Transposon Tn5 Target Specificity: Preference for Insertion at G/C Pairs

Jennifer K. Lodge, Kathleen Weston-Hafer and Douglas E. Berg

Departments of Microbiology and Immunology and of Genetics, Washington University Medical School, St. Louis, Missouri 63110
Manuscript received December 18, 1987
Revised copy accepted July 28, 1988

ABSTRACT

The procaryotic transposon Tn5 inserts into many different sites within a single gene, but some sites (hotspots) are targeted repeatedly. Hotspots are not closely related in sequence, but most have G/C pairs at the ends of the nine base pairs duplicated by Tn5 insertion. In pBR322, the major hotspot coincides with the "-10 region" of the tet promoter. We mutated the G/C pairs at this hotspot and assayed for insertion into hotspot I, resistance to tetracycline, and plasmid supercoiling. We found that changing the G/C pairs to A/T pairs reduced the frequency of insertion into the hotspot by at least fivefold. The reduction in hotspot use caused by these G/C to A/T changes was not attributable to changes in plasmid supercoiling or tet promoter strength.

TRANSPORTABLE elements (transposons) are DNA segments that can move to new locations in a genome without the need for extensive sequence homology between donor and target DNAs. Many different elements have been identified and all show some preference for insertion at particular sites (KLAER et al. 1980; Tu and COHEN 1980; HALLING and KLECKNER 1982; BERG, SCHMANDT and LOWE 1983; ZERBIB et al. 1985; SENGSTAG and ARBER 1987; WADDELL and CRAIG 1988). Transposons generally contain a transposase gene, whose protein product, in concert with host factors, is necessary for movement of the element (for example, see CRAIGIE, ARNDT-JOVIN and MIZUUCHI 1985; MORISATO and KLECKNER 1987; PHADNIS and BERG 1987; YIN and REZNKOFF 1987). These transposon-specific proteins probably act at the ends of the element (CRAIGIE and MIZUUCHI 1985; MCKOWN et al. 1987; MORISATO and KLECKNER 1987) and may participate in target site selection as well. Host factors have also been postulated to play a role in the insertion specificity of certain elements, most notably integration host factor for IS1 (GAMAS et al. 1987).

A short segment of target DNA is duplicated during transposition, probably the result of staggered nicks made in the target DNA by transposition proteins and repair of the gaps by DNA polymerase I (SASAKAWA, UNO and YOSHIKAWA 1981; SYVANEN, HOPKINS and CLEMENTS 1982). The present study tests the idea that base pairs at the postulated cutting site are important in the insertion specificity of the procaryotic kanamycin-resistance transposon Tn5.

Tn5 (for review, see BERG et al. 1988) inserts into many dissimilar sites in a single gene, but some sites (hotspots) are chosen repeatedly. For example, in one study, one-fourth of all insertions into the tet gene of pBR322 were at a single site (hotspot I) and another one-fourth were distributed among four other hotspots (BERG, SCHMANDT and LOWE 1983). Tn5 generates nine bp direct repeats of target sequence during transposition, and the duplications made at known hotspots are shown in Table 1. The only obvious feature common to these hotspots are G-C or C-G base pairs (referred to hereafter as G/C pairs) at each end of the target duplication. Seven of nine hotspots have G/C pairs at both ends and the remainder (2 of 9) have a G/C pair at only one end. None have A/T pairs at both ends. A weaker yet significant preference for G/C pairs is also seen in insertion sites that are not known to be hotspots (BOSSI and GIAMPETTI 1981; COLLINS, VOLCKAERT and NEVERS 1982; Egelhoff et al. 1985; LUPSKI et al. 1984, 1986; MCKINNON et al. 1985; SCHALLER 1978; SCHOLFIELD and WATSON 1986). Nearly half (19 of 43) of these sites have G/C pairs at both ends, most others (22 of 43) have a G/C pair at one end and an A/T pair at the other, and only a few (2 of 43) have A/T pairs at both ends. The preference for G/C pairs, especially at known hotspots, suggests that they participate in target site selection.

The present study tests the importance of G/C pairs at the major hotspot in pBR322 by analyzing the effects of single base pair changes at each end of the hotspot I sequence in pBR322 on Tn5 insertion into the mutant plasmids.

