Volume 11 Number 19 1983 Nucleic Acids Research

Analysis of a bacterial hygromycin B resistance gene by transcriptional and translational fusions and by DNA sequencing

Kevin R.Kaster, Stanley G.Burgett, R.Nagaraja Rao and Thomas D.Ingolia*

Lilly Research Laboratories, 307 E. McCarty Street, Indianapolis, IN 46285, USA

Received 24 June 1983; Revised and Accepted 9 September 1983

ABSTRACT

We have characterized hygromycin B and apramycin resistance genes from an E. coli plasmid. We have localized the coding and control regions of these genes by deletion of DNA fragments from plasmids containing the genes. It was found that polypeptides with apparent molecular weights of 33,000 and 31,500 daltons are encoded by the apramycin resistance gene and polypeptides with apparent molecular weights of 42,500 and 41,500 daltons are encoded by the hygromycin B resistance gene. DNA sequence analysis identified a typical promoter sequence upstream of the genes. Deletion of this promoter eliminated both resistance phenotypes, and hygromycin B resistance could be restored by substitution of a promoter from a foreign gene. The region known to be necessary for hygromycin B resistance contained an open reading frame large enough to encode the hygromycin B resistance gene product. This open reading frame was fused with the amino terminus of β-galactosidase. This hybrid gene conferred hygromycin resistance to E. coli, and expression of resistance was under IPTG control.

INTRODUCTION

A dominant, selectable genetic marker functional in heterologous systems would be valuable for a variety of reasons. For example, it could be used on shuttle vectors to select for cells containing foreign DNA or as a vehicle for isolating promoters and control sequences from complex DNA. Ideally, the gene encoding the dominant marker would still be active after fusion with sequences from the host organism which promote expression.

One candidate for a widely useful dominant marker is the hygromycin B resistance gene from E. coli. Hygromycin B is an aminoglycoside antibiotic that inhibits protein synthesis and kills bacteria, fungi, and higher eukaryotes (1,2). A gene was recently isolated from E. coli which confers hygromycin resistance to this organism by phosphorylating the antibiotic (Rao, R.N., N.E. Allen, J.N. Hobbs, Jr., W.E. Alborn, Jr., H.A. Kirst and J.W. Paschal. Submitted to Antimicrobial Agents and Chemotherapy). During the course of this earlier work, the gene was moved to a multicopy plasmid derived from pBR322. We wanted to characterize the drug resistance genes on

this hybrid plasmid so that we could test the utility of the genes or derivatives of the genes as dominant selectable markers.

MATERIALS AND METHODS

Bacterial Strains and Plasmids

The strains used were E. coli K12 derivatives RR1 (3), JA221, CSR603 (4, 5), DH1 (6), and BE904 (a lac Δ mutant; Kuhstoss, S., H. Hsiung, R. Belagaje, S.G. Burgett and R.N. Rao, manuscript in preparation). All plasmids were derivatives of pBR322 (7).

Bacterial Culture Conditions

Drug-resistant bacteria were selected on TY plates (8) containing the appropriate antibiotics at the following concentrations: Ampicillin (Ap), 50 μg/ml; apramycin (Am), 100 μg/ml; hygromycin B (Hm), 200 μg/ml; and tetracycline (Tc), 25 μg/ml. Assessment of Hm resistance was made by patch tests from cells grown on a different antibiotic except in the case of cells containing pTI105, which could be plated directly after transformation on hygromycin B-containing plates. β-galactosidase assays were performed according to Miller (8).

Maxicell Labelling

CSR603 cells containing plasmids were labelled with ^{35}S-methionine according to Sancar, et al. (4,5) except that after UV irradiation the cells were grown overnight in the presence of 100 μg/ml cycloserine.

Electrophoresis

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed according to Ames (9).

Formaldehyde-agarose gels were run as described by Zaret and Sherman (10). RNA blots were performed as outlined by Thomas (11).

DNA Sequencing

DNA sequence analysis was performed using the chemical cleavage method of Maxam and Gilbert (12). DNA labelling was carried out as described previously (13).

Conversion of Sac I site to Bam HI site.

After digestion with SacI, 4μg of pKC222 was incubated for 15 minutes at 37°C in 5 μl of 50 mM Tris (pH7.5), 10mM $MgCl_2$, 20μM each of dCTP, dATP, TTP and dGTP with 3 units of DNA polymerase I large fragment. After heat inactivation of the polymerase, BamHI linkers were ligated onto the blunt-ended DNA molecules, digested with Bam HI, and religated. After digestion with SacI to lower the number of parental plasmids, the DNA was

used to transform strain RR1 to ampicillin resistance. Sequence analysis of one of the resulting plasmids, pTI104, showed that the treatment with DNA polymerase I large fragment had removed the 3'-extension generated by SacI.

Removal of 3' Extensions with T4 DNA Polymerase.

The 3' extension left by restriction enzyme HphI was removed with T4 DNA polymerase. About 1 μg of restricted DNA was incubated in 10 μl of 33 mM Tris.HCl, pH 7.8, 67 mM potassium acetate, 10 mM magnesium acetate, 0.5 mM dithiothreitol, 0.1 mg/ml bovine serum albumin, and all four dXTP's at 150 μM at 37 degrees C for 5 minutes with 3 units of T4 DNA polymerase. The reaction was quenched at 65 degrees C for 5 minutes after adding 1 μl of 50 mM EDTA.

Materials

All enzymes were obtained from either Boehringer-Mannheim, Bethesda Research Labs or New England Biolabs. Apramycin and hygromycin B were obtained from Eli Lilly and Company. Ampicillin, tetracycline, cycloserine, and agarose were from Sigma Chemical Company. [^{35}-S]-methionine was from Amersham. 5-bromo-4-chloro-3-indolyl-B-D-galactoside (XG) was purchased from Bachem. BamHI linkers were from Collaborative Research.

