# Discovery and Characterization of <sup>a</sup> New Transposable Element, Tn4811, in Streptomyces lividans 66

CARTON W. CHEN,<sup>1\*</sup> TIN-WEIN YU,<sup>1</sup>† HUI-MING CHUNG,<sup>1</sup>‡ AND CHUN-FEN CHOU<sup>2</sup>

Institute of Genetics<sup>1</sup> and Institute of Microbiology and Immunology,  $2$ National Yang-Ming Medical College, Shih-Pai, Taipei, Taiwan, Republic of China

Received 29 July 1992/Accepted 28 September 1992

Transposition of a new 5.4-kb transposon, Tn4811, of Streptomyces lividans to the melC operon of Streptomyces antibioticus on plasmid pU702 was discovered. The nucleotide sequence of this copy of Tn4811, which contained an imperfect (9 of 11 bp) terminal inverted repeat, five putative Streptomyces coding sequences for an oxidoreductase and its transcription regulator, and three transposition-related proteins, was determined. SLP<sup>-</sup> strains of S. lividans contained one copy (A) of Tn4811, while SLP2<sup>+</sup> strains contained an additional copy  $(B)$  on the SLP2 plasmid. The nucleotide sequences at three insertion junctions of Tn4811 were determined. Copy B lacked <sup>41</sup> bp from the left end. At the other five junctions the duplication of <sup>a</sup> putative 3-bp target sequence (TGA) was observed. A sequence of less than 3 kb homologous to Tn4811 was present in S. antibioticus. DNA homologous to Tn4811 was not detected in <sup>14</sup> other Streptomyces species.

Structural instability of genomic DNA is <sup>a</sup> widespread feature of gram-positive, filamentous, soil bacteria of the genus Streptomyces (32). Many genetic traits undergo spontaneous loss in laboratory cultures at frequencies of  $10^{-4}$  to  $10^{-2}$ . Exposure to certain stresses such as UV irradiation, DNA intercalating agents, cold temperature, or protoplasting and regeneration increases the frequency of these instabilities. The mutations involved in the instabilities are attributed to large deletions of chromosomal DNA, which are frequently accompanied by tandem amplifications of particular sequences nearby (32). The genetic principle(s) underlying the structural instability is not clear. Possible mechanisms involved are homologous recombination, site-specific recombination, and transposition. Homologous recombination systems in Streptomyces species and their contribution to genetic instability have received little study. Tsai and Chen (41) isolated a rec mutant defective in intraplasmid recombination but not in chromosomal recombination (25). Chou and Chen (11), in their investigation of an unstable arg gene, found no effect of this rec mutation on its instability. Site-specific recombination and transposition, although representing forms of fluidity of genomic DNA, have not been implicated in the instability of other DNA sequences (8).

About one-third of Streptomyces species can produce melanin pigment (44). The melanin (melC) operon of Streptomyces antibioticus has been cloned (24), sequenced (3), and widely used in many recombinant vectors. The melC operon in the Streptomyces species examined is genetically unstable, undergoing spontaneous deletions at relatively high frequencies—about  $10^{-3}$  (32). Similar to those observed in other unstable genes in Streptomyces species, the deletions of melC were frequently accompanied by extensive tandem amplifications of specific sequences (18, 32).

The structure of the melC sequence on recombinant plasmids is also relatively unstable, being frequently observed in many laboratories. In this study we characterized the structural instabilities of melC on a recombinant plasmid, pIJ702, and discovered a new 5.4-kb transposon, designated Tn4811 (independently discovered by Tanaka et al. [38]). The nucleotide sequence of Tn4811, which exhibited an imperfect terminal inverted repeat and several potential coding sequences with putative transpositional and accessory functions, was determined. Two copies of Tn4811 were found in wild-type Streptomyces lividans. Putative duplications of target sequences were detected at the insertion junctions. Screening of 15 other Streptomyces species for sequences homologous to Tn4811 by hybridization revealed only a less than 3-kb homolog in S. antibioticus.

## MATERIALS AND METHODS

Bacterial cultures, plasmids, and phages. The bacterial strains and plasmids used in this study are in Tables 1 and 2, respectively. S. lividans TK64 argX was an Arg<sup>-</sup> mutant of TK64 that required arginine or ornithine for growth (11). Streptomyces albus B3381 and Streptomyces griseorubens IAM0060 were a gift from Ying-Chieh Tsai (National Yang-Ming Medical College, Taipei, Taiwan). S. antibioticus strains were grown on S agar or in S liquid medium (9) at 30°C. S. lividans was grown on R2YE agar or in TSB or YEME liquid medium (22) at 30°C. Escherichia coli cultures were grown on L agar or in L broth (27) at 37°C. VCSM13 helper phage was from Stratagene (La Jolla, California).

Enzymes and reagents. Restriction enzymes and T4 DNA ligase were from Boehringer Mannheim (Mannheim, Germany), Bethesda Research Laboratories (Gaithersburg, Md.), and New England Biolabs (Beverly, Mass.) and were used according to the specifications supplied. Thiostrepton was a gift from S. J. Lucania of Squibb Institute for Medical Research (Princeton, N.J.). Sequenase and TaqI polymerase sequencing kits were from United States Biochemical (Cleveland, Ohio) and Promega (Madison, Wis.), respec-

<sup>\*</sup> Corresponding author.

t Present address: John Innes Institute, John Innes Center, Norwich NR4 7UH, United Kingdom.

f Present address: Department of Genetics and Development, Columbia University, New York, NY 10032.

<b>Bacterial strain</b>	Genotype or characteristics	Source <sup>a</sup> or reference
E. coli XL-1 Blue	recAl lac endAl gyrA96 thi hsdR17 supE44 relA1 (F' proAB lacI <sup>q</sup> lacZ $\Delta M$ 15 Tn10)	<b>Stratagene</b>
<b>Streptomycetes</b>		
S. lividans		
1326	Wild type $(SLP2^+ SLP3^+)$	22
<b>TK19</b>	$SLP2^ SLP3^+$	22
<b>TK20</b>	$SLP2^+ SLP3^-$	22
TK21	$SLP2^-$ SLP3 <sup>-</sup>	22
<b>TK23</b>	$SLP2^-$ SLP3 <sup>-</sup> spc-1	22
TK24	$SLP2^-$ SLP3 <sup>-</sup> str-6	22
<b>TK64</b>	$SLP2^-$ SLP3 <sup>-</sup> pro-2 str-6	22
TK64 argX	$SLP2^-$ SLP3 <sup>-</sup> pro-2 str-6 argX	11
JT46	$SLP2^ SLP3^-$ pro-2 str-6 rec-46	41
S. alboniger ATCC 12461	Wild type	<b>CCRC</b>
<i>S. albus</i> B3381	Wild type	Y.-C. Tsai
S. antibioticus IMRU 3720	Wild type	24
S. argenteolus ATCC 11009	Wild type	<b>CCRC</b>
S. cattleya IFO 14057	Wild type	<b>CCRC</b>
S. coelicolor M130	uraAl argAl strAl SCP1 <sup>-</sup> SCP2 <sup>-</sup>	22
S. glaucescens ATCC 23622	Wild type	<b>CCRC</b>
S. griseorubens IAM0060	Wild type	Y.-C. Tsai
S. griseus ATCC 10137	Wild type	<b>CCRC</b>
S. hydrogenans ATCC 19631	Wild type	<b>CCRC</b>
S. kanamyceticus ATCC 12853	Wild type	<b>CCRC</b>
<i>S. paryulus</i> 2283 (from ATCC 12434)	str-1	22
S. moderatus ATCC 23443	Wild type	<b>CCRC</b>
S. nogalater ATCC 27451	Wild type	<b>CCRC</b>
S. sparsogenes ATCC 25498	Wild type	<b>CCRC</b>

TABLE 1. Bacterial strains used in this study

<sup>a</sup> CCRC, Culture Collection and Research Center, Food Industry Research and Development Institute, Hsinchu, Taiwan.

recombinant DNA techniques were the methods of Sam-<br>brook et al. (29). Preparation, transformation, and regener-

