or overlapping DNA segments that spanned a region of 25 kb were subcloned into the poorly replicating *S. erythraea* vector pWHM3 (*10*), and the resulting plasmids were integrated into the wild-type chromosome by single reciprocal recombination (*11*). Except for the *ermE*-distal fragment (Fig. 2A), all integrants exhibited an EryA phenotype (*12*). Because a mutant phenotype is expected when the cloned fragment is internal to a transcription unit, these results indicate that *eryA* extends for 32 to 35 kb and is likely to consist of one or possibly two transcripts. No integration events were obtained in a 1.6-kb segment of *eryA*, but this segment is fully internal to *orf1*; see Fig. 2C. Nucleotide sequencing (*13*) of the 35-kb segment established that *eryA* consists of six repeated units. Comparison of the 35-kb segment of sequenced DNA with a 4.5-kb subsegment of *eryA* indicates that the sizes of each of the six repeated units is 4.3 to 6.5 kb and that the sequences have a similarity of 64 percent or higher (Fig. 2B).

Three large open reading frames (*orf's*) are present in *eryA*, each consisting of two repeated units (Fig. 2C). The ORF organization of *eryA* supports the possible existence of large transcripts, which is indicated by gene disruption experiments. The putative start for *orf1* is within the 3-kb segment (full bar in Fig. 2A) where one of the *eryA* boundaries was mapped by gene disruption. A segment of 1.44 kb apparently not involved in 6dEB synthesis separates *orf1* and *orf2*, whereas *orf2* and *orf3* are contiguous; *orf3* corresponds to ORFA described independently (*14*). The deduced amino acid sequences of the three Orf's were compared to FAS and PKS systems, and putative FAS-like domains in *eryA* are depicted at their approximate map positions (Fig. 2C). A total of seven ACP's, six KS's, eight AT's, six KR's, one DH, and one ER have been identified and are organized into the six repeated units (Fig. 2B). The segment of *eryA* encoding a repeated unit is designated a "module" (Fig. 2C). Starting from the 5′ end of *eryA*, the first module encodes putative activities in the order AT, ACP, KS, AT, KR, and ACP. The second, third, and fifth modules consist of KS, AT, KR, and ACP. The fourth contains KS, AT, DH, ER, KR, and ACP, and the sixth, KS, AT, KR, ACP, and AT. In summary, *eryA* consists of six modules that encode FAS-like activities, and each Orf contains two modules.

Each of the putative domains in the six modules contains the active site motifs reported to be conserved in FAS and PKS systems (*15*). Specifically, the acyltransferase GHSxG (*15*; x = any amino acid) motif embedded in a hydrophobic region (where S is the serine involved in the formation of the acyl-enzyme intermediate) is present in each of the eight proposed *eryA* AT's. Similarly, the highly conserved sequence GPxxxxxTACSS around the cysteine residue that participates in thioester formation can be detected in the six putative KS domains of *eryA*. The pantotheine-binding serine, present in most ACP's in the LGxDSLxxVE motif, is found in the seven *eryA* ACP's identified, and the "fingerprint" region GxxGxx-AxxA of NADPH (reduced form of nicotinamide adenine dinucleotide phosphate)-dependent reductases (*16*) is maintained in similar surroundings in the *eryA* KR domains. The functional domains for KR, DH and ER were identified as described below.

Our experiments extend to the three *eryA* Orf's recent findings (*14*), which showed that the predicted polypeptide designated Orf3 here resembles the type I eukaryotic FAS systems, wherein the functions required for fatty acid synthesis reside within a single polypeptide. This contrasts with the type II–like organization (monofunctional polypeptides) of PKS's involved in the synthesis of aromatic polyketides in *Streptomyces* (*17*). A comparison of the segment of Orf2 that corresponds to module 4 with FAS from rat (*18*) is illustrated in Fig. 3. It can be seen that, with the exception of a gap of about 400 amino acids in Orf2 and of 600 in FAS, the two sequences align end to end. This indicates an identical order of functional domains in both type I FAS and Orf2 and identifies the putative domains for DH and ER. Functional domains for the KS, AT, KR, and ACP specified by module 4 were identified by multiple alignments of the six modules (*13*) (Fig. 3). The exact extent of the carboxyl-terminal and amino-terminal domains for DH and ER, respectively, could not be established, but the best conserved regions between module 4 and rat FAS are located at the two ends of the DH + ER domain (Fig. 3), suggesting that the active site domains for DH and ER are located in the amino-terminal and carboxyl-terminal portion of this segment, respectively (*13*). A similar comparison of the segment of Orf3 corresponding to module 6 with rat FAS (Fig. 3) indicates that the last domain present at the end of Orf3 corresponds to the TE of FAS, as has been previously pointed out (*14*).

