JOURNAL OF BACTERIOLOGY, July 1991, p. 4454–4463
0021-9193/91/144454-10$02.00/0

Vol. 173, No. 14

Nucleotide Sequence Analysis Reveals Linked *N*-Acetyl Hydrolase, Thioesterase, Transport, and Regulatory Genes Encoded by the Bialaphos Biosynthetic Gene Cluster of *Streptomyces hygroscopicus*

A. RAIBAUD,[1] M. ZALACAIN,[1] T. G. HOLT,[1]† R. TIZARD,[2] AND C. J. THOMPSON[1]*

Institut Pasteur, 28 rue du Docteur Roux, 75015 Paris, France,[1] and Biogen Research Corp., Cambridge, Massachusetts 02142[2]

Received 16 January 1991/Accepted 10 May 1991

Nucleotide sequence analysis of a 5,000-bp region of the bialaphos antibiotic production (*bap*) gene cluster defined five open reading frames (ORFs) which predicted structural genes in the order *bah*, ORF1, ORF2, and ORF3 followed by the regulatory gene, *brpA* (H. Anzai, T. Murakami, S. Imai, A. Satoh, K. Nagaoka, and C. J. Thompson, J. Bacteriol. 169:3482–3488, 1987). The four structural genes were translationally coupled and apparently cotranscribed from an undefined promoter(s) under the positive control of the *brpA* gene product. S1 mapping experiments indicated that *brpA* was transcribed by two promoters ($brpAp_1$ and $brpAp_2$) which initiate transcription 150 and 157 bp upstream of *brpA* within an intergenic region and at least one promoter further upstream within the *bap* gene cluster ($brpAp_3$). All three transcripts were present at low levels during exponential growth and increased just before the stationary phase. The levels of the $brpAp_3$ band continued to increase at the onset of stationary phase, whereas $brpAp_1$- and $brpAp_2$-protected fragments showed no further change. BrpA contained a possible helix-turn-helix motif at its C terminus which was similar to the C-terminal regulatory motif found in the receiver component of a family of two-component transcriptional activator proteins. This motif was not associated with the N-terminal domain conserved in other members of the family. The structural gene cluster sequenced began with *bah*, encoding a bialaphos acetylhydrolase which removes the *N*-acetyl group from bialaphos as one of the final steps in the biosynthetic pathway. The observation that Bah was similar to a rat and to a bacterial (*Acinetobacter calcoaceticus*) lipase probably reflects the fact that the ester bonds of triglycerides and the amide bond linking acetate to phosphinothricin are similar and hydrolysis is catalyzed by structurally related enzymes. This was followed by two regions encoding ORF1 and ORF2 which were similar to each other (48% nucleotide identity, 31% amino acid identity), as well as to GrsT, a protein encoded by a gene located adjacent to gramicidin S synthetase in *Bacillus brevis*, and to vertebrate (mallard duck and rat) thioesterases. The amino acid sequence and hydrophobicity profile of ORF3 indicated that it was related to a family of membrane transport proteins. It was strikingly similar to the citrate uptake protein encoded by the transposon Tn*3411*.

Bialaphos, a linear tripeptide produced as a secondary metabolite by *Streptomyces hygroscopicus*, is composed of two L-alanine residues and the glutamate analog phosphinothricin (PT) (4, 29). After cleavage of bialaphos by intracellular nonspecific peptidases, PT acts as an inhibitor of glutamine synthetase (4, 29). For this reason, bialaphos has herbicide and antibiotic activities. The biosynthetic pathway has been well characterized by analyzing intermediates accumulated in cultures of blocked mutants (19). Primary metabolic substrates are converted to the demethylated form of PT, which is then condensed with two alanine residues. Although the mechanism of this condensation has not been studied biochemically, mutants blocked in this step have been isolated (18). Genes which restore productivity to mutants having defects in either demethylated PT synthesis or peptide bond formation (bialaphos antibiotic production [*bap*]) have been cloned and localized to a 35-kb segment of the genome (Fig. 1) (18, 41). This region includes a gene which confers bialaphos resistance (*bar*) encoding a PT or demethylated PT acetyltransferase (58). Mutants unable to export bialaphos have not been reported.

Expression of *bap* genes begins at the approach of stationary phase (22) and requires the *brpA* DNA sequence (2); the critical role of *brpA* was demonstrated in a *brpA* mutant (NP57). In NP57 there was no detectable transcription of most *bap* genes (2), and analysis by sodium dodecyl sulfate-polyacrylamide gel electrophoresis revealed a dramatic difference in the profile of total cellular proteins relative to its parent (2). Two-dimensional gel electrophoretic analysis of a *brpA* mutant showed that it was deficient in the synthesis of at least 27 proteins implicated in the biosynthesis of bialaphos (22).

Here we report the nucleotide sequence of a portion of the *bap* gene cluster which identifies gene products whose similarity to lipases, thioesterases, membrane transport proteins, and regulatory proteins suggests roles in the activation and terminal steps of bialaphos biosynthesis.

MATERIALS AND METHODS

Bacterial strains and growth conditions. The bialaphos-producing strain *S. hygroscopicus* ATCC 21705 (29) and *S. hygroscopicus* HP5-29, a bialaphos-over-producing derivative (2), were obtained from the Meiji Seika Culture Collection. Seed cultures were inoculated by placing a single colony of *S. hygroscopicus* grown on NE agar (41) into S1 medium (43) (10 ml in a 30-ml culture tube equipped with a stainless-steel spring to break up mycelial clumps) and

* Corresponding author.
† Present address: Merck Sharp and Dohme Research Laboratories, Rahway, NJ 07065.

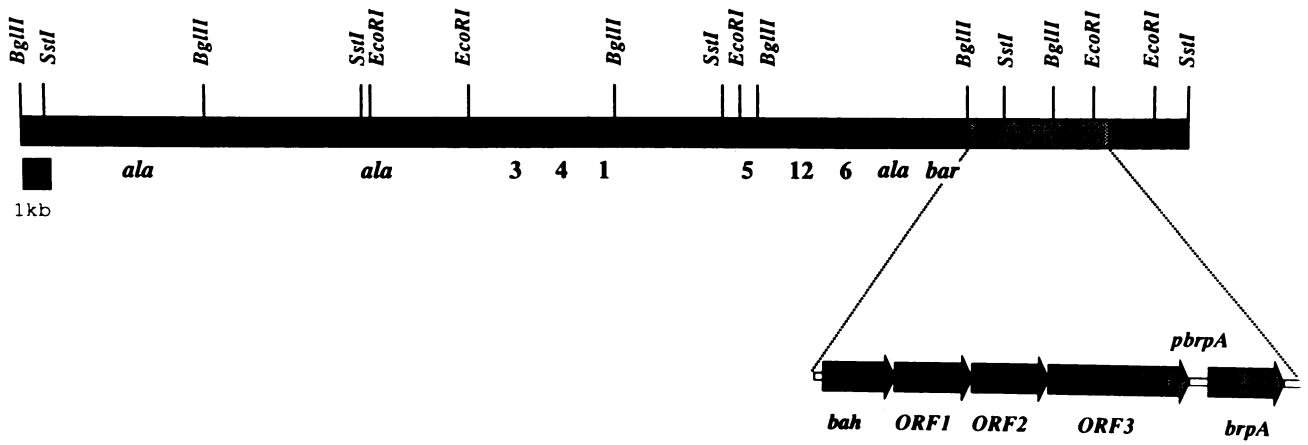


FIG. 1. Bialaphos biosynthetic gene cluster. The locations of the *bap* genes which catalyze steps in the biosynthetic pathway (indicated by numbers as defined by Hidaka et al. [20]) were mapped by Murakami et al. (41) and Hara et al. (18). *bar*, bialaphos resistance gene which catalyzes step 10 in the pathway (31, 58); *ala*, genes involved in peptide condensation; *brpA*, bialaphos regulatory protein. Codon usage analysis (shown in Fig. 2) of the DNA sequence indicated that regions *bah*, ORF1, ORF2, ORF3, and *brpA* probably encoded functional structural genes. *pbrpA*, *brpA* promoter region.

incubating on a rotary shaker for 2 days at 30°C. Baffled shake flasks (3 liters) containing 250 ml of S1 medium were inoculated with 5 ml of this seed culture and incubated on a rotary shaker at 30°C.

DNA manipulations. Cosmid pBG9 was obtained from *S. hygroscopicus* ATCC 21705 (41); cosmid pMSB13-3 was obtained from *S. hygroscopicus* HP460 (41). Fragments were subcloned into pUC19 and then into M13 (63) for sequencing reactions.

