Local DNA Sequence Control of Deletion Formation in *Escherichia coli* Plasmid pBR322

Ujjala DasGupta,[1] 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 March 31, 1986
Revised copy accepted October 7, 1986

ABSTRACT

The specificity of deletion formation was studied using tests involving reversion of palindromic insertion mutations. Insertions of a Tn*5*-related transposon at 13 sites in the ampicillin-resistance (*amp*) gene of plasmid pBR322 were shortened to a nested set of perfect palindromes, 22, 32 and 90 bp long. We monitored frequencies of reversion to Amp^r, which is the result of deletion of the palindrome plus one copy of the flanking 9 bp direct repeats (which had been formed by transposition). Revertant frequencies were found to depend on the location and the sequence of the palindromic insert. Changing a 45-kb interrupted palindrome to a 22-bp perfect palindrome stimulated deletion formation by factors of from fourfold to 545-fold among the 13 sites, while elongation of the perfect palindrome from 22 to 90 bp stimulated deletion formation by factors of from eight- to 18,000-fold. We conclude that deletion formation is strongly affected by subtle features of DNA sequence or conformation, both inside and outside the deleted segment, and that these effects may reflect specific interactions of DNA processing proteins with template DNAs.

DELETIONS are frequent among both spontaneous and induced mutations. Insights into mechanisms by which they arise have come from the mapping and sequencing of frameshift mutations and deletions, and from studies of transposon excision.

In studies that preceded the advent of direct DNA sequencing it was deduced that frameshift mutations occur preferentially in repeated DNA sequences. It was proposed that they arise by replication errors in which misaligned pairing of complementary template and nascent DNA strand sequences result in the failure to copy, or the repeated copying of, one or a few bases of the template (STREISINGER *et al.* 1966; STREISINGER and OWEN 1985). Analyses of forward mutations in an F′ episome indicated that deletions less than 200 bp long are formed preferentially, and that the endpoints of these and also larger (>700 bp) deletions usually coincide with direct repeats of 4 bp or more. These data were also explained on the basis of replication errors due to slipped mispairing of complementary sequences analogous to the frameshift mispairing model (ALBERTINI *et al.* 1982; FARABAUGH *et al.* 1978; GALAS 1978). The endpoints of deletions in other replicons including small multicopy plasmids also generally coincide with direct repeats (COLLINS, VOLCKAERT and NEVERS 1982; JONES, PRIMROSE and EHRLICH 1982; LOPEZ *et al.* 1984). Short, weakly matched, inverted repeats were also found near certain deletion endpoints. This led to a suggestion that palindromes are also deletionogenic (ALBERTINI *et al.* 1982; GLICKMAN and RIPLEY 1984; SCHAAPER, DANFORTH and GLICKMAN 1986). This putative intrinsic instability of short palindromes is surprising in light of the importance of palindromes in regulatory sites and stable (transfer and ribosomal) RNAs (PLATT 1981; NOLLER 1984; CALENDAR and GOLD 1985). Other studies showed that sites nicked or cleaved by proteins such as DNA gyrase or the M13 gene 2 product are deletion hotspots (MARVO, KING and JASKUNAS 1983; IKEDA, KAWASAKI and GELLERT 1984; IKEDA 1986; MICHEL and EHRLICH 1986). It is likely therefore, that deletions arise by several different mechanisms, some involving DNA replication, and others DNA breakage.

Useful information on deletion formation has also come from studies of the reversion of Tn*5*- and Tn*10*-induced insertion mutations, in which reversion results from loss of the transposon plus one copy of the flanking 9-bp direct repeats (precise excision). This excision was not correlated with transposition to new sites, nor dependent on element-specific transposition functions, and thus was considered a type of spontaneous deletion event (EGNER and BERG 1981; FOSTER *et al.* 1981; NAG *et al.* 1985; J. RIKKERS and D. E. BERG unpublished data). In other tests, shortening the approximately 1.5-kb long inverted repeats of Tn*5* or Tn*10* or reversing them (to give direct repeats) markedly decreased reversion frequencies, thus indicating that long inverted repeats are deletionogenic (EGNER

Abbreviations: *amp*, *tet* and *kan*, genes encoding resistance to ampicillin, tetracycline and kanamycin, respectively; superscript r, resistance; superscript s, sensitivity; bp, base pair(s); kb, kilobase pair(s).

[1] Present address: Biophysics Division, Indian Institute of Chemical Biology, Calcutta, 700032, India.

and BERG 1981; FOSTER *et al.* 1981). The deletion of Tn*5* and Tn*10* is more frequent in F′ episomes than in the bacterial chromosome, and is due to F′ transfer among cells in the population (BERG, EGNER and LOWE 1983; SYVANEN *et al.* 1986). Because F′ transfer results in the formation of long single stranded regions, and earlier tests had shown Tn*5* to be extremely unstable in single-stranded fd phage (HERRMANN *et al.* 1978), we concluded that single strandedness as well as palindromy promotes the formation of large deletions, perhaps by replication errors as diagrammed in Figure 1A (EGNER and BERG 1981).

The simplest versions of models such as those in Figure 1 imply that direct and inverted repeats will be important determinants of deletion frequency, and of the location of deletion endpoints. However, the stabilities of Tn*5* insertion mutations also seemed to be strongly dependent on the insertion site (EGNER and BERG 1981; BERG, EGNER and LOWE 1983; NAG *et al.* 1985). This indicated that other features of DNA sequence, less distinctive than direct and inverted repeats, contribute to the formation of large deletions.

Tests based on the reversion of insertion mutations are analogous to the classical reversion analyses of point mutations (MARON and AMES 1983), and we have continued to use this approach to learn how the location and sequence of DNA segments affect their stabilities. In the present studies we focused on the formation of small deletions using palindromic inserts in the *amp* gene of plasmid pBR322. Our results indicate that frequency of deletion of a segment depends on its location and sequence, and is independent of *recA* function.

MATERIALS AND METHODS

General procedures: Bacterial growth, plasmid DNA extraction, restriction analysis and cloning, and DNA sequencing procedures have been described (BIRNBOIM and DOLY 1979; MANIATIS, FRITSCH and SAMBROOK 1982; MAXAM and GILBERT 1980; NAG *et al.* 1985).

Bacterial strains and plasmids: All bacterial strains used are derivatives of *Escherichia coli* K12. Strain MC1061 (from M. CASADABAN) was the host for most plasmid constructions and reversion tests; GM119 (*dam*$^-$) (ARRAJ and MARINUS 1983) was used to grow DNAs to be cleaved with the methylation-sensitive enzyme *Bcl*I. The isogenic strains AB1157 (*recA*$^+$) and JC2924 (*recA*56) (from A. J. CLARK via H. HUANG) were used to test the *recA*-independence of deletion events.

