On the Deletion of Inverted Repeated DNA in *Escherichia coli*: Effects of Length, Thermal Stability, and Cruciform Formation *in Vivo*

Richard R. Sinden, Guoxing Zheng, Robert G. Brankamp and Karen N. Allen

Department of Molecular Genetics, Biochemistry, and Microbiology, University of Cincinnati College of Medicine, Cincinnati, Ohio 45267–0524

Manuscript received March 23, 1991
Accepted for publication August 19, 1991

ABSTRACT

We have studied the deletion of inverted repeats cloned into the *Eco*RI site within the CAT gene of plasmid pBR325. A cloned inverted repeat constitutes a palindrome that includes both *Eco*RI sites flanking the insert. In addition, the two *Eco*RI sites represent direct repeats flanking a region of palindromic symmetry. A current model for deletion between direct repeats involves the formation of DNA secondary structure which may stabilize the misalignment between the direct repeats during DNA replication. Our results are consistent with this model. We have analyzed deletion frequencies for several series of inverted repeats, ranging from 42 to 106 bp, that were designed to form cruciforms at low temperatures and at low superhelical densities. We demonstrate that length, thermal stability of base pairing in the hairpin stem, and ease of cruciform formation affect the frequency of deletion. In general, longer palindromes are less stable than shorter ones. The deletion frequency may be dependent on the thermal stability of base pairing involving approximately 16–20 bp from the base of the hairpin stem. The formation of cruciforms *in vivo* leads to a significant increase in the deletion frequency. A kinetic model is presented to describe the relationship between the physical-chemical properties of DNA structure and the deletion of inverted repeats in living cells.

INVERTED repeated sequences are found naturally in prokaryotic and eukaryotic DNA. Many inverted repeats reflect dimeric protein binding sites or specific sequences that may be involved in the initiation of DNA replication. The sequence arrangement of inverted repeated, or palindromic, DNA allows the sequence to exist in one of two alternative forms: an interstrand base paired linear form, or a cruciform structure with intrastrand base pairing within the individual strands of DNA.

Although inverted repeats exist naturally in the DNA of many organisms, it is difficult to maintain inverted repeats greater than 150 bp in length in plasmid DNA in *Escherichia coli*. LILLEY (1981) demonstrated that it was impossible to clone long inverted repeats into *E. coli*. The inability to clone long inverted repeats and the genetic instability associated with inverted repeats have been reported by a large number of investigators (COLLINS 1980; BETZ and SADLER 1981a; LEACH and STAHL 1983; HAGAN and WARREN 1983; MIZUUCHI *et al.* 1982; WELLER *et al.* 1985; LOCKSHON and GALLOWAY 1986). The instability of inverted repeats may be due to any one of a number of possibilities [see RIPLEY (1990) for review] . First, inverted repeats may be deleted as a result of intermolecular or intramolecular recombination (WARREN and GREEN 1985; SINGER and WESTLYE 1988). Second, the deletion of inverted repeats may result from the formation of a cruciform structure and the subsequent processing by nucleases (DRAKE, RIPLEY and GLICKMAN 1983; LEACH and STAHL 1983; GLICKMAN and RIPLEY 1984). Third, in the case of inverted repeats containing direct repeats, misalignment between the direct repeats may be stabilized by the formation of a hairpin or other DNA secondary structures (RIPLEY and GLICKMAN 1983; DRAKE, RIPLEY and GLICKMAN 1983; GLICKMAN and RIPLEY 1984). Cloned inverted repeats are, by the nature of cloning, flanked by direct repeats. Thus, the instability of cloned inverted repeats may reflect a mechanism involving the deletion between direct repeats (STREISINGER *et al.* 1966; ALBERTINI *et al.* 1982). The model involving slipped misalignment directed by a hairpin arm has been supported by the results from a number of investigations (WILLIAMS and MÜLLER 1987; SINGER and WESTLYE 1988; BALBINDER, MACVEAN and WILLIAMS 1989; WESTON-HAFER and BERG 1989, 1991).

Many different palindromic sequences have been shown to adopt a cruciform structure in negatively supercoiled DNA *in vitro* (MIZUUCHI *et al.* 1982; LILLEY 1980; PANAYOTATOS and WELLS 1981; SINDEN, BROYLES and PETTIJOHN 1983; COUREY and WANG 1983). The rate of the transition from the linear form of the inverted repeat to the cruciform structure varies as a function of base composition, temperature, salt concentration, and superhelical density (SINGLETON 1983; SINDEN and PETTIJOHN 1984; COUREY and

WANG 1988; SULLIVAN and LILLEY 1986, 1987; ZHENG and SINDEN 1988). Under certain conditions, flanking DNA sequences can have a dominant influence on the cruciform transition *in vitro* (SULLIVAN and LILLEY 1986; SULLIVAN, MURCHIE and LILLEY 1988). GELLERT, O'DEY and MIZUUCHI (1983) and COUREY and WANG (1983) suggested that cruciforms may not exist *in vivo* from analyses of inverted repeats that required high temperatures (50–60°) and high superhelical densities ($\sigma < -0.05$) for cruciform formation. An initial assay for cruciforms *in vivo* suggested that cruciforms were not likely to exist in living *E. coli* cells (SINDEN, BROYLES and PETTIJOHN 1983). However, recent quantitative analysis of inverted repeats "torsionally tuned" to form cruciforms at low temperatures and at low superhelical densities has shown that cruciforms can exist at levels as high as 50% in plasmids in living *E. coli* cells (ZHENG *et al.* 1991). Other indirect evidence for the existence of cruciforms *in vivo* has come from analysis of the nuclease sensitivity of DNA *in vivo* (PANAYOTATOS and FONTAINE 1987) and from the effects on transcription from a palindromic promoter (HORWITZ and LOEB 1988). Cruciforms have also been suggested to exist in eukaryotic nuclei since cruciform-specific antibodies bind to nuclei at certain times during the cell cycle (WARD *et al.* 1990).