A model for 6dEB synthesis. We propose that each of the *eryA* modules participates in 6dEB formation and is specific for one of the six elongation steps required for production of the full-length chain. The FAS-like functions specified by each module constitute one

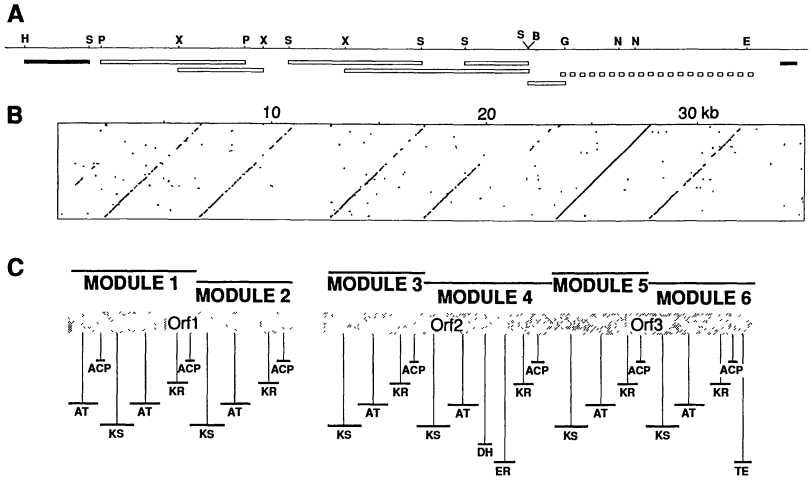

Fig. 2. Organization of *eryA*. (**A**) Gene disruptions. Solid bars refer to EryA$^+$ integrants; open bars to EryA integrants. The dotted bar refers to previously identified *eryA* DNA (*7, 8*). Only the restriction sites that define the fragments employed for gene disruptions or those mentioned in the text are shown according to the abbreviations B, Bam HI; E, Eco RI; G, Bgl II; H, Hind III; N, Nco I; P, Pvu II; S, Sst I; and X, Xho I. The Eco RI site is 11.5 kb downstream of the Bam HI site of *ermE* (*9*). DNA fragments were subcloned into pWHM3 (*10*) by standard methods (*33*), and integrative transformants of erythromycin-producing *S. erythraea* ER720 (*29*) were isolated as described (*11*). The integrated nature of the incoming plasmid and the characterization of the Ery phenotype were performed as described (*7*). (**B**) Repeated units within *eryA*. Comparison of a 4.5-kb subsegment to the 35-kb segment sequenced with the use of COMPARE-DOTPLOT programs (*34*) with a window of 50 and a stringency of 32. (**C**) Extension of the six modules, Orf's and putative FAS-like domains. The nucleotide sequence that corresponds to the three polypeptides can be obtained from GenBank under accession numbers M63676 (for Orf1) and M63677 (for Orf's 2 and 3). Putative FAS-like domains within *eryA* ORF's are identified by bars labeled ACP (acyl carrier protein), AT (acyltransferase), DH (dehydratase), ER (enoylreductase), KR (ketoreductase), KS (ketoacyl-ACP synthase), and TE (thioesterase). Bars are placed at their approximate positions; the length reflects the approximate extent of the domains.

functional synthase unit (SU) and the synthesis of 6dEB proceeds by a FAS-like mechanism. At each biosynthetic step, the acyl-KS is formed, the ACP within the same SU is charged with the proper extender unit, and condensation takes place. The resulting β-ketoacyl–ACP undergoes the extent of processing determined by the particular functional domains present within the SU. The processed acyl chain is then transferred from the ACP to the KS of the next SU, which carries all the functions required for the next elongation step. After six cycles, the full-length C_{15} chain has been synthesized through the concerted action of the six SU's, each carrying the specificity for the proper extender unit and for the extent of processing of the β-carbonyl.

