JOURNAL OF BACTERIOLOGY, July 1995, p. 3673–3679 Vol. 177, No. 13
0021-9193/95/$04.00+0
Copyright © 1995, American Society for Microbiology

A *Sorangium cellulosum* (Myxobacterium) Gene Cluster for the Biosynthesis of the Macrolide Antibiotic Soraphen A: Cloning, Characterization, and Homology to Polyketide Synthase Genes from Actinomycetes

THOMAS SCHUPP,[1]* CHRISTIANE TOUPET,[1] BERNARD CLUZEL,[1] SNEZANA NEFF,[1] STEVEN HILL,[2] JAMES J. BECK,[2] AND JAMES M. LIGON[2]

Pharma Research, Ciba-Geigy AG, CH 4002 Basel, Switzerland,[1] and Agricultural Biotechnology Research Unit, Ciba-Geigy Corporation, Research Triangle Park, North Carolina 27709[2]

Received 27 February 1995/Accepted 20 April 1995

A 40-kb region of DNA from *Sorangium cellulosum* So ce26, which contains polyketide synthase (PKS) genes for synthesis of the antifungal macrolide antibiotic soraphen A, was cloned. These genes were detected by homology to *Streptomyces violaceoruber* genes encoding components of granaticin PKS, thus extending this powerful technique for the identification of bacterial PKS genes, which has so far been applied only to actinomycetes, to the gram-negative myxobacteria. Functional analysis by gene disruption has indicated that about 32 kb of contiguous DNA of the cloned region contains genes involved in soraphen A biosynthesis. The nucleotide sequence of a 6.4-kb DNA fragment, derived from the region with homology to granaticin PKS genes, was determined. Analysis of this sequence has revealed the presence of a single large open reading frame beginning and ending outside the 6.4-kb fragment. The deduced amino acid sequence indicates the presence of a domain with a high level of similarity to β-ketoacyl synthases that are involved in polyketide synthesis. Other domains with high levels of similarity to regions of known polyketide biosynthetic functions were identified, including those for acyl transferase, acyl carrier protein, ketoreductase, and dehydratase. We present data which indicate that soraphen A biosynthesis is catalyzed by large, multifunctional enzymes analogous to other bacterial PKSs of type I.

Polyketides are a large group of structurally diverse secondary metabolites, several of which have applications as antibiotics, immunosuppressants, anticancer agents, and veterinary products. In recent years genetic studies have yielded detailed information on the organization and function of genes involved in the biosynthesis of polyketides in microorganisms (17). Cloning, sequencing, and functional analysis of these genes have improved our understanding of the genetic programming of polyketide synthases (PKSs) leading to a variety of different polyketides (15, 16). Among the prokaryotes, however, these genetic studies have focused on actinomycetes, which are gram-positive soil bacteria that are a very rich source for polyketides. Another group of bacteria that produce a large number of secondary metabolites, including several polyketides, are the gram-negative myxobacteria, which, like the actinomycetes, are soil bacteria that have a complex life cycle with multicellular differentiation (26). However, knowledge of the genetic control of polyketide synthesis in these microorganisms is still very limited. Genes of *Myxococcus xanthus* that are involved in the biosynthesis of the macrolide antibiotic TA (identical to myxovirescin A [31]) have been cloned and were found to be clustered on a 36-kb chromosomal segment (33). To date, no sequence data or detailed functional analysis of this gene cluster has been published.

To obtain more information on the genetic control of polyketide biosynthesis in myxobacteria and to detect potential similarities to well-studied systems of actinomycetes, we have undertaken and describe here the cloning and analysis of genes for the synthesis of the macrolide soraphen A. Soraphen A is a highly active antifungal antibiotic produced by the myxobacterium *Sorangium cellulosum* So ce26 (12). Soraphen A inhibits the growth of a variety of fungi with a spectrum of activity that is particularly suited for the control of fungal plant pathogens. A very interesting feature of soraphen A is its novel mode of action. It is a strong inhibitor of the fungal acetyl coenzyme A carboxylase and therefore disrupts the synthesis of fatty acids, leading to growth inhibition and finally to the death of the fungal cell (25, 32). The chemical structure of soraphen A has been elucidated (4). It is a macrocyclic polyketide containing an 18-member lactone ring with a characteristic unsubstituted phenyl ring attached to it (Fig. 1).

Several *Streptomyces* genes coding for the synthesis of polyketides have been found to have significant homology to each other (21), and this homology has been used to detect and clone polyketide genes from other members of the genus *Streptomyces* (2, 19, 21, 28, 34). We were interested to determine whether a significant level of nucleotide sequence homology for polyketide biosynthesis genes extends also to other groups of bacteria, such as the myxobacteria. Streptomycetes and myxobacteria have a similar, high G+C content in their DNA (about 70%). This fact increases the chance of finding DNA homology among genes that are involved in similar enzymatic reactions for chain elongation (ketosynthase) and processing of the keto groups (ketoreductase, dehydrase, and enoylreductase) involved in polyketide biosynthesis. For this reason we have chosen the experimental approach of using DNA frag-

* Corresponding author. Mailing address: Ciba-Geigy AG, Pharma Research, K-681.3.46, CH-4002 Basel, Switzerland. Phone: 61 696 5051. Fax: 61 696 9301.

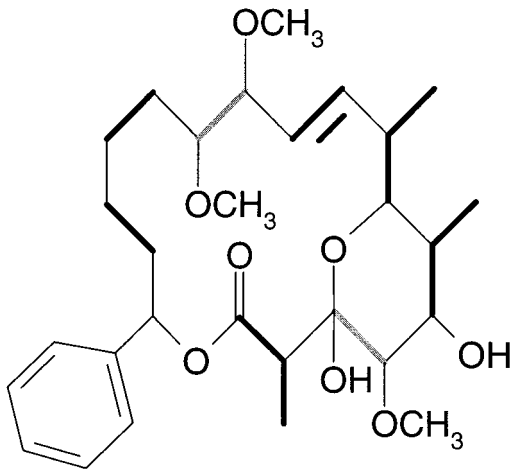

FIG. 1. Chemical structure of soraphen A, showing the derivation of each carbon atom in the lactone ring (3a). Carbons connected by thick black and thick grey lines are derived from acetate and glycerol, respectively; the three methyl side groups, the carbons to which they are connected, and one adjacent carbon, indicated by branched thick black lines, are derived from propionate; and C-1 and the phenyl group are derived from phenylalanine.

ments coding for a *Streptomyces* PKS as a hybridization probe for the isolation of genes from the myxobacterium *S. cellulosum* that are involved in the biosynthesis of the polyketide soraphen A.