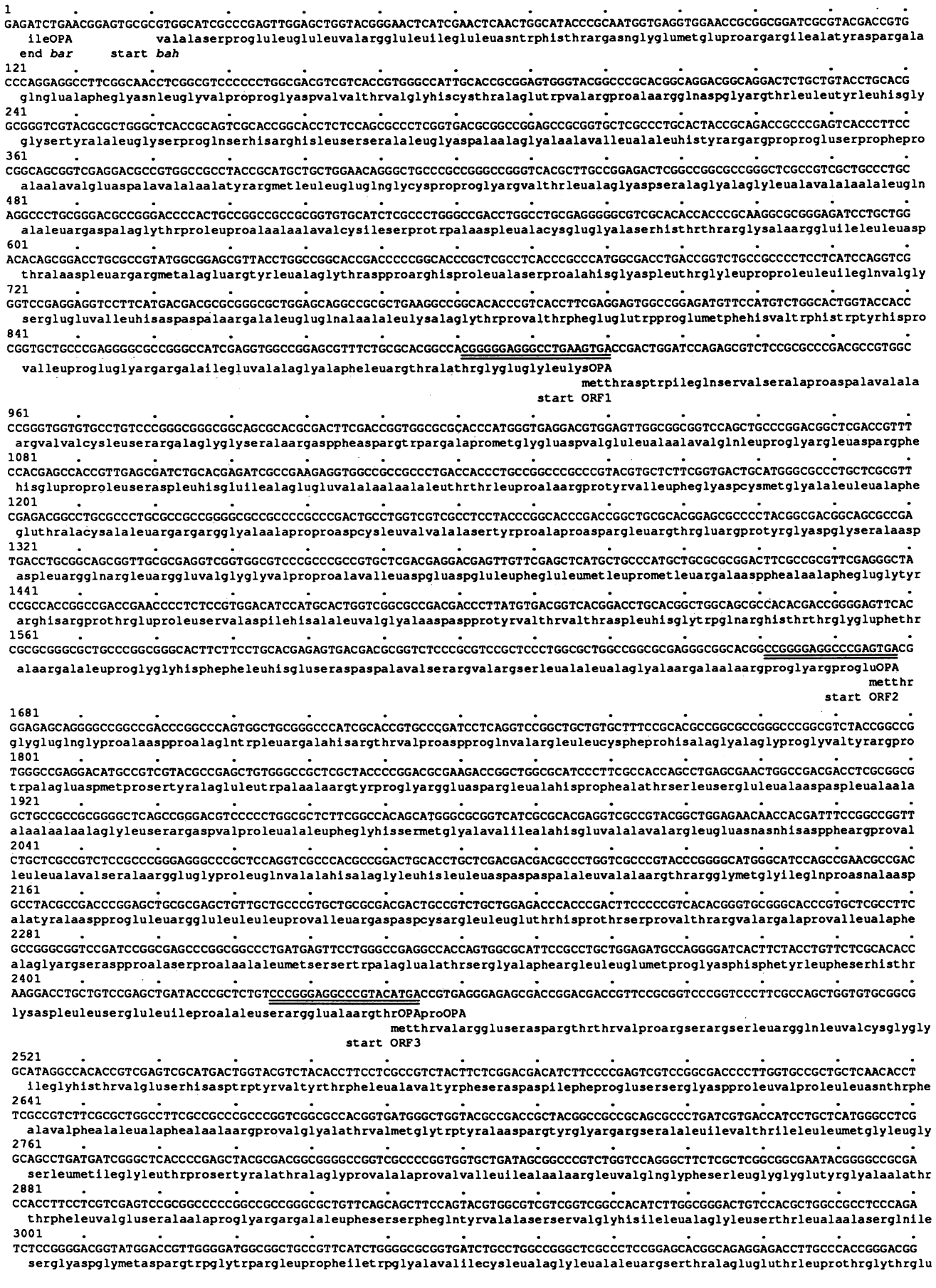
Alkaline lysis methods were used for the isolation of *Streptomyces* (23) and *Escherichia coli* plasmids (39). Chromosomal DNA was isolated from *Streptomyces* spp. as described by Hopwood et al. (23), with the following modifications. Mycelia were grown in S1 medium supplemented with 0.6 M glycine and lysed in a buffer containing 10 mM Tris HCl (pH 8), 10% sucrose, 50 mM EDTA, 4 mg of achromopeptidase (Wako Pure Chemicals, Ltd.) per ml, and 50 mg of lysozyme (Sigma Chemical Co.) per ml.

All plasmid and genomic DNAs were purified by centrifugation in CsCl-ethidium bromide gradients. DNA was sequenced by the dideoxy-nucleotide chain termination method (48), using modified T7 DNA polymerase (Sequenase) as specified by the instructions of the supplier (U.S. Biochemical Corp.). Primers described in this report were prepared with a Cyclone DNA synthesizer (Biosearch, Inc.). The M13 universal primers were routinely used, but in some cases 17-mer primers from the determined sequences were synthesized. Both DNA strands were completely sequenced with dGTP, and most regions were resequenced with dITP to overcome compression problems due to the high G+C content of *Streptomyces* DNA.

Purification of total RNA and S1 nuclease mapping. RNA from *Streptomyces* spp. was extracted with 2× Kirby mixture to disrupt the mycelia used in S1 mapping experiments (23). To prepare the probe, an *Sph*I-*Bam*HI fragment containing *brpA* and the region upstream was subcloned from pMSB13-3 (42) into M13mp19. An oligonucleotide primer (5′-AGCGCGGACTCCAGGGCTTCCAGCAGCCGG-3′) representing a sequence within *brpA* (opposite strand of nucleotides 4145 to 4174 in Fig. 2) was 5′-endlabeled with [γ-^{32}P]ATP, hybridized to an *Sph*I-*Bam*HI (nucleotides 3786 to 4262 in Fig. 2) single-stranded template, and extended with Sequenase. A single-stranded probe was released from the template by cleaving the *Hin*dIII or *Nar*I site in M13 and heating for 10 min at 85°C in 40% formamide–5 mM EDTA, purified on a 5% polyacrylamide–8 M urea gel, and eluted overnight at 42°C in 0.3 M sodium acetate. The probe was coprecipitated with 40 μg of RNA, resuspended in 20 μl of hybridization buffer [50% formamide, 1 mM EDTA, 0.4 M NaCl in 40 mM PIPES [piperazine-*N*,*N*′-bis(2-ethanesulfonic acid)]-HCl, pH 6.4], heated for 10 min at 85°C, and allowed to hybridize overnight at 42°C. After the mix was digested with 400 U of S1 (Bethesda Research Laboratories) in 300 μl of S1 buffer (23) for 45 min at 37°C, the reaction was stopped by the addition of 2.5 M ammonium acetate–50 mM EDTA (75 μl) and the nucleic acids were recovered by precipitation in ethanol.

Computer-assisted analyses. A program developed by Staden (56) was used to identify functional open reading frames (ORFs) on the basis of codon usage. A data base representing codon frequency found in previously sequenced streptomycete structural genes (*aph*, *vph*, *mel*, *endoH*, *amy*, *mmy*, *tipA*, *tsr*, *carB*, and *bar* [all references can be found in references 42 and 51]) was used to scan the DNA sequence in windows of 49 codons. The ORFs which were identified were used to screen the PGtrans (version 61) and NBRF (version 26) data bases, using the FastP program and the PAM250 mutational data matrix (44). Each alignment was assessed for its statistical significance by comparing its optimized alignment score with a distribution of scores obtained with 500 randomized sequences having the same amino acid composition to generate a z value (observed score − mean of shuffled scores/standard deviation of shuffled scores). A z value of >10 indicates a significant match (34). Hydrophobicity profiles were determined by the method of Kyte and Doolittle (32), with a window of 11 residues, and DNA Strider software provided by C. Marck. Autoradiograms displaying the results of S1 mapping experiments were quantified with a Molecular Dynamics laser adapted by Protein Databases Inc. for use with software they developed to quantify optical density.

Nucleotide sequence accession number. The nucleotide sequence determined has been submitted to Genbank and has received the accession number M64783.