The asymmetry of the structure of 6dEB and the differences among the functional domains specified by each module facilitates the assignment of modules to particular steps in the synthesis. Because a methylene group is present at the C-7 position of the 6dEB ring (Fig. 1), corresponding to biosynthesis step 4, this step must be carried out by the SU determined by module 4 (SU4), the only module that specifies DH and ER. Furthermore, the additional AT and ACP domains determined by module 1 are required for formation of propionyl-KS (via propionyl-ACP); thus module 1 is likely to carry out step 1. Similarly, the presence in module 6 of a domain that corresponds to the TE of FAS suggests the participation of this module in transacylation of the full-length C_{15} acyl chain (step 6). Thus, the different components of putative FAS-like activities from the various *eryA* modules suggest that modules 1, 4, and 6 participate in biosynthesis steps 1, 4, and 6, respectively. In addition, although module 3 carries a segment that corresponds to a KR domain, we propose that this KR is nonfunctional, as it lacks the highly conserved VSRRG motif and 9 out of 25 amino acid residues that are found to be invariant in the other five *eryA* KR domains, as well as in the KR domains from chicken (*19*) and rat FAS (Fig. 4). This suggests that SU3 participates in synthesis step 3. Modules 2 and 5 may also participate in the corresponding synthesis steps, so that the genetic order of the six modules corresponds to the order in which the corresponding synthase units are employed in the six elongation steps.

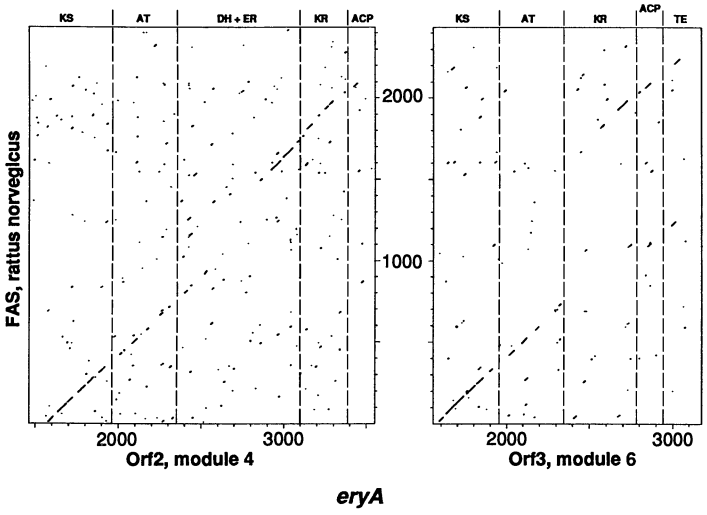

Fig. 3. Comparison of *eryA* with FAS from rat. The translations of the *eryA* segments that correspond to module 4 (left) and module 6 (right) are compared to FAS from rat (*18*) with the use of COMPARE/DOTPLOT programs (*34*) with a window of 30 and a stringency of 16. The proposed extent of each functional domain within *eryA* is marked by vertical dashed lines.

Our model (Fig. 5) displays the proposed intermediates formed after completion of each cycle as enzyme-bound acyl-ACP thioesters. The first ACP from module 1 is acylated by propionate. After condensation of propionate with (2*R*)-mmCoA and β-keto reduction, the first step is completed with formation of the C_5 β-hydroxyacyl-ACP by SU1. The C_5 intermediate is then transferred to the KS of SU2, where (2*S*)-mmCoA is condensed, β-keto reduction takes place, and the C_7 intermediate is transferred from Orf1 to the KS from module 3 in Orf2. Here, (2*R*)-mmCoA is condensed by SU3 and, on the resulting C_9 β-ketothioester, SU4 adds a new (2*R*)-mmCoA unit, with the subsequent full cycle of β-keto reduction, dehydration, and enoylreduction yielding the C_{11} intermediate. The two final elongation steps are carried by SU5 and SU6 in Orf3, both employing (2*S*)-mmCoA units and requiring β-keto reduction, with formation of the C_{13} and then full-length C_{15} chain. Finally, the TE from module 6 transacylates the C_{15} acyl chain from SU6, possibly to CoA, with the resulting acyl-CoA acted upon by a cyclase, which has been proposed to be located 4 kb downstream of module 6 (*20*). Our model is consistent with the results of feeding experiments (*5*) and with the hypothetical macrolactone precursors isolated from some macrolide producers (*6*).