MATERIALS AND METHODS

Strains and plasmids. The *S. cellulosum* strains used were the wild-type strain So ce26 (12) and its streptomycin-resistant mutant SJ3 (18). *Escherichia coli* HB101 and JM101 were used for cloning, and *E. coli* ED8767 (23) was used for conjugative plasmid transfer. *E. coli* vectors used for cloning and sequencing were pHC79 (14), pBR322 (6), pUC18 and pUC19 (24), and pBluescript (Stratagene, La Jolla, Calif.). For conjugative plasmid transfer the *E. coli* vectors pUZ8 (13), pSUP2021 (29), and pCIB132 were used. pCIB132 is a derivative of pSUP2021 which was constructed by us by excision and subsequent reversal of the 5-kb *Not*I fragment followed by removal of the resulting 3-kb *Bam*HI fragment, leaving a unique *Bam*HI site for cloning.

Media and growth conditions. *E. coli* was grown at 37°C in Luria broth or on Luria broth agar, with the appropriate antibiotic. *S. cellulosum* was grown at 30°C on solid medium S42 (18) or in liquid medium G55, which contained, per liter, 8 g of potato starch (Noredux, Cerestar Italia S. p. a., Milan, Italy), 8 g of potato dextrin (Blattmann, Wädenswil, Switzerland), 2 g of glucose, 6 g of defatted soybean meal (Mucedola S. r. l., Settimo Milanese, Italy), 2 g of yeast extract (Fould & Springer, Maison Alfort, France), 1 g of aspartic acid, 1 g of $CaCl_2 \cdot 2H_2O$, 1 g of $MgSO_4 \cdot 7H_2O$, 12 g of HEPES (*N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid), and 0.008 g of Fe(III)-EDTA. The pH of the medium was adjusted to 7.5 with NaOH before autoclaving. Fifty grams of adsorber resin XAD1180 (Rohm and Haas, Frankfurt, Germany) per liter was added to medium G55 when it was used as the production medium for soraphen.

Genetic procedures. Standard genetic techniques with *E. coli* and for in vitro DNA manipulations were as described earlier (22). Isolation of total DNA from *S. cellulosum*, Southern blot analysis, and matings between *E. coli* and *S. cellulosum* for plasmid transfer were as described previously (18). *S. cellulosum* transconjugants were selected on S42 agar containing 20 mg of phleomycin (Cayala, Toulouse, France) per liter and 300 mg of streptomycin (Sigma, St. Louis, Mo.) per liter.

Construction of an *S. cellulosum* genomic library. Genomic DNA of strain So ce26 was partially digested with *Sau*3A, treated with calf intestinal phosphatase, and size fractionated on sucrose density gradients. DNA fragments of 35 to 45 kb were ligated to *Bam*HI-restricted pHC79 cosmid DNA. The ligated DNA was packaged into lambda phage particles by using an in vitro DNA packaging system from Boehringer Mannheim Biochemicals (Indianapolis, Ind.), and these particles were used to transfect *E. coli* HB101.

DNA sequencing. DNA sequencing was accomplished by the dideoxynucleotide chain termination method (27) with an automated sequencer (model 373A; Applied Biosystems, Foster City, Calif.). DNA fragments were subcloned into pBluescript, and sequencing was performed with the universal primers specific for the opposite sides of the cloning sites. Subsequent sequencing was performed with oligonucleotide primers that were designed for 3′ regions of newly obtained DNA sequence. Oligonucleotides were synthesized with an Applied Biosystems model 394 automated DNA synthesizer. In all cases, the sequences of both DNA strands were determined.

Computer analysis of DNA and protein sequences. Primary DNA sequence data were analyzed and assembled by using software from DNASTAR (Madison, Wis.). DNA and protein sequences were analyzed with the University of Wisconsin Genetics Computer Group programs (11).

Assay of soraphen production. Cells derived from a single colony of *S. cellulosum* were cultivated at 30°C for 7 days as a surface culture on an area of 1 to 2 cm^2 on S42 agar (with appropriate antibiotics added for transconjugants). The cells were harvested with a plastic loop, transferred into 5 ml of medium G55 in a 20-ml flask, and incubated at 30°C on a rotary shaker (180 rpm) for 5 to 6 days (all further incubations were at 30°C and 180 rpm). The whole culture was then transferred into 50 ml of medium G55 in a 200-ml flask and incubated for 3 days. Five milliliters of this second preculture was transferred into a 200-ml flask with 50 ml of medium G55 containing the resin XAD1180. After 4 and 7 days of incubation, 2.5 ml of each of the two feeding solutions, i.e., 6% yeast extract (Fould & Springer) and 10% Bacto-maltose (Difco) plus 20% potato dextrin (Blattmann) was added. After 14 days of incubation, the resin of the whole culture was harvested on a polyester sieve (B 420-47-N; Satorius, Göttingen, Germany) and eluted with 50 ml of isopropanol by shaking for 1 h at 180 rpm in a 200-ml shake flask. The isopropanol eluate was centrifuged for 5 min in an Eppendorf centrifuge, and the clear supernatant was analyzed by high-pressure liquid chromatography (HPLC) for its soraphen A content. Reverse-phase HPLC analysis was performed with a silica-based stationary phase (Lichropher 100 RP-18 [5 mm]; Merck, Darmstadt, Germany) in a 125- by 4-mm column and with a phosphate buffer (pH 6.5) acetonitrile gradient as the mobile phase. Soraphen A was detected by its absorption at 210 nm.

Nucleotide sequence accession number. The DNA sequence reported here has been deposited in the GenBank and EMBL data libraries under accession number U24241.

RESULTS

Cloning of *S. cellulosum* DNA homologous to the *graI* region of *Streptomyces violaceoruber*. Several polyketide-producing streptomycetes have been demonstrated by Southern blot analysis to contain DNA sequences that are homologous to the *actI* region encoding part of the *Streptomyces coelicolor* PKS for actinorhodin biosynthesis (21). This homology has been used to identify and clone, among others, the PKS genes (*gra*) of the granaticin producer *Streptomyces violaceoruber* (28). To demonstrate the existence of sequences in *S. cellulosum* So ce26 that are homologous to PKS genes from actinomycetes, we performed a chromosomal Southern blot by using as a probe the 4.8-kb *Bam*HI fragment encoding open reading frames (ORFs) 1 to 4 of the *graI* region (28). At a low stringency several *Sal*I DNA fragments of *S. cellulosum* So ce26 which hybridized weakly to this probe were visible (Fig. 2A). This experiment demonstrated that in *S. cellulosum* there are DNA regions which show a degree of homology to *graI* that could allow their identification and cloning by DNA hybridization.