The plasmids used were pBR322 (Amp^r Tet^r) (MANIATIS, FRITSCH and SAMBROOK 1982) and the pBR322 derivative pGC1 (Amp^r but Tet^s) (MYERS, LERMAN and MANIATIS 1985). The sequence of the polylinker segment of our isolate of pGC1 (from R. MYERS) had three differences from the published sequence: four instead of three contiguous Gs in the *Xho*I-*Sal*I interval; five instead of four contiguous Cs in the *Bam*HI-*Bgl*II interval; and three instead of two contiguous Gs in the *Nco*I-*Eco*RI interval.

Plasmids with revertible insertion mutations: Mutations in the pBR322 *amp* gene were generated by NAG *et al.* (1985) using the Tn*5*-related inverse transposon diagrammed in Figure 2. This transposon, like Tn*5*, generates 9-bp direct repeats of target sequences during insertion; Amp^r revertants result from deletion of the entire insert plus one copy of the direct repeats of *amp* gene sequences. Tests of 34 insertion mutants revealed reversion frequencies ranging from 10^{-10} to 10^{-7} (NAG *et al.* 1985). Thirteen mutants including six representative of high and low reverting insertions were chosen for tests of the effect of position and palindrome length on deletion frequency. The pBR322 coordinates and sequences at the sites of insertion in *amp* are given in Table 1.

***In vitro* manipulation of insertion mutations.** While both Tn*5*-wild type and the Tn*5* based inverse transposon used to generate insertion mutations contain terminal inverted repeats of the 1534-bp IS*50* elements, the inverse transposon contains *Bcl*I and *Bgl*II restriction sites conveniently close to its ends (Figure 2). Plasmids containing uninterrupted 22- and 32-bp palindromes were generated by digestion with *Bcl*I and *Bgl*II, respectively (these enzymes do not cleave pBR322). An uninterrupted 90-bp palindrome was generated by the *in vitro* insertion of duplicate copies of a short segment from plasmid pGC1 into the central *Bgl*II site of the 32-bp palindrome (Figure 3A). The sequences of the palindromic inserts are given in Table 2.

The sequence of the 90-bp palindrome at A-10 was verified by DNA sequencing (MAXAM and GILBERT 1980) using DNA fragments end-labeled at the central *Xho*I site. The presence of the 90-bp insert at 11 of the other 12 sites was verified by electrophoresis of DNA fragments end-labeled at *Bgl*II sites within the inverted repeats as in NAG *et al.* (1985). The palindrome first constructed at site A-24 was found by end-labeling and DNA sequencing to contain an internal 5-bp deletion, and is hereafter designated A-24δ. A perfect 90-bp palindrome at this site was also constructed and verified.

Measurement of deletion frequencies. The frequencies of revertants observed in a population will reflect the rate at which reversion occurs, and also differences in growth rate that favor one plasmid or cell type over another (PHADNIS and BERG 1985). To assess the effects of selection we estimated reversion frequencies using fluctuation tests (LURIA and DELBRUCK 1943). For ease, young single colonies were used (Figure 4) instead of the traditional liquid cultures. Bacteria carrying the plasmids to be tested for reversion to Amp^r were streaked on an L-agar tetracycline plate and grown at 37° until the colonies were just visible (about 7 hr of growth) and contained about 5×10^5 viable cells per colony. The fraction of colonies which contained revertant subclones and relative revertant frequencies were estimated by streaking entire colonies on L-agar ampicillin plates (Figure 4). These control experiments showed that the quantitation of deletion frequency described below was not seriously biased by selection.

To quantitate the frequencies of deletion of an insert plus one copy of its direct repeats, the reversion of the Amp^s mutations to Amp^r was selected as follows: Tet^r Amp^s bacteria were streaked on L-agar tetracycline plates (12 μg/ml) and incubated overnight at 37°. Young single colonies were inoculated from these plates into 2 ml L-broth plus tetracycline, and grown to stationary phase (14 hr). Aliquots were then spread on L-agar ampicillin plates (250 μg/ml) and incubated at 37°. This represents about 32 generations of clonal growth from the single founding cell.

RESULTS

Reversion of insertion mutations: a model system to study deletion formation. Initial studies had shown