Evidence for the model. We have proposed that each module is involved in a single elongation step, that the choice of extender unit and the extent of processing of the β-carbonyl at each step are determined by the specificity of the corresponding SU and by the functional domains it contains, respectively, and that the linear order of modules corresponds to the succession of elongation steps. Thus, a selective mutation in a given segment of *eryA* should affect a single biochemical event in a predicted elongation step. According to the model, the KR of SU5 is responsible for the introduction of the hydroxyl group appearing at C-5 in the completed polyketide. If such reduction does not take place, and if the succeeding steps in the synthesis, including lactonization, do occur, the compound 5,6-

Fig. 4. Alignments of segments of KR domains from *eryA* Orf's with KR domains from FAS systems. The six *eryA* segments identified by the module that encodes them were manually aligned with a combination of BESTFIT and LINEUP programs (*34*). The resulting consensus was used against the two FAS sequences with PROFILE-SEARCH (*34*). Invariant residues are marked by dashes and gaps by dots. The best conserved segments between *eryA* Orf's and FAS's are enclosed by boxes. The shaded box (top) refers to the proposed NADPH (*16*) binding sites. Bars below alignment of KR's indicate major deviations from possible consensus in the KR from module 3.

```
chicken   KSYIITGGLGGFGLELAQWLI.ERGAQKLVLTSRSGIRTGYQAKCVREWKALGIQVLVSTSDVGTLEGTQLLIEEALKLG.PVGGIFNLAVVL
rat       KSYII---LGGF-LEL-RW-V.LR-AQRLV-TS-S-IRTGYQAKHVR-WRRQ-IHVLVSTSNVSSLEGARALIAEATKLG.PVGGVFNL-MVL
module 6  GTALV---TGAL-GHV-RH-A.RC-VEDLV-VS-R-VDAPGAAELEA-LVAL-AKTTITACDVADREQLSKLLEELRGQGRPVRTVVHT-GVP
module 5  GTVLV---TGGI-AHV-RW-A.RS-AEHLV-LG-R-ADAPGASELRE-LTAL-TGVTIAACDVADRARLEAVLAAERAEGRTVSAVMHA-GVS
module 4  GTVLI---TGTL-RLL-RH-VTEH-VRHLL-VS-R-ADAPGSDELRA-IEDL-ASAEIAACDTADRDALSALLDG.LPR..PLTGVVHA-GVL
module 2  GTILV---TAGL-AEV-RW-A.GR-AEHLA-VS-R-PDTEGVGDLTA-LTRL-ARVSVHACDVSSREPVRELVHGLIEQGDVVRGVVHA-GLP
module 1  GTVLV---TGGV-GQI-RW-A.RR-APHLL-VS-S-PDADGAGELVA-LEAL-ARTTVAACDVTDRESVRELLGG.IGDDVPLSAVFHA-ATL
module 3  GTVLV--AASPV-DQLVRW-A.DR-AERLV-.....AGACPGDDLLAAVEEA-A.SAVVAQDAA...ALREALG.....DEPVTALVHAGTLT

chicken   KDAMIENQTPELFWEVNKPKYSGTLHLDWVTRKKCPDLD..YFVVFSSVSCGRGNAGQSNYGFANSAMERICEQRHHDGLPGLAVQWGAIGDV
rat       RDAMLENQTPELFQDVNKP-YNGTLN-DRATREACPELD..Y-VA---VSCGRGNAGQSN-GFA-STMERICEQ-RHD-LPGLAVQ--AIGDV
module 6  ESRPLHEIGE..LESVCAA-VTGARL-D....ELCPDAE..T-VL---GAGVWGSANLGA-SAA-AYLDALAHR-RAE-RAATSVA--AWAGE
module 5  TSTPLDDLTEAEFTEIADV-VRGTVN-D....ELCPDLD..A-VL---NAGVWGSPGLAS-AAA-AFLDGFARR-RSE-APVTSIA--LWAGQ
module 4  ADGLVTSIDEPAVEQVLRA-VDAAWN-H....ELTANTGLSF-VL---AASVLAGPGQGV-AAA-ESLNALAAL-RTR-LPAKALG--LWAQA
module 2  QQVAINDMDEAAFDEVVAA-AGGAVH-D....ELCSDAE..L-LL---GAGVWGSARQGA-AAG-AFLDAFARH-RGR-LPATSVA--LWAAG
module 1  DDGTVDTLTGERIERASRA-VLGARN-H....ELTRELDLTA-VL---FASAFGAPGLGG-APG-AYLDGLAQQ-RSD-LPATAVA--TWAGS
module 3  NFGSISEVAPEEFAETIAA-TALLAV-D....EVLGDRAVEREVYC--VAGIWGGAGMAA-AAGSAYLDALAEHHRAR-RSCTSVA-TPWALP
```