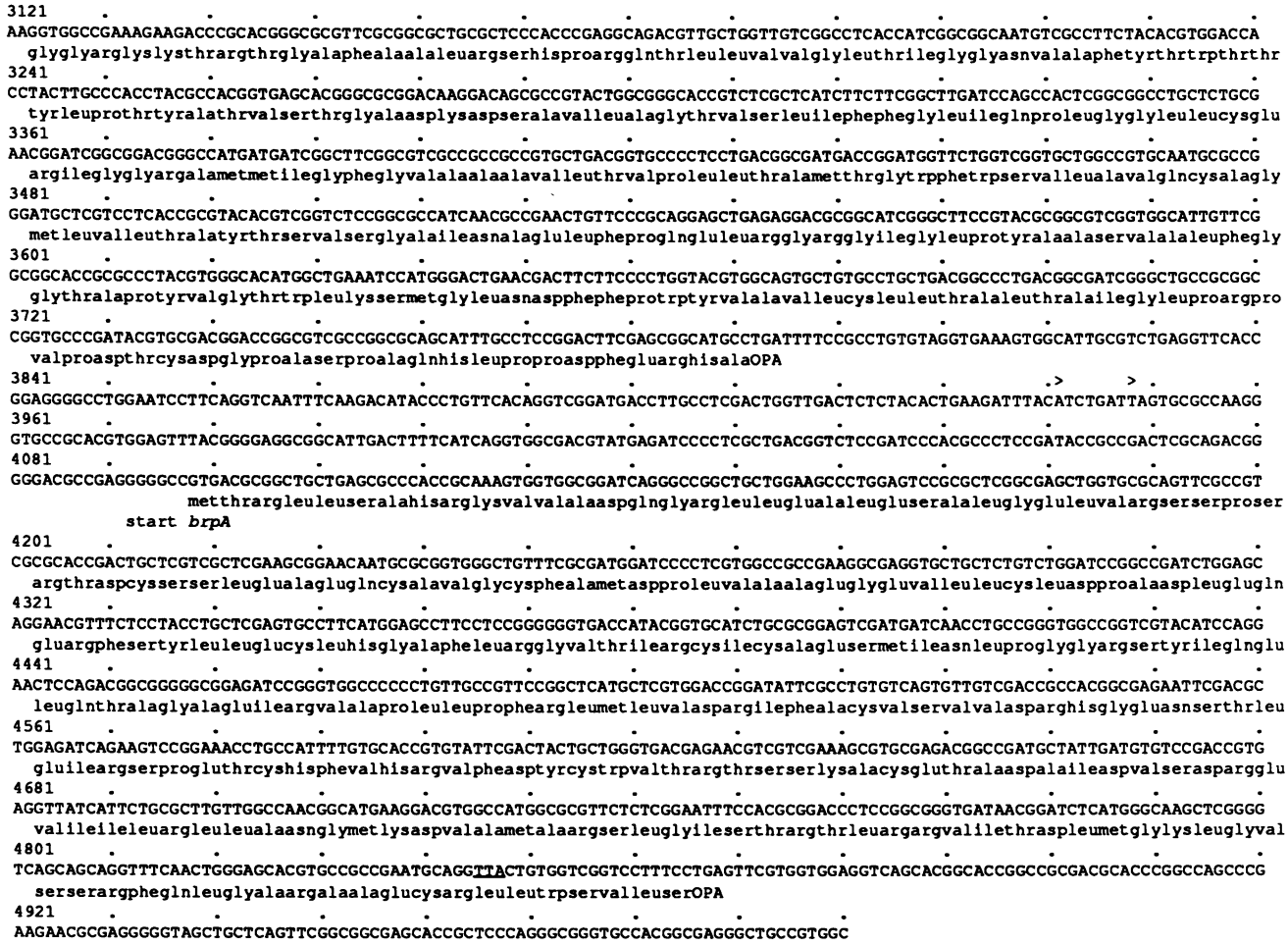


FIG. 2. Nucleotide sequence of a region of the *bap* gene cluster. The sequence begins at the *Bgl*II site, which includes the *bar* stop codon (58) and extends through the *brpA* gene. It includes DNA which was cloned from both wild-type *S. hygroscopicus* ATCC 21705 (nucleotides 1 to 1256 and 3721 to 5000) and a derivative which overproduced bialaphos (HP460; nucleotides 1257 to 4145). Although the restriction endonuclease map for over 36 sites is identical in the *bap* gene cluster (41; unpublished observations) and the sequences of the *brpA* promoter region in both strains are identical, we cannot exclude the possibility that mutations have occurred in the HP460 DNA. Probable coding sequences were predicted by codon preference (Fig. 3). Conserved regions (Table 1) which included presumptive ribosome binding sites are double underlined. The translational start site of *brpA* was ambiguous since start sites predicted by codon usage (indicated in Fig. 2) were near a GTG which was not preceded by conventional ribosome binding sites. The first GTG start codon preceded by sequences similar to other ribosome binding sites was chosen as the preferred translational start site; however, multiple initiation sites cannot be ruled out. The only TTA codon found in the five ORFs was located near the C terminus of *brpA* (single underline). The *brpA* mRNA start sites suggested by S1 mapping (Fig. 6) are indicated (>).

RESULTS AND DISCUSSION

ORFs predicted by the nucleotide sequence. We sequenced 5,000 bp of the *bap* gene cluster (Fig. 1) which spanned the region downstream of the previously reported *bar* gene sequence (58) through *brpA* (Fig. 2). Five ORFs were predicted based on the codon usage found in previously characterized streptomycete structural genes (Fig. 3): *bah*, ORF1, ORF2, ORF3, and *brpA*. A 250-bp noncoding region was detected between ORF3 and *brpA*. Although the expectation that ORF1, ORF2, and ORF3 encode functional proteins was strongly supported by their significant similarity to other known genes in the data banks (described below), mutants are not yet available to confirm their roles in bialaphos biosynthesis.

Translational coupling of *bah*, ORF1, ORF2, and ORF3.

TABLE 1. Conserved nucleic acid sequences in the 5′ region of *bah*, ORF1, ORF2, and ORF3[a]

ORF	Sequence	ORF
bah	C G G G G G A G G G C C T G A A G T G A	ORF1
ORF1	C C G G G G A G G C C C G A G T G A	ORF2
ORF2	C C C G G G A G G C C C G T A C A T G A	ORF3

[a] The presumed translational stop codon is underlined. It is preceded by the translational start codon (bold print) of the following gene.

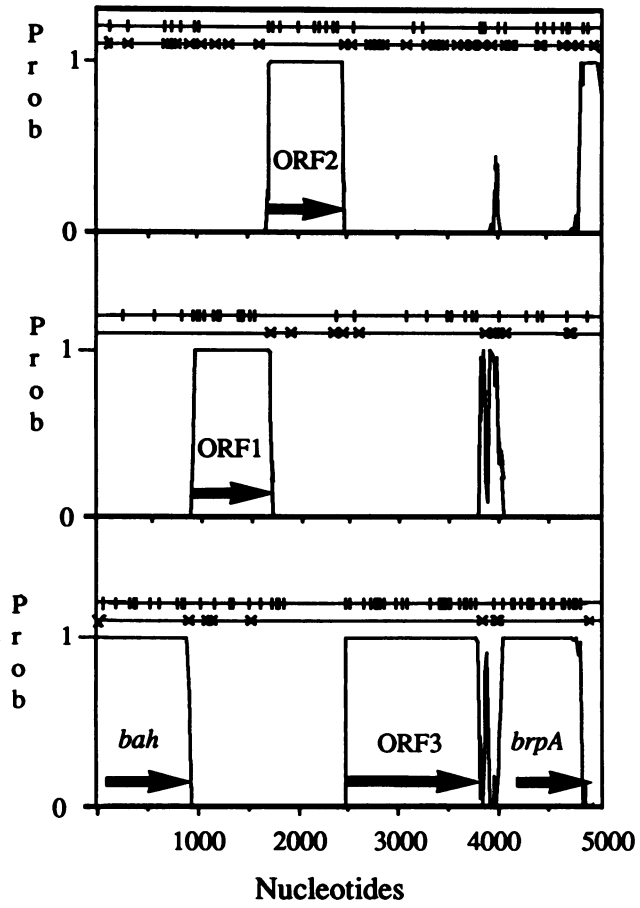


FIG. 3. Predicted protein coding regions. Codon preference analysis, using a program developed by Staden (56) and data base representing codons found in *Streptomyces* structural genes (see Materials and Methods), suggested the coding sequences *bah*, ORF1, ORF2, ORF3, and *brpA*. Translational start (hatch marks) and stop (×) codons are indicated immediately above the codon usage profile.