**DNA sequencing.** pLUS729 was digested with BamHI into (Table 2) were sequenced directly without subcloning.<br>three fragments, subcloned into pBluescript II KS(-) and **Computer analyses of DNA and protein sequences.** three fragments, subcloned into pBluescript II  $KS(-)$  and **Computer analyses of DNA and protein sequences.** DNA propagated in E. coli XL1-B (6). These subclones were and protein sequences were analyzed by the University of plasmid DNA was rescued by using VSCM13 helper phage

tively. KS and SK sequencing primers were from Stratagene. and sequenced with the dideoxy method of Sanger et al. (30)  $[\alpha^{-32}P]$ d-eoxynucleoside triphosphate and  $[\alpha^{-35}S]dATP$  were by using the Sequences or the *TaqI* Po  $[\alpha^{-32}P]$ d-eoxynucleoside triphosphate and  $[\alpha^{-35}S]dATP$  were by using the Sequenase or the TaqI Polymerase Sequencing from Amersham (Buckinghamshire, England). Kit. Primers were either the SK or KS primer or synthetic om Amersham (Buckinghamshire, England). Kit. Primers were either the SK or KS primer or synthetic<br> **DNA manipulations and genetic transformations.** Standard oligonucleotides based on the sequences determined. Oligooligonucleotides based on the sequences determined. Oligo-<br>nucleotides were synthesized with an oligonucleotide synbrook et al. (29). Preparation, transformation, and regener-<br>ation of *Streptomyces* protoplasts were by the methods of pLUS797, pLUS798, and pLUS799 DNA containing the ation of *Streptomyces* protoplasts were by the methods of pLUS797, pLUS798, and pLUS799 DNA containing the various insertion iunctions on pBluescript II KS(+)/KS(-) various insertion junctions on pBluescript II  $KS(+)/KS(-)$ 

propagated in E. coli XL1-B (6). These subclones were and protein sequences were analyzed by the University of subsequently dissected with other restriction enzymes and Wisconsin Genetic Computing Group (UWGCG) package subsequently dissected with other restriction enzymes and Wisconsin Genetic Computing Group (UWGCG) package further subcloned into 25 recombinant plasmids (based on (version 7.0), PC/Gene (IntelliGenetics, release 6.60), o further subcloned into 25 recombinant plasmids (based on (version 7.0), PC/Gene (IntelliGenetics, release 6.60), or the pBluescript II vectors) for sequencing. Single-stranded GeneWorks (IntelliGenetics, release 2.01). The GeneWorks (IntelliGenetics, release 2.01). The protein and DNA data bases used for homology search were SwissProt

TABLE 2. Plasmids used in this study

Plasmid	Description	Source or reference
SLP <sub>2</sub>	50-kb linear conjugative plasmid of S. lividans	10
pBluescript II $KS(+)/KS(-)$	E. coli phagemid cloning vectors, bla lacZ $\alpha$	Stratagene
pIJ702	Streptomyces plasmid pIJ350 containing melC of S. antibioticus	24
pLUS729	pIJ702 melC2::Tn4811	This study
pLUS791	pBluescript II $KS(-)$ carrying the 5.0-kb KpnI-SstI fragment of Tn4811 from pLUS729	This study
pLUS796	pBluescript II $KS(-)$ carrying the 0.45-kb KpnI-EcoRI fragment encompassing the left terminus of copy $A$ of Tn4811	This study
pLUS797	pBluescript II $KS(+)$ carrying the 0.40-kb BamHI-PstI fragment containing the right terminus of copy $\vec{A}$ of Tn4811	This study
pLUS798	pBluescript II $KS(+)$ carrying the 4.0-kb <i>EcoRI-ClaI</i> fragments of SLP2 containing the left terminus of copy $B$ of Tn4811	This study
pLUS799	pBluescript II $KS(+)$ carrying the 4.7-kb <i>EcoRI-ClaI</i> fragments of SLP2 containing the right terminus of copy $B$ of Tn4811	This study

(version 18.0) and EMBL (version 27.0). Prosite (version 4.00 in PC/Gene) was used for searching for signals and motifs in proteins.

# RESULTS AND DISCUSSION

 $melC$  on pLJ702 was structurally unstable. Recombinant plasmid pIJ702 (24) contains the cloned melC sequence of S. antibioticus, which confers melanin production to Mel<sup>-</sup> hosts, S. lividans, and others. Spontaneous Mel<sup>-</sup> variants were detected at high frequencies in pIJ702 transformants of S. lividans TK64. The frequencies were in the order of  $10^{-3}$ in fresh cultures from single Mel' reisolates or transformants but appeared to rise as the cultures were progressively propagated. In some cases the Mel<sup>-</sup> phenotype was deduced to be due to an unidentified mutation in the S. lividans host, because plasmid DNA isolated from these cultures still conferred melanin production to secondary transformants. This genetic determinant also appeared to be unstable; mutations in it were readily detected and accumulated in cultures propagated without reisolation (up to a frequency of 30% in very old cultures) (40). These mutants were not explored further. In other cases, however, the mutations were evidently on the plasmid (see below), because secondary transformants by these plasmids were  $Mel^-$ 

Analysis of plasmids isolated from these  $Mel^-$  transformants showed that rearrangements had occurred in many of them. Of 11 Mel<sup>-</sup> plasmids analyzed, 3 had 1.0- to 1.8-kb deletions including part of  $melC$ , 2 showed no apparent changes in size, and 6 contained a 5.4-kb insert.

Insertional mutations by a new transposable element, Tn4811. Digestion of the six plasmids with a 5.4-kb insert with *BclI* produced the same seven fragments (not shown). Four of the restriction sites were on pIJ702 (24), and the other three must be on the insert. It thus appeared that either the insert had integrated at approximately the same position on these six plasmids or these plasmids were of the same origin. One of them, pLUS729 (Fig. 1), was chosen for further characterization. Restriction mapping of pLUS729 showed that the insertion was between the SstI and SalI sites within the  $melC2$  (tyrosinase) gene.

The size and restriction map of the inserted sequence were identical to <sup>a</sup> mobile DNA element in <sup>a</sup> preliminary report by Tanaka et al. (36, 38). This element was originally designated IS456, but because of the complexity of its genetic content, we (Tanaka's and our laboratories) now regard it a type II transposon (see below) and have renamed it Tn4811.

Determination of nucleotide sequence of Tn4811. The copy of Tn4811 on pLUS729 was subcloned into 25 phagemids and sequenced in both directions by using synthetic oligonucleotide primers. The Tn4811 sequence was determined by comparison of the compiled sequence with the melC sequence. Between the putative duplicated target se-<br>quences, TGA (see below), the total length of Tn4811 is 5,396 bp (Fig. 2). The average  $G+C$  content is 68%. There is an imperfect inverted terminal repeat (a match of 9 of 11 bp).

Codon preference analysis (16; data not shown), based on the preference for G/C in the third position of codons by typical Streptomyces genes (5), revealed five putative coding sequences, open reading frame <sup>1</sup> (ORF1) to ORF5. Prediction with the TESTCODE algorithm (14; data not shown) also identified the same regions as coding sequences. The sizes of the proteins encoded by ORF1 to ORF5 were 336, 138, 320, 297, and 304 amino acids, respectively.

ORF4 and ORF5 as a putative accessory unit. The protein data bases were searched for sequences similar to the



FIG. 1. Restriction map of pLUS729. The melC1 and melC2 sequences are indicated by the thick solid arrows, and the tsr gene is indicated by the thin solid arrow. Tn4811 is indicated by the gray double-headed arrow. The two sequenced restriction fragments that contain the insertion junctions are indicated by the two outermost arcs. Abbreviations for restriction enzymes: Ba, BamHI; Sa, Sall; Bc, BclI; Cl, ClaI; Pv, PvuII; Ss, SstI; Ps, PstI; Sp, SphI; Ec, EcoRI; Kp, KpnI; Bg, BglII.

potential protein products encoded by the five ORFs. Only the ORF4 and ORF5 products found significantly similar protein sequences. The ORF4 product is similar to <sup>a</sup> group of oxidoreductases (Fig. 3), including the protochlorophyllide reductase (Pcr) of barley (33), the glucose dehydrogenase of Bacillus megaterium (19), the ribitol dehydrogenase of Klebsiella aerogenes (28), and the NodG protein of Rhizobium  $meliloti$  (which is homologous to  $\beta$ -ketoacyl reductase [34]). The similarity among these oxidoreductases was stronger at the N termini. The ORF4 product was most similar to Pcr, with an identity of 35% and a similarity of 48% (Fig. 3). Although the putative ORF4 product is likely also an oxidoreductase, its function is not clear.

The ORF5 product is similar to a group of bacterial transcription regulators of the AraC (arabinose operon regulator) family. Figure 4 shows an optimized alignment of the ORF5 product with five members of the AraC family. AraC of Erwinia carotovora exhibited the highest homology with the ORES product (26% identity and 50% similarity). Along the polypeptide sequences the strongest similarity lies in the C termini that contain potential helix-turn-helix DNA-binding domains (17), also seen in the ORES product.