RESULTS

Genetic Analysis of the Plasmid-Borne Genes

Resistance genes for hygromycin B (Hm) and apramycin (Am) are contained on pKC222, shown in figure 1a. This plasmid is derived from pBR322 (via pKC7) and includes a BglII-SalI fragment from a naturally occurring plasmid cloned into the BglII and SalI sites of pKC7 (14, Rao, R.N., N.E. Allen, J.N. Hobbs, Jr., W.E. Alborn, Jr., H.A. Kirst and J.W. Paschal. Submitted to Antimicrobial Agents and Chemotherapy). Hygromycin B resistance is mediated by an aminocyclitol phosphotransferase that inactivates by covalent addition of a phosphate to the 4-position of hygromycin B (Rao, R.N., N.E. Allen, J.N. Hobbs, Jr., W.E. Alborn, Jr., H.A. Kirst and J.W. Paschal. Submitted to Antimicrobial Agents and Chemotherapy). The gene is abbreviated as aph(4). Apramycin resistance is mediated by an aminocyclitol acetyltransferase that acetylates the nitrogen at the 3-position of apramycin (14). The gene is abbreviated as aac(3)IV.

Deletion experiments showed that information encoding Hm^r is on the left (BglII) side and information for Am^r is on the right (SalI) side of the insert in pKC222 (figure 1a). Deletion of a SacI piece (fragment 1) from the left side of the insert eliminates Am^r but not Hm^r. Deletion

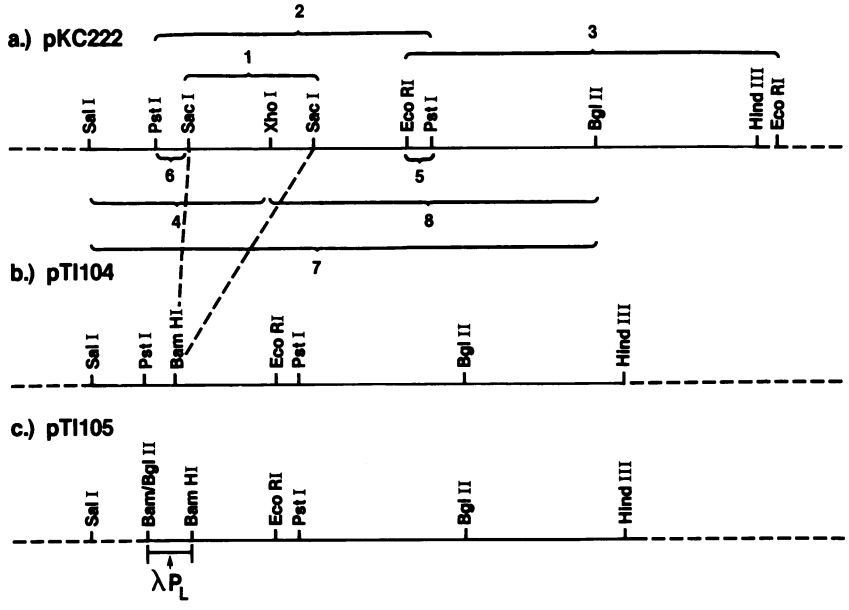

Figure 1. Plasmids encoding Hm^r and/or Am^r

A) pKC222 is derived from pBR322 via pKC7 and contains a SalI-BglII fragment isolated from a naturally occurring plasmid, pKC203, that confers resistance to apramycin and hygromycin B. (Rao, R.N., N.E. Allen, J.N. Hobbs, Jr., W.E. Alborn, Jr., H.A. Kirst and J.W. Paschal. Submitted to Antimicrobial Agents and Chemotherapy).

B) The BamHI site in pTI104 is derived from a 10 base pair long BamHI linker inserted into pKC222 after cutting with SacI and removal of the 3' extension with DNA polymerase.

C) The SalI-BamHI fragment of pTI104 is replaced by a Sal-BamHI fragment containing the λ leftward promoter. The BamHI site is inserted into λp_L at the HaeIII site upstream of the translational initiation site for the λ "N" protein and downstream from the transcriptional start site for "N" in RNA (16).

of an EcoRI piece (fragment 3) from the right side of the insert eliminates Hm^r but not Am^r. Subcloning the PstI piece (fragment 2) of pKC222 into pBR322 allows expression of Am^r but not Hm^r, which is consistent with the results of the deletion experiments.

Further subcloning experiments from pKC222 suggested that expression of Hm^r by the aph(4) gene requires information from the aac(3)IV gene. A SalI-XhoI subclone (fragment 4) did not express resistance to either Am or Hm, and neither did an Xho-BglII subclone (fragment 8), yet the sum of the two subclones (fragment 7) encoded resistance to both Hm and Am. Since

deletion of the SacI piece (fragment 1) still permits Hm^r, yet the XhoI-BglII piece (fragment 8) is not sufficient for Hm^r, it appears that information to the left of the SacI site (including fragment 6 in figure 1a) is required for expression of Hm^r.

Physical Characterization of the Am^r and Hm^r Gene

The two most likely explanations for the importance of information to the left of the leftmost SacI site (figure 1a) for expression of Hm^r are: 1) the two resistances are conferred by one polypeptide with distinct domains; or 2) the two resistances are on separate proteins translated from a single mRNA.