ORFs *bah*, ORF1, ORF2, and ORF3 were adjacent and overlapped each other such that a TG sequence was used in both the initiation (GTG or ATG) and termination (TGA) triplets. These overlapping regions were part of a conserved 20-bp sequence (Table 1). Although sequences located 5′ to structural genes are usually similar to each other due to their partial complementarity to the 3′ end of the 16S rRNA (54), the conservation of sequence observed here was notably more extensive. Genes containing their ribosome binding sites in an upstream gene might be "translationally coupled" to ensure equimolar translation of the two genes (reviewed in reference 16). Translational coupling is often associated with genes whose products are physically associated, such as multienzyme complexes or virions. Stoichiometric translation might be ensured by the similarity of the potential ribosome binding regions. Since *brpA* is required for transcription of *bah* (2) and ORF3 (data not shown), these translationally coupled genes are probably all transcribed from a *brpA*-dependent promoter located upstream of *bah*.

Regulation of *brpA*. When the predicted translation product of *brpA* (28,200 Da) was compared with the data bases, the 10 best matches were to regulatory components of two-component transcriptional control systems (Fig. 4) used by diverse bacteria to transduce metabolic signals and activate gene expression (1, 17, 57). One component, the sensor, has a membrane-spanning N-terminal region and is able to autophosphorylate at a conserved site in the C terminus in response to a specific external signal; we have not yet identified a sensor component which could modulate the activity of *brpA*. Typically, the phosphate group is then transferred to the N-terminal region (receiver motif) of a second component (regulator), which then serves as a transcriptional activator. Although the receiver motif is the most distinguishing conserved feature of the family, it was not found in BrpA. Instead, the N terminus of BrpA was characterized by three hydrophobic regions (Fig. 5) which probably indicate transmembrane domains or protein-protein hydrophobic interactions. Data base searches using various peptide sequences within this region were negative. That BrpA did not have the conserved phosphorylation site implied that it is not subject to posttranslational activation, that it is able to activate itself, or that it uses a previously unknown receiver structure. The sequence of the C terminus of BrpA showed that it is a member of the subfamily of transcriptional activator proteins, which includes UhpA,

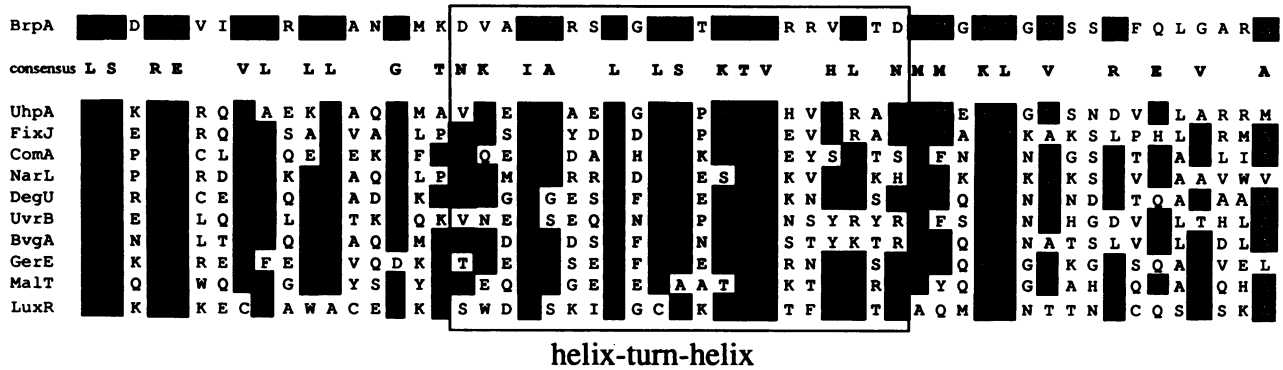


FIG. 4. Homology between the deduced amino acid sequence of *brpA* and two-component transcriptional regulator proteins. The C-terminal region of the predicted *brpA* gene product is similar to the C-terminal regions of a family of procaryotic regulator proteins (see Gross et al. [17] for references to all sequences except BvgA [3]). Shaded regions indicate positions having similar amino acids (R = K; D = E; F = Y = W; M = I = L = V) in at least six of the ten proteins used to define the consensus sequence. A region which contains a possible helix-turn-helix DNA binding consensus sequence is enclosed in a box.

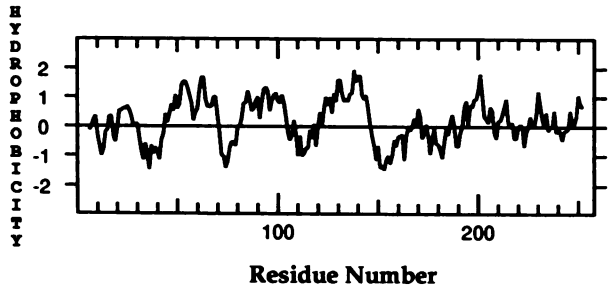


FIG. 5. Hydrophobicity profile of BrpA as calculated by the method of Kyte and Doolittle (32).

FixJ, ComA, NarL, DegU, UvrB, BvgA, GerE, MalT, and LuxR (17). This conserved region includes a potential helix-turn-helix motif (Fig. 4).

In *S. coelicolor*, expression of antibiotic biosynthetic genes may be limited by the availability of $tRNA_{UUA}$ (33). The corresponding TTA codon, found only in a few secondary metabolic genes (9), was present near the C terminus of *brpA* (Fig. 2).

S1 protection experiments (Fig. 6A) indicated that *brpA* transcription was growth phase dependent and suggested that transcription was initiated from at least three different promoters. The termini of two RNA-protected fragments mapped within an intergenic region 150 ($brpAp_1$) and 157 ($brpAp_2$) bp upstream of the presumed *brpA* translational start site. More detailed experiments are needed to demonstrate rigorously that both represent authentic mRNA initiation sites rather than processing or degradation products;