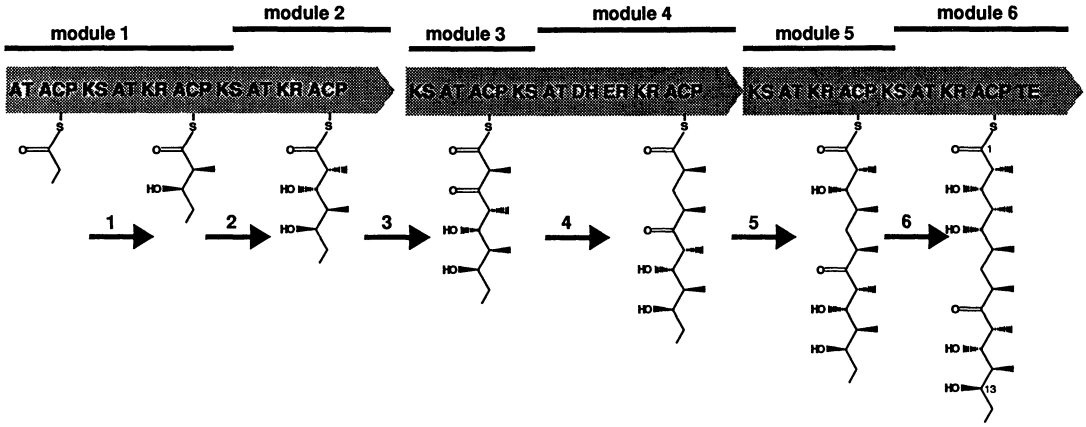

Fig. 5. Model for 6-deoxyerythronolide B formation. The starter unit and the intermediates formed at the end of each elongation cycle are shown as the corresponding ACP thioesters. The FAS-like functions responsible for each elongation step are shown inside each Orf, with the corresponding modules above the three Orf's. Numbered arrows refer to the six elongation steps. Lactonization of the C_{15} chain between carbons 1 and 13 results in the formation of 6-deoxyerythronolide B (Fig. 1).

dideoxy-5-oxoerythronolide B should be produced. An 813-bp in-frame deletion was introduced in vitro into the DNA segment of module 5 that specifies the KR domain, and the mutant allele was introduced into *S. erythraea* to replace the wild-type counterpart in the chromosome (*21*). The resulting strain AKR5 produced 5,6-dideoxy-3α-mycarosyl-5-oxoerythronolide B as the major component, as well as minor amounts of 5,6-dideoxy-5-oxoerythronolide B (Fig. 6). This indicates that the KR of SU5 is responsible for β-keto reduction in synthesis step 5 and provides direct evidence for the model. It also indicates that hydroxylation at C-6, which normally takes place on 6dEB, does not occur on the corresponding 5-deoxy-5-oxo derivative and confirms the substrate flexibility of the enzyme machinery which does not require hydroxyl groups at C-6 as recently reported (*20*) or at C-5 and C-6, as our data show. Because desosamine attachment occurs at the C-5-hydroxyl group of the lactone ring, 5,6-dideoxy-3α-mycarosyl-5-oxoerythronolide B could not be further processed, and in this way the compound was accumulated by strain AKR5.

Minor quantities of the compounds 2-norerythromycins have been isolated from the *S. erythraea eryA* strain 9EI41 when transformed with a cosmid library of heterologous DNA. This family of compounds resulted from the incorporation of malonate instead of methylmalonate at the sixth condensation step (*22*). Strain 9EI41 carries an 87-bp in-frame deletion in the segment of module 6 that specifies the AT domain (*7*), thus this result is consistent with the involvement of the AT from module 6 in synthesis step 6, as predicted by the model, and suggests that the DNA that complemented the *eryA* mutation in strain 9EI41 encoded a malonyl-CoA transferase. The yield of 5-deoxy-5-oxoerythronolide B is about 100-fold greater than that of 2-norerythromycin and thus highlights the requirement for precise interaction among the various domains in the multifunctional polypeptides. The observation of alternate compounds illustrates the lack of absolute specificity of the KS of SU6, which condensed a malonate into the growing chain in this case and employed a β-keto-C_{13} substrate in the case of strain AKR5.