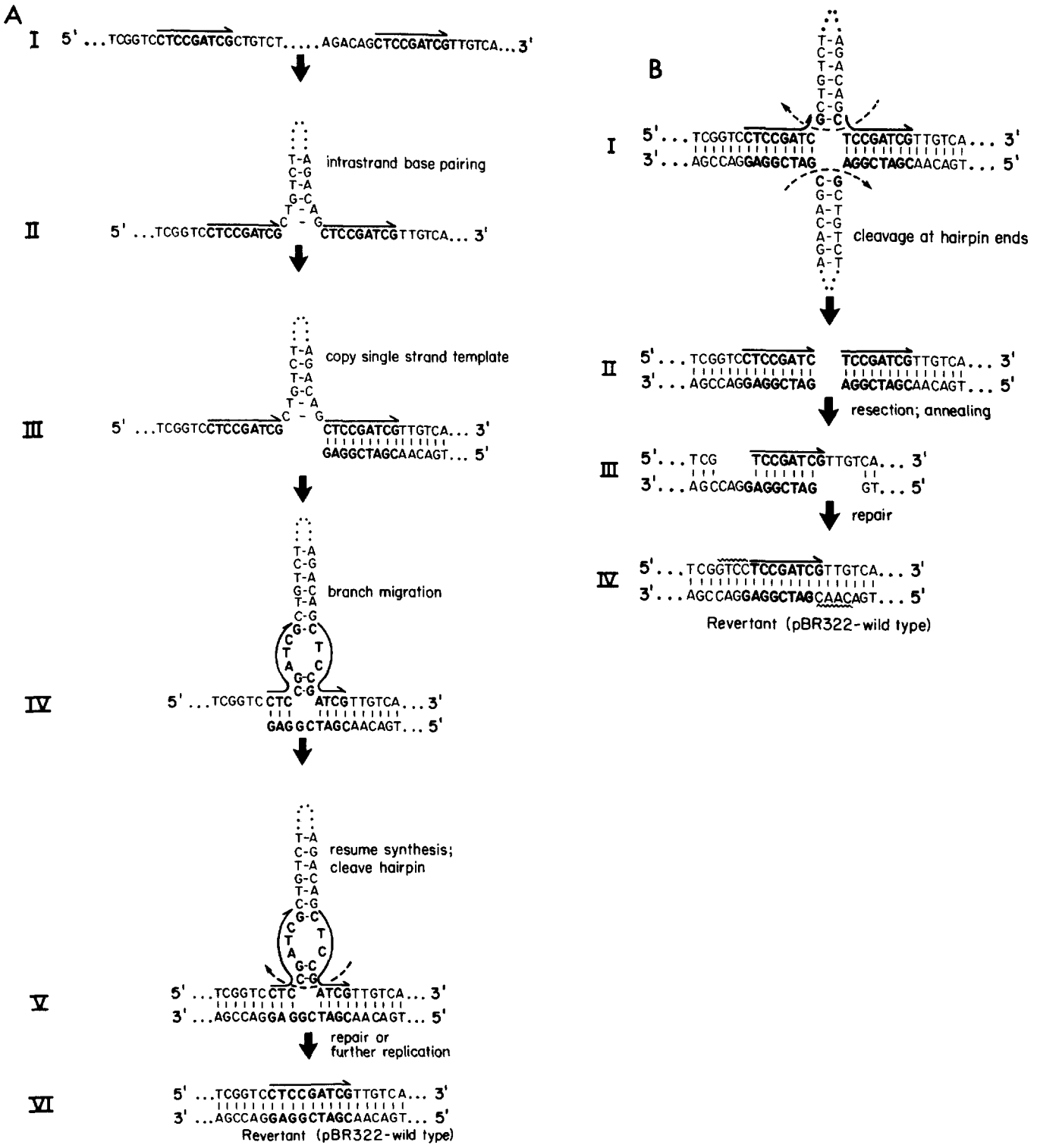

FIGURE 1.—Models of deletion formation. The specific versions drawn depict the deletion of a palindrome plus one copy of a pair of flanking direct repeats. The sequences shown are those at the deletion prone A-10 site, and the overlining indicates the nine bp direct repeats used as deletion endpoints (see Tables 1 and 2). A, Deletion as the result of an error in replication (adapted from STREISINGER *et al.* 1966; CHOW, DAVIDSON and BERG 1974; FARABAUGH *et al.* 1978; EGNER and BERG 1981; ALBERTINI *et al.* 1982; STREISINGER and OWEN 1985). (I) Single stranded regions are formed by replication, by exonucleolytic digestion, or by helix unwinding proteins such as those encoded by the genes *rep* or *recBC*. (II) Intramolecular base pairing generates hairpin structures in single stranded DNA. Hairpins may also be extruded from supercoiled double stranded DNA (see part B). (III) The presence of the hairpin may impede copying of the template strand. (IV) The slowing of DNA synthesis at the hairpin facilitates separation of the end of the nascent DNA strand from the first copy of the direct repeats and its chance reassociation with the second copy of the direct repeat in the template strand. (V) The resumption of synthesis in this alternate configuration stabilizes the mispairing and prevents subsequent copying of the extruded segment. The disordered conformation of the template may provoke cleavage by a conformation-specific nuclease (dashed arrow). (VI) Double stranded DNA with the sequence of pBR322 wild type is regenerated by repair synthesis after cleavage of the hairpin, or, in the absence of cleavage, by a second round of replication. B, Deletion as the result of DNA breakage without extensive replication (adapted from CHOW, DAVIDSON and BERG 1974; GLICKMAN and RIPLEY 1984). (I) A palindromic DNA sequence in covalently closed double stranded circular DNA is extruded into a cruciform structure, possibly in response to DNA supercoiling. (II) The product of cleavage by a conformation-specific nuclease operating at the cruciform ends. (III) The product of limited resection by a 3′ → 5′ exonuclease, and then annealing of complementary sequences from the nine bp direct repeats. (IV) The product of repair synthesis. Newly incorporated nucleotides are indicated by jagged lines.

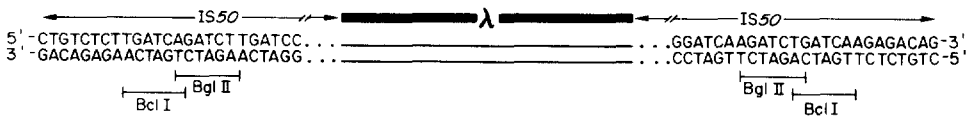

FIGURE 2.—Sequences at the termini of the 45-kb Tn*5*-related inverse transposon which was used to generate insertion mutations in the *amp* gene of pBR322. The inverted terminal repeats of this transposon are in the opposite sense to that found in the Tn*5*-wild type; its ends, here labeled I, are buried within Tn*5* (see NAG *et al.* 1985). λ indicates the approximately 42 kb of phage λ DNA between the inverted 1534-bp IS*50* elements. *Bgl*II and *Bcl*I designate restriction sites used in insertion mutant plasmid constructions.

TABLE 1

Sites of insertion in the *amp* gene of plasmid pBR322[a]

Site designation	pBR322 position[b]	Insertion mutant sequence[c]
A-40	3388–3396	· · TACGATACGGGAGctgt · · · acagATACGGGAGGGCT · ·
A-32	3409–3417	· · ATCTGGCCCCAGTctgt · · · acagGGCCCCAGTGCTG · ·
A-24	3411–3419	· · CTGGCCCCAGTGCctgt · · · acagCCCCAGTGCTGCA · ·
A-1	3416–3424	· · CCCAGTGCTGCAActgt · · · acagGTGCTGCAATGAT · ·
A-19	3418–3426	· · CAGTGCTGCAATGctgt · · · acagGCTGCAATGATAC · ·
A-21	3430–3438	· · GATACCGCGAGACctgt · · · acagCCGCGAGACCCAC · ·
A-18	3441–3449	· · ACCCACGCTCACCctgt · · · acagACGCTCACCGGCT · ·
A-22	3730–3738	· · CGGTCCTCCGATCctgt · · · acagCCTCCGATCGTTG · ·
A-10	3731–3739	· · GGTCCTCCGATCGctgt · · · acagCTCCGATCGTTGT · ·
A-41	3893–3901	· · AGTTGCTCTTGCCctgt · · · acagGCTCTTGCCCGGC · ·
A-37	3941–3949	· · AGAACTTTAAAAGctgt · · · acagCTTTAAAAGTGCT · ·
A-13	3995–4003	· · AGGATCTTACCGCctgt · · · acagTCTTACCGCTGTT · ·
A-15	4015–4023	· · ATCCAGTTCGATGctgt · · · acagAGTTCGATGTAAC · ·

[a] Insertion mutations in the *amp* gene were generated by *in vivo* transposition of the Tn*5*-related element diagrammed in Figure 2 and sequenced as described (NAG *et al.* 1985).