A cosmid library of *S. cellulosum* So ce26 DNA was constructed in the cosmid vector pHC79. By screening the library with the *graI* probe, 36 of 1,300 colonies gave a weak hybridization signal. The plasmid DNAs of these clones were isolated and analyzed by Southern blotting. One clone, p98/1, revealed a 1.8-kb *Sal*I fragment which gave a clear hybridization signal with the *graI* probe (Fig. 2A). This fragment was the same size as one of the *Sal*I fragments from *S. cellulosum* genomic DNA that hybridized to the same probe. In contrast to the other clones that contained hybridizing bands, p98/1 also gave the same hybridization signal with a 1.3-kb *Nco*I subfragment of *graI*, representing only ORF2.

To verify the presence of a *Sal*I fragment of identical size in the chromosome of *S. cellulosum* So ce26, a second Southern blot was performed with the 1.8-kb *Sal*I fragment of p98/1 as the probe. This experiment confirmed the presence of an identical genomic fragment (Fig. 2B), indicating that we have cloned an original chromosomal region of *S. cellulosum*. In addition, this blot demonstrated the presence of two smaller fragments with high levels of homology to the 1.8-kb fragment

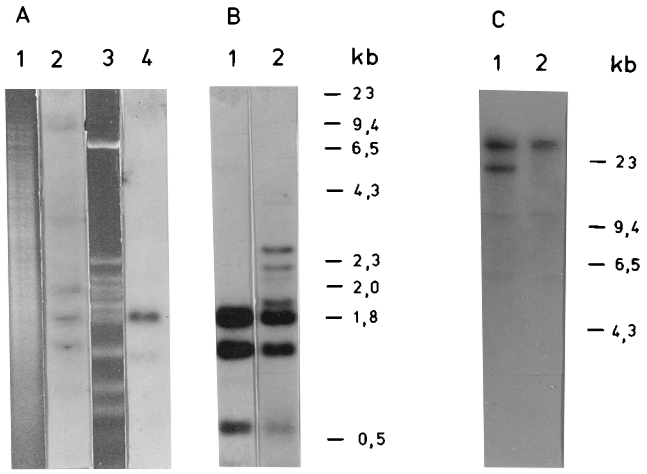

FIG. 2. Southern hybridization analysis of DNA fragments from *S. cellulosum* So ce26. (A) *Sal*I-cleaved *S. cellulosum* So ce26 genomic DNA and p98/1 plasmid DNA in ethidium bromide-stained agarose gels (lanes 1 and 3, respectively) and after hybridization with the ^{32}P-labelled *graI* gene fragment from *Streptomyces violaceoruber* (lanes 2 and 4, respectively). (B) Hybridization of the ^{32}P-labelled 1.8-kb *Sal*I DNA fragment derived from plasmid p98/1 to *Sal*I-cleaved p98/1 plasmid DNA and *S. cellulosum* genomic DNA (lanes 1 and 2, respectively). (C) Hybridization of the ^{32}P-labelled 1.8-kb *Sal*I DNA fragment to *Eco*RV-cleaved genomic DNAs from the *S. cellulosum* So ce26 transconjugant blocked in soraphen A biosynthesis by integration of the recombinant plasmid containing the 6.4-kb *Pvu*I fragment and from the parental strain SJ3 (lanes 1 and 2, respectively). Washing of the filters was in 0.2× SSC (1× SSC is 0.15 M NaCl plus 15 mM sodium citrate) at 55, 60, and 70°C for panels A, B, and C, respectively. Marker positions indicated at the right in panels B and C are from λ DNA cleaved with *Hin*dIII and from the 1.8-kb hybridizing fragment.

in both plasmid p98/1 and the genomic DNA of *S. cellulosum* So ce26.

The cloned *S. cellulosum* chromosomal region in p98/1 was further characterized by restriction enzyme analysis. The restriction map of the *S. cellulosum* insert in p98/1 is shown in Fig. 3. From the sizes of the different restriction fragments, a total length of approximately 39.6 kb for the cloned DNA in p98/1 can be calculated. The 1.8-kb *Sal*I fragment that gave the hybridization signal in the Southern blot experiments was found to be internal to the 6.4-kb *Pvu*I fragment in the cloned region (Fig. 3).

Functional test by gene disruption. The involvement of the cloned region in soraphen biosynthesis was tested by gene disruption. For these experiments we used the gene transfer and plasmid vector system developed by us for *S. cellulosum* (18). Four *Pvu*I fragments (12.5, 6.4, 4.0, and 3.8 kb) of the p98/1 insert (Fig. 3, upper shaded bars) were subcloned into the *Pvu*I site of the mobilizable plasmid pSUP2021 and introduced by transformation into *E. coli* ED8767 containing the helper plasmid pUZ8, which supplied the transfer functions in *trans* for plasmid mobilization. As a control plasmid, pSJB55 (18), containing a random 3.5-kb *Pvu*I genomic fragment derived from *S. cellulosum* So ce26 and cloned into pSUP2021, was used. The resulting *E. coli* ED8767 derivatives containing the recombinant pSUP2021-derived plasmids were used as donors for mating experiments with *S. cellulosum* SJ3, a streptomycin-resistant mutant of *S. cellulosum* So ce26 that allows counter selection with streptomycin after mating as the recipient.

All four *Pvu*I fragments gave stable bleomycin-resistant *S. cellulosum* transconjugants at frequencies similar to those obtained with the control plasmid, pSJB55. The soraphen productivities of these transconjugants were determined in shake flask cultures by HPLC analysis. The results are summarized in Table 1 and show clearly that the four *Pvu*I fragments yield transconjugants with blocked soraphen biosynthesis. Previous molecular characterization of transconjugants had shown that pSUP2021-derived recombinant plasmids were maintained in *S. cellulosum* only after integration into the chromosome by homologous recombination between the cloned insert and the homologous region in the chromosome (18). In the case of transconjugants with the recombinant plasmid containing the 6.4-kb *Pvu*I fragment, we confirmed its integration into the chromosome by Southern blot analysis (Fig. 2C). Hybridization of the 1.8-kb *Sal*I fragment derived from the 6.4-kb *Pvu*I fragment to *Eco*RV-cleaved genomic DNA from the parent strain So ce26 revealed the presence of a single homologous fragment 29 kb in size, as expected from the restriction map of p98/1 (Fig. 3). However, hybridization to genomic DNA derived from a transconjugant blocked in soraphen biosynthesis indicated the presence of two homologous *Eco*RV fragments (Fig. 2C). This is consistent with the integration of pSUP2021 containing a single *Eco*RV site and the 6.4-kb *Pvu*I fragment into the chromosome of So ce26 within the 6.4-kb *Pvu*I chromosomal fragment.