MATERIALS AND METHODS

Strains: Bacterial strains are derivatives of Escherichia coli K-12. DB1572 is a lacZ124::Tn5 recA rpsL. tna Drtpe3 trpR derivative of W3110 (BACHMANN 1972). MC1061 is F araD139 (ara-leu)7697 ΔlacX74 galU galK hisD179 rpsL16. (CASADABAN and COHEN 1980) obtained from H. V. Huang. BD2062 is a tyrA sfg2008::Tn10 ung′1 nadB7 derivative of AB1157 (DeWitt and Adelberg 1962) obtained from B. K. DUNCAN. DM800 (Sternglanz et al. 1981) is a Δ(top-
The product was cleaved with HindIII and BamHI and position 31, the transformants were pooled and plasmid DNA was extracted, cleaved with HindIII and transformed into MC1061. Plasmid DNAs from individual isolates were restricted with HindIII, filling in the 5' extensions using the large fragment of DNA polymerase I (ZOLLER and SMITH 1982). Mutants were identified by colony hybridization and DNA sequencing.

<table>
<thead>
<tr>
<th>Target duplication</th>
<th>Designation Frequency Gene</th>
<th>Target size (kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCTTTAATG&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Hotsp I 40/150 tet&lt;sup&gt;b&lt;/sup&gt;</td>
<td>1.2</td>
</tr>
<tr>
<td>GTCAAGGCA&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Hotsp II 17/150 tet&lt;sup&gt;b&lt;/sup&gt;</td>
<td>1.2</td>
</tr>
<tr>
<td>CCTCTGATG&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Hotsp III 4/75 tet</td>
<td>1.2</td>
</tr>
<tr>
<td>CGCCAGTGC&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Hotsp IV 7/75 tet</td>
<td>1.2</td>
</tr>
<tr>
<td>GTCTCTGTC&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Hotsp V 5/75 tet</td>
<td>1.2</td>
</tr>
<tr>
<td>GGGTTGATGC&lt;sup&gt;a&lt;/sup&gt;</td>
<td>79 2/8 CS&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.28</td>
</tr>
<tr>
<td>GCAAGGGG&lt;sup&gt;a&lt;/sup&gt;</td>
<td>8 2/8 CS</td>
<td>0.28</td>
</tr>
<tr>
<td>GCCTCGGG&lt;sup&gt;a&lt;/sup&gt;</td>
<td>2B2 2/14 nodABD&lt;sup&gt;c&lt;/sup&gt;</td>
<td>3.0</td>
</tr>
<tr>
<td>GCCCGAGCT&lt;sup&gt;a&lt;/sup&gt;</td>
<td>245 2/15 nodABCDEF&lt;sup&gt;c&lt;/sup&gt;</td>
<td>3.7</td>
</tr>
</tbody>
</table>

<sup>a</sup>The 9 bp duplicated by Tn5 insertion are shown. The exceptions to G/C pairs at the first and ninth base pair are underlined.

<sup>b</sup>From BERG, SCHMANDT and LOWE (1985).

<sup>c</sup>From BERG, SCHMANDT and LOWE (1985).

General procedures: Standard techniques were used for bacterial transformation, plasmid DNA isolation and restriction mapping (MANIATIS, FRITSCH and SAMBROOK 1982). DNA sequencing was performed using either the MAXAM-GILBERT (1980) or the dideoxy (SANGER, NICKLEN and GILBERT (1980) method as modified for double stranded DNA (ZAGURSKI et al. 1985). Oligonucleotides were made either by the phosphite method (CROCKETT 1985) or on an Applied Biosystems 380A DNA synthesizer.

Plasmid construction: All plasmids made for this study are derivatives of pBR322 (SUTCLIFFE 1978; FEDEN 1985). pBR322-31A was made by bisulfite mutagenesis (Figure 1a) which preferentially deaminates cytosines in single stranded DNA, converting them to uracils. pBR322 DNA was cleaved with HindIII, treated with sodium bisulfite stranded DNA, converting them to uracils. pBR322 DNA was cleaved with HindIII and transformed into the complementary HindIII site of pBR322. Mutants were identified by site-directed mutagenesis as described by MANDECKI et al. (1984). Insertions of Tn5 in the chromosome of R. STERNGLANZ.

