The Influence of Primary and Secondary DNA Structure in Deletion and Duplication Between Direct Repeats in *Escherichia coli*

Thuan Q. Trinh*[,1] and Richard R. Sinden[†]

**Department of and Molecular Genetics, Biochemistry, and Microbiology, University of Cincinnati College of Medicine, Cincinnati, Ohio 45267–0524, and [†]Institute of Biosciences and Technology, Center for Genome Research, Texas A&M University, Houston, Texas 77030*

Manuscript received August 19, 1992
Accepted for publication November 16, 1993

ABSTRACT

We describe a system to measure the frequency of both deletions and duplications between direct repeats. Short 17- and 18-bp palindromic and nonpalindromic DNA sequences were cloned into the *Eco*RI site within the chloramphenicol acetyltransferase gene of plasmids pBR325 and pJT7. This creates an insert between direct repeated *Eco*RI sites and results in a chloramphenicol-sensitive phenotype. Selection for chloramphenicol resistance was utilized to select chloramphenicol resistant revertants that included those with precise deletion of the insert from plasmid pBR325 and duplication of the insert in plasmid pJT7. The frequency of deletion or duplication varied more than 500-fold depending on the sequence of the short sequence inserted into the *Eco*RI site. For the nonpalindromic inserts, multiple internal direct repeats and the length of the direct repeats appear to influence the frequency of deletion. Certain palindromic DNA sequences with the potential to form DNA hairpin structures that might stabilize the misalignment of direct repeats had a high frequency of deletion. Other DNA sequences with the potential to form structures that might destabilize misalignment of direct repeats had a very low frequency of deletion. Duplication mutations occurred at the highest frequency when the DNA between the direct repeats contained no direct or inverted repeats. The presence of inverted repeats dramatically reduced the frequency of duplications. The results support the slippage-misalignment model, suggesting that misalignment occurring during DNA replication leads to deletion and duplication mutations. The results also support the idea that the formation of DNA secondary structures during DNA replication can facilitate and direct specific mutagenic events.

THE accurate transmission of the genetic information is, in part, dependent on the complementary nature of the DNA double helix. Complementary base pairing in the proper tautomeric or ionic form and the correct alignment of the two strands is required to reproduce the genetic information during DNA replication, and pass information on from generation to generation. However, certain DNA sequences like reiterated bases (STREISINGER *et al.* 1966), direct repeats (FARABAUGH *et al.* 1978) and palindromes (RIPLEY 1982; GLICKMAN and RIPLEY 1984) present an enormous challenge to the accurate alignment of the two complementary DNA strands. Specific DNA sequences may form secondary structures that deviate from the normal duplex form of DNA and promote misalignment of one DNA strand on its complement, or promote pairing within a single DNA strand (DRAKE, GLICKMAN and RIPLEY 1983).

Spontaneous frameshift mutations have been observed in bacteriophage T4 at reiterated base pairs (STREISINGER *et al.* 1966; STREISINGER and OWEN 1985). In the T4 lysozyme gene, a run of five or six A residues provides a chance for "slipped mispairing" of adjacent residues between one strand and its complement. The slipped mispaired or misalignment model was proposed by STREISINGER *et al.* (1966) to explain the mechanism of addition and deletion of one base (a + 1 or −1 frameshift mutation) within a homopolymeric run of several bases. The same type of mutation has also been seen in the deletion and addition of one copy of the three tandem direct repeat sequence CTGG located in the *LacI* gene (FARABAUGH *et al.* 1978). The Streisinger model has been extended to explain the deletion of DNA located between direct repeats concomitant with one copy of the direct repeat (FARABAUGH *et al.* 1978; GALAS 1978). This type of mutation has been attributed to errors of replication caused by the interstrand misalignment of nonadjacent direct repeats (ALBERTINI *et al.* 1982).

Palindromic (inverted repeated) and quasi-palindromic DNA sequences have also been implicated in spontaneous mutagenesis. The normal interstrand duplex conformation of these sequences can unpair and alternate DNA conformations that involve intrastrand base pairing can form. Cruciform structures can form within palindromic DNA *in vitro* (LILLEY 1981; PAN-

[1] Present address: Life Technologies, Inc., Molecular Biology Research and Development, Gaithersburg, Maryland 20877.

AYOTATOS and WELLS 1981; SINDEN, BROYLES, and PETTIJOHN 1983) as well as *in vivo* (HORWITZ and LOEB 1988; PANAYOTATOS and FONTAINE 1987; ZHENG *et al.* 1991). The protruding hairpins may be the cause of the instability of palindromes (COLLINS 1980; BETZ and SADLER 1981a; WILLIAMS and MÜLLER 1987; DASGUPTA, WESTON-HAFER and BERG 1987). Recently, the existence of cruciforms *in vivo* has been shown to correlate with an increase in their deletion frequency (SINDEN *et al.* 1991). In addition to perfect palindromes, many regions of DNA when single stranded have symmetrical elements sufficient to promote the formation of complex structures in which interstrand misalignment may occur (GLICKMAN and RIPLEY 1984). These complex intermediate structures are believed to be responsible for diverse types of mutations including deletion, addition, duplication and various frameshift mutations (SCHAAPER, DANFORTH and GLICKMAN 1986; RIPLEY 1982). Aberrant DNA processing and synthesis at these complex secondary structures have been proposed to explain most of these complex mutations (RIPLEY 1982; RIPLEY and GLICKMAN 1983; GLICKMAN and RIPLEY 1984). Recombination cannot be solely responsible for deletion and duplications since some direct repeat-directed and palindrome-directed mutations have been shown to be independent of RecA (FARABAUGH *et al.* 1978; COLLINS 1980; DASGUPTA, WESTON-HAFER and BERG 1987; SINDEN *et al.* 1991).

In some instances there is an overlap between direct repeat-directed and palindrome-directed mutations. The mutations S74, S23 and S32 in the *Escherichia coli lacI* gene (GLICKMAN and RIPLEY 1984) and the *eG348* mutation in the T4 lysozyme gene (RIPLEY and GLICKMAN 1983) are classic examples. In such cases, it is proposed that the interstrand misalignment of direct repeats is stabilized by the intrastrand base pairing of the palindrome (RIPLEY and GLICKMAN 1983). In these cases a three-way junction could form that may present a stable misaligned intermediate for replication (TRINH and SINDEN 1991). Deletion via this secondary structure misalignment model involves a forward slippage and mispairing of the progeny strand facilitated by the secondary structure formation in the template strand. The model also predicts precise duplications since there is the possibility of secondary structure formation in the progeny strand after synthesis of the palindrome. Duplication requires a backward slippage of the progeny strand, misalignment of the direct repeat, followed by resynthesis of the palindromic DNA.

We have designed a system to measure the frequency of deletions and duplications between direct repeats. The construction of specific DNA inserts between direct repeats allows examination of the role of DNA secondary structure in the formation of deletion and duplications.

MATERIALS AND METHODS

Media, bacterial strains and plasmid vectors: Luria broth contained 10.0 g bacto-tryptone, 5.0 g yeast extract, 10.0 g NaCl, and 1.5 ml 1 N NaOH per 1000 ml H_2O. K medium consisted of M9 buffer (1 g of NH_4Cl, 11 g of $Na_2HPO_4 \cdot 7H_2O$ and 3 g KH_2PO_4 per liter of distilled water) with the addition of the following sterile components: 20 ml 10% casamino acids, 10 ml 20% glucose, 0.4 ml 0.1 M $CaCl_2$, 0.2 ml 1 M $MgSO_4$ and 0.2 ml 0.01% thiamine per 200 ml M9 buffer.

Bacterial strains used were *E. coli* K12 derivatives, HB101 (F^-, *hsdS20*, *mcr*B *mrr*, *recA13*, *ara-14*, *proA2*, *lacY1*, *galK2*, *rpsL20* (Sm^r), *xyl-5*, *mtl-1*, *supE44*) and DH5 (F^-, *recA1*, *supE44*, *endA1*, *hsdR17*(r_k^-,m_k^+), *thi-1*, *gyrA96*, *relA1*). Plasmid pBR325 harbors the genes coding for resistance to the antibiotics tetracycline (Tet), ampicillin, and chloramphenicol (Cm) (BOLIVAR 1978; PRENTSKI *et al.* 1981). The chloramphenicol acetyltransferase (CAT) gene contains an *Eco*RI site, which is unique to the plasmid (ALTON and VAPNEK 1979). Plasmid pBR325 has replication elements derived from pMP1, a ColE1-like plasmid (BETLACH *et al.* 1976), which has a unidirectional origin of replication (BOLIVAR 1978; SCOTT 1984). Cloning a 17-bp fragment or an 18-bp fragment with TAA stop codon into the *Eco*RI site of plasmid pBR325 inactivates the CAT gene, allowing for the selection of plasmids containing the insert. Plasmid pJT7 is derived from pBR325 and contains a single deletion (of a GC base pair) just upstream of the *Eco*RI site at position 4772 (BETZ and SADLER 1981b). To construct plasmid pBR523, the region between the two *Asu*II sites flanking the CAT gene was reversed. This reverses the orientation of the CAT gene within the pBR plasmid.