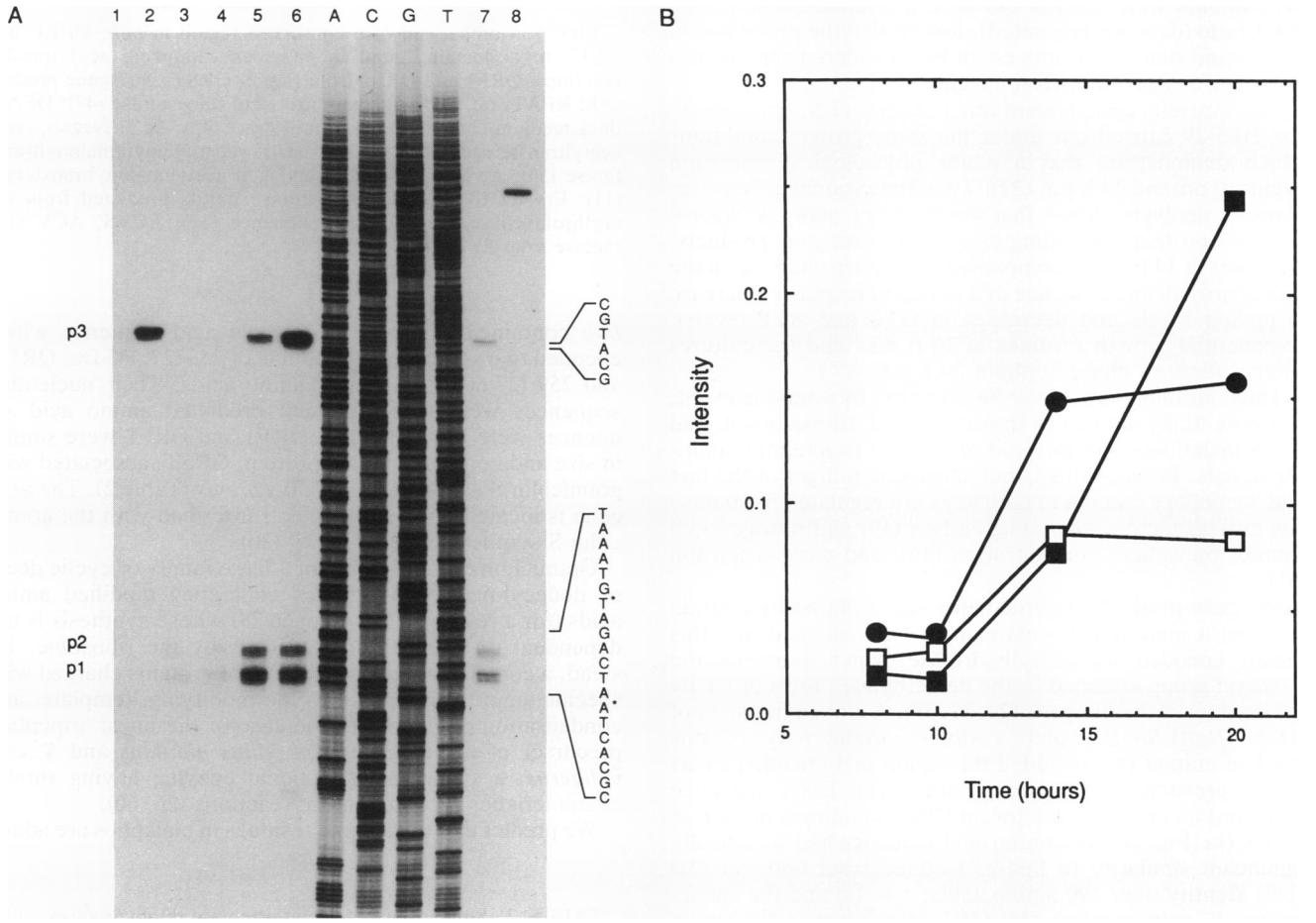


FIG. 6. S1 nuclease studies of the *brpA* promoter. (A) S1 mapping experiments were carried out as described in Materials and Methods to locate *brpA* promoters with respect to the nucleotide sequence. Two single-stranded DNA probes were prepared by using the same oligonucleotide primer and cleaving with restriction enzymes which cut in M13 either 5 (*Hin*dIII) or 235 (*Nar*I) bp upstream of the vector-*S. hygroscopicus* DNA junction (an *Sph*I site). The *Nar*I-generated probe, which could not be purified in large quantities to be used in all experiments, served as a control to demonstrate that undigested probe was not present after S1 digestion; mRNA which traverses the *Sph*I site generates an S1-protected fragment which was easily resolved from the undigested probe on the 6% polyacrylamide gels used here. The *Hin*dIII probe (lane 2) was hybridized to *S. lividans* RNA (lane 1; i.e., to verify that the probe was completely digested in the absence of *brpA* mRNA) or *S. hygroscopicus* HP5-29 RNA isolated from cultures 8 (lane 3), 10 (lane 4), 14 (lane 5), and 20 (lane 6) h p.i. The *Nar*I probe (lane 8) was hybridized only with RNA extracted from *S. hygroscopicus* HP5-29 at 14 h p.i. (lane 7). Lanes A, C, G, and T contain the products of sequencing reactions carried out with the same template and oligonucleotide primer. (B) The abundance of the three protected fragments at different stages of growth was quantified by scanning the autoradiogram to obtain total optical density in each band: ●, $brpAp_1$; □, $brpAp_2$; ■, $brpAp_3$.

however, streptomycete genes are often controlled by tandem promoters (10). In the case of the four promoters which control the *dagA* gene, it has been demonstrated that each is controlled by a different holoenzyme (7). A correlation between the sequence of streptomycete promoters (51), which are characterized by their extreme diversity, and different sigma factor specificities (6) is not yet available. These promoters could also be controlled by the same sigma factor but activated by alternative regulatory proteins. A third transcript was observed which traversed the upstream *Sph*I site located near the 3′ end of ORF3 ($brpAp_3$).

All three protected fragments were present at low levels at 8 and 10 h postinoculation (p.i.) and increased at 14 h p.i. The levels of the $brpAp_3$ band continued to increase at the onset of stationary phase (20 h p.i.), whereas $brpAp_1$- and $brpAp_2$-protected fragments showed no further change (Fig. 6B). Although these S1 experiments cannot be used to compare the absolute levels of the different mRNA species observed, the fact that the same results were obtained when experiments were carried out with a fivefold-lower probe/RNA ratio (data not presented) showed that the probe was in excess and that the results could be considered semiquantitative for any given mRNA species.

These results complement other studies of *S. hygroscopicus* HP5-29 carried out under the same growth conditions which demonstrate that a major physiological transition begins at around 14 h p.i. (22). Two-dimensional gel electrophoretic analysis shows that synthesis of many *S. hygroscopicus* proteins, including *brpA*-dependent gene products, increases at 14 h p.i. accompanied by a sharp decrease in the rate of growth and evidence of a stringent response (increase in ppGpp levels and decreases in GTP and ATP levels). Exponential growth resumes at 16 h p.i., and the cultures enter stationary phase at about 20 h p.i.

Thus, although *brpA* may be activated by a unique event, it is potentially subject to transcriptional, translational, and posttranslational controls and may serve to integrate multiple signals. Perhaps this is not surprising in light of the fact that secondary metabolic pathways are regulated by numerous growth factors, including carbon (13) and nitrogen (52) source, phosphate concentration (40), and oxygen tension (61).

***bah* gene product.** Intermediates accumulated by a strain carrying a mutation downstream of *bar* indicated that this region encoded an acetylhydrolase which removes the *N*-acetyl group attached to the demethylated form of PT by the product of the *bar* gene (24, 31, 58). A cloned fragment of DNA (*Bgl*II-*Sph*I) which restored productivity in this blocked mutant (41) included the region that encoded a part of the presumed *bah* gene product. The DNA sequence predicted an initiation site for an ORF encoding a protein of 32,076 Da (Fig. 2). This amino acid sequence had statistically significant similarity to lipases isolated from both rat (21) (24% identity over 190 amino acids; $z = 15$) and the eubacterium *A. calcoaceticus* (46) (26% identity over 206 amino acids; $z = 20$) (Fig. 7). A conserved region in all three proteins contained the motif GXSXG, which is common to lipases (14) and synthases involved in similar acyl transfer reactions (50) (Table 2). The *A. calcoaceticus* gene *est* has esterase activity which allows growth on triglycerides, including triacetyl glycerol. The observation that Bah was similar to a rat and a bacterial (*A. calcoaceticus*) lipase probably reflects the fact that the ester bonds of triglycerides and the amide bond linking acetate to PT are similar and hydrolysis is catalyzed by structurally related enzymes.

ORF1 and ORF2 gene products. The region downstream of *bah* contained a duplicated nucleic acid sequence which encoded two similar gene products of 253 (27,245 Da, ORF1) and 259 (27,665 Da, ORF2) amino acids. Their nucleotide sequences were 48% identical; predicted amino acid sequences were 31% identical. ORF1 and ORF2 were similar in size and homologous to a protein, GRST, associated with gramicidin S synthesis in *Bacillus brevis* (Table 2). The *grsT* gene is located upstream and cotranscribed with the gramicidin S synthetase I gene, *grsB* (30).

Gramicidin S is a member of a large family of cyclic deca- or dodecyl-peptide antibiotics containing modified amino acids (for a review, see reference 28) whose synthesis is not dependent on translation of mRNA by the ribosome. Instead, a complex composed of several proteins charged with specific amino acids provides the modifying, template, and condensation reactions. In the case of the linear tripeptide precursor of penicillin in *Aspergillus nidulans* and *S. clavuligerus*, a single multifunctional enzyme having similar characteristics catalyzes these reactions (26, 60).

We predict that the alanine residues in bialaphos are added

ORF1	L	F	**G**	D	**C**	M	**G**	A	L	L	A
ORF2	L	F	**G**	H	**S**	M	**G**	A	V	I	A
GRST	F	L	**G**	H	**S**	M	**G**	A	L	I	S
DFAT	L	F	**G**	H	**S**	F	**G**	S	F	V	S
RFAT	F	F	**G**	H	**S**	F	**G**	S	Y	I	A
Ac Tr (yeast)	A	T	**G**	H	**S**	Q	**G**	L	V	T	A
P/M Tr (yeast)	F	A	**G**	H	**S**	L	**G**	E	Y	A	A
Ac/M Tr (chicken)	I	L	**G**	H	**S**	Q	**G**	L	V	T	A
EryA (1)	V	V	**G**	H	**S**	Q	**G**	E	I	A	A
EryA (2)	V	I	**G**	H	**S**	Q	**G**	E	I	A	A
EryA (3)	V	A	**G**	H	**S**	A	**G**	A	L	M	A
BAH	L	A	**G**	D	**S**	A	**G**	A	G	L	A
EST	I	S	**G**	D	**S**	C	**G**	A	N	L	H
LIPR	L	A	**G**	D	**S**	A	**G**	G	N	L	C
ACVS	F	I	**G**	W	**S**	F	**G**	G	I	L	A

FIG. 7. Comparison of a conserved region in *bah*, ORF1, and ORF2 to a domain found in enzymes catalyzing acyl transfer reactions. ORF1 and ORF2 from Fig. 2. GRST, *grsT* gene product (30); RFAT, rat medium-chain fatty acid thioesterase (47); DFAT, duck medium-chain fatty acid thioesterase (45); Ac Tr (yeast), yeast acetyltransferase (50); P/M Tr (yeast), yeast palmytl/malonyltransferase (50); Ac/M Tr (chicken), chicken acetyl/malonyltransferase (11); EryA(1 to 3), three thioesterase motifs predicted from the erythromycin synthetase gene sequence (12); ACVS, ACV synthetase from *A. nidulans* (55).