The divergently oriented ORF4 and ORF5 are separated by <sup>188</sup> bp of relatively low G+C content. This region is likely to contain the promoter and regulatory sequences involved in the divergent transcription of the two putative genes. The distances between the translation start sites in the known divergent promoters of prokaryotes range from 103 to 611 bp (2). Beck and Warren (2) reported that in the 45 compiled regions of divergent transcription in prokaryotes, 18 were of the regulatory-structural type. With one exception ( $micF\text{-}ompC$  of E. coli), in which the regulator RNA

TIR<br>GIPGATICTTT GTGATCCCCG GTTCTGAGTA GCCTTGCGTC CTAGGCTGGC GTGGTGTCCA 60 CAGGTCAGTA CTGCTGCCAC CGCTGGGAGG CCCGCCGGTT CAGCCACTGA AGAGCCAGCT Q V <sup>8</sup> T A A T A G R P A G <sup>8</sup> A T <sup>Z</sup>E P A 2940 304 GGGGTACCGG GCGGGCCGGG GAAOGGGACT GTCGGGCTTG TGGCGAAOGG CTGAGGCCGG 120 CGAGCTGCCC CGCTGTCAGC CCGGCAGCCT CAGCCAGCCG GTCGGTTCTG AGCCCGTCGG<br>R A A P L S A R Q P Q P A G R P \* GTTCGCGGCC TGATGCGCTG TTCTGTTCCT CGGCATGTcG TGccAGGCAG TGGCGGGCGG 180  $rac{000}{320}$ GATACTTGTC GOGCACCCGG GCGATCTTGC CGATCCATAG TTCTTCCTCG ATGAAGGGCC AACGCCGGTT ACGCAAGCGC CTGGCGGCCG TCCGGGGCGG TGCCGGTGAG GTGGAGTGTC 240 3060 GACGCGCATA CCOCTOCCGG AGGACGTCCA GATGCTCCTG GGAGCATGCG GTGACGAGGC 3120 GCACACCGTC ATGTGCCAGG TCATGCACAG AGACCGCCGA CGAGTCCGGC ACCGCATACC 3180 3240 GCATTCGCCG GCCTTCOGCG ATGGCCTGCT GGCACACGTC GCACACCACG TTGTCCCCCG 3300 TCGGGTCCTC GCACGCTCGC CTGGATTCCA CCGCGGCGGC CGGTCATCCG GGCCGAGGG AGTAGGC CGCATC . G A S L  $540$ 294 3360 TGTCAGCGCC GTACTCCCAG AGCCGGTCTG CTGCCTCCGG GTCGAGTGCG TGGGCGGCTA<br>ACAGTCGCGG CATGAGGGTC TCGGCCAGAC GACGGAGGCC CAGCTCACGC ACCCGCCGAT<br>T D A G Y E N L R D A A EP D L A H A A 274 3420 CGCCGCCGGG CTGGACGTCC CCGTCCTCGA CGGTCCGGGC CTCCTGGTTG TCTTCGAAGT<br>GCGGCGGCCC GACCTGCAGG GGCAGGAGCT GCCAGGCCG GAGGACCAAC AGAAGCTTCA<br>V G G P O V D G D E V T R A E O N D E F 254 3480 AGCGGCGGT GACGCCGTTG AGGAGAGGAG AGGOCGCCAG CAGGACGGAG GTGGCCGCCC<br>TCGCCGGCCA CTGCGGCAAC TCCTCTCTCTCTCCTGCGGOTC GTCCTGGCTCC<br>Y R G T V G N L LP S A A L L V S T A A 234 3540 CCTGTTCGGG AGTCTTGTAG TACGGCAGCG GCTTCACGTT GCCCTGGTCG TCCATCACGC<br>GGACAAGCCC TCAGAACATC ATGCCGTCGC CGAAGTGCAA CGGGACCAGC AGGTAGTGCG<br>G Q Z P T K Y Y P L P K V N G Q D D R V V V 214 3600 CGAATGCGCG CATGGTCTCG TCGTCGACGT GCCGCTGCAG ACGGGTGAGG ATGTAGCCGG GCTTACGCOC GTACCAGAGC AGCAGCTOCA CGGCSACGTC TOCCCACTCC TACATCGGCC <sup>G</sup> P A <sup>R</sup> S <sup>T</sup> E <sup>D</sup> <sup>D</sup> <sup>V</sup> <sup>H</sup> R <sup>Q</sup> <sup>L</sup> R <sup>T</sup> <sup>L</sup> <sup>I</sup> <sup>Y</sup> <sup>G</sup> 194 3660 GGTTGAGCGC GTTGGCCGTG ATCCCGTCGG CTGCCCAGCG GCGGGCACCC ACCGTGAACA<br>CCAACTCGGCG CAACCGGGCAC TAGGGCAGGC GACGGGATCG CGCGCATTGGG TGGCACTTGT<br>P N L A N A T I G D A A W R R A G V T F 174 3720 GGACGTCGGC GGTCTTGGAC TGCCCGTAGG CCGCCCAGGG GTCGTACGGC CGCCGTGCGA<br>CCTGCAGCCG CCAGAACCTG ACGGGCATCC GGCGGGTCCC CAGCATGGCG GCGGCACGCT<br>L V D A T K S N G Y A N P D Y P R R A 154 AATGAGGATC TTCGAAGTCG AAGGGCGTGC CCAGGTGGGC GCCGGAGCTG ACGACCACGA<br>TTACTCCTAG AAGCTTCAGC TTCCCGCACG GGTCCACCG CGGCCTCGAC TGCTGGTG<br>F H P D E P D F P T G L H A G S S V V V 3780 134 3840 TGCGGGCGGA GCCGGCGTCC CGCAGAGCCG CGTGCAGGCC AGTGGCCAGA GCGAAGTGGC<br>ACGCCCGCCT CGGCCGCAGG GCGTCTCGGC GCACGTCGG GCACCGGTCT CGCTTCACGG<br>I R A S G A D R L A A H L G T A L A F H 114 3900 CGAGGTAGTT GGTGGCAAGC TGCATTTCCC AGCCGTACGG GGTGAGGGTG CGTGTGGGAA<br>GCTCCATCAA CCACCGTTCG ACGTAAAGGG TCGGCATGCC CCACTCCCAC GCACCCTT<br>G L Y N T A L O N E W G Y P T L T R T P 94 3960 GGGCCATGAT CCCCGCGTTC GCGACGAGGA TGTCGAGAGG CCCGCGCCAT GCCCGCGCGA<br>CCCGGTACTA GGGGCGCAGAG CGCTGCTCCT ACAGCTCTCC GGGCGCGTA CGGGC<br>L A S I G A N A V L I D L P G R W A R A 74 4020 AGCTGTCGAC GGAGGCCACA TCGGAAAGAT CGAGGGCCTC TGCGTGCACC CGGCCCGCAC<br>TCGACAGCTG CCTCGGGTGT AGCCTTTCTA GCTCCCGGAG ACGCACGTGG GCGGGGCTG<br>F S D V S A V D S L D L A B A H V R G A 54 4080 CGGCGGGGCC CGCCTCCTGG ACCAGGGGCT CGGCGGACTG CGGGTGGCGC GTGGCGATGG<br>GCCGGCGGCCG GCGGAOGACCC TGGTCCCCGA GCGGCCTGAC GCCGACGGGCG CACCGGTTACC<br>GCAAAAAEQ V L P ZAS 0 P E R TAI 34 4 140 TGACCTCCGC GCCGGCGGCG GCCAGGCCC GGACGGTTTC GGCGCCGAGA CCCGAGGCAC<br>ACTOGAGGCG GGGCGGCGC CGGTCCCGGG CCTOCCAAG CCGGCGTT GGGCTCCGTG<br>T V Z A G A A A L A R V T Z A G L G S A 14 4200 CGCCGGTGAC CACGGCACGG CGGCCGGTGA GATCCACGCC GTCGATCACC TCCTGCGCGC<br>GCGGCCACTG GTGCCGTGCC GCCGGCCACT CTAGGTGCGG CAGCTAGT<mark>GG AGG</mark>  $\overline{1}$ G G T V V A R R G T L D M<br>GGOCGTGAGA GTCGAAGGGG GTAGTCAGGA GGOCGGAGTC GTTTCGCGCG TGCCTGCCGC 4260 CGTCGGGATT CGTCGTCATG CGTCCACGCT ACGATGAGGC GACTGGATOC TCTATGACTG 4320 4380<br>5 AGAGTCCTTC TCCTTTTAGC GATCGTCCAA TCCC:~ TGAGTGTGGA TCCTCTTGAG rbs 5 D P L <sup>E</sup>

