DNA
Volume 8, Number 4, 1989
Mary Ann Liebert, Inc., Publishers
Pp. 000–000

LABORATORY METHODS

Using Transposon Tn*5* Insertions to Sequence Bacteriophage T4 Gene *11*

BENJAMIN K. BARRETT and PETER B. BERGET

ABSTRACT

A simple and rapid method of creating an overlapping set of deletions in cloned DNA in preparation for sequencing has been developed. The method is based on a positive selection for Tn*5* transposition into the cloned DNA fragment on a high-copy-number filamid, resolution of potential filamid dimers by filamentous phage infection, and the use of Tn*5* both as a "portable" restriction enzyme site for *in vitro* DNA deletion and a "portable" sequencing primer binding site to initiate DNA sequencing reactions using a custom primer complementary to the outside ends of IS*50*. This new method has been utilized to sequence bacteriophage T4 gene *11*, encoding the T4 baseplate protein gp11. The coding sequence of gene *11* is 657 bp in length, and predicts a primary structure of 219 amino acids that agrees well with the biochemical data previously obtained. DNA sequence around gene *11* suggests that the expression of genes *10*, *11*, and *12* of phage T4 are translationally coupled. Plasmids carrying deletions generated using this method have been used to map genetically five amber alleles of gene *11*. These amber alleles were sequenced to confirm the proposed reading frame. The five amber alleles actually represent two different mutational changes at either codon 206 or 207, changing these adjacent glutamine codons to amber. The position of these amber alleles lends support to earlier studies identifying the carboxyl terminus of gp11 as essential in the interaction of P11 with baseplate protein P10 (Plishker and Berget, 1984).

INTRODUCTION

Several sophisticated methods to generate deletions in cloned DNA have been developed to facilitate DNA sequencing. All of these methods rely on bringing a "universal primer" binding site adjacent to unknown DNA sequences by deleting DNA between the universal primer binding site located in the vector sequences and some "random" position in the unknown cloned DNA sequence. A topologically equivalent manipulation would be to introduce universal primer binding sites at random throughout an unknown cloned DNA sequence. This manipulation could be done either *in vitro* or *in vivo*. It occurred to us that a transposable genetic element could be useful to effect this manipulation *in vivo*. We have explored this alternative method using the transposable element Tn*5*.

Tn*5* is a composite transposon in which nearly identical inverted repeats of the 1500-bp transposable element IS*50* (Rothstein and Reznikoff, 1981) bracket a unique sequence of about 3,000 bp encoding, among other things, a neomycin phosphotransferase gene which confers resistance to kanamycin upon cells carrying this transposon (Berg *et al.*, 1982). Sasakawa *et al.* (1982) have shown that cells harboring high-copy-number plasmids carrying Tn*5* are resistant to higher levels of kanamycin than isogenic cells containing a single Tn*5* insertion in the bacterial chromosome. Moreover, they demonstrated that this hyperresistant phenotype provided a direct selection for cells in which Tn*5* had transposed from the chromosome (or a single copy episome) to a high-copy-number plasmid. This selection scheme was used in this study to isolate essentially random insertions into a cloned fragment of bacteriophage T4

Department of Biological Sciences, Carnegie Mellon University, Pittsburgh, PA 15213-3890.

DNA carrying gene *11*. The positions of these insertions were determined by restriction endonuclease analyses. Deletions in these plasmids were done using both restriction enzyme sites in the vector sequences and those brought into the plasmid by Tn*5*. These deletion plasmids were used for DNA sequencing and genetic studies.

MATERIALS AND METHODS

Bacterial strains, plasmids, and phages

The *E. coli* strains used in this study were KK2186 [Δ(*lac-pro*)*thi strA supE endA sbcB* r^-m^+/F′ *traD36 proAB lacI*^{q} Δ(*lacZ*)M15]; PB759 [*zzz*::Tn*5 araD*139 Δ(*ara-leu*)7697 Δ(*lac*)X74 *galU galK* r^-m^+ *rpsL srl* Δ(*recA*)/F′ *lac pro zzf*::Tn*10*dCam; Bb (nonpermissive host for amber mutants); and CR63 (amber-suppressing host). The following bacteriophage T4D strains were used in this study: *11⁻am*N93, *11⁻am*N108, *11⁻am*N128, *11⁻am*NG115, *11⁻am*NG369, and *10⁻am*NG153. Unless otherwise specified, bacteria were grown in LB Broth (Miller, 1972) with appropriate antibiotics (100 μg/ml sodium ampicillin, 25 μg/ml kanamycin sulfate). Bacteriophage T4 procedures were as described previously (Berget and King, 1978). The Bluescribe M13⁺ plasmid was purchased from Stratagene, Inc. The interference-resistant M13 phage, IR1, was obtained from G. Weinstock.

DNA manipulations

General methods for DNA manipulations were carried out as described in Maniatis *et al.* (1982). Enzymatic reactions of DNA were performed as recommended by the suppliers of the enzymes. DNA fragments used in cloning experiments were separated by electrophoresis on 1.0% low-melting-temperature agarose, the bands excised, and DNA phenol-extracted from the agarose slices. The Sequenase DNA sequencing kit was purchased from U.S. Biochemical Corp. Restriction endonucleases and T4 ligase were purchased from Boehringer-Mannheim Biochemicals. one-kilobase DNA size standards were purchased from Bethesda Research Laboratories. Oligonucleotide syntheses were performed at the Pittsburgh DNA Synthesis Facility (University of Pittsburgh). Sequences of the gene *11* oligonucleotides are as follows: oligo A,5′-AATAGCTGATTCTATG; oligo B, 5′-GTCTTTATATGAATTG; oligo C,5′-TAACGAGATTCATTAG.