To further demonstrate that blocked soraphen biosynthesis in the transconjugants is a direct result of gene disruption, we screened for a soraphen-producing phenotype in one transconjugant derived from integration of the 6.4-kb *Pvu*I fragment. After several generations of unselective growth in liquid culture and plating of appropriate dilutions on S42 agar, 1 of 11,000 colonies was found in an agar diffusion test with *Saccharomyces cerevisiae* as the test organism to have reverted to a soraphen-producing phenotype. This colony was subsequently cultivated in liquid culture and tested by HPLC analysis for soraphen production (Table 1) and plasmid loss (Southern blot). The results of these analyses demonstrated that soraphen production was restored to nearly the same level as that of the parental strain SJ3 and that pSUP2021 sequences were completely absent in this revertant. These analyses clearly demonstrate a direct correlation between soraphen production and gene disruption in the 6.4-kb *Pvu*I region of the *S. cellulosum* chromosome.

Size of the region involved in soraphen biosynthesis in the cloned insert of plasmid p98/1. The four *Pvu*I fragments that were shown to be involved in soraphen biosynthesis cover a region of 29 kb (Fig. 3, upper shaded bars). In order to determine the limits of the soraphen A biosynthetic gene region of *S. cellulosum* So ce26, further gene disruption experiments were performed with other gene fragments. Three *Bgl*II fragments (2.9, 3.2, and 2.9 kb in size) situated near the right end of the insert of p98/1 (fragments B1, B2, and B3 in Fig. 3, respectively) and the 1.3-kb *Bam*HI fragment at the far left end were analyzed. The fragments were subcloned into the unique *Bam*HI site of the mobilizable plasmid pCIB132, a derivative of pSUP2021. Plasmid transfer to *S. cellulosum* with *E. coli* ED8767 containing pUZ8 was as described for the experiment with the *Pvu*I fragments. As outlined in Fig. 3 (lower shaded bars), the *Bam*HI fragment located at the extreme left border of the cloned region and one of the three *Bgl*II fragments yielded transconjugants with blocked soraphen biosynthesis. The two *Bgl*II fragments near the right end of the p98/1 insert (B2 and B3; open bars in Fig. 3), however, yielded transconjugants with normal soraphen production.

Sequence analysis of the 6.4-kb *Pvu*I fragment. In order to further characterize the cloned region of *S. cellulosum* So ce26 and to analyze potential similarities to PKS genes from actinomycetes, the nucleotide sequence of the 6.4-kb *Pvu*I fragment was determined (see Materials and Methods). The nu-

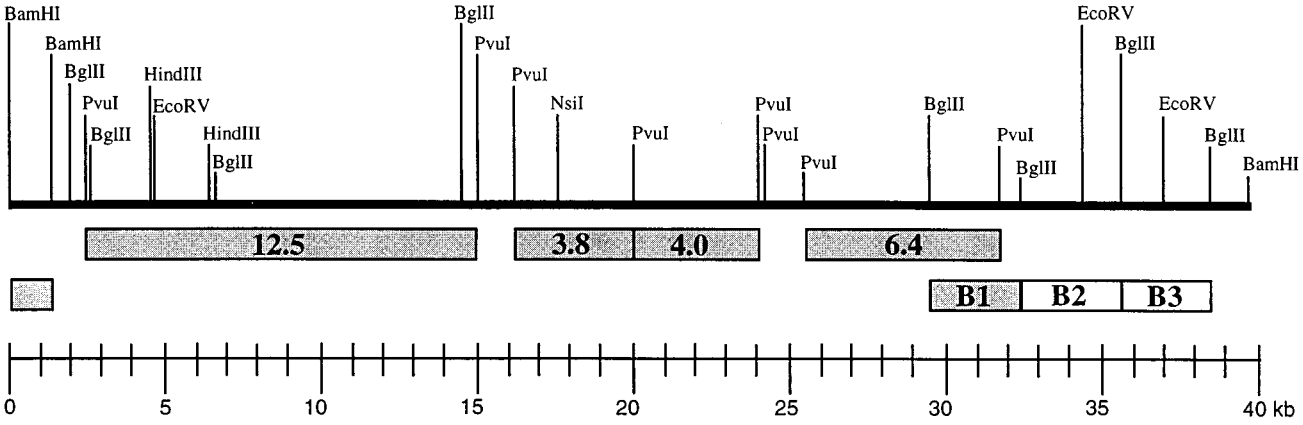

FIG. 3. Restriction map of the genomic region of *S. cellulosum* So ce26 cloned in plasmid p98/1. The shaded bars below the restriction map indicate the four *Pvu*I fragments (upper row) and a *Bam*HI fragment and one *Bgl*II fragment (B1) (lower row) demonstrated by gene disruption to be involved in soraphen biosynthesis. The size of each of the *Pvu*I fragments in kilobase pairs is indicated inside each bar. The open bars (B2 and B3) indicate the two *Bgl*II fragments that resulted in no disturbance of soraphen biosynthesis upon gene disruption. Additional *Pvu*I sites between the 6.4-kb *Pvu*I fragment and the right end of the cloned region were not mapped, and all *Bam*HI sites are not shown. A scale is shown below the map.

cleotide sequence is shown in Fig. 4. It consists of 6,373 bp and has a high G+C content (70.6%), which is typical for myxobacteria. Comparison of this sequence with the available nucleotide sequences from the GenBank/EMBL data bank by using the program FASTA (11) revealed significant similarities with PKS genes from actinomycetes. The highest score was obtained with the *eryA* locus of *Saccharopolyspora erythraea* (10), coding for the synthesis of 6-deoxyerythronolide B (the precursor of erythromycin), and with genes from *Streptomyces antibioticus* coding for the synthesis of the polyketide antibiotic oleandomycin (30). The degree of homology between these sequences was analyzed with the program BESTFIT (11), using a local homology algorithm. In the *eryA* locus several regions of about 1.8 kb with 60 to 63% identity were found, and in the oleandomycin PKS gene region, encoding ORFs 5 and 6, two regions of about 1 kb with identities of 57 to 59% were found. The *graI* region, which was used as a probe to identify clone p98/1, contains only shorter stretches of about 200 bp with 60 to 65% identity.