CCGAGTGCGG GGTCCGCTGG GTCGTCGGCG TCGATCGGCG CTCGGACGCC GTCTACTGCT 300 CGCGCCGGTG TGTGGTGAGG GCCTGGAGGC GCOGGAAGGA AACGTTCGGC GAACGGGCAC 360 AGTGACGGCA ACGCGAACAC ATCCTGAGTT CAAGTT¶IGT CAGAAACTGG ACATCCGTGG 420 TCATTCGTTG TGATCAGTGG ATTTACGGTC TGACTCTCGG TGATGAGCGG GTCGTGTCCG 480 CTCACGGAC GTCAGCTCCG TGCTGCGGCC CGTATGGGTG GTEGAACCACC ATGCCGGAGG rbs M P oar \_0- AGATCGACAG GGTTGGAAGT GTCAGTCAGC GCCGCTACGA GCAGATCGTG GCCGAGTTGC <sup>600</sup> <sup>I</sup> <sup>D</sup> <sup>R</sup> V <sup>G</sup> <sup>S</sup> <sup>V</sup> <sup>8</sup> <sup>Q</sup> <sup>R</sup> <sup>R</sup> Y E Q <sup>I</sup> <sup>V</sup> A <sup>Z</sup> L <sup>R</sup> <sup>24</sup> GGGAAGTGGT CGAACAGCAG ACACAGGGCT CGTTCACGAT CGGGAACCGC GCGCTGGAGA 660<br>E V V E Q Q T Q G S P T I G D R A L E I 44 TCGAGCCGAT GCGTGAGCGG GGCGGTGGCC AGCAGGTTGC TCCGGGCCAG GAGTTGTTCA 720 Z <sup>P</sup> M R <sup>Z</sup> R G G G 0 Q V A P G Q <sup>E</sup> L <sup>P</sup> T <sup>64</sup> CGGTCAGCGA GACOCTGCAC CGTCTTGCCG AGGACATCGG ACTGGCCTAC AGGACCGTGG 780 V <sup>S</sup> <sup>S</sup> <sup>T</sup> L <sup>S</sup> R L A <sup>Z</sup> D <sup>I</sup> G L A Y R <sup>T</sup> V 8 <sup>84</sup> AAAAGGCGAG GTGGACGGCG TCTCGGTGGC CGAAGGACAA GCGGCAGAAG GGCGTGTCGT 840 <sup>K</sup> A R W T A <sup>0</sup> R W <sup>P</sup> K <sup>D</sup> <sup>K</sup> R Q <sup>K</sup> <sup>G</sup> V <sup>S</sup> <sup>F</sup> <sup>104</sup> TTCGGGTCCA CAGAGTCCTG GCCCAGATCG CGGATGAGGC TGAGCGGTTC GCGACCATCG 900<br>R V H R V L A Q I A D E A B R F A T I A 124 ACCCCC CGCGGGCAAG ACGCGGTGGA CGGGGGACGA GGCCAACCGC AAGGTGGTC 960<br>Accession Caracter Tent CDE A NR K V G R 144  $\frac{1}{\sqrt{100}}$   $\frac{1}{\sqrt{100}}$  binding  $\frac{1}{\sqrt{100}}$  is the stationary conduction control conductions  $\frac{1}{\sqrt{100}}$  or  $\frac{1}{\sqrt{100}}$  is the stationary conduction  $\frac{1}{\sqrt{100}}$  or  $\frac{1}{\sqrt{100}}$  is the stationary conduc GGGACGAGGA CGTCGCTGCG GTGGTGACCA GCGACTTCCT CAAACGTCCG ACGGTGGCGG 1080 <sup>D</sup> <sup>8</sup> <sup>D</sup> <sup>V</sup> A <sup>A</sup> V <sup>V</sup> T <sup>8</sup> D <sup>F</sup> <sup>L</sup> K <sup>R</sup> <sup>P</sup> T <sup>V</sup> A <sup>A</sup> <sup>184</sup> CCAAGGTCTC CGACCAGGAC AAGGTCCGGG TGGTGGAGGA GTTCACCCGC GACGAGCGCG 1140 <sup>K</sup> <sup>V</sup> <sup>8</sup> D <sup>G</sup> <sup>K</sup> <sup>V</sup> <sup>R</sup> <sup>V</sup> <sup>V</sup> <sup>8</sup> <sup>P</sup> <sup>T</sup> <sup>R</sup> <sup>D</sup> <sup>8</sup> <sup>R</sup> <sup>V</sup> <sup>204</sup> TCGCCAGCCA GGTGACCACC GGTCTGCTGC GTCGCCCTGC AGTGGCGTAC AAGGCAATGA <sup>1200</sup> A <sup>S</sup> <sup>Q</sup> V T <sup>T</sup> G <sup>L</sup> <sup>L</sup> <sup>R</sup> <sup>R</sup> <sup>P</sup> <sup>S</sup> V <sup>A</sup> <sup>Y</sup> K A <sup>M</sup>S <sup>224</sup> GCGACGACAC CGCCCGCCAC CAGGTCAATC AGGCTCAGGT CGAGCGAGGT CGGCAGGCCC 1260 D D <sup>T</sup> <sup>A</sup> R H Q V N Q A <sup>0</sup> V <sup>Z</sup> R G R Q A R <sup>244</sup> GCGAGCACTT CGAGGACACC AATCCGGTTG CCCCGGCGGT CCGTCACATC GACCGGACGG 1320<br>E H F E D T N P V A P A V R H I D R T V 264 TGGAGTTCCT GGACCTGGTC ACCGCCTGTC ACTCGTTCGT GGCCGCGGCC GGCOGGGCGG 1380 <sup>E</sup> P <sup>L</sup> D L <sup>V</sup> T <sup>A</sup> <sup>C</sup> <sup>H</sup> S <sup>F</sup> <sup>V</sup> A A <sup>A</sup> G <sup>R</sup> A <sup>V</sup> <sup>284</sup> TCCCCGGGCT GCGTGATCGC ACCCTGGGCG AGGATGAACG CACCATTGTC CACGAGAACG 1440 P G L R D R T L G <sup>5</sup> D 8 R T <sup>I</sup> V <sup>S</sup> 8 <sup>S</sup> V 304 TGGCGAAGGT ACOGGCGACG CTCGACTGGA TCGAGACCGC GGTCGACACC GGCAAGGTCG 1500 A <sup>K</sup> <sup>V</sup> <sup>R</sup> A <sup>T</sup> L <sup>D</sup> <sup>W</sup> <sup>I</sup> <sup>E</sup> T <sup>A</sup> V <sup>D</sup> <sup>T</sup> G K <sup>V</sup> <sup>D</sup> ACATGGACGG CGAACTGGCT CGCATGCTGC **CACGGCGAGTA GCCGTGCCTC GCTCACGGCG** 1560<br>R D G K L A R M L R G K M P R S R R 6 rbe O 22 CACCCGAGGAGAGA CACCATCGGA CACCATCGGA CACCATCGGA CACCAGGCAGGA CACCAGGCAGGA CACCACCOG T<br>CACCATCGA A R R S A D T I R F V L F G A R<br>K G D A A R R S A D T I R F V L F G A R SCCGGCCGGC TCTGCGATGC ACCAGCTGAT CAGGGCCAGC GGGTTGTCGG ACAGTCAGGT 1680<br>P A G S A M B Q L I R A S G L S D S Q V 46 CAGATCGGGG CTOTCGGCCC TGCGCGACGA GGCGGCGGCC AAGGGCTGGC CACCGCTGAT 1740<br>R 8 G L 8 A L R D B A A A K G W P P L I 66 CTGGACCCGG ACCGACGGCT ATCAGCTCGA CGCTGAACGG GCGGCGCTGG AGTCATACGA 1800<br>W T R T D G Y Q L D A S R A A L S S Y G 86 ACGGGCGGTG GTCAGGGAGA AGCTGACCCA GTTCCOCCGG TTCATCACCG GGACCGTCAC 1860<br>R A V V R B K L T Q P R R P I T G T V T 106 CCCGCATGCC GCGGCCCACC CGAACGACAA GTGGGTCAAG CACATCGTCG CTCAGCTCAA 1920<br>P H A A A H P N D K W V K H I V A Q L N 126 CTCCATCGAA TCCACCCTCG ACCTCATCGC CAGTGCCTGA GCCTTCCTCG CGGGGGCCGA 1980<br>8 I 8 S T L D L I A 8 A \* CCTGCTCTGC CGGTCGGCCC CCGCCGCAGG TCAGTGACCGC GAGGAGAAAT GCGTCACCGC 2040 rbs M R (1)<br>CTCTATCCCT CGGACATGAC GGACGCAGAG TGGGCACTGG TGGAACCGCGCGCGGGAG TGGGCACTG TGGCGCAGE P 24<br>L Y P S D M T D A E W A L V B P L L P P 24 CCGGCCTGTG ACACGGCCCG CGGCGGGCGG CCGGAGAAAG ATCCCCGGGCG CGAGATCGTC 2160<br>P A C D T A R G G R P B K H P R R B I V 44 SACGCGATTC GATACGTTGT GGACACCGGC TGCAAATGGA GGGCCCTACC TGCGGACTTC 2220 CCTCCATGGA GGACAGTCTG GGGCTTCATG GCCCGCTGGG CGGCGGTCGG AGTCATCGGC 2280<br>P P W R T V W G F M A R W A A V G V I G 84 CAGCTCCGT ACGCCCTGGC CCAACGAATT CGCCGTCATGGCCGAGG GCCAAGAGCC 2340 Q <sup>L</sup> <sup>R</sup> <sup>D</sup> A <sup>L</sup> A Q <sup>R</sup> <sup>I</sup> <sup>R</sup> <sup>R</sup> <sup>D</sup> M G <sup>R</sup> <sup>G</sup> <sup>P</sup> R A <sup>104</sup> GTCGCGACGA TCATCGACTC TCAGTCGGTG AAAGCCGT CGACCGTCGG CAAGGACAGC 2400 V A T <sup>I</sup> <sup>I</sup> D <sup>8</sup> Q <sup>S</sup> V K A A <sup>8</sup> T V G K D <sup>8</sup> <sup>124</sup> CGCGGCTATG ACCGGAA AAGAATCAAC GGCGCAARC GGCACATGGT GGTCGACACC <sup>2460</sup> R G Y D A G X R <sup>I</sup> N G R X R H M V V D T 144 AAGGGCTGC CGCTGATGGT GATGGTCACC CCGGCCGACC TGAGCC CGCTCC <sup>2520</sup> K G L P LS V M V T P A D L D T 8 A V A <sup>164</sup> AAGGAAGTCC TCTTCCGACT TCGCCTGACG CACCCCGAGA TCACCCTCGT CTGGGCCGAC 2580 K B V L F R L R L T H P B I T L V W A D 184 TCCGCCTATG CGGGCAAGCT CGTGACCTGG GCAAAGAAGC ATCTGAACCT CACGATCAAG 2640<br>3 A Y A G K L V T W A K K H L M L T I K 204