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

[c] Upper case, pBR322 sequence; lower case, insert sequence. The sequences of the 9 bp direct repeats of *amp* sequence are underlined.

that the reversion of insertion mutations induced by the Tn*5*-related inverse transposon diagrammed in Figure 2 is due to deletion of the inserted segment plus a copy of the 9-bp direct repeats formed during insertion, that the excision of such transposons is independent of Tn*5*-encoded functions, and that it is not correlated with movement to new sites (EGNER and BERG 1981; NAG *et al.* 1985). In experiments which led to the current work, the frequencies of reversion of mutations due to this 45-kb inverse transposon in the *amp* gene of pBR322 ranged from 10^{-10} to 10^{-7} (NAG *et al.* 1985).

To better characterize the formation of smaller deletions, we converted the 45-kb insert into a nested set of perfect palindromes of 22, 32 and 90 bp at the three least and the three most deletion-prone of the 34 sites studied previously (Figures 2 and 3; Table 2). The data in Table 3 show that conversion of the 45-kb interrupted palindrome to a 22-bp perfect palindrome increased the frequencies of Amp^r revertants (deletions) at each site tested. The extent of increase was site-dependent, ranging from about fourfold (site A-10) to about 450-fold (site A-13). The 32- and 90-bp perfect palindromes were generally more deletion prone, and the extent of increase was also site dependent: relative to the 22-bp palindrome, the increase in deletion frequency of the 32-bp palindrome ranged from negligible to 116-fold, and the 90-bp palindrome ranged from nine- to 18,000-fold.

An imperfect 85-bp palindrome, containing one 45-bp and one 40-bp inverted repeat (the result of deletion of 5 bp, Table 2) was found at one site. This variant, called A-24δ, was sevenfold less deletion prone than the perfect 90-bp palindrome at the same site (A-24 *vs.* A-24δ, Table 3). Thus, our data suggest that deletion frequencies increase with palindrome length, and that they are markedly reduced by small mismatches in a palindrome. Superimposed on these effects of palindromy, the range of reversion frequencies for identical palindromes at different sites makes it clear that sequences flanking the segment undergoing deletion are also major determinants of deletion frequency.

Accumulation of revertants reflects rates of deletion formation, not selection. Because an *amp*^{r} allele is dominant to an *amp*^{s} allele, the observed differences in revertant frequencies might have been attributable to differences in plasmid copy number. However, electrophoresis of extracts indicated that plasmid DNA levels are not affected by palindrome location or length in the range of 22 to 90 bp (data not shown). Additional control experiments indicated that the observed spectrum of revertant frequencies is not attributable to inadvertent growth rate selection prior to plating for Amp^r revertants: (i) The 90-bp palindrome at site A-10 had no detectable effect on bacterial growth monitored by either colony size or mixed growth in liquid culture in comparison to the same

