JOURNAL OF BACTERIOLOGY, Oct. 1988, p. 4634–4651
0021-9193/88/104634-18$02.00/0
Copyright © 1988, American Society for Microbiology
Vol. 170, No. 10

Complete Nucleotide Sequence of the *Streptomyces lividans* Plasmid pIJ101 and Correlation of the Sequence with Genetic Properties

KEVIN J. KENDALL[1] AND STANLEY N. COHEN[1,2]*

Departments of Genetics[1] and Medicine,[2] Stanford University School of Medicine, Stanford, California 94305

Received 29 April 1988/Accepted 24 June 1988

The complete nucleotide sequence of the multicopy *Streptomyces* plasmid pIJ101 has been determined and correlated with previously published genetic data. The circular DNA molecule is 8,830 nucleotides in length and has a G+C composition of 72.98%. The use of a computer program, FRAME, enabled identification in the sequence of seven open reading frames, four of which, *tra* (621 amino acids [aa]), *spdA* (146 aa), *spdB* (274 aa), and *kilB* (177 aa), appear to be genes involved in plasmid transfer. At least two of the above genes are predicted to be transcribed by known promoters that are regulated in *trans* by the products of the *korA* (241 aa) and *korB* (80 aa) loci on the plasmid. The segment of the plasmid capable of autonomous replication contains one large open reading frame (*rep*; 450 aa) and a noncoding region presumed to be the origin of replication. Four other small (less than 90 aa) open reading frames are also present on the plasmid, although no function can be attributed to them. The sequence of the pIJ101 replication segment present in several widely used cloning vectors (e.g., pIJ350 and pIJ702) has also been determined, so that the complete nucleotide sequences of these vectors are now known.

Streptomyces spp. are gram-positive, mycelial soil bacteria having a complex life cycle that involves differentiation and sporulation. Plasmids are widespread in *Streptomyces* species and potentially may have special features that allow them to exist in these structurally and developmentally complex organisms. In addition, *Streptomyces* DNA has an unusually high G+C content, typically about 73% (14); thus, the organization of DNA sequences required for autonomous replication and for the control of gene expression may well be different from that found in procaryotic organisms such as *Escherichia coli*, which has an average G+C composition of 50%.

pIJ101 is a circular multicopy *Streptomyces* plasmid of 8.9 kilobases (kb). Originally found in *S. lividans* ISP 5434, it has a broad host range, and a number of widely used cloning vectors have been derived from it (16, 19, 21). More is presently known about the basic biology of pIJ101 than is known for any other *Streptomyces* plasmid. Its essential replication functions have been localized to a 2.1-kb DNA segment (21), and other loci affecting plasmid maintenance and plasmid transfer have been described (16, 20, 21). Despite its relatively small size, pIJ101 is capable of transferring itself to plasmidless recipient cells at a frequency approaching 100% and in doing so can mobilize the host chromosome at a frequency of 10^{-3} (21). Previous studies (20) have identified four loci involved in plasmid transfer and two other loci that control the expression of transfer-related functions.

We report here the complete nucleotide sequence of pIJ101 and correlate this with previously published reports describing genetic loci on the plasmid (20, 21).

MATERIALS AND METHODS

Bacterial strains and plasmids. The *Streptomyces* plasmids pIJ101 (21) and pIJ350 (21) originally were obtained from the John Innes Institute, Norwich, England. Both plasmids were introduced individually into the *S. lividans* strain TK64 *pro-2 str-6* SLP2$^-$SLP3$^-$ (17) by transformation, and cells derived from single transformants were used for this study to ensure that the DNA used for sequencing was of clonal origin. The M13 sequencing vectors mp18 and mp19 (31) were used throughout this study. Some pIJ101-derived DNA fragments were subcloned by using the plasmid vector pUC19 (31) before further manipulation. The *E. coli* strain JM109 [*recA1 endA1 gyrA96 thi hsdR17 supE44 relA1* λ^- Δ(*lac-proAB*) (F′ *traD36 proAB lacI*^{q} *Z*Δ*M15*)] (31) was used as host for the M13 and pUC19 vectors.

DNA isolations and manipulations. Standard cloning techniques and plasmid DNA isolation protocols were used for both *E. coli* (22) and *S. lividans* (16). Occasionally, restriction endonuclease-generated DNA fragments to be sequenced were isolated from low-melting-point agarose gels as described previously (20). Single-stranded M13 DNA sequencing templates were prepared essentially as described by Messing (23) except that the polyethylene glycol-precipitated phage pellets were resuspended in TE containing 0.5% sodium dodecyl sulfate to facilitate phage lysis.

DNA sequencing. M13 clones were sequenced by a modification of the method of Sanger et al. (26). A 1-μg amount of single-stranded M13 template DNA was annealed with 2.5 ng of the 17-mer universal sequencing primer (New England BioLabs, catalog no. 1211) at 65°C for 30 min in a total volume of 10 μl of 7 mM Tris hydrochloride (pH 7.5)–60 mM NaCl–7 mM β-mercaptoethanol. Then 10 μCi of [^{32}P]dATP (Amersham, catalog no. PB.10384) and 1 U of freshly diluted DNA polymerase I Klenow fragment (Boehringer Mannheim) were added to this mixture, and 2.5 μl of the resulting solution was placed on the lip of each of four tubes containing 2 μl of the appropriate dNTP^0 mixes. The dNTP^0 mixes contained 10 mM Tris hydrochloride (pH 7.5), 20 mM $MgCl_2$, and deoxynucleoside triphosphate (dNTP) and dideoxy-NTP (ddNTP) nucleotides as follows. dATP^0 was 100 nM dGTP, 100 nM dCTP, 100 nM dTTP, and 75 nM ddATP; dGTP^0 was 2.5 nM dGTP, 100 nM dCTP, 100 nM dTTP, and 38 nM ddGTP; dCTP^0 was 100 nM dGTP, 2.5 nM dCTP, 100 nM dTTP, and 20 nM ddCTP; dTTP^0 was 100 nM dGTP, 100 nM dCTP, 1.25 nM dTTP, and 225 nM ddTTP. The solutions were mixed to start the reaction by centrifugation, and

* Corresponding author.

incubation was allowed to proceed at 55°C for 15 min. After this time, 2 μl of chase solution (2 mM each all four dNTPs) was added, and the reaction mixes were incubated for a further 15 min at 55°C. The reactions were terminated by the addition of 4 μl of 95% formamide–25 mM EDTA containing 0.15 mg of bromophenol blue and 0.15 mg of xylene cyanol per ml and boiled for 2 min prior to loading on a polyacrylamide sequencing gel.

Sequencing gels were 0.35 mm thick and were run with a Bethesda Research Laboratories S2 sequencing apparatus. The gels were prepared essentially as described by Bankier et al. (1), except that freshly deionized 40% (wt/vol) formamide was routinely incorporated into the gels to alleviate problems of compression caused by G+C-rich secondary structures. Runs of 90 min and 4 h on a 6% gel at 45 mA (2,300 to 3,000 V) were usually sufficient to resolve 200 to 400 nucleotides per sequence.

Computer-assisted sequence analysis. Data base searches and some sequence analysis were performed with the Bionet National Computing Resource (Intelligenetics, Mountain View, Calif.). Other analyses and the FRAME program (5) were implemented with standard algorithms on a Commodore Amiga 1000 computer.

RESULTS

Sequencing strategy. The sequence of pIJ101 was determined by using modifications of the dideoxy chain termination method (26) as described in Materials and Methods. Plasmid DNA was partially digested with the restriction endonucleases *Sau*3A, *Hpa*II, *Taq*I, and *Hin*PI, the randomly produced fragments were cloned by insertion into the appropriate site of M13mp18 or M13mp19, and 150 such clones were subjected to sequencing. The nucleotide sequences of the clones were then assembled into a single long sequence with the aid of a computer. Subsequently, any gaps remaining in the sequence were filled by analysis of specific DNA clones generated by suitable restriction enzymes that cleaved near the sequence of interest. The complete sequence was determined for both strands of the plasmid, and each base pair was sequenced an average of six times.

Because of the high G+C content of *Streptomyces* DNA (14), several modifications were made to the original sequencing protocol to overcome the problems of polymerase "stalling" and band compression on gels because of the high stability of G+C-rich secondary structures. The most significant changes made were to incubate the sequencing reactions at 50 to 55°C in a low-salt (12.5 mM NaCl) buffer and to incorporate formamide (up to 40%) into the polyacrylamide gels.

The complete nucleotide sequence of pIJ101 is presented in Fig. 1. The plasmid is a circular DNA molecule of 8,830 base pairs (bp) having a G+C composition of 72.98%. In this and all other figures, the sequence has been split at the unique *Bam*HI site that was previously considered base 1 (21). Also shown in the figure are the predicted amino acid sequences of the most significant postulated open reading frames (ORFs) (see Results), the most significant regions of dyad symmetry (see Results), and the location of certain restriction endonuclease cleavage sites.

Analysis of the nucleotide sequence of pIJ101 for ORFs. Because of the high G+C composition of *Streptomyces* DNA, protein-encoding ORFs tend to have a biased codon usage such that coding sequences have about 70 mol% G+C in the first position of the codon triplet, about 50 mol% G+C in the second position, and over 90 mol% G+C in the third position (5). This observation has been exploited by the computer program FRAME (5), which graphically displays the G+C composition of the first, second, and third bases of the triplet codons over a window of *t* codons. The graphic output of a program similar to FRAME for pIJ101 is presented in Fig. 2.

In this figure, the heavy line is the plot of the G+C composition at every third base with a window of 50 codons starting at base 1 of the plasmid. Similarly, the thin line is the plot starting at base 2, and the dotted line is the plot starting at base 3. Regions of the plasmid that do not encode proteins would not be expected to have any bias in G+C composition due to codon usage, and thus all three plots would be expected to lie close together. Such regions can be seen centered around bases 1500 and 7400. On the other hand, in protein-coding regions the plots should diverge such that the plot corresponding to the second base of the codon triplets would have a much lower average G+C composition than expected and the plot corresponding to the third base would have an average G+C composition of 80 to 100%. In Fig. 2, the clearest example of such a region is the DNA segment believed to encode the *korA* gene product (20). The plots within this region can be seen to diverge at about base 5900 and to converge at about 6750. The third plot (dotted line) averages about 70% G+C and thus corresponds to the first position of the codon triplets. Suitable start and stop codons are present at positions 6031 and 6754, respectively (graphically represented above the plot), suggesting that a protein of 242 amino acids is translated from left to right on the sequence. Combined with previous genetic evidence for the existence and localization of *korA* (20), these data provide powerful, but not conclusive, evidence that a protein product is encoded by this region. Similar interpretations can be made for other regions of the sequence, as discussed below. Note that most of the predicted ORFs would be translated from right to left on the sequence. Statistics concerning the ORFs selected from Fig. 1 are summarized in Table 1.

(i) Possible ORFs in the essential replication region. It has been shown previously that the 2.1-kb *Sac*II fragment of pIJ101 (bp 8491 to 1972) is capable of autonomous replication in several *Streptomyces* species (21). The frame plot indicates that this region could encode two proteins. The largest of these, here designated *rep*, can be seen starting at about the *Not*I site at bp 1128, reading right to left, and terminating just after the *Apa*I site at bp 8592. (Note that this 456-codon frame crosses the *Bam*HI site at which the circular plasmid was linearized.) Inserts into or truncations of this region abolish the replication ability of pIJ101-derived plasmids (21; K. J. Kendall, unpublished observations). A second possible ORF is contained within the *Sac*II fragment that has been designated orf56, as it could encode a 56-amino-acid protein. The FRAME plot for this reading frame (centered at about base 1800) indicates that the second and third codons deviate significantly from the average G+C content of the region, and suitable start and stop codons are positioned at the points of plot divergence and convergence. However, no function can be assigned to this region from the results of previous genetic studies, and thus it is unknown whether this ORF actually encodes a functional protein product.

