SPECIFICITY OF TRANSPOSON Tn*5* INSERTION

DOUGLAS E. BERG,*† MARGARET A. SCHMANDT* AND JOHN B. LOWE*‡

*Departments of *Microbiology and Immunology, †Genetics and ‡Laboratory Medicine, Washington University Medical School, St. Louis, Missouri 63110*

Manuscript received July 5, 1983
Revised copy accepted August 20, 1983

ABSTRACT

Genetic mapping studies had shown that the bacterial transposon Tn*5* can insert into many sites in a gene, but that some sites are preferred. To begin understanding Tn*5*'s insertion specificity at the molecular level, we selected transpositions of Tn*5* from the *Escherichia coli* chromosome to the plasmid pBR322 and analyzed the resultant pBR322::Tn*5* plasmids by restriction endonuclease digestion and DNA sequencing. Seventy-five insertions in the *tet* gene were found at 28 sites including one major hotspot (with 21 insertions) and four lesser hotspots (with four to ten insertions each). All five hotspots are within the first 300 of the 1250-base pair (bp) *tet* gene. In contrast, 31 independent insertions in the *amp* gene were found in at least 27 distinct sites.—Tn*5* generates 9 bp target sequence duplications when it transposes. Such transposon-induced duplications are generally taken to indicate that cleavages of complementary target DNA strands are made 9 bp apart during transposition. DNA sequence analysis indicated that GC base pairs occupy positions 1 and 9 in the duplications at each of the five hotspots examined, suggesting a GC-cutting preference during Tn*5* transposition.

TRANSPOSABLE elements insert into numerous sites in a genome without need for extensive DNA sequence homology. For most elements, however, insertion sites are nonrandomly distributed, and the distribution of sites seems to be element specific (MILLER *et al.* 1980; GALAS, CALOS and MILLER 1980; BERG, WEISS and CROSSLAND 1980; HALLING and KLECKNER 1982; TU and COHEN 1980; LICHTENSTEIN and BRENNER 1982; SAINT-GIRONS *et al.* 1981; SENGSTAG and ARBER 1983; KLAER *et al.* 1981; ENGLER and VAN BREE 1981). The transposase proteins, upon which movement depends, are believed to bind distinctive DNA sequences at the ends of their respective elements. It is parsimonious to consider that these same transposase proteins also participate in insertion site selection, albeit with a specificity which is different or lower than that used in identifying the termini of transposable elements (for reviews, BERG and BERG 1981, 1983; SHAPIRO 1983; KLECKNER 1981).

Tn*5*, whose insertion specificity is assessed here, is a composite element in which terminal inverted repeats of smaller transposition modules named IS*50*L (left) and IS*50*R (right) (BERG *et al.* 1982a) bracket a central segment encoding kanamycin resistance (Figure 1); these *IS50* elements are not homologous to any of the IS elements indigenous to *Escherichia coli* K-12 (BERG and DRUMMOND 1978). Tn*5* has proven useful for detailed studies of the mechanism and

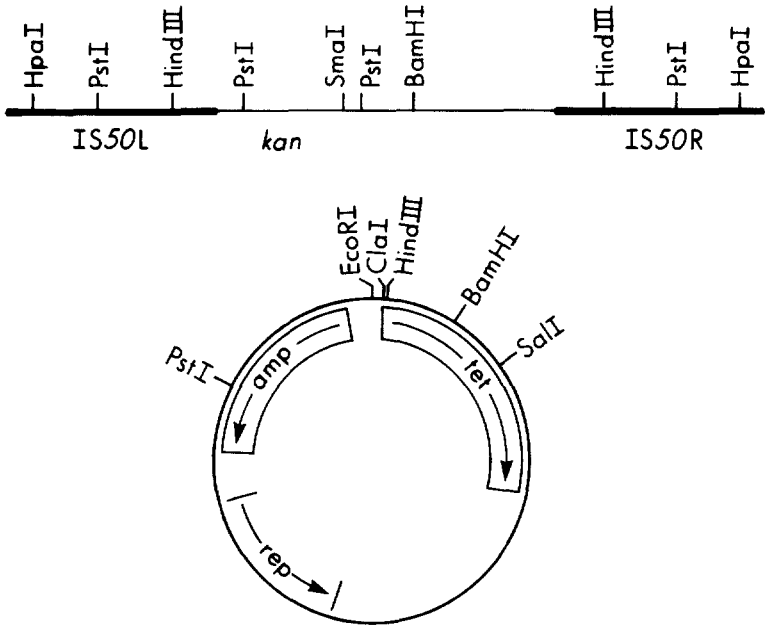

FIGURE 1.—Restriction maps of Tn*5* and pBR322. Top, The 5700-bp Tn*5* element. Thickened lines, IS*50* elements present as 1534-bp terminal inverted repeats (BERG *et al.* 1975, 1982a). The positions in base pairs of restriction endonuclease cleavage sites from the outside ends of the IS*50* elements are: *Hpa*I, 186; *Pst*I, 681 and *Hind*III, 1196. The *Bam*HI and *Sma*I sites in the central region of Tn*5* are separated by about 500 bp. Tn*5* is not cut by *Eco*RI or *Cla*I (from JORGENSEN, ROTHSTEIN and REZNIKOFF 1979; AUERSWALD, LUDWIG and SCHALLER 1980; COLLINS, VOLCKAERT and NEVERS 1982). Bottom, pBR322. The positions of restriction sites in the 4363-bp plasmid are *Eco*RI, 4361; *Cla*I, 24; *Hind*III, 29; *Bam*HI, 375; *Sal*I, 651; *Pst*I, 3613. The plasmid is not cleaved by *Hpa*I or *Sma*I. The positions and directions of transcription of the loci conferring resistance to ampicillin and tetracycline and also the RNA primer for DNA replication are shown (SUTCLIFFE 1978; PEDEN 1983).

control of transposition and has become a superb tool for molecular-genetic analyses and manipulations in diverse bacterial species (BERG and BERG 1981, 1983). The restriction endonuclease and DNA sequence analyses reported here show that sites of Tn*5* insertion in plasmid pBR322 are nonrandomly distributed. Several Tn*5* insertion hotspots have been identified: more than one-tenth of all Tn*5* insertions in this 4363-bp plasmid occur at the most prominent site.

MATERIALS AND METHODS

All bacterial strains are derivatives of *E. coli* K-12. MC1061 is F⁻ Δ(*ara-leu*) *hsr* (CASADABAN and COHEN 1980). DB1572 is F⁻ *recA lac*Z124::Tn*5* Δ*trp*E5 (LOWE and BERG 1983). DB1891 is F⁻ Δ*trpE5 hfl*-1 Δ*srl-recA* 306 (SASAKAWA *et al.* 1982). Monomeric plasmid pBR322 (Figure 1) and the dimeric form of pBR322 derived from it by homologous recombination *in vivo* were introduced into DB1572 by transformation.