Isolation and mapping of Tn5 insertions: Transpositions of Tn5 from a site in the lacI locus of DB1572 in the tet gene in pBR322 were obtained by plating cells on LN plates (10 g/liter NaCl, 10 g/liter N-Z-Amine, 5 g/liter yeast extract, 15 g/liter Bacto-agar, 4.8 mM NaOH) containing 250 μg/ml neomycin, which enriches for multiple copies of Tn5. Colonies from each plate were pooled, plasmid DNA extracted and transformed into MC1061. To enrich for mutations at or near hotspot I were initially identified by a characteristic ~550 bp band on an agarose gel, and analyzed further by end-labeling plasmid DNA at either the Clal or EcoRI sites, subcutting with HindIII which cleaves 486 bp from the ends of Tn5) followed by electrophoresis in agarose gels. Inserts at or near hotspot I were identified by a characteristic ~550 bp band on an agarose gel, and analyzed further by end-labeling plasmid DNA at either the Clal or EcoRI sites, subcutting with HindIII which cleaves 6 bases from the end of Tn5), and electrophoresing on a polyacrylamide-urea sequencing gel. This strategy allows the mapping of insertions with single base pair resolution.

Assays of tetracycline resistance: Single colonies of DB1572 containing each plasmid were grown overnight in LN broth with no antibiotics. One-hundredth volume was inoculated into fresh medium and grown until the OD<sub>600</sub> nm was ~0.4. Ten microliters were added to 100 μl of broth containing different concentrations of tetracycline. OD<sub>620</sub> nm was measured after 7 hr of growth.

**TABLE 1**

<table>
<thead>
<tr>
<th>K10</th>
<th>Stbl3</th>
<th>Stbl3</th>
<th>Stbl3</th>
<th>Tet</th>
<th>Tet</th>
<th>Tet</th>
<th>Cm</th>
<th>Cm</th>
<th>Cm</th>
</tr>
</thead>
<tbody>
<tr>
<td>245</td>
<td>50</td>
<td>0.28</td>
<td>50</td>
<td>218</td>
<td>986</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 1.** Mutagenesis of pBR322 positions 31 and 39. (a) Bisulfite mutagenesis of position 31. Of 10 mutants sequenced, each had position 31 changed from G-C to A-T. (b) Oligonucleotide mutagenesis of position 39. The dashed line represents in vitro DNA synthesis.
growth at 37°C in a Titer Tek plate reader. Nine replicas of each plasmid carrying strain were assayed and the values were averaged.

**Analysis of plasmid supercoiling:** Fresh overnight cultures of the top strain DM800 carrying the different mutant plasmids were diluted 1:250 into 50 ml of LN broth and grown for 4 hr until OD_{600 nm} was ~0.5. The cells were pelleted, resuspended in 1 ml broth, transferred to an Eppendorf tube, and repelleted. They were resuspended in 200 μl STET buffer (5% Triton, 8% sucrose, 50 mM Tris pH 8, 50 mM EDTA) containing 1 mg/ml lysozyme, placed in a boiling water bath for 40 sec and centrifuged for 15 min at 4°C. The pellet of cell debris was removed and 200 μl isopropanol was added to the supernatant which was then centrifuged for 2 min. The precipitate was rinsed, dried, and resuspended in 100 μl TE (10 mM Tris, pH 7.5, 1 mM EDTA). One-tenth of the sample was electrophoresed in a 1% agarose gel containing 12 μg/ml chloroquine for 12 hr at 3 V/cm (Pruss 1985).

**RESULTS AND DISCUSSION**

To test whether G/C pairs at the ends of a hotspot are important, we changed the terminal G/C pairs of hotspot I of pBR322 (positions 31 and 39) and used these mutant plasmids as targets for Tn5 transposition. Over 60 independent insertions of Tn5 into the tet gene of each mutant plasmid were mapped to determine the frequencies of insertion into hotspot I. The G/C to A/T substitutions at positions 31 and 39 each dramatically reduced the ability of Tn5 to insert into hotspot I. Figure 2 shows that the frequency of Tet\(^+\) insertions that were in hotspot I was reduced from 26% with wild-type pBR322 to between 0% and 5% with the G/C to A/T mutants. No new hotspots close to hotspot I were found. When the G-C at positions 31 or 39 was changed to a C-G (reversing the strand orientation of the G/C pair) the frequencies of insertion into hotspot I were reduced, but by less than twofold.