(ii) ORFs involved in plasmid transfer. (a) Repressor proteins. Previous studies demonstrated the existence of two loci, *korA* and *korB*, that are required in *trans* to control the lethal effects of two other loci, *kilA* and *kilB* (20). *korA* was localized to a 1.1-kb *Sal*I-*Bcl*I fragment (coordinates 5634 to 6756). *korB* was localized to an 800-bp *Bcl*I-*Fnu*DII frag-

```
     BamHI                          SmaI   XhoI
1    GGATCCTCGTTGCCGTCCTTGCCTTCGGCGGCCCGGGTCGCCTCGAGGTCGAGGGCGCGGCGGGTGACCGCGTGCCATCCGTCCTCGGTCACGGCGACCCCGGCCCGCAGCTCCCCGCCG
     CCTAGGAGCAACGGCAGGAACGGAAGCCGCCGGGCCCAGCGGAGCTCCAGCTCCCGCGCCGCCCACTGGCGCACGGTAGGCAGGAGCCAGTGCCGCTGGGGCCGGGCGTCGAGGGGCGGC
      AspGluAsnGlyAspLysGlyGluAlaAlaArgThrAlaGluLeuAspLeuAlaArgArgThrValAlaHisTrpGlyAspGluThrValAlaValGlyAlaArgLeuGluGlyGlyAsp

                                                                                                                  SmaI
121  TCGGCGTCGGCCGCCAGGAGCAGATCGAGGTCGTCGGCCTCGGTGTCGCCGCCGTCGAGCCCGAGCATCTGCCGCAGGTAGCGGGTCCATTCGATGGCCCGGCGTCCCCGGGTTGCCCGC
     AGCCGCAGCCGGCGGTCCTCGTCTAGCTCCAGCAGCCGGAGCCACAGCGGCGGCAGCTCGGGCTCGTAGACGGCGTCCATCGCCCAGGTAAGCTACCGGGCCGCAGGGGCCCAACGGGCG
      AlaAspAlaAlaLeuLeuLeuAspLeuAspAspAlaGluThrAspGlyGlyAspLeuGlyLeuMetGlnArgLeuTyrArgThrTrpGluIleAlaArgArgGlyArgThrAlaArgGlu

241  TCGTACTCGTGCCAGCGCGAGAGGTTCCACTCCAGCGAGCCGACCCCGGCGGCGTCGTCCTCGGTCATGCCGCCGGTCAGGTCCCCGATCCGTCCGAGGAGTTCGAACGGGGCGACGTTC
     AGCATGAGCACGGTCGCGCTCTCCAAGGTGAGGTCGCTCGGCTGGGGCCGCCGCAGCAGGAGCCAGTACGGCGGCCAGTCCAGGGGCTAGGCAGGCTCCTCAAGCTTGCCCCGCTGCAAG
      TyrGluHisTrpArgSerLeuAsnTrpGluLeuSerGlyValGlyAlaAlaAspAspGluThrMetGlyGlyThrLeuAspGlyIleArgGlyLeuLeuGluPheProAlaValAsnGly

361  CCGCCGGTCGCCGTCTTGAGGTCGGCGCGGGCGAGTTCGAGGGCGGGCGCCTTCCCGTCCTGGGTCTTGGCGATGTACTCGGCGAGGTCGTTGGCGTCGCGCTCGGTCTCCAGCCGCTTG
     GGCGGCCAGCGGCAGAACTCCAGCCGCGCCCGCTCAAGCTCCCGCCCGCGGAAGGGCAGGACCCAGAACCGCTACATGAGCCGCTCCAGCAACCGCAGCGCGAGCCAGAGGTCGGCGAAC
      GlyThrAlaThrLysLeuAspAlaArgAlaLeuGluLeuAlaProAlaLysGlyAspGlnThrLysAlaIleTyrGluAlaLeuAspAsnAlaAspArgGluThrGluLeuArgLysPhe

       SalI                                       FspI
481  AAGTCGACGCCGTGCCGGTCGTCGGGCGTGAAGGCGGGGTTGACCTTGCGCAGGGCGGCGGTCCACACGGACCGCCAGTGCCCCTGCCACTCGTCGAGCGCGGCGCCGGTCGGCTCGAAG
     TTCAGCTGCGGCACGGCCAGCAGCCCGCACTTCCGCCCCAACTGGAACGCGTCCCGCCGCCAGGTGTGCCTGGCGGTCACGGGGACGGTGAGCAGCTCGCGCCGCGGCCAGCCGAGCTTC
      AspValGlyHisArgAspAspProThrPheAlaProAsnValLysArgLeuAlaAlaThrTrpValSerArgTrpHisGlyGlnTrpGluAspLeuAlaAlaGlyThrProGluPheThr

                                                                   PvuI
601  GTGGCGACGATCTGCTTCGCGGACCGCTCCCCCTCGGTCCGGCCGCCGACCAGGACGATCGCGTGGATGTGCGGGTGCCAGCCGTTGATCTGCCCCACGGTGACTTCGGTCGCGCGGATC
     CACCGCTGCTAGACGAAGCGCCTGGCGAGGGGGAGCCAGGCCGGCGGCTGGTCCTGCTAGCGCACCTACACGCCCACGGTCGGCAACTAGACGGGGTGCCACTGAAGCCAGCGCGCCTAG
      AlaValIleGlnLysAlaSerArgGluGlyGluThrArgGlyGlyValLeuValIleAlaHisIleHisProHisTrpGlyAsnIleGlnGlyValThrValGluThrAlaArgIleMet

                                                                                                     BclI
721  ATGCCGACGTACCCGATCCGGTCTCGGATGCCCTCGCGGTCGGCGGCCCGGTGCCCGTCCTTGGCCCGGCGTCCGGCCCACGTGCCGCCCGTGATCAGTCGCTGGTAGGCGCCCGGCCGC
     TACGGCTGCATGGGCTAGGCCAGAGCCTACGGGAGCGCCAGCCGCCGGGCCACGGGCAGGAACCGGGCCGCAGGCCGGGTGCACGGCGGGCACTAGTCAGCGACCATCCGCGGGCCGGCG
      GlyValTyrGlyIleArgAspArgIleGlyGluArgAspAlaAlaArgHisGlyAspLysAlaArgArgGlyAlaTrpThrGlyGlyThrIleLeuArgGlnTyrAlaGlyProArgArg

841  CGGGGGCTGTCCGGCGTCTTCCGGGTGCCCTGGAGGGCGTCCATGAGGTCCGCGAGCCGGTCCGTGTGCCCATGGCGGGCCGTGAAGGTGACCAGGTAGGCGGTCCCCCCGCGCTTGATC
     GCCCCCGACAGGCCGCAGAAGGCCCACGGGACCTCCCGCAGGTACTCCAGGCGCTCGGCCAGGCACACGGGTACCGCCCGGCACTTCCACTGGTCCATCCGCCAGGGGGGCGCGAACTAG
      ProSerAspProThrLysArgThrGlyGlnLeuAlaAspMetLeuAspAlaLeuArgAspThrHisGlyHisArgAlaThrPheThrValLeuTyrAlaThrGlyGlyArgLysIleTrp

961  CACTCGACCACGGCGGCGGTGATCTCCTCGGCCCGCTTGTGCCGGATCGTGGCGGCGCAGACCGGGCAGAGCCAGATCCGCCCGCACCGCATCAGGCCCAGGACCACGGACGTTCCGGCC
     GTGAGCTGGTGCCGCCGCCACTAGAGGAGCCGGGCGAACACGGCCTAGCACCGCCGCGTCTGGCCCGTCTCGGTCTAGGCGGGCGTGGCGTAGTCCGGGTCCTGGTGCCTGCAAGGCCGG
      GluValValAlaAlaThrIleGluGluAlaArgLysHisArgIleThrAlaAlaCysValProCysLeuTrpIleArgGlyCysArgMetLeuGlyLeuValValSerThrGlyAlaAla

                                                          NotI
1081 GCCGTCTGGGCGACGATCACGCCGGAGGCAGGGTCCATCAGGGCGCGGCCGCAGCCCTTGCACGCGGCGTCCCCGCTGATCCGCCACAGCGTCCGGCGGCGGCTGTACCGGGCGGCTTTC
     CGGCAGACCCGCTGCTAGTGCGGCCTCCGTCCCAGGTAGTCCCGCGCCGGCGTCGGGAACGTGCGCCGCAGGGGCGACTAGGCGGTGTCGCAGGCCGCCGCCGACATGGCCCGCCGAAAG
      ThrGlnAlaValIleValGlySerAlaProAspMet
                                    rep

1201 CGCAGTCGGGCAGCCTCGGTCCGCGACGTGCTTCCTACTTCCCAGAGGCTGTCGCCTCTCGGGCTCTCCCCATCCACCCCGTCCGGAGAAACCGCAGGTCGGAGGGGTGCGGGAAACTCT
     GCGTCAGCCCGTCGGAGCCAGGCGCTGCACGAAGGATGAAGGGTCTCCGACAGCGGAGAGCCCGAGAGGGGTAGGTGGGGCAGGCCTCTTTGGCGTCCAGCCTCCCCACGCCCTTTGAGA

                                                                      PvuI
1321 GTTGTTTCTTTCCCAAGGTGTTCGCTTTTGCCTCGGGCGGCATCTCGCGTCACACGCGCGATCGCCCGCTTCGCTGCCATCCGGCAGCGGTCTGAGCAGTAGATACGCGGCCGTTTGCCC
     CAACAAAGAAAGGGTTCCACAAGCGAAAACGGAGCCCGCCGTAGAGCGCAGTGTGCGCGCTAGCGGGCGAAGCGACGGTAGGCCGTCGCCAGACTCGTCATCTATGCGCCGGCAAACGGG
```

FIG. 1. Complete nucleotide sequence of pIJ101. The sequence is numbered starting at the first base of the recognition sequence of the unique *Bam*HI site. The translated amino acid sequences of the predicted proteins encoded by the plasmid (see text) are shown. The arrows indicate the positions of the most significant regions of dyad symmetry, as discussed in the text.

ment, which has since been further reduced to the 316-bp *Apa*I-*Spe*I fragment (bp 6929 to 7245) (K. J. Kendall and D. Stein, unpublished results). Subsequent work has shown that *korA* and *korB* repress transcription from certain pIJ101 promoters, including those responsible for transcription of the *kilA* and *kilB* loci (D. Stein, K. J. Kendall, and S. N. Cohen, unpublished results). As indicated above, the FRAME plot of the *korA* region indicated a likely ORF reading from left to right. The FRAME plot of the *korB* region suggested the existence of an ORF starting at bp 7208, terminating at bp 6968, and encoding an 80-amino-acid protein.