LN broth contained 10 g of Humko-Sheffield NZ amine, 5 g of Difco yeast extract and 10 g of NaCl per liter adjusted to pH 7.2. Media was solidified with 15 g Bacto Difco agar per liter. Antibiotics were used at the following concentrations: ampicillin, 250 μg/ml; tetracycline, 12.5 μg/

ml; kanamycin, 50 μg/ml or neomycin, 250 μg/ml (to select strains hyperresistant to aminoglycosides).

Standard molecular-genetic techniques were used (MANIATIS, FRITSCH and SAMBROOK 1982). Plasmid DNAs were generally extracted from stationary phase LN broth cultures or from pools of colonies grown for 36 hr on LN-neomycin (250 μg/ml) agar using the alkaline-SDS rapid lysis procedure. DNAs used for end labeling and DNA sequence analysis were generally purified using the cleared lysis procedure and centrifugation to equilibrium in CsCl-ethidium bromide. Plasmid DNAs were introduced into new hosts by transformation of $CaCl_2$-shocked cells. Restriction endonuclease digestions, ligations with T4 DNA ligase and repair syntheses using the Klenow fragment of DNA polymerase I were carried out according to the suppliers' recommendations (New England Biolabs and Bethesda Research Laboratories). Intact and restriction endonuclease-digested DNAs were electrophoresed in Tris-acetate-buffered 0.7% agarose slab gels or in polyacrylamide slab gels and photographed after staining with ethidium bromide.

The junctions between pBR322 and Tn*5* at the five hotspots diagrammed in Figure 4 were sequenced using the chemical modification method (MAXAM and GILBERT 1980). The 3′ ends of DNA fragments were labeled by the filling reaction of DNA polymerase I (New England Biolabs) in the presence of the appropriate [α-^{32}P]deoxynucleoside triphosphate (Amersham) ([α-^{32}P]-dGTP for labeling *Cla*I-cleaved DNA; [α-^{32}P]-dTTP for labeling *Hind*III-cleaved DNA and [α-^{32}P]-dCTP for labeling *Bam*HI cleaved-DNA; ROBERTS 1981). The DNAs were digested with a second restriction endonuclease and electrophoresed in a preparative polyacrylamide gel, and the fragments to be sequenced were eluted electrophoretically from the appropriate gel slice prior to base-specific chemical cleavage, electrophoresis and autoradiography.

RESULTS

Transposition of Tn5 to pBR322: Preliminary experiments indicated that cells harboring Tn*5* in the multicopy pBR322 plasmid are resistant to higher levels of neomycin than isogenic cells containing Tn*5* inserted at a single chromosomal locus, and that the hyperresistant phenotype could be used to enrich for transposition of Tn*5* from the *E. coli* chromosome to pBR322 (SASAKAWA *et al.* 1982). The survival of strain DB1572 (*recA*$^-$ *lac*::Tn*5*) harboring pBR322 on high-neomycin medium was generally 10^{-5} to 10^{-6}. Agarose gel electrophoresis of plasmid DNA extracted from pools of hyperresistant colonies showed that 5–50% (varying from clone to clone) was 10 kb, the size of pBR322::Tn*5*, and that the remainder was 4.4 kb, the size of pBR322. pBR322::Tn*5* plasmids were isolated by transforming MC1061 to Kan^r with plasmid DNAs from each of 90 independent pools of hyperresistant colonies. Ten to 12 Kan^r transformants were tested from each transformation. About 40% of the Kan^r transformants were sensitive to tetracycline and 10% to ampicillin. [The *tet* and *amp* determinants are about 1250 and 900 bp in size, respectively (SUTCLIFFE 1978; PEDEN 1983).] To ensure independence, no more than one Kan^r transformant of each phenotype (Amp^r Tet^s, Tet^r Amp^s or Tet^r and Amp^r) was saved from each of the 90 pools.

Mapping sites of Tn5 insertion: Tn*5* is known to insert nonrandomly, although it does so into many sites within a gene (BERG 1977; SHAW and BERG 1979; BERG, WEISS and CROSSLAND 1980; MILLER *et al.* 1980; M. HOWE, personal communication). To try to understand how targets for Tn*5* insertion are selected we looked for hotspots for Tn*5* insertion in pBR322. This plasmid may be divided into four regions (*amp*, *tet*, *rep* and "other"; see Figure 1). Although insertions in *amp* and *tet* are easiest to analyze because of the antibiotic sensi-

tivity resulting from Tn*5* insertion, it was important to test whether there might be prominent insertion hotspots elsewhere in pBR322.

Dimeric plasmids were used to look for insertions in the indispensable 550-bp *rep* region. Since in these plasmids all sequences are duplicated, one copy of the 550-bp *rep* region is dispensable; any pBR322/pBR322::Tn*5* dimer carrying Tn*5* in one *rep* region should be viable and recognizable by its failure to give rise to viable monomeric pBR322::Tn*5* derivatives.

Transpositions of Tn*5* were generated in the *recA*$^-$ strain DB1572 carrying stable dimeric pBR322. The resultant pBR322/pBR322::Tn*5* mixed dimers were selected by transformation of *rec*A$^-$ strain DB1891 to Kan^r with plasmid DNA extracted from hyperresistant populations. Each of the 16 independent Kan^r plasmid DNAs tested was, as expected, about 15 kb in size and gave rise to two fragments after *Eco*RI digestion, 4.4 kb (pBR322) and 10 kb (pBR322::Tn*5*) in length. The *Eco*RI-digested DNAs were ligated at low concentrations to favor intramolecular circularization and used to transform MC1061 to Amp^r. In each case about half of the transformants were Kan^r, and these contained plasmids 10 kb in size. Thus, Tn*5* had not inserted into the essential *rep* region in any of 16 independent transposition events.

Restriction endonuclease digestions were used to look for major hotspots outside of the *tet*, *amp* and *rep* loci. Approximate mapping was based on the principle that, as a consequence of the inverted orientation and the near-identity of Tn*5*'s terminal IS*50* elements, the distances between any fixed point in pBR322 and the sites in IS*50* recognized by enzymes such as *Hpa*I, *Pst*I and *Hind*III (Figure 1) depend on the location but not the orientation of Tn*5* (see Figure 2, bottom). The DNAs of 12 independent Amp^r Tet^r pBR322::Tn*5* plasmids were digested with *Hind*III (cleavage 1196 bp from the outside end of each *IS50* element and also in the *tet* promoter, pBR322 position 29), and the sizes of the resultant fragments were measured using agarose gel electrophoresis. The fusion fragments of each of the 12 plasmids tested were different in size, indicating that each plasmid contained Tn*5* at a different site. Thus, there seem to be no preeminent hotspots for Tn*5* insertions in *rep* or in the regions of Tn*5* collectively designated "other."