Implications of the model. The proposed model explains how the synthesis of an asymmetric polyketide such as 6dEB is pro-

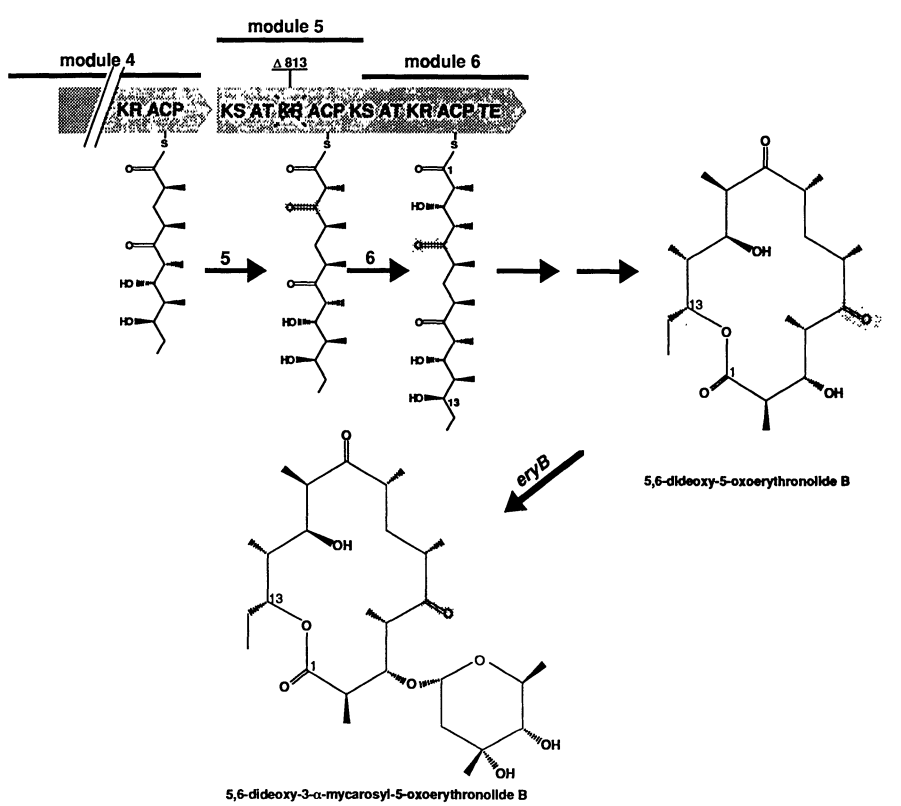

Fig. 6. Pathway for formation of 5,6-dideoxy-3α-mycarosyl-5-oxoerythronolide B. The altered Orf3 that carried an inactive KR in SU5 and the corresponding 813-bp deletion in strain AKR5 are shown on top. The keto group that remained after elongation step 5 is highlighted throughout the synthesis. Strain AKR5 was fermented for 3 days according to conditions described (*22*) and the broth was extracted three times with ethyl acetate. The dried extract was digested in toluene and chromatographed on a column of silica gel (Merck Kiesel gel 60, 70-230 mesh) and eluted with a step gradient of isopropanol in toluene. The 5 and 7 percent isopropanol eluates were combined and concentrated to an oily residue (800 mg), which was digested in a solvent system consisting of *n*-heptane:chloroform:ethanol (10:10:1, by volume) and chromatographed on a Sephadex LH-20 column that has been packed and developed with the same solvent system. Fractions (9 to 13) were combined, concentrated, and crystallized from a mixture of *n*-heptane and ethyl acetate to give 5,6-dideoxy-3α-mycarosyl-5-oxoerythronolide B (45.2 mg, melting point 163° to 164°C). Structure was deduced by spectral analysis and confirmed by single crystal x-ray to diffraction analysis. Fractions (15 to 17) were combined and crystallized from ethyl acetate to yield 5,6-dideoxy-5-oxoerythronolide B (5.3 mg) with spectral characteristics identical to those of an authentic sample (*32*).

grammed in the producing microorganism. Rather than through the utilization of a hypothetical enzyme capable of altering its specificity in response to variations in length and side groups of the growing acyl chain, the correct polyketide is made through the concerted action of six SU's, each devoted to a single FAS-like cycle. The final molecular structure is programmed by (i) use of a single and distinct module encoding the information required for each elongation step and (ii) selection of the correct SU to be transacylated by the growing chain.