ACCGTCAGCC GCCCGAAGGA CACCTCGGGC TGGGTTCTGT TGCCCCGCCG CTGGGTGGTC 2700<br>T V S R P K D T 8 G W V L L P R R W V V GAACGGAGCC TGGCCTGGAT GATGAACGCC CGCCGTCATG CCCGAGATTA CGAGCGGCTG 2760<br>Z R S L A N M M M A R R H A R D Y E R L 244 ATCCAGCACT CCGAGGCCCT GATCACCTGG GCAGCGATCA CGGTGATGAC CAAGCGTCTG 2820<br>I Q H 8 B A L I T W A A I T V M T K R L 264 ACCCGCACGG GTCCCACCGG CTGGTCGAAG AAGCCGAAGG CAACG CTCATCCCCG <sup>2880</sup> T R T G P T G W S <sup>K</sup> <sup>K</sup> <sup>P</sup> <sup>K</sup> A T A <sup>D</sup> <sup>S</sup> <sup>8</sup> <sup>P</sup> <sup>284</sup>

GACGTGTTCC TCATCGGCGG CGGCTTCTCC TTCGGCACCC GCGCCCAGGA GCTGCTGCTC<br>D V P L I G G G P S P G T R A Q B L L L GACAGGCTGC CCCCGATCGT CCATGTGCCC GCCGACACCC CGCACGCGGA GACGGTGCAG<br>D R L P P I V H V P A D T P H A B T V Q TGGGCACTGA CCGCCATCGA CCAGGAGCTG ACACATCGGC CAATGGCCTC CACCCTGATC W A L T A <sup>I</sup> D Q E L T <sup>H</sup> R <sup>P</sup> M A <sup>8</sup> T L <sup>I</sup> GCGGAGCACC TGGCCGTCAT CATGCTCGTT CATGTGCTGC GCCTGCACCT CGAACGCGCG<br>A E H L A V I M L V H V L R L H L B R A CCGCACGCGG TGTCAGGCTG GCTGGCCGGC CTCGCAGATC CCGTGGTCGC CACAGCGCTG<br>P H A V S G W L A G L A D P V V A T A L ACCTGCCTGC ACCGCGATCC GGCGCGTTCT TGGACCGTGG CCGACCTGGC CGACACCGCC<br>T C L H R D P A R S W T V A D L A D T A GCGGTGTCCC GTTCCACCCT GGCCGCCCGC TTCAAAGCCA CAGTCGGCCA AGGGCCATTG<br>A V S R S T L A A R F K A T V G Q G P L GAATACCTCA CGCGATGGCG GATCGAGCTC GCCGCCCGCC AGCTACGGGA AGGCAACGCA<br>E Y L T R W R I E L A A R Q L R E G N A ACACTCGCCT CCATCGCCCA CTCCGTGGGA TACGGATCCG AAAGCGCCCT CAGCGTCGCC<br>T L A S I A H 8 V G Y G 8 S S A L 8 V A TTCAAAAGGG TCCTGGGAAT GCCGCCGGGC GACTACCGCA AACATCCCAC GATGCCTTGA<br>P K R V L G M P P G D Y R K E P T M P \* TGCTTCGAGG AGACGCGTTA CCACAGCIGG GCAACCGATG CACTCTGCAT CACCATGCAT TOCITICSHIP INTERFERED ACCACCGTG AGGGCGCCTA CTTCG AACTATCGCT COGCTCCTA AACCACCGTG AGGGCGCCTA CTTCG AACTATCH

GATGTACTGA CCCTGCTGAA GACCCGCAGC CATCTCTCAG CGAGCCTGGT CGCGGAGGA D V L T L L K T R S H L S A S L V A G G CGCTGGGCCG TGCGGTTCGA CGCCCCCCGC GTGGTGAAGT TCAATGCCGT TCGCCGCGGC <sup>R</sup> W A <sup>V</sup> <sup>R</sup> P <sup>D</sup> A P <sup>R</sup> V <sup>V</sup> <sup>K</sup> P N A V <sup>R</sup> <sup>R</sup> G ACCTGCCAGC TGGACGTCGA CGGGATCGAC GAGCCGATCG ACCTGGCCGA GGGTGACTGC T C <sup>Q</sup> <sup>L</sup> E <sup>V</sup> <sup>D</sup> <sup>G</sup> <sup>I</sup> <sup>D</sup> G <sup>P</sup> <sup>I</sup> <sup>D</sup> L A S G <sup>D</sup> <sup>C</sup> TACCTCCTGA CACGGCCACG CTCCTTCACG CTCCGCAGOC ATCCGGAAAC TGCTCCTGTC Y L L T R P R <sup>S</sup>O T L R <sup>S</sup> D P S T A P V GATGGCGGAG TCGTGTTCGC AAGAGCCGAA GACGGCATCG CCAGAGCGGG CCAGGGCGAC <sup>D</sup> <sup>G</sup> <sup>G</sup> <sup>V</sup> <sup>V</sup> <sup>F</sup> <sup>A</sup> R A <sup>E</sup> <sup>D</sup> <sup>G</sup> <sup>I</sup> <sup>A</sup> R A <sup>G</sup> 0 <sup>G</sup> <sup>D</sup>

(GenBank Accession No. Z11519)

**i** 

FIG. 2. Nucleotide sequence and potential protein products of Tn4811 isolated from pLUS729. The nucleotide sequence between the putative duplicated target sequences (TGA) is shown. The imperfect terminal inverted repeats (TIR) and the putative ribosome-binding sequences (rbs) are boxed. The five putative coding sequences (ORF1 to ORF5) and their translation products are indicated. The asterisks denote translation termination codons. The putative motif A of an ATP and GTP binding sequence in the ORF1 product is underlined. For ORF4, which runs in the reverse orientation, both strands are shown.

ORF4	MLTGRRAVVTGGASGLGAETVRALAAAGA-	30
Per	AP SPVTTSPGSTASSPSGKKTLROG <b>VVVITGASSGLGLAAAKALA</b> ETGKW	100
NodG	MF-ELTGRKALVTGASGAIGGAIARVLHAOCA-	31
GdhA	MYTDLKDKVVVITGGSTGLGRAMAVRFGOEEA-	
		32
DekbR	MKHSVSSMNTSLSGKVAAITGAASGIGLECARTLLCAGA-	39
	$\cdot$	
ORF4	EVTIATRHPOSAEPLVOEAAAAGAGRVHAEALDLSDVASVDSFARAWRG-	79
Pcr	HVVMACROFLKASK-AAKAAGMADGSYTVMHLDLASLDSVRQFVDAFRRA	149
NodG	-- IVGLHGTQIEKLETL -- ATELGORVKLFPANLANRDEVKALGORAEAD	77
GdhA	KVVINYYNNEEEALDAKKEVEEAGGQAIIVQGDVTKEEDVVMLVOTAIKE	82
DekbR	KVV--LIDREGEKLNKL--VAELGONAFALQVDLVOADQVDNLLQGILOL	85
	a. $\cdot$ .	
ORF4	--PIDILVANAGIMALPTRT--LTPYGNIMOLATNYLCHFALATGLHAAL	125
Pcr	EMPLDVLVCMAAIYRPTARTPTFTADGHEMSVGVMHLGHFLLARLLMEDL	199
NodG	LEGVDILVNNAGITK-DGLFLHMADPDNDIVLEVNLTAMFRLTREITQQM	126
GdhA	FGTLDVMINNAGVEN-PVPSHELSLDNWNKVIDTNLTGAFLGSREATKYF	131
DekbR	TGRIDIFHANAGAYI-GGPVAEGDPDVNDRVLHININAAFRCVRSVLPHL	134
	**. $\mathcal{M}_{\mathrm{max}}$ $\star$ $\sim$ $\sim$ $\cdot$	
ORF4	RDAG--SARIVVVSSGAHLGTPFDFEDPHFAR------------------	155
Pcr	OKSDYPSRRMVIVGSITGNSNTLAGNVPPKASLGDLRGLAGGLSGASGSA	249
NodG		138
GdhA		
DekbR		144
		146
ORF4	-----RPYDPWAAYGOSKTADVLFTVGARRW--AADGITANALNPGYI--	
Pcr		196
	MIDGDESFDGAKAYKDSKVCNMLTMOEFHRRYHEETGITFSSLYPGCIAT	299
NodG	SVAGAIGNPGOTNYCASKAGMIGFSKSLAOEI-ATRNITVMCVAPGFIES	187
GdhA	SVHEMIPWPLFVHYAASKGGMXLMTETLALEY-APKGIRUMNIGPGAMMT	193
DekbR	VIAGVVIW--EPVYTASKFAVOAFVHTTRROV-AOYGVRVGAVLPGPVVT	193
	 $\mathbf{L}$	
ORF4	--LIRLORHVDDETTRAFGVMDDOGNVKPLPYYKTPEOGAATSVLLAASP	244
Pcr	TGLFREHIPLFRTLFPPFOKFVTKG------FVSEAESGKRLAOVVAEPV	343
NodG	--------AM-TDKL-NHKOKEKIMVAIPIHRMGTGTEVASAVA------	221
GdhA	--------PINAEKFADPEORADVESMIPMGYIGKPEEVAAVAA------	229
DekbR	--------ALLDOW-----PKAKMDEALADGSLMOPIEVAESVL------	224
	$\cdot$	
ORF4	LLNGVTGRYFEDNOEARTVEDGDVOPGGVAAHALDPEAADRLMEYGADTL	294
Pcr	LTKS--GVYWSWMKDSASFENQ------LSQEASDPEKARKVWELSEKLV	385
NodG	--------YLASDHAAYVTGOTIHVNGGAAM---------------[-----	245
GdhA	--------FLASSQASYVTGITLFADGGMTK-------------YPSFQA	258
DekbR	--------TMVTRSKNVTVRDIVILPNSVDL-------------------	247
ORF4	SAG 297	
Pcr	<b>GLA</b> 388	
NodG	--- 245	
GdhA	GRG 261	
DekbR	--- 247	