The first hypothesis was tested by examining the polypeptide products of several plasmids using E. coli maxicells (4,5), which allow identification of plasmid encoded polypeptides. Figure 2 shows that the aph(4) and aac(3)IV genes encode distinct polypeptides. The hygromycin resistance gene encodes a polypeptide of about 42,000 daltons, and the apramycin resistance gene encodes a polypeptide of about 32,000 daltons. In E. coli cells containing the plasmid pKC222, which confers resistance to Hm, Am, and Ap, the predominant plasmid-encoded polypeptides are a doublet at about 32,000 daltons and another doublet at about 42,000 daltons. In cells containing pKC237, which is derived from pKC222 by deletion of an EcoRI piece (fragment 3 of figure 1a) and confers Am^r and Ap^r but not Hm^r, the doublet at 42,000 daltons is absent. In cells containing pKC241, which is derived from pKC222 by deletion of a SacI piece (fragment 1 of figure 1a) and confers Ap^r and Hm^r but not Am^r, the doublet at 32,000 is completely missing. (The remaining band near 32,000 comigrates with the Ap^r gene products of pBR322).

The second hypothesis to explain the requirement for aac(3)IV gene DNA in order to express the aph(4) gene and obtain Hm^r is that the genes are part of a polycistron with the aac(3)IV gene upstream from the aph(4) gene. This hypothesis predicts that a large mRNA species complementary to both the aac(3)IV and aph(4) genes should be found in cells containing pKC222.

Consistent with this idea is the finding that probes from both the aph(4) and aac(3)IV genes hybridize to large RNA molecules from E. coli cells containing pKC222 but not from cells without plasmid (data not shown).

Analysis of the Transcriptional Activator Region

The region containing the aac(3)IV and aph(4) genes should contain a promoter within the PstI-SacI piece (fragment 6 of figure 1a) of pKC222 based on the results of genetic experiments described earlier. Since the

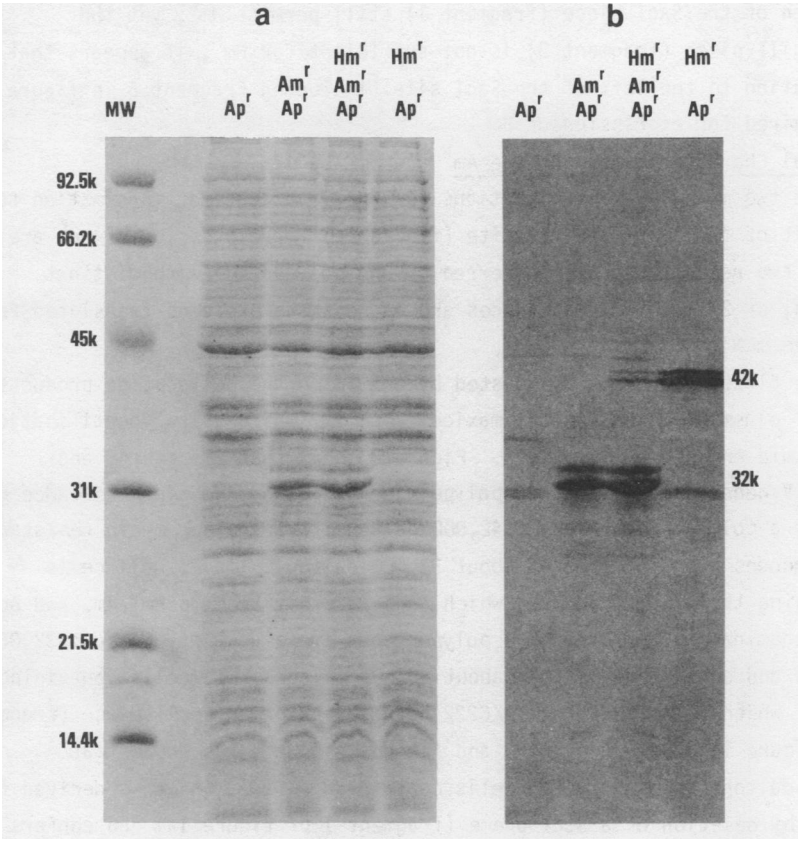

Figure 2. Identification of Plasmid Encoded Polypeptides Using E. coli Maxicells

Plasmids were transformed into the E. coli strain CSR603, a uvrA mutant. Plasmid encoded polypeptides were examined by labelling cells with [^{35}S]-methionine after UV inactivation and overnight growth in cycloserine. Panel a:Coomassie Blue stained 12.5 percent SDS-polyacrylamide gel. The first lane shows molecular weight markers of 92.5K, 66.2K, 45K, 31K, 21.5K and 14.4K daltons. The following lanes show total cellular protein from CSR603 containing these drug resistances: Ap^r) cells contain pBR322; Am^r Ap^r) cells contain pKC237, a derivative of pKC222 missing an EcoRI fragment (fragment 3 of figure 1a); Hm^r Am^r Ap^r) cells contain pKC222 (figure 1a); Hm^r Ap^r) cells contain pTI105 (figure 1c).
Panel b: the autoradiogram of the gel shown in panel a, which should identify the plasmid encoded polypeptides (4,5). The film (XAR-5, Kodak) was exposed for 24 hours at room temperature.

subcloned PstI piece (fragment 2) conferred Am^r, it most likely contains a promoter. Also, since deletion of a SacI piece (fragment 1) still permitted expression of Hm^r, the promoter must not be within the SacI piece.