The current model for palmitic acid formation in type I FAS systems calls for transacylation of the growing chain after each cycle between two distinct but identical polypeptides (*2*). Six different SU's participate in 6dEB synthesis, however. Some mechanism must thus ensure that the correct SU is transacylated after each cycle by the required substrate. The fidelity of this transacylation may depend on appropriate steric juxtaposition of ACP and KS domains of adjacent SU's or on the substrate specificity of the KS. The observation that a β-hydroxyl C_{13} chain is transferred from the ACP of SU5 to the KS of SU6 in wild-type Orf3, whereas the same transfer occurs with a β-keto C_{13} chain by AKR5 Orf3, suggests that the precise intramolecular transfer is independent of the acyl chain substrate structure. The close proximity of ACP and KS domains that are part of two adjacent SU's in the primary structure of each polypeptide may provide the necessary steric juxtaposition required for correct intramolecular transacylation. Intermolecular transfer could also result from having a single KS domain sterically available to each ACP domain if the three polypeptides assemble in a specific manner in a hypothetical PKS complex. It is possible that such precise geometric arrangement of the three polypeptides is facilitated by the colinearity of genetic and biochemical order in 6dEB formation.

The presence of an AT specific for propionyl-CoA in SU1 and the formation of 2-norerythromycin in a strain defective in the AT of SU6 suggest a direct function for the AT in the choice of the correct extender unit. Furthermore, the apparent lack of specificity of the KS and ACP domains of SU6 for the extender unit employed, (2*S*)-mmCoA in the wild-type and malonylCoA in the 2-norerythromycin–producing derivative of strain 9EI41, suggests that these domains do not function in extender unit choice.

Homopolymeric aromatic polyketides made through successive condensations of the same extender unit are synthesized by type II PKS's in bacteria (*17*) and by a type I enzyme in fungi (*23*). Synthesis of these simple polyketides requires the participation of no more than six polypeptides in the type II systems (*3*). According to our model, the fidelity of 6dEB synthesis is ensured by the involvement of one SU for each elongation step. It is difficult to conceive how the synthesis of this complex polyketide, which requires the proper interactions among six different SU's for a total of 28 functional FAS-like domains, could be achieved in a type II system. The occurrence in prokaryotes of type I or type II systems may thus relate to the complexity of the polyketide made. The preliminary evidence suggesting that *S. erythraea* FAS is a type II enzyme (*24*) is consistent with this proposal and indicates that type I and type II systems may coexist within one organism. The proposed model for the synthesis of 6dEB and the organization of the corresponding PKS in large multifunctional polypeptides are analogous to the nonribosomal template synthesis of peptide antibiotics and to the organization of the relative synthetases (*25*).

Genes for the formation of the macrolides tylosin (*26*), spiramycin (*27*), and avermectin (*28*) also contain repeated units. We propose, therefore, that the genes for the biosynthesis of macrolides and possibly of other complex polyketides, for example polyenes and polyethers, are also organized in modules in a manner similar to *eryA*. If, as illustrated by the example of the KR from module 5 in *eryA*, each SU exhibits some substrate flexibility, an alteration of a functional domain of the PKS may result in the formation of a novel polyketide of predicted structure. Modular organization of PKS genes and colinearity of the genetic and biochemical order permits the opportunity to make select genetic alterations in polyketide-producing organisms and thus offers possibilities for the tailored production of polyketide structures.