A cruciform structurally resembles a Holliday recombination intermediate. Enzymes that recognize and cleave cruciform structures have been identified and purified from both eukaryotic and prokaryotic sources (MIZUUCHI *et al.* 1982; DEMASSEY, WEISBERG and STUDIER 1987; WEST and KÖRNER 1985; TAYLOR and SMITH 1990). If a cruciform forms *in vivo* it may be recognized as a recombination intermediate and cut by a Holliday structure resolvase. This could explain the instability of inverted repeats and the difficulty in cloning inverted repeats greater than 150 bp in *E. coli*. However, we have shown that cruciforms can exist at very high levels *in vivo* (ZHENG *et al.* 1991), suggesting that cruciforms in *E. coli* are not obligatorily cut by Holliday structure resolvases. The results of TAYLOR and SMITH (1990) suggest that the RecBCD nuclease, which can cut cruciforms *in vitro*, might only cut cruciforms that it participates in forming. Since RecBCD requires a free end to get on DNA, cruciform formation in a plasmid or within a topological domain might not be susceptible to RecBCD nuclease cleavage.

This report presents a detailed analysis of the relationship between the DNA sequence, the rates and superhelical density dependence of cruciform formation, the existence of cruciforms *in vivo*, and the genetic stability of the various inverted repeats in *E. coli*. To facilitate analysis of the mutagenic potential of inverted repeats and to minimize influences of neighboring DNA sequence, we have analyzed a number of inverted repeats inserted into the same site of the CAT gene in plasmid pBR325. This analysis provides insight into the relationship between the formation of hairpins and cruciforms *in vivo*, the stability of such structures, and the molecular mechanisms involved in the deletion of inverted repeated DNA sequences.

MATERIALS AND METHODS

Bacteria and growth conditions: Bacterial strains used were HB101 (F⁻ *lac*Y1 *mcr*B *mrr hsd*S20 *rec*A13 *leu*B6 *ara*-14 *pro*A2 *lac*Y1 *gal*K2 *xyl*-5 *mtl*-1 *rps*L20 *sup*E44), JTT1 (*pyr*F287 *nir*A *icl*R7 *gal*25 *rsp*L195 *rec*A56), RS2 (as JTT1, with *top*A10), W3110 (λ⁻, IN(rrnD-rrnE) and SD108 (derived from W3110 with *acr*A13 *trp*E65 *pyr*F286 *gyr*B225). Strains JTT1, RS2 and SD108 were described by STERNGLANZ *et al.* 1981 and DINARDO *et al* 1982. Cells were grown in K medium or Luria broth. K medium consisted of M9 buffer (1 g NH_4Cl, 5.8 g Na_2HPO_4 and 3.0 g KH_2PO_4 per liter of distilled water) plus (per liter) 0.03 g $MgSO_4$, 0.07 g $CaCl_2$, 10 g glucose and 10 g casamino acids (Difco). Luria broth consisted of 10 g Bacto-tryptone, 10 g NaCl and 5 g yeast extract per liter of distilled water. For reversion analyses, cells were plated on Luria agar plates or X-gal plates (described by BETZ *et al.* 1986) containing 25 μg/ml tetracycline (Tet) for determination of the viable cell number and on plates containing 25 μg/ml chloramphenicol (Cam) and tetracycline for determination of the number of Cam^r revertants.

Plasmids and inverted repeats: The DNA sequence of the nonpalindromic and palindromic inserts used are listed in Figures 1 and 2. These sequences were inserted into the *Eco*RI site within the chloramphenicol transacetylase (CAT) gene of plasmid pBR325. Inverted repeats F1, F2, S1, S2 and S3B were described previously (ZHENG and SINDEN 1988). These inverted repeats were named historically. The S series was so named because these inverted repeats were chemically synthesized. The F series was so named because they were derived from a fragment of SV40 DNA. The SF series is a hybrid of the center of F1 and F2 inserted into S1. The npF series was so named because the centers of these F series derived sequences are *non*palindromic. pBRS1F1 was constructed by inserting the sequence CCCGGCCGCGGCCGGG into the central ApaI site of inverted repeat pBRS1 (the S1 inverted repeat cloned into pBR325). pBRS1F2 was constructed by inserting the sequence CCCATATATATATGGG into the central ApaI site of inverted repeat pBRS1. Inverted repeats F6S, F8S, F10S, F12S, F14S and F14C were constructed by replacing the central 16-bp *Apa*I fragment within the F2 inverted repeat with the central 20-bp *Apa*I fragments shown in Figure 1. These inverted repeats were cloned into pBR325 to make plasmids pBRF6S, pBRF6S, pBRF10S, pBRF12S, pBRF14S and pBRF14C. All sequences were confirmed by DNA sequencing.