(b) Other transfer-related ORFs. Previously we identified four genetic loci at which frame-disrupting inserts affected the ability of pIJ101 derivatives to transfer into recipient strains lacking pIJ101 (20). One of these loci, designated

```
                                                 Apa I
1441 GGTGTGTGGGCAATTGCGGTCCCGCAGTGGCAGCGGGGCCCGGCGGGCCGATCTGGCAATGCCTCGGCATCGCTCCGTACTCTGGGCACGAGCAACGTTCCTGTCTCGCCCGGCTAAGGG
     CCACACACCCGTTAACGCCAGGGCGTCACCGTCGCCCCGGGCCGCCCGGCTAGACCGTTACGGAGCCGTAGCGAGGCATGAGACCCGTGCTCGTTGCAAGGACAGAGCGGGCCGATTCCC

1561 GCGCGAGTCTGGGAGCGGACGGGTCGGAGGTGCGAAGTCCGGCCCGTTGCTCTTTGGTCTGGTGGGAATCCTGGCACCAATCGGGCCAGAGGTTCCCTCCGCCACTCCCGACGCCCCTTG
     CGCGCTCAGACCCTCGCCTGCCCAGCCTCCACGCTTCAGGCCGGGCAACGAGAAACCAGACCACCCTTAGGACCGTGGTTAGCCCGGTCTCCAAGGGAGGCGGTGAGGGCTGCGGGGAAC

                                                                                                  Bal I
1681 GGGCTGGTGTGACTTGGAGGGCCGAAGAGAGCCCCGCCGGGTATCCGGCGGGGCTTTGACGTGCGGTCAGTGCGTGTGTCGGCGAGCGATGGCCACGAGGCCCTGGAAGCCGAGCGGTCC
     CCCGACCACACTGAACCTCCCGGCTTCTCTCGGGGCGGCCCATAGGCCGCCCCGAAACTGCACGCCAGTCACGCACACAGCCGCTCGCTACCGGTGCTCCGGGACCTTCGGCTCGCCAGG
                                                                       ***HisThrHisArgArgAlaIleAlaValLeuGlyGlnPheGlyLeuProGly

1801 GGCGAAGTCGGCCCAGTCGCAACCGGGCTCAGCGCAGTGGGCGGACCAGCCACCGCCGTTGGGGTCCTGGACCAGGTTCACGGTCCCCTCGGTCAGGCGTCCGTCGAAGTCGGTCATGGT
     CCGCTTCAGCCGGGTCAGCGTTGGCCCGAGTCGCGTCACCCGCCTGGTCGGTGGCGGCAACCCCAGGACCTGGTCCAAGTGCCAGGGGAGCCAGTCCGCAGGCAGCTTCAGCCAGTACCA
     AlaPheAspAlaTrpAspCysGlyProGluAlaCysHisAlaSerTrpGlyGlyGlyAsnProAspGlnValLeuAsnValThrGlyGluThrLeuArgGlyAspPheAspThrMet
                                                                                                                   orf56

                                                       Sac II
1921 CGGTCTCCTGGTGGGTGGGGGCGGGGCGCCAGCACGAAGTGCCGGCGCCCCGCGGGGGTTGGTCGGGTCAGGCGCCGAACCGGCGGGCGGCGGCGGCGACCAGGCCGTCGGCGGCGGCCA
     GCCAGAGGACCACCCACCCCCGCCCCGCGGTCGTGCTTCACGGCCGCGGGGCGCCCCCAACCAGCCCAGTCCGCGGCTTGGCCGCCCGCCGCCGCCGCTGGTCCGGCAGCCGCCGCCGGT
                                                                       ***AlaGlyPheArgArgAlaAlaAlaAlaValLeuGlyAspAlaAlaAlaMet

2041 TGGCGCGGTCGCGGTCGGTGGTGAGGGCGGTGCGGTCGGCGGCGGCCCGCAGGTCGTAGGCCGCTTGGGCGGCGGCGGTCGCTGCGGGGGCGAGGGCGGGGGCGAGCACCGACACGGTGG
     ACCGCGCCAGCGCCAGCCACCACTCCCGCCACGCCAGCCGCCGCCGGGCGTCCAGCATCCGGCGAACCCGCCGCCGCCAGCGACGCCCCCGCTCCCGCCCCCGCTCGTGGCTGTGCCACC
      AlaArgAspArgAspThrThrLeuAlaThrArgAspAlaAlaAlaArgLeuAspTyrAlaAlaGlnAlaAlaAlaThrAlaAlaProAlaLeuAlaProAlaLeuValSerValThrThr

2161 TGAGGGGCGCGGTGATCGCGGAGCGGGTGGCGTGGGACTCGGTGCGGGCCGCCTCGTACGCCTCGGGGGAGGCGCCGGTCAGGCGCAGGTCCTCGCGCACCCACATGGCGCGGCGGTGGT
     ACTCCCCGCGCCACTAGCGCCTCGCCCACCGCACCCTGAGCCACGCCCGGCGGAGCATGCGGAGCCCCCTCCGCGGCCAGTCCGCGTCCAGGAGCGCGTGGGTGTACCGCGCCGCCACCA
      LeuProAlaThrIleAlaSerArgThrAlaHisSerGluThrArgAlaAlaGluTyrAlaGluProSerAlaGlyThrLeuArgLeuAspGluArgValTrpMetAlaArgArgHisAsp

2281 CGGCGAGGGCGGCGGCGAGGGCGGCGACGGCCTGGACGGCCGCGTCGCGGCGGGAGTCGTCGCGGGTGGTGCGGGCGGTGCGGTGCTGGAGGAGTCCGGCGAGGCCGGTCCCGGCGAGCG
     GCCGCTCCCGCCGCCGCTCCCGCCGCTGCCGGACCTGCCGGCGCAGCGCCGCCCTCAGCAGCGCCCACCACGCCCGCCACGCCACGACCTCCTCAGGCCGCTCCGGCCAGGGCCGCTCGC
      AlaLeuAlaAlaAlaLeuAlaAlaValAlaGlnValAlaAlaAspArgArgSerAspAspArgThrThrArgAlaThrArgHisGlnLeuLeuGlyAlaLeuGlyThrGlyAlaLeuThr

                                                                                                   Fsp I
2401 TGCCGATCACGGCAATGAGCGTGGTCACCATGTGAGCCCCCTGGCGTCGTGTCCGTCTGCCTACGTGTATCAGTCTGACACGCACGTGTCAGGTTGCGCAATGGGTAGGCCCCGCCGGTT
     ACGGCTAGTGCCGTTACTCGCACCAGTGGTACACTCGGGGGACCGCAGCACAGGCAGACGGATGCACATAGTCAGACTGTGCGTGCACAGTCCAACGCGTTACCCATCCGGGGCGGCCAA
      GlyIleValAlaIleLeuThrThrValMet
                 kilB

                                                                                                 Sal I
2521 TCCGGCGGGGCCATCCGTCATGCGGCGGTGCTGAGTCGGGCGGCGAGCTGGTGGGTGGCGGCGAGGATGACGCGGATCTCCGCGCCGCCGTCGACCGCCGCGAGGGCGGGGGCCGGGGCG
     AGGCCGCCCCGGTAGGCAGTACGCCGCCACGACTCAGCCCGCCGCTCGACCACCCACCGCCGCTCCTACTGCGCCTAGAGGCGCGGCGGCAGCTGGCGGCGCTCCCGCCCCCGGCCCCGC
                        ***AlaAlaThrSerLeuArgAlaAlaLeuGlnHisThrAlaAlaLeuIleValArgIleGluAlaGlyGlyAspValAlaAlaLeuAlaProAlaProAlaPro

                                                                          Pvu I
2641 GGTGCCGGCGGATTGGTCGGGGTCGGGCGGCGGCGGAGGGCGACGGCCAGGACCATGGCGGCGATCGCCACGGTGGCGAGGTAGGCGAGGGCGATCACGCGGCCACCTCCCTCGTGACGC
     CCACGGCCGCCTAACCAGCCCCAGCCCGCCGCCGCCTCCCGCTGCCGGTCCTGGTACCGCCGCTAGCGGTGCCACCGCTCCATCCGCTCCCGCTAGTGCGCCGGTGGAGGGAGCACTGCG
       AlaProProAsnThrProThrProArgArgArgLeuAlaValAlaLeuValMetAlaAlaIleAlaValThrAlaLeuTyrAlaLeuAlaIleVal
                                                                                                      ***AlaAlaValGluArgThrValArg
                                                                                                  orf66

                    Pst I                                                                    Sal I
2761 GGCGCCAGGACGCCTGCAGCCGTGCCTCCCCGGCGGAGTAGCCCATGTCCCGGAAGCGTCCGGCGGCGACCCGGTACGACCCGGAGTCGACCAGGCCGGACATGATCCCGTCGATCTCCT
     CCGCGGTCCTGCGGACGTCGGCACGGAGGGGCCGCCTCATCGGGTACAGGGCCTTCGCAGGCCGCCGCTGGGCCATGCTGGGCCTCAGCTGGTCCGGCCTGTACTAGGGCAGCTAGAGGA
      ArgTrpSerAlaGlnLeuArgAlaGluGlyAlaSerTyrGlyMetAspArgPheArgGlyAlaAlaValArgTyrSerGlySerAspValLeuGlySerMetIleGlyAspIleGluGlu
```

```
2881 CGTTGCTGAGCGGGACCACGTCGGCCGACCCCTCGATGGCGGCCGGGACGGGCTCCGTGGGTCCCCAGACGGGCAGGCGCCAGGAGGGCGGAGACGTGGGCGTGACACGAGAGGTGACGC
     GCAACGACTCGCCCTGGTGCAGCCGGCTGGGGAGCTACCGCCGGCCCTGCCCGAGGCACCCAGGGGTCTGCCCGTCCGCGGTCCTCCCGCCTCTGCACCCGCACTGTGCTCTCCACTGCG
      AsnSerLeuProValValAspAlaSerGlyGluIleAlaAlaProValProGluThrProGlyTrpValProLeuArgTrpSerProProSerThrProThrValArgSerThrValCys

                                                                                                         Sma I
                                  Sal I                                                          Sma I     Apa I
3001 AGGTCAGAGCGCGTGCCGTGCCCTCGCCCTCGCGGGCCTCCTCCAGCGTCGACCAGGCCGACGCCAGCTCCCCGGCGGTCTCGGCCTGGACCCGGGCGACCCGGGCCCCGGCGCCGGTCA
     TCCAGTCTCGCGCACGGCACGGGAGCGGGAGCGCCCGGAGGAGGTCGCAGCTGGTCCGGCTGCGGTCGAGGGGCCGCCAGAGCCGGACCTGGGCCCGCTGGGCCCGGGGCCGCGGCCAGT
      ThrLeuAlaArgAlaThrGlyGluGlyGluArgAlaGluGluLeuThrSerTrpAlaSerAlaLeuGluGlyAlaThrGluAlaGlnValArgAlaValArgAlaGlyAlaGlyThrVal

3121 CGGCCTCCAGCCGGGTGACCTCCGTACCGGCCTCGGCCCGGAGCTGAGCGCGGGCCACGGCCGCCCTGTCGCGGGCGTCCTGCTGGATGTCGCGGATCTGGTCCAGGGCGGTCGGGGTGA
     GCCGGAGGTCGGCCCACTGGAGGCATGGCCGGAGCCGGGCCTCGACTCGCGCCCGGTGCCGGCGGGACAGCGCCCGCAGGACGACCTACAGCGCCTAGACCAGGTCCCGCCAGCCCCACT
      AlaGluLeuArgThrValGluThrGlyAlaGluAlaArgLeuGlnAlaArgAlaValAlaAlaArgAspArgAlaAspGlnGlnIleAspArgIleGlnAspLeuAlaThrProThrLeu

                       Apa LI
3241 GCGCGGTCCGCTCCCAGAGCCCGTGCACGAGCCAGAGCGCCTTGGCGGCGAGGGGGAGCCAGGCGACCGCGAGCCAGGCCCCGGCCGACTCCTCACCCAGCGCGTGCGCGACCAGGACGC
     CGCGCCAGGCGAGGGTCTCGGGCACGTGCTCGGTCTCGCGGAACCGCCGCTCCCCCTCGGTCCGCTGGCGCTCGGTCCGGGGCCGGCTGAGGAGTGGGTCGCGCACGCGCTGGTCCTGCG
      AlaThrArgGluTrpLeuGlyHisValLeuTrpLeuAlaLysAlaAlaLeuProLeuTrpAlaValAlaLeuTrpAlaGlyAlaSerGluGluGlyLeuAlaHisAlaValLeuValGly

                                                                                          Bal I
3361 CGGTCGCCACGGCGCCGAATCCCCACCCGACGCCGGTGATCGCGCGGCTGTGGTCGCCCCGCGCGGCGAGGCGCCGCTCGTACGCGAGGGTGGCCAGCCATCCGCCGTCGAGCCCGAGGC
     GCCAGCGGTGCCGCGGCTTAGGGGTGGGCTGCGGCCACTAGCGCGCCGACACCAGCGGGGCGCGCCGCTCCGCGGCGAGCATGCGCTCCCACCGGTCGGTAGGCGGCAGCTCGGGCTCCG
      ThrAlaValAlaGlyPheGlyTrpGlyValGlyThrIleAlaArgSerHisAspGlyArgAlaAlaLeuArgArgGluTyrAlaLeuThrAlaLeuTrpGlyGlyAspLeuGlyLeuGly

3481 CGACGACCAAGGCAACGGCCCACGGCATGGCCTCCCCGAGCCACATCACGATCACGACCACGGTCAGGACCATGGACACCGCCGTCATGGCGACGGCCGGGGCGGCGGTCATCGAACGCT
     GCTGCTGGTTCCGTTGCCGGGTGCCGTACCGGAGGGGCTCGGTGTAGTGCTAGTGCTGGTGCCAGTCCTGGTACCTGTGGCGGCAGTACCGCTGCCGGCCCCGCCGCCAGTAGCTTGCGA
      ValValLeuAlaValAlaTrpProMetAlaGluGlyLeuTrpMetValIleValValValThrLeuValMetSerValAlaThrMetAlaValAlaProAlaAlaThrMetSerArgLys

              Bcl I
3601 TCTTCTCCATGATCACTTCCCCTTCCGGGCCTTGAGCGAGATGGACAGGCCGACCCCGGCGGGGGCGGCGGCGGCGAGCGCGGTCGCCGTGGTGGCGGCGGTCTGGAGCAGGAGGCAGAG
     AGAAGAGGTACTAGTGAAGGGGAAGGCCCGGAACTCGCTCTACCTGTCCGGCTGGGGCCGCCCCCGCCGCCGCCGCTCGCGCCAGCGGCACCACCGCCGCCAGACCTCGTCCTCCGTCTC
      LysGluMet
               ***LysGlyLysArgAlaLysLeuSerIleSerLeuGlyValGlyAlaProAlaAlaAlaAlaLeuAlaThrAlaThrThrAlaAlaThrGlnLeuLeuLeuCysLeu
      spdB

3721 CGTCATGACGACGCCGACCGAGCCGCCAGCGGCGACCAGCGCCAGGAGGATCGGGCCGGTCCAGTCGCGGGCCTCGGCCTGGTGGTGGACGACGACGACCGTGCGGGGGTCGGTGCCGTC
     GCAGTACTGCTGCGGCTGGCTCGGCGGTCGCCGCTGGTCGCGGTCCTCCTAGCCCGGCCAGGTCAGCGCCCGGAGCCGGACCACCACCTGCTGCTGCTGGCACGCCCCCAGCCACGGCAG
     ThrMetValValGlyValSerGlyGlyAlaAlaValLeuAlaLeuLeuIleProGlyThrTrpAspArgAlaGluAlaGlnHisHisValValValValThrArgProAspThrGlyAsp

                                                 Fsp I
3841 CGGGATCAGGTGGGCCGGAATGTGACCGGCCCTCATCTGCGCGTCGGGGGTGCGCATCAGGCCGACACCTCCCCGACCTGCGTCACGACGGACAGGGAGCCGTCCTCGGACCGCAGGACC
     GCCCTAGTCCACCCGGCCTTACACTGGCCGGGAGTAGACGCGCAGCCCCCACGCGTAGTCCGGCTGTGGAGGGGCTGGACGCAGTGCTGCCTGTCCCTCGGCAGGAGCCTGGCGTCCTGG
     ProIleLeuHisAlaProIleHisGlyAlaArgMetGlnAlaAspProThrArgMet
                                                     ***AlaSerValGluGlyValGlnThrValValSerLeuSerGlyAspGluSerArgLeuValGlu
                                                spdA

                 Pvu II                                                                                 Pvu I
3961 TCGCCCGCGTCGAGCAGCTGCTTGACCGCCTTCGACACGGAGCCCTTGTTGATGCCGGTCACGGTGGCGACGTCGGCGACGGTCGTAGCGCCCGTGCCGATCGCGGCTGCGACCTTCTCC
     AGCGGGCGCAGCTCGTCGACGAACTGGCGGAAGCTGTGCCTCGGGAACAACTACGGCCAGTGCCACCGCTGCAGCCGCTGCCAGCATCGCGGGCACGGCTAGCGCCGACGCTGGAAGAGG
       GlyAlaAspLeuLeuGlnLysValAlaLysSerValSerGlyLysAsnIleGlyThrValThrAlaValAspAlaValThrThrAlaGlyThrGlyIleAlaAlaAlaValLysGluArg

                                         Sac II
4081 CGGTTGGTCGGCGCCTTGGTCGGCTGGGCCGGGACCTCAGCCGCGGGGGCGGTCTCCTTCACCAGCCGGAGCGGGGCCGGGGCGGAGGCGCCGGCACTTCGTGCAGGCGACTCCTGGCGA
     GCCAACCAGCCGCGGAACCAGCCGACCCGGCCCTGGAGTCGGCGCCCCCGCCAGAGGAAGTGGTCGGCCTCGCCCCGGCCCCGCCTCCGCGGCCGTGAAGCACGTCCGCTGAGGACCGCT
       AsnThrProAlaLysThrProGlnAlaProValGluAlaAlaProAlaThrGluLysValLeuArgLeuProAlaProAlaSerAlaGlyAlaSerArgAlaProSerGluGlnArgArg

                                                                                                                 Sma I
4201 CGCCAGACCGGCCGGTCGGGGAGCGCGATCACGTCGGCCGGGGAGAACGCGCGGGTGTTGATCGGGTGAGGCTGCACCTTCGGGCCGGACCGGAGCATCGCAACCCCGGGCATGGGCAGC
     GCGGTCTGGCCGGCCAGCCCCTCGCGCTAGTGCAGCCGGCCCCTCTTGCGCGCCCACAACTAGCCCACTCCGACGTGGAAGCCCGGCCTGGCCTCGTAGCGTTGGGGCCCGTACCCGTCG
       TrpValProArgAspProLeuAlaIleValAspAlaProSerPheAlaArgThrAsnIleProHisProGlnValLysProGlySerArgLeuMetAlaValGlyProMetProLeuGlu
```

4321 TCGTGCGCGTGCCAGCCCTTCTCCGTCGCGTCCTCGCCGAACACCACGCGGGACTCCCCGGAGGTGCTGAGCGCGAGGGCGGCCCGGTAGGTGATCTGCGCGCTGATCTGCGGGTCGATG
AGCACGCGCACGGTCGGGAAGAGGCAGCGCAGGAGCGGCTTGTGGTGCGCCCTGAGGGGCCTCCACGACTCGCGCTCCCGCCGGGCCATCCACTAGACGCGCGACTAGACGCCCAGCTAC
HisAlaHisTrpGlyLysGluThrAlaAspGluGlyPheValValArgSerGluGlySerThrSerLeuAlaLeuAlaAlaArgTyrThrIleGlnAlaSerIleGlnProAspIleGly

4441 CCGCCCTTGGCGTCCATGGTCGGCTTCTGCGTGGCCCAGATCAGGATGATCTCGGCCGCCCGTGCCATCCGGGCCAGCGTGCTCAGGTTCTCCATGATCCGGGACCAGTCCGGGTCACCC
GGCGGGAACCGCAGGTACCAGCCGAAGACGCACCGGGTCTAGTCCTACTAGAGCCGGCGGGCACGGTAGGCCCGGTCGCACGAGTCCAAGAGGTACTAGGCCCTGGTCAGGCCCAGTGGG
GlyLysAlaAspMetThrProLysGlnThrAlaTrpIleLeuIleIleGluAlaAlaArgAlaMetArgAlaLeuThrSerLeuAsnGluMetIleArgSerTrpAspProAspGlyPro

Sma I Bal I
4561 GGCTCTTCCTTCGACCCCTTCGCCCGGGTCTTCTTGGCCATGGCGATGACCTCGGCGCCCTCGTCGATGAACACCGTGATCCGGGGCCGCTCCGGGCTGATCTGGATCACGTCCTGGCCG
CCGAGAAGGAAGCTGGGGAAGCGGGCCCAGAAGAACCGGTACCGCTACTGGAGCCGCGGGAGCAGCTACTTGTGGCACTAGGCCCCGGCGAGGCCCGACTAGACCTAGTGCAGGACCGGC
GluGluLysSerGlyLysAlaArgThrLysLysAlaMetAlaIleValGluAlaGlyGluAspIlePheValThrIleArgProArgGluProSerIleGlnIleValAspGlnGlyArg

Pvu I
4681 CGCGGGATCAGTTCAAGCCGCTCGTGCATCTCCTCGACCAGCTCGTCGGTCACGTCCAGGACGTCTTCGATCGAGATGGCCGTGCGAGCCCGGTGCTGCCAGTTGATCGCCTCGACCCGC
GCGCCCTAGTCAAGTTCGGCGAGCACGTAGAGGAGCTGGTCGAGCAGCCAGTGCAGGTCCTGCAGAAGCTAGCTCTACCGGCACGCTCGGGCCACGACGGTCAACTAGCGGAGCTGGGCG
ProIleLeuGluLeuArgGluHisMetGluGluValLeuGluAspThrValAspLeuValAspGluIleSerIleAlaThrArgAlaArgHisGlnTrpAsnIleAlaGluValArgLys