FIG. 3. Similarity of the ORF4 product to oxidoreductases. The protein product of ORF4 was aligned with the amino acid sequences of Pcr of barley (33), nodulation protein G (NodG) of R meliloti (13), glucose dehydrogenase (GdhA) of B. megaterium (19), and ribitol dehydrogenase (DekbR) of K. *aerogenes* (28) by using the CLUSTAL program of PC/GENE. A K-tuple value of 1, <sup>a</sup> gap penalty of 5, a window of 15, a filtering level of 1.5, an open gap cost of 10, and a unit gap cost of 10 were used. The bold characters denote homologous amino acids. Among the aligned sequences, perfectly conserved residues are indicated by an asterisk below, and well conserved residues are indicated by a dot. Pcr is most similar to the ORF4 product, with an identity score of 35% and <sup>a</sup> similarity of 48% by using the GAP program of the UWGCG package (gap weight of 3.0, length weight of 0.1, average match of 0.54, and average mismatch of 0.396). The first 74 amino acids of Pcr is supposedly a transit peptide (33). Pcr is shown starting from amino acid 51, whereas the others start at amino acid 1.

directly regulates a gene located elsewhere, all the regulatory polypeptides act within the intergenic region. It is most likely that the ORF5 product also regulates the divergent transcription of ORF4 and ORF5. The notion that ORF4 and ORF5 form <sup>a</sup> functional unit is supported by our finding of <sup>a</sup> 2.2-kb sequence in S. antibioticus containing highly conserved sequences of ORF4 and ORF5 in the same arrangement (45) (see below).

ORF1 to ORF3 as possible transposition proteins. Although no siguificant similarity to protein sequences in the data bases was observed for ORF1 to ORF3, <sup>a</sup> low level of similarity of the putative ORF1 and ORF2 products to some DNA-binding proteins and retroviral DNA polymerases was detected. ORF1 contained a consensus ATP/GTP-binding motif A  $(43)$  or P-loop  $(31)$ , A/G-X<sub>4</sub>-G-K-S/T, at residues 124 to 131 (Fig. 2). Proteins possessing this motif include several bacterial helicases (such as DnaB, RecB, Rho, and UvrD), viral proteins essential for replication (UL5 of herpes simplex virus, BBLF4 of Epstein-Barr virus, PS3 of human cytomegalovirus, and the product of gene 55 of varicellazoster virus), and the yeast PIF protein involved in mito-

ORF <sub>5</sub>	MDPLEDVLTLLKTR--SHLSASLVAGGRWAVRFDAPRVVKTNAVRRGT	46
RhaR	MAFCNMANLLNVFVRHIANNQLRSLAEVATVAHQLKLLKDDFFASDQQAV	50
RhaS	<b>MTVLHSVDFFPSGNASV</b>	17
	MAETOMDPLLPGYS--FNAHLVAGLTPIEANGY------LDFT-------	35
AraC		
VirF	MASLEIIKLEW--------	11
ORF <sub>5</sub>	C---OLEVDGIDEPI-DLAEGDCYLLTRPRSFTLRSDPETAPVDGGVVFA	92
RhaR	AVADRYPODVFAEHTHDFCELV--IVWRGNGLHVLMDRPYRITRGDLFYI	98
	AIEPRIPOADFPERHHDFHEIV--IVEHGTGIHVFNGOPYTITGGTVCFV	65
RhaS		
AraC	-IDRPLGMKGYI--------LN--LTIRGEGVINMQGKQFVCRPGDILLF	74
VirF	------ATPIFKVVENSQDGL--YILLQGQISWQNSSQTYDLDEGNMLFL	53
	₩	
ORF <sub>5</sub>	RAED--GIARAGOGDDVF-LIGGGTSFGTRAOELL-LDRLPPIVHVPADT	138
RhaR	HADDKESYAS---VNDLV-LONIIY-CPERLKLNLDWOGAIPGFNASAGO	143
<b>RhaS</b>	RDHDRHLYEH---TDNLC-LTMVLYRSPDRFOFLAGLNOLLPOELDGOYP	111
AraC	PPGEIHHYGRHPDASEWY-HQWVYFRPRAYWQEWLTWPAIFAO------T	117
VirF	R---RGSYAVRCGTKEPCOLLWIPLPGSFLSTFLHRFGSLLSEIRRDNAT	100
ORF5	PH---AETVOWALTAIDOELTERPMASTLIAEHLAVIMLVHVLRL-HLER	184
RhaR	PENRLGSMGMAQARQVIGQLEHESSQHVPFANEMAELLFGQLVML--LNR	191
	SEMEVNHSVLOOVROLVAOMEOOEGENDLPSTASREILFMOLLLL--LRK	159
RhaS		
AraC	GFFRPDEAHQPHFNELFGQIINAGQGEGRYSELLAINLLEQL--L--LRR	163
VirF	PKPLLIFNISPILSQSIQNLCAI-LERSDFPSVLTQLRIEELLLLLLAFSS	149
	$\ddot{\phantom{a}}$	
ORF <sub>5</sub>	<b>APHAVSGWLAGLADPVVATALTCLHRD-PARSWTVADLADTAAVSRSTLA</b>	233
RhaR	HRYTSDSLPPTSSETLLDKLITRLAASLKSP-FALDKFCDEASCSERVLR	240
RhaS	SSLO-ENL--ENSASRLNI.LLAWLEDHFADE-VNWDAVADOFSLSLRTLH	205
		212
AraC	MEAINESLHPPM-DNRVRDACOYISDHLADSNFDIASVAOHVCLSPSRLS	
VirF	OGALFLSALRHLGNRPEERLOKFMEENYLO-GWKLSKFAREFGMGLTTFK	198
	×.	
	h 'n $t -$	
ORF <sub>5</sub>	ARFKATVGOGPLEYLTRWRIELAAROLREGNATLASIAHSVGYGSESALS	283
RhaR	OOFROOTGMTINOXLROVRVCHAOYLLOESRLLISDISTECGFEDSNYFS	290
<b>RhaS</b>	ROLKOOTGLTPORYLNRLRLMKARHLLRHSEASVTDIAYRCGFSDSMHFS	255
AraC	HLFROOLGISVLSWREDORISOAKLLLSTTRMPIATVGRNVGFDDOLYFS	262
VirF	ELFGTVYGISPRAWISERRILYAHOLLLNGKMSIVDIAMEAGFSSQSYFT	248
	٠ ۰. $\sim$ $\sim$ $\sim$ $\sim$ $\sim$ $\ddot{\phantom{0}}$	
ORF <sub>5</sub>	<b>VAFKRVLOPPGDYRKHPTMP</b> 304	
RhaR	<b>VVFTRETGMTPSQWRHLNSQKD</b> 312	
	TLFRREFNWSPRDIROGRDGFLO 278	
RhaS		
AraC	<b>RVFKKCTGASPSEFRAGCE</b> 281	
VirF	271 OSYRRRFGCTPSQARLTKIATTG	
	计选择文件	

FIG. 4. Similarity of the ORF5 product to the AraC family of transcription regulators. The CLUSTAL program of PC/GENE was used to align the ORF5 product with four proteins with the highest similarity: the rhamnose operon transcription activators, RhaR and RhaS, of E. coli (39), the arabinose operon regulator AraC of E. carotovora (26), and the virulence regulon transcriptional activator VirF of Yersinia species (12). The parameters and notations used are as in Figure 3. The predicated helix-turn-helix (h <sup>t</sup> h) DNA-binding domains are indicated by the overline. AraC is most similar to the ORF5 product, with an identity score of 26% and a similarity of 50% by using the GAP program of the UWGCG package (gap weight of 3.0, length weight of 0.1, average match of 0.54, and average mismatch of 0.396).

chondrial DNA recombination (21). All these proteins are known or presumed to be involved in nucleic acid replication and/or recombination. The ORF1 product may take part in the replication and/or recombination process during transposition.