Isolation of T4 *11⁻* amber phage genomic DNA for sequencing was accomplished as follows. A 100-ml culture of CR63 at midlog phase was infected with the amber phage of interest (moi 5) and grown overnight with shaking at 30°C. The next day, the cells were lysed with chloroform and pelleted by centrifugation. The supernatant was centrifuged at 15,000 rpm for 90 min to pellet the phage. Phage were gently resuspended overnight in TM (10 m*M* Tris-HCl, 5 m*M* $MgCl_2$). The phage suspension was layered onto a CsCl block gradient; after centrifugation for 2 hr at 35,000 rpm, the phage band was removed and dialyzed overnight against cold TM. This procedure typically yielded 2 ml of dialysate with a titer of 5×10^{12} phage/ml. A 0.5-ml amount of the dialysate was incubated at 65°C for 15 min in 0.1% $NaDodSO_4$ and 5 m*M* EDTA pH 8. The DNA was extracted with 0.5 ml of phenol, phenol/chloroform (50:50), and chloroform, consecutively. An equal volume of isopropanol was added and the DNA was precipitated at −20°C for 15 min. The DNA was pelleted by centrifugation and washed with 70% ethanol. The pellet was dried under vacuum and resuspended in 100 μl of TE. The concentration of the DNA was estimated by measuring A_{260}. Typically, 2.5×10^{12} phage yield about 350 μg of DNA.

Thirty micrograms of T4 genomic DNA was prepared for double-stranded sequencing. The DNA was digested to completion with *Taq* I which cleaves T4 hydroxymethylcytosine-containing DNA but does not cleave in gene *11*. The digested DNA was subjected to phenol/ether extraction and resuspended in H_2O. The DNA was denatured and annealed with gene *11* primers using the alkaline-denaturing protocol of Kraft *et al.* (1988). DNA sequencing was performed by the dideoxy chain-terminating method (Sanger *et al.*, 1977) using the modified T7 polymerase (Tabor and Richardson, 1987) provided in the Sequenase kit. Single-stranded DNA from filamid lysates was prepared for sequencing according to the procedures outlined in the M13 Cloning/Sequencing Manual provided by BRL.

Bacterial growth and genetic manipulations

For isolation of Tn*5* insertions into the cloned T4 DNA, a modification of the selection developed by Sasakawa *et al.* (1982) was used. The isolation procedure described below should eliminate the possibility that siblings of an original insertion event are present in the final plasmid pools and that all potential plasmid dimers are resolved into monomers. All plasmids described in this report carry an M13 phage origin of replication and single-stranded DNA packaging, as well as a plasmid origin of replication and a selectable drug-resistance marker for resistance to ampicillin. Such plasmids are termed "filamids." The presence of the M13 origin of replication allows both the production of monomeric plasmid transducing particles in M13 helper infections as well as the isolation of single-stranded DNA from these mixed lysates. A filamid lysate was generated by infecting 1×10^9 cells harboring the ampicillin-resistant plasmids of interest with M13 interference resistant (IR1) phage at a moi of 20. This lysate was diluted serially and the dilutions were used to transduce the plasmid into *E. coli* PB759, which contains the Tn*5* transposon in its chromosome. Transductants were selected by plating on ampicillin-containing plates. Single colonies of transductants were picked and grown overnight to saturation in BL broth containing ampicillin and kanamycin. A 0.1-ml amount from each of the individual overnight cul-

tures was spread onto plates containing 250 μg/ml of kanamycin and 100 μg/ml of ampicillin to select kanamycin hyperresistant bacteria. As a negative control, 0.1 ml of an overnight culture of PB759 was plated similarly. The plates were incubated at 37°C for 2 days and usually contained 50–1,000 colonies per plate after this period of incubation. The selection for bacteria which were hyperresistant to kanamycin was repeated for each independent pool of bacteria in liquid culture. The colonies from each plate were scrape-harvested in 5 ml of LB broth. The suspensions were centrifugated for 15 min at 4,000 rpm, and the cell pellets were resuspended in 5 ml of LB broth. Five milliliters of LB broth containing 100 μg/ml of ampicillin and 250 μg/ml of kanamycin was inoculated with a sufficient volume (typically 0.1 ml) of washed cells to yield a final cell density of about 10^7 cells/ml. The cultures were grown overnight at 37°C. Then, 0.5-ml aliquots of each culture were infected with 5×10^9 particles of M13 IR1 (moi 5) for 15 min at room temperature. LB broth was added to a final volume of 5 ml and the culture was grown overnight at 37°C. The cultures were centrifuged for 10 min at 10,000 rpm, after which the supernatants were heat-sterilized at 65°C for 20 min. Each of the independently derived filamid lysates was diluted serially, and 100 μl of each dilution was mixed with 2×10^8 cells of KK2186. These transduction mixes were spread onto ampicillin-containing plates, which were incubated overnight at 37°C. This transduction step in the protocol effectively eliminates any dimeric plasmids that may have been generated by recombination in the host strain. The ampicillin-resistant colonies were replica-plated onto plates containing kanamycin and ampicillin. Only those plasmids that had received a Tn*5* insertion should be able to transduce KK2186 to kanamycin resistance. The resultant colonies were streak-purified once onto plates containing ampicillin and kanamycin. These colonies were grown overnight in BL broth containing ampicillin at 37°C and used for small-scale plasmid preparations (Birnboim and Doly, 1979) followed by restriction analysis.

Spot tests for marker rescue of T4 *11*⁻ amber mutants were conducted in the following manner. Overnight cultures of the strains harboring the deletion plasmids were diluted 10-fold in LB broth, and a T4 phage containing an amber allele of gene *11* (*am*N93) was added to the culture at a moi of 1. The phage were allowed to adsorb to the cells for 20 min at room temperature. These infected cultures were spotted separately in 10-μl aliquots on two soft agar lawns, one containing strain CR63 (amber suppressing) and one containing strain Bb (nonpermissive). Positive marker rescue was indicated by a clearing of the spot on Bb. As negative controls, phage were spotted directly onto the lawns and also allowed to infect KK2186 harboring no plasmids. As a positive control, phage were allowed to infect KK2186 carrying plasmid pBB1 which carries the entire gene *11*.

To determine recombination frequencies between the *11*⁻ amber alleles, cultures of CR63 at midlog phase were infected with each amber phage alone (moi 10) and with each in pairwise combination with the other amber alleles (moi 5 each). After incubation at 30°C for 90 min with shaking, the cells were lysed with chloroform. The lysates were titered for total phage on CR63 and wild-type phage on Bb.