Table 1. Sequence of the Apramycin Resistance Gene Promoter

```
....GACTATTTGCAACAGTGCCGTTGATCGTGCTATGATCGACTGATGTCATCAGCGGTGGAGTGCAATGTCGTGCAATACGAATGGCGAAAAGCCGAGCTC
      *** *  *          ** *** **                  shine-  met                              SacI
    tgttgaca-ttt       atttgttataatg   cat          dalgarno
     -35      6-9 b.p.     -10      4-7
                                    b.p.
```

The upper case letter are sequences obtained by Maxam-Gilbert sequencing (12) from pKC222 (figure 1a). The sequence is derived from fragment 6, figure 1a, and reads left to right (note the SacI site, which delineates the right end of fragment 6).
The lower case letters underneath the pKC222 sequence are consensus sequences for bacterial promoters (22).
Asterisks indicate matches between the apramycin resistance gene promoter and the E. coli consensus sequence.

Therefore, since the SacI fragment is contained within the PstI fragment, the promoter must be between the PstI and SacI sites. This prediction was tested by primary sequence analysis and by promoter exchange using known promoters.

The primary sequence analysis of the PstI-SacI fragment demonstrated typical E. coli promoters and translational initiation signals upstream of the SacI sites, as shown in Table 1. These sequences would act in a left to right direction based on the orientation shown in figure 1a, as expected from the genetic studies.

In order to test the function of this proposed promoter, it was necessary to change the SacI site to a BamHI site. This was accomplished by removing the 3'-extension generated by SacI digestion and ligating 10-mer BamHI linkers to the flush ends. The resulting plasmid, pIT104, is shown in figure 1B.

The proposed promoter sequence, now contained on a PstI-BamHI fragment of pIT104, was able to promote β-galactosidase expression when fused to a fragment of the lacZ gene lacking activator sequences. The PstI-BamHI fragment was ligated to pMC1403, a plasmid containing the lacZ gene with a BamHI site at the codon for amino acid 8 (15). As shown in table 2, the BamHI site of pMC1403 should have the same translational reading frame as that proposed for the Pst-BamHI fragment from pIT104. In fact, cells missing the genomic copy of the lacZ gene but harboring this hybrid plasmid produced dark blue colonies on X-G ampicillin plates (8), indicating that the PstI-BamHI fragment could indeed drive transcription and translation. In a complementary experiment, it was found that a characterized promoter could drive expression of Hm^r when spliced into the BamHI site of pIT104. A promoter from bacteriophage λ, λP_L, was ligated into pIT104 at the SalI-BamHI sites. The BamHI site in λP_L was artificially inserted at a HaeIII site in the untranslated leader region (115 bases downstream from the

Table 2. Fusion of the Apramycin Resistance Gene with β-galactosidase

pTI104	ATG/TCG/TGC/AAT/ACG/AAT/GGC/GAA/AAG/CCG/ccg/gat/ccg
β-gal	CCG/GAT/CCG

The upper line represents the sequence obtained from pTI104 near the BamHI site (figure 1b). This sequence is the same as the 3' end of the sequence shown in Table 1 except that the SacI site 3'-extension has been eliminated and a 10 base pair long BamHI linker has been inserted. The lower case letters represent nucleotides from the linker. The lower line represents the reading frame at the BamHI site of β-galactosidase in pMC1403 (15).

mRNA start site) and would not bring translational activator sequences with it. (16; Kuhstoss, S., H. Hsiung, R. Belagaje, S. G. Burgett and R. N. Rao, manuscript in preparation.) This hybrid plasmid, pIT105, shown in figure 1c, efficiently conferred Hm^r to E. coli.

DNA Sequence Analysis of the Hygromycin B Resistance Gene.

Previous experiments showed that the hygromycin B phosphotransferase was encoded within a 1.5 kb DNA fragment bounded by a SacI site and a BglII site (figure 1). In order to localize the gene more precisely, the primary sequence of this region was determined.

The sequencing strategy is shown in figure 3, and the primary sequence in figure 4. Computer analysis of the primary sequence identified a large open reading frame bounded by the SacI site on the 5' end and the BglII site on the 3' end. The amino acid sequence predicted for the polypeptide encoded by this open reading frame is shown in figure 4.

The open reading frame postulated to encode the hygromycin B phosphotransferase is preceded by a GGGGG sequence 9 bases upstream from the ATG which is postulated to initiate translation of the polypeptide. This sequence and the spacing are appropriate for a ribosome binding site in E. coli (17).

Fusion of the Hygromycin B Resistance Gene with β-Galactosidase

We fused the protein coding region of β-galactosidase with the putative coding region of the hygromycin B resistance gene. We utilized the unique HphI restriction site in the hygromycin B gene, which cuts 10 bp downstream of the initiating ATG (as shown in figure 5), as the junction site for the fusion. The 3' overhang of the HphI site was eliminated with T4 DNA polymerase (see Materials and Methods), and the blunt ended molecules were ligated to the β-galactosidase coding sequence on pUC8 (18). This β-galactosidase sequence has multiple restriction sites, including a unique

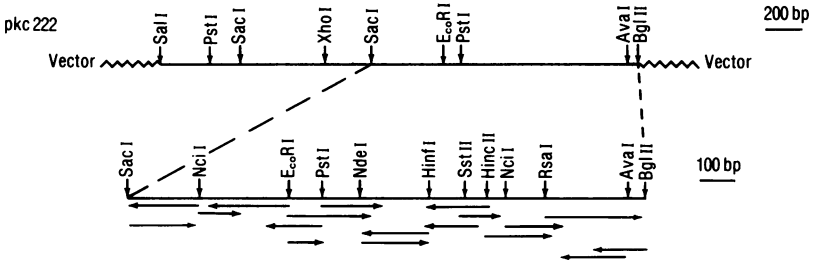

Figure 3. Sequencing Strategy Used for the Hygromycin B Resistance Gene
The primary sequence was determined using the chemical cleavage method of Maxam and Gilbert (12). The length of the arrows indicates the distance sequenced from the labelled end.