Measurement of reversion frequencies: For reversion analyses, cells were grown with shaking at 37° in 10–25 ml of media that was inoculated with 1–3 ml of an overnight culture. The overnight culture, which was started from a single colony or a frozen stock culture of cells, was grown without shaking at 37° to an $OD_{650} \approx 1$. For experiments involving F14C and F6S in strain RS2, frozen cultures were used to inoculate the overnight cultures to avoid a high frequency of pretest reversion. For the S, SF, F and npF

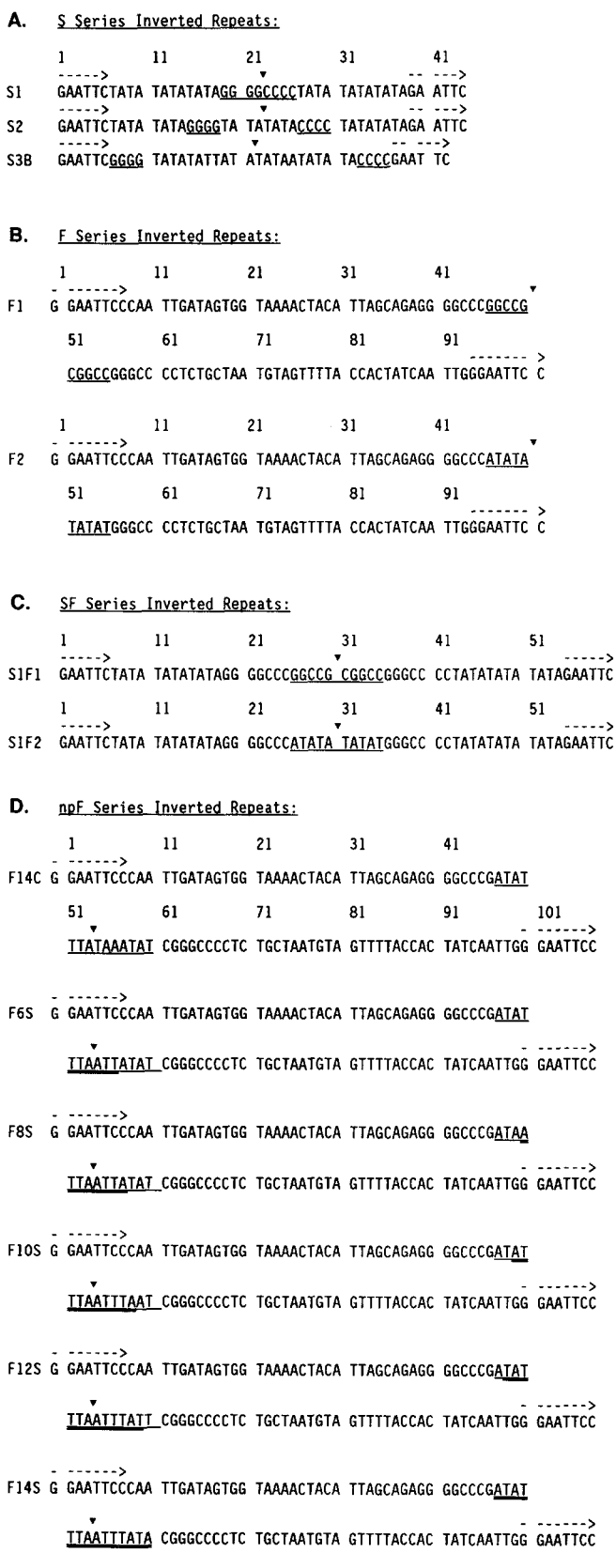

FIGURE 1.—DNA sequences of the S, F, SF and npF series inverted repeats. **A**, Shown above are the S series inverted repeats (ZHENG and SINDEN 1988). The G_4 and C_4 blocks are underlined to emphasize the different positions of these sequences. In all sequences the flanking direct repeats are indicated by the dashed overline with an arrow. The direct repeats for the S series are the, 6-bp *Eco*RI sites. The centers of the perfect inverted repeats are indicated with a ▼. **B**, In the F series inverted repeats (ZHENG and SINDEN 1988) the central 10 bp of G + C and A + T rich sequence are underlined. The numbering in the F and npF series is from the G of the *Eco*RI site although the preceding G, which is part of the 8-bp direct repeat (overlined arrow) is shown. **C**, The SF series inverted repeats contain the same 10-bp G + C and A + T rich centers found in the F series inverted repeats. The SF series inverted repeats are flanked by 6 bp direct repeats of the *Eco*RI site. **D**, In the npF series inverted repeats the central 14- bp A + T rich sequences are underlined. In F6S, F8S, F10S, F12S and F14S the central 6, 8, 10, 12 and 14 bp are nonpalindromic, respectively. This is indicated by double underlining. The doubly underlined center sequences have mirror repeat symmetry. The npF series inverted repeats are flanked by 8-bp direct repeats.

series inverted repeats (Table 1), overnight cultures (grown in Luria broth or K medium) were diluted to an OD_{650} = 0.05–0.1 and the culture was grown at 37° with shaking at 250 RPM until an OD_{650} = 0.5–0.8 was reached. These cells were collected by centrifugation, serially diluted and plated on Tet plates for viable cells counts. An equivalent of 5 ml of the original culture was plated on Tet/Cam plates for determination of the number of Cam^r revertants. For plasmids that gave high frequencies of deletion (F14C, for example), dilutions of the culture were plated on the Tet/Cam plates to give about 200 revertants per plate. This assay is similar to that described by BALBINDER (1988) and BALBINDER, MACVEAN and WILLIAMS (1989). For the experiments involving the *lac* operator inverted repeats and the nonpalindromic inserts (Table 2) overnight cultures grown in K medium were plated directly for the reversion analyses.