The nucleotide sequence of the 6.4-kb *Pvu*I fragment was analyzed for ORFs, using the program FRAMES and CODONPREFERENCE (11). This analysis revealed clearly that the entire fragment represents one protein-coding region which is part of a larger ORF, with a start point and an endpoint outside the 6.4-kb *Pvu*I fragment. The frequency of G+C base pairs in the third position of each codon in the presumed ORF from the 6.4-kb fragment is 83%, which is typical of an ORF from an organism with a high overall moles percent G+C content. In addition, the codon preference value for this ORF was determined to be 0.98, compared with a value of 0.76 for a random sequence from a GC-rich sequence, further supporting the existence of a single ORF. Within the deduced amino acid sequence of this large ORF, several regions that have a high degree of similarity to the PKS biosynthetic domains of the *eryA* genes of *Saccharopolyspora erythraea* (9) were identified. These include (listed by the first and last amino acids in each domain with amino acid numbers from Fig. 4), a ketoreductase (KR) domain (G-52 to G-231), an acyl carrier protein (ACP) domain (L-337 to L-410), a β-ketoacyl synthase (KS) domain (E-434 to L-857), an acyltransferase (AT) domain (F-970 to V-1302), a dehydratase (DH) domain (A-1324 to R-1503), a second KR domain (G-1819 to R-1999), and the N-terminal portion of a second ACP domain (L-2065 to end). Further analysis of the 424 amino acids of the KS domain indicated that it has 53% identity with the *eryAIII* KS domain in module 6 (Fig. 5). When conservative amino acid replacements are considered, the level of similarity is 70.2%. This high degree of amino acid similarity demonstrates that this portion of the encoded protein is likely to be a KS domain. The results of similar comparisons between the other putative PKS domains in the 6.4-kb *Pvu*I fragment and the corresponding domains in module 6 of *eryAIII* also revealed high degrees of identity and similarity. The putative KR and ACP domains located 5′ to the KS domain in the 6.4-kb *Pvu*I fragment were found to be 56.2 and 41.9% identical and 69.1 and 60.8% similar, respectively, to the KR and ACP domains of *eryAIII*. Similarly, the putative AT and KR domains located 3′ to the KS domain in the 6.4-kb *Pvu*I fragment had 48.0 and 48.3% identity and 64.8 and 67.6% similarity, respectively, to the corresponding domains of *eryAIII*. The only DH domain present in the *eryA* gene system is in module 4 of *eryAII*. Comparison of the deduced amino acid sequence of the putative DH domain in the 6.4-kb *Pvu*I fragment with the *eryAII* DH domain revealed 43.6% identity and 63.0% similarity between the two. The high level of amino acid similarity between the putative domains identified in the 6.4-kb *Pvu*I fragment

TABLE 1. Soraphen A production by *S. cellulosum* So ce26 transconjugants obtained by plasmid integration through homologous integration of *Pvu*I fragments subcloned from p98/1

Strain and *Pvu*I fragment	No. of individual colonies tested	Soraphen production (mg/liter)[a]
Transconjugants		
12.5 kb	20	0–4
3.8 kb	12	0–5
4.0 kb	12	0–4
6.4 kb	35	0–3
3.5 kb[b]	10	150–180
SJ3[c]	10	180–200
Revertant[d]	1	120

[a] HPLC values of up to 5 mg/liter are within baseline and therefore are not significant.
[b] Control plasmid pSJB55 with random 3.5-kb *Pvu*I fragment.
[c] Parental strain without gene disruption.
[d] Revertant colony from transconjugants blocked in soraphen A biosynthesis.

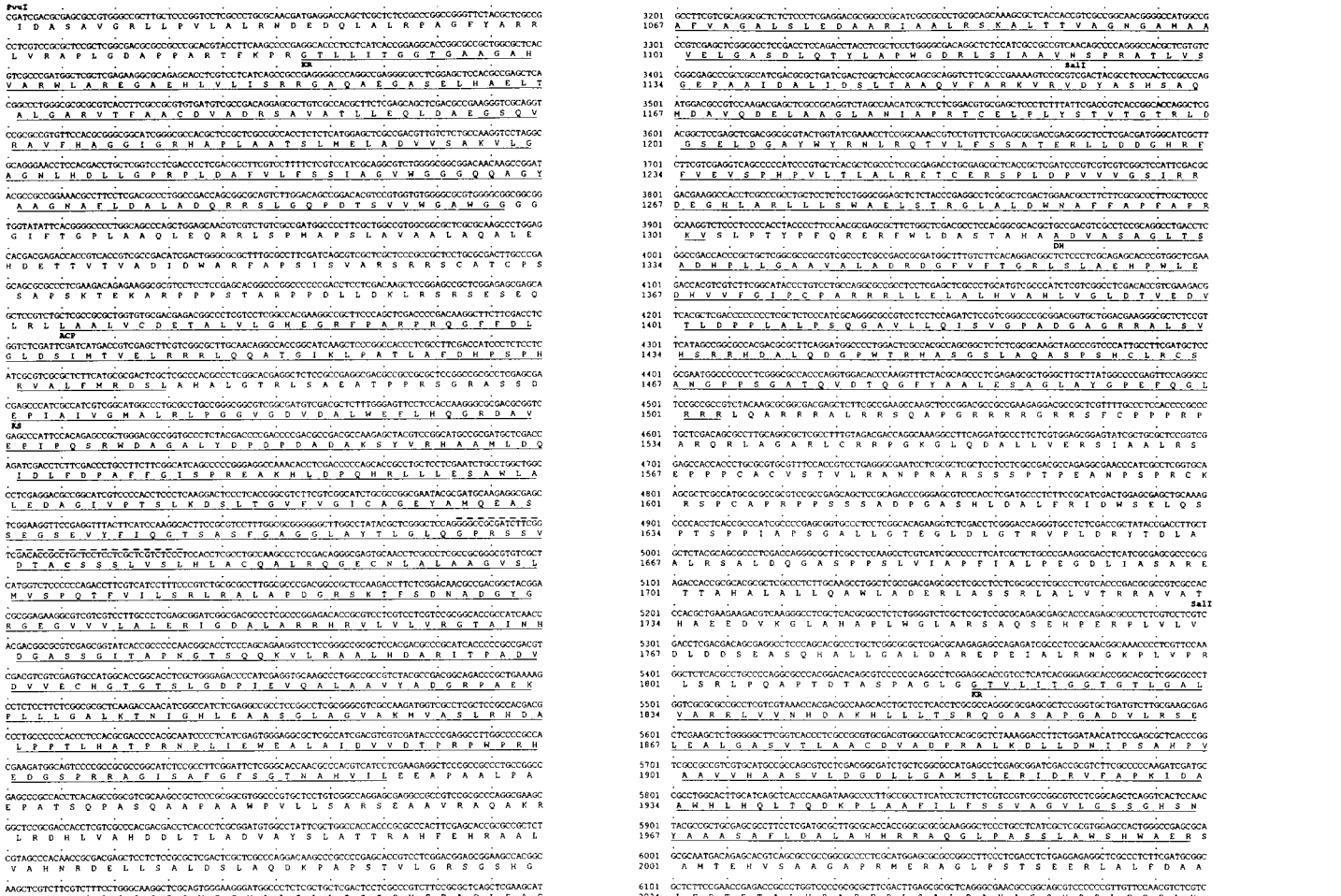

FIG. 4. Nucleotide sequence of the 6.4-kb *Pvu*I fragment from plasmid p98/1 and the deduced amino acid sequence for the large ORF contained within. The approximate locations of the catalytic domains are indicated by underlining the amino acid sequence of each domain and designating its identity at the beginning. The nucleotide sequence of the presumed active site of the KS domain is overlined with a broken line, and the catalytic site C residue is in boldface. The positions of the two *Sal*I sites that define the ends of the 1.8-kb fragment found to be homologous to the *graI* gene fragment of *Streptomyces violaceoruber* are indicated.