Furthermore, the predicted pIs of the ORF2 and ORF3 products were very high (10.9 and 11.2, respectively). This is consistent with the fact that many characterized transposase proteins are very basic (15). Thus, we tentatively suppose that these are genes involved in transposition and its regulation. The lack of homology between these ORFs and other transposon-encoded products is not uncommon. The putative transposition-related proteins encoded by Streptomyces transposable elements, ISJJO (7), IS117 (20), and IS493 (35), also showed no homology to other known transposases.

Putative transcriptional and translational signals. The basic features of Tn4811, including the distributions of restriction sites, potential coding sequences, and G+C content, are summarized in Fig. 5. Two distinct regions of relatively low G+C contents were discerned: one between ORF4 and ORF5 (53%  $G+C$ ) and the other at 170 to 60 bp upstream from ORF1 (50%  $G+C$ ). The low  $G+C$  is typical for putative promoter-containing regions in Streptomyces species. We propose that these regions contain the promoter sequences.

ORF3, ORF4, and ORF5 had a putative 5-bp ribosomebinding sequence (1, 4), GGAGG, located 4, 11, and <sup>6</sup> bp, respectively, upstream from the initiation codon. ORF1 had



FIG. 5. Overview of the organization of Tn4811. The top panel shows the unique cleavage sites for several restriction enzymes. The middle panel shows the G+C content averaged over 50 nucleotides. The two particularly low  $\tilde{G}$ +C regions proposed to be promoter regions are indicated by the gray horizontal bars. In the lower panel, solid arrows at the termini indicate the imperfect inverted repeats. The five potential coding sequences are shown by the open arrows together with their sizes (in amino acids) and putative functions. The high pI values of ORF2 and ORF3 are indicated.

<sup>a</sup> 3-bp GGA sequence <sup>5</sup> bp upstream, and ORF2 had <sup>a</sup> 4-bp GAGG sequence <sup>9</sup> bp upstream, as <sup>a</sup> potential ribosomebinding sequence. The initiation codon of ORF2 was separated by only 2 bp from the termination codon of ORFi (Fig. 2).

Occurrence of Tn4811. The 5.0-kb KpnI-SstI fragment that spanned more than 90% of Tn4811 was subcloned on pBluescript II  $KS(+)$  to yield pLUS791 and used as a hybridization probe in subsequent experiments. Southern hybridization (Fig. 6) revealed a single copy of Tn4811 in the genome of TK64, which had been genetically characterized, but not physically proven, to be plasmidless (23). Thus, the location of copy  $A$  cannot be certain. On the other hand, an additional copy of Tn4811 was present in wild-type S. lividans 1326 and its  $SLP2$ <sup>+</sup> derivatives (see legend to Fig. 6). We designated the copy of Tn4811 in TK64 copy  $A$  and designated the other copy  $B$ . Copy  $B$  appeared to cosegregate with SLP2. We have recently isolated SLP2 DNA and confirmed the presence of Tn4811 on it (10).

The restriction and hybridization patterns of the two copies of Tn4811 were indistinguishable. The restriction maps of the region spanning the two copies were determined by a series of Southern hybridization analyses (Fig. 6). Unexpectedly, the homology between the two copies extended further from the right end of Tn4811, whereas the sequences flanking the left ends were different. The homology to the right extended for about 10 kb and approximately coincides with the right end of SLP2 DNA (10).

Tn4811 in S. lividans appeared quite stable. During this and other studies, we have never detected its movement on the chromosome (by using the not-very-sensitive Southern hybridization assay) or to a plasmid other than the ones reported here. Tanaka et al., however, reported the transpositions of Tn4811 to plasmids (38) and the rearrangement of Tn4811-containing sequences in ethidium bromide-induced mutants of S. lividans 1326 (37).

Fifteen other Streptomyces species (Table 1) were probed for the presence of sequences homologous to Tn4811. Only S. antibioticus contained a sequence (less than 3.0 kb) of significant homology (data not shown). This S. antibioticus sequence, designated Vi, is homologous to the ORF4 and ORF5 sequences of Tn4811 (45).

The junction sequences of three Tn4811 insertions. A cosmid clone containing the copy  $B$  sequence was isolated from <sup>a</sup> library of SLP2 DNA (10) by hybridization with <sup>a</sup> Tn4811 probe. A cosmid clone containing the genomic copy A sequence was isolated from a library of TK64 (SLP2<sup>-</sup>). Junction fragments from these two insertions were subcloned in pBluescript II KS(+) (Table 2) and their nucleotide sequences were determined. These junction sequences together with those for copy  $A$  on pLUS729 are aligned and shown in Fig. 7.

Copy  $A$  of Tn4811 from the genomic DNA of TK64 and the copy from pLUS729 are identical in their terminal sequences, which is consistent with the fact that the copy on pLUS729 originated from the TK64 genome. Interestingly, the putative (duplicated) target sequence (TGA) of the insertion on TK64 genome is identical to that on pLUS729. Copy  $B$  isolated from SLP2 DNA is 41 bp shorter from the left terminus than copy A. Again, the right junction is flanked by the TGA triplet.

The insertion of Tn4811 into a TGA (opal) termination codon is not likely to affect the expression of the gene. This potentially allows Tn4811 to transpose with a reduced probability of damaging the host and being detected by genetic screening. In the insertional inactivation of melC on pLUS729, the putative target was an out-of-frame TGA triplet. If the insertion were at the (in-frame) termination a.





FIG. 6. Two copies of Tn4811. (a) Genomic DNA was digested with the restriction enzymes indicated and hybridized with the radioactively labeled 5.0-kb KpnI-SstI fragment from pLUS791. The sizes of the hybridizing fragments are indicated. All the S. lividans strains in Table <sup>1</sup> were tested, but only selected results are shown here. All the SLP2<sup>+</sup> strains gave the same hybridization pattern as 1326 (with both copies of  $Tn4811$ ), whereas all the SLP2<sup>-</sup> strains, like TK64, had copy  $A$  only. Hybridization to the 11.0-kb  $BcI$ fragments was very weak but real. The hybridizing 6.6-kb BclI fragment in TK20 DNA was due to incomplete digestion. (b) The restriction maps of regions spanning the two copies of Tn4811 based on the Southern hybridization data. Tn4811 is represented by the open bars, and the neighboring sequences are represented by the solid lines. The probe and the fragments that hybridized (sizes in kilobases) are indicated. Abbreviations for restriction sites are as in the legend to Fig. 1.

codon TGA, its presence might not have been detected. In this connection, TGA is by far the most frequently used termination codon in Streptomyces species (42). Analyses of more insertion sites will be necessary to determine the precise target specificity of Tn4811.

## ACKNOWLEDGMENTS

This work was supported by research grants from National Science Council to C.W.C. (NSC78-0412-BO10-08 and NSC79-0412- B010-08). C.W.C. was a recipient of a research award from the Medical Research and Advancement Foundation in memory of Chi-Shuen Tsou.

We thank David A. Hopwood and Stanley N. Cohen for helpful comments and suggestions for improvement of the manuscript, M.



FIG. 7. Junction sequences at the three insertion sites of Tn4811. The junction sequences of copy  $A$  in melC on pLUS729, copy  $A$  on the TK64 genome (cloned in pLUS796 and pLUS797), and copy B on SLP2 (cloned in pLUS798 and pLUS797) were determined. The relevant nucleotide sequences at these junctions are aligned. The Tn4811 sequences are shown in boldface type and numbered, and the flanking sequences are in plain type. The sequence shared by all three copies was shown only in one (middle) line. The imperfect terminal repeats of Tn4811 are boxed. The putative target sequences (TGA) are underlined.

Tanaka for exchange of information and discussion, and Helen M. Kieser and Ying-Chieh Tsai for providing the Streptomyces strains.