Plasmid DNA purification and sequence analysis: Plasmid DNAs were purified using a modified procedure of that described by HOLMES and QUIGLEY (1981). Briefly, cell cultures were quickly chilled to 4°, washed once with chilled M9 buffer, and resuspended in 400 μl of 8% sucrose, 5% Triton X-100, 50 mM EDTA, 50 mM Tris, pH 8.1. An aliquot of 10 μl of lysozyme (10 mg/ml) was added, and the sample was incubated 1 min at room temperature and then heated in a boiling water bath for 1 min. Then 0.4 ml phenol:$CHCl_3$ (1:1) and 0.4 ml NTE buffer (0.1 M NaCl, 10 mM Tris pH 7.6, 5 mM EDTA) were added, the tube inverted several times, and the DNA was extracted into the aqueous phase by incubation for 30 min at 65°. The samples were centrifuged, the aqueous phase was removed and extracted three times with phenol, three times with $CHCl_3$:isoamylalcohol (24:1). The DNA was precipitated by the addition of one volume of isopropanol. The DNA was collected by centrifugation, redissolved in 200 μl TEN buffer (10 mM Tris pH 7.6, 50 mM NaCl, 1 mM EDTA) and digested with 5 units/ml T1 RNAse and 10 μg/ml pancreatic RNAse at 37° for 60 min. The DNA was reextracted with both phenol and $CHCl_3$:isoamylalcohol and precipitated by the addition of two volumes of ethanol.

The DNA sequence of the revertants was analyzed at three levels. First, revertant plasmids were analyzed on agarose gels (SINDEN and PETTIJOHN 1984) to determine if the revertant plasmid was a monomer or dimer. Dimer plasmids were not analyzed further. Monomer plasmids were cut with *Alu*I and analyzed on a 20 × 10 × 0.15 cm 5% polyacrylamide gel in 40 mM Tris borate buffer (pH 8.3) (SINDEN, BROYLES and PETTIJOHN 1983). This analysis could resolve the size of the original 129-bp *Alu*I fragment containing the *Eco*RI site at a resolution of ±2–3 bp. Since

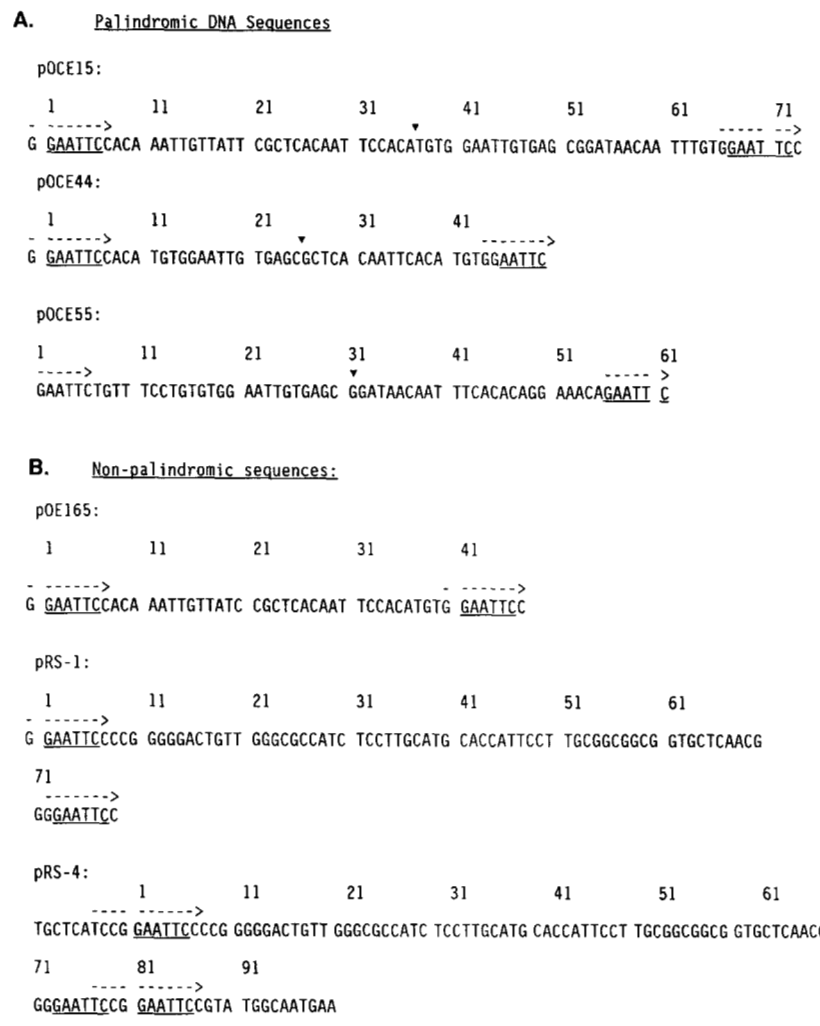

FIGURE 2.—*Lac* operator and other DNA sequences. For all sequences the *Eco*RI sites are underlined, the direct repeats are overlined with a dashed arrow, and the sequences are numbered from the G of the first *Eco*RI site. **A**, pOCE15 contains a 66-bp palindromic *lac* operator sequence (BETZ and SADLER 1981a). pOCE44 contains a 44-bp perfectly symetrical *lac* operator (SADLER, SASMOR and BETZ 1983). pOCE55 contains a 55-bp quasi-palindromic *lac* operator (SADLER and TECKLENBERG 1981). **B**, pOCE165 contains a 40-bp fragment containing the natural *lac* operator sequence (SADLER *et al.* 1978). pRS-1 contains a 74-bp insert consisting of the 64-bp *Hae*III fragment from pBR322 ligated to 5′GGAATTCC *Eco*RI linkers (BALBINDER, MACVEAN and WILLIAMS 1989). pRS-4 contains the same 64-bp *Hae*III fragment as pRS-1 but it is cloned with two copies of the *Eco*RI linker at the 3′ end of the fragment (BALBINDER, MACVEAN and WILLIAMS 1989).