For more detailed analyses of the specificity of Tn*5* insertion in pBR322 we focused on 106 insertions in the *amp* and in the *tet* genes (31 and 75, respectively). Preliminary mapping was based on the results of *Hind*III cleavage; digestions of representative pBR322 *amp*::Tn*5* and *tet*::Tn*5* plasmids are shown in Figure 2 (the numbers above each lane indicate the position in pBR322 of the Tn*5* insertion). Although it was apparent from these digests that Tn*5* had inserted into numerous sites in *amp* and in *tet*, some nonrandomness was also evident.

To map sites of Tn*5* insertion in *amp* with higher resolution, plasmid DNAs were digested with *Pst*I (cleavage within *amp*, and also 681 bp from the ends of Tn*5*) or in selected cases with both *Pst*I and *Hpa*I (cleavage 186 bp from the ends of Tn*5*) and electrophoresed in a polyacrylamide gel (see Figure 3). Whenever two plasmid isolates seemed to contain Tn*5* at about the same location, digests of these DNAs were electrophoresed again in adjacent lanes

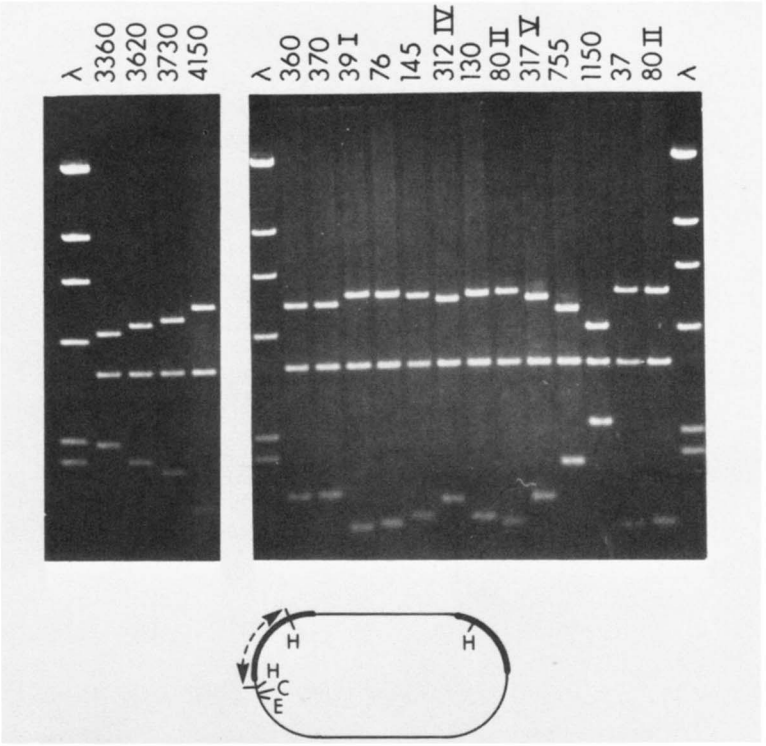

FIGURE 2.—Mapping Tn*5* insertions by *Hind*III digestion. Top, *Hind*III-digested pBR322::Tn*5* plasmid DNAs electrophoresed in a 0.7% agarose slab gel. The number above each lane refers to the position (base pairs) of Tn*5* insertion in pBR322. Left, DNAs from *amp*::Tn*5* plasmids. Right, DNAs from *tet*::Tn*5* plasmids. *Hind*III digested λ DNA provides the size standard. Bottom, Diagram of the positions of *Hind*III cleavage sites in a pBR322::Tn*5* plasmid; the short fusion fragment useful for estimating the position of Tn*5* insertion is indicated by a dashed arrow. E, C and H designate *Eco*RI, *Cla*I and *Hind*III sites, respectively. Digests of plasmids with insertions at hotspots I, II, IV and V are indicated by roman numerals.

to facilitate direct comparisons. The 31 *amp*::Tn*5* plasmids studied contained insertions of Tn*5* at 27 widely (but not randomly) distributed sites (Figure 4).

To map closely spaced insertions in *tet* with high resolution we carried out double digestions: *Hpa*I or *Xho*I (cleavage 186 and 486 bp from the ends of Tn*5*, respectively) and *Hind*III, *Cla*I, *Bam*HI or *Sal*I, depending on the approximate location of Tn*5* estimated using *Hind*III. Plasmids in which Tn*5* was near the *Bam*HI site or the *Sal*I site of pBR322 were also digested with *Bam*HI plus *Eco*RI (Figure 5) or *Sal*I plus *Eco*RI, to distinguish insertions clockwise and counterclockwise from these sites. For resolution at the level of a single base pair, sets of DNAs containing Tn*5* at apparent hotspots were 3′ end labeled after *Cla*I digestion (sites I, II and III in Figure 4) or *Bam*HI digestion (sites IV and V in Figure 4) using DNA polymerase and [α-^{32}P]-deoxynucleoside triphosphates, cleaved with *Hpa*I, electrophoresed in a polyacrylamide DNA sequencing gel and visualized by autoradiography (Figure 6).

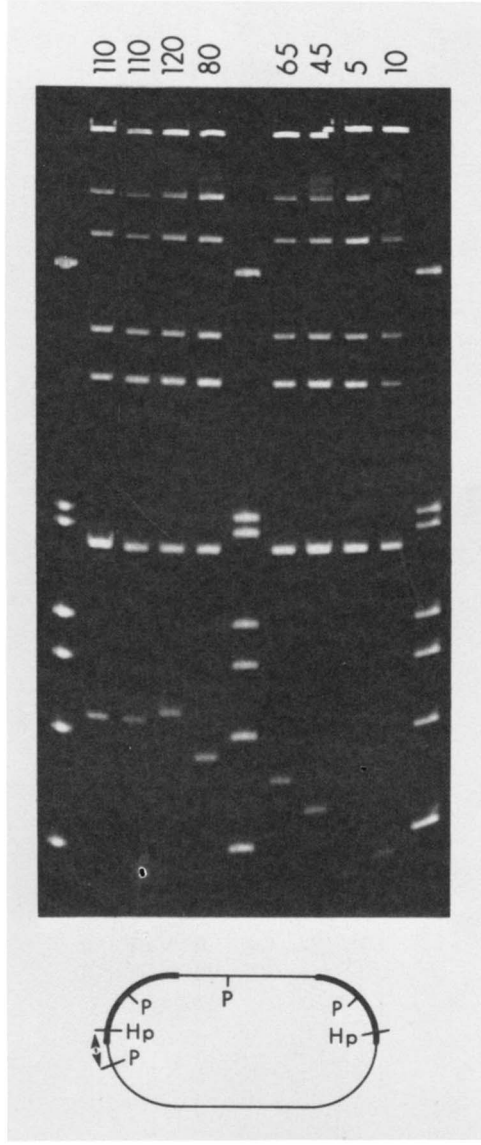

FIGURE 3.—Fine structure mapping of *amp*::Tn*5* mutants. Top, DNAs of *amp*::Tn*5* plasmids digested with *Pst*I plus *Hpa*I and electrophoresed in an 8% polyacrylamide gel. The numbers above each lane refer to the estimated distance in base pairs of the Tn*5* insertion from the *Pst*I site in *amp*. Bottom, Diagram of sites in pBR322::Tn*5* cleaved by *Pst*I and by *Hpa*I. A fusion fragment used to estimate the position of Tn*5* insertion is indicated by the arrow. P and Hp designate the *Pst*I and *Hpa*I sites, respectively.