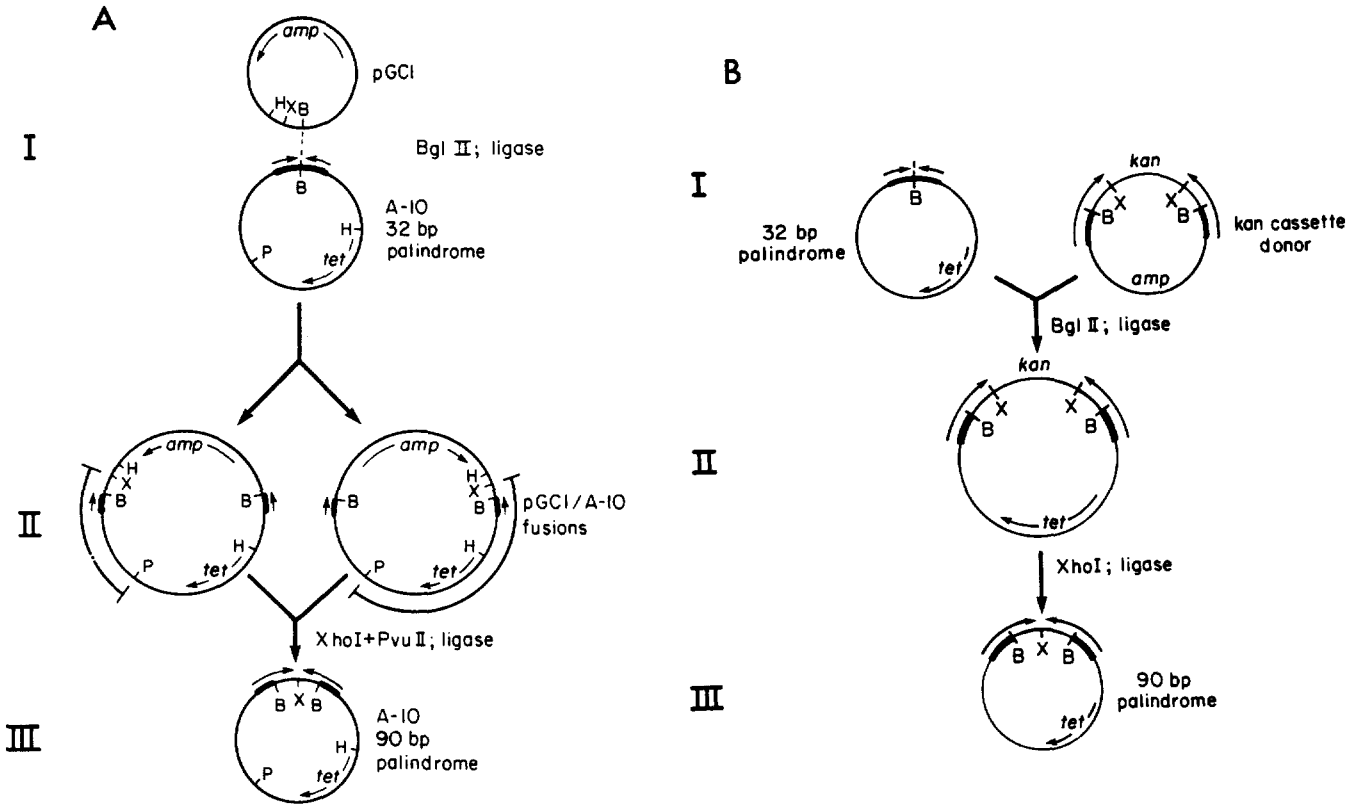

FIGURE 3.—Conversion of a 32-bp palindrome to a 90-bp palindrome. Restriction endonuclease cleavage sites used are *Bgl*II (B), *Xho*I (X), *Hin*dIII (H), and *Pvu*II (P). The inverted repeats are indicated by shortened arrows. A, Construction of the 90-bp palindrome using pGC1. (I) The DNAs of pBR322 derivatives with a 32-bp palindrome in the *amp* gene and of plasmid pGC1 were cleaved with *Bgl*II, ligated, and used to transform strain MC1061 to an Amp^r Tet^r phenotype. (II) Fusion plasmids were identified by size (7 kb). Representative plasmids were digested with *Hin*dIII to identify fusions in each orientation [for site A-10, *Hin*dIII fragment sizes of 3200 and 4000 bp (*left*) and 1000 and 6200 bp (*right*)]. The DNAs of fusion plasmids of each orientation were digested with *Xho*I and *Pvu*II, mixed, ligated, and used to transform MC1061 to a Tet^r (Amp^s) phenotype. (III) Plasmids containing the desired 90-bp palindrome were identified by size and by restriction endonuclease digestion (The *Xho*I site at the palindrome center was resistant to *Xho*I in our preparations unless the DNAs were linearized with a second enzyme, probably because the palindromic DNAs were extruded as cruciforms *in vitro*). This procedure was used to form a 90-bp uninterrupted palindrome at six sites: A-10, A-13, A-15, A-22, A-37 and A-41. B, Construction using a 90-bp palindrome at another site. (I) The 90-bp palindrome constructed in panel A was marked by insertion of a *Xho*I fragment from Tn*5* which contains the *kan^r* gene and in which the *Bgl*II site near the begining of the *kan^r* gene had been inactivated. (II) *Bgl*II digestion and ligation was used to move the central segment of the *kan^r*-marked palindrome to sites in Amp^s insertion mutant plasmids with the 32-bp palindrome. (III) The *kan^r* marker was removed using *Xho*I. This procedure was used to place the 90 bp palindrome at the other seven sites (A-40, A-32, A-24, A-1, A-19, A-21 and A-18).

palindrome at the 10^4-fold more stable A-37 site. (ii) The 90-bp palindrome at A-10 did not cause a significant accumulation of plasmid-free segregants (<0.2% after 30 generations of unselected growth). (iii) A fluctuation test (LURIA and DELBRUCK 1943) modified for use with single colonies (MATERIALS AND METHODS) established that the frequency of revertants measured in young stationary phase cultures reflects their rate of formation. Figure 4 shows that after about 19 generations of growth from a single cell every colony with the 90-bp palindrome at site A-10 contained many revertants. In contrast, only two-thirds of the colonies with the 90-bp palindrome at A-41 and one-tenth of the colonies with the imperfect 85-bp palindrome at A-24 contained revertants, and then only one or a few Amp^r papillae per colony. These results are in accord with the revertant frequencies observed at 30 generations (Table 3), and rule out explanations in which the deletion frequency is constant, and only the fitness of insertion mutant plasmids or of the cells carrying them varies.

Insertions at neighboring sites can differ markedly in deletion frequency. Each of the six insertion sites studied in Table 3 is separated by at least 50 bp from the other sites. To assess the effects of small changes in location on deletion frequencies we studied insertions at six additional sites found within 30 bp of site A-24. The results (Table 4, top) indicated that the location of an insert did not affect its deletion frequency in any systematic way. As at other sites (Table 3), the 90-bp perfect palindrome was most deletion prone, and the 45-kb interrupted palindrome was least deletion prone.

Tests with an additional site (A-22), just 1 bp away from the most deletion-prone site (A-10), showed that a 1-bp shift can drastically affect DNA stability (Table

TABLE 2

Sequences of palindromic inserts[a]

22 bp	5′CTGTCTCTTGA/TCAAGAGACAG
32 bp	5′CTGTCTCTTGATCAGA/TCTGATCAAGAGACAG
90 bp	5′CTGTCTCTTGATCAGATCTGGGGGATCCTCTAGAGTCGACCCCTC/GAGGGGTCGACTCTAGAGGATCCCCCAGATCTGATCAAGAGACAG
85 bp[b]	5′CTGTCTCTTGATCAGATCTGGGGGATCCTCTAGAGTCGACCCCTC/GAGGGGTCGACTCTAGAGGATCCCCCA-δ-GATCAAGAGACAG

[a] The slash (/) indicates the center of symmetry which consist of *BclI*, *BglII* and *XhoI* sites in the 22-, 32- and 90-bp palindromes, respectively.

[b] The 85-bp sequence, A-24δ, is an imperfect palindrome containing 40- and 45-bp inverted repeats. The deletion endpoints in the 40-bp component correspond to 5′GATC direct repeats, and the placement of the -δ- (to indicate the position of the 5-bp deletion in this repeat) is arbitrary.

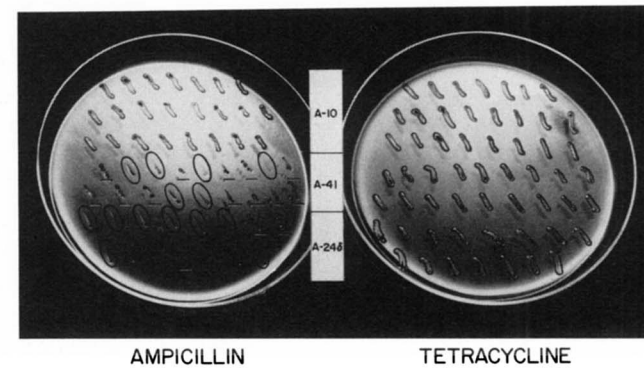

FIGURE 4.—Colony fluctuation test to estimate the time of appearance of Amp^r revertants. Colonies derived from single cells of strain MC1061 carrying plasmids with insertions of the 90-bp palindrome at sites A-10 and A-41 and the nearly perfect palindrome (one 40-bp and one 45-bp inverted repeat) at A-24 (labeled A-24δ) were grown about 19 generations on tetracycline-containing medium, and picked to ampicillin-containing and then to tetracycline-containing agar media. The elipses and horizontal bars indicate clones without Amp^r revertants, and with just a few Amp^r revertants, respectively.

4, bottom). The frequencies of deletion of the different size palindromes at A-10 and A-22 differed by between sixfold and 3,000-fold. Conversion of the 22-bp palindrome to 90 bp stimulated deletion about 18,000-fold at A-10, but only 38-fold at A-22. Thus, a small change in position can radically alter the stability of a given DNA segment.