HincII site, downstream of the translation initiating ATG sequence. The reading frame postulated to encode the hygromycin B phosphotransferase polypeptide would be maintained by ligating together these blunt-ended molecules, as shown in figure 5. The resulting plasmid, named pKC307, was subjected to DNA sequence analysis and the sequence predicted in figure 5 was verified.

This hybrid gene conferred hygromycin B resistance to E. coli. The hygromycin B resistance was more efficiently conferred in the presence of IPTG, an inducer of β-galactosidase activity, than in the absence of IPTG, as shown in table 3. It should be noted that while the efficiency of plating on hygromycin B plates is low in both cases, efficiency of plating increased about 100 fold in the presence of IPTG.

DISCUSSION

Our analysis of a hygromycin B resistance gene from an E. coli plasmid shows it to be linked to an apramycin resistance gene as part of a polycistron. The apramycin resistance gene (aac(3)IV) precedes the hygromycin B resistance gene (aph(4)) on the primary transcript. The evidence for this configuration is threefold. First, the expression of the aph(4) gene requires information in front of the aac(3)IV gene. That is, the absence of a SacI-PstI piece of DNA, fragment 6 of figure 1a, eliminates expression of both genes, the absence of DNA downstream (fragment 1) eliminates expression of Am^r, and the absence of a piece of DNA even farther downstream (fragment 3) eliminates expression of Hm^r. Second, the two resistance genes encode separate polypeptides as shown by maxicell

```
GAGCTCATGAGCGGAGAACGAGATGACGTTGGAGGGGCAAGGTCGCGCTGATTGCTGGGGC
        10        20        30        40        50        60

AACACGTGGAGCGGATCGGGGATTGTCTTTCTTCAGCTCGCTGATGATATGCTGACGCTCA
        70        80        90       100       110       120

ATGCCGTTTGGCCTCCGACTAACGAAAATCCCGCATTTGGACGGCTGATCCGATTGGCACGG
       130       140       150       160       170       180

                                                    MET LYS LYS
CGGACGGCGAATGGCGGAGCAGACGCTCGTCCGGGGGCAATGAGAT      ATG AAA AAG
    190       200       210       220       230

    5                   10                  15
PRO GLU LEU THR ALA THR SER VAL GLU LYS PHE LEU ILE GLU LYS PHE
CCT GAA CTC ACC GCG ACG TCT GTC GAG AAG TTT CTG ATC GAA AAG TTC
240           250           260           270           280

20                  25                  30                  35
ASP SER VAL SER ASP LEU MET GLN LEU SER GLU GLY GLU GLU SER ARG
GAC AGC GTC TCC GAC CTG ATG CAG CTC TCG GAG GGC GAA GAA TCT CGT
  290           300           310           320           330

                40                  45                  50
ALA PHE SER PHE ASP VAL GLY GLY ARG GLY TYR VAL LEU ARG VAL ASN
GCT TTC AGC TTC GAT GTA GGA GGG CGT GGA TAT GTC CTG CGG GTA AAT
     340           350           360           370           380

            55                  60                  65
SER CYS ALA ASP GLY PHE TYR LYS ASP ARG TYR VAL TYR ARG HIS PHE
AGC TGC GCC GAT GGT TTC TAC AAA GAT CGT TAT GTT TAT CGG CAC TTT
        390           400           410           420           430

        70                  75                  80
ALA SER ALA ALA LEU PRO ILE PRO GLU VAL LEU ASP ILE GLY GLU PHE
GCA TCG GCC GCG CTC CCG ATT CCG GAA GTG CTT GAC ATT GGG GAA TTC
          440           450           460           470

    85                  90                  95
SER GLU SER LEU THR TYR CYS ILE SER ARG ARG ALA GLN GLY VAL THR
AGC GAG AGC CTG ACC TAT TGC ATC TCC CGC CGT GCA CAG GGT GTC ACG
480           490           500           510           520

100                 105                 110                 115
LEU GLN ASP LEU PRO GLU THR GLU LEU PRO ALA VAL LEU GLN PRO VAL
TTG CAA GAC CTG CCT GAA ACC GAA CTG CCC GCT GTT CTG CAG CCG GTC
  530           540           550           560           570

                120                 125                 130
ALA GLU ALA MET ASP ALA ILE ALA ALA ALA ASP LEU SER GLN THR SER
GCG GAG GCC ATG GAT GCG ATC GCT GCG GCC GAT CTT AGC CAG ACG AGC
     580           590           600           610           620

            135                 140                 145
GLY PHE GLY PRO PHE GLY PRO GLN GLY ILE GLY GLN TYR THR THR TRP
GGG TTC GGC CCA TTC GGA CCG CAA GGA ATC GGT CAA TAC ACT ACA TGG
        630           640           650           660           670

        150                 155                 160
ARG ASP PHE ILE CYS ALA ILE ALA ASP PRO HIS VAL TYR HIS TRP GLN
CGT GAT TTC ATA TGC GCG ATT GCT GAT CCC CAT GTG TAT CAC TGG CAA
          680           690           700           710

    165                 170                 175
THR VAL MET ASP ASP THR VAL SER ALA SER VAL ALA GLN ALA LEU ASP
ACT GTG ATG GAC GAC ACC GTC AGT GCG TCC GTC GCG CAG GCT CTC GAT
720           730           740           750           760

180                 185                 190                 195
GLU LEU MET LEU TRP ALA GLU ASP CYS PRO GLU VAL ARG HIS LEU VAL
GAG CTG ATG CTT TGG GCC GAG GAC TGC CCC GAA GTC CGG CAC CTC GTG
  770           780           790           800           810
```

```
                200                 205                 210
HIS ALA ASP PHE GLY SER ASN ASN VAL LEU THR ASP ASN GLY ARG
CAC GCG GAT TTC GGC TCC AAC AAT GTC CTG ACG GAC AAT GGC CGC
     820          830          840          850          860

                215                 220                 225
ILE THR ALA VAL ILE ASP TRP SER GLU ALA MET PHE GLY ASP SER GLN
ATA ACA GCG GTC ATT GAC TGG AGC GAG GCG ATG TTC GGG GAT TCC CAA
           870          880          890          900

           230                 235                 240
TYR GLU VAL ALA ASN ILE PHE PHE TRP ARG PRO TRP LEU ALA CYS MET
TAC GAG GTC GCC AAC ATC TTC TTC TGG AGG CCG TGG TTG GCT TGT ATG
 910          920          930          940          950

       245                 250                 255
GLU GLN GLN THR ARG TYR PHE GLU ARG ARG HIS PRO GLU LEU ALA GLY
GAG CAG CAG ACG CGC TAC TTC GAG CGG AGG CAT CCG GAG CTT GCA GGA
    960          970          980          990          1000

    260                 265                 270
SER PRO ARG LEU ARG ALA TYR MET LEU ARG ILE GLY LEU ASP GLN LEU
TCG CCG CGG CTC CGG GCG TAT ATG CTC CGC ATT GGT CTT GAC CAA CTC
      1010         1020         1030         1040         1050

275                 280                 285                 290
TYR GLN SER LEU VAL ASP GLY ASN PHE ASP ASP ALA ALA TRP ALA GLN
TAT CAG AGC TTG GTT GAC GGC AAT TTC GAT GAT GCA GCT TGG GCG CAG
         1060         1070         1080         1090         1100

                295                 300                 305
GLY ARG CYS ASP ALA ILE VAL ARG SER GLY ALA GLY THR VAL GLY ARG
GGT CGA TGC GAC GCA ATC GTC CGA TCC GGA GCC GGG ACT GTC GGG CGT
           1110         1120         1130         1140

           310                 315                 320
THR GLN ILE ALA ARG ARG SER ALA ALA VAL TRP THR ASP GLY CYS VAL
ACA CAA ATC GCC CGC AGA AGC GCG GCC GTC TGG ACC GAT GGC TGT GTA
 1150         1160         1170         1180         1190

       325                 330                 335
GLU VAL LEU ALA ASP SER GLY ASN ARG ARG PRO SER THR ARG PRO ARG
GAA GTA CTC GCC GAT AGT GGA AAC CGA CGC CCC AGC ACT CGT CCG AGG
    1200         1210         1220         1230         1240

    340
ALA LYS GLU
GCA AAG GAA TAGAGTAGATGCCGACCGAACAAGAGCTGATTTCGAGAACGCCTCAGCCAG
      1250         1260         1270         1280         1290         1300

CAACTCGCGCGAGCCTAGCAAGGCAAATGCGAGAGAACGGCCTTACGCTTGGTGGCACAGTTC
     1310         1320         1330         1340         1350         1360

TCGTCCACAGTTCGCTAAGCTCGCTCGGCTGGGTCGCGGGAGGGCCGGTCGCAGTGATTCAGG
  1370         1380         1390         1400         1410         1420         1430

CCCTTCTGGATTGTGTTGGTCCCCAGGGCACGATTGTCATGCCCACGCACTCGGGTGATCTGA
         1440         1450         1460         1470         1480         1490

CTGATCCCGCAGATTGGAGATCGCCGCCCGTGCCTGCCGATTGGGTGCAGATCT
      1500         1510         1520         1530         1540
```