and similar domains in the *eryA* gene system strongly suggest that the domains in the former have functions similar to those of the corresponding domains in the latter.

Further confirmation of the function of the KS domain was derived from the discovery of a region within it that is highly similar to the sequence around the presumed active-site cysteine in other KSs (overlined with a broken line in Fig. 4). A comparison of the amino acid sequence of this region with the sequence of the presumed active-site domains from KSs from several diverse sources is shown in Fig. 6.

DISCUSSION

We have demonstrated by Southern hybridization that there is homology between genomic DNA from the gram-negative myxobacterium *S. cellulosum* So ce26, which produces the polyketide soraphen A, and *Streptomyces violaceoruber* genes coding for components of granaticin PKS (*graI*) (28). Furthermore, we have utilized this homology to identify and isolate a genomic clone containing DNA from *S. cellulosum* So ce26 that contains genes that have a role in the biosynthesis of soraphen A. This clone, designated p98/1, was determined to contain approximately 40 kb of genomic DNA. In order to define the functional soraphen A biosynthetic regions in this clone, we utilized a gene disruption system for this organism (18) to demonstrate that 32 kb of the cloned DNA in p98/1 has a role in soraphen A biosynthesis. This analysis indicates that p98/1 contains the right boundary (relative to Fig. 3) of a large genetic region that directs the synthesis of soraphen A but that it does not contain the left boundary.

A 6.4-kb *Pvu*I fragment with homology to the *graI* gene fragment was sequenced (Fig. 4). Analysis of this sequence revealed the presence of a single, continuous ORF beginning and ending outside the fragment (Fig. 4). Within the deduced amino acid sequence, several domains with high degrees of homology to known PKS biosynthetic domains of the *eryA* genes of *Saccharopolyspora erythraea* were identified. In the order of their occurrence, these were the KR, ACP, KS, AT, DH, second KR, and second ACP domains. Analysis of these domains indicated high levels of identity with the corresponding domains of *eryAIII* from *Saccharopolyspora erythraea* (Fig. 5), which is involved in the biosynthesis of 6-deoxyerythrono-

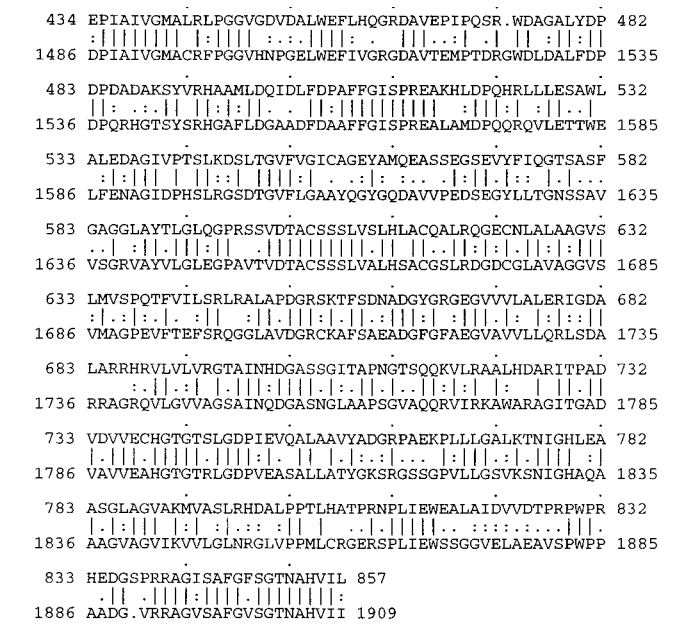

FIG. 5. Comparison of the deduced amino acid sequence from the KS-homologous domain from the 6.4-kb *Pvu*I fragment of plasmid p98/1 (top sequence) with the KS domain from the synthase unit 6 of the *eryAIII* gene from *Saccharopolyspora erythreae* (bottom sequence). The degree of identity between the two sequences is indicated as follows: complete identity, vertical bar; conservative replacement, two dots; neutral replacement, one dot; and no similarity, no symbol. Numbers at the left and right margins indicate the amino acid positions at the beginning and end of each row, respectively, for the *S. cellulosum* PKS (first amino acid in the 6.4-kb *Pvu*I fragment is assigned the number 1; see Fig. 4) and for *eryAIII* (GenBank accession number M63677).

lide B, the precursor in the synthesis of the polyketide antibiotic erythromycin. Additional evidence supporting the conclusion that the KS-like region in the 6.4-kb *Pvu*I fragment encodes a KS domain is the presence of an amino acid sequence within it that is highly similar to the presumed active-site domain that is present in all KSs examined to date (Fig. 6). The presumed active-site cysteine residue in this region is present in a location relative to the other amino acid residues that is identical to that in other KSs.

So. cellulosum PKS	G P R S S V D T A **C** S S S L V S L
eryA ORF3 module 6	G P A V T V D T A **C** S S S L V A L
eryA ORF3 module 5	G P A M T V D T A **C** S S G L T A L
S. antibioticus PKS	G P A V T V D T A **C** S S S L V A L
gra ORF1	G P V T M V S D G **C** T S G L D S V
tcm ORF1	G P V T V V S T G **C** T S G L D A V
whiE ORFIII	G P V Q T V S T G **C** T S G L D A V
6-MSAS	G P S T A V D A A **C** A S S L V A I
FAS *E.coli fabB*	G V N T S I S S A **C** A T S A H C I
FAS rat	G P S I A L D T A **C** S S S L L A L

FIG. 6. Alignments of segments of the deduced protein sequence around the presumed active-site cysteine (boldface) in KSs, including soraphen A PKS of *S. cellulosum* (this paper), 6-deoxyerythronolide B PKS (*ery*) (7, 9), PKS of *Streptomyces antibioticus* (29), granaticin PKS (*gra*) (27), tetracenomycin C PKS (*tcm*) (5), spore pigment (*whiE*) (8), 6-methylsalicylic acid synthase (6-MSAS) (3), *E. coli* FAS (*fabB*) (20), and rat FAS (1).