Sal I
4801 TTGGGGTCGACGACGACCAGGCGGTGATCGGCGTACTCCGAGCCTTCCGCCAGCAGGGCGCGGGTGGACCAGGACTTGCCCGAGCCGGACGTACCGGCGATCAGCATCCGGCGCCCGAGC
AACCCCAGCTGCTGCTGGTCCGCCACTAGCCGCATGAGGCTCGGAAGGCGGTCGTCCCGCGCCCACCTGGTCCTGAACGGGCTCGGCCTGCATGGCCGCTAGTCGTAGGCCGCGGGCTCG
ProAspValValValLeuArgHisAspAlaTyrGluSerGlyGluAlaLeuLeuAlaArgThrSerTrpSerLysGlySerGlySerThrGlyAlaIleLeuMetArgArgGlyLeuPro

Sal I Bcl I
4921 GGCACCTGCACCGGCTCGCCGGTCACCGTGTCGACTCCCCACGGGGCGCCGGGCGTCCAGCCGGTCAGGTCGATCCCGTCGGCCGCGCTGCGGGTCCGCAGCGTGATCACGGCGCGGTCG
CCGTGGACGTGGCCGAGCGGCCAGTGGCACAGCTGAGGGGTGCCCCGCGGCCCGCAGGTCGGCCAGTCCAGCTAGGGCAGCCGGCGCGACGCCCAGGCGTCGCACTAGTGCCGCGCCAGC
ValGlnValProGluGlyThrValThrAspValGlyTrpProAlaGlyProThrTrpGlyThrLeuAspIleGlyAspAlaAlaSerArgThrArgLeuThrIleValAlaArgAspGly

5041 CCGTGCGATCCGGCCTTGATCTCCATGCGGAGGTCGGTCCGGGCTCCGAGCAGGGCGCGGATCTCCTCGTGCTTGGCCTTGAAGGCGGACGGCTTCCACCGGCCGTCCAGGCGCACCGTG
GGCACGCTAGGCCGGAACTAGAGGTACGCCTCCAGCCAGGCCCGAGGCTCGTCCCGCGCCTAGAGGAGCACGAACCGGAACTTCCGCCTGCCGAAGGTGGCCGGCAGGTCCGCGTGGCAC
HisSerGlyAlaLysIleGluMetArgLeuAspThrArgAlaGlyLeuLeuAlaArgIleGluGluHisLysAlaLysPheAlaSerProLysTrpArgGlyAspLeuArgValThrThr

5161 GTGACCAGCCCGGCCGGGGTGACCTGAACCGGCGTCGTCACCGTGCCTACGAGGCCGCGCTCGTCGGCGTGCTGAGCCCAGTACGACGGGTCGAGCCGCTCGACCAGGCGCCGCTCCTCC
CACTGGTCGGGCCGGCCCCACTGGACTTGGCCGCAGCAGTGGCACGGATGCTCCGGCGCGAGCAGCCGCACGACTCGGGTCATGCTGCCCAGCTCGGCGAGCTGGTCCGCGGCGAGGAGG
ValLeuGlyAlaProThrValGlnValProThrThrValThrGlyValLeuGlyArgGluAspAlaHisGlnAlaTrpTyrSerProAspLeuArgGluValLeuArgArgGluGluAla

5281 GCTGACAGTTCGTCCGCGACCGCGACCGCGACCTTGCGGCGGCCGATCACCAGCGCGGCGACGTTGGCCGCGATGAGTGCGAGCGAGCCCGCGACCGGCCAGGCCCCGGCCTGGTCCCAG
CGACTGTCAAGCAGGCGCTGGCGCTGGCGCTGGAACGCCGCCGGCTAGTGGTCGCGCCGCTGCAACCGGCGCTACTCACGCTCGCTCGGGCGCTGGCCGGTCCGGGGCCGGACCAGGGTC
SerLeuGluAspAlaValAlaValAlaValLysArgArgGlyIleValLeuAlaAlaValAsnAlaAlaIleLeuAlaLeuSerGlyAlaValProTrpAlaGlyAlaGlnAspTrpGly

5401 CCGAGAGCGGAGATGCCGGCGCCCGGCACCGCGTGCATGGCGTGCGCCTGAACGGCTCGCGCCCGGAGGGTCGCCGGGTAGTCCTTGCGCGCGGCCTTGAGCTTGGTCTTCGCCTGGTCG
GGCTCTCGCCTCTACGGCCGCGGGCCGTGGCGCACGTACCGCACGCGGACTTGCCGAGCGCGGGCCTCCCAGCGGCCCATCAGGAACGCGCGCCGGAACTCGAACCAGAAGCGGACCAGC
LeuAlaSerIleGlyAlaGlyProValAlaHisMetAlaHisAlaGlnValAlaArgAlaArgLeuThrAlaProTyrAspLysArgAlaAlaLysLeuLysThrLysAlaGlnAspArg

Sal I Sal I
5521 CGGTGCGTGCGGGCCGCCTTGTCCGCCGTGCGAGCCGCCCGACGCACGGTCGACAGCGGGTTCTTGGACGCCGCCCGCGCGGCCGTCCGCTGGCTCTTGGCCGTGGCCGCCGTCGACCGC
GCCACGCACGCCCGGCGGAACAGGCGGCACGCTCGGCGGGCTGCGTGCCAGCTGTCGCCCAAGAACCTGCGGCGGGCGCGCCGGCAGGCGACCGAGAACCGGCACCGGCGGCAGCTGGCG
HisThrArgAlaAlaLysAspAlaThrArgAlaAlaArgArgValThrSerLeuProAsnLysSerAlaAlaArgAlaAlaThrArgGlnSerLysAlaThrAlaAlaThrSerArgAla

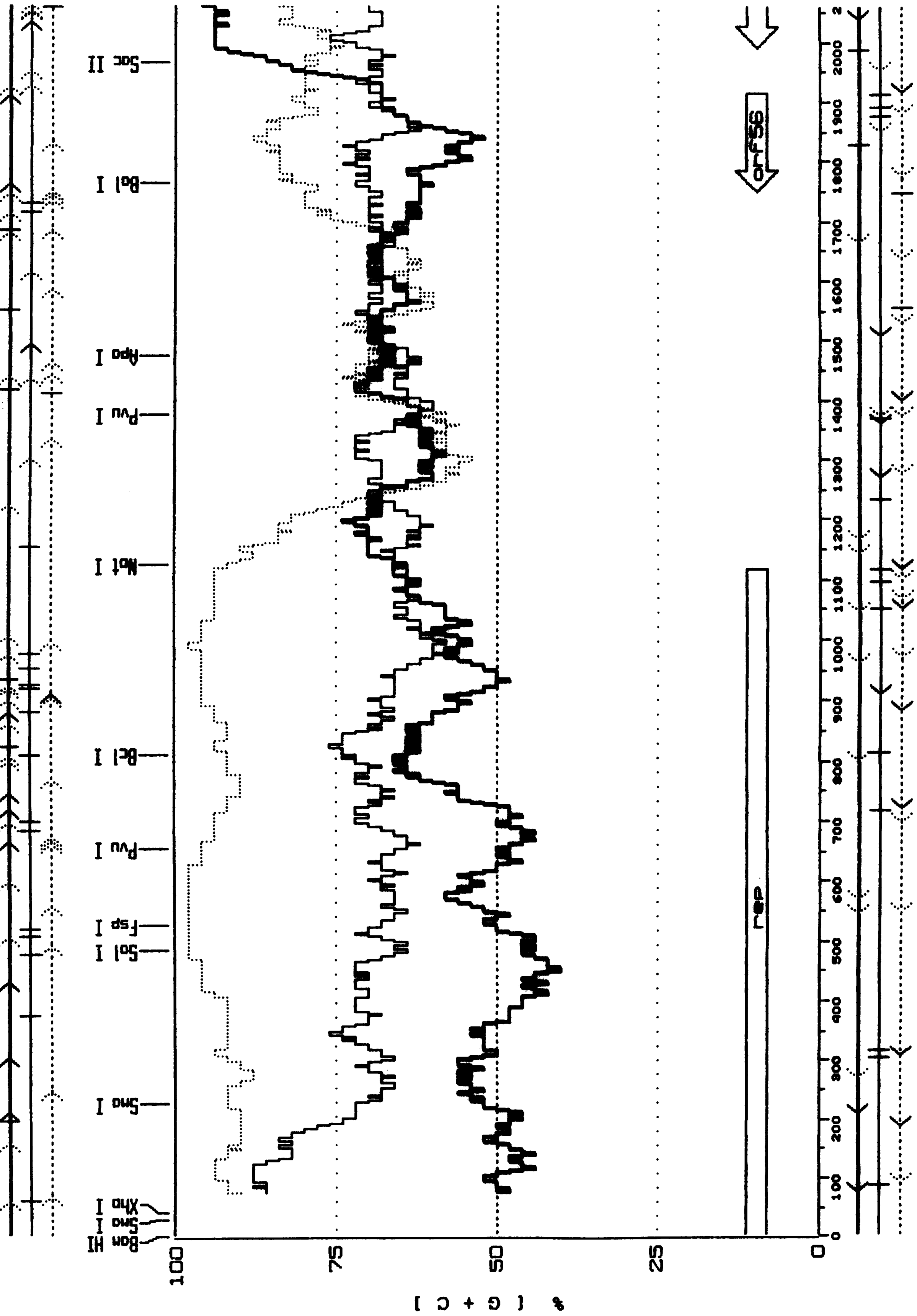
5641 GCGGAGTTGTAGCCCTTCTGGGCGTCCATCAGCGCCTTCAGGTGCTCCGGGGTCCGGAGCGCCATGCGGCGGTCGGCCTCGGCGTCCCAGCGGGCCGCGAACGGCGCGAGGGTCGGGGCC
CGCCTCAACATCGGGAAGACCCGCAGGTAGTCGCGGAAGTCCACGAGGCCCCAGGCCTCGCGGTACGCCGCCAGCCGGAGCCGCAGGGTCGCCCGGCGCTTGCCGCGCTCCCAGCCCCGG
SerAsnTyrGlyLysGlnAlaAspMetLeuAlaLysLeuHisGluProThrArgLeuAlaMetArgArgAspAlaGluAlaAspTrpArgAlaAlaPheProAlaLeuThrProAlaMet

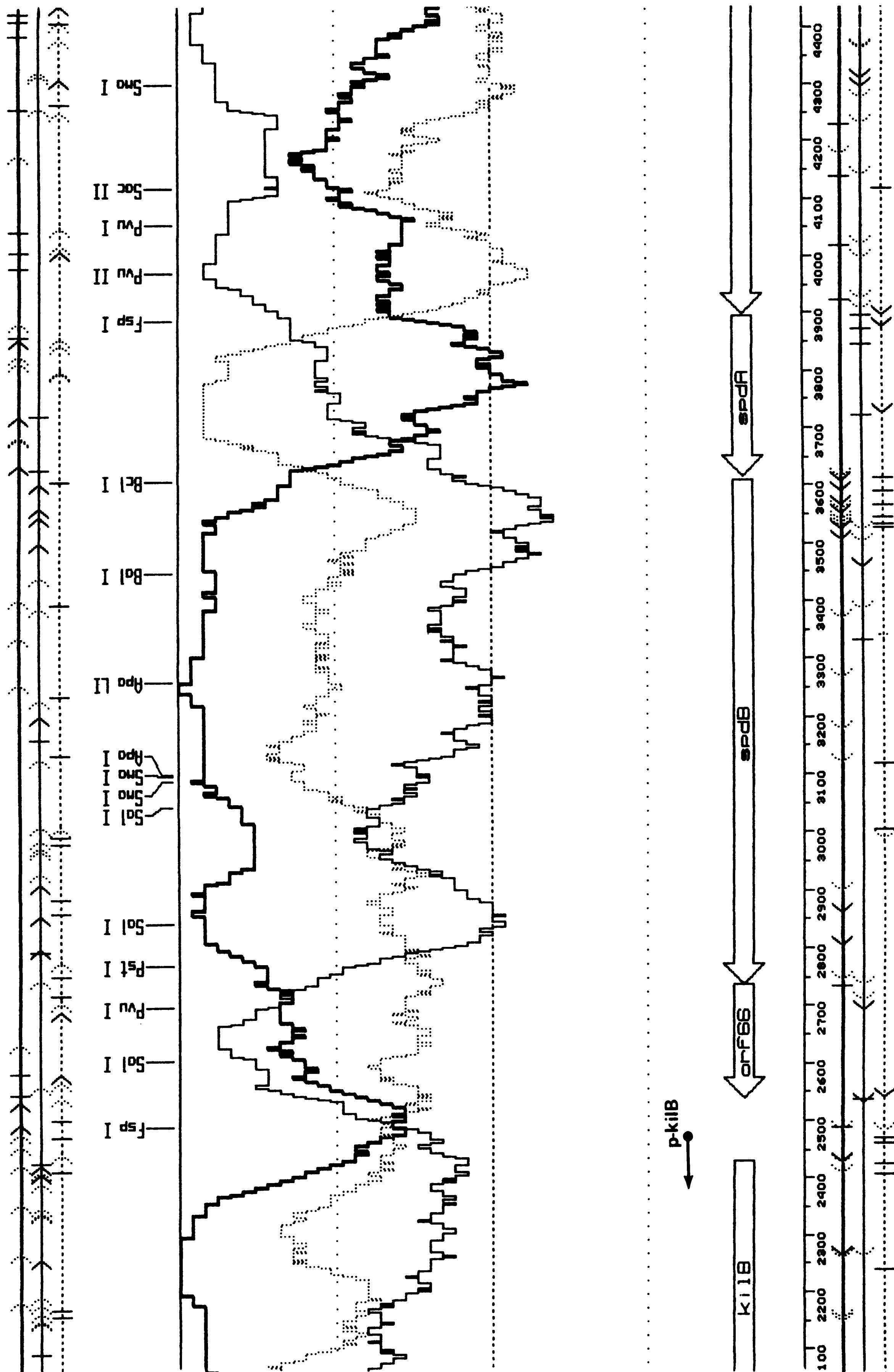
5761 ATGGCGTCGAGCGCGTTCGTGATCGCTCCGGCGGCGTTCTTGGAGCCGGTTGCGATCTTCGTAGAGACTGCCTTCGGGTCCACGGTTTGTCCTTTCGCGAGGGACGTGGATCTAGGGCCG
TACCGCAGCTCGCGCAAGCACTAGCGAGGCCGCCGCAAGAACCTCGGCCAACGCTAGAAGCATCTCTGACGGAAGCCCAGGTGCCAAACAGGAAAGCGCTCCCTGCACCTAGATCCCGGC