RESULTS

Cloning of T4 phage DNA

Construction of plasmids containing T4 DNA is depicted in Fig. 1. Plasmid pDO301 (Oliver *et al.*, 1981),containing a portion of bacteriophage gene *12* and the complete sequence of genes *10* and *11*, was digested with *Pst* I and *Eco* RI. A 2.3-kb fragment, containing gene *11* and partial sequences of genes *10* and *12*, was cloned into pBR322 and yielded plasmid pCS13. Subsequently, a 2.33-kb *Pst* I-*Hind* III fragment of pCS13 was subcloned into Bluescribe M13⁺ using the same restriction sites in the vector yielding plasmid pJZ1. This plasmid contains the 3′ end of gene *10*, the 5′ end of gene *12*, and the complete coding sequence of gene *11*. A 2.3-kb *Pst* I-*Eco* RI fragment of plasmid pCS13 was also subcloned into Bluescribe M13⁺ using the same restriction sites in the vector yielding plasmid pBB1. This plasmid contains T4 DNA identical to pJZ1, although in opposite orientation.

Isolation and restriction mapping of Tn5 insertions into plasmids

Independent transductants of pJZ1 and pBB1 in *E. coli* PB759 were subjected to the hyperresistance selection scheme described in Materials and Methods. Our selection technique guaranteed the isolation of independent, essentially random insertion events, since each original ampicillin-resistant transductant was propagated separately. The strains harboring plasmids with Tn*5* insertions were cultured and were subjected subsequently to small-scale plasmid preparation, followed by restriction analysis.

A Tn*5* insertion into a plasmid increases the size of the plasmid by 5.8 kb. Moreover, the additional restriction sites introduced by the transposon allow the location of the insertion within the plasmid to be ascertained by restriction enzyme mapping (see Fig. 2). The presence of the Tn*5* insertions was determined initially by restriction digests utilizing *Hind* III and *Eco* RI. In pJZ1, this double digest yields two major fragments, one of 3.2 kb corresponding to the vector backbone, and one of about 2.4 kb corresponding to the original T4 DNA insert (another small fragment ≤30 bp is also generated and ignored in these analyses). In those plasmids which had Tn*5* insertions within the cloned T4 DNA, the 2.4-kb fragment is disrupted by the presence of two additional *Hind* III sites within Tn*5*. These insertions cause the disappearance of the 2.4 kb fragment with the simultaneous production of three additional fragments, one of 3.4 kb derived wholly from within Tn*5* and two fragments whose summed

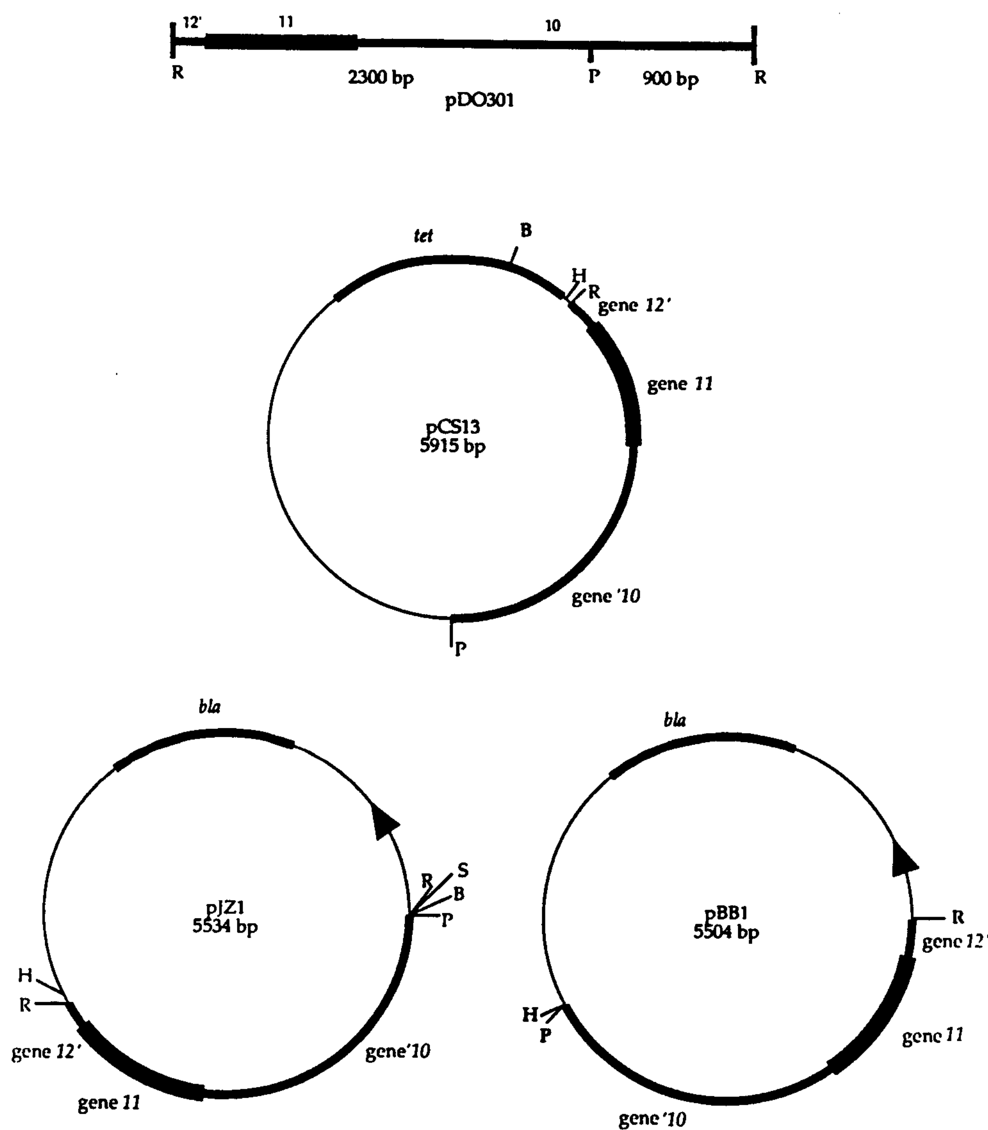

FIG. 1. Diagrams of plasmids used in this study. The top line represents the fragment of T4 DNA cloned in pDO301 (Oliver *et al.*, 1981). The direction and location of the M13 origin of replication are denoted by the large arrowhead. Restriction endonucleases are represented as follows: P, *Pst* I; R, *Eco* RI; H, *Hind* III; B, *Bam* HI; S, *Sal* I.

lengths should equal that of the original 2.4-kb fragment plus 2.4 kb as a result of the two 1.2-kb *Hind* III fragments from the IS*50* regions of the transposon (Fig. 2).