The synthesis of polyketides is known to be similar to the process of fatty acid synthesis (17). In bacteria and higher plants, fatty acid synthesis is catalyzed by separate polypeptides which carry out the condensation event (KS) and the ketoreduction, dehydration, and enoyl reduction of the β-carbonyl group on each new two-carbon extender unit. This system is the type II fatty acid synthase (FAS). However, in vertebrates and fungi, fatty acid synthesis is catalyzed by distinct domains on large, multifunctional polypeptides, and this system has been termed the type I FAS (17). The same two types of genetic organization have evolved for the synthesis of polyketides. However, in contrast to the case with fatty acid synthesis, the structural organization for PKS genes seems to depend not on the life form of the producer organism but rather on the chemical structure of the polyketide synthesized. All polyketide biosynthetic systems for macrolide antibiotics from bacteria that have been studied to date are similar to the type I FAS, since they also use large multifunctional polypeptides (17). For example, in *Saccharopolyspora erythraea*, the biochemical domains for the synthesis of 6-deoxyerythronolide B are organized on three large polypeptides, typical of type I PKS (10). Each contains two synthase units, or modules, each of which contains the domains necessary for the addition and specific processing of a single two-carbon extender unit.

The results from this study indicate that the soraphen A biosynthetic genes have a modular organization similar to that of the *eryA* genes and other modular polyketide biosynthetic systems of the type I class. This is supported by the presence of an ORF larger than 6.4 kb in size that has been shown to contain domains that are typically found in PKS biosynthetic genes, including a KS domain. In the *eryA* gene system, the domains within each biosynthetic module are organized in a conserved order: KS–AT–β-carbonyl group-processing domains (if required) (KR, DH, and/or enoyl reductase)–ACP (10). The biosynthetic domains within the 6.4-kb *Pvu*I fragment were found to have a similar order (KR-ACP-KS-AT-DH-KR-ACP). The first two domains, KR and ACP, likely represent the C terminus of one biosynthetic module, while the subsequent KS, AT, DH, KR, and ACP domains make up a second, complete module.

Examination of the structure of soraphen A and consideration of the pattern of incorporation of building units (Fig. 1) indicate the potential for eight biosynthetic modules. The β-carbonyl-processing domains of the complete biosynthetic module in the 6.4-kb *Pvu*I fragment, the DH and KR domains, should result in the incorporation of an enoyl bond in the soraphen A ring structure. There is one enoyl bond between carbons 8 and 9 of the lactone ring of soraphen A, and therefore it is likely that the complete biosynthetic module in the 6.4-kb *Pvu*I fragment is responsible for processing these carbons derived from acetate. From the structure of soraphen A it appears that there are three O-methylation events required in the synthesis of soraphen A. The location and characteristics of the other biosynthetic modules and the O-methyltransferase genes have yet to be determined.

S. cellulosum So ce26 is in the order *Myxobacterales*, the gliding bacteria, which are unicellular, gram-negative chemoorganotrophs. Bacteria of the order *Actinomycetales*, on the other hand, are mycelial, gram-positive heterotrophs and are not closely related to the myxobacteria. However, the striking similarity of the *sorA* genes of *S. cellulosum* at the nucleotide and organizational levels to the type I PKS biosynthetic genes from the actinomycetes suggests that there has been a genetic exchange between these diverse bacteria that has resulted in the dispersal of these genes in nature. One genetic feature shared by the myxobacteria and actinomycetes is a high moles

percent G+C content (67 to 71 and 69-73 mol%, respectively). This would facilitate the expression of genetic material exchanged between these two groups of bacteria. An intriguing question is how the basic type I PKS genetic system evolves in one organism, after transfer from another, to direct the biosynthesis of a new metabolite with unique structural and biological properties.

As far as we are aware, this work represents the first molecular analysis of genes from a myxobacterium that have a role in the biosynthesis of a polyketide antibiotic and thus offers the first opportunity to compare these systems with the better characterized polyketide biosynthetic systems from the actinomycetes. Further work is needed to locate the left border of the soraphen A biosynthetic gene region and to characterize the complete gene region.

ACKNOWLEDGMENTS

We are grateful to J. Heim, S. Jaoua, T. Kieser, H. Reichenbach, J. Ryals, and P. Urech for helpful discussions. We thank D. A. Hopwood for providing *graI* DNA fragments. We also thank J. DiMaio and C. Boyce for assistance in DNA sequencing and K. Hess and H. B. Jenny for performing the HPLC analysis.

REFERENCES

1. **Amy, C. M., A. Witkowski, J. Naggert, B. Williams, Z. Randhawa, and S. Smith.** 1989. Molecular cloning and sequencing of cDNAs encoding the entire rat fatty acid synthase. Proc. Natl. Acad. Sci. USA **86:**3114–3118.
2. **Arrowsmith, T. J., F. Malpartida, D. H. Sherman, A. Birch, D. A. Hopwood, and J. A. Robinson.** 1992. Characterisation of *act*I-homologous DNA encoding polyketide synthase genes from the monensin producer *Streptomyces cinnamonensis*. Mol. Gen. Genet. **234:**254–264.
3. **Beck, J., S. Ripka, A. Siegner, E. Schiltz, and E. Schweizer.** 1990. The multifunctional 6-methylsalicylic acid synthase gene of *Penicillium patulum*. Its gene structure relative to that of other polyketide synthases. Eur. J. Biochem. **192:**487–498.

3a. **Bedorf, N., K. Gerth, H. Reichenbach, and G. Höfle.** Personal communication.