REFERENCES AND NOTES

1. A. J. Birch and F. W. Donavan, *Aust. J. Chem.* 6, 360 (1953).
2. S. J. Wakil, *Biochemistry* **28**, 4523 (1989).
3. D. A. Hopwood and D. H. Sherman, *Annu. Rev. Genet.* **24**, 37 (1990).
4. S. Omura and Y. Tanaka, in *Macrolide Antibiotics: Chemistry, Biology and Practice*, S. Omura, Ed. (Academic Press, New York, 1984), p. 199.
5. D. E. Cane and C. Yang, *J. Am. Chem. Soc.* **109**, 1255 (1987); S. Yue, J. S. Duncan, Y. Yamamoto, C. R. Hutchinson, *ibid.*, p. 1253.
6. K. Kinoshita, S. Takenaka, M. Hayashi, *J. Chem. Soc. Chem. Commun.* **1988**, 943 (1988); M. L. B. Huber *et al.*, *Antimicrob. Agents Chemother.* **34**, 1535 (1990).
7. J. S. Tuan *et al.*, *Gene* **90**, 21 (1990).
8. J. M. Weber *et al.*, *J. Bacteriol.* **172**, 2372 (1990).
9. M. J. Bibb, G. R. Janssen, J. M. Ward, *Gene* **38**, 215 (1985); H. Uchiyama and B. Weisblum, *ibid.* **38**, 103 (1985).
10. J. Vara, M. Lewandowska-Skarbek, Y.-G. Wang, S. Donadio, C. R. Hutchinson, *J. Bacteriol.* **171**, 5872 (1989).
11. J. M. Weber and R. Losick, *Gene* **68**, 173 (1988).
12. Plasmid DNA was transformed into *S. erythraea* ER720 (*29*) and two integrants from each transformation were analyzed for the production of bioactive compounds and for the presence of ethyl acetate–extractable metabolites by thin-layer chromatography (TLC) as described (7). The EryA phenotype was characterized by the lack of known erythromycin intermediates and by the absence of detectable bioactivity in the ethyl acetate extract (7).
13. M. J. Staver, S. Donadio, J. S. Tuan, L. Katz, in preparation.
14. J. Cortes, S. H. Haydock, G. A. Roberts, D. J. Bevitt, P. F. Leadlay, *Nature* **348**, 176 (1990).
15. For a recent compilation of active site motifs for these activities, see (*3*) and (*14*). Abbreviations for the amino acid residues are A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
16. N. S. Scrutton, A. Berry, R. N. Perham, *Nature* **343**, 38 (1990).
17. M. J. Bibb, S. Biro, H. Motamedi, J. F. Collins, C. R. Hutchinson, *EMBO J.* **8**, 2727 (1989); D. H. Sherman *et al.*, *ibid.*, p. 2717.
18. M. Schweizer, K. Takabayashi, T. Laux, K.-F. Beck, R. Schreglmann, *Nucleic Acids Res.* **17**, 567 (1989).
19. Z. Yuan, W. Liu, G. G. Hammes, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 6328 (1988); K. P. Holzer, W. Liu, G. G. Hammes, *ibid.* **86**, 4387 (1989); S. S. Chirala *et al.*, *J. Biol. Chem.* **264**, 3750 (1989).
20. J. M. Weber, J. O. Leung, S. J. Swanson, K. B. Idler, J. B. McAlpine, *Science* **252**, 114 (1991).
21. The 9.5-kb Bam HI–Xho I segment between coordinates 22.0 and 31.5 (Fig. 2A) was subcloned into the Bam HI–Sal I fragment of pUC19 (*30*) to generate pAIBX9. The internal 813-bp Nco I fragment of this plasmid was removed and in the resulting construct the pUC19 portion was replaced with pWHM3 (*10*) by Eco RI–Hind III digestion. The resulting plasmid pAKR5.2 was introduced into *S. erythraea* ER720 protoplasts and integrative transformants were isolated (*11*). Thiostrepton-sensitive colonies (five) were isolated from one integrative transformant as described (*31*) and three were found upon genomic Southern (DNA) blot hybridization to have retained the deleted *eryA* segment in the chromosome. These colonies, which upon TLC analysis (7) showed no visible erythromycin intermediates but produced a product that corresponded to authentic 5,6-dideoxy-5-oxoerythronolide B (*32*), were named strain AKR5.
22. J. B. McAlpine *et al.*, *J. Antibiot. (Tokyo)* **40**, 1115 (1987); in preparation.
23. J. Beck, S. Ripka, A. Seigner, E. Schlitz, E. Schweizer, *Eur. J. Biochem.* **192**, 487 (1990).
24. R. S. Hale, K. N. Jordan, P. F. Leadlay, *FEBS Lett.* **224**, 133 (1987).
25. H. Kleinkauf and H. von Dohren, *Annu. Rev. Microbiol.* **41**, 259 (1987); *Eur. J. Biochem.* **192**, 1 (1990).
26. R. H. Baltz and E. T. Seno, *Annu. Rev. Microbiol.* **42**, 547 (1988).
27. M. A. Richardson *et al.*, *J. Bacteriol.* **172**, 3790 (1990).
28. D. J. MacNeil, K. M. Gewain, C. L. Ruby, T. MacNeil, P. H. Gibbons, *J. Cell. Biochem.* **14A**, (suppl.) 92, (1990).
29. J. DeWitt, *J. Bacteriol.* **164**, 969 (1985).
30. J. Vieira and J. Messing, *Gene* **19**, 259 (1982).
31. S. Donadio and C. R. Hutchinson, *Gene*, in press.
32. J. R. Martin and T. J. Perun, *Biochemistry* **7**, 1728 (1968).
33. T. Maniatis, E. F. Fritsch, J. Sambrook, *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982).
34. J. Devereux, P. Haeberli, O. Smithies, *Nucleic Acids Res.* **12**, 387 (1984).

35. We thank D. Brown for help with fermentations; D. Whittern and P. Hill for spectral determinations; S. Spanton for x-ray diffraction analysis; and C. R. Hutchinson, R. Losick, and P. Youngman for critical reading of the manuscript.

25 January 1991; accepted 20 March 1991