Oligo-deoxyribonucleotide synthesis and fragment preparation: 6, 7, 10, 11, 14, 17 and 18 bases oligo-deoxyribonucleotides and DNA primers for chain termination sequencing were synthesized using β-phosphoramidite chemistry with the Pharmacia LKB Biotechnology, Inc. Gene Assembler. The oligonucleotides were purified from 20% denaturing polyacrylamide gels (containing 8.3 M urea) run at 45 mA in TBE buffer as described (MAXAM and GILBERT 1980). After electrophoresis DNA bands were visualized by UV shadowing, cut out, masticated and eluted overnight at room temperature in 3 volumes of 1 M NaOAc, 2 mM EDTA. The elute solution was centrifuged through a glass wool pad in a 1.5-ml microcentrifuge tube and then precipitated with 4 volumes ethanol.

Oligo-deoxyribonucleotides were phosphorylated with [γ-^{32}P]-ATP (MANIATIS, FRITSCH and SAMBROOK 1982). Equimolar amounts of nonpalindromic complementary strands were annealed in TEN buffer (10 mM Tris, 50 mM NaCl and 0.1 mM EDTA) by heating to 95° and allowing the solution to slowly cool to room temperature (22–24°) or, in some cases, to very slowly cool to 4° overnight is a cold room. To minimize the self pairing or hairpin formation of palindromic oligonucleotides, 17- and 18-bp palindromic fragments were synthesized in three parts with the most palindromic strand being made as two oligonucleotides.

Plasmid construction: Plasmids pBR325n1, pBR325n2, pBR325n3, pBR325p1, pBR325p2, pBR325p3 and pBR325p4 were constructed by cloning DNA fragments 17n1, 17n2, 17n3, 18p1, 17p2, 17p3 and 17p4, respectively, into the *Eco*RI site of vector pBR325 (Table 1). Plasmids pJT7n1, pJT7n2, pJT7n3, pJT7p4 and pJT7p5

TABLE 1

DNA sequences

Sequence[a]	Insert name	Base composition of the insert		Flanking direct repeats		
		A + T	C + G	Size	Spacing	Sequence
ctcatccggAATTCGTCTGATGCACGaattccgt	n1	9	8	6	11	GAATTC
ctcatccggAATTCCGGTATATGCCGaattccgt	n2	9	8	8	9	GAATTCCG
ctcatccggAATTCCCCTATATCCCGaattccgt	n3	9	8	7	10	GAATTCC
ctcatccggAATTCTaattccgt	6n	5	1	5	1	AATTC
ctcatccggAATTCCTATAGGAATTCTaattccgt	p1	13	5	6	12	AATTCC
ctcatccggAATTCTTATAAGAATTCaattccgt	p2	14	3	5	12	AATTC
ctcatccggAATTCCCCGGGGAATTCaattccgt	p3	8	9	6	11	AATTCC
ctcatccggAATTCCCCATTATGGGGaattccgt	p4	9	8	8	9	GGAATTCC
ctcatccggAATTCGTTACGAATTCAaattccgt	p5	12	5	5	13	AATTC

[a] The sequence of the region containing a 17- or 18-bp insert is shown. The synthesized insert is shown in capital letters; the pBR325 sequence is shown in the small letters.

were made by cloning the appropriate DNA fragments into the *Eco*RI site of plasmid pJT7 (Table 1). Plasmid pBR6n was created by removing the 12-bp *Eco*RI fragment from plasmid pBRp1. The sequence of the inserts in all plasmids constructed was confirmed by DNA sequence analysis. Table 1 shows the DNA sequence of the inserts in the orientation studied.

Plasmid purification and molecular biological procedures: For small scale plasmid preparations, bacterial cultures were inoculated from overnight cultures (which were inoculated from single colonies) into 10–15 ml K medium containing 10 μg/ml tetracycline. Cultures were incubated at 37° with vigorous shaking (300 rpm) until an OD_{650} 0.7–0.9 was reached. The culture was then amplified by the addition of spectinomycin (300 μg/ml) and continued incubation with shaking overnight. DNA was initially purified as described by HOLMES and QUIGLEY (1981). Plasmid DNA then was further purified as described by SINDEN *et al.* 1991.

For plasmid purification from a large scale culture, 5 ml of an overnight culture was inoculated into 500 ml K media containing 10 μg/ml Tet at 37°. The culture was incubated with shaking at 300 rpm until an $OD_{650} = 0.7$ was reached. Spectinomycin was added to 300 μg/ml and the culture was incubated for an additional 10–14 hr for plasmid amplification. Covalently closed circular plasmid was purified from CsCl gradients as described by SINDEN, CARLSON and PETTIJOHN (1980).

For the recovery of DNA from agarose after digestion with restriction enzymes, DNA was precipitated, redissolved, and loaded into a 0.8% low melting agarose (SeaPlaque) gel. The gel was run at 5 V/cm in TBE buffer (89 mM Tris, 89 mM boric acid and 2.5 mM Na_2 EDTA) to separate the DNA fragments. A specific band was cut from the gel and the DNA was recovered from agarose as described by WEISLANDER (1979). Briefly, the agarose slices were placed in 5 volumes of 20 mM Tris-HCl, pH 7.6, and 1 mM EDTA, and melted by incubation at 65° for several minutes. The solution was adjusted to 0.3 M NaOAc at room temperature and extracted with phenol, phenol:chloroform (1:1) and finally chloroform:isoamylalcohol (24:1). The DNA then was precipitated by the addition of two volumes of ethanol.

DNA cloning and other standard molecular biology procedures were done as described by MANIATIS, FRITSCH, and SAMBROOK (1982). Restriction enzyme digestions, T4 DNA ligase and T4 polynucleotide kinase reactions were performed as suggested by the manufacturer. The preparation of competent cells and plasmid transformations were performed as described by HANAHAN (1983).

Reversion to chloramphenicol resistance: To estimate the Cm^r reversion frequency, single colonies were inoculated into 10 ml K medium containing 10 μg/ml tetracycline and the culture incubated overnight at 37°. A 25 ml culture was then inoculated at an $OD_{650} = 0.1$ and grown at 37° with shaking at 250 RPM to an $OD_{650} = 0.6–0.8$. Cells were then pelleted by centrifugation and resuspended in 1 ml of M9 buffer. Viable cells were estimated by plating serial dilutions of the cells on Luria broth plates containing 25 μg/ml Tet. Reversion to Cm^r was determined by plating 0.2–0.25 ml cells (the equivalent of about 5 ml of the original cells culture) on Luria broth plates containing 25 μg/ml each of tetracycline (Tet) and chloramphenicol (Cm). Colonies were counted after 48 hr of incubation at 37°. Control experiments showed that potential growth of revertants was not prevented by the large number of cells plated on each Tet/Cm plate.

Colony hybridization: Colony hybridization was used to detect the deletion of 8 bp that could be caused by the misalignment of direct repeat CCGG in plasmid pBRn2. Pin-head size Cm^r revertant colonies containing a single plasmid population were transferred to a fresh like Tet/Cm plate, allowed to grow, and hybridization performed as described by HANAHAN and MESELSON (1983). The oligonucleotide probe AATTCGGCATATACCGG was hybridized to the target sequence at 45° as suggested by SZOSTAK *et al.* (1980).

There are multiple copies of pBR325 plasmids in *E. coli* cells, and a single Cm^r plasmid may make bacteria resistant to Cm. Therefore, revertant cells will harbor two plasmid populations: the parental and Cm^r revertant plasmids. This mixed population of plasmids would cause a false positive result in the colony hybridization analysis since both the parental and the Cm^r plasmid with 8 bp deleted from the insert will hybridize to the probe with the same efficiency (as shown below in Figure 2). Before colony hybridization, an *E. coli* strain containing a single Cm^r plasmid population was selected. For this, DNA from 50 original individual revertant colonies was purified and transformed into DH5 competent cells and the resulting Cm^r colonies were grown on Tet/Cm plates. Cm^r colonies grown from these 50 plates should contain a plasmid population composed of a single revertant type, with complete deletion or partial deletion of the insert.