Figure 4. Primary sequence of the Hygromycin B Resistance Gene.
The sequence is reported from the rightmost SacI site in figure 4 to the BglII site in a 5' to 3' direction. The predicted amino acid sequence encoded by the large open reading frame found within this region is also included.

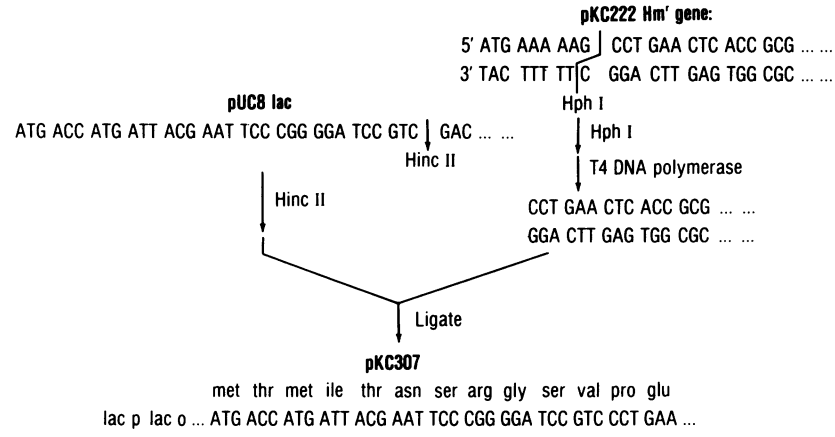

Figure 5. Fusion of the Hygromycin B Resistance Gene with lacZ Coding Sequence

The hygromycin B resistance gene was cut at a HphI restriction site, which cuts the upper strand after base 236 (figure 5). After removing the 3' extension with T4 DNA polymerase, the blunt ended molecules were ligated to the lac Z gene on pUC8 (18) cut at a HincII site. The predicted sequence of the fused construct, shown as pKC307, was verified by DNA sequence analysis.

analysis (figure 2) of plasmid-encoded proteins. Third, a large RNA species exists in E. coli cells harboring the relevant plasmid which hybridizes to probes from both genes. The transcriptional activator sequence responsible for the expression of the polycistron has been identified and characterized. Primary sequence analysis of the SacI-PstI fragment (fragment 6 of figure 1a) demonstrates a typical E. coli promoter sequence (table 1). When the SacI-PstI fragment is fused in frame with a lacZ gene lacking its promoter and first 8 amino acids, active β-galactosidase protein is produced (table 2). Also, if this SacI-PstI sequence is excised and replaced by a known promoter sequence, the leftward promoter from bacteriophage lambda, the resistance gene is once again expressed. Thus, physical and functional evidence support the conclusion that the SacI-PstI fragment contains the promoter for the aac(3)IV and aph(4) genes.