***rec-A*-independence of deletion formation.** The formation of deletions is generally found to be *recA*-independent (FRANKLIN 1967; ROSS, SWAN and KLECKNER 1979; GHOSAL and SAEDLER 1979; EGNER and BERG 1981; COLLINS, VOLCKAERT and NEVERS 1982; BERG, EGNER and LOWE 1983), although there has been a report of *recA*$^+$-stimulation (ALBERTINI *et al.* 1982). The data presented in Table 5 show that the formation of deletions in our insertion mutant plasmids is not affected by the *recA* allele.

DISCUSSION

Earlier studies of the sequences of deletion endpoints and of transposon excision had indicated that the control of deletion formation is complex (see Introduction). Direct repeats and inverted repeats promote deletion events. While deletions are usually smaller than 200 bp, at least in an F′ episome, larger (>700 bp) deletions typically also end in direct repeats. Large deletions are more frequent in an F′ episome than in the bacterial chromosome, possibly because the conjugal transfer of single DNA strands is deletionogenic.

Our choice of reversion tests to analyze the deletion process, rather than the traditional sequence analyses of forward mutations, was dictated by several considerations: (i) *a priori* knowledge of the deletion endpoints in reversion tests eliminates the need to sequence each deletion; (ii) rare, as well as frequent,

TABLE 3

Amp^r revertant frequencies as a function of palindrome length and position[a]

Site designation	pBR322 position[b]	Interrupted palindrome[c] 45 kb (1534 bp IR)	Uninterrupted palindromes[d] 22 bp	32 bp	90 bp	Relative deletion frequencies 22 bp ÷ 45 kb	32 bp ÷ 22 bp	90 bp ÷ 22 bp
A-10	3731–3739	2.5×10^{-7}	9.5×10^{-7}	1.1×10^{-4}	1.7×10^{-2}	3.8	116	18,000
A-41	3893–3901	1.4×10^{-7}	5.7×10^{-7}	1.2×10^{-6}	1.5×10^{-4}	4.1	2.1	260
A-24	3411–3419	1.0×10^{-7}	4.0×10^{-6}	5.0×10^{-6}	3.6×10^{-5}	40	1.3	9
A-24δ[e]					5.3×10^{-6}			
A-15	4015–4023	2.2×10^{-10}	1.0×10^{-7}	2.0×10^{-7}	4.1×10^{-6}	455	2.0	41
A-13	3995–4003	1.1×10^{-10}	6.0×10^{-8}	1.5×10^{-7}	4.3×10^{-6}	545	2.5	72
A-37	3941–3949	1.1×10^{-10}	1.4×10^{-8}	3.4×10^{-8}	4.4×10^{-6}	127	2.4	314

[a] The Amp^r revertant frequencies were determined in plasmid-carrying derivatives of strain MC1061, and are based on the median revertant frequencies of at least five separate subclones (frequencies generally varied by less than a factor of two).
[b] Refers to positions of insertion in the *amp* gene of plasmid pBR322. See Table 1 for nucleotide sequences.
[c] A λ-IS*50* transposon in which inverted repeats of IS*50* elements (1534 bp) bracket the λ phage genome (Figure 2).
[d] See Table 2 for DNA sequences.
[e] A-24δ designates an imperfect palindrome at site A-24 in which one of inverted repeat is 40 bp long and the other is 45 bp (Table 2).

TABLE 4

Amp^r revertant frequencies at closely linked sites[a]

Site designation	Relative pBR322 position[b]	Interrupted palindrome 45 kb (1534 bp IR)	Uninterrupted palindromes 22bp	32 bp	90 bp	Relative deletion frequencies 22 bp ÷ 45 kb	32 bp ÷ 22 bp	90 bp ÷ 22 bp
A-40	−23	3.4×10^{-8}	2.2×10^{-7}	4.1×10^{-7}	5.3×10^{-5}	6.4	1.9	241
A-32	−2	3.0×10^{-8}	3.8×10^{-7}	1.0×10^{-6}	2.9×10^{-6}	13	2.6	8
A-24	0	1.0×10^{-7}	4.0×10^{-6}	5.0×10^{-6}	3.6×10^{-5}	40	1.3	9
A-1	+5	1.5×10^{-8}	1.0×10^{-7}	7.5×10^{-7}	2.7×10^{-6}	6.7	7.5	27
A-19	+7	2.9×10^{-8}	9.0×10^{-8}	2.1×10^{-7}	2.2×10^{-6}	3.1	3.1	24
A-21	+19	1.5×10^{-7}	2.3×10^{-6}	2.9×10^{-6}	5.1×10^{-5}	15	1.2	22
A-18	+30	7.6×10^{-9}	1.0×10^{-7}	2.2×10^{-7}	1.9×10^{-6}	13	2.0	19
Range deletion frequencies:		20-fold	44-fold	23-fold	28-fold			
A-10	+320	2.5×10^{-7}	9.5×10^{-7}	1.1×10^{-4}	1.7×10^{-2}	3.8	116	18,000
A-22	+319	3.8×10^{-9}	1.5×10^{-7}	3.5×10^{-7}	5.7×10^{-6}	40	2.3	38
Ratio of deletion frequencies at A-10 ÷ A-22:		66	6.3	310	3,000			

[a] The frequencies of Amp^r revertants were determined in plasmid carrying derivatives of strain MC1061, as in Table 3. The data for sites A-10 and A-24 are taken from Table 3.
[b] Position relative to site A-24 (pBR322 coordinates 3411–3419; see Table 1).

events can be easily detected and measured by counting revertant colonies; (iii) identical DNA segments can be placed throughout a gene and changed at will, thereby permitting systematic analyses of the parameters affecting deletion frequency; and (iv) sequences so unstable that they would have been purged from the genome during evolution are readily generated as insertion mutations and can be analyzed by reversion.

As in several other studies, the formation of our deletions was *recA*-independent (Table 5). Because they involve 9-bp direct repeats, whereas homologies of at least 20 bp are needed for legitimate recombination (SHEN and HUANG 1986; SINGER *et al.* 1982), the deletions we studied do not arise by homologous recombination. We hypothesize that most arise by replication errors (Figure 1A). As originally proposed for frameshift mutations (STREISINGER *et al.* 1966; STREISINGER and OWEN 1985) and deletions in F′ episomes (FARABAUGH *et al.* 1978), misaligned pairing of template and nascent DNA strands during replica-

TABLE 5

Deletion frequencies in isogenic *recA*$^+$ and *recA*$^-$ strains[a]

Site	Palindrome length (bp)	Median revertant frequencies *recA*$^+$	*recA*$^-$
A-10	22	9.5×10^{-7}	1.7×10^{-6}
	32	1.5×10^{-4}	2.2×10^{-4}
	90	2.3×10^{-2}	2.2×10^{-2}
A-22	22	6.1×10^{-8}	6.6×10^{-8}
	32	1.7×10^{-7}	1.8×10^{-7}
	90	8.5×10^{-6}	6.0×10^{-6}
A-37	90	2.7×10^{-6}	2.8×10^{-6}

[a] Median frequencies of Amp^r revertants were measured as in Tables 3 and 4, but using plasmid carrying derivatives of the isogenic strains AB1157 (*recA*$^+$) and JC2924 (*recA*$^-$).

tion could result in bypassing part of the template. Such replication errors should be stimulated by palindromy because intrastrand pairing brings the flanking direct repeats together, and because hairpin structures interfere with the copying of single stranded templates (KORNBERG 1982).