Analysis of Cm^r reversion events: For DNA sequence

analysis it was necessary to isolate the revertant plasmid from any parental plasmid present in the original revertant. This is because reversion of one copy of the plasmid per cell should be sufficient to make the cell Cm^r. Plasmid DNA was purified from chloramphenicol resistant revertant colonies, and the DNA was transformed into competent DH5 cells. Cells were plated on plates containing Tet and Cm to select for colonies containing only a Cm^r plasmid. Single colonies from these Tet/Cm plates were grown in 10 ml K medium and DNA was purified as described above. The revertants were characterized by digestion with *Alu*I, which allowed analysis of the 129-bp *Alu*I fragment containing the *Eco*RI site. The size of the insert in the *Eco*RI site could be determined within about 2–3 bp from analysis of the *Alu*I digest. The DNA sequence of the region was determined for several revertants from each insert; 12–24 independent revertants were examined in for each plasmid.

DNA sequencing reactions were done using the dideoxy nucleotide chain termination method (SANGER, NICKEN and COULSON 1977) using reverse transcriptase as described (ZAGURSKY *et al.* 1985) except that the reaction was carried out at 45°. Alternatively, for some experiments a Sequenase (modified T7 DNA polymerase) kit (United States Biochemical) was used according to the manufacturer's instruction with the exception that NaOH/NaOAc was used for primer annealing (ZAGURSKY *et al.* 1985). In addition, the sequencing reactions were done at 42° for 12 min. The sequencing products were resolved on a denaturing 8% polyacrylamide gel (containing 8.3 M urea) in TBE buffer as described (MAXAM and GILBERT 1980).

RESULTS

An assay for deletions and duplications between direct repeats: We have designed the genetic system shown in Figure 1 for the selection of deletion and duplication mutations between direct repeats. Palindromic and nonpalindromic 17- and 18-bp DNA fragments were cloned into the unique *Eco*RI site located in the chloramphenicol acetyl transferase (CAT) gene of plasmid pBR325. A 17-bp insert creates a +2 frameshift mutation in the CAT gene causing a chloramphenicol sensitive (Cm^s) phenotype. The p1 18-bp insert contains an in frame TAA stop codon. A complete deletion of the 17- or 18-bp insert restores the reading frame, resulting in a chloramphenicol resistant (Cm^r) phenotype. A duplication event producing a 34-bp insert will not restore the reading frame and would result in a Cm^s phenotype. To select for duplication events we utilized plasmid pJT7, which is identical to pBR325 with the exception that it contains a −1 frameshift mutation (a deletion of a C:G base pair) just upstream of the *Eco*RI site in the CAT gene (BETZ and SADLER 1981b). pJT7 and derivatives containing a 17-bp insert are Cm^s. A duplication of a 17-bp insert at the *Eco*RI site will restore the reading frame of the CAT gene and produce a Cm^r phenotype (Figure 1). BETZ and SADLER (1981b) showed that a 39-bp insert at the *Eco*RI site resulted in a functional CAT protein. Provided there is no stop codon, the CAT gene can functionally tolerate an insert of up to 51 bp (T. Q. TRINH and R. R. SINDEN, unpublished data).

Measurement of deletions in pBR325: The frequency of deletion for the 17- and 18-bp insert sequences shown in Table 1 are listed in Table 2. Cm^r reversion frequencies in pBR325 plasmids containing the various inserts were obtained in two *recA* strains HB101 and DH5. In all cases, the Cm^r reversion frequencies were similar between these two strains. The reversion frequencies varied from $<0.1 \times 10^{-9}$ to more than 50×10^{-9}. The reversion frequencies of the nonpalindromic (n) inserts varied about 10-fold, whereas the reversion frequencies of the palindromic (p) inserts varied nearly 350-fold.

Plasmid pBR6n was created by digesting pBR325p1 with *Eco*RI and religating. This resulted in the loss of a 12-bp *Eco*RI fragment and production of a construct with a 5-bp AATTC direct repeat separated by 1 bp. This effectively created a 6-bp insert with an in-frame TAA stop codon, which resulted in the Cm^s phenotype. Reversion to Cm^r resistance occurred at a frequency of 4.6×10^{-9}.

Restriction digestion with *Alu*I was used to analyze the nature of the insert in a number of randomly selected revertants. As shown in Table 2, 14 of 24 or 58% of the n1 revertants showed a deletion of the insert. DNA sequencing was done on several deletion revertants of this class to confirm the precise deletion. For the n2 and n3 nonpalindromic inserts, 75% and 91% of the revertants had precise deletions of the insert. None of the 6n revertants analyzed contained a deletion. All were an A to C transversion changing the TAA stop codon to a TCA codon. All revertants analyzed from plasmids containing the p1 and p4 inserts, which reverted at a high frequency, had precise deletions of the insert. All revertants from the p2 insert, which had a low frequency of Cm^r reversion, had complete deletions. The sequences of the "other" revertant class of mutations were not determined. From *Alu*I digestion analysis, many of this "other" class contained +1 or −2 frame shift mutations since the size of the *Alu*I band containing the insert was very similar to that from the original Cm^s plasmid. These events were not analyzed by DNA sequencing. The "estimated deletion and duplication frequencies," which represent the Cm^r reversion frequency multiplied by the fraction of the revertants containing the precise deletion or duplication mutation, are shown in Table 2.

In some instances two *Alu*I bands containing the *Eco*RI site were present; the original band corresponding to an insert and that corresponding to deletion. Analysis on agarose gels showed that this was a result of a dimer plasmid containing one pBR325 molecule with an intact CAT gene and one molecule with a deletion in the CAT gene. The formation of deletions between short direct repeats in some plasmids in *E. coli* is associated exclusively with dimer formation

A. Deletion Assay in pBR325	Insert Size	Phenotype
1. Sequence of the *Eco*RI region in pBR325		
TGATGATGAATGCTCATCCGGAATTCCGTATGGCAATGAAAGACGG	0	Cm^r
2. Sequence of *Eco*RI region in pBR325n1		
---------->		
TGATGATGAATGCTCATCCGGaattcgtctgatgcacgAATTCCGTATGGCAATGAAAGACGG	17 bp	Cm^s
3. Deletion of the insert from pBR325n1		
TGATGATGAATGCTCATCCGGAATTCCGTATGGCAATGAAAGACGG	0	Cm^r
B. Duplication Assay in pJT7		
1. Sequence of the *Eco*RI region in pJT7		
TGATGATGAATGCTCATCGGAATTCCGTATGGCAATGAAAGACGG	0	Cm^s
2. Sequence of the *Eco*RI region in pJT7n2		
---------->		
TGATGATGAATGCTCATCGGaattccggtatatgccgAATTCCGTATGGCAATGAAAGACGG	17 bp	Cm^s
3. Duplication of the insert in pJT7n2.		
----------><----------		
TGATGATGAATGCTCATCGGaattccggtatatgccgaattccggtatatgccgAATTCCGT		
ATGGCAATGAAAGACGG	34 bp	Cm^r

FIGURE 1.—System for selection of deletion and duplication mutations. (A) Deletion assay in pBR325. (1) Sequence of the *Eco*RI region in pBR325. pBR325 contains the CAT gene and has a chloramphenicol resistant (Cm^r) phenotype. (2) Sequence of the *Eco*RI region in pBR325n1. Insertion of the 17-bp n1 insert destroys the reading frame of the CAT gene and produces a Cm^s phenotype. The sequence of the n1 insert is shown in lower case letters. (3) Deletion of the insert from pBR325n1. On precise deletion of the insert, the original reading frame of the CAT gene is restored, resulting in a Cm^r phenotype. (B) Duplication assay in pJT7. (1) Sequence of the *Eco*RI region in pJT7. pJT7 contains a −1 deletion of a C 2 or 3 bp 5′ to the *Eco*RI site (GAATTC). Consequently, pJT7 has a Cm^s phenotype. (2) Sequence of the *Eco*RI region in pJT7n2. The insertion of a 17-bp fragment still results in a Cm^s phenotype. (3) Duplication of the insert in pJT7n2. Duplication of the insert produces a 34-bp insert which, with the −1 deletion mutation present in the plasmid, results in a 33-bp insert, which restores the reading frame of the CAT gene. Thus, duplication produces a Cm^r phenotype and is selectable on chloramphenicol plates. The CAT gene will tolerate at least a 51-bp insert and still produce a functional chloramphenicol acetyltransferase gene product. The *Eco*RI sites are underlined in the figure above. The n1 and n2 17-bp inserts are indicated by an arrow above the sequence.

(MAZIN *et al.* 1991). Dimers were seen rarely (<5%), except in the n1 and n2 inserts where they constituted about 10% of the revertant population. In one case examined when Cm^r revertant dimers were present, monomers were also present in other colonies of the transformation.