The 75 *tet*::Tn*5* insertions were found distributed among 28 sites but very nonrandomly. There were five hotspots with 21, ten, seven, five and four transposition events, respectively. The remaining 28 insertions were distributed among at least 23 sites (Figure 4). Each of the five hotspots corresponded to a single nucleotide position (see Figure 6 and Table 1).

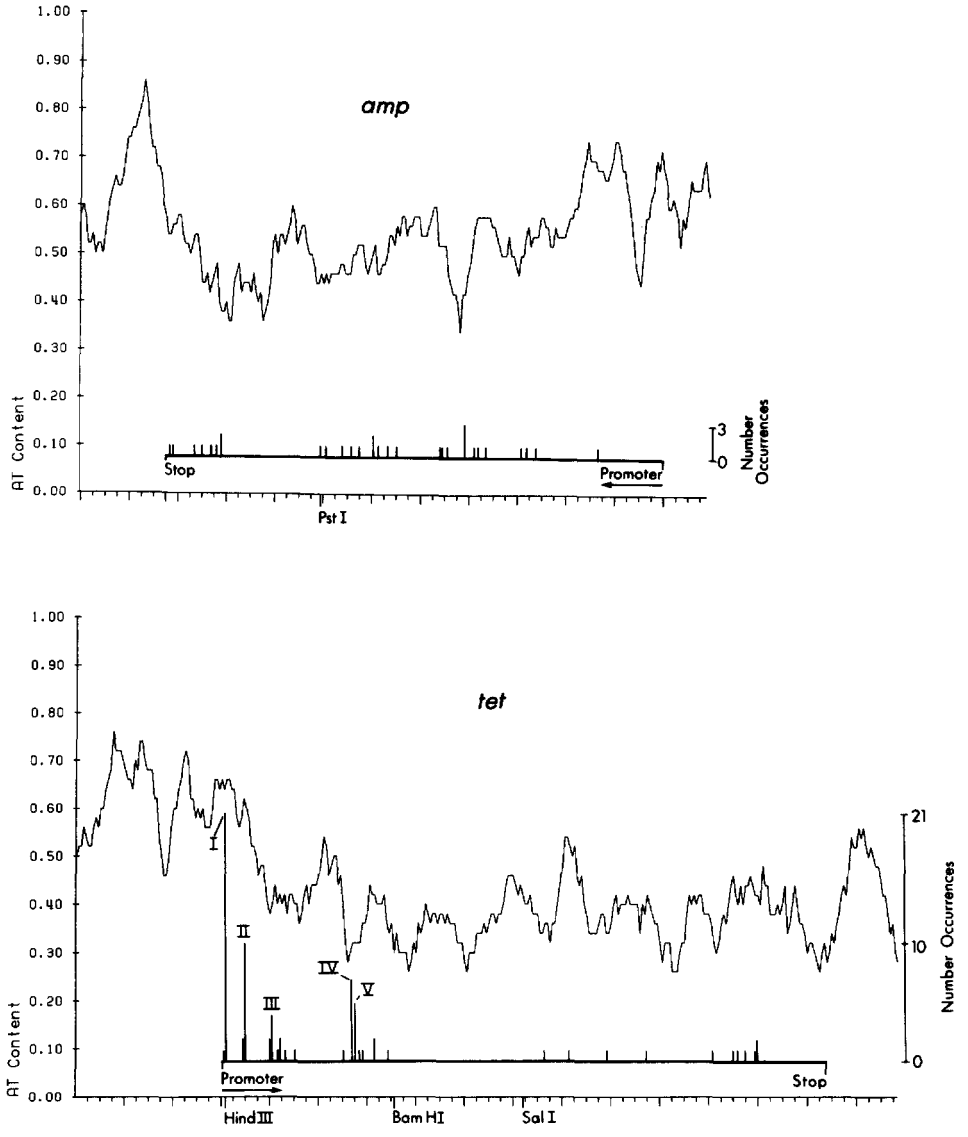

FIGURE 4.—Map of Tn5 insertions in the *amp* and *tet* genes of pBR322 and of the base composition of these genes. The insertion map was determined by restriction endonuclease digestion of pBR322::Tn5 plasmids as described in the text. The approximate positions of the promoter and stop codon for the *amp* gene are 4327 and 3295 bp, respectively. The promoter and stop codon of the *tet* gene (which is in the opposite orientation) are 12 and 1276 bp, respectively (SUTCLIFFE 1978; PEDEN 1983). The number of insertions at any site is given on the vertical axis at right. The marks along the horizontal axis are spaced 25 bp apart. The DNA sequences at the insertion hotspots in *tet* labeled I, II, III, IV and V have been determined (see Table 2). Other sites that are presented as containing two or three independent insertions of Tn5 have not been analyzed in detail and may actually be a cluster of closely spaced sites, each containing only one insert; the restriction endonuclease digestions used to characterize those insertions would not distinguish sites a few pairs apart. The adenine plus thymine content was determined using 50-bp intervals, spaced 5 bp apart (*i.e.*, the AT content of 1→50, then 6→55, then 11→60, etc.) by G. CARLE.