Those candidates that met this initial mapping criteria were subjected to *Pst* I digestion to locate more accurately the transposon within the cloned region. *Pst* I cleaves the transposon at four sites, and the parental plasmids at one site. Digestion with *Pst* I yields five fragments in those plasmids containing a transposon. The mobilities of these fragments were compared to those of size standards, and the fragment lengths were used to estimate the distance to the transposon insertions from the external *Pst* I site of the plasmid. With two simple restriction endonuclease digestions, both the presence of the insertion in the cloned fragment and its relative location were identified quickly.

Creating deletions

The primer utilized in our sequencing protocol is complementary to the DNA sequence located between the 10th and 30th nucleotides from the outside ends of the inverted terminal repeats of Tn*5* and has the sequence 5'-GTTCAGGACGCTACTTGTGT. It has been shown previously that attempting to use such a primer for sequencing from Tn*5* insertions in single-stranded DNA before deleting one of the IS*50* regions was unsuccessful (G. Weinstock, personal communication), on the assumption that the IS*50* regions of Tn*5* are inverted terminal repeats. Once single-stranded sequencing templates are made, the complementary sequences of the two IS*50* elements could hybridize with one another, displacing the sequencing primer. Thus,

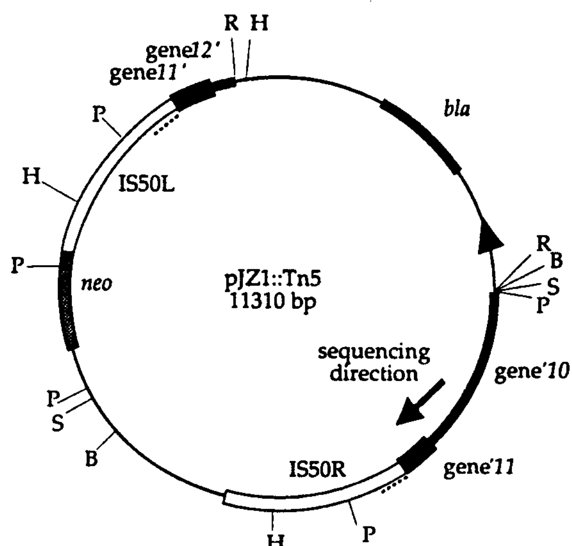

FIG. 2. Schematic representation of a Tn*5* insertion into plasmid pJZ1. Sequencing direction is indicated by the arrow. Tn*5* primer binding sites in IS*50* (white boxes) are indicated by dotted lines. The neomycin resistance gene of Tn*5* is indicated by the hatched box. Restriction endonuclease abbreviations are defined in Fig. 1.

it became necessary to delete the IS*50* that, upon production of the single-stranded template, contained the sequence identical to our primer. Because the sequencing direction off of the single-stranded DNA obtained by infecting cells carrying the Bluescribe plasmid is "clockwise," as depicted in Fig. 2, the IS*50* region deleted was always the one proximal to the external Bluescribe *Eco* RI site (Fig. 2). The deletions were accomplished by digesting the pJZ1 candidates with *Pst* I, followed by dilution and ligation. The locations of these deletions are shown in Fig. 3.

Localizing gene 11 *in the cloned DNA*

To delineate the general location of gene *11* on the cloned fragment, the deletion plasmids were used in marker rescue experiments (see Materials and Methods). As expected, spots of the *11*⁻ amber-infected cells on the CR63 lawn were cleared by lysis in all cases. However, only phage that were allowed to undergo recombination by infecting hosts harboring a plasmid-borne wild-type allele of gene *11* showed clearing on Bb (data not shown). Thus, it was determined that the wild-type allele of *11*⁻ am N93 was somewhere between the deletion end points of plasmids pBB6 and pBB7. As the length of gene *11* had been previously been estimated to be 650 bp, it was apparent that the gene was entirely between deletion plasmids pBB2 and pBB7.