The separate polypeptides encoded by the resistance genes both appear as doublets on an SDS-polyacrylamide gel (figure 2). The polypeptide products of the aac(3)IV gene correspond to apparent molecular weights of 31,500 and 33,000 daltons, and the polypeptides made by the aph(4) gene correspond to apparent molecular weights of 41,500 and 42,500. This molecular weight estimate for the aph (IV) gene product agrees with the estimation (42,000

Table 3. The Influence of IPTG on Expression of Hm Resistance from pKC307.

		Titer on TY	Titer on Ty plus ampicillin plus hygromycin B*	Efficiency of Plating on Drug Plates
Cells Grown Without IPTG	no IPTG in plates	5.6×10^9	1.9×10^4	3.4×10^{-6}
	1mM IPTG	5.6×10^9	3.2×10^5	5.7×10^{-5}
Cells Grown with 1mM IPTG	no IPTG in plates	6.4×10^9	3.7×10^4	5.8×10^6
	1mM IPTG in plates	6.4×10^9	4.5×10^6	7.0×10^{-4}

Bacteria were grown overnight in TY broth (8), with or without 1 mM IPTG, then diluted and plated on plates with or without 1 mM IPTG. Plates contained TY plus 1.5 percent agar. When added IPTG was present at 1 mM, ampicillin was present at 100 µg/ml, and hygromycin B was present at 200 µg/ml.

*Control cells without plasmid acquire spontaneous hygromycin B resistance under these conditions at a frequency of less than 1 in 10^9 cells.

daltons) based on molecular sieving (Santerre, R.F., N.E. Allen, J.N. Hobbs, Jr., R.N. Rao and R.J. Schmidt. Manuscript in preparation), which implies that the enzyme exists in a monomeric state. The presence of doublets suggests that the polypeptide may be excreted into the periplasmic space, with the larger polypeptides corresponding to the molecules still inside the cell and retaining a signal polypeptide, and the smaller polypeptides corresponding to processed molecules lacking the signal polypeptide. This interpretation is consistent with the finding that the enzymatic activities associated with Hm^r and Am^r can be osmotically shocked from the cells (Rao, R.N., N.E. Allen, J.N. Hobbs, Jr., W.E. Alborn, Jr., H.A. Kirst, and J.W. Paschal. Antimicrobial Agents and Chemotherapy, in press).

DNA Sequence of the Hygromycin Resistance Gene.

Since it was known that the hygromycin resistance gene was contained within the second SacI and BglII sites of pKC222 (figure 1), the primary sequence of this region was determined. The sequencing strategy is shown in figure 3, and the primary sequence itself in figure 4. One large open reading frame was found, and the amino acid sequence encoded within this

open reading frame is also shown in figure 4. No other open reading frame is within 40 percent of the size of this large open reading frame, which is large enough to encode a polypeptide with a molecular weight of 38,236 daltons. This predicted weight is about 10 percent lower than the value of about 42,000 daltons determined by gel electrophoresis for the hygromycin phosphotransferase polypeptide (figure 2). The reason for the discrepancy is not known, although it could be due to anomolous migration on SDS-polyacrylamide gels or posttranslational modification.

There are several noteworthy features of the sequence reported in figure 4. First, the predicted site of initiation of translation is preceded by a GGGGG sequence ending 9 bp before the ATG. This sequence and the spacing are typical of ribosome binding sites in E. coli (17). Another feature of the sequence is that there is a large open reading frame in the opposite strand, bounded by stop codons at bases 1348 and 143 in figure 5. However, there are no ATG codons preceded by ribosome binding sites in this reading frame. The significance of this open reading frame is not known.

The predicted amino acid sequence of the Hygromycin B phosphotransferase polypeptide encoded on the aph(4) gene was compared with that of the neomycin phosphotransferase polypeptides encoded on Tn5 (19) and Tn903 (20). No obvious similarities were found, even in the regions where the resistance genes from Tn5 and Tn903 show extensive homology. This result is not surprising since neomycin and related compounds are not substrates for the hygromycin B phosphotransferase polypeptide encoded on the aph(4) gene (Rao, R.N., N.E. Allen, J.N. Hobbs, Jr., W.E. Alborn, Jr., H.A. Kirst and J.W. Paschal. Antimicrobial Agents and Chemotherapy, in press).

Fusing the hygromycin B resistance gene to lac Z

We constructed protein fusions in order to test the functionality of the open reading frame predicted to be responsible for hygromycin B resistance. The putative hygromycin B resistance gene was cut at the unique HphI site which excises the DNA encoding the first three amino acids. After removing the 3' overhang, the DNA was joined to the amino terminus of a modified β-galactosidase gene carried on pUC8 (18). The resulting hybrid plasmid, outlined in figure 5, substitutes 11 amino acids for the three removed from the hygromycin resistance gene.