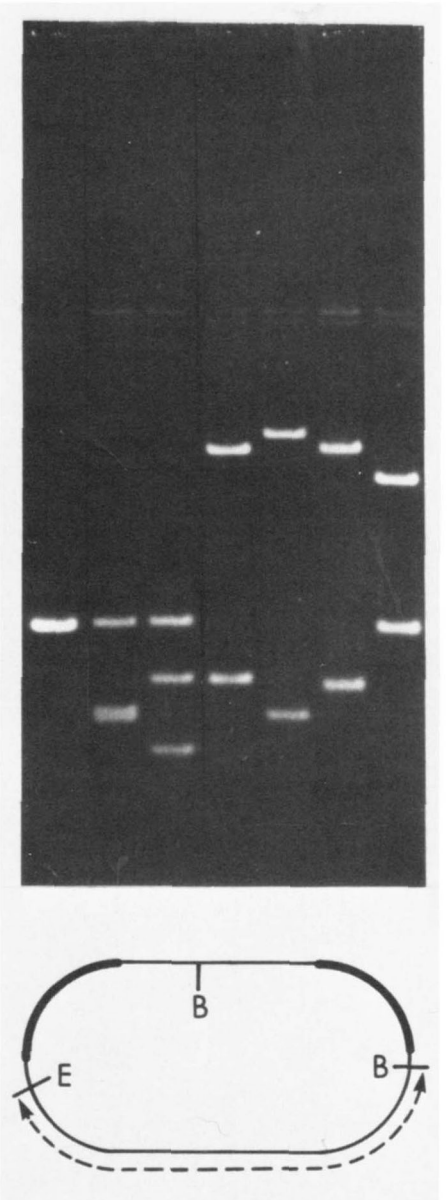

FIGURE 5.—Tn*5* localized *vis a vis* the *Bam*HI site in *tet*. Top, Each DNA was digested with *Bam*HI and with *Eco*RI and electrophoresed in a 0.7% agarose gel. Lane 1, pBR322; lanes 2 and 3, pBR322 plasmids which contain Tn*5* counterclockwise from the *Bam*HI site; lanes 4 to 7, plasmids which contain Tn*5* clockwise from the *Bam*HI site. Note that three DNA fragments are generated by *Eco*RI plus *Bam*HI digestion of any pBR322::Tn*5* plasmid (diagram at bottom). These are most evident in lane 3 and can also be seen in lane 2 where the faster moving band is a doublet. For insertions clockwise from the pBR322 *Bam*HI site only two bands are evident, because the third is small (*Eco*RI-*Bam*HI, 375 bp) and has migrated off this gel. Bottom, Diagram of Tn*5* inserted between the *Bam*HI (B) and *Eco*RI (E) sites of pBR322. The dashed arrow indicates the 4-kb pBR322 fragment generated by digestion of plasmids containing Tn*5* in the 375-bp segment of the *tet* gene between the *Eco*RI and *Bam*HI sites.

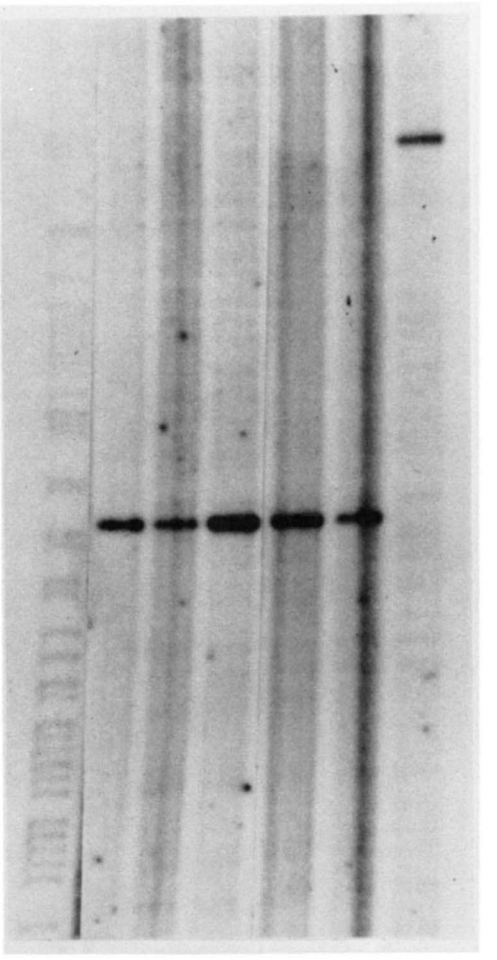

FIGURE 6.—Tn*5* insertions at site I occupy a single nucleotide position. pBR322 *tet*::Tn*5* DNAs, digested with *Cla*I, 3′ end labeled with [α^{32}P]-dGTP, digested with *Hpa*I and electrophoresed in a 8% polyacrylamide-urea gel prior to autoradiography. Lanes 2–6 contain DNAs from site I insertion mutants (Figure 4). The identical position of the labeled band in lanes 2–6 shows that the Tn*5* element in each plasmid is inserted at the same site. Lane 1 contains a section of a Maxam-Gilbert sequencing ladder showing that single base pair differences can be distinguished in this gel; lane 7, the DNA of a site II mutant (Tn*5* inserted 41 bp away from site I), also digested with *Cla*I, end labeled and digested with *Hpa*I.

Orientation of tet::*Tn*5 *insertions*: The IS*50*R component of Tn*5* encodes transposase, whereas IS*50*L contains the *kan* gene promoter but does not encode transposase in the *E. coli* strain used (ROTHSTEIN *et al.* 1980; ROTHSTEIN and REZNIKOFF 1981). A comparison of single and double digestions (*Bam*HI and *Bam*HI plus *Sma*I) was used to determine whether there are preferred orientations of Tn*5* insertion, especially at the hotspots. Because *Bam*HI cleaves pBR322 and Tn*5* each at one site and *Sma*I cleaves Tn*5* to the left of the *Bam*HI site, but does not cleave pBR322 (Figure 1), Tn*5*'s orientation can be determined after identifying the *Bam*HI fragment whose size is decreased by

*Sma*I digestion (Figure 7). Of the 21 insertions at hotspot I, 12 were in the R orientation (IS*50*R closest to the *tet* promoter) and nine were in the L orientation (IS*50*L closest to the *tet* promoter). Tn*5* was also found in both orientations at each of the four other hotspots, and in general Tn*5* seemed not to prefer a particular orientation when it inserted in *tet*. In contrast among *amp*::Tn*5* mutants the R orientation seemed to be preferred (22R:9L) (Table 1).

Target sequences at hotspots: To precisely identify hotspots for Tn*5* insertion the DNA sequences of one pBR322-Tn*5* junction fragment from each of eight plasmids (one containing Tn*5* in each orientation at sites I, II and III and single representatives of insertions at sites IV and V) were determined. The