Alternatively, some or all of the deletions we studied may arise by DNA cleavage (Figure 1B) (CHOW, DAVIDSON and BERG 1974; GLICKMAN and RIPLEY 1984), not replication. Cleavage might occur preferentially where a palindrome had been extruded as a cruciform structure. A viable deletion mutant might be rescued following limited exonucleolytic degradation of linear DNA by annealing of the ends at complementary sequences from the direct repeats, repair synthesis to fill gaps and ligation.

Our reversion tests showed deletion frequencies to be remarkably dependent on both the site and the size of the palindromic insertion. There was a 400-fold spread in deletion frequencies of the identical 22-bp palindrome among the 13 insertion sites tested and a 10,000-fold spread in deletion frequencies of the 90-bp palindrome (Tables 3 and 4). Therefore, sequences outside a given DNA segment must be significant determinants of its stability. Because previous data showed that the reversion of Tn*5* induced mutations also vary from site to site in two larger replicons, an F'*lac* episome and the *E. coli* chromosome (EGNER and BERG 1981; and unpublished), we suggest that effects of context on deletion formation seen here are not unique to insertions in small multicopy plasmids, nor to the perfect palindromes used in this study.

The finding that small deletions are formed preferentially (GALAS 1978), and that a *ssb* mutation stimulates deletion of the 9-kb Tn*10* element, but not a 50-bp Tn*10* remnant (LUNDBLAD and KLECKNER 1984) suggests mechanistic differences in the formation of large and of small deletions. The weak correlation between deletion frequencies of the 45-kb insert and of the shorter perfect palindromes from the same sites (Table 3) might also reflect the operation of different deletion mechanisms.

At each site, changing the palindrome from 22 to 90 bp stimulated deletion formation, but with an efficiency which was site dependent. There was a 450-fold difference in the extent of stimulation between the adjacent A-10 and A-22 sites, and 2,200-fold range of stimulation among the 13 sites tested. The increased deletion frequency associated with longer inverted repeats at each site is consistent with deletion frequency being a direct function of inverted repeat length *per se* (over the range of 22–90 bp and for the nested palindrome family studied). Nevertheless, the variation in extent of increase indicates that other, subtle, aspects of insert sequence or conformation also affect stability.

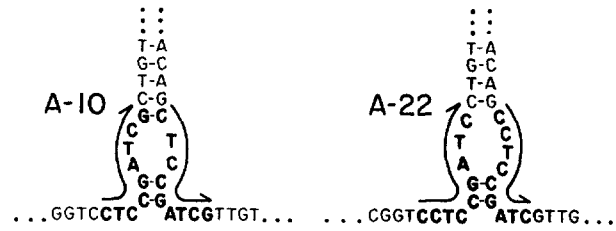

FIGURE 5.—Comparison of possible hairpin configurations of insertions at the adjacent sites A-10 and A-22. Note that one bp of the A-10 direct repeat (outlined) is also part of the perfectly matched palindrome, and that the unpaired loop in the A-10 insertion contains three and two bases, whereas the unpaired loop of the A-22 insertion contains three and four bases.

The observed location-dependence of deletion frequency may reflect the specificity of certain DNA binding proteins for particular DNA sequences or conformations. These might be proteins which participate in DNA replication, which recognize or stabilize cruciform structures, or which mediate DNA cleavage, exonucleolytic digestion or repair. For example, the drastic effect of a one bp difference in position on deletion frequency seen at sites A-10 and A-22 is not so bizzare when the effects of possible intrastrand pairing within the nine bp direct repeats are considered. This pairing yields a 3:2 mismatch (ATC opposite TC) at A-10 but a 3:4 mismatch (ATC opposite CCTC) at A-22 (Figure 5). These short loops should differ in stability, and might be acted on very differently by proteins of the replication complex (PAPANICOLAOU, LECOMTE and NINIO 1986) or by conformation-specific endonucleases (KEMPER *et al.* 1984).

It is likely that deletions arise by both replicative and break-join mechanisms (Figure 1), and that the absolute and relative frequencies of these two classes of events vary, depending on: physiological parameters such as nucleotide pools, DNA supercoiling, or the levels of replication and repair enzymes; the enzymology of replication of the DNA molecule under study; the extent and closeness of match of inverted repeats and also of direct repeat components; and the actual DNA sequences of both the segment undergoing deletion and the surrounding regions. The reversion tests developed here are well suited for systematically evaluating the importance of many of these parameters in the deletion process.

We thank M. REAGAN and D. WARNOCK for participation in the present experiments, and C. M. BERG and H. HUANG for stimulating discussions and critical readings of the manuscript. This research was supported by United States Public Health Service Research grants AI-14267 and GM-37138 to D.E.B.

LITERATURE CITED

ALBERTINI, A. M., M. HOFER, M. P. CALOS and J. H. MILLER, 1982 On the formation of spontaneous deletions: the importance of short sequence homologies in the generation of large deletions. Cell **29:** 319–328.

ARRAJ, J. A. and M. G. MARINUS, 1983 Phenotypic reversal in *dam* mutants of *Escherichia coli* K-12 by a recombinant plasmid containing the *dam*$^+$ gene. J. Bacteriol. **153:** 562–565.

BERG, D. E., C. EGNER and J. B. LOWE, 1983 Mechanism of F factor enhanced excision of transposon Tn*5*. Gene **22:** 1–7.

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

CALENDAR, R. A. and L. GOLD (Editors), 1985 Sequence specificity in transcription and translation. In: *UCLA Symposia on Molecular and Cellular Biology* (New Series), Vol. 30. Alan R. Liss, New York.

CHOW, L. T., N. DAVIDSON and D. E. BERG, 1974 Electron microscope study of the structures of λ*dv* DNAs. J. Mol. Biol. **86:** 69–89.

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.

EGNER, C. and D. E. BERG, 1981 Excision of transposon Tn*5* is dependent on the inverted repeats but not the transposase function of Tn*5*. Proc. Natl. Acad. Sci. USA **78:** 459–463.