This hybrid gene, present in pKC307, is active in E. coli and the activity is stimulated by addition of IPTG, a gratuitous inducer of the β-galactosidase promoter. As shown in table III, cells containing pKC307 are partially resistant to 200 μg/ml hygromycin B. That is, when cells are

plated on plates containing hygromycin B, about 1 cell in 10^5 grew into a visible colony. Since cells containing intact hygromycin B resistance gene on pIT105 or pKC222 give a 100 percent efficiency of plating on hygromycin B plates, the hybrid construct seems less efficient than the parental. The reduced efficiency could be due to lower specific activity and/or reduced stability of the hybrid polypeptide, or other problems with transcription or translation.

The reduced plating efficiency on hygromycin B plates can be partially overcome by addition of IPTG, a gratuitous inducer of transcription of the β-galactosidase gene. As shown in table III, the efficiency of plating on hygromycin B-containing plates is increased 100 fold by inclusion of IPTG.

These results support our prediction that the open reading frame depicted in figure 4 encodes the hygromycin B phosphotransferase polypeptide. It could also be argued that the β-galactosidase gene is simply providing a transcriptional activator, and that the correct open reading frame lies elsewhere downstream. This is not likely because construction of hybrid hygromycin B resistance genes that delete the same three amino acid codons but shift the reading frame do not confer resistance whereas similar constructs that maintain the reading frame do allow resistance (data not shown).

It was noted earlier that the hygromycin B phosphotransferase polypeptide might be a secreted protein, since the activity can be osmotically shocked from the periplasma and a doublet is observed on SDS-polyacrylamide gels (see figure 2). If this is the case one would expect to find a signal sequence at the beginning of the predicted amino acid sequence. The anticipated characteristics of a signal sequence are 1) a positively charged region at the start of the signal sequence and 2) a region of hydrophobic amino acid (21). The amino acid sequence shown in figure 5 meets the first requirement nicely, since the initiating methionine is followed by two lysines. However, the following amino acid sequence is not as hydrophobic as that found in other signal sequences in E. coli (21). It should be noted, however, that cells containing the hybrid hygromycin B resistance gene which have had the two lysine residues replaced by eleven other amino acids do not seem to have hygromycin B phosphotransferase activity in their periplasm. No phosphotransferase activity can be detected in extracts made by osmotically shocking the cells, but phosphotransferase activity can be detected in cell lysates (data not shown). Thus, the ability of the cells to secrete the polypeptide into the periplasm may be

abolished in the hybrid genes. The relationship, if any, between secretion of the polypeptide and conferral of hygromycin B resistance is not known.

The ability to form fusion polypeptides with the hygromycin B phosphotransferase protein has important implications for its possible use as a dominant selectable marker in heterologous systems. Ligation of appropriate promoters and translational initiation sequences from other organisms in front of the hygromycin resistance gene might allow the gene to be used as a dominant selectable marker in that organism, or even to isolate these activator sequences in the first place.

ACKNOWLEDGEMENTS

We thank Stu Kuhstoss for his gift of the λP_L sequences, Dr. R. H. Baltz for his thoughtful reading of the manuscript, and Cheryl Alexander for typing the manuscript. We also thank Linda Gritz and Julian Davies for sending sequence data prior to publication, which helped us uncover a sequencing error.

*To whom correspondence should be addressed

REFERENCES

1. Pittenger, R.C., Wolfe, R.N., Hoehn, M.M., Marks, P.N., Daily, W.A. and McGuire, J.M. (1953). Hygromycin I. Antibiotics and Chemotherapy. 3, 1268-1282.
2. Gonzalez, A., Jimenez, A., Vazquez, D., Davies, J.E. and Schindler, D. (1978) Biochem. Biophys. Acta 521, 459-469.
3. West, R.W., Neve, R.L. and Rodriquez, R.L. (1979) Gene 7, 271-288.
4. Sancar, A., Hack, A.M. and Rupp, W.D. (1979) J. Bacteriol. 137, 692-693.
5. Sancar, A. and Rupert, C. S. (1978) Nature (London) 272, 471-472.
6. Wozney, J., Hanahan, D., Morimoto, R., Boedtker, H. and Doty, P. (1981) Natl. Acad. Sci. U.S.A. 89, 712-716.
7. Sutcliffe, J.G. (1979) Cold Spring Harbor Symposia on Quantitative Biology. Vol. XLIII. DNA: Replication and recombination. p.77-90.
8. Miller, J.H. (1972) Experiments in Molecular Genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
9. Ames, G.F. -L. (1974) J. Biol. Chem. 249, 634-644.
10. Zaret, K.S. and Sherman, F. (1982) Cell 28, 563-573.15.
11. Thomas, P.S. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 5201-5205.
12. Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 560-564.
13. Ingolia, T.D. and Craig, E.A. (1981) Nucleic Acids Res. 9, 1627-1642.
14. Davies, J.E. and O'Connor, S. (1978) Antimicrob. Agents Chemother. 14, 69-72.
15. Casadaban, M.J., Chou, J. and Cohen, S.N. (1980) J. Bacteriol. 143, 971-980.

16. Franklin, N.C. and Bennett, G.N. (1979) Gene 8, 107-119.
17. Scherer, G.F.E., Walkinshaw, M.D., Arnott, S., and Morre, D.J. (1980) Nucleic Acids Res. 8, 3895-3907.
18. Vieira, J. and Messing, J. (1982) Gene 19, 259-265.
19 Beck, E., Ludwig, G., Auerswald, E.A., Reiss, B. and Scholler, H. (1982) Gene 19, 327-336.
20. Oka, A., Sugisaki, H., and Tanakami, M. (1981) J. Mol. Biol. 147, 217-226.
21. Michaelis, S. and Beckwith, J. (1982) Ann. Rev. Microbiol. 36.
22. Rosenberg, M., and Court, D. (1979) Ann. Rev. Genet. 13, 319-353.