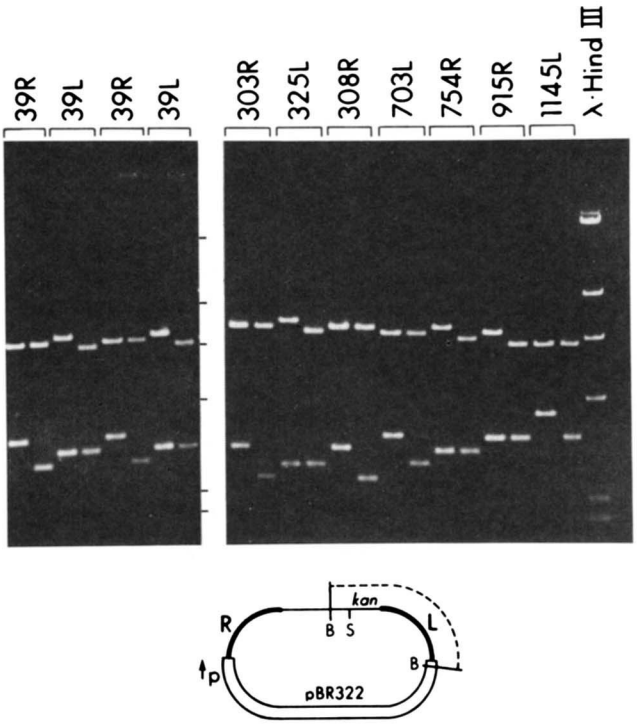

FIGURE 7.—Orientation of Tn*5* insertions. Top, Plasmid DNAs were digested with *Bam*HI and with *Bam*HI plus *Sma*I (left and right members of each pair, respectively), and the DNAs were electrophoresed in 0.7% agarose gels. R and L refer to the orientation inferred from these digests, and the numbers 39, 303, 325, etc., refer to the nucleotide position in pBR322 at which Tn*5* is inserted (39, 303 and 308 correspond to hotspots I, IV and V, respectively). The positions to which λ-*Hind*III fragments had migrated are drawn after the first four lanes, and the migration of *Hind*III-digested λ DNA is also shown in the rightmost lane. Note that the interpretation of the orientation of an insertion in *tet* obtained from these digests depends on the position of the insertion relative to the *Bam*HI site. For example, the plasmids designated 308R and 703L are in opposite orientations, but in both cases the smaller *Bam*HI fragment is shortened by *Sma*I digestion, a consequence of their positions on either side of the *Bam*HI site. Bottom, Diagram of a pBR322::Tn*5* plasmid containing Tn*5* upstream of the *Bam*HI site in *tet* (position 375) and in the R orientation *vis a vis* the *tet* promoter (p). The smaller *Bam*HI fragment contains the *Sma*I site.

TABLE 1

Orientation of Tn5 insertions[a]

Location of insertions[b]	IS*50* element closest to promoter	
	R	L
tet, hotspot		
I	12	9
II	3	7
III	1	3
IV	4	3
V	3	2
tet, not hotspots	14	14
Total in *tet*	37	38
Total in *amp*	22	9

[a] The orientation of Tn*5* elements was determined by digestion of pBR322::Tn*5* plasmid DNAs with *Bam*HI and with *Bam*HI plus *Sma*I as outlined in Figure 7 and described in the text.

[b] Hotspots I, II, III, IV and V refer to preferred sites of insertion proceeding clockwise from the *tet* promoter (the *Hind*III site) as shown in Figure 4.

critical section of a polyacrylamide-sequencing gel containing the site of fusion of Tn*5* and pBR322 at hotspot I is shown in Figure 8. The DNA sequence determinations indicate that GC base pairs occupy the first and ninth positions of the target sequence duplications at each of the five Tn*5* insertion hotspots analyzed (Table 2).

DISCUSSION

Tn*5* belongs to a group of transposable elements whose members, although probably not closely related, all generate direct duplications of 9 bp of target sequences when they transpose. Such duplications are thought to reflect the positioning of cuts in complementary strands of target DNAs early during transposition and, hence, to provide clues to determinants of insertion specificity. Tn*5* exhibits a complex pattern of specificity in its choice of insertion sites. A tendency to insert repeatedly into a few preferred sites (hotspots) is superimposed on a general ability to insert into a large number of other sites. Of 75 independent insertions of Tn*5* in the *tet* locus of pBR322, 21 were at a single site, 26 were distributed among four lesser hotspots and the remaining 28 insertions were found in at least 23 different sites. No major hotspot for Tn*5* insertion was found in the *amp* gene, but otherwise the distribution (31 insertions in at least 27 sites) resembles that seen in *tet* (Figure 4). Because of the complexity of these patterns, several signals in the DNA sequence at or near a preferred site may each contribute to the probability of insertion there.

At each of the five hotspots in the *tet* locus of pBR322 the duplications both begin and end with GC base pairs (Table 2). GC base pairs at positions 1 and 9 were also seen at three of five other sequenced sites of Tn*5* insertion (SCHALLER 1978; BOSSI and CIAMPI 1981) and at one of three sequenced sites of IS*50* insertion (BERG *et al.* 1982a). Since the duplications at the other three sites contain only one terminal GC base pair, there seems to be a preference,

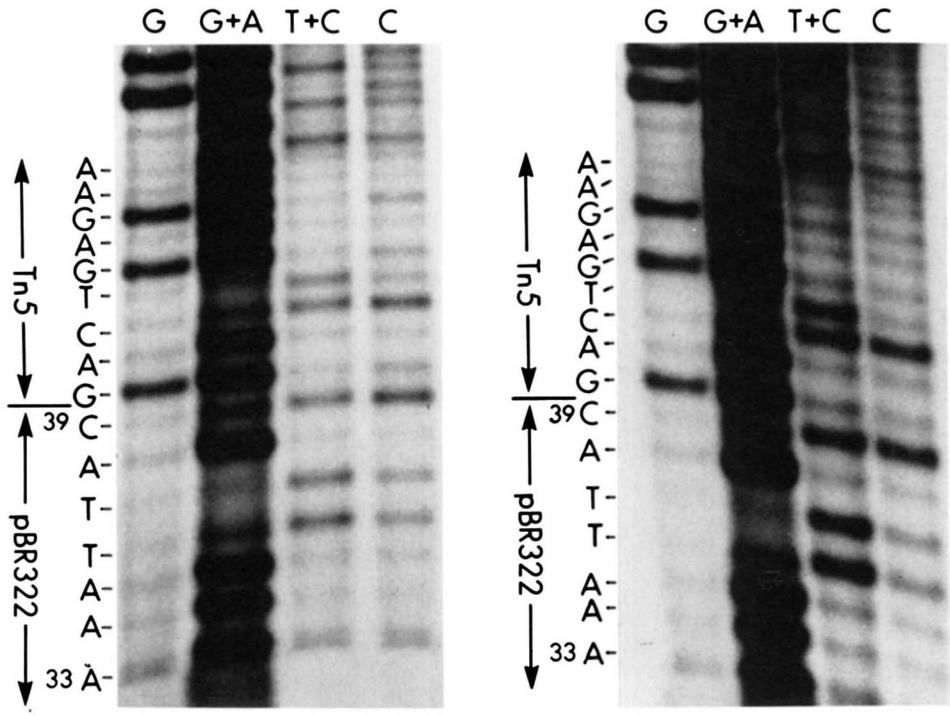

FIGURE 8.—Determination of the DNA sequence at hotspot I. Autoradiogram of the section of a sequencing gel showing the pBR322-Tn*5* junctions for two independent insertions at hotspot I, one in each orientation. Because the DNAs were 3′ end labeled for sequencing, the sequence written next to the gel is the complement of the standard presentation of pBR322 and IS*50* sequences which would be 5′TTTAATG/CTGACTCTT3′ going from pBR322 into Tn*5*. The pBR322 sequence presented as 3′AAATTAC5′ corresponds to positions 33 through 39.

TABLE 2

Hotspots for Tn5 insertion

Hot-spot	No. se-quenced	Sequence at insertion site[a]	Fre-quency	pBR322 position[b]
I	2	TATCATCGATAAGCTTTAATGCGGTAGTTTATC	21/75	31–39
II	2	ATTGCTAACGCAGTCAGGCACCGTGTATGAAAT	10/75	72–80
III	2	TCGGCACCGTCACCCTGGATGCTGTAGGCATAG	4/75	129–137
IV	1	GGCCGCCGCCCAGTCCTGCTCGCTTCGCTACTT	7/75	303–311
V	1	GCTTTGGCCGCCGCCCAGTCCTGCTCGCTTCGC	5/75	308–316

[a] DNA sequences of hotspots for Tn*5* insertion in pBR322 were determined as described in the text with reference to the known sequence of pBR322. Only one of the two complementary DNA strands is shown (extending 5′ to 3′, left to right). The central underlined nine base pairs of the displayed sequences of pBR322 were duplicated by Tn*5* insertion.