plasmids exist in multiple copies in *E. coli*, a reversion event in one plasmid should confer chloramphenicol resistance to the cell. Thus, a mixed plasmid population might be observed. This was frequently evident on analysis of restriction digests or DNA sequencing from a newly reverted colony. To obtain a pure culture of a revertant for DNA sequence analysis, plasmid from revertant cells was retransformed, and DNA purified from resulting Cam^r colonies was sequenced. Plasmids subjected to sequence analysis were analyzed by the dideoxy sequencing method basically as described previously (KOCHEL and SINDEN 1988). More than 24 independent revertants were analyzed for the results shown in Table 2. Fewer revertants were examined for the results shown in Table 1. Six to twelve revertants were analyzed for the S series sequences and about three to six, for most of the npF series sequences.

Analyses of DNA yields have not revealed observable differences (less than a factor of two) in plasmid copy numbers for plasmids carrying the *lac* operator, S, SF or F series inverted repeats. Within the npF series inverted repeats, there was occasionally a decrease in the yield of plasmid as the instability of the inverted repeat increased, consistent with the observation of WARREN and GREEN (1985). In these experiments, in which plasmid was purified from cells incubated at stationary phase for 12–24 hr, the differences were less than a factor of five for F14C and F14S. We have not done a careful analysis of copy number, in specific reversion analyses, to correct deletion frequencies. A lower copy number, would result in an underestimation of the effect of palindromic sequence on deletion. However, this does not change (in fact it strengthens) the conclusions from the npF series inverted repeats.

RESULTS AND DISCUSSION

We have analyzed reversion frequencies for a large number of inverted repeats in several genetic backgrounds. Data are shown in Tables 1, 2 and 5. In summary, we have identified several factors that appear to be important in defining the frequency of deletion of inverted repeats. (1) The thermodynamic stability of base pairing in the hairpin arm is important in defining the frequency of deletion. Within perhaps the first 20 bp of a hairpin stem, a higher deletion frequency correlated with increasing stability of base pairing. (2) Within a series of inverted repeats of equal length, both the rate of cruciform formation (k_c) and the superhelical density dependence of cruciform formation (σ_c) effect the frequency of deletion. Inverted repeats with a rapid rate of cruciform formation at a low level of negative supercoiling have a higher rate of deletion. (3) The existence of cruciforms *in vivo* is very important in determining the frequency of deletion. The formation of cruciforms *in vivo* leads to an

TABLE 1

Cam^r reversion frequencies for the S, F, SF and npF series inverted repeats

Insert	Size of inverted repeat (bp)	Size of direct repeat (bp)	Strain			Ratio: RS2/JTT1
			HB101[a]	JTT1[b]	RS2 *topA10*[b]	
S Series	42	6				
S1			1.15×10^{-10}			
S2			1.77×10^{-9}			
S3B			1.63×10^{-8}			
SF Series	60	6				
S1F1				7.4×10^{-9}	1.76×10^{-8}	2.4
S1F2				1.37×10^{-8}	2.5×10^{-8}	1.8
F Series	100	8				
F1			6.2×10^{-7}			
F2			1.45×10^{-6}			
npF Series	106	8				
F14C				6.37×10^{-8}	9.25×10^{-5}	1450.0
F6S				9.13×10^{-8}	6.12×10^{-6}	67.0
F8S				3.98×10^{-8}	6.93×10^{-7}	17.4
F10S				6.86×10^{-8}	4.26×10^{-7}	6.2
F12S				6.14×10^{-8}	5.3×10^{-7}	8.6
F14S				5.29×10^{-8}	3.45×10^{-7}	6.5

[a] Cells grown to late log phase in Luria broth and plated for reversion analysis.

[b] Cells grown to mid log phase in K media and plates for reversion analysis. The reversion frequencies are the average, each from four to ten independent experiments. In all cases when individual series of inverted repeats were compared, all analyses were done at the same time to ensure that all cultures experienced identical growth conditions.

increase in the deletion frequency. (4) The length of the inverted repeat appears to be important. In general, a higher the deletion frequency was observed for longer inverted repeats. This conclusion can be made for selected cases in which the length of the flanking direct repeats are identical. The length of the direct repeat flanking the inverted repeat is likely important although constructs were not designed to test this specifically. In several cases examined, longer lengths of both the inverted repeat and the flanking direct repeats correlated with an increased frequency of deletion.

Measurement of Cam^r reversion frequencies: The system for measuring the frequency of deletion of inverted repeats utilizes the CAT gene in plasmid pBR325. This system has been used by BETZ and SADLER (1981b) for selection of *lac* operator mutants and by BALBINDER, MACVEAN and WILLIAMS (1989) for analysis of one of the *lac* operator inverted repeats and two of the nonpalindromic inserts studied here. We have inserted inverted repeats into the *Eco*RI site in the CAT gene. This generates an inverted repeat that includes the two *Eco*RI sites. This can be considered an inverted repeated DNA sequence flanked by direct repeats since the two *Eco*RI sites constitute two direct repeats. Insertion of an inverted repeat results in a chloramphenicol sensitive phenotype (Cam^s) due to the introduction of a frameshift mutation or, for in-frame insertions, the addition of sufficient amino acids to disrupt the CAT gene. The Cam^r reversion frequencies were determined by plating a known number of cells on plates containing chloramphenicol and counting the viable Cam^r revertant colonies. Table 2 provides examples of the frequencies obtained from individual experiments. In most cases the numbers varied within a factor of two. Occasionally a high frequency was obtained, probably due to the occurrence of a "jackpot" in the inoculating culture. In the case of the *lac* operator inverted repeats, cells harboring the original plasmids containing a *lac* operator stained blue on X-gal plates. Reversion to Cam^r was accompanied by formation of white (or blue and white sectored) colonies indicative of the loss of the *lac* operator sequence (BETZ and SADLER 1981b).