4. **Bedorf, N., D. Schonburg, K. Gerth, H. Reichenbach, and G. Höfle.** 1993. Isolation and structure elucidation of soraphen A_{1a}, a novel antifungal macrolide from *Sorangium cellulosum*. Liebigs Ann. Chem. **9:**1017–1021.
5. **Bibb, M. J., S. Biró, H. Motamedi, J. F. Collins, and C. R. Hutchinson.** 1989. Analysis of the nucleotide sequence of the *Streptomyces glaucescens tcmI* genes provides key information about the enzymology of polyketide antibiotic biosynthesis. EMBO J. **8:**2727–2736.
6. **Bolivar, F., R. L. Rodriguez, P. J. Greene, M. C. Beltach, H. L. Heynecker, H. W. Boyer, J. H. Cross, and S. Falkow.** 1977. Construction and characterization of new multipurpose cloning vehicles. II. A multipurpose cloning system. Gene **2:**95–113.
7. **Cortes, J., S. F. Haydock, G. A. Roberts, D. J. Bevitt, and P. F. Leadlay.** 1990. An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of *Saccharopolyspora erythraea*. Nature (London) **348:**176–178.
8. **Davis, N. K., and K. F. Chater.** 1990. Spore colour in *Streptomyces coelicolor* A3(2) involves the developmentally regulated synthesis of a compound biosynthetically related to polyketide antibiotics. Mol. Microbiol. **4:**1679–1691.
9. **Donadio, S., and L. Katz.** 1992. Organization of the enzymatic domains in the multifunctional polyketide synthase involved in erythromycin formation in *Saccharopolyspora erythraea*. Gene **111:**51–60.
10. **Donadio, S., M. J. Staver, J. B. McAlpine, S. J. Swanson, and L. Katz.** 1991. Modular organization of genes required for complex polyketide biosynthesis. Science **252:**675–679.
11. **Genetics Computer Group.** 1991. Sequence analysis software package, version 7. University of Wisconsin Genetics Computer Group, Madison.
12. **Gerth, K., N. Bedorf, H. Irschik, G. Höfle, and H. Reichenbach.** 1994. The soraphens: a family of novel antifungal compounds from *Sorangium cellulosum* (Myxobacteria). J. Antibiot. **47:**23–31.
13. **Hedges, R. W., and M. Matthew.** 1979. Acquisition by *Escherichia coli* of plasmid-borne β-lactamases normally confined to *Pseudomonas* spp. Plasmid **2:**269–278.
14. **Hohn, B., and J. Collins.** 1980. A small cosmid for efficient cloning of large DNA fragments. Gene **11:**291–298.
15. **Hopwood, D. A., and C. Khosla.** 1992. Genes of polyketide secondary metabolic pathways in microorganisms and plants. Ciba Found. Symp. **171:**88–112.
16. **Hopwood, D. A., C. Khosla, D. H. Sherman, M. J. Bibb, S. Ebert-Khosla, E.-S. Kim, R. McDaniel, W. P. Revill, R. Torres, and T.-W. Yu.** 1993. Towards an understanding of the programming of aromatic polyketide synthases: a genetic-driven approach, p. 267–275. *In* G. D. Hegeman, R. H. Baltz, and P. L. Skatrud (ed.), Genetics and molecular biology of industrial microorganisms. American Society for Microbiology, Washington, D.C.
17. **Hopwood, D. A., and D. H. Sherman.** 1990. Molecular genetics of polyketides and its comparison to fatty acid biosynthesis. Annu. Rev. Genet. **24:**37–66.
18. **Jaoua, S., S. Neff, and T. Schupp.** 1992. Transfer of mobilizable plasmids to *Sorangium cellulosum* and evidence for their integration into the chromosome. Plasmid **28:**157–165.
19. **Kakinuma, S., Y. Takada, H. Ikeda, H. Tanaka, S. Omura, and D. A. Hopwood.** 1991. Cloning of large DNA fragments, which hybridize with actinorhodin biosynthesis genes, from kalafungin and nanaomycin A methyl ester producers and identification of genes for kalafungin biosynthesis of the kalafungin producer. J. Antibiot. **44:**995–1005.
20. **Kauppinen, S., M. Siggaard-Anderen, and P. von Wettstein-Knowles.** 1988. β-Ketoacyl-ACP synthase I of *Escherichia coli*: nucleotide sequence of the *fabB* gene and identification of the cerulenin binding residue. Carlsberg Res. Commun. **53:**357–370.
21. **Malpartida, F., S. E. Hallam, H. M. Kieser, H. Motamedi, C. R. Hutchinson, M. J. Butler, D. A. Sugden, M. Warren, C. McKillop, C. R. Bailey, G. O. Humphreys, and D. A. Hopwood.** 1987. Homology between *Streptomyces* genes coding for synthesis of different polyketides used to clone antibiotic biosynthesis genes. Nature (London) **325:**818–821.
22. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
23. **Murray, N. E., W. J. Brammar, and K. Murray.** 1977. Lambdoid phages that simplify the recovery of in vitro recombinants. Mol. Gen. Genet. **150:**53–61.
24. **Norrander, J., T. Kempe, and J. Messing.** 1983. Construction of improved M13 vectors. Gene **26:**101–106.
25. **Pridzun, L.** 1992. Untersuchungen zum Wirkungsmechanismus von Soraphen A. Doctoral thesis. Technische Universität Braunschweig, Braunschweig, Federal Republic of Germany.
26. **Reichenbach, H., K. Gerth, H. Irschik, B. Kunze, and G. Höfle.** 1988. Myxobacteria: a source of new antibiotics. Trends Biotechnol. **6:**115–121.
27. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
28. **Sherman, D. H., F. Malpartida, M. J. Bibb, H. M. Kieser, and D. A. Hopwood.** 1989. Structure and deduced function of the granaticin-producing polyketide synthase gene cluster of *Streptomyces violaceoruber* Tü22. EMBO J. **8:**2717–2725.
29. **Simon, R., U. Priefer, and A. Pühler.** 1983. A broad host range mobilization system for in vivo genetic engineering: transposon mutagenesis in gram negative bacteria. Bio/Technology **1983:**784–790.
30. **Swan, D. G., A. M. Rodriguez, C. Vilches, C. Mendez, and J. A. Salas.** 1994. Characterisation of a *Streptomyces antibioticus* gene encoding a type I polyketide synthase which has an unusual coding sequence. Mol. Gen. Genet. **242:**358–362.
31. **Trowitzsch, W., V. Wray, K. Gerth, and G. Höfle.** 1982. Structure of myxovirescin A, a new macrocyclic antibiotic from gliding bacteria. J. Chem. Commun. **1982:**1340–1342.
32. **Vahlensieck, H. F., L. Pridzun, H. Reichenbach, and A. Hinnen.** 1994. Identification of the yeast Acc 1 gene product (acetyl-CoA carboxylase) as the target of the polyketide fungicide soraphen A. Curr. Genet. **25:**95–100.
33. **Varon, M., N. Fuchs, M. Monosov, S. Tolchinsky, and E. Rosenberg.** 1992. Mutation and mapping of genes involved in production of the antibiotic TA in *Myxococcus xanthus*. Antimicrob. Agents Chemother. **36:**2316–2321.
34. **Yu, T.-W., M. J. Bibb, W. P. Revill, and D. A. Hopwood.** 1994. Cloning, sequencing, and analysis of the griseusin polyketide synthase gene cluster from *Streptomyces griseus*. J. Bacteriol. **176:**2627–2634.