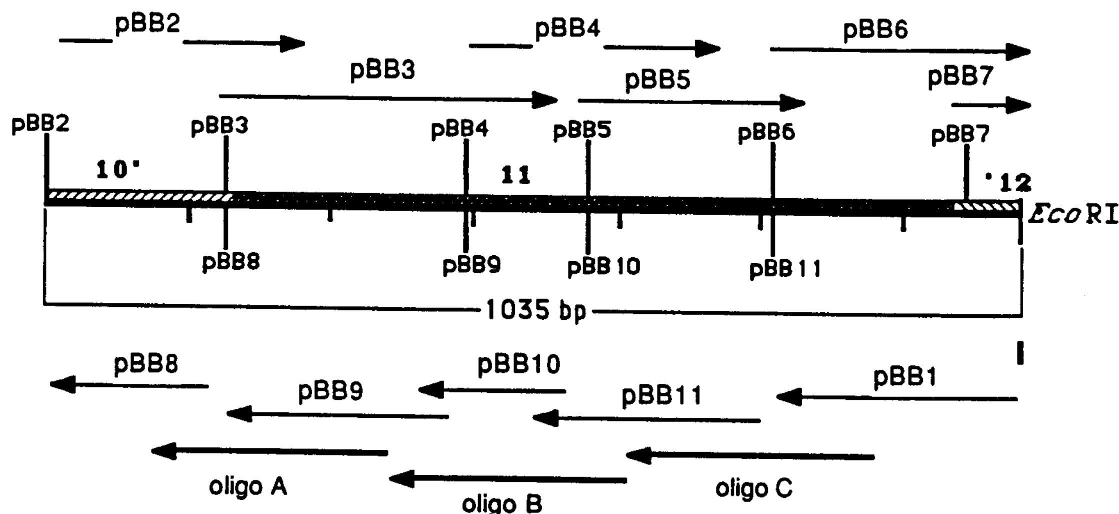

FIG. 3. Placement of Tn*5* insertions/deletions on the cloned gene *11* fragment. The locations of genes *10*, *11*, and *12* are denoted by hatched boxes. Only those deletions used for sequencing are shown. Vertical lines through the DNA fragment indicate points of Tn*5* insertions. Deletion plasmids are missing DNA from the point of Tn*5* insertion rightwards to the *Eco* RI site for those indicated above the DNA segment. Those plasmids indicated below the DNA segment are missing all T4 DNA leftwards. Arrows above the fragment represent sequence information generated from insertions into plasmid pJZ1. Light arrows below the fragment represent sequence data obtained from insertions inverted to obtain sense strand sequence. Heavy arrows below the fragment represent sequence data obtained from synthetic oligonucleotide primers. Short vertical lines below the fragment mark 150-bp sections. Sequence obtained from plasmid pBB1 using the M13 "universal" primer is indicated.

Localizing amber alleles

Our deletion collection was also used in an effort to determine the genetic location of each *11⁻* amber mutant allele in our T4 phage stock collection. Using the same spot test described above, it was determined that the five amber mutants tested showed exactly the same pattern of marker rescue from the set of deletion plasmids. This indicated that the mutations were clustered in the same region of the gene, between deletion plasmid pBB6 and the *Eco* RI site. Recombination studies were conducted to determine how far apart the alleles were. The percent recombination between each allele was calculated. The data indicated that the amber alleles could be divided into two groups, one containing *11⁻* amber alleles N108 and N93 and the other containing alleles N128, NG115, and NG369. For example, the recombination between amber alleles NG369 and N128 was $< 10^{-3}$%, whereas that between amber alleles (N108 and N128) and (N108 and NG369) was 0.0016% and 0.01%, respectively. Thus, the percent recombination between alleles in the two groups ranged from 1.6- to 10-fold greater than that between alleles in the same group, which was typically equivalent to background (reversion) levels. As a control, percent recombination was determined for a representative of each group and an outside marker (*10⁻ am*NG153). In both cases, the percent recombination was 800-fold greater than that between alleles in the same class (4.5–4.8%). Thus, the two classes of *11⁻* amber alleles were distinctly separate, though very close to one another.

Completing the deletion collection

The scheme for selecting cells which were hyperresistant to kanamycin had yielded a set of Tn*5* insertions from which deletions were made that allowed complete sequencing of the antisense (noncoding) strand of gene *11* (see Fig. 3). However, a large area on the sense (coding) strand was not accessible to sequencing using these deletions. Thus, it became necessary to manipulate the existing insertions in such a manner as to obtain the complementary sequence. As stated previously, our sequencing primer binds to both inverted terminal repeats of Tn*5*. By simply subcloning in the proper orientation that region of our insertion plasmids that was adjacent to and included the "other" primer binding site, we would be able to sequence the coding strand. Of course, this new sequence information would be in the opposite direction to the sequence which could be obtained from a particular starting deletion. This subcloning was accomplished by digesting selected members of the original Tn*5* insertion group with *Bam* HI and *Sal* I and cloning the appropriate fragment into a similarly digested Bluescribe plasmid.

Sequencing

We determined the DNA sequence of gene *11* using the dideoxy chain-termination method (Sanger *et al.*, 1977) as applied to bacteriophage M13. The DNA sequence of the cloned insert from deletion plasmid pBB2 to the *Eco* RI site is presented in Fig. 4. The entire sequence was determined on both strands; in addition, the complete sequence of the coding strand was confirmed using three synthetic oligonucleotide primers. A potential ribosome binding site (Shine and Dalgarno, 1974) is underlined. Starting at nucleotide position 207 and extending through nucleotide position 866, there is a single open reading frame, commencing with ATG and terminating with TGA, which would encode a polypeptide of 219 amino acids with a predicted molecular weight of 23,680 daltons. That this is the open reading frame coding for gene *11* product is supported by several facts. First, the molecular weight of the predicted polypeptide agrees well with that of the gp11 monomer, 26,000 daltons (as determined by $NaDodSO_4$ polyacrylamide gel electrophoresis (King and Laemmli, 1973). Second, the amino-terminal portion of the deduced protein sequence matches that derived by Edman degradation of the purified protein (NH_2-X-X-Leu-Asn-Asn-Lys-Ala-Gly-Val-Ile..., where the first residue is probably Met and the second residue may be Ser, Thr, or Cys; F. Wold, personal communication). Third, sequencing of amber alleles confirmed this reading frame as that of gene *11*.

Several other aspects of the sequence should be noted. A portion of the 3′ end of gene *10* is included in Fig. 4. The 207-bp region shown comprises about 8.5% of the calculated length of gene *10*. The 5′ end of gene *12* is also indicated. From the sequence around the boundaries of genes *10*, *11*, and *12* where overlapping start and stop codons are found, it is apparent that the expression these genes could be coupled translationally. This is consistent with the known transcription and polarity in this region of the T4 genome (Stahl *et al.*, 1970). The sequence of gene *12* has already been published (Selivanov *et al.*, 1988). The exact sites of the Tn*5* transposon insertions were determined by repeating the sequencing of the deletion plasmids using a second primer that is complementary to the −25 to −45 region of the IS*50* sequence of Tn*5*. The nucleotide position of each insert is indicated in Fig. 4.

Having defined the sequence of gene *11*, we wanted to determine the specific nucleotide changes present in our collection of five amber mutants. DNA sequencing from double-stranded DNA templates made from T4 amber mutant phage was performed on all five amber phage, using genomic DNA and three synthetic oligonucleotide primers (see Materials and Methods). The location and specific nucleotide changes of the amber mutants are shown in Fig. 4.