On DNA sequence analysis of spontaneous Cam^r revertants, in all cases examined, the revertant plasmids contain a restored, intact CAT gene. The inverted repeat and one copy of the *Eco*RI site had been precisely deleted. In-frame insertions in the *Eco*RI site of the CAT gene can produce a functional CAT protein if the insertions are less than 51 bp (BETZ and SADLER 1981b; T. Q. TRINH and R. R. SINDEN, unpublished results). Thus, deletion of the entire inverted repeat is not necessary to produce a Cam^r phenotype. However, the precise deletion of the entire inverted repeat occurred in all cases examined.

The thermodynamic stability of base pairing in the hairpin arm effects the frequency of deletion of inverted repeats: The thermal stability of the cruciform or hairpin arm may be important in mutagenesis. The deletion frequency appeared to increase with increasing stability of intrastrand base pairing. Within

TABLE 2

Cam^r reversion frequencies for *lac* operator palindromic and nonpalindromic inserts in the CAT gene of pBR325

Plasmid (insert)	Strain HB101	Strain JTT1	Strain RS2 *top*A10	Ratio RS2/JTT1	Strain W3110	Strain SD108 *gyr*B	Ratio SD108/W3110
A. Nonpalindromic inserts							
pRS-1 (64 bp)							
	5.5×10^{-8}	5.6×10^{-8}	7.1×10^{-8}	1.3	5.6×10^{-8}	2.8×10^{-7}	4.8
	3.9×10^{-8}	1.4×10^{-7}	2.2×10^{-7}	1.5			
	9.2×10^{-8}	1.5×10^{-7}	3.7×10^{-7}	2.5			
	8.2×10^{-8}	1.9×10^{-7}	2.0×10^{-7}	1.1			
	(6.7×10^{-8})[a]	(1.4×10^{-7})	(2.0×10^{-7})	(1.4)			
pRS-4 (64-bp fragment with 3 *Eco*R1 linkers)							
	1.6×10^{-6}	5.4×10^{-6}	4.9×10^{-6}	0.9	3.0×10^{-6}		
	2.4×10^{-6}	5.5×10^{-6}	7.6×10^{-6}	1.4	3.8×10^{-6}		
	3.8×10^{-6}	5.0×10^{-6}	6.3×10^{-6}	1.2			
	1.2×10^{-6}	1.4×10^{-6}	1.8×10^{-6}	1.3			
	(2.3×10^{-6})	(4.3×10^{-6})	(5.2×10^{-6})	(1.2)	(3.4×10^{-6})		
pOE165 (40-bp *lac* operator)							
	8.4×10^{-8}	3.9×10^{-8}	1.3×10^{-8}	0.3	4.2×10^{-8}	6.3×10^{-8}	1.4
	9.3×10^{-8}	3.6×10^{-8}	1.8×10^{-8}	0.5	5.6×10^{-8}	3.1×10^{-8}	0.7
	3.8×10^{-8}	1.2×10^{-7}	2.1×10^{-8}	0.2	2.9×10^{-8}	1.8×10^{-8}	0.6
	1.4×10^{-7}						
	1.7×10^{-7}						
	(1.1×10^{-7})	(6.5×10^{-8})	(1.7×10^{-8})	(0.3)	(4.2×10^{-8})	(3.7×10^{-8})	(0.9)
B. Palindromic *lac* inserts							
pOCE15 (66-bp perfect palindrome)							
	6.4×10^{-7}	2.4×10^{-7}	5.9×10^{-6}	25	3.5×10^{-7}	3.1×10^{-8}	0.09
	6.6×10^{-7}	1.1×10^{-7}	1.4×10^{-5}	130	4.1×10^{-7}	6.6×10^{-8}	0.16
	5.0×10^{-7}	1.0×10^{-7}	3.1×10^{-6}	31			
	4.5×10^{-7}	1.3×10^{-7}	3.3×10^{-6}	25			
	5.2×10^{-7}						
	(5.5×10^{-7})	(1.5×10^{-7})	(6.6×10^{-6})	(44)	(3.8×10^{-7})	(4.9×10^{-8})	(0.12)
pOCE44 (44-bp perfect palindrome)							
	1.1×10^{-6}	6.6×10^{-7}	6.4×10^{-6}	9.7	1.1×10^{-6}	5.8×10^{-7}	0.5
	1.1×10^{-6}	1.8×10^{-6}	1.1×10^{-5}	6.3	5.6×10^{-7}	4.0×10^{-7}	0.7
	7.7×10^{-7}	9.5×10^{-7}	3.0×10^{-5}	32	1.5×10^{-6}	8.7×10^{-7}	0.6
		1.3×10^{-6}	1.1×10^{-5}	8.5			
		1.2×10^{-6}	1.4×10^{-5}	12			
	(9.9×10^{-7})	(1.2×10^{-6})	(1.4×10^{-5})	(12)	(1.1×10^{-6})	(6.2×10^{-7})	0.6
pOCE55 (55-bp imperfect palindrome)							
	9.4×10^{-7}	1.4×10^{-7}	3.8×10^{-6}	27	10×10^{-7}	5.9×10^{-8}	0.06
	1.2×10^{-6}	3.3×10^{-7}	2.5×10^{-5}	76	5.0×10^{-7}	2.1×10^{-7}	0.4
	5.8×10^{-7}	1.9×10^{-6}	2.9×10^{-5}	15	1.8×10^{-7}	9.6×10^{-8}	0.5
		8.5×10^{-7}	4.2×10^{-5}	49	5.6×10^{-7}	1.2×10^{-7}	0.2
		7.7×10^{-7}	1.9×10^{-5}	25			
	(9.1×10^{-7})	(7.9×10^{-7})	(2.3×10^{-5})	(29)	(5.6×10^{-7})	(1.2×10^{-7})	(0.2)
pOCE17 (two copies of the 66-bp perfect palindrome)							
	4.9×10^{-3}						
	9.3×10^{-3}						
	1.4×10^{-3}						
	(5.2×10^{-3})						