TABLE 2. Similarities between vertebrate thiohydrolases and bacterial genes associated with nonribosomal peptide synthesis[a]

ORF or thioesterase	Similarity (% identity)			
	ORF2	GRST	RFAT	DFAT
ORF1	31	31	27	23
ORF2		28	25	26
GRST			32	31
RFAT				39

[a] GRST, ORF adjacent to the gramicidin S synthetase I gene (30); DFAT, duck fatty acid thioesterase (45); RFAT, rat fatty acid thioesterase (47). z values for all pairwise combinations were >30.

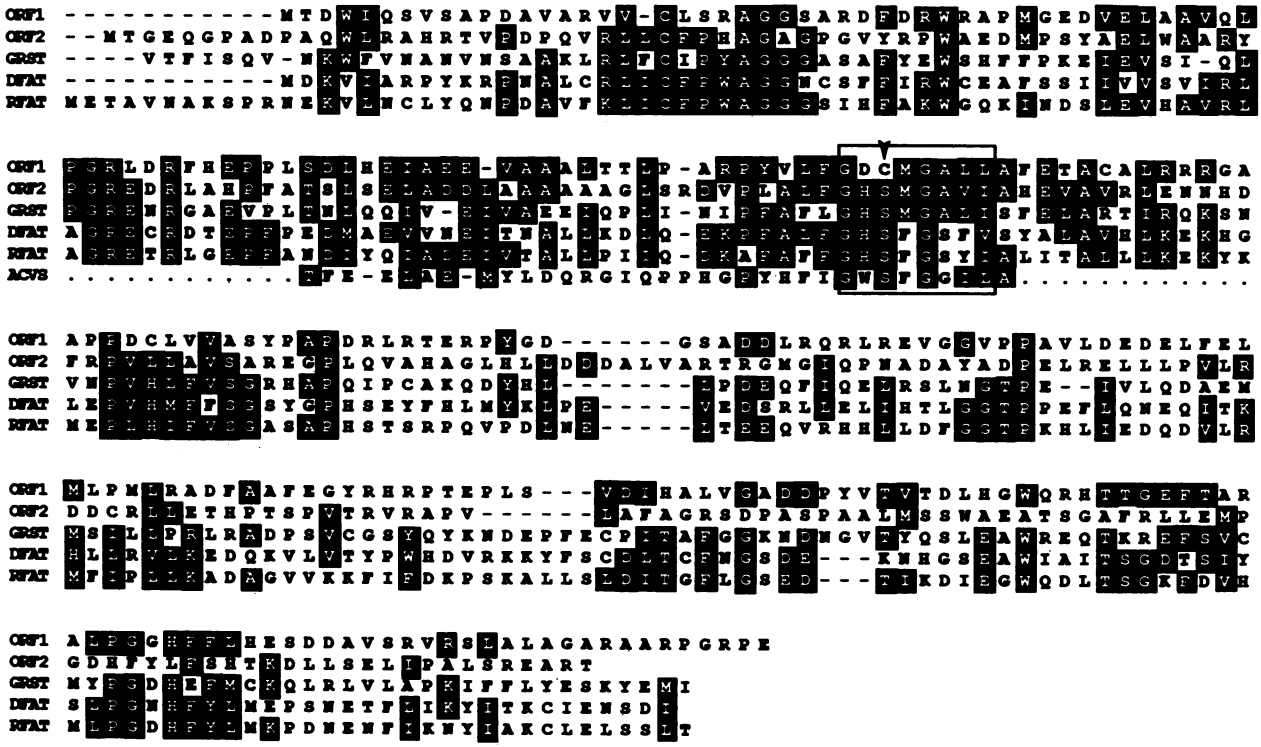


FIG. 8. ORF1 and ORF2 are similar to each other and to vertebrate thioesterases, as shown by alignment of the predicted amino acid sequences of ORF1 and ORF2 with GRST (an ORF adjacent to the gramicidin S synthetase I gene [30]), DFAT (duck fatty acid thioesterase [45]), and RFAT (rat fatty acid thioesterase [47]). Shaded regions indicate positions where similar amino acids (R = K; D = E; F = Y = W; M = I = L = V) were found in three of the five proteins. The proposed reaction center is enclosed in a box, and the active-site serine (replaced by a C residue in ORF1) is indicated by an arrow. A similar sequence was found in the multifunctional ACV synthase, ACVS (55).

to PT by such an enzyme complex which includes one or more peptidyl synthetases. Mutants which cannot carry out this reaction have been isolated. Three of the four mutations isolated, perhaps representing a defective synthetase(s), map to a ca. 15-kb region adjacent to the genes required for the synthesis of PT.

The data base search revealed that ORF1 and ORF2, as shown previously for GRST (30), were also similar to predicted vertebrate thioesterase II enzymes from duck (DFAT [45]) and rat (RFAT [47]) which can transfer and thus catalyze release of fatty acids from multifunctional synthetase complexes (Table 2 and Fig. 8). The sequence GXSXG, described above as a conserved acyl transfer domain found in *bah*, lipases, fatty acid syntheses, and a polyketide synthetase, also forms part of the presumed active center of the thioesterases. Thus, GRST, ORF1, and ORF2 probably take part in catalyzing transfer of the growing peptide chain to another site on the synthetase or hydrolyzing the thioester bond to release the completed product. Another possible thioesterase domain is found in the multifunctional *Aspergillus nidulans* δ-(L-α-aminoadipyl)-L-cysteinyl-D-valine (ACV) synthetase (55, 59) (Fig. 7 and 8). Although purified ACV synthetase from *Aspergillus nidulans* was reported to synthesize and release ACV (60), the enzyme purified from *S. clavuligerus* was associated with two proteins of 283 and 32 kDa. Jensen et al. suggested that the 32-kDa protein might be a thioesterase (26).

The unexpected observation of two independent thioesterase genes in the bialaphos cluster may suggest that some bialaphos condensation or release reactions are not catalyzed by domains on one multifunctional enzyme but rather by separate proteins. Clarification of the role these thioesterases play in bialaphos biosynthesis awaits biochemical and genetical analyses.

In any case, our observations further support the prediction made by Lipmann (35) that peptide and fatty acid enzyme complexes have striking similarities and may have common evolutionary origins. Many polyketide antibiotics are also synthesized on multifunctional enzymes similar to fatty acid synthase (12). All three processes involve synthesis on multienzyme complexes with the growing acyl chain transferred to active sites on the same, or on different proteins within the complex by a 4′-phosphopantetheine arm (5, 12, 35, 36, 53). These transfer reactions require cleavage of thioester bonds; thioesterase motifs are found either as an integral part of the synthetase or as separate proteins. We speculate that this system might also allow transfer of intermediates among polyketide, polypeptide, and fatty acid multifunctional enzymes to catalyze the ordered condensation of residues in antibiotics having mixed-chain backbones.

ORF3 gene product. The nucleotide sequence of ORF3 predicted a hydrophobic gene product of 448 amino acids having a molecular mass of 47,203 Da. Its amino acid sequence reflected structural motifs characteristic of many integral membrane proteins which transport solutes. The hydrophobicity profile (Fig. 9), included 12 alternating hydrophobic and hydrophilic domains. An N-terminal signal sequence (62) was not present. When the predicted gene product of ORF3 was compared with the data bases, the best

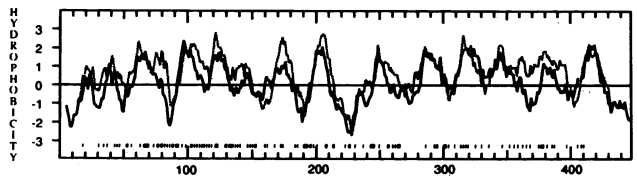


FIG. 9. Hydrophobicity profiles (plots) of ORF3 (solid line) and the citrate transporter (CIT; broken line). The positions of identical amino acids are indicated below the hydrophobicity profiles. z value = 53.

match (32% identity over 419 amino acids) was to a protein responsible for citrate uptake in *E. coli* (25, 49). The hydrophobicity profiles of the two proteins were remarkably similar (Fig. 9). Matching amino acids found throughout the proteins were not associated preferentially with either hydrophobic or hydrophilic domains (Fig. 9). Other significant matches were to human glucose transporters (15, 27), the *E. coli* arabinose transport protein (*araE* [38]), and a putative transport protein in *Leishmania* spp. (8), all members of a broad family of membrane transport proteins (37).