Hotspot I is in a complex region of the plasmid. It coincides with the "−10 region" of the tet promoter and contains the transcription start site of the antitet promoter (Brosius, Cate and Perlmutter 1982). It has been shown that mutations in the tet promoter affect plasmid supercoiling (Pruss and Drilca 1986), and a supercoiled target is thought to be important for efficient insertion of Tn5 (Isberg and Syvanen 1982). The single base changes at positions 31 and 39 are outside the consensus (TATAAT) for the "−10 region," but it had seemed possible that they could affect the function of the tet promoter, thus affecting insertion into hotspot I only indirectly. Similarly, it also seemed possible that the point mutations changed Tn5 insertion indirectly by changing plasmid supercoiling. To test these possibilities, we assayed tetracycline resistance (as a measure of promoter strength) and plasmid supercoiling of the mutant plasmids.

Figure 3 shows the tetracycline resistance of Db1572 containing each of the mutant plasmids. Although each point mutation caused a decrease in resistance to tetracycline, the level of resistance was the same among the three plasmids with mutations at 31 and between the two plasmids with mutations at 39. So even though a G to C change gave the same decrease of resistance to tetracycline as a change to A or T, the change to A or T caused a much greater effect on insertion into the hotspot. It is possible that the reduced promoter activity may cause a twofold overall suppression of insertion into hotspot I, but this cannot account for the large reductions found with mutants 31-A, 31-T and 39-T.

Certain mutations in the tet gene alter plasmid DNA supercoiling, as monitored in bacteria lacking DNA topoisomerase I, the enzyme that relaxes negative supercoils (Pruss and Drilca 1986). The supercoiling effects seen in the top bacterias probably reflect the different DNA topologies present, but not observable, in normal top\(^+\) bacteria. We examined the supercoiling of the five mutant plasmids (Figure 4). Each of the three point mutations at position 31 reduced supercoiling to the same extent, whereas neither of the point mutations at position 39 affected supercoiling appreciably. This result indicates that the specific
reductions in hotspot I use in mutants 31A, 31T and 39T are not attributable to the effects of these mutations on target DNA supercoiling.

These results support the hypothesis, formed by scanning sequences of sites of Tn5 insertion, that G/C pairs are important for efficient selection of target sites by Tn5. The dramatic decrease of insertion into hotspot I in plasmids with the G/C to A/T mutations results from the base pair changes themselves and not from the secondary effects on promoter function or plasmid supercoiling.

G/C pairs cannot be absolutely essential for Tn5 insertion. Many Tn5 insertion sites contain A/T pairs at one end of the duplicated sequence and a few have A/T pairs at both ends. Also, two of the mutants described here, in which T-A replaced G-C at the first or ninth base pair of hotspot I, reduced but did not eliminate insertion into that site. Several possible mechanisms could account for the frequent presence of G/C pairs at the ends of the target duplication. Transposase might bind preferentially at the sequence G/CNNNNNNNG/C. Secondly, it might cleave most efficiently at G/C pairs nine bp apart (Galas, Calos and Miller 1980). Finally, transposase might bind and cleave potential targets indiscriminately, but insertion might often be abortive at sites not terminating in G/C pairs. In each case the net effect would be equivalent: the transpositions recovered would preferentially have G/C at the first and ninth base pairs.

Other features of target DNAs must also influence site selection. There are, for example, hundreds of sets of G/C pairs spaced nine bp apart in pBR322 that are not used as hotspots of Tn5 insertion. It is likely that DNA topology or transcription of the tet promoter are additional factors in making hotspot I a preferred target, because mutations outside the nine base pairs of hotspot I which affect the promoter also affect insertion into the hotspot (J. K. Lodge and D. E. Berg, unpublished data). We are currently studying the contributions that sequences adjacent to hotspot I make to its repeated use as an insertion site of Tn5.

We are grateful to J. B. Lowe for participating in initial experiments, to T. Kazic, J. Majors and J. Vovvodic for helpful comments on the manuscript and M. Michener for assistance with the figures. This work was supported by grants from the Monsanto Company, the U.S. Health Service (GM-37138) and the National Science Foundation (DMB-8608193) to D.E.B. J.K.L. was sup-
ported by a Graduate and Professional Opportunities Program Fellowship and K.W.H. was supported by the Division of Biology and Biomedical Sciences at Washington University.

LITERATURE CITED


Communicating editor: G. Mosig