Analysis of an 8-bp deletion in pBRn2: When the n2 insert is cloned in the orientation shown, there is a CCGC sequence flanking the insert that is directly repeated 4 bp away, within the n2 insert. The sequence and repeats are shown in Figure 2. Deletion of one direct repeat and the intervening DNA should result in an 8-bp deletion, which should restore the reading frame of the gene. The resulting deletion should contain a sequence identical to that of the insert cloned into the original plasmid. The synthesized strand therefore provides a convenient hybridization probe for this 8-bp deletion. Among 50 Cm^r revertant colonies analyzed, three showed a positive hybridization. DNA sequence analysis revealed the expected deletion (TRINH 1990). This result demonstrates that the 8-bp deletion between 4-bp direct repeats occurred, although at a frequency of 6% that of the larger 17-bp deletion between 8-bp direct repeats. The 3 "other" class Cm^r revertants analyzed in the *Alu*I digests shown in Table 2 had 1- to 2-bp insertions or deletions.

Measurement of duplications in pJT7: Duplication mutations were detected in plasmid pJT7. The Cm^r reversion frequencies were determined for plasmids containing several different inserts in strains HB101 and DH5. The Cm^r reversion frequencies varied more than 100-fold for various inserts. As shown in Table 2, the Cm^r reversion frequencies of the nonpalindromic inserts varied from $4–79 \times 10^{-9}$. For the n1 nonpalindromic insert 24/29 Cm^r revertants (83%) were duplications. With insert n2, 21/28 revertants (75%) were duplications. In contrast, no duplications were detected in 11 n3 revertants analyzed; of these, six were large deletions and five small deletions. Of the 24 revertants of the p4 insert analyzed, only one duplication was detected.

DISCUSSION

The use of plasmids pBR325 and pJT7 containing inserts in the *Eco*RI site provide an opportunity to

TABLE 2

Cm^r reversion frequencies

	pBR325						pJT7					
	Cm^r reversion frequency[a] × 10^{-9}		Nature of the reversion event			Estimated deletion frequency[f] × 10^{-9}	Cm^r reversion frequency[a] × 10^{-9}		Nature of the reversion event			Estimated duplication frequency[f] × 10^{-9}
Insert	DH5	HB101	Deletion	Duplication	Other		DH5	HB101	Deletion	Duplication	Other	
n1	7.4	5.6	14	0	10[b]	3.8	14	8.3	0	24	5	9.1
n2	46	35	9	0	3[c]	30.0	4.0	4.0	2	21	5	3.4
n3	34	50	32	0	3[d]	38.2	55	79	6	0	5	<6.1
6n	ND	4.6	0	0	9[e]	<0.5	ND	ND				ND
p1	43	44	12	0	0	43.5	NA	NA				ND
p2	ND	<0.1	12	0	0		ND	ND				
p3	0.7	<0.1					ND	ND				
p4	32	36	24	0	0	34	4.3	4.6	2	1	21	0.19
p5	ND	ND					<0.5	<0.5				

[a] The reversion frequencies represent the average of eight to 12 independent experiments in which the individual frequencies varied by less than two- to threefold.

[b] "Other" mutations included five +1 or −2 frameshift and five small deletions of about −5 bp. These were not sequenced.

[c] "Other" mutations included +1 or −2 frameshifts. Some revertants of the n2 insert were on dimer plasmids, one copy contained the reversion event and the other, the original sequence containing the insert.

[d] +1 or −2 frameshifts.

[e] Sequenced 6n revertants all contained an A to C transversion destroying the TAA stop codon (TAA to TCA).

[f] "Estimated deletion (or duplication) frequency" is the chloramphenicol Cm^r reversion frequency multiplied by the percent deletions divided by 100. The Cm^r reversion frequencies used are the average of the frequencies in HB101 and DH5. The percent deletions represents the percent of the total revertants analyzed that contained deletion events.

ND, not done; NA, not applicable.

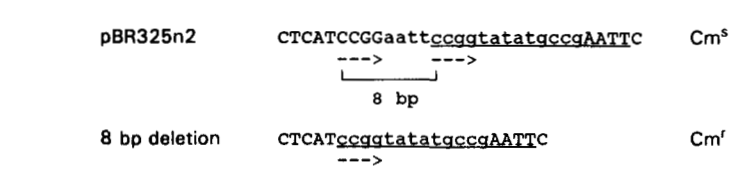

FIGURE 2.—Analysis of an 8-bp deletion in pBR325n2. The sequence of pBR325n2 and the region containing the 8-bp deletion are shown above. The n2 insert is shown in lower case letters. An 8-bp deletion between the direct repeats CCGG will restore the reading frame of the CAT gene, producing a Cm^r phenotype. The 17-bp region that is underlined shows a sequence of one of the strands used to clone the n2 insert. This 17-bp sequence is present in both pBR325n2 and the derivative containing the 8-bp deletion. Consequently this fragment can be used as hybridization probe to identify the revertants containing the 8-bp deletion.

measure the frequency of both deletions and duplications. A functional CAT protein can contain a variable number and composition of amino acids at the region encoded by the *Eco*RI recognition sequence GAATTC. The events responsible for restoring a functional chloramphenicol transacetylase protein from a gene containing a 17-bp insert include small frameshifts (+1, −2 bp), duplications (up to an insert of at least 51 bp) and precise deletions of the insert between the *Eco*RI direct repeats. By analyzing a number of different sequences inserted into this region of DNA we can examine the effect of subtle differences in the "mutagenic potential" of the DNA between direct repeats. These differences include primary base sequence or base composition (A + T or G + C content), the potential for slipped mispairing and the potential for secondary structure (hairpin) formation. Although this model system will allow one to analyze differences in the length and spacing of direct repeats between which deletion or duplication occurs, these factors were not a focus of this initial investigation. The primary goal of this initial investigation was to explore the applicability of this approach for detecting deletions and duplications and to investigate the role of DNA secondary structure in the formation of deletion and duplication mutations.

Deletion between direct repeats: It is well known that deletions occur between direct repeats (ALBERTINI *et al.* 1982; see RIPLEY 1991 for review). The frequency of deletion between direct repeats increases with increasing length of the direct repeat (DASGUPTA, WESTON-HAFER and BERG 1987, WILLIAMS and MÜLLER 1987, WESTON-HAFER and BERG 1989) as well as a decrease in the spacing of the direct repeats (SINGER and WESLYE 1988). A recent comprehensive study by PIERCE, KONG and MASKER (1991) demonstrated a log relationship between the frequency of deletion (of 76 bp) between direct repeats and the length of the direct repeats above 8 bp in a bacteriophage T7 system. There was also a substantial difference in the deletion frequency as the length of the direct repeat varied from 5–8 bp. The sequences analyzed here maintain a 9–13 bp spacing of 5–8 bp direct repeats. We have not observed as large a dependence of the deletion frequency on the length of the direct repeats between 6–8 bp as that observed by

Direct repeats in the non-palindromic inserts

```
n1   ctcatccggAATTCGTCTGATGCACGaattccgt
          1----->          1----->          17 bp

n2   ctcatccggAATTCCGGTATATGCCGaattccgt
          1------->         1------->        17 bp
       2ccg    2CCG      2CCG                8, (10, Cm^S), 17 bp
                  3TAT
                    3TAT                     2 bp
       4ggcc   4GGCC                         8 bp

n3   ctcatccggAATTCCCCTATATCCCGaattccgt
          1------>          1------>         17 bp
       2ccg               2CCG               17 bp
                  3TAT
                    3TAT                     2 bp
                 4CCC    4CCC                8 bp
                5CCCC    6CCC                +1 bp
               7TCCC    7TCCC                (9 bp, Cm^S)
```

Direct and inverted repeats in the palindromic inserts

```
p1   ctcatccggAATTCCTATAGGAATTCTaattccgt
              ----->            ----->
             -------><--------

p2   ctcatccggAATTCTTATAAGAATTCaattccgt
              ---->             ---->
             -------><--------

p3   ctcatccggAATTCCCCGGGGAATTCaattccgt
              ----->            ----->
               ------><--------

p4   ctcatccggAATTCCCCATTATGGGGaattccgt
             ------->          ------->
            ------------> <------------

p5   ctcatccggAATTCGTTACGAATTCAaattccgt
              ---->             ---->
             -------> <-------
```

FIGURE 3.—Organization of direct and inverted repeats in the mutation insert constructs. Direct repeats in the nonpalindromic inserts: The regions containing the three inserts n1, n2 and n3 are shown. The synthesized inserts are shown in capital letters. The direct repeats that include the *Eco*RI sites are underlined with arrows and numbered. The n1 insert creates 6-bp direct repeats that are separated by 11 bp; the 5′ (or 3′) bases of the direct repeats are 17 bp apart. The n2 construct has several other direct repeats indicated by the numbered sequences below the sequence. For example, the second direct repeat, CCG, is present three times. Deletion of 8 bp or 17 bp would restore the reading frame of CAT gene. A 10-bp deletion could also occur between the rightmost two direct repeats. However, this would not restore the reading frame of the CAT gene. The n3 sequence contains more direct repeats than the other two nonpalindromic inserts. There are six different direct repeat sequences in which deletion between them could restore the CAT gene reading frame. One deletion event, 7, would not restore the CAT gene reading frame. Repeats 5 and 6 are runs of Cs in which the insertion of a single base would restore the CAT gene frame. Direct and inverted repeats in the palindromic inserts: The bottom part of the figure shows the direct and inverted repeats in the palindromic inserts. The direct repeats are indicated by the two arrows facing to the right, below the sequence. The palindromic region is indicated by facing arrows below the direct repeat arrows.