The similarity of the ORF3 gene product to proteins involved in carbon source uptake suggests that these proteins have similar functions. The close association of ORF3 with genes catalyzing the terminal steps in bialaphos biosynthesis implies that it may be involved in the export of bialaphos or may allow the uptake of carbon sources for conversion to bialaphos. Indeed, these two processes might be linked, as is the case of the antiporter transport systems, to allow an exchange of an external carbon source for bialaphos.

Conclusions. *Streptomyces* spp. appear to utilize specialized regulatory cascades to control genes for the synthesis of a vast array of structurally heterogeneous molecules. Nevertheless, in the case of bialaphos biosynthesis, these apparently specialized functions are catalyzed by genes similar to lipases, thioesterases, sugar transport proteins, and regulatory proteins commonly used for primary metabolism in other organisms. In addition, analysis of these genes allows predictions with respect to the synthesis of the bialaphos tripeptide using known thio-template mechanisms and also to the requirement of a bialaphos export system.

ACKNOWLEDGMENTS

We thank H. Anzai, O. Hara, C. R. Hutchinson, S. Imai, M. Maraheil, T. Murakami, K. Nagaoka, E. Takano, F. Kunst, J. Parkinson, T. Msadek, and H. von Döhren for invaluable contributions of ideas and Wolfgang Wohlleben for critical sequence analysis. The thoughtful suggestions of J. Davies made this a better paper.

This work was supported by Meiji Seika Kaisha, Ltd., Biogen, Inc., Pfizer, Inc., and the Institut Pasteur.

REFERENCES

1. **Albright, L. M., E. Huala, and F. M. Ausubel.** 1989. Prokaryotic signal transduction mediated by sensor and regulatory protein pairs. Annu. Rev. Genet. **23**:311–336.
2. **Anzai, H., T. Murakami, S. Imai, A. Satoh, K. Nagaoka, and C. J. Thompson.** 1987. Transcriptional regulation of bialaphos biosynthesis in *Streptomyces hygroscopicus*. J. Bacteriol. **169**: 3482–3488.
3. **Arico, B., J. F. Miller, C. Roy, S. Stibitz, D. Monack, S. Falkow, R. Gross, and R. Rappuoli.** 1989. Sequences required for expression of *Bordetella pertussis* virulence factors share homology with prokaryotic signal transduction proteins. Proc. Natl. Acad. Sci. USA **86**:6671–6675.
4. **Bayer, E., K. H. Gugel, K. Hagele, H. Hagenmaier, S. Jessipow, W. A. Köning, and H. Zähner.** 1972. Stoffwechselprodukte von Mikroorganismen. Phosphinothricin und Phosphinothricyl-alanyl-alanine. Helv. Chim. Acta **55**:224–239.
5. **Bibb, M. J., S. Biro, 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 biosynthesis. EMBO J. **8**:2727–2736.
6. **Buttner, M. J.** 1989. RNA polymerase heterogeneity in *Streptomyces coelicolor* A3(2). Mol. Microbiol. **3**:1653–1659.
7. **Buttner, M. J., A. M. Smith, and M. J. Bibb.** 1988. At least three different RNA polymerase holoenzymes direct transcription of the agarase gene (*dagA*) of *Streptomyces coelicolor* A3(2). Cell **52**:599–607.
8. **Cairns, B. R., M. W. Collard, and S. M. Landfear.** 1989. Developmentally regulated gene from *Leishmania* encodes a putative membrane transport protein. Proc. Natl. Acad. Sci. USA **86**:7682–7686.
9. **Chater, K. F.** 1989. Multilevel regulation of *Streptomyces* differentiation. Trends Genet. **5**:372–377.
10. **Chater, K. F., and D. A. Hopwood.** 1989. Antibiotic biosynthesis in *Streptomyces*, p. 129–150. *In* D. A. Hopwood and K. F. Chater (ed.), Genetics of bacterial diversity. Academic Press, Inc. (London), Ltd., London.
11. **Chirala, S. S., R. Kasturi, M. Pazirandeh, D. T. Stolow, W.-Y. Huang, and S. J. Wakil.** 1989. A novel cDNA extension procedure. J. Biol. Chem. **264**:3750–3757.
12. **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 synthetase of *Saccharopolyspora erythraea*. Nature (London) **348**:176–178.
13. **Demain, A.** 1989. Carbon source regulation of idiolite biosynthesis in actinomycetes, p. 127–134. *In* S. Shapiro (ed.), Regulation of secondary metabolism in actinomycetes. CRC Press, Boca Raton, Fla.
14. **Docherty, A. J. P., M. W. Bodmer, A. S., R. Verger, C. Riviere, P. A. Lowe, A. Lyons, J. S. Emtage, and T. J. R. Harrys.** 1985. Molecular cloning and nucleotide sequence of rat lingual lipase cDNA. Nucleic Acids Res. **13**:1891–1903.
15. **Fukumoto, H., S. Seino, H. Imura, Y. Sieno, R. L. Eddy, Y. Fukushima, M. G. Byers, T. B. Shows, and G. I. Bell.** 1988. Sequence and tissue distribution and chromosomal localization of mRNA encoding a human glucose transporter-like protein. Proc. Natl. Acad. Sci. USA **85**:5434–5438.
16. **Gold, L., and G. Stormo.** 1987. Translational initiation, p. 1302–1307. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
17. **Gross, R., B. Aricò, and R. Rappuoli.** 1989. Families of bacterial signal-transducing proteins. Mol. Microbiol. **3**:1661–1667.
18. **Hara, O., H. Anzai, S. Imai, Y. Kumada, T. Murakami, R. Itoh, E. Takano, A. Satoh, and K. Nagaoka.** 1988. The bialaphos biosynthetic genes of *Streptomyces hygroscopicus*: cloning and analysis of the genes involved in the alanylation step. J. Antibiot. **41**:538–547.
19. **Hidaka, T., S. Imai, O. Hara, H. Anzai, T. Murakami, K. Nagaoka, and H. Seto.** 1990. Carboxyphosphonoenolpyruvate phosphonomutase, a novel enzyme catalyzing C-P bond formation. J. Bacteriol. **172**:3066–3072.
20. **Hidaka, T., H. Seto, and S. Imai.** 1989. Biosynthetic mechanism of C-P bond formation. Isolation of carboxyphosphonoenolpyruvate and its conversion to phosphinopyruvate. J. Am. Chem. Soc. **111**:8012–8013.
21. **Holm, C., T. G. Kirchgessner, K. L. Svenson, A. J. Lusis, and P. Berfrage.** 1988. Nucleotide sequence of rat adipose hormone sensitive lipase cDNA. Nucleic Acids Res. **16**:9879.
22. **Holt, T. G., C. Chang, C. Laurent-Winter, T. Murakami, J. I. Garrels, J. E. Davies, and C. J. Thompson.** Unpublished data.
23. **Hopwood, D. A., M. J. Bibb, K. F. Chater, T. Kieser, C. J. Bruton, H. M. Kieser, D. J. Lydiate, C. P. Smith, J. M. Ward, and H. Schrempf.** 1985. Genetic manipulation of streptomyces.