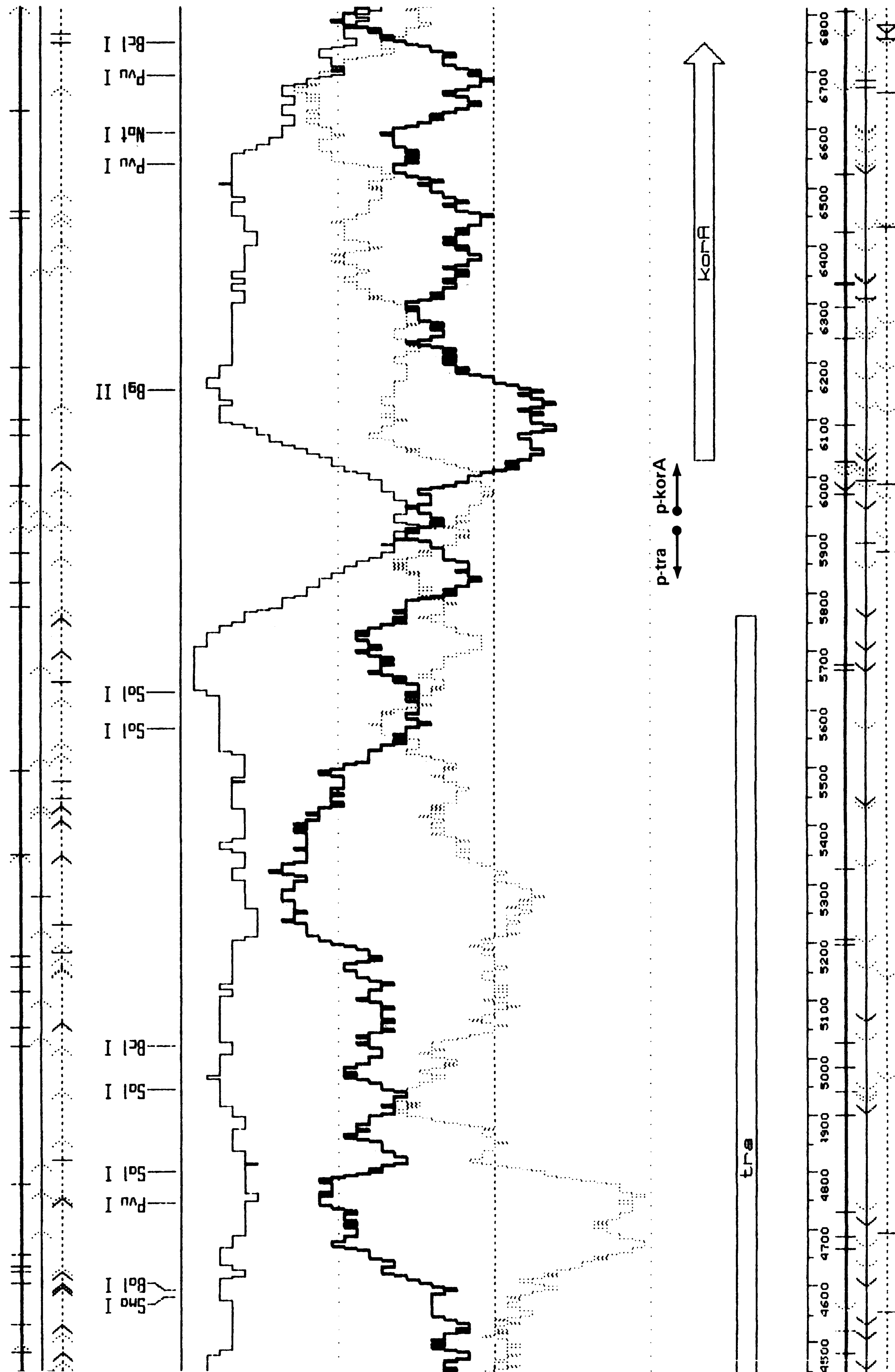
tra

```
5881 GAGACCGTTCACGCGGTCTCCGGTCCGCCCCGTTTCCGGGGCTGTGTGTGGCGTCGAACAAGGTCCATACTGTGGTCGCACAGTTGCTGTGTCAAGGCATACACTGTGCTAGACAGCTAC
     CTCTGGCAAGTGCGCCAGAGGCCAGGCGGGGCAAAGGCCCCGACACACACCGCAGCTTGTTCCAGGTATGACACCAGCGTGTCAACGACACAGTTCCGTATGTGACACGATCTGTCGATG

                            korA
                                MetSerLeuGluArgThrProProTyrLeuGlnValValAlaAlaLeuLysAlaLysIleValSerGlyGluLeuLysHisGlyAspThr
6001 ACACCGCGCACCACACTCGAAGGAGTCGTCATGTCCCTGGAGCGCACGCCCCCGTACCTCCAAGTCGTCGCCGCGCTGAAGGCAAAGATCGTCAGCGGGGAGCTGAAGCACGGGGACACG
     TGTGGCGCGTGGTGTGAGCTTCCTCAGCAGTACAGGGACCTCGCGTGCGGGGGCATGGAGGTTCAGCAGCGGCGCGACTTCCGTTTCTAGCAGTCGCCCCTCGACTTCGTGCCCCTGTGC

                                     Bgl II
     LeuProSerValArgAspLeuAlaAlaGlnTyrGluIleSerThrAlaThrAlaGlnLysValHisArgThrLeuLysAlaGluGlyLeuAlaGluAlaLysGlnGlySerAlaThrThr
6121 CTGCCGTCCGTGCGGGACCTCGCGGCGCAGTACGAGATCTCGACCGCCACGGCCCAGAAGGTCCACCGGACGCTGAAGGCGGAAGGGCTGGCGGAGGCGAAGCAGGGCAGCGCGACCACG
     GACGGCAGGCACGCCCTGGAGCGCCGCGTCATGCTCTAGAGCTGGCGGTGCCGGGTCTTCCAGGTGGCCTGCGACTTCCGCCTTCCCGACCGCCTCCGCTTCGTCCCGTCGCGCTGGTGC

     ValSerThrArgArgThrLeuHisArgThrAlaAlaAspArgLeuGluSerAlaLeuSerThrGlyArgIleTyrAlaAspGlyGluTyrAlaValIleThrSerAlaAlaLeuAlaGlu
6241 GTCAGCACGCGACGGACCCTGCACCGGACCGCAGCCGACCGGCTGGAGTCGGCGCTCAGCACGGGCCGGATCTACGCGGACGGGGAGTACGCGGTCATCACCAGCGCCGCCCTTGCCGAG
     CAGTCGTGCGCTGCCTGGGACGTGGCCTGGCGTCGGCTGGCCGACCTCAGCCGCGAGTCGTGCCCGGCCTAGATGCGCCTGCCCCTCATGCGCCAGTAGTGGTCGCGGCGGGAACGGCTC

     ProProGluTrpValAlaAspLeuLeuGlyThrGluSerGlyGlnAlaValArgArgGluArgValThrHisSerAlaAspAspGlnProValSerAlaSerValSerTrpPheSerAla
6361 CCGCCCGAGTGGGTGGCCGATCTCCTCGGCACCGAGAGCGGCCAGGCCGTGCGACGCGAGCGCGTCACCCACTCAGCCGACGACCAGCCGGTGAGCGCCAGCGTGAGCTGGTTCTCCGCA
     GGCGGGCTCACCCACCGGCTAGAGGAGCCGTGGCTCTCGCCGGTCCGGCACGCTGCGCTCGCGCAGTGGGTGAGTCGGCTGCTGGTCGGCCACTCGCGGTCGCACTCGACCAAGAGGCGT

                                                                       Pvu I                                       Not I
     AspLeuAlaGluThrValProAlaLeuLeuValArgAspArgIleIleGlyGlyThrProSerAlaIleGluAlaAlaThrGlyArgArgAlaValAlaThrGluGluAlaThrThrAla
6481 GACCTCGCGGAGACCGTGCCCGCCCTCCTGGTCCGCGACCGGATCATCGGCGGCACCCCGTCCGCGATCGAGGCAGCCACCGGCCGCCGGGCCGTCGCCACCGAGGAAGCCACCACGGCG
     CTGGAGCGCCTCTGGCACGGGCGGGAGGACCAGGCGCTGGCCTAGTAGCCGCCGTGGGGCAGGCGCTAGCTCCGTCGGTGGCCGGCGGCCCGGCAGCGGTGGCTCCTTCGGTGGTGCCGC

                                                                                                     Pvu I
     AlaAlaAlaThrGluAspGlnAlaAlaLeuLeuGlyValAlaAlaGlyAlaProValLeuLeuSerArgAsnValTyrValAspAlaGlnGlyAspThrIleGluValGlyGluSerVal
6601 GCCGCCGCGACCGAGGACCAGGCCGCGCTTCTGGGCGTAGCCGCAGGCGCCCCAGTCCTGCTCTCACGCAACGTCTACGTGGACGCTCAGGGCGACACGATCGAGGTCGGGGAGTCCGTC
     CGGCGGCGCTGGCTCCTGGTCCGGCGCGAAGACCCGCATCGGCGTCCGCGGGGTCAGGACGAGAGTGCGTTGCAGATGCACCTGCGAGTCCCGCTGTGCTAGCTCCAGCCCCTCAGGCAG

                                     Bcl I
     AlaProAlaGlyArgTrpArgValHisArgAsp***
6721 GCTCCGGCGGGCCGCTGGCGCGTCCACCGGGACTGATCACTCGCCCATTGAGAAGCCCCGTCAGGCACCGCCCGACGGGGCTTCTTCACGTCCAGACGACGTGGTTTCCGGGGGTTGCCC
     CGAGGCCGCCCGGCGACCGCGCAGGTGGCCCTGACTAGTGAGCGGGTAACTCTTCGGGGCAGTCCGTGGCGGGCTGCCCCGAAGAAGTGCAGGTCTGCTGCACCAAAGGCCCCCAACGGG

                                                                                                 Apa I
6841 ACCCGGTTGAGCCGTTGCACCCCGGTTGACCCCCCGCAACCGGGTCTGACCTGCGAAGTTGCAAGGTTGCGACGGTTTCCAGGAGGGGGCCCTACGCGTGCGCGCGCGAGGAAGCCATTT
     TGGGCCAACTCGGCAACGTGGGGCCAACTGGGGGGCGTTGGCCCAGACTGGACGCTTCAACGTTCCAACGCTGCCAAAGGTCCTCCCCCGGGATGCGCACGCGCGCGCTCCTTCGGTAAA

                                                             Sma I
6961 TTGATCTAGCTTCCGGAGCCCTTCTCGGCCTCCGCCTTGGCCCACGCCCGGGCCGTGTTCGGAGCGACCGCCGTCACCTCGTTGATGCGGCGGTAGGGGACCTCCATCCGGACCGCCTCC
     AACTAGATCGAAGGCCTCGGGAAGAGCCGGAGGCGGAACCGGGTGCGGGCCCGGCACAAGCCTCGCTGGCGGCAGTGGAGCAACTACGCCGCCATCCCCTGGAGGTAGGCCTGGCGGAGG
        ***SerGlySerGlyLysGluAlaGluAlaLysAlaTrpAlaArgAlaThrAsnProAlaValAlaThrValGluAsnIleArgArgTyrProValGluMetArgValAlaGluAla

7081 GCGATCAGCGGGCGCAGCTCCTTCTCAATCTCGTCCAGCTCCGCGAGGAGCTTGATCCGCTGCTGCCCCAACGGCTTCAGCGCCGCCTCTGCCTCGGCCCGGATCTCGCCCGGTGTCTTT
     CGCTAGTCGCCCGCGTCGAGGAAGAGTTAGAGCAGGTCGAGGCGCTCCTCGAACTAGGCGACGACGGGGTTGCCGAAGTCGCGGCGGAGACGGAGCCGGGCCTAGAGCGGGCCACAGAAA
       IleLeuProArgLeuGluLysGluIleGluAspLeuGluAlaLeuLeuLysIleArgGlnGlnGlyLeuProLysLeuAlaAlaGluAlaGluAlaArgIleGluGlyProThrLysGln

                                             Fsp I Spe I
7201 TGCGTCATGAAGTCATCCTGACCGACTGTGTCAGTCTGCGCAACTAGTTCAGGCTGCGTTTTTTGCGGTACAACTTTCCCTACGTCATCAAGGCGGCCCGCGAGCGGGCCGCGCGGCCCG
     ACGCAGTACTTCAGTAGGACTGGCTGACACAGTCAGACGCGTTGATCAAGTCCGACGCAAAAAACGCCATGTTGAAAGGGATGCAGTAGTTCCGCCGGGCGCTCGCCCGGCGCGCCGGGC
       ThrMet
     korB

7321 GCCCACGGCCCGGCCGACGCTCCTGTCTTCGCCCCGCTCCGGCCCGGCCGCCGACCGGCCCGCGCACACGACGGGGGCGGCATCGGTGGGCGGTTTACGTGGCGCCTGCTCCGCCGACTG
     CGGGTGCCGGGCCGGCTGCGAGGACAGAAGCGGGGCGAGGCCGGGCCGGCGGCTGGCCGGGCGCGTGTGCTGCCCCCGCCGTAGCCACCCGCCAAATGCACCGCGGACGAGGCGGCTGAC
```

```
7441 CGGGCATCGCCGTTGTGCTCGCCGACAGGGCAGCGGGGAGGGGTGGGGGACTCGCGGCCCTACGCGGCCGTCTGAGCGCCTGTCAGCCTCCCGGAGCGCCGTACCCCCGCCGTCGCGGTG
     GCCCGTAGCGGCAACACGAGCGGCTGTCCCGTCGCCCCTCCCCACCCCCTGAGCGCCGGGATGCGCCGGCAGACTCGCGGACAGTCGGAGGGCCTCGCGGCATGGGGGCGGCAGCGCCAC

                                                                                                     orf85
                                                                                       MetArgTrpArgProSerSerTrpSerAlaValArgLeuArgArgGlyArgSerSerGly
7561 CTGAGCCGCGTGAGGCGACCCTGAGCCCCGTCGTGGGTTGCTGGGGAGCACCTGCTGCCGCGATGAGGTGGCGGCCGTCGAGCTGGTCAGCCGTGCGGCTCCGTCGTGGCCGGTCATCCG
     GACTCGGCGCACTCCGCTGGGACTCGGGGCAGCACCCAACGACCCCTCGTGGACGACGGCGCTACTCCACCGCCGGCAGCTCGACCAGTCGGCACGCCGAGGCAGCACCGGCCAGTAGGC
                                                                                                                       ***GlyAla

           Pvu I                                                                      Sac II
      CysProIleValValGlyLysMetProAlaGluProAlaAspLeuAspArgAspGlyTyrProAlaGlyArgGlyArgAlaSerValGluGlyGluGlyGlyArgHisAlaAspArgHis
7681 GCTGCCCGATCGTGGTGGGCAAGATGCCGGCGGAACCGGCGGACCTCGACCGCGACGGCTATCCGGCGGGCCGCGGGCGGGCCTCCGTAGAGGGCGAGGGCGGGCGCCATGCCGACCGCC
     CGACGGGCTAGCACCACCCGTTCTACGGCCGCCTTGGCCGCCTGGAGCTGGCGCTGCCGATAGGCCGCCCGGCGCCCGCCCGGAGGCATCTCCCGCTCCCGCCCGCGGTACGGCTGGCGG
      AlaArgAspHisHisAlaLeuHisArgArgPheArgArgValGluValAlaValAlaIleArgArgAlaAlaProProGlyGlyTyrLeuAlaLeuAlaProAlaMetGlyValAlaVal

                     Sac I
      GlyGlyProGlnAlaGlnGluLeuProHisAspGlnArgProAlaAspGlnGluGlnCysGlyGlnHisCysArg***
7801 ACGGCGGGCCACAGGCCCAGGAGCTCCCGCACGATCAGCGTCCCGCCGACCAGGAGCAGTGCGGCCAGCACTGCCGCTAACGCCTGGTCCTGGTCCCGGTCCTGGTGGTGCATCAGTCCT
     TGCCGCCCGGTGTCCGGGTCCTCGAGGGCGTGCTAGTCGCAGGGCGGCTGGTCCTCGTCACGCCGGTCGTGACGGCGATTGCGGACCAGGACCAGGGCCAGGACCACCACGTAGTCAGGA
      AlaProTrpLeuGlyLeuLeuGluArgValIleLeuThrGlyGlyValLeuLeuLeuAlaAlaLeuValAlaAlaLeuAlaGlnAspGlnAspArgAspGlnHisHisMet
                                                                                                                 orf79

          Bcl I                  Bcl I
7921 CCCCGTGATCACTTCGGCACCCACCGTAGTGATCACCCCCGACAGCGGATCAAGGGGTTTGCGGGTCCCGGTCGGCGCCGGGCGGGGGAGGCAGGAGCCGCCGACGCTGCCTCTGGGACG
     GGGGCACTAGTGAAGCCGTGGGTGGCATCACTAGTGGGGGCTGTCGCCTAGTTCCCCAAACGCCCAGGGCCAGCCGCGGCCCGCCCCCTCCGTCCTCGGCGGCTGCGACGGAGACCCTGC

                                     Pst I
8041 GGCCGGACGGCAGGGGGACCGGCGGCCGGGCGAGCTGCAGCCGGGGGTCCGGCAGGGCCGGAGCGGGCGGAACCGTGCTCTGACCTGCGGCCCGAGTTTCGTCACGTGACGGAATGGAAG
     CCGGCCTGCCGTCCCCCTGGCCGCCGGCCCGCTCGACGTCGGCCCCCAGGCCGTCCCGGCCTCGCCCGCCTTGGCACGAGACTGGACGCCGGGCTCAAAGCAGTGCACTGCCTTACCTTC

                                                                          Pvu I
8161 GCTGCTGCATTTCGTCACGTGACGTATCTCGGCGAGCGACTGCCGACGCCACGGCGGACACGATCGCCTCGCGCTGGCGCCGGGCCTCGTACGCCCGCTGGCGGCAGGAGCGGCGGCAGT
     CGACGACGTAAAGCAGTGCACTGCATAGAGCCGCTCGCTGACGGCTGCGGTGCCGCCTGTGCTAGCGGAGCGCGACCGCGGCCCGGAGCATGCGGGCGACCGCCGTCCTCGCCGCCGTCA

8281 AGTCCCGGCTCCGGCCGACGCCGGATTGCTTGATCTCCGAGCCGCACCAGGCGCAGAGCTTCGCGCCGTCGGCGTCCCTGGGGGTGGTGGTGCTCATGGCCGACGACCGTACGCGGCACG
     TCAGGGCCGAGGCCGGCTGCGGCCTAACGAACTAGAGGCTCGGCGTGGTCCGCGTCTCGAAGCGCGGCAGCCGCAGGGACCCCCACCACCACGAGTACCGGCTGCTGGCATGCGCCGTGC

                                   Kpn I                          Sma I                      Sac II
8401 TCTCGTAGCGAGGCGAGTCGGGCGCGAGGTACCGCCTGCACGAAGTGCCGGCGGGGCCGACCCCGGGCGAGTAATCCCAGGATTACTCCCGCGGCTTCGACCCCGGCCGCCGTCGCCGCG
     AGAGCATCGCTCCGCTCAGCCCGCGCTCCATGGCGGACGTGCTTCACGGCCGCCCCGGCTGGGGCCCGCTCATTAGGGTCCTAATGAGGGCGCCGAAGCTGGGGCCGGCGGCAGCGGCGC

                                                                            Apa I
8521 TACGTCACCGACCCCCGCCGTACGTCACCGGGATGACGTACGGCGGGGGGGAGCGAGTTAGTGCGAAGTGGGCCCACTTGCGAGCCGGGCGATGTGCCGGGCGGCCCGCTCCTGGCGGTC
     ATGCAGTGGCTGGGGGCGGCATGCAGTGGCCCTACTGCATGCCGCCCCCCCTCGCTCAATCACGCTTCACCCGGGTGAACGCTCGGCCCGCTACACGGCCCGCCGGGCGAGGACCGCCAG
                                                                  ***HisSerThrProGlySerAlaLeuArgAlaIleHisArgAlaAlaArgGluGlnArgAsp

8641 GTCGGCGTCGTCGTCCTGGTCGTCGTCCTGCTCTCGCCGTCGGCGTGCAGTTGCTTCCTCGCGGCGCTGGGCGAGGGCGGCGAGCATGTCGGCGTACGCCTCGGCCACCTCCCCCGCCGT
     CAGCCGCAGCAGCAGGACCAGCAGCAGGACGAGAGCGGCAGCCGCACGTCAACGAAGGAGCGCCGCGACCCGCTCCCGCCGCTCGTACAGCCGCATGCGGAGCCGGTGGAGGGGGCGGCA
     AspAlaAspAspAspGlnAspAspAspGlnGluArgArgArgArgAlaThrAlaGluGluArgArgGlnAlaLeuAlaAlaLeuMetAspAlaTyrAlaGluAlaValGluGlyAlaThr

8761 GAGCACCACCACTGTGTCGGCCGCGTCGGCCAGCGCCAGGACCTCCCGCACCCGTTCGCCCACGGCCGCC
     CTCGTGGTGGTGACACAGCCGGCGCAGCCGGTCGCGGTCCTGGAGGGCGTGGGCAAGCGGGTGCCGGCGG
     LeuValValValThrAspAlaAlaAspAlaLeuAlaLeuValGluArgValArgGluGlyValAlaAlaPro
                                                                    rep
```