[a] Average values are shown in parentheses.

the S series, inverted repeat S3B, which forms cruciforms most readily *in vitro* (ZHENG and SINDEN 1988), was the most prone to deletion (Table 1). As discussed below, the rate and superhelical density dependence of cruciform formation are important as they may affect the level of cruciforms *in vivo*. The high frequency of deletion of S3B, however, may not be entirely dependent on the rate of cruciform formation but may result from an increased thermal stability within the S3B hairpin arm compared to S1 and S2. The overall thermal stability of the linear form of the inverted repeats S1, S2 and S3B are nearly identical since they contain essentially identical base composition. However, as shown in Table 3 and Figure 3, the stability of the hairpin arms differs significantly. This variation is due to the arrangement of the G_4 and C_4 blocks (underlined in Figure 1) and the $(AT)_n$ tracts. S3B has a long AT rich center which creates the most

TABLE 3

Length and T_m values of hairpin stems

Inverted repeat	Length of the hairpin stem (bp)[a]	T_m (°C)[b]
S1	21	56.6
S2	21	62.6
S3B	20	64.5
S1F1	29	68.7
S1F2	29	62.7
F1	48	71.3
F2	48	68.2
F14C	52	68.2
F6S	50	68.7
F8S	49	68.9
F10S	48	69.6
F12S	47	69.9
F14S	46	69.5
66 bp *lac*	35	68.6
55 bp *lac*[c]	26	69.6
44 bp *lac*	24	70.2

[a] Length of the hairpin stem includes a G/C base pair flanking the *Eco*RI site but does not include the 4 bp which will be unpaired at the tip (loop) of the hairpin stem.

[b] T_m values are calculated as described by ZHENG and SINDEN (1988).

[c] Calculations for the 55-bp *lac* operator inverted repeat do not include the G:T mismatch in the stem nor the 9 bp at the tip of the loop which would probably not be involved in base pairing.

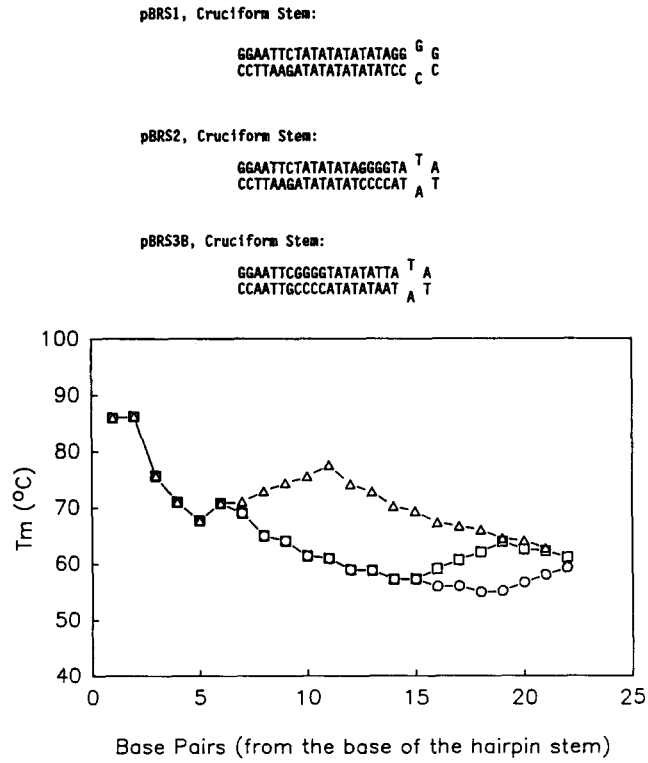

FIGURE 3.—Thermal stability of the S series hairpin arms. Top, The cruciform stems of the S series inverted repeats (S1, S2 and S3B) are shown. Each sequence includes a palindromic G/C pair flanking the *Eco*RI site. Bottom, The T_m values were calculated as described previously (ZHENG and SINDEN 1988) and are plotted as a function of length up the hairpin stem. The T_m calculations do not include the four bases that will be unpaired at the loop of the hairpin stem. ○, S1; □, S2; △, S3B.

thermally stable base pairing of the hairpin arm of the S series (this is most pronounced around 8–18 bp). In contrast, the hairpin stem of S1 has the G_4C_4 sequence at the tip (including the loop) of the hairpin arm. Thus, in S1, the hairpin stem is predominantly AT rich. A plot of T_m of the hairpin stem as a function of increasing number of base pairs within the hairpin stem is shown in Figure 3. It is clear that the relative stability of the hairpin stem in the range of 16–20 bp correlates with the relative values of the deletion frequencies. Within the first 6 bp, the thermal stability is identical for all three inverted repeats. Above 8 bp, however, S3B is the most stable.