DISCUSSION

We have investigated the use of Tn*5* insertions to provide "randomly positioned" primer binding sites for DNA sequencing. In our hands we have found this technique to be particularly successful and easy. Though Berg *et al.* (1983) and Lodge *et al.* (1988) have reported that Tn*5* transposes into plasmid DNA in a nonrandom manner, for our purposes the distribution of insertion sites was sufficiently random to create a collection of overlapping dele-

```
Thr Glu Thr Asp Glu Glu Val Leu Ile Val Asp Glu Asn Gly Ser Val Ile Val Gly Gly
ACT GAA ACA GAT GAA GAA GTT CTA ATA GTT GAT GAA AAT GGA TCA GTC ATT GTT GGT GGG      60

Cys Gln Tyr Asp Pro Asp Glu Ser Gly Pro Ile Tyr Thr Lys Tyr Arg Glu Ala Lys Ala
TGT CAA TAC GAT CCA GAT GAA TCC GGT CCA ATT TAC ACT AAA TAC CGT GAA GCT AAA GCA     120

Ser Thr Asn Ser Thr His Thr Pro Pro Thr Ser Ile Thr Asn Ile Gln Pro Tyr Ile Thr
TCT ACT AAC TCT ACT CAC ACT CCG CCA ACA TCA ATA ACT AAC ATT CAA CCA TAT ATT ACA     180

Val Tyr Arg Trp Ile Arg Ile Ala  *
                                  Met Ser Leu Leu Asn Asn Lys Ala Gly Val Ile        11
GTT TAT CGT TGG ATA AGG ATT GCA TA  ATG AGT TTA CTT AAT AAT AAA GCG GGA GTT ATT     239
               ^

Ser Arg Leu Ala Asp Phe Leu Gly Phe Arg Pro Lys Thr Gly Asp Ile Asp Val Met Asn      31
TCC CGC TTA GCC GAT TTT CTT GGT TTT AGA CCT AAA ACT GGC GAC ATT GAT GTA ATG AAT     299

Arg Gln Ser Val Gly Ser Val Thr Ile Ser Gln Leu Ala Lys Gly Phe Tyr Glu Pro Asn      51
CGT CAA TCA GTC GGG TCA GTG ACA ATA TCT CAA TTA GCG AAA GGA TTT TAT GAA CCA AAC     359

Ile Glu Ser Ala Ile Asn Asp Val His Asn Phe Ser Ile Lys Asp Val Gly Thr Ile Ile      71
ATA GAA TCA GCT ATT AAT GAC GTT CAT AAT TTT TCT ATA AAA GAC GTT GGC ACA ATT ATT     419

Thr Asn Lys Thr Gly Val Ser Pro Glu Gly Val Ser Gln Thr Asp Tyr Trp Ala Phe Ser      91
ACT AAT AAA ACT GGT GTT TCT CCT GAG GGT GTT TCT CAA ACT GAT TAT TGG GCA TTT TCT     479
                           ^

Gly Thr Val Thr Asp Asp Ser Leu Pro Pro Gly Ser Pro Ile Thr Val Leu Val Phe Gly     111
GGA ACT GTA ACA GAC GAT TCT CTT CCT CCG GGT TCT CCT ATT ACG GTA TTA GTA TTT GGT     539

Leu Pro Val Ser Ala Thr Thr Gly Met Thr Ala Ile Glu Phe Val Ala Lys Val Arg Val     131
CTT CCA GTT TCA GCA ACA ACT GGA ATG ACG GCA ATT GAG TTT GTT GCA AAA GTT CGC GTT     599
                                    ^

Ala Leu Gln Glu Ala Ile Ala Ser Phe Thr Ala Ile Asn Ser Tyr Lys Asp His Pro Thr     151
GCA CTA CAA GAA GCT ATT GCG TCA TTT ACT GCT ATC AAT TCA TAT AAA GAC CAT CCA ACT     659

Asp Gly Ser Lys Leu Glu Val Thr Tyr Leu Asp Asn Gln Lys His Val Leu Ser Thr Tyr     171
GAT GGT AGT AAA TTA GAA GTT ACT TAT TTA GAT AAT CAA AAA CAT GTA TTA AGC ACA TAT     719

Ser Thr Tyr Gly Ile Thr Ile Ser Gln Glu Ile Ile Ser Glu Ser Lys Pro Gly Tyr Gly     191
TCT ACA TAT GGA ATA ACT ATT TCC CAA GAA ATT ATA TCT GAG TCT AAG CCT GGC TAT GGT     779
                                                   ^

Thr Trp Asn Leu Leu Gly Ala Gln Thr Val Thr Leu Asp Asn Gln Gln Thr Pro Thr Val     211
ACA TGG AAT TTA TTG GGC GCA CAA ACT GTA ACT TTA GAT AAT CAG CAG ACT CCT ACA GTA     839
                                                        T   T

                                Met  Ser Asn Asn Thr Tyr Gln His Val Ser Asn Glu
Phe Tyr His Phe Glu Arg Thr Ala  *                                                  220
TTT TAT CAT TTT GAG AGA ACA GCA TGA GT AAT AAT ACA TAT CAA CAC GTT TCT AAT GAA      898

Ser Arg Tyr Val Lys Phe Asp Pro Thr Asp Thr Asn Phe Pro Pro Glu Ile Thr Asp Val
TCT CGT TAT GTA AAA TTT GAT CCT ACC GAT ACG AAT TTT CCA CCG GAG ATT ACT GAT GTT     958

Gln Ala Ala Ile Ala Ala Ile Ser Pro Ala Gly Val Asn Gly Val Pro Asp Ala Ser Ser
CAG GCT GCT ATA GCA GCC ATT TCT CCT GCT GGA GTA AAT GGA GTT CCT GAT GCA TCG TCA    1018
                      ^

Thr Thr Lys Gly Ile
ACA ACA AAG GGA ATT C                                                               1035
```