PIERCE, KONG, and MASKER (1991). In many cases the repeat length differences in our system are relatively small and may have only a minor influence on the frequency of deletion between direct repeats. As described below, the primary sequence arrangement of the bases that can lead to potential slippage or secondary structure formation may have a major influence on the frequency of mutation. The frequency of deletion between 5-bp direct repeats was very low in agreement with the results of PIERCE, KONG, and MASKER (1991).

Deletion of nonpalindromic sequences: The three nonpalindromic sequences analyzed, n1, n2 and n3, were designed to have an identical length of 17 bp; identical base composition, 8 G:A base pairs and 9 A:T base pairs; and similar spacing of the direct repeats (9–11 bp). The direct repeats of all inserts begin with the sequence GAATTC. However, the frequency of deletion between these inserts varied considerably. The n1 insert had the lowest frequency of deletion. Compared with the n1 insert, the Cm^r reversion frequencies were increased about sixfold for the n2 insert and five- to ninefold for the n3 insert. Considering the estimated deletion frequency shown in Table 2, there was approximately an eight- and 10-fold difference between the frequency of deletion for the n1 inserts compared with n2 and n3, respectively. The eightfold difference between n1 and n2 may reflect, in part, the greater length of the direct repeat for n2. However, the slightly greater deletion frequency of n3 compared with n2 cannot be explained by direct repeat length, which is shorter for n3. n1, n2 and n3 vary in terms of the potential for misalignment. As shown in Figure 3, n1 has no internal direct repeats. n2 has two internal 3-bp direct repeats as well as 3-bp and 4-bp repeats between the insert and flanking region that could lead to a Cm^r phenotype on deletion. n3 has a 3-bp internal and a 3-bp external direct repeat in which deletion between them could lead to a Cm^r phenotype. In addition, n3 has two runs of C:G base pairs in which +1 or −2 frameshifts could occur to produce a Cm^r phenotype. Although the +1 or −2 frameshifts would revert the Cm^s phenotype, only 3/23 n3 revertants and 3/12 n2 revertants had small deletions. (These revertants were not sequenced to determine the identity of the reversion event.) In fact, the frequency of short deletions was highest for the n1 insert and lowest for the n3 insert (Table 2). The hybridization analysis of the n2 insert showed that 6% of the Cm^r revertants contained the 8-bp deletion

FIGURE 4.—Misalignments in the p4 insert. Structure A shows the sequence of the region surrounding the insert in the pBR325p4 construct. The 17-bp p4 insert is shown in capital letters. The 8-bp direct repeats are indicated by arrows above the sequence whereas the inverted repeated is shown by facing arrows below the sequence. Misalignment could occur by replication up to the hairpin stem shown in structure B. Structure C shows continued replication through the hairpin stem, resulting in a synthesis of the first 8-bp direct repeat. In structure D the first 8-bp direct repeat in the progeny strand pairs with the second copy in the template strand, resulting in the formation of a 17-bp loop in the template strand. This 17-bp loop contains several short inverted repeated elements that could form two very short hairpin stems shown in E. The leftmost of these constitutes the *Eco*RI site whereas the rightmost represent a 9-bp direct repeat which is part of the p4 insert. This may not represent a very stable or likely structure since a loop of 3–4 bp is needed at the end of the hairpin loop. Structure F shows a DNA molecule that may represent a stable deletion intermediate. In structure F, the hairpin stem forms by initiating pairing at the center of symmetry of the inverted repeat. This pairing could initiate within the loop shown in D accompanied by branch migration of the loop to the right as compared with D. In addition, this structure could also form from the structure shown in C by extrusion of the short hairpin stem concomitant with a leftward movement of the base of the hairpin stem relative to the direct repeat in the progeny strand. This would be accompanied by pairing of the first direct repeat in the progeny strand with parts of the first and second direct repeats in the template strand.

between CCGC. This would correspond to an estimated deletion frequency of this event of 2.7×10^{-9}.

The various frequencies of deletion of the nonpalindromic inserts should reflect the molecular mechanisms and biochemical events occurring during the process of spontaneous mutagenesis. The structures proposed for the misalignment of the n1 insert that lead to deletion have been shown previously (TRINH and SINDEN 1991). There are several models that might explain the higher frequency of deletion in n3 and n2 compared to n1. One possibly is that the *in vivo* frequency at which polymerase pauses and dissociates is higher at the end of the first direct repeat of the n2 or n3 inserts compared to the n1 insert. KUNKEL (1990) and PAPANICOLAOU and RIPLEY (1991) have shown that replication pause sites are frequently associated with mutations. A second model is one in which multiple opportunities for the formation of short misalignments increases the potential for slippage between the flanking direct repeats in the n2 and n3 insert. Any one of a number of slippages in the template or progeny strand may provide an *initiating event* that destabilizes the polymerase complex and allows a greater opportunity for a more extensive

FIGURE 5.—The formation of three-way junctions in the palindromic p1, p2 and p3 insert constructs. (A) pBR325p1. (1) The sequence and symmetry elements of the region containing the p1 insert are shown. The 18-bp p1 (and 17-bp p2 and p3) insert is shown in capital letters. Direct repeats are shown by the dashed arrows above, and the inverted repeat is shown by the facing arrows below the sequence. (2) The intermediate following replication of first direct repeat is shown. (3) Misalignment between the first direct repeat the progeny strand and the second direct repeat in the template strand is shown in which the palindromic region forms a perfect hairpin stem resulting in the formation of a three-way junction. The junction is perfect with the exception of a G:A mismatch involving the G of the *Eco*RI site. In panels B and C, similar structures can form; however, the sequences of p2 and p3 are 1 base shorter, which results in a unpaired G in the progeny strand on the formation of the three-way junction. The presence of this extra G, which would have to be either extrahelical or intercalated into the three-way junction, may act to destabilize the pairing of the first direct repeat of the progeny strand with the second direct repeat of the template strand. This destabilizing effect may be responsible for the low deletion frequency observed for the p2 and p3 constructs. The proposed three-way junctions do not contain mispaired bases (only the unpaired G).

misalignment to occur. Once this larger misalignment involving the flanking direct repeats has occurred, exonucleolytic proofreading would digest the nascent strand back to the end of the first direct repeat which would now be perfectly paired with the second copy of the direct repeat in the template strand.

Deletion of palindromic sequences: The deletion between direct repeats for inserts in which the DNA sequence between the direct repeats was palindromic varied considerably from <0.1–40×10^{-9} revertants per cell. All p1, p2 and p4 revertants examined had deleted one copy of the direct repeat together with the intervening inverted repeat. Sequence p4 has perfect inverted repeat symmetry between and including the *Eco*RI sites (Figure 3). The frequency of deletion for the p4 insert was about 34×10^{-9}. It can form the hairpin structure and the other misaligned intermediates shown in Figure 4. One potential misaligned intermediate contains a loop of DNA like that shown for the nonpalindromic inserts (Figure 4D). Another possible intermediate, structure E, contains two short hairpins that can form in DNA sequences present in the loop of structure D. One very short hairpin contains the *Eco*RI sequence as a bulge at the 3′ side of a second short hairpin (Figure 4E). It is known that large inverted repeats are unstable (BETZ and SADLER 1981a; LILLEY 1981; WELLER *et al.* 1985; LOCKSHON and GALLOWAY 1986) and that the formation of cruciforms *in vivo* can lead to deletion of the inverted repeat, presumably via a misaligned structure as