[b] "pBR322 position" refers to the 9 bp of pBR322 sequence duplicated by Tn*5* insertion.

but not an actual requirement, for cutting at GC base pairs during Tn*5* insertion.

A cutting preference for GC base pairs had been proposed earlier to guide IS*1* and Tn*9* insertion: 20 of 26 insertion sites, including five of six used

repeatedly contain GC base pairs at positions 1 and 9 of the target sequence duplications (GALAS, CALOS and MILLER 1980). Tn*10*, in contrast, exhibits no such cutting preference (HALLING and KLECKNER 1982).

Since there are hundreds of 9-bp segments in pBR322 which begin and end with GC base pairs, it is tempting to consider other features of the DNA sequence at or near sites of frequent Tn*5* insertion which might also contribute to their serving as hotspots. With Tn*10*, for which the choice of potential sites is much more restricted than with Tn*5*, a symmetrical consensus sequence, positions 2 to 4 and 6 to 8 in the 9-bp duplication, has been proposed as a major determinant of insertion specificity (HALLING and KLECKNER 1982). Inspection of the sequences in Table 2 indicates that no such consensus sequence is likely to guide Tn*5* to its preferred hotspots.

Other sequence features that might affect the distribution of insertions can be sought. For example, overall base sequence could be important. AT richness and palindromic sequences facilitate transient opening of the DNA helix *in vivo*. Tn*9* and its component IS*1* elements seem to insert preferentially into AT-rich regions (despite their preference for GC termini) (GALAS, CALOS and MILLER 1980), as does Tn*3* (TU and COHEN 1980), an unrelated element which makes target sequence duplications of only 5 bp. In contrast, some of the hotspots for Tn*5* insertion are GC rich (Table 2) and seem not to coincide with the sites of highest AT content (see Figure 4) or palindromic sequences (LILLEY 1980). Homology to sequences at the termini of Tn*5* might be considered important, permitting base-pairing interactions between Tn*5* and target sequences to stabilize a transposition intermediate, or selection of the target by the transposase protein. For example, near one preferred region of Tn*3* insertion is a sequence in which 15 of 17 base pairs of target sequence match 15 of 18 base pairs at the Tn*3* termini (TU and COHEN 1980). Similarly, a preferred region for Tn*9* insertion contains a perfect match to 7 bp at the ends of Tn*9* (GALAS, CALOS and MILLER 1980). An inspection of the sequences of three insertions of Tn*5* in the leader region of the histidine operon and two in a nonessential region of phage fd led BOSSI and CIAMPI (1981) to suggest that Tn*5*-target homology is also a determinant of insertion specificity. They had picked out a 12-bp segment in Tn*5* extending from 15 to 26 bp from its ends, which resembled sequences found near these five sites of insertion (matches of eight to ten of 12 base pairs). When we scanned the DNA sequence of pBR322 (SUTCLIFFE 1978) we found 19 segments that exhibit an eight or nine of 12-bp match to the 12-bp Tn*5* segment focused on by BOSSI and CIAMPI. However, because only one of these 19 segments (pBR322 sequences 318→329) is near sites of frequent Tn*5* insertion (hotspots IV and V, Figure 4) matches between the DNA sequence at the insertion site and this region of Tn*5* are unlikely to play an important role in target selection.

The ends of each IS*50* component of Tn*5* consist of a hyphenated (8/9 bp) inverted repeat (BERG *et al.* 1982a), although recent mutational analyses of Tn*5*'s ends indicates that the actual recognition sequence extends about 19 bp into Tn*5* (SASAKAWA, CARLE and BERG 1983). Even though no sequences in pBR322 closely match Tn*5*'s ends, 15 segments exhibit a 7- of 9-bp match, three exhibit a 12- of 19-bp match and nine exhibit an 11- of 19-bp match.

Of all of these possible insertion specificity determinants, only one, an 11- of 19-bp match at pBR322 positions 59→41, is near a hotspot. We conclude that the distribution of these matches to Tn*5*'s ends *vis a vis* the distribution of Tn*5* insertion hotspots provide uncertain support, at best, for the concept that Tn*5*-pBR322 sequence similarities play a role in target site selection. The occasional matches may, in fact, be quite fortuitous.

Transposase may still participate in insertion site selection, even though there is no special sequence similarity between the Tn*5* insertion hotspots in pBR322 and the ends of Tn*5*. In our model for transposase structure there are two separate DNA-binding domains. The domain responsible for recognition of Tn*5* sequences is buried within the protein complex, whereas the domain responsible for target recognition is on the outside (BERG *et al.* 1982b). Because these two domains are likely to be different in structure, the DNAs they recognize need not, *a priori*, be similar in sequence.

The various transposable elements that now coexist in the Gram-negative bacterial gene pool appear to choose their insertion sites using several different sets of rules or DNA sequence signals. This is consistent with their heterogeneity in other fundamental characteristics: DNA sequence and functional organization, the regulation of transposition and the mechanism of transposition (see BERG 1983; BERG and BERG 1983; KLECKNER 1981). Such patterns of diversity could reflect either selective pressures favoring divergent evolution from common ancestors or convergent evolution from different immobile ancestral genes.

We are grateful to M. BITTNER and G. CARLE for help with computer searches of pBR322 sequences and to C. M. BERG, G. CARLE, H. HUANG and C. SASAKAWA for critical insights and suggestions. This work was supported by United States Public Health Service research grants AI 14267 and AI 18980 and American Cancer Society postdoctoral fellowship PF-2089 to J. L.

LITERATURE CITED

AUERSWALD, E., G. LUDWIG and H. SCHALLER, 1980 Structural analysis of Tn*5*. Cold Spring Harbor Symp. Quant. Biol. **45:** 107–113.

BERG, C. M. and D. E. BERG, 1981 Transposons as tools in microbial genetics. pp. 107–116. In: *Microbiology 1981*, Edited by D. SCHLESSINGER. American Society for Microbiology, Washington, D.C.

BERG, D. E., 1977 Insertion and excision of the transposable kanamycin resistance determinant Tn*5*. pp. 205–212. In: *DNA Insertion Elements, Plasmids, and Episomes*, Edited by A. I. BUKHARI, J. A. SHAPIRO and S. L. ADHYA. Cold Spring Harbor Press, Cold Spring Harbor, New York.

BERG, D. E., 1983 Structural requirement for IS*50*-mediated gene transposition. Proc. Natl. Acad. Sci. USA **80:** 792–796.