Sequence Name: pIJ101-Bam t = 50
% [G + C]
100
75
50
25
0
Bam HI
Sma I
Xho I
Sma I
Sal I
Fsp I
Pvu I
Bcl I
Not I
Pvu I
Kpn I
Bal I
Sac II
rep
orf56


Sma I
Sac II
Pvu I
Pvu II
Fsp I
Bcl I
Bgl I
ApaLI
Apa I
Sma I
Sma I
Sal I
Sal I
Pst I
Pvu I
Sal I
Fsp I
spdA
spdB
orf66
kilB
p-kilB
100
2200
2300
2400
2500
2600
2700
2800
2900
3000
3100
3200
3300
3400
3500
3600
3700
3800
3900
4000
4100
4200
4300
4400


Bcl I
Pvu I
Not I
Pvu I
Bgl II
Sal I
Sal I
Bcl I
Sal I
Sal I
Pvu I
Bal I
Sma I
korA
p-korA
p-tra
tra
4500
4600
4700
4800
4900
5000
5100
5200
5300
5400
5500
5600
5700
5800
5900
6000
6100
6200
6300
6400
6500
6600
6700
6800

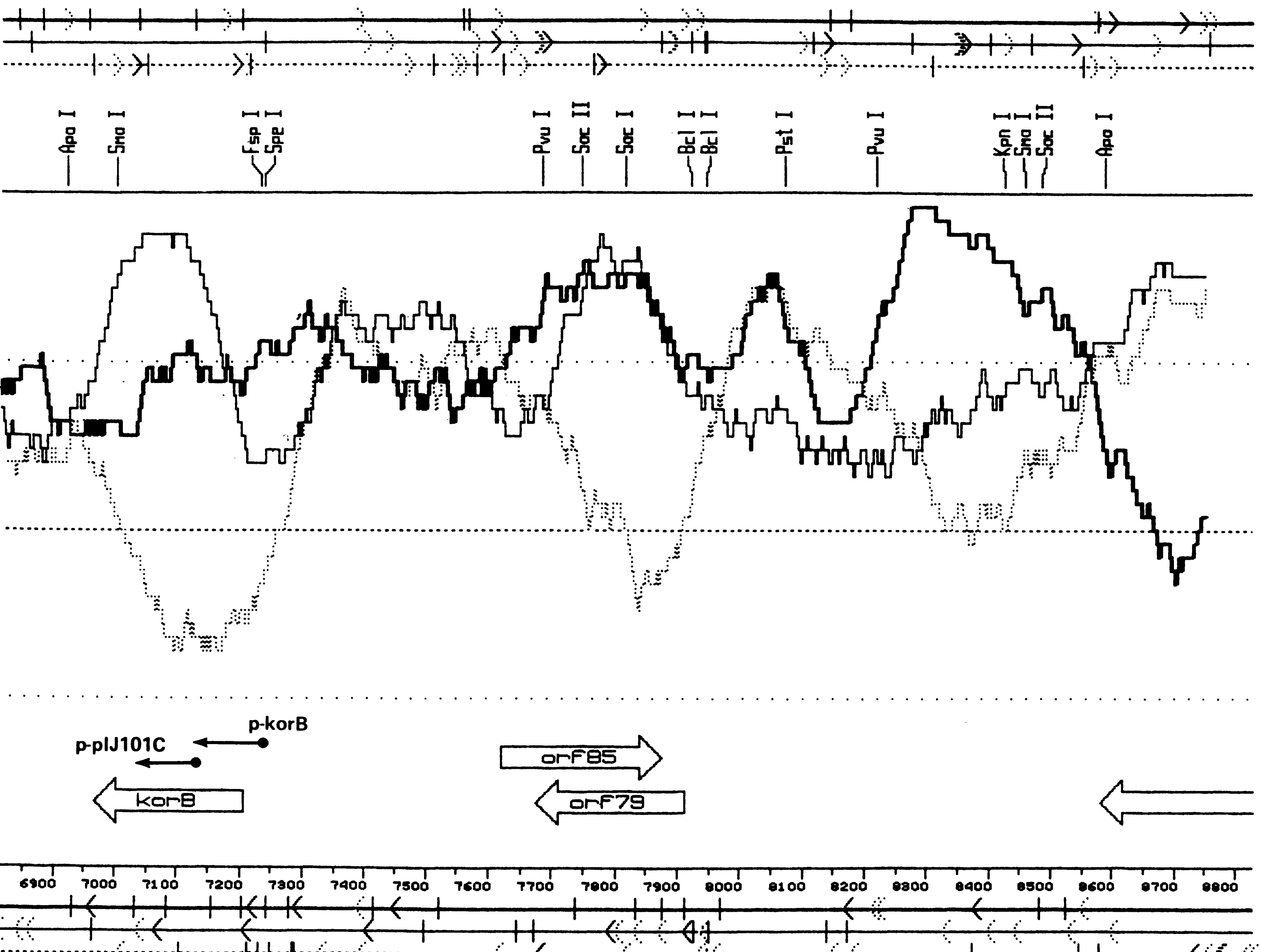


FIG. 2. FRAME plot of pIJ101. The average mole% G+C over a 50-codon window is plotted for the first (▬), second (——), and third (┈┈) positions in each triplet. The positions of AUG (〉) and GUG (⁚〉) start codons and UAA, UGA, and UAG stop codons (❙) for each possible reading frame are indicated above (translation, left to right) and below (translation, right to left) the plot. Also included are certain restriction sites and the positions of and direction of transcription from five known promoters on the plasmid (•→).

TABLE 1. Putative ORFs of pIJ101

Name	Predicted no. of codons	Predicted mol wt	% G+C at codon position:		
			1st	2nd	3rd
rep	456	57,752	75	52	93
korA	241	29,610	72	54	91
korB	80	10,223	69	45	91
tra	621	77,504	70	53	94
spdA	94	11,188	72	54	95
spdB	292	35,830	73	57	94
kilB	148	17,700	77	64	96
orf56	56	6,937	70	52	85
orf66	66	7,711	79	61	92
orf79	79	9,840	87	48	86
orf85	85	10,841	79	60	86

kilA, was defined in terms of a genetic function carried by a 1,126-bp *BclI-BglII* fragment (bp 5026 to 6157) that could not be introduced into *S. lividans* unless a particular Kil override locus (*korA*) was also present. Insertions into *kilA* or into an adjacent locus (designated *tra*) that is devoid of any Kil function (20) abolished the pocking phenotype characteristic of plasmid transfer (6, 21). The FRAME plot indicated the presence of a single long ORF within the DNA segment spanning both the *kilA* and *tra* loci, starting at about bp 5751 and terminating at bp 3899. A number of possible start codons for the ORF are present near the point of plot divergence; the AUG start at bp 5762 was chosen as being the most likely, yielding an ORF that could encode a protein of 621 amino acids. This ORF extends well beyond the *BclI* site at bp 5026 which marks the genetic limit of the *kilA* locus. Thus, if the KilA phenotype is due to a protein product, the *kilA* locus cannot code for a complete protein but must instead encode a product that is a truncation of the *tra* protein. Recent evidence (K. J. Kendall, unpublished observations) has demonstrated that the KilA phenotype does not result from truncation of the putative *tra* protein product. Thus, the entire ORF has been designated *tra*.

Two other transfer-related loci, *spd* and *kilB*, were genetically defined as regions in which frame-disrupting insertions resulted in very small pocks, suggesting that plasmid transfer was occurring at a greatly reduced frequency (20, 21). The *spd* (spread) locus was found to lie between the *SalI* site at bp 2848 and the *PvuII* site at bp 3977. Two ORFs appeared to be present in this region. The first, *spdA*, could use an AUG start codon (bp 3897) that overlaps the UGA termination codon of the *tra* ORF. This ORF would be 94 codons in length. A second ORF, *spdB*, followed *spdA* closely in the same direction. While the AUG start codon at bp 3610 was chosen as the most likely start, it should be noted that many alternative in-frame start sites are present in this region, including a GUG codon at bp 3616 that would overlap the UGA termination codon of the *spdA* ORF. The frame analysis indicated that the *spdB* ORF would contain 291 codons and terminate at bp 2737. Insertions into either of these ORFs produce the Spd phenotype (20, 21).

kilB has been defined as a genetic function localized to an 880-bp *SacII-SalI* fragment (bp 1972 to 2848) that could not be introduced into *Streptomyces* in the absence of either of the *korA* or *korB* repressors; insertions into this region give the Spd phenotype (20), as indicated above. The FRAME plot indicated the existence of a 148-codon ORF starting at an AUG (bp 2431) and terminating at bp 1990.