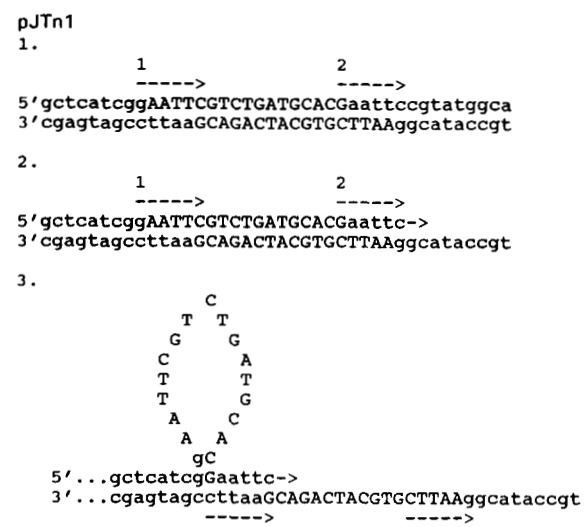

FIGURE 6.—Duplication in the nonpalindromic inserts. (1) The sequence of the *Eco*RI region of the pJTn1 construct. The 17-bp insert is shown in capital letters and the direct repeats are indicated by dashed arrows above the sequence. (2) The replication intermediate in which pausing has occurred following synthesis of the second direct repeat is shown. (3) Structure 3 shows a misaligned intermediate in which the progeny strand has unwound from the template strand and pairing has occurred between the second direct repeat in the template strand in the first direct repeat in the progeny strand. Continued replication would result in resynthesis of 17-bp insert producing the duplication mutation following a second round of DNA replication.

shown in Figure 4E (also see SINDEN *et al.* 1991). However, because these hairpin stems are so short, structure E may not represent a stable secondary structure. A second misaligned intermediate structure F contains a three-way junction, which stabilizes misalignment of a 4-bp direct repeat. Structure F may provide a reasonable template for continued replication. However, it is uncertain just how close the third strand (the hairpin stem) can be to the 3′ end of the progeny strand and still be extendable by DNA polymerase. Deletion readily occurs between 5-bp direct repeats flanking inverted repeats in bacteriophage T4 (PIERCE, KONG and MASKER 1991). It is not known if DNA polymerase would have difficulty extending a 4 base "primer" associated with a three-way junction.

The 17-bp p4 insert has the same base composition as n1, n2 and n3. Considering that it has 8-bp direct repeats and a deletion frequency similar to that for n2 (also with 8-bp direct repeats), this result might seem to argue that the palindrome has little effect in determining the frequency of mutation. However, in all 24 p4 revertants analyzed only the precise deletion was observed. Thus, the palindrome is directing a very specific mutagenic event at a high frequency. The comparison of the results for the p1 and n1 inserts suggests the importance of a hairpin arm in stabilizing and directing deletion between direct repeats. p1 is an asymmetric palindromic insert described previously (TRINH and SINDEN 1991). It can form a near perfect three-way junction where a hairpin stem stabilizes misalignment of a 6-bp direct repeat (Figure 5A). The deletion frequency of the p1 insert is significantly higher than the two other inserts with 6-bp direct repeats: the nonpalindromic insert n1 and the palindromic p3 insert. A possible explanation for the low reversion frequency of the p3 insert is discussed below. Moreover, we have analyzed sequences similar to that for p1, but with longer direct repeats and the Cm^r reversion frequency was increased more than 10-fold (W. A. ROSCHE and R. R. SINDEN, unpublished data).

In summary, a palindromic sequence between direct repeats may strongly direct a very specific mutagenic event. If polymerase pauses or dissociates at the site of inverted repeats (possibly as a result of encountering a hairpin in the template), there exists the possibility for misalignment. Following polymerase dissociation and on breathing, the hairpin conformation has a probability of formation. The formation of the hairpin in the template arm can lead to a stabilized misalignment of the direct repeats.

The formation of a stable secondary structure may also prevent or preclude the formation of a stable misalignment event that can lead to mutation. This may explain the very low frequency of deletion of the p2 and p3 inserts. Sequences p2 and p3 were designed with 5- and 6-bp direct repeats, respectively. One difference that distinguishes p1 from p2 and p3 is the removal of one base pair (from the sequence arrangement of p1) that prevents the precise organization of a perfect three-way junction. To form a three-way junction, an extrahelical G must form in the progeny strand of both p2 and p3 as shown in Figure 5, B and C. Without extruding the extrahelical base the direct repeat in the progeny strand will not properly align with its complementary sequence in the template strand. p1 and p3 share an identical 6-bp direct repeat with a similar spacing of 12 and 11 bp, respectively. The extrahelical base may prevent the formation of a stable misaligned progeny strand that provides a substrate for continued replication. Alternatively, if the misalignment occurs, the structure must be not readily extendable by DNA polymerase.

The low reversion frequencies of the p2 and p3 inserts demonstrate that the presence of the inverted repeat alone is not sufficient for directing a high frequency of deletion. It is known that DNA polymerase will pause at hairpins (WEAVER and DEPAMPHILIS 1984) and that sites of pausing are associated with mutations (KUNKEL 1990; PAPANICOLAOU and RIPLEY 1991). The specific DNA secondary structure that can form may be an important determinant responsible for a particular frequency of deletion.

Duplications between direct repeats: *Effect of nonpalindromic DNA between direct repeats on duplication:*

Duplications between direct repeats can occur by slippage between the first direct repeat in the template strand and the second copy of the direct repeat in the progeny strand following synthesis of the second direct repeat (Figure 6). The slippage is followed by continued synthesis that will lead to a duplication of the direct repeat and intervening DNA. In general, it is believed that duplications occur less frequently than deletions (FARABAUGH *et al.* 1978; ALBERTINI *et al.* 1982; SCHAAPER, DANFORTH and GLICKMAN 1986; SCHAAPER and DUNN 1991). This would seem reasonable since a larger region of DNA must denature to allow the slippage to occur. In the case of the 17- to 18-bp inserts studied here, as many as 24 bp must unwind to permit the misalignment. On the other hand, for deletions to occur only 5–8 bp must unwind to allow misalignment. The results shown in Table 2 show duplication frequencies in the range of 0.2–9 × 10^{-9}.

It is clear that the sequence and symmetry of the DNA between the direct repeats can dramatically influence the frequency of duplication. (In the constructs analyzed here, there appears to be no relationship between the length of the direct repeat and the frequency of duplication). The duplication frequency of the three nonpalindromic inserts was very different. The n1 insert, which contained no internal direct repeats, had the highest frequency of duplication; an estimated frequency of 9.1 × 10^{-9}, which was slightly *higher* than the deletion frequency for this insert. We find it difficult to rationalize a physical basis for this observation in light of the reasoning that unwinding leading to duplication may be less likely than unwinding leading to deletion. The result may suggest that the unwinding event that occurs at the 3′ end of the n1 progeny strand is *always* large enough to include both direct repeats and the region between the direct repeats (*e.g.*, >20 bp). Were this the case, there may be an equal chance for misalignment leading to duplication after replicating the second direct repeat as for deletion after replicating the first direct repeat. The duplication frequency for the n2 insert was nearly 10-fold lower than its corresponding deletion frequency and two- to threefold lower than the duplication frequency for the n1 insert. It is possible that the multiple internal direct repeats provide nucleation events that stop the complete dislocation of the progeny strand relative to the template strand. The n3 insert showed very high frequency of Cm^r reversion in plasmid pJT7. Analysis of 11 revertants failed to detect a duplication event. Alhtough more revertants would need to be analyzed to accurately estimate a duplication frequency, the results clearly show that large and small deletions occur at a much higher frequency than duplications in the n3 construct. Thus, in the n3 construct the multiple internal direct repeats greatly increase the deletion frequency (consistent with the results for the nonpalindromic inserts in pBR325). We have not yet undertaken a sequence analysis of the n3 revertants from pJT7.

In summary, whereas multiple internal direct repeats correlated with an increase in the frequency of deletions between flanking direct repeats, the multiple internal direct repeats correlated with a decrease in the duplication frequency. This may be consistent with the idea that multiple misalignments will occur before the second direct repeat in the progeny strand has completely dislocated and paired with the first direct repeat in the template.

Effect of palindromic DNA between direct repeats on duplication: The frequency of duplications of palindromic DNA between direct repeats was in general very low. In the case of the p4 insert the Cm^r reversion frequency it was about 4 × 10^{-9}, which was eight- to ninefold lower than its deletion frequency in pBR325. However, only one of 24 revertants analyzed had a duplication, resulting in an estimated duplication frequency of about 0.17 × 10^{-9}. This value is more than 150-fold lower than the deletion frequency for this insert. Palindromic insert p5 was designed to provide a three-way junction structure much like that of p1 (Figure 7A, 4). The frequency of Cm^r reversion of this insert in pJT7 was also very low and revertants were not analyzed. The results for the p5 and p4 inserts strongly suggest that the frequency of duplication between direct repeats is very low when the intervening DNA contains inverted repeat symmetry. It is possible for the p4 insert to form a misaligned intermediate containing an unpaired loop of DNA (see Figure 7B, 4). This structure is analogous to those formed by n1, n2 and n3. In spite of this potential structure (which for the nonpalindromic inserts corresponds with a reasonably high duplication frequency of 3–9 × 10^{-9}), the duplication frequency of p4 was very low. This may suggest that the palindromic symmetry of DNA between the direct repeats has a dominant impact on the nature of the mutational event.