The importance of the stability of base pairing in the hairpin stem is also suggested from analysis of the deletion frequencies of the 66-, 55- and 44-bp *lac* inverted repeats (Table 2). These inverted repeats do not exist appreciably as cruciforms *in vivo* (SINDEN, BROYLES and PETTIJOHN 1983; R. R. SINDEN, unpublished results) and are, by far, the least active cruciform-forming sequences studied here. The 66-bp inverted repeat forms cruciforms relatively easily *in vitro* ($\sigma_c < -0.05$) (SINDEN and PETTIJOHN 1984). We have failed to observe cruciform transitions in the 44- and 55-bp sequences under conditions similar to that required for cruciform formation in the 66-bp *lac* operator sequence. (R. R. SINDEN and S. E. TIMMONS, unpublished results). However, the deletion frequencies in wild-type cells (JTT1, HB101 and W3110) correlate roughly with the predicted thermal stabilities of hairpin arms as shown in Figure 4 and Table 3. The 66 bp inverted repeat in pOCE15 is genetically the most stable and has the lowest thermal stability. The 44-bp inverted repeat in pOCE44 has the highest deletion frequency and it is more thermally stable than the 66-bp sequence. The relative thermal stabilities best correlate with the deletion frequencies within the range of 15–18 bp. Since these sequences are different lengths these data are not entirely conclusive. However, since the length of the inverted repeat appears to be important, as discussed below the inverse relationship between length and deletion frequency for the *lac* operator inverted repeats may support the argument that thermal stability of the stem is also an important factor in determining deletion frequency. Additional experiments are in progress to specifically test this hypothesis.

The frequency of deletion is dependent on the rate of cruciform formation: The reversion frequencies for the S and F series inverted repeats are shown in Table 1 and listed in Table 4 with the k_c and σ_c values. There was a large difference in the reversion frequencies of the S and F series inverted repeats and, as discussed below, this may reflect a difference in the length of the direct repeat. Within the S or F series inverted repeats there was a correlation between the

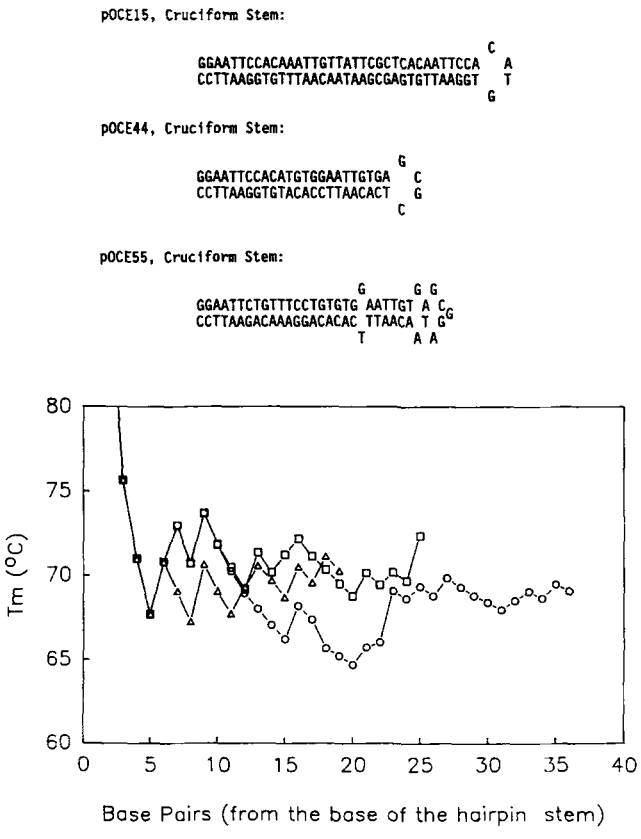

FIGURE 4.—Thermal stability of the *lac* operator hairpin arms. Top, The cruciform stems of the 66-, 44- and 55-bp *lac* operator inverted repeats are shown (pOCE15, pOCE44 and pOCE55, respectively). Each sequence includes a palindromic G/C pair flanking the *Eco*RI site. Bottom, The T_m values were calculated as described previously (ZHENG and SINDEN 1988) and are plotted as a function of length up the hairpin stem. The T_m calculations do not include the four bases that will be unpaired at the loop of the hairpin stem in pOCE15 and pOCE44. In pOCE55 the T_m calculation stops at the 19th base pair just before G/T mismatch since the stability of the end of the hairpin past this point is uncertain. ○, pOCE15; □, pOCE44; △, pOCE55.

reversion frequency and the rate of cruciform formation as shown in Table 4. Inverted repeat F2 formed cruciforms 35 times more rapidly than F1 at the same superhelical density. The deletion frequency of F2 is about 2.3 times higher than that for F1. Because of the length of the hairpin stem there may be little effect of the total thermal stability of the stem on the frequency of deletion. The thermal stability is identical for the first 43 bp of the hairpin stem. Including the AT or GC rich center, the T_m of F2 is lower than that for F1, respectively (Table 3). Based on thermal stability alone, F1 might be expected to be genetically less stable.

Although the rates of cruciform formation have not been precisely measured for the npF series inverted repeats, the rates for this series correspond to the deletion frequencies in a fashion consistent with the results for the S and F series. The T_m values of the

TABLE 4

Comparison of deletion frequencies for the F and S series inverted repeats in relation to the superhelical density dependence