(c) Unidentified ORFs. In addition to the ORFs mentioned above, the FRAME plot suggested that other ORFs may be present on the plasmid. The most prominent of these are indicated in Fig. 2.

The very short ORF, orf56 (bp 1917 to 1749), on the 2.1-kb *SacII* fragment containing the minimal replicon has been mentioned previously. This ORF follows *kilB* so closely (75 bp) that it is possible that both ORFs belong to the same operon. If this is so, then it is unlikely that orf56 is required for replication, as the putative *kilB* promoter (see below) is not present on the *SacII*-generated minimal replicon.

A short ORF, orf66, immediately follows *spdB*. A GUG start codon preceded by a possible ribosome-binding site overlaps the UGA termination codon of *spdB*. Frame-disrupting insertions into the *SalI* site (bp 2612) do not have detectable phenotypic effects on the plasmid (20). Thus, the function, if any, of this ORF is unknown.

The FRAME analysis of the region between the *PvuI* site at bp 7689 and the *ApaI* site at bp 8592 is difficult to interpret. Between bp 7700 and 7900, the third plot (dotted line) indicated an extremely low G+C composition. However, both of the other plots showed much higher than average G+C content. Although overlapping bidirectional reading frames have not been identified in *S. lividans*, this pattern is what one might expect if reading frames are encoded on both strands; both would have a common low G+C second base in each triplet, with the first base somewhat higher in G+C than expected for a coding region and the third base somewhat lower in G+C than expected. A suitable start and stop codon is present on each strand for both possible ORFs; thus, the ORFs have been tentatively designated orf85 and orf79, respectively.

Between the *BclI* site at bp 7952 and the *ApaI* site at bp 8592, all three reading frames diverge significantly from 73% G+C. In particular, the plot of the region between *PvuI* (bp 8223) and *ApaI* (bp 8451) strongly suggested the presence of an ORF starting at about bp 8150 and ending at about bp 8500. However, a number of stop codons are present in that ORF, so that any proteins encoded would be extremely short. Conversely, a long ORF could exist running in the opposite direction, starting at bp 8600 and finishing at bp 8150. However, such an ORF would have a G+C composition of 87, 62, and 70% at the first, second, and third codons, respectively, strongly contradicting the expected compositions of protein-encoding ORFs of *Streptomyces* spp. Thus, no ORF has been assigned to this region.

Analysis of ORFs in pIJ101. Although the FRAME program is an extremely powerful tool for identifying protein-encoding reading frames when applied to *Streptomyces* DNA, there are some limitations in its use. The reading frame amino termini indicated by the program are approximate rather than exact, and identifying the protein start codon can be difficult in the absence of substantiating protein sequence data, particularly in view of the frequent use of GUG protein start codons for *Streptomyces* proteins (9, 29, 32). The choice of initiation codons in the identified ORFs on pIJ101 was determined in our analysis in part by the proximity of possible protein start codons to regions of complementarity with the 3′ end of the 16S rRNA of *S. lividans* (4). Table 2 lists the DNA sequences immediately preceding the most likely initiation codons for each of the ORFs identified. Note that there are two entries for the *tra* ORF. We have assumed that the AUG codon at bp 5762 is the most likely initiation site based on the FRAME plot, although there is no good ribosome-binding site sequence preceding this codon and a reasonable ribosome-binding site exists 5′ to a GUG start codon situated in frame 81 bp further upstream. Thus, it is entirely possible that the GUG codon cited is used.

TABLE 2. Ribosome-binding sites preceding the initiation codons of putative ORFs of pIJ101[a]

ORF	Sequence
rep	CAAGGGCTGCGGCCGCGCCCTG**ATG**
korA	CACCACACTCGAAGGAGTCGTC**ATG**
korB	CACAGTCGGTCAGGATGACTTC**ATG**
tra (AUG start)	GATCACGAACGCGCTCGACGCC**ATG**
tra (GUG start)	GTCCCTCGCGAAAGGACAAACC**GTG**
spdA	AGGTCGGGGAGGTGTCGGCCTG**ATG**
spdB	AGGCCCGGAAGGGGAAGTGATC**ATG**
kilB	GACACGACGCCAGGGGGCTCAC**ATG**
orf56	CCCACCCACCAGGAGACCGACC**ATG**
orf66	GCGTCACGAGGGAGGTGGCCGC**GTG**
orf79	GAAGTGATCACGGGGAGGACTG**ATG**
orf85	CTGGGGAGCACCTGCTGCCGCG**ATG**

[a] Bases exhibiting complementarity to the 3′ end of 16S rRNA (5′-GAUC ACCUCCUUUCU-3′) are underlined (including G · U base pair). Putative initiation codons are indicated by boldface type.

Although most of the ORFs show some regions of complementarity to the 3′ end of 16S rRNA immediately preceding the proposed initiation codons, we note the lack of any complementarity preceding the AUG of the *rep* ORF. However, none of the other possible GUG or AUG codons around the beginning of the *rep* ORF display greater complementarity.

A compilation of the predicted molecular weight, number of codons, and G+C usage at the first, second, and third codons of each of the ORFs is presented in Table 1. It can be seen that, with the exception of the overlapping orf79 and orf85, all of the proposed ORFs have a G+C composition fully characteristic of *Streptomyces* reading frames. As a further check for the validity of the selected reading frames, each ORF was examined for the presence of codons known to be used rarely in *Streptomyces* species (data not shown). No unusual codon usage characteristics could be seen for any of the presumptive ORFs.

Having determined the likely coding capacity of the various ORFs present on pIJ101, we analyzed each of the predicted amino acid sequences to determine any structural or sequence similarities to previously characterized proteins of known function. Thus, all of the predicted amino acid sequences were scanned for homologies with the Protein Identification Resource and Swiss Protein data bases with the Intelligenetics IFIND and XFASTP programs. No significant homologies were found. The predicted proteins were also subjected to searches for (i) homology to DNA-binding domains of DNA-binding proteins (24), (ii) homology to nucleotide-binding fold domains of proteins with ATPase activity (30), (iii) Chou-Fasman and Robson-Garnier plots for secondary structure (10, 15), and (iv) hydrophobic moment plots to determine the likelihood of association with membranes (13).

The *korA* and *korB* loci both produced diffusible products that decreased transcription from certain promoters present on pIJ101 (D. Stein, K. J. Kendall, and S. N. Cohen, unpublished observations). Thus, by analogy with other procaryotic systems, it is most likely that the loci code for repressor proteins. Neither of the predicted amino acid sequences for the *kor* proteins shows significant homology to any other known proteins. However, each sequence was scanned for regions of homology to the consensus DNA-binding α-helix-turn-α-helix structural motif found in many procaryotic repressor proteins (24). A reasonable match was found starting at amino acid 194 in the *korA* sequence (Fig. 3). The best match in the *korB* protein occurred within the segment starting at amino acid 13 (Fig. 3). However, subsequent work (D. Stein, personal communication) has demonstrated that alterations in the amino acid sequence of *korB* around residues 66 to 70 severely affect the function of the protein. An extremely weak match to the DNA-binding domain consensus sequence spans this region starting at amino acid 56 (Fig. 3).

Hydrophobic moment plots (13) were drawn for all of the proteins. Three very strongly hydrophobic domains were found in the amino-terminal half of the *spdB* protein, each long enough to span a membrane; the first domain (45 amino acids in length) is located starting at amino acid 9, the second (20 amino acids) starts at amino acid 73, and the third (24 amino acids) starts at amino acid 96. Similar domains were also found in the predicted amino acid sequences of *spdA* (28 amino acids starting at amino acid 41) and *kilB* (21 amino acids starting at amino acid 1), the two other loci required for the Spd$^+$ phenotype. A very strongly hydrophobic domain is also present in the predicted amino acid sequence of orf66 (21 amino acids starting at amino acid 1).

The predicted *tra* protein has no significant hydrophobic domains or homology to DNA-binding domains. However, a good match to the consensus nucleotide-binding fold domain of ATP-requiring enzymes (30) is present in the sequence (Fig. 4). This raises the possibility that ATP may be required for *tra* function.

No significant structural features or homologies were found in the predicted amino acid sequence of the *rep* ORF.

Location of promoters on pIJ101. Studies in this laboratory (D. Stein, K. J. Kendall, and S. N. Cohen, unpublished results) and previously published results from other laboratories (7, 8, 11) have demonstrated the existence of at least five promoters on pIJ101. The positions of these promoters and their directions of transcription are indicated in Fig. 2. Note that we have renamed the previously identified pIJ101A (7) and pIJ101B (8) promoters the *kilB* and *tra* promoters, respectively, in accordance with their location on the genetic map. The FRAME analysis showed that all of these promoters (except that of pIJ101C [11]) lie upstream of major identified ORFs.

Location of regions of dyad symmetry on pIJ101. The nucleotide sequence of pIJ101 was searched for regions of dyad symmetry. The locations of the most significant of these (those having a match of at least 10 out of 12 nucleotides) are indicated in Fig. 1 and tabulated in Table 3. Certain inverted repeat structures have commonly been found to have transcription termination activity (reviewed in reference 25); one of the regions of dyad symmetry, centering around bp 6788, has already been shown to be able to terminate transcription in both *Streptomyces* species and *E. coli* (12). This segment lies between the convergently transcribed *korA* and *korB* genes and may thus act as a transcriptional terminator for either or both of these genes.

Significant inverted repeat structures were also found downstream of the *rep* (bp 8551 and 8478), orf85 (bp 7941), orf66 (bp 2520 and 2484), *kilB* (bp 1957), and orf56 (bp 1723) ORFs. Of the remaining significant regions of dyad symmetry, one lies in the divergent reading frame region between the *korA* and *tra* genes, three lie within the *tra* ORF, and one lies within the *spdB*-coding ORF.

Non-protein-coding sequences involved in replication. All known procaryotic plasmids contain a DNA locus that is required in *cis* for plasmid replication to occur. This region is the site of initiation of DNA synthesis and thus is referred to as the replication origin (*ori*). The only region of the repli-

korA 194 Thr-Glu-Asp-Gln-**ALA**-Ala-Leu-Leu-**GLY**-Val-Ala-Ala-Gly-Ala-Pro-Val-Leu-Leu-Ser-Arg

Frequency (%) 11 13 16 0 83 2 16 47 75 41 8 5 8 25 0 5 5 11 2 16

korB 13 Ala-Glu-Ala-Ala-Leu-Lys-Pro-Leu-**GLY**-Gln-Gln-Arg-Ile-Lys-Leu-Leu-Ala-Glu-Leu-Asp

Frequency (%) 5 13 8 8 0 8 0 47 75 0 8 22 0 0 33 0 5 8 25 0

korB 56 Asn-Glu-Val-Thr-**ALA**-Val-Ala-Pro-Asn-Thr-Ala-Arg-Ala-Trp-Ala-Lys-Ala-Glu-Ala-Glu

Frequency (%) 0 13 0 8 83 0 8 5 2 2 8 22 8 0 2 0 5 8 0 2

λcro 16 Gln-Thr-Lys-Thr-**ALA**-Lys-Asp-Leu-**GLY**-Val-Tyr-Gln-Ser-Ala-Ile-Asn-Lys-Ala-Ile-His

Frequency (%) 27 11 22 5 83 5 0 47 75 41 0 22 36 25 30 2 19 5 8 5

lacI 6 Leu-Tyr-Asp-Val-**ALA**-Glu-Tyr-Ala-**GLY**-Val-Ser-Tyr-Gln-Thr-Val-Ser-Arg-Val-Val-Asn

Frequency (%) 11 8 16 27 83 22 2 5 75 41 33 2 2 44 27 36 33 8 8 22

CAP 170 Arg-Gln-Glu-Ile-Gly-Gln-Ile-Val-**GLY**-Cys-Ser-Arg-Glu-Thr-Val-Gly-Arg-Ile-Leu-Lys

Frequency (%) 19 11 19 16 8 8 8 8 75 0 33 22 8 44 27 2 33 13 25 8

FIG. 3. Putative DNA-binding domains in pIJ101-encoded repressor proteins. Each predicted amino acid sequence was scanned for homology to the DNA-binding domains of 37 DNA-binding proteins (compiled from Pabo and Sauer [24]; and K. J. Kendall, unpublished results). The numbers underneath the amino acid sequences indicate the frequency with which that amino acid appears at the same relative position in the 37 DNA-binding domains. The conserved alanine and glycine residues are highlighted when they appear in the sequences. The best scores for *korA* and *korB* occurred starting at amino acids 194 and 13, respectively. However, the region of *korB* starting at amino acid 56 has also been included because preliminary evidence (D. Stein, unpublished results) suggests that this region may be involved in DNA binding. Also included for comparison are the DNA-binding domains of the λ *cro*, *lacI*, and CAP proteins (24).

cation-competent *Sac*II fragment that did not appear to contain an ORF was the segment between coordinates 1200 and 1700 (Fig. 2). It thus seems likely that this region is the origin of replication. Deng, Kieser, and Hopwood (Mol. Gen. Genet., in press) and H. Schrempf (personal communication) recently found a second replication locus that functions to initiate lagging-strand DNA synthesis and mapped this locus to a site within the noncoding region that we have identified between coordinates 7300 and 7600. Neither putative *ori* locus contains significant stretches of directly repeated nucleotide sequences, which are a commonly found feature of replication origins in procaryotic plasmids (reviewed in reference 27).