When one considers that a large region must unwind to allow duplication to occur (as shown in Figures 6 and 7) it is perhaps not surprising that the frequency of duplication for palindromic inserts is low. This is because although one might consider the formation of a hairpin in the progeny strand to promote slippage that would lead to a duplication (Figure 7, A4 and B5), the template strand can form a hairpin as well. In terms of kinetics and thermodynamics, the probability of hairpin formation should be identical for both the template and progeny strands (in the absence of bound proteins). Thus, if a hairpin forms in the progeny strand it may likely from as well in the template strand. As shown in Figure 7 (structures A5 and B6), this simply results in the pairing of the second copies

Duplication in palindromic inserts

A. pJTp5

1.

```
          ----->            ---->
5'gctcatcggAATTCGTTACGAATTCAaattccgtatgg
3'cgagtagccttaaGCAATGCTTAAGTTTAAggcatacc
          -------> <-------
```

2.

```
          ----->            ---->
5'gctcatcggAATTCGTTACGAATTCAaattc->
3'cgagtagccttaaGCAATGCTTAAGTTTAAggcatacc
          -------> <-------
```

3.

```
          ----->            ---->
          AATTCGTTACGAATTCAaattc->
        cgg------> <-------
      tc
    a
5'gctc
3'cgagtagccttaaGCAATGCTTAAGTTTAAggcatacc
          -------> <-------
```

4.

```
          T
        T  A
         GC
         CG
         TA
         TA
         AT
         AT
         gC---->
5'gctcatcgAaattc->
3'cgagtagccttaaGCAATGCTTAAGTTTAAg
          -------> <-------
```

5.

```
          T
        T  A
         GC
         CG
         TA
         TA
         AT
         AT
         gC
5'gctcatcgAaattc==>
3'cgagtagcTTTAAggcataccgt
         cG
         tA
         tA
         aT
         aT
         GC
         CG
        A  T
         T
```

B. pJTp4

1.

```
          ------->          ------->
5'gctcatcggAATTCCCCATTATGGGGaattccgtatgg
3'cgagtagccttaaGGGGTAATACCCCTTAAggcatacc
          ------------> <------------
```

2.

```
          ------->          ------->
5'gctcatcggAATTCCCCATTATGGGGaattcc->
3'cgagtagccttaaGGGGTAATACCCCTTAAggcatacc
```

3.

```
          ------->          ------->
            AATTCCCCATTATGGGGaattcc->
          gg---------> <------------
        cg--
      tc
    a
5'gctc
3'cgagtagccttaaGGGGTAATACCCCTTAAggcatacc
          ------------> <------------
```

4.

```
          C
        C  C
      C    A
     T      T
     T      T
     A      A
      A    T
       g  G
        gG------>
5'gctcatcGGaattcc->
3'cgagtagccttaaGGGGTAATACCCCTTAAggcatacc
          <-------
```

5.

```
            T
          T  A
           AT
           CG
           CG
           CG
           CG
           Ta
           Ta
5'gctcatcggAAttcc->
3'cgagtagccttaaGGGGTAATACCCCTTAAggcatacc
          <-------
         ------------> <------------
```

6.

```
            T
          T  A
           AT
           CG
           CG
           CG
           CG
           Ta
           Ta
5'gctcatcggAAttcc->
3'cgagtagccttAAggcatacc
           aT
           aT
           GC
           GC
           GC
           GC
           TA
          A  T
           A
```

of the direct repeat of the progeny and template strands. Branch migration will then reform a linear molecule from this cruciform structure since negative supercoiling may be compromised by the nicks and gaps present at the replication fork.

This discussion has focused on the impact of symmetry elements within the intervening DNA on the deletion or duplication between direct repeats. The intervening DNA can have a predominant effect on directing the nature of the mutation. We realize, however, that this may not be the only factor influencing the frequency and spectrum of mutation. There may also be a relationship between sequences flanking the *Eco*RI site and the sequence in the insert. In many cases we have tried to preserve the A + T and G + C content to minimize large changes in the thermal stability of the DNA. However, changes in the relationship between primary sequence may be important in ways we have not considered. For example, different inserts may change the probability of polymerase pausing or dissociation as it traverses the region. In the case of the palindromic inserts, however, potential pausing, which may occur with near equal frequency in all constructs (if polymerase stops at hairpins), does not appear to be the major determinant. Different sequence arrangements may also influence where an Okazaki fragment may be initiated, thus affecting the positioning of the 3′ OH end of a lagging strand fragment. Mutations involving DNA secondary structure may occur at a higher frequency at either the 3′ or 5′ ends of a lagging fragment. Understanding the nature of mutation is clearly complex since so many different variables are likely to influence the frequency and spectrum of mutation. However, the type of analysis described here and by others (SINDEN *et al.* 1991; WESTON-HAFER and BERG 1989, 1991; DASGUPTA, WESTON-HAFER and BERG 1987; PIERCE, KONG and MASKER 1990; BALBINDER, MACVEAN and WILLIAMS 1989; PAPANICOLAOU and RIPLEY 1989, 1991) begins to provide insight into the biological impact of the many variables that are responsible for directing spontaneous mutation.

We thank KAREN ALLEN, EILEEN SCHWAB, JAMES BROWNING and WILBURN COUCH for technical assistance with some of the experiments. This work was supported by grants NP490 from the American Cancer Society and ES 05508 from the National Institutes of Environmental Health Sciences.

LITERATURE CITED

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

ALTON, N. K., and D. VAPNEK, 1979 Nucleotide sequence analysis of the chloramphenicol resistance transposon Tn9. Nature **282:** 864–869.

BALBINDER, E., C. MACVEAN and R. E. WILLIAMS, 1989 Overlapping direct repeats stimulate deletions in specifically designed derivatives of plasmid pBR325 in *Escherichia coli*. Mutat. Res. **214:** 233–252.

BETLACH, M. C., V. HERSHFIELD, L. CHOW, W. BROWN, H. M. GOODMAN and H. W. BOYER, 1976 A restriction endonuclease analysis of the bacterial plasmid controlling the *Eco*RI restriction and modification of DNA. Fed. Proc. **34:** 2037–2043.

BETZ, J. L., and J. R. SADLER, 1981a Variants of a cloned synthetic lactose operator. I. A palindromic dimer lactose operator derived from one strand of the cloned 40-base pair operator. Gene **13:** 1–12.

BETZ, J. L., and J. R. SADLER, 1981b Variants of a cloned synthetic lactose operator. II. Chloramphenicol-resistant revertants retaining a lactose operator in the CAT gene of plasmid pBR325. Gene **15:** 187–200.

BOLIVAR, F., 1978 Construction and characterization of new cloning vehicles, III. Derivatives of plasmid pBR322 carrying unique *Eco*RI sites for selection of *Eco*RI generated recombinant DNA molecules. Gene **4:** 121–136.

COLLINS, J., 1980 The instability of palindromic DNA in *Escherichia coli*. Cold Spring Harbor Symp. Quant. Biol. **45:** 409–416.

DASGUPTA, U., K. WESTON-HAFER and D. E. BERG, 1987 Local DNA sequence control of deletion formation in *Escherichia coli* plasmid pBR322. Genetics **115:** 41–49.

DRAKE, J. W., B. W. GLICKMAN and L. S. RIPLEY, 1983 Updating

FIGURE 7.—Duplication in palindromic inserts. (A)(1) The sequence of the insert in pJTp5 is shown in capital letters with direct repeats indicated by dashed arrows above the sequence and the inverted repeat indicated by the facing arrows below the sequence. (2) The structure following replication of the second direct repeat is shown. To form a duplication, the entire region containing both direct repeats must unwind as shown in structure 3. Once unwinding has occurred, hairpin formation can readily form because the single strand regions contain inverted repeated symmetry. This could result in the formation of a hairpin stem as shown in structure 4, in which second direct repeat in the progeny strand pairs with first direct repeat of the template strand. However, structure 4 may not form at reasonable frequency from the intermediate shown in structure 3. This is because both strands contain the inverted repeat symmetry, and hairpin formation should occur in both strands with equal probability, resulting in a cruciform structure as shown in structure 5. In this cruciform structure, the second direct repeat of progeny strand is paired with the second direct repeat in the template strand. This does not represent a misalignment and would not result in deletion or duplication on replication. Branch migration of structure 5 would simply return the DNA structure to that shown in structure 2. (B) This panel shows structures that may form in the pJTp4 construct. Structure 1 shows the sequence and symmetry elements of the p4 insert. Structure 2 shows a replication intermediate following replication of both direct repeats. Structure 3 shows the unwound intermediate required for misalignment to allow the duplication event. Structure 4 shows a possible misalignment between the second direct repeat of the progeny strand with the first direct repeat of the template strand. In this case, because of the position of the inverted repeat with respect to the direct repeats, a loop of DNA would form much like that observed for the nonpalindromic inserts. Branch migration by 4 bp to the right could produce the perfect three-way junction shown in structure 5 in which a hairpin stem forms and misalignment involves a 4-bp direct repeat rather than the 8-bp repeat shown in structure 4. However, once structure 5 has formed or on formation of structure 4, the template strand (which also contains the inverted repeat) can likewise form a hairpin stem resulting in the formation of a cruciform structure as shown in structure 6. Following branch migration this cruciform could reform structure 2. Continued synthesis from either structures 2 or 6 would not result in a duplication mutation.

the theory of mutation. Am. Sci. **71:** 621–630.