FARABAUGH, P. I., U. SCHMEISSNER, M. HOFER and J. H. MILLER, 1978 Genetic studies on the *lac* repressor. VII. On the molecular nature of spontaneous hotspots in the *lacI* gene of *Escherichia coli*. J. Mol. Biol. **126:** 847–863.

FOSTER, T. J., V. LUNDBLAD, S. HANLEY-WAY, S. M. HALLING and N. KLECKNER, 1981 Three Tn*10*-associated excision events: relationship to transposition and role of direct and inverted repeats. Cell **23:** 215–227.

FRANKLIN, N. C., 1967 Extraordinary recombination events in *Escherichia coli*: their independence of the *rec*$^+$ function. Genetics **35:** 699–707.

GALAS, D. J., 1978 Analysis of sequence repeats in the *lacI* gene of *Escherichia coli*. J. Mol. Biol. **126:** 858–863.

GHOSAL, D. and H. SAEDLER, 1979 IS2-61 and IS2-611 arise by illegitimate recombination from IS2-6. Mol. Gen. Genet. **176:** 233–238.

GLICKMAN, B. W. and L. S. RIPLEY, 1984 Structural intermediates of deletion mutagenesis: a role for palindromic DNA. Proc. Natl. Acad. Sci. USA **81:** 512–516.

HERRMANN, R., K. NEUGEBAUER, H. ZENTGRAF and H. SCHALLER, 1978 Transposition of a DNA sequence determining kanamycin resistance in the single-stranded genome of bacteriophage fd. Mol. Gen. Genet. **159:** 171–178.

IKEDA, H., 1986 Bacteriophage T4 DNA topoisomerase mediates illegitimate recombination *in vitro*. Proc. Natl. Acad. Sci. USA **83:** 922–926.

IKEDA, H., I. KAWASAKI and M. GELLERT, 1984 Mechanism of illegitimate recombination: common sites for recombination and cleavage mediated by *E. coli* DNA gyrase. Mol. Gen. Genet. **196:** 546–549.

JONES, I. M., S. B. PRIMROSE and S. D. ERLICH, 1982 Recombination between short direct repeats in a *recA* host. Mol. Gen. Genet. **188:** 486–489.

KEMPER, B., JENSCH, M., V. DEPKA-PRONDZYNSKI, H.-J. FRITZ, U. BORGMEYER and K. MIZUUCHI, 1984 Resolution of Holliday structures by endonuclease VII as observed in interactions with cruciform DNA. Cold Spring Harbor Symp. Quant. Biol. **49:** 815–825.

KORNBERG, A., 1982 *Supplement to DNA Replication*. W. H. Freeman, San Francisco.

LOPEZ, P., M. ESPINOSA, B. GREENBERG and S. A. LACKS, 1984 Generation of deletions in pneumococcal *mal* genes cloned in *Bacillus subtilis*. Proc. Natl. Acad. Sci. USA **81:** 5189–5193.

LUNDBLAD, V. and N. KLECKNER, 1984 Mismatch repair mutations of *Escherichia coli* K12 enhance transposon excision. Genetics **109:** 3–19.

LURIA, S. and M. DELBRUCK, 1943 Mutations of bacteria from virus sensitivity to virus resistance. Genetics **28:** 491–511.

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

MARON, D. M. and B. N. AMES, 1983 Revised methods for the Salmonella mutagenicity test. Mutat. Res. **113:** 173–215.

MARVO, S. L., S. R. KING and S. R. JASKUNAS, 1983 Role of short regions of homology in intermolecular illegitimate recombination events. Proc. Natl. Acad. Sci. USA **80:** 2452–2456.

MAXAM, A. and W. GILBERT, 1980 Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65:** 499–560.

MICHEL, B. and S. D. EHRLICH, 1986 Illegitimate recombination at the replication origin of bacteriophage M13. Proc. Natl. Acad. Sci. USA **83:** 3386–3390.

MYERS, R. M., L. S. LERMAN and T. MANIATIS, 1985 A general method for saturation mutagenesis of cloned DNA fragments. Science **229:** 242–247.

NAG, D. K., U. DASGUPTA, G. ADELT and D. E. BERG, 1985 IS*50*-mediated inverse transposition: specificity and precision. Gene **34:** 17–26.

NOLLER, H. F., 1984 Structure of ribosomal RNA. Annu. Rev. Bioch. **53:** 119–162.

PAPANICOLAOU, C., P. LECOMTE and J. NINIO, 1986 Mnemonic aspects of *Escherichia coli* DNA polymerase I. Interaction with one template influences the next interaction with another template. J. Mol. Biol. **189:** 435–448.

PHADNIS, S. H. and D. E. BERG, 1985 *recA*-independent recombination between repeated *IS50* elements is not caused by an IS*50*-encoded function. J. Bacteriol. **161:** 928–932.

PLATT, T., 1981 Termination of transcription and its regulation in the tryptophan operon of *E. coli*. Cell **24:** 10–23.

ROSS, D. G., J. SWAN and N. KLECKNER, 1979 Physical structures of Tn*10*-promoted deletions and inversions: role of 1400 bp inverted repetitions. Cell **16:** 733–738.

SCHAAPER, R. M., B. N. DANFORTH and B. W. GLICKMAN, 1986 Mechanisms of spontaneous mutagenesis: an analysis of the spectrum of spontaneous mutation in the *Escherichia coli lacI* gene. J. Mol. Biol. **189:** 273–284.

SHEN, P. and H. V. HUANG, 1986 Homologous recombination in *Escherichia coli*: dependence on substrate length and homology. Genetics **112:** 441–457.

SINGER, B. S., L. GOLD, P. GAUSS and D. H. DOHERTY, 1982 Determination of the amount of homology required for recombination in bacteriophage T4. Cell **31:** 25–33.

STREISINGER, G. and J. (EMRICH) OWEN, 1985 Mechanisms for spontaneous and induced frameshift mutation in bacteriophage T4. Genetics **109:** 633–659.

STREISINGER, G., Y. OKADA, J. EMRICH, J. NEWTON, A. TSUGITA, E. TERZAGHI and M. INOUYE, 1966 Frameshift mutations and the genetic code. Cold Spring Harbor Symp. Quant. Biol. **31:** 77–84.

SYVANEN, M., J. D. HOPKINS, T. J. GRIFFIN IV, T.-Y. LIANG, K. IPPEN-IHLER and R. KOLODNER, 1986 Stimulation of precise excision and recombination by conjugal proficient F′ plasmids. Mol. Gen. Genet. **203:** 1–7.

Communicating editor: J. ROTH