Nucleotide sequences of cloning vectors derived from pIJ101. A number of cloning vectors that employ the pIJ101 replication origin have been constructed. Most of the commonly used vectors (e.g., pIJ702 [19] and pIJ364 [21]) use the replication region of pIJ350 (21). pIJ350 was derived from two adjacent *Bcl*I fragments of pIJ102, a spontaneous in vivo deletion derivative of pIJ101 (21); the replication-competent segment was ligated to a 1,030-bp *Bcl*I fragment from *S. azureus*, conferring resistance to thiostrepton (28). During the course of our study, the two *Bcl*I fragments in pIJ350 were sequenced to allow determination of the extent of the pIJ101 replication region present in the various vectors. It was found that pIJ350 consists of the *Bcl*I (bp 7952)-*Bcl*I (bp 6796) fragment of pIJ101 in which bases 2086 to 6650 have been deleted. The only homologies between the endpoints of the deletion are a run of three C residues (Fig. 5). Thus, the replication region of pIJ350 consists of 3,109 bp; these data

Domain A.

tra 283 **GLY**-Arg-Arg-Met-Leu-Ile-Ala-**GLY**-Thr-Ser-**GLY-SER-GLY-LYS-SER**-Trp-**SER**

consensus Gly- - - - - - -Gly- - -Gly-Ser-Gly-Lys-Ser- -Thr

Thr Thr Ser

Domain B.

tra 308 Ser-Glu-Tyr-Ala-Asp-His-Arg-**LEU-VAL-VAL-VAL-ASP**-Pro

consensus Arg- - -Gly- - - - H - H - H - H -Asp-Asp

(hydrophobic) Glu

FIG. 4. Putative nucleotide-binding domain within the *tra* protein. The two-domain consensus nucleotide-binding fold sequence from Walker et al. (30) is aligned underneath the segments of the *tra* amino acid sequence that show the best homology.

TABLE 3. Inverted repeat sequences in pIJ101[a]

Start coordinate	Sequence	Free energy (kcal/mol)
1710	AGCCCCGCCGG TGAT CCGGCGGGGCT	−35.7
1941	GCGGGGCGCCAGCAC GAA GTGCCGGCGCCCCGC	−34.0
2471	CAGTCTGACACG CA CGTGTCAGGTTG	−12.0
2504	GGTAGGCCCCGCCGG TTT CCGGCGGGGCCATCC	−38.2
3083	GCCTGGACCCGGG CGA CCCGGGCCCCGGC	−24.1
4154	GGAGGCGCCGGCAC TTC GTGCAGGCGACTCC	−16.3
4471	GTGGCCCAGATCA GGA TGATCTCGGCCGC	−6.6
5244	CGACGGGTCGAGC C GCTCGACCAGGCG	−15.8
5875	GGGCCGGAGACCG TTCACG CGGTCTCCGGTCC	−30.2
6771	AGAAGCCCCGTCAGGC ACC GCCCGACGGGGCTTCT	−27.2
7925	GTGATCACTTCGG CACCCA CCGTAGTGATCAC	−19.4
8452	CGGGGCCGACCCCGGGCGAGTAATCC CA GGATTACTCCCGCGGCTTCGACCCCG	−38.8
8532	CCCCCGCCGTACGTCA CCGGGA TGACGTACGGCGGGGG	−44.1

[a] The complete sequence of pIJ101 was scanned for inverted repeats with a minimum of 10 matches out of 12 and a maximum inter-stem-loop of 10 bases. To simplify the search, only "perfect" matches were considered, i.e., no "loopouts" were permitted in the putative stem structures. Matched bases on each arm of the inverted repeats are indicated by underlining. The approximate free energy for each of the structures for RNA at 25°C is indicated.

allow the complete sequences of plasmids pIJ350 (21), pIJ702 (21), and pHYG1 (20) to be constructed from the previously published sequences of the thiostrepton (6), tyrosinase (2), and hygromycin (31) resistance genes, respectively. The collated sequences are available on request.

DISCUSSION

The complete nucleotide sequence of the *S. lividans* plasmid pIJ101 has been determined and analyzed for ORFs, and the results were correlated with previously published genetic data. The circular plasmid is 8,830 nucleotides in length and has a 72.98 mol% G+C composition, which is typical of *Streptomyces* DNA (14). Seven ORFs have been identified and correlated with specific phenotypic properties. At least four additional short ORFs for which no function can be assigned may be present on the plasmid.

Of the identified ORFs, two, *korA* and *korB*, encode repressors that control expression of several promoters identified on the plasmid (D. Stein, K. J. Kendall, and S. N. Cohen, unpublished results). Although the predicted amino acid sequences of these ORFs show no regions of homology to any other known repressor proteins, the derived amino acid sequence for *korA* contains an α-helix-turn-α-helix motif common to the DNA-binding domains of many DNA-binding proteins (24). Preliminary studies have identified electrophoretic protein bands corresponding in molecular weight to those predicted for both the *korA* and *korB* proteins (D. Stein, personal communication).

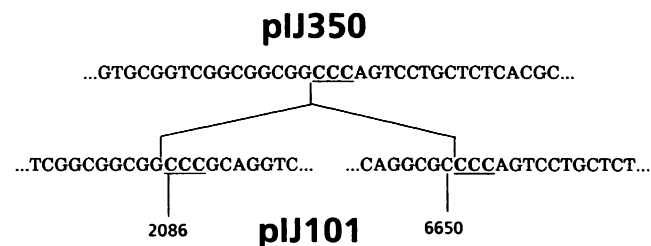


FIG. 5. Sequence of the deletion endpoints of pIJ101 present in pIJ350. The pIJ350 replication region is derived from an in vivo deletion of bases 2086 to 6650 of pIJ101. The sequence across the junction region is shown for both plasmids. A run of three C residues common to both ends of the deletion is highlighted.

Four presumptive ORFs involved in plasmid transfer have been identified. *tra* could encode a large protein of M_r 77,000. Insertions into this ORF completely prevent the pocking phenotype indicative of plasmid transfer (20, 21). No other frame-disrupting insertions into pIJ101 have such dramatic effects, suggesting that this ORF encodes the major protein required for plasmid transfer. No clues as to the mechanism by which transfer occurs can be deduced from the primary amino acid sequence. The predicted protein has no significant hydrophobic domains, suggesting that it probably is not associated with membranes, and it has very little similarity to other known proteins or to DNA-binding domains; however, it does have a region bearing some resemblance to known nucleotide-binding fold domains such as those found in ATPases (30).

Insertions into the other three ORFs (*spdA*, *spdB*, and *kilB*) produce the Spd phenotype. It has been suggested that the Spd phenotype results from the inability of plasmids to undergo intramycelial transfer (18). The putative *spdB* protein contains three strongly hydrophobic domains in its amino-terminal half, suggesting that it may be associated with membranes. Single strongly hydrophobic domains are also present in *spdA* and, to a lesser extent, in *kilB*.

A large ORF is present in the 2.1-kb *Sac*II fragment, which has been shown to be sufficient for autonomous replication of pIJ101 derivatives in *Streptomyces* species (21). The presumptive *rep* ORF provides very little structural information about the protein, which has no discernible similarities to known *rep* proteins of other plasmids. It is not clear whether this protein alone (plus, presumably, a functional *ori* DNA sequence) is sufficient for replication, as a second small ORF (orf56) is also present on the *Sac*II fragment. The approximate location of the presumed pIJ101 *ori* sequence can be deduced from an examination of the FRAME plot, which indicates only one noncoding region within the *Sac*II fragment. A second origin used for lagging-strand replication may be present in a noncoding region upstream of the *korB* gene in a segment that is not essential for autonomous replication (Deng et al., in press; H. Schrempf, personal communication). An additional region worthy of mention is centered around the *Kpn*I site at bp 8430; insertions into this site and deletions removing this site appear to destabilize pIJ101 derivatives (16; S. Lee, personal communication). The FRAME plot around this site is not consistent with a protein product being encoded by the

region, although the G+C composition is clearly not random. There may be trivial reasons for this discrepancy (e.g., sequencing errors or clonally derived deletions in the plasmid preparations used for sequencing). Alternatively, an ORF that does not conform to the usual *Streptomyces* ORF characteristics could be present in this region.

Five promoters have been localized on the pIJ101 sequence (Fig. 2). Of the three previously reported promoters, pIJ101A (7) appears to be the promoter for the *kilB* gene and pIJ101B (8) appears to be the promoter for the *tra* gene. Transcription from the latter promoter possibly extends through the clustered *spdA*, *spdB*, and orf66 reading frames. The reported pIJ101C promoter (11) lies in the middle of the *korB* ORF; the function of this promoter is not immediately obvious. Two other promoters located upstream of the *korA* and *korB* ORFs, respectively, have been identified (D. Stein and K. J. Kendall, unpublished observations). Thus, it appears that the *tra* and *korA* genes are transcribed divergently from their respective promoters, consistent with the orientation of the ORFs assigned to these genetic functions. Transcription from both promoters is controlled by the product of the *korA* gene (D. Stein, K. J. Kendall, and S. N. Cohen, unpublished results).

The only published pIJ101 segment known to have transcription-terminating activity in *Streptomyces* spp. is located between the convergently transcribed *korA* and *korB* genes (12) (Fig. 2). This terminator functions in both orientations in *Streptomyces* spp. and presumably acts to prevent read-through interference of transcription for both genes. Its sequence consists of a G+C-rich region of dyad symmetry (ΔG for RNA of -27.2 kcal) but lacks the stretches of poly(U) commonly seen with rho-independent terminators in *E. coli* (25). Similar regions of strong dyad symmetry can be found downstream of the *rep*, *kilB*, orf56, orf85, and *tra*/*spdA*/*spdB*/orf66 ORFs, suggesting that these may be involved in transcription termination.

Determination of the primary nucleotide sequence of pIJ101 is a further step in elucidating the biology of this plasmid. In addition to facilitating the construction of pIJ101-derived cloning vectors, the sequence should allow progress in understanding the mechanisms of replication and transfer of *Streptomyces* species plasmids and should enable detailed studies of the genetic switches involved in the regulation of pIJ101 gene expression.

ACKNOWLEDGMENTS

These studies were supported by Public Health Service grant AI08619 to S.N.C. from the National Institutes of Health.

We thank David Stein, Stephen Lee, and Michael Brasch for helpful discussions and Mark Buttner for communicating unpublished information and for helpful comments.

LITERATURE CITED

1. **Bankier, A. T., K. M. Weston, and B. G. Barrell.** 1987. Random cloning and sequencing by the M13/dideoxynucleotide chain termination method. Methods Enzymol. **155**:51–93.
2. **Bernan, V., D. Fipula, W. Herber, M. J. 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.
3. **Bibb, M. J., M. J. Bibb, J. M. Ward, and S. N. Cohen.** 1985. Nucleotide sequences encoding and promoting expression of three antibiotic resistance genes indigenous to *Streptomyces*. Mol. Gen. Genet. **199**:26–36.
4. **Bibb, M. J., and S. N. Cohen.** 1982. Gene expression in *Streptomyces*: construction and application of promoter-probe 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. **Bibb, M. J., and D. A. Hopwood.** 1981. Genetic studies of the fertility plasmid SCP2 and its SCP2* variants in *Streptomyces coelicolor* A3(2). J. Gen. Microbiol. **126**:427–442.
7. **Buttner, M. J., and N. L. Brown.** 1985. RNA polymerase-DNA interactions in *Streptomyces*: in vitro studies of a *Streptomyces lividans* plasmid promoter with *Streptomyces coelicolor* RNA polymerase. J. Mol. Biol. **185**:177–188.
8. **Buttner, M. J., and N. L. Brown.** 1987. Two promoters from the *Streptomyces* plasmid pIJ101 and their expression in *Escherichia coli*. Gene **51**:179–186.
9. **Buttner, M. J., I. M. Fearnley, and M. J. Bibb.** 1987. The agarase gene (*dagA*) of *Streptomyces coelicolor* A3(2): nucleotide sequence and transcriptional analysis. Mol. Gen. Genet. **209**:101–109.
10. **Chou, P. Y., and G. D. Fasman.** 1978. Prediction of the secondary structure of proteins from their amino acid sequence. Adv. Enzymol. **47**:45–147.
11. **Deng, Z., T. Kieser, and D. A. Hopwood.** 1986. Expression of a *Streptomyces* plasmid promoter in *Escherichia coli*. Gene **43**: 295–300.
12. **Deng, Z., T. Kieser, and D. A. Hopwood.** 1987. Activity of a *Streptomyces* transcriptional terminator in *Escherichia coli*. Nucleic Acids Res. **15**:2665–2675.
13. **Eisenberg, D., R. M. Weiss, and T. C. Terwilliger.** 1982. The helical hydrophobic moment: a measure of the amphiphilicity of a helix. Nature (London) **299**:371–374.
14. **Enquist, L. W., and S. G. Bradley.** 1971. Characterization of deoxyribonucleic acid from *Streptomyces venezuelae* species. Dev. Ind. Microbiol. **12**:225–236.
15. **Garnier, J., D. J. Osguthorpe, and B. Robson.** 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol. **120**: 97–120.
16. **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, England.
17. **Hopwood, D. A., T. Kieser, H. M. Wright, and M. J. Bibb.** 1983. Plasmids, recombination and chromosome mapping in *Streptomyces lividans* 66. J. Gen. Microbiol. **129**:2257–2269.
18. **Hopwood, D. A., D. J. Lydiate, F. Malpartida, and H. M. Wright.** 1985. Conjugative sex plasmids of *Streptomyces*, p. 615–634. *In* D. R. Helinski, S. N. Cohen, D. B. Clewell, D. A. Jackson, and A. Hollaender (ed.), Plasmids in bacteria. Plenum Press, New York.
19. **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.
20. **Kendall, K. J., and S. N. Cohen.** 1987. Plasmid transfer in *Streptomyces lividans*: identification of a *kil-kor* system associated with the transfer region of pIJ101. J. Bacteriol. **169**:4177–4183.
21. **Kieser, T., D. A. Hopwood, H. M. Wright, and C. J. Thompson.** 1982. pIJ101, a multicopy broad host range *Streptomyces* plasmid: functional analysis and development of DNA cloning vectors. Mol. Gen. Genet. **185**:223–238.
22. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
23. **Messing, J. M.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20–78.
24. **Pabo, C. O., and R. T. Sauer.** 1984. Protein-DNA recognition. Annu. Rev. Biochem. **53**:293–322.
25. **Platt, T.** 1986. Transcription termination and the regulation of gene expression. Annu. Rev. Biochem. **55**:339–372.
26. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequenc-

ing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
27. **Scott, J. R.** 1984. Regulation of plasmid replication. Microbiol. Rev. **48**:1–23.
28. **Thompson, C. J., J. M. Ward, and D. A. Hopwood.** 1982. Cloning of antibiotic resistance and nutritional genes in *Streptomyces*. J. Bacteriol. **151**:668–677.
29. **Uchiyama, H., and B. Weisblum.** 1985. *N*-Methyl transferase of *Streptomyces erythraeus* that confers resistance to the macrolide-lincosamide-streptogramin B antibiotics: amino acid sequence and its homology to cognate R-factor enzymes from pathogenic bacilli and cocci. Gene **38**:103–110.
30. **Walker, J. E., M. Saraste, M. J. Runswick, and N. J. Gay.** 1982. Distantly related sequences in the α- and β-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. **1**:945–951.
31. **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.
32. **Zalacain, M., A. Gonzalez, M. C. Guerrero, R. J. Mattaliano, F. Malpartida, and A. Jimenez.** 1986. Nucleotide sequence of the hygromycin B phosphotransferase gene from *Streptomyces hygroscopicus*. Nucleic Acids Res. **14**:1565–1581.