FARABAUGH P. J., U. SCHMEISSNER, M. HOFER and J. H. MILLER, 1978 Genetic studies of the *lac* repressor on the molecular nature of spontaneous hotspots in the *lacI* gene of *Escherichia coli* J. Mol. Biol. **126:** 847–857.

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

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

HANAHAN, D., 1983 Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. **166:** 557–580.

HANAHAN, D., and M. MESELSON, 1983 Plasmid screening at high colony density, pp. 333–342 in *Methods in Enzymology*, Vol. 100, edited by R. WU, L. GROSSMAN and K. MOLDAVE. Academic Press, New York.

HOLMES, D. S., and M. QUIGLEY, 1981 A rapid boiling method for the preparation of bacterial plasmid. Anal. Biochem. **114:** 193–197.

HORWITZ, M. Z., and L. A. LOEB, 1988 An *E. coli* promoter that regulates transcription by DNA superhelix-induced cruciform extrusion. Science **241:** 703–705.

KUNKEL, T. A., 1990 Misalignment-mediated DNA synthesis errors. Biochemistry **29:** 8003–8011.

LILLEY, D. M. J., 1981 *In vivo* consequences of plasmid topology. Nature **292:** 380–382.

LOCKSHON D., and D. A. GALLOWAY, 1986 Cloning and characterization of *oriL2*, a large palindromic DNA replication origin of herpes simplex virus type 2. J. Virol. **58:** 513–521.

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

MAXAM, A. M., and W. GILBERT, 1980 Sequencing end-labeled DNA with base specific chemical cleavage, pp. 499–559 in *Methods in Enzymology*, Vol. 65, edited by L. GROSSMAN and K. MOLDAVE. Academic Press, New York.

MAZIN, A. V., A. V. KUZMINOV, G. L. DIANOV and R. I. SALGANIK, 1991. Mechanisms of deletion formation in *Escherichia coli* plasmids II. Deletions mediated by short direct repeats. Mol. Gen. Genet. **228:** 209–214.

PANAYOTATOS, N., and A. FONTAINE, 1987 A native cruciform DNA structure probed in bacteria by recombinant T7 endonuclease. J. Biol. Chem. **262:** 11364–11368.

PANAYOTATOS, N., and R. D. WELLS, 1981 Cruciform structures in supercoiled DNA. Nature **281:** 466–470.

PAPANICOLAOU, C., and L. S. RIPLEY, 1989 Polymerase-specific differences in the DNA intermediates of frameshift mutagenesis. *In vitro* synthesis errors of *Escherichia coli* DNA polymerase I and its large fragment derivative. J. Mol. Biol. **207:** 335–353.

PAPANICOLAOU, C., and L. S. RIPLEY, 1991 An *in vitro* approach to identifying specificity determinants of mutagenesis mediated by DNA misalignments. J. Mol. Biol. **221:** 805–821.

PIERCE, J. C., D. KONG and W. MASKER, 1991 The effect of the length of direct repeats and the presence of palindromes on deletion between directly repeated DNA sequences in bacteriophage T7. Nucleic Acids Res. **14:** 3901–3905.

PRENTSKI, P., F. KARCH, S. IIDA and J. MEYER, 1981 The plasmid cloning vector pBR325 contains a 482 base-pair-long inverted duplication. Gene **14:** 289–299.

RIPLEY, L. S., 1982 Model for the participation of quasi-palindromic DNA sequences in frameshift mutation. Proc. Natl. Acad. Sci. USA **79:** 4128–4132.

RIPLEY, L. S., 1991 Frameshift mutation: determinants of specificity Annu. Rev. Genet. **24:** 189–213.

RIPLEY, L. S., and B. W. GLICKMAN, 1983 Unique self-complementarily of palindromic sequences provides DNA structural intermediates for mutation. Cold Spring Harb. Symp. Quant. Biol. **47:** 851–861.

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

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

SCHAAPER, R. M., and R. L. DUNN, 1991 Spontaneous mutation in the *Escherichia coli lacI* gene. Genetics **129:** 313–326.

SCOTT, J. R., 1984 Regulation of plasmid replication. Microbiol. Rev. **48:** 1–23.

SINDEN, R. R., J. O. CARLSON and D. E. PETTIJOHN, 1980 Torsional tension in the DNA double helix measured with trimethylpsoralen in living *E. coli* cells: analogous measurements in insect and human cells. Cell **21:** 773–783.

SINDEN, R. R., S. S. BROYLES and D. E. PETTIJOHN, 1983 Perfect palindromic *lac* operator DNA sequence exists as a stable cruciform structure in supercoiled DNA *in-vitro* but not *in-vivo*, Proc. Natl. Acad. Sci. USA **80:** 1797–1801.

SINDEN, R. R., G. ZHENG, R. G. BRANKAMP and K. N. ALLEN, 1991 On the deletion of inverted repeated DNA in *Escherichia coli*: effects of length, thermal stability, and cruciform formation *in vivo*. Genetics **129:** 991–1005.

SINGER, B. S., and J. WESTLYE, 1988 Deletion formation in Bacteriophage T4. J. Mol. Biol. **202:** 233:243.

STREISINGER, G., and J. OWEN, 1985 Mechanisms of 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 Harb. Symp. Quant. Biol. **31:** 77–84.

SZOSTAK, J. W., J. I. STILES, B. K. TYE, D. CHIU, F. SHERMAN and R. WU, 1980 Oligonucleotide hybridization, pp. 419 in *Methods in Enzymology*, Vol. 68, edited by R. WU. Academic Press, New York.

TRINH, T. Q., 1990 The role of DNA secondary structure in deletion and duplication between direct repeats in *E. coli*: differential mutagenesis between the leading and lagging strands. Ph. D. Thesis, University of Cincinnati.

TRINH, T. Q., and R. R. SINDEN, 1991 Preferential DNA secondary structure mutagenesis in the lagging strand of replication in *E. coli*. Nature **352:** 544–547.

WEAVER, D. T., and M. L. DEPAMPHILIS (1984) The role of palindromic and non-palindromic sequences in arresting DNA synthesis *in vitro* and *in vivo*. J. Mol. Biol. **180:** 961–986.

WEISLANDER, L., 1979 A simple method to recover intact high molecular weight RNA and DNA after electrophoretic separation in low gelling temperature agarose gels. Anal. Biochem. **98:** 305–309.

WELLER, S. K., A. SPADARO, J. E. SCHAFFER, A. W. MURRAY, A. M. MAXAM and P. A. SCHAFFER, 1985 Cloning, sequencing, and functional analysis of ori_L, a herpes simplex virus type 1 origin of DNA synthesis. Mol. Cell. Biol. **5:** 930–942.

WESTON-HAFER, K., and D. E. BERG, 1989 Palindromy and the location of deletion endpoints in *Escherichia coli*. Genetics **121:** 651–658.

WESTON-HAFER, K., and D. E. BERG, 1991 Deletions in plasmid pBR322: replication slippage involving leading and lagging strands. Genetics **127:** 649–655.

WILLIAMS W. L., and U. R. MÜLLER, 1987 Effects of palindrome size and sequence on genetic stability in the bacteriophage ΦX174 genome. J. Mol. Biol. **196:** 743–755.

ZAGURSKY, R. J., K. BAUMEISTER, N. LOMAX and M. L. BARMAN, 1985 Rapid and easy sequencing of large linear double-stranded DNA and supercoiled plasmid DNA. Gene Anal. Tech. **2:** 89–94.

ZHENG, G., T. J. KOCHEL, R. W. HOEPFNER, S. E. TIMMONS and R. R. SINDEN, 1991 Torsionally tuned cruciform and Z-DNA probes for measuring unrestrained supercoiling at specific sites in DNA of living cells. J. Mol. Biol. **221:** 107–129.

Communicating editor: N. R. DRINKWATER