FIG. 4. Nucleotide sequence of bacteriophage T4 gene *11*. The complete sequence of the anti-sense (mRNA-like) strand of gene *11* is shown in addition to the 3′ end of gene *10* and the 5′ end of gene *12*. The predicted amino acid sequence is given above the nucleotide sequence. Italicized nucleotides indicate the positions of translational start and stop codons of gene *11*. Underlined nucleotides locate potential ribosome binding sites in front of genes *11* and *12*. Carets indicate Tn*5* insertion points of deletions utilized in the sequencing protocol. The locations of amber mutations (codons 206 and 207) are indicated by boldface type in the wild-type sequence, with the specific nucleotide changes to amber shown below the sequence. The locations of synthetic oligonucleotide primer binding sites used for confirmation of the sequence are indicated by broken lines.

tions suitable for sequencing and genetic studies. If appropriate precautions are taken to ensure that the isolated insertions are true independent events, the method permits rapid development of a pool of deletion plasmids suitable for DNA sequencing and genetic studies. Moreover, our selection procedure decreases the accumulation of Tn*5* inserts into the plasmid backbone; the probability that an insertion into the backbone would disrupt a region essential for plasmid survival (*i.e.*, β-lactamase production, the plasmid origin of replication or the M13 origin of replication, and DNA packaging) is high. Those few backbone insertion events that are still viable can be distinguished rapidly and eliminated from consideration by the simple restriction endonuclease analysis. In this study, 100% of those insertions isolated by the described selection proved to be localized to the cloned fragment. Furthermore, by using M13- mediated transduction to transfer plasmids containing Tn*5* insertions to the new host, we eliminate the possibility of isolating dimeric plasmids which could conceivably interfere with subsequent genetic analysis and DNA manipulations.

The utility of this method may depend upon the nature of the cloned DNA and its effect on plasmid copy number. In a separate set of experiments, the hyperresistant selection scheme was utilized to isolate Tn*5* insertions into a fragment of bacteriophage T4 DNA containing genes *7*, *8*, and *9* cloned into the same BlueScribe vector with approximately the same success rate (P. Berget, unpublished results). However, an attempt to use this method to isolate Tn*5* insertions into a fragment of bacteriophage P22 DNA containing genes *7*, *16*, and *20* cloned into BlueScribe proved to be unsuccessful (V. Stark, personal communication). Thus, it is important to consider that the hyperresistant selection may be limited by certain characteristics of the cloned DNA.

One step in our procedure that could be simplified arises from the necessity of removing one of the IS*50* inverted repeat sequences by *in vitro* manipulation before DNA sequencing can proceed. In fact, the general utility of this method depends on the restriction map of the cloned DNA fragment and the vector in that one must choose the restriction sites used to make the *in vitro* deletions appropriately. To address this limitation, we have constructed an altered transposon Tn*5* that contains a filamentous phage origin of replication. Once this modified transposon inserts into the fragment of interest cloned into a vector which also has such an origin, the resulting plasmid contains two filamentous phage origins. This results in a streamlined procedure in that *in vivo* deletion of DNA between the two origins occurs spontaneously upon M13 infection of cells carrying such plasmids. This obviates the need for enzymatic digestions and religations to create deletions (B. Barrett, M. Mojumdar, and P. Berget, manuscript in preparation). It is also possible that the use of a transposable element which does not have inverted repeats (such as Mu bacteriophage) could be used to circumvent this *in vitro* manipulation.

We have sequenced bacteriophage T4 gene *11* to continue the analysis of the complex interactions involved in the assembly of the T4 baseplate. Our sequence predicts a protein of 219 amino acid residues with a predicted molecular weight of 23,680 daltons, in close agreement with that determined for the gp11 monomer (King and Laemmli, 1973). In addition, the first part of the predicted sequence agrees well to the amino-terminal protein sequence of gp11.

We sequenced five amber mutations of gene *11* that shared a consistent pattern of marker rescue from our collection of deletion plasmids. Surprisingly, the five independently derived amber mutations represent only two distinct mutations. These are C−T transitions in two adjacent Gln codons at residues 206 and 207. The location of these amber mutations both supports the genetic data and confirms the open reading frame of the sequence. The location of these amber mutations also supports the hypothesis that the carboxyl terminus of the protein is essential for its interaction with gp10 during the assembly of the T4 baseplate (Plishker and Berget, 1984). These two distinct *11*⁻ amber alleles results in the synthesis of gp11 missing only 13 or 14 carboxy-terminal amino acid residues, yet both are defective in assembly. Furthermore, the location of these amber mutations in gene *11* provides an explanation for the paradoxical anti-gp11 serum blocking data reported by Plishker *et al.* (1983). In these experiments it was shown that *11*⁻ lysates made from the amber mutants described above unexpectedly blocked the phage-inactivating activity of anti-gp11 serum. It seems logical now to assume that this serum blocking occurs because of the large size of the amber fragment generated by these amber mutants.

Based on native polyacrylamide gel electrophoresis analysis, Plishker and Berget (1984) suggested that the loss of assembly activity of P11 upon proteolytic digestion with trypsin, subtilisin, or carboxypeptidase Y was concomitant with the loss of positive charged residues from the carboxyl terminus of the gp11 monomers. However, the predicted protein sequence does not support this hypothesis. There is not a sufficient number of charged residues in the carboxyl terminus of gp11 to account for the regular nature of the altered mobilities. Since native polyacrylamide gel electrophoresis separates proteins on the basis of both charge and hydrodynamic shape, the increased mobility of gp11 after proteolysis is likely due to both disruption of the protein conformation and loss of peptides. In addition, it is apparent that the amino-terminal and carboxy-terminal residues (tyrosine and valine, respectively) suggested by previous work on the protein (Plishker and Berget, 1984) are in error.

It should also be noted that our sequence, which extends through the 5′ end of gene *12* to the *Eco* RI site, does not completely agree with the published sequence of gene *12* (Selivanov *et al.*, 1988). We have carefully reexamined our sequence and confirmed it on both strands for this region. In the paper by Selivanov *et al.* (1988), the CAC codon at nucleotide positions 123–125 should read CAG. Thus, the proposed histidine residue is actually a glutamine residue. Moreover, the *Eco* RI site shown at nucleotide position 163 is in error; a deletion of 30 bp is evident in their sequence. The sequence in that region should read as we have determined it from nucleotide positions 1,000 to 1,035.