#### **REFERENCES**

- $\frac{1}{2}$   $\frac{1}{2}$  1. Baylis, H. A., and M. J. Bibb. 1988. Transcriptional analysis of coelicolor A3(2). Mol. Microbiol. 2:569-579.
	- 2. Beck, C. F., and R. A. J. Warren. 1988. Divergent promoters, a common form of gene organization. Microbiol. Rev. 52:318- 326.
	- 3. Bernan, V., D. Filpula, W. Herber, M. Bibb, and E. Katz. 1985. The nucleotide sequence of the tyrosinase gene from Streptomyces antibioticus and characterization of the gene product. Gene 37:101-110.
	- 4. Bibb, M. J., and S. N. Cohen. 1982. Gene expression in Streptomyces: construction and application of promoter-probe plasmid vectors in Streptomyces lividans. Mol. Gen. Genet. 187:265-277.
	- 5. Bibb, M. J., P. R. Findlay, and M. W. Johnson. 1984. The relationship between base composition and codon usage in bacterial genes and its use for the simple and reliable identification of protein-coding sequences. Gene 30:157-166.
	- 6. Bullock, W. O., J. M. Fernandez, and J. M. Short. 1987. XL1-B: a high efficiency plasmid transforming recA Escherichia coli strain with beta-galactosidase selection. BioTechniques 5:376- 379.
	- 7. Burton, C. J., and K. F. Chater. 1987. Nucleotide sequence of IS110, an insertion sequence of Streptomyces coelicolor A3(2). Nucleic Acids Res. 15:7053-7065.
	- 8. Chater, K. F., D. J. Henderson, M. J. Bibb, and D. A. Hopwood. 1988. Genetic flux in Streptomyces coelicolor and other streptomycetes and its possible relevance to the evolution of mobile antibiotic resistance determinants, p. 8-42. In A. J. Kingston, S. M. Kingsman, and K. F. Chater (ed.), Transposition. Cambridge University Press, Cambridge.
	- 9. Chater, K. F., D. A. Hopwood, T. Kieser, and C. J. Thompson. 1982. Gene cloning in Streptomyces. Curr. Top. Microbiol. Immunol. 96:69-95.
	- 10. Chen, C. W., T.-W. Yu, Y.-S. Lin, H. Kieser, and D. A. Hopwood. Submitted for publication.
	- 11. Chou, C.-F., and C. W. Chen. Submitted for publication.
	- 12. Cornelis, G., C. Sluiters, C. L. de Rouvroit, and T. Michiels. 1989. Homology between VirF, the transcriptional activator of

the Yersinia virulence regulon, and AraC, the Escherichia coli arabinose operon regulator. J. Bacteriol. 171:254-262.

- 13. Debelle, F., and S. R. Sharma. 1986. Nucleotide sequence of Rhizobium meliloti RCR2011 genes involved in host specificity of nodulation. Nucleic Acids Res. 14:7453-7472.
- 14. Fickett, J. W. 1982. Recognition of protein coding regions in DNA sequences. Nucleic Acids Res. 10:5303-5318.
- 15. Galas, D. J., and M. Chandler. 1989. Bacterial insertion sequences, p. 109-162. In D. E. Berg and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.
- 16. Gribskov, M., J. Devereux, and R. R. Burgess. 1984. The codon preference plot: graphic analysis of protein coding sequences and prediction of gene expression. Nucleic Acids Res. 12:539- 549.
- 17. Harrison, S. C., and A. K. Aggarwal. 1990. DNA recognition by proteins with the helix-turn-helix motif. Annu. Rev. Biochem. 59:933-969.
- 18. Haüsler, A., A. Birch, W. Krek, J. Piret, and R. Hütter. 1989. Heterogeneous genomic amplification in Streptomyces glaucescens: structure, location and junction sequence analysis. Mol. Gen. Genet. 217:437-446.
- 19. Heilman, H. J., H. J. Maegert, and H. G. Gassen. 1988. Identification and isolation of glucose dehydrogenase genes of Bacillus megaterium M1286 and their expression in Escherichia coli. J. Biochem. 174:485-490.
- 20. Henderson, D. J., D. J. Lydiate, and D. A. Hopwood. 1989. Structural and functional analysis of the mini-circle, a transposable element of Streptomyces coelicolor A3(2). Mol. Microbiol. 3:1307-1318.
- 21. Hodgman, T. C. 1988. A new superfamily of replicative proteins. Nature (London) 333:23. (Erratum, 333:578.)
- 22. 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. John Innes Foundation, Norwich, England.
- 23. Hopwood, D. A., T. Kieser, H. M. Wright, and M. J. Bibb. 1983. Plasmids, recombination, and chromosomal mapping in Streptomyces lividans 66. J. Gen. Microbiol. 129:2257-2269.
- 24. Katz, E., C. J. Thompson, and D. A. Hopwood. 1983. Cloning and expression of the tyrosinase gene from Streptomyces antibioticus in Streptomyces lividans. J. Gen. Microbiol. 129:2703- 2714.
- 25. Kieser, H. M., D. J. Henderson, C. W. Chen, and D. A. Hopwood. 1989. A mutation of Streptomyces lividans which prevents intraplasmid recombination has no effect on chromosomal recombination. Mol. Gen. Genet. 220:60-64.
- 26. Lei, S. P., H. C. Lin, L. Heffernan, and G. Wilcox. 1985. araB gene and nucleotide sequence of the araC gene of Erwinia carotovora. J. Bacteriol. 164:717-722.
- 27. Miller, J. H. 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- 28. Morris, R. M., D. H. Williams, G. G. Midwinter, and B. S. Hartley. 1974. A mass-spectrometric sequence study of the enzyme ribitol dehydrogenase from Klebsiella aerogenes. Biochem. J. 141:701-713.
- 29. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- 30. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463-5467.
- 31. Saraste, M., P. R. Sibbald, and A. Wittinghofer. 1990. The P-loop-a common motif in ATP- and GTP-binding proteins. Trends Biochem. Sci. 15:430-434.
- 32. Schrempf, H., A. Kessler, V. Bronneke, W. Dittrich, and M. Betzler. 1989. Genetic instability in streptomycetes, p. 133-140. In C. L. Hershberger, S. W. Queeners, and G. Hegeman (ed.), Genetics and molecular biology of industrial microorganisms. American Society for Microbiology, Washington, D.C.
- 33. Schulz, R., K. Steinmueller, M. Klass, C. Forreiter, S. Rasmussen, C. Hiller, and K. Apel. 1989. Nucleotide sequence of a cDNA coding for the NADPH-protochlorophyllide oxidoreductase (PCR) of barley (Hordeum vulgare L.) and its expression in Escherichia coli. Mol. Gen. Genet. 217:355-361.
- 34. Sheldon, P. S., R. G. 0. Kekwick, C. Sidebottom, C. G. Smith, and A. R. Slabas. 1990. 3-Oxyacyl-(acyl-carrier protein) reductase from avocado (Persea americana) fruit mesocarp. Biochem. J. 271:713-720.
- 35. Solenberg, P. J., and S. G. Burgett. 1989. Method for selection of transposable DNA and characterization of <sup>a</sup> new insertion sequence, IS493, from Streptomyces lividans. J. Bacteriol. 171:4807-4813.
- 36. Tanaka, M. (Toho University, Funabashi, Chiba, Japan). 1991. Personal communications.
- 37. Tanaka, M., N. Takabe, F. Kato, and Y. Koyama. 1991. Analysis of mutants and genetic rearrangements in the region containing IS456(Tn4811), abstr. P1-056. Abstr. 8th Int. Symp. Biol. Actinomycetes. University of Wisconsin, Madison.
- 38. Tanaka, M., H. Tanaka, F. Kato, and Y. Koyama. 1988. IS-like element from Streptomyces lividans, abstr. P5-20, p. 157. Abstr. 7th Int. Symp. Biol. Actinomycetes. Academic Societies Japan, Tokyo.
- 39. Tobin, J. F., and R. F. Schleif. 1987. The positive regulation of the Escherichia coli L-rhamnose operon is mediated by the products of tandemly repeated regulatory genes. J. Mol. Biol. 196:789-799.
- 40. Tsai, J. F.-Y., and C. W. Chen. Unpublished data.
- 41. Tsai, J. F.-Y., and C. W. Chen. 1987. Isolation and characterization of Streptomyces lividans mutants deficient in intraplasmid recombination. Mol. Gen. Genet. 208:211-218.
- 42. Wada, K., S. Aota, R. Tsuchiya, F. Ishibashi, T. Gojobori, and T. Ikemura. 1990. Codon usage tabulated from the GenBank genetic sequence data. Nucleic Acids Res. 18:2367-2411.
- 43. Walker, J. E., M. Saraste, M. J. Runswick, and N. J. Gay. 1982. Distantly related sequenced in the  $\alpha$ - and  $\beta$ -subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and <sup>a</sup> common nucleotide binding fold. EMBO J. 1:945-951.
- 44. Williams, S. T., M. Goodfellow, G. Alderson, E. M. H. Wellington, P. H. A. Sneath, and M. J. Sackin. 1983. Numerical classification of Streptomyces and related genera. J. Gen. Microbiol. 129:1743-1813.
- 45. Yu, T.-W., and C. W. Chen. Submitted for publication.