A laboratory manual. The John Innes Foundation, Norwich, United Kingdom.
24. **Imai, S., H. Seto, T. Sasaki, T. Tsuruoka, H. Ogawa, A. Satoh, S. Inouye, T. Niida, and N. Otake.** 1985. Studies on the biosynthesis of bialaphos (SF-1293). 6. Production of N-acetyldemethylphosphinothricin and N-acetylbialaphos by blocked mutants of *Streptomyces hygroscopicus* SF-1293 and their roles in the biosynthesis of bialaphos. J. Antibioti. **38**:687–690.
25. **Ishiguro, N., and G. Sato.** 1985. Nucleotide sequence of the gene determining plasmid-mediated citrate utilization. J. Bacteriol. **164**:977–982.
26. **Jensen, S. E., A. Wong, M. J. Rollins, and D. W. S. Westlake.** 1990. Purification and partial characterization of δ-(L-α-aminoadiptyl)-L-cysteinyl-D-valine synthetase from *Streptomyces clavuligerus*. J. Bacteriol. **172**:7269–7271.
27. **Kayano, T., H. Fufumoto, R. L. Eddy, Y. Fan, M. G. Byers, T. B. Shows, and G. I. Bell.** 1988. Evidence for a family of human glucose transporter-like proteins: sequence and gene localization of a protein expressed in fetal skeletal muscle and other tissues. J. Biol. Chem. **262**:15245–15248.
28. **Kleinkauf, H., and H. von Döhren.** 1987. Biosynthesis of peptide antibiotics. Annu. Rev. Microbiol. **41**:259–289.
29. **Kondo, Y., T. Shomura, Y. Ogawa, T. Tsuruoka, H. Watanabe, K. Totsukawa, T. Suzuki, C. Moriyama, J. Yoshida, S. Inouye, and T. Niida.** 1973. Studies of a new antibiotic SF1293. I. Isolation and physicochemical and biological characterization of SF1293 substances. Sci. Rep. Meiji Seika Kaisha **13**:34–41.
30. **Krätzschmar, J., M. Krause, and M. A. Marahiel.** 1989. Gramicidin S biosynthesis operon containing the structural genes *grsA* and *grsB* has an open reading frame encoding a protein homologous to fatty acid thioesterases. J. Bacteriol. **171**:5422–5429.
31. **Kumada, Y., H. Anzai, E. Takano, T. Murakami, O. Hara, R. Itoh, S. Imai, A. Sato, and K. Nagaoka.** 1988. The bialaphos resistance gene (*bar*) plays a role in both self-defense and bialaphos biosynthesis, in *Streptomyces hygroscopicus*. J. Antibiot. **41**:1838–1845.
32. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
33. **Lawlor, K. J., H. A. Baylis, and K. F. Chater.** 1987. Pleiotropic morphological and antibiotic deficiencies result from mutations in a gene encoding a tRNA-like product in *Streptomyces coelicolor* A3(2). Genes Dev. **1**:1305–1310.
34. **Lipman, D. J., and W. R. Pearson.** 1985. Rapid and sensitive protein similarity searches. Science **227**:1435–1441.
35. **Lipmann, F.** 1971. Attempts to map a process evolution of peptide biosynthesis. Science **173**:875–884.
36. **Lynen, F.** 1980. On the structure of fatty acid synthetase of yeast. Eur. J. Biochem. **112**:431–442.
37. **Maiden, M. C. J., E. O. Davis, S. A. Baldwin, D. C. M. Moore, and P. J. F. Henderson.** 1987. Mammalian and bacterial sugar transport proteins are homologous. Nature (London) **325**:641–643.
38. **Maiden, M. C. J., M. C. Jones-Mortimer, and P. J. F. Henderson.** 1988. The cloning, DNA sequence, and overexpression of the gene *araE* coding for arabinose-proton symport in *Escherichia coli* K12. J. Biol. Chem. **263**:8003–8010.
39. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor, N.Y.
40. **Martín, J. F.** 1989. Molecular mechanisms for the control by phosphate of the biosynthesis of antibiotics and other secondary metabolites, p. 213–238. *In* S. Shapiro (ed.), Regulation of secondary metabolism in actinomycetes. CRC Press, Boca Raton, Fla.
41. **Murakami, T., H. Anzai, S. Imai, A. Satoh, K. Nagaoka, and C. J. Thompson.** 1986. The bialaphos biosynthetic genes of *Streptomyces hygroscopicus*: molecular cloning of the gene cluster. Mol. Gen. Genet. **205**:42–50.
42. **Murakami, T., T. Holt, and C. J. Thompson.** 1989. Thiostrepton-induced gene expression in *Streptomyces lividans*. J. Bacteriol. **171**:1459–1466.
43. **Ogawa, H., S. Imai, A. Satoh, and M. Kojima.** 1983. An improved method for the preparation of *Streptomyces* and *Micromonospora* protoplasts. J. Antibiot. **36**:184–186.
44. **Pearson, W. R., and D. J. Lipman.** 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA **85**:2444–2448.
45. **Poulose, A. J., L. Rogers, T. M. Cheesebrough, and P. E. Kolattukudy.** 1985. Cloning and sequencing of cDNA of S-acyl fatty acid synthetase thioesterase from uropygial gland of mallard duck. J. Biol. Chem. **260**:15953–15958.
46. **Reddy, P. G., R. Allon, M. Mevarech, S. Mendelovitz, Y. Sato, and D. L. Gutnick.** 1989. Cloning and expression in *Escherichia coli* of an esterase-coding gene from the oil-degrading bacterium *Acinetobacter calcoaceticus* RAG-1. Gene **76**:145–152.
47. **Safford, R., J. de Silva, C. Lucas, J. H. C. Windust, J. Sheddon, C. M. James, C. M. Sidebottom, A. R. Sladbas, M. P. Tombs, and S. G. Hughes.** 1987. Molecular cloning and sequence analysis of complementary DNA encoding rat mammary gland medium-chain S-acyl fatty acid synthetase thio ester hydrolase. Biochemistry **26**:1358–1364.
48. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
49. **Sasatsu, M., T. K. Misra, L. Chu, R. Laddaga, and S. Silver.** 1985. Cloning and DNA sequence of a plasmid-determined citrate utilization system in *Escherichia coli*. J. Bacteriol. **164**: 983–993.
50. **Schweizer, M., L. M. Roberts, H. Möltke, K. Takabayashi, M. Möllerer, B. Hoffmann, G. Müller, M. Kötting, and E. Schweizer.** 1986. The pentafunctional fas-1 gene of yeast—its nucleotide sequence and order of catalytic domains. Mol. Gen. Genet. **203**:479–486.
51. **Seno, E. T., and R. H. Baltz.** 1989. Structural organization and regulation of antibiotic biosynthesis and resistance genes in actinomycetes, p. 1–48. *In* S. Shapiro (ed.), Regulation of secondary metabolism in actinomycetes. CRC Press, Boca Raton, Fla.
52. **Shapiro, S.** 1989. Nitrogen assimilation in actiomycetes and the influence on nitrogen nutrition on actinomycete secondary metabolism, p. 135–214. *In* S. Shapiro (ed.), Regulation of secondary metabolism in actinomycetes. CRC Press, Boca Raton, Fla.
53. **Sherman, D. H., F. Malpartida, M. J. Bibb, H. Kieser, M. J. Bibb, and D. A. Hopwood.** 1989. Structure and deduced function of the granaticin-producing polyketide synthetase gene cluster of *Streptomyces violaceoruber* Tü22. EMBO J. **8**:2717–2725.
54. **Shine, J., and L. Dalgarno.** 1975. Determinant of cistron specificity in bacterial ribosomes. Nature (London) **254**:34–38.
55. **Smith, D. J., A. J. Earl, and G. Turner.** 1990. The multifunctional peptide synthetase performing the first step of penicillin biosynthesis in *Penicillium chrysogenum* is a 421,073 dalton protein similar to *Bacillus brevis* peptide antibiotic synthetases. EMBO J. **9**:2743–2750.
56. **Staden, R.** 1984. Codon preference and its use in identifying protein coding regions in long DNA sequences. Nucleic Acids Res. **12**:521–538.
57. **Stock, J. B., A. M. Stock, and J. M. Mottonen.** 1990. Signal transduction in bacteria. Nature (London) **344**:395–400.
58. **Thompson, C. J., N. R. Movva, R. Tizard, R. Crameri, J. E. Davies, M. Lauwereys, and J. Botterman.** 1987. Characterization of the herbicide-resistance gene *bar* from *Streptomyces hygroscopicus*. EMBO J. **6**:2519–2523.
59. **von Döhren, H.** Personal communication.
60. **van Liempt, H., H. von Döhren, and H. Kleinkauf.** 1989. δ-(L-α-Aminoadipyl)-L-cysteinyl-D-valine synthetase from *Aspergillus nidulans*. J. Biol. Chem. **264**:3680–3684.
61. **Votruba, J., and Z. Vanek.** 1989. Physiochemical factors affecting actinomycete growth and secondary metabolism, p. 263–281. *In* S. Shapiro (ed.), Regulation of secondary metabolism in actinomycetes. CRC Press, Boca Raton, Fla.
62. **Watson, E. E.** 1984. Compilation of published signal sequences. Nucleic Acids Res. **12**:5145–5165.
63. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33**:103–119.