Because bacteriophage T4 gene *11*-containing DNA had few convenient restriction sites, we decided to develop a method suggested by the work of Berg *et al.* (1983) and Sasakawa *et al.* (1982) to create a collection of deletion plasmids. Jacobs *et al.* (1985) and Egelhoff *et al.*)1985) have used Tn*5* insertions to map nodulation genes in *Rhizobium meliloti* and have used the inverted repeat ends of Tn*5* as primer binding sites to sequence portions of the *nod* gene cluster. Their procedures differ from ours in several ways. Fragments of all of their Tn*5* insertions were either subcloned after isolation onto M13 vectors for sequencing or sequenced directly from fragments isolated from agarose gels. In our protocol, the DNA to be sequenced was cloned originally into a filamid vector so that single-stranded DNA suitable for sequencing could be produced when desired. Furthermore, Jacobs and Egelhoff did not employ a hyperresistance selection protocol in the isolation of their original Tn*5* insertions. The hyperresistance selection has proved extremely useful in the generation of Tn*5* insertions into plasmids for sequencing. Finally, we used infection by helper M13 phage to resolve potential heterogeneous filamid dimers. In our experiments, the sequence of the entire cloned gene *11* fragment was covered by the overlapping sequence generated by six Tn*5* insertions. With a minimum of enzymatic manipulations we were able to sequence both strnads completely.

ACKNOWLEDGMENTS

The authors thank Colleen Sari and Jian Qing Zhang for constructing plasmids pCS13 and pJZ1, respectively. The authors also thank Christopher C.Q. Chin and Finn Wold from the Analytical Chemistry Center of the University of Texas Health Science Center at Houston for determining the sequence of the amino terminus of gp11, and George Weinstock of the University of Texas Medical School at Houston for providing the secondary Tn*5* primer. This research was supported by the National Science Foundation grant PCM 8314241 and the National Institutes of Health grant GM38302 to P.B.B.

REFERENCES

BERG, D.E., JOHNSRUD, L., McDIVITT, L., RAMABHADRAN, R., and HIRSCHEL, B.J. (1982). The inverted repeats of Tn*5* are transposable elements. Proc. Natl. Acad. Sci. USA **79**, 2632–2635.

BERG, D.E., SCHMANDT, M.A., and LOWE, J.B. (1983). Specificity of transposon Tn*5* insertion. Genetics **105**, 813–828.

BERGET, P.B., and KING, J. (1978). Isolation and characterization of precursors in bacteriophage T4 baseplate assembly. I. The complex of gene *10* and gene *11* products. J. Mol. Biol. **124**, 469–486.

BIRNBOIM, H.C., and DOLY, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**, 1513–1523.

EGELHOFF, T.T., FISHER, R.F., JACOBS, T.W., MULLIGAN, J.T., and LONG, S.R. (1985). Nucleotide sequence of *Rhizobium meliloti* 1021 nodulation genes: *nodD* is read divergently from *nodABC*. DNA **4**, 241–248.

JACOBS, T.W., EGELHOFF, T.T., and LONG, S.R. (1985). Physical and genetic map of a *Rhizobium meliliti* nodulation gene region and nucleotide sequence of *nodC*. J. Bacteriol. **162**, 469–476.

KING, J., and LAEMMLI, U.K. (1973). Bacteriophage T4 tail assembly: Structural proteins and their genetic identification. J. Mol. Biol. **75**, 315–337.

KRAFT, R., TARDIFF, J., KRAUTER, K.S., and LEINWAND, L.A. (1988). Using mini-prep plasmid DNA for sequencing double stranded templates with Sequenase. Biotechniques **6**, 544–547.

LODGE, J.K., WESTON-HAFER, K., and BERG, D.E. (1988). Transposon Tn*5* target specificity: Preference for insertion at G/C pairs. Genetics **120**, 645–650.

MANIATIS, T., FRITSCH, E.F., and SAMBROOK, J. (1982). *Molecular Cloning: A Laboratory Manual.* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).

MILLER, J.H. (1972). *Experiments in Molecular Genetics.* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).

OLIVER, D.B., MALAMY, M.H., and GOLDBERG, E.B. (1981). Cloned genes for bacteriophage T4 late functions are expressed in *Escherichia coli*. J. Mol. Biol. **152**, 267–283.

PLISHKER, M.F., and BERGET, P.B. (1984). Isolation and characterization of precursors in bacteriophage T4 baseplate assembly. III. The carboxy-termini of protein 11 are required for assembly activity. J. Mol. Biol. **178**, 699–709.

PLISHKER, M.F., CHIDAMBARAM, M., and BERGET, P.B. (1983). Isolation and characterization of precursors in bacteriophage T4 baseplate assembly. II. Purification of the protein products of genes *10* and *11* and the *in vitro* formation of the P(10/11) complex. J. Mol. Biol. **170**, 119–135.

ROTHSTEIN, S.J., and REZNIKOFF, W.S. (1981). The functional differences in the inverted repeats of Tn*5* are caused by a single base pair non-homology. Cell **23**, 191–199.

SANGER, F., NICKLEN, S., and COULSON, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**, 5463–5467.

SASAKAWA, C., LOWE, J.B., McDIVITT, L., and BERG, D.E. (1982). Control of transposon Tn*5* transposition in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **79**, 7450–7454.

SELIVANOV, N.A., PRILIPOV, A.G., and MESYANZHINOV, V.V. (1988). Nucleotide and deduced amino acid sequence of bacteriophage T4 gene *12*. Nucleic Acids Res. **16**, 2334.

SHINE, T., and DALGARNO, L. (1974). The 3′-terminal sequence of *E. coli* 16S ribosomal RNA: Complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA **71**, 1342–1346.

STAHL, F.W., CRASEMANN, J.M., YEGIAN, C., STAHL, M.M., and NAKATA, A. (1970). Co-transcribed cistrons in bacteriophage T4. Genetics **64**, 157–170.

TABOR, S., and RICHARDSON, C.C. (1987). DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc. Natl. Acad. Sci. USA **84**, 4767–4771.

Address reprint requests to:
Dr. Peter B. Berget
4400 Fifth Avenue
Pittsburgh, PA 15213-3890

Received for publication November 29, 1988, and in revised form January 10, 1989.