BERG, D. E. and C. M. BERG, 1983 The prokaryotic transposable element Tn*5*. Biotechnology. **1:** 417–435.

BERG, D. E., J. DAVIES, B. ALLET and J.-D. ROCHAIX, 1975 Transposition of R factor genes to bacteriophage λ. Proc. Natl. Acad. Sci. USA **72:** 3628–3632.

BERG, D. E. and M. DRUMMOND, 1978 Absence of DNA sequences homologous to transposable element Tn*5* (Kan) in the chromosome of *Escherichia coli* K-12. J. Bacteriol. **136:** 419–422.

BERG, D. E., L. JOHNSRUD, L. MCDIVITT, R. RAMABHADRAN and B. J. HIRSCHEL, 1982a The

inverted repeats of Tn*5* are transposable elements. Proc. Natl. Acad. Sci. USA **79:** 2632–2635.

BERG, D. E., J. B. LOWE, C. SASAKAWA and L. MCDIVITT, 1982b The mechanism and control of Tn*5* transposition. In: *Fourteenth Stadler Genetics Symposium*, Edited by G. REDEI, University of Missouri Press, Columbia, Missouri. In press.

BERG, D. E., A. WEISS and L. CROSSLAND, 1980 Polarity of Tn*5* insertion mutations in *Escherichia coli.* J. Bacteriol. **142:** 439–446.

BOSSI, L. and M. S. CIAMPI, 1981 DNA sequences at the sites of three insertions of the transposable element Tn*5* in the histidine operon of *Salmonella.* Mol. Gen. Genet. **183:** 406–408.

CASADABAN, M. J. and S. N. COHEN, 1980 Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli.* J. Mol. Biol. **138:** 179–207.

COLLINS, J., G. VOLCKAERT and P. NEVERS, 1982 Precise and nearly-precise excision of the symmetrical inverted repeats of Tn*5*: common features of *recA*-independent deletion events in *Escherichia coli.* Gene **19:** 139–146.

ENGLER, J. A. and M. P. VAN BREE, 1981 The nucleotide sequence and protein-coding capacity of the transposable element IS*5*. Gene **14:** 155–163.

GALAS, D. J., M. P. CALOS and J. H. MILLER, 1980 Sequence analysis of Tn*9* insertions in the *lacZ* gene. J. Mol. Biol. **144:** 19–41.

HALLING, S. M. and N. KLECKNER, 1982 A symmetrical six-base-pair target sequence determines Tn*10* insertion specificity. Cell **28:** 155–163.

JORGENSEN, S. J., ROTHSTEIN and W. S. REZNIKOFF, 1979 A restriction enzyme cleavage map of Tn*5* and location of a region encoding neomycin resistance. Mol. Gen. Genet. **177:** 65–72.

KLAER, R., S. KUHN, E. TILLMAN, H.-J. FRITZ and P. STARLINGER, 1981 The sequence of IS*4*. Mol. Gen. Genet. **181:** 169–175.

KLECKNER, N., 1981 Transposable elements in prokaryotes. Annu. Rev. Genet. **15:** 341–404.

LICHTENSTEIN, C. and S. BRENNER, 1982 Unique insertion site of Tn*7* in the *E. coli* chromosome. Nature **297:** 601–603.

LILLEY, D. M. J., 1980 The inverted repeat as a recognizable structural feature in supercoiled DNA molecules. Proc. Natl. Acad. Sci. USA **77:** 6468–6472.

LOWE, J. B. and D. E. BERG, 1983 A product of the transposase gene of Tn*5* inhibits Tn*5* transposition. Genetics **103:** 581–592.

MANIATIS, T., E. F. FRITSCH and J. SAMBROOK, 1982 *Molecular Cloning.* Cold Spring Harbor Press, Cold Spring Harbor, New York.

MAXAM, A. and W. GILBERT, 1980 Sequencing end-labeled DNA with base-specific hemical cleavages. Methods Enzymol. **65:** 499–580.

MILLER, J. H., M. P. CALOS, D. GALAS, M. HOFER, D. E. BUCHEL and B. MULLER-HILL, 1980 Genetic analysis of transpositions in the *lac* region of *Escherichia coli.* J. Mol. Biol. **144:** 1–18.

PEDEN, K. W. C., 1983 Revised sequence of the tetracycline resistance gene of pBR322. Gene **22:** 277–280.

ROBERTS, R. J., 1981 Restriction and modification enzymes and their recognition sequences. Nucleic Acids Res. **9:** r75–r96.

ROTHSTEIN, S. J., R. A. JORGENSEN, K. POSTLE and W. S. REZNIKOFF, 1980 The inverted repeats of Tn*5* are functionally different. Cell **19:** 795–805.

ROTHSTEIN, S. J., and W. S. REZNIKOFF, 1981 The functional differences in the inverted repeats of Tn*5* are caused by single base pair nonhomology. Cell **23:** 191–199.

SAINT GIRONS, I., H.-J. FRITZ, C. SHAW, E. TILLMAN and P. STARLINGER, 1981 Integration

specificity of an artificial kanamycin transposon constructed by the *in vitro* insertion of an internal Tn*5* fragment into IS*2*. Mol. Gen. Genet. **183:** 45–50.

SASAKAWA, C., G. F. CARLE and D. E. BERG, 1983 Sequences essential for transposition at the termini of IS*50*. Proc. Natl. Acad. Sci. USA. In press.

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

SCHALLER, H., 1978 The intergenic regions and origins for filamentous phage DNA replication. Cold Spring Harbor Symp. Quant. Biol. **43:** 401–408.

SENSTAG, C. and W. ARBER, 1983 IS*2* insertion is a major cause of spontaneous mutagenesis of the bacteriophage P1: nonrandon distribution of target sites. EMBO J. **2:** 67–71.

SHAPIRO, J. A. (Editor), 1983 *Mobile Genetic Elements*. Academic Press, New York.

SHAW, K. J. and C. M. BERG, 1979 *Escherichia coli* K-12 auxotrophs induced by insertion of the transposable element Tn*5*. Genetics **92:** 741–747.

SUTCLIFFE, J. G., 1978 Complete nucleotide sequence of the *Escherichia coli* plasmid pBR322. Cold Spring Harbor Symp. Quant. Biol. **43:** 77–90.

TU, C. P. D. and S. N. COHEN, 1980 Translocation specificity of the Tn*3* element: characterization of sites of multiple insertions. Cell **19:** 151–160.

Corresponding editor: G. MOSIG

Note added in proof: We have generated a GC to AT transition at pBR322 position 31, which corresponds to the first of the 9 base pairs duplicated by Tn*5* insertion at hotspot I. None of 60 *tet*::Tn*5* derivatives of this mutated plasmid contained Tn*5* at hotspot I, in contrast to 21 of 75 with pBR322-wild type (J. Lodge and D. E. Berg, unpublished results). This outcome supports the hypothesis that a cutting preference for GC base pairs helps guide Tn*5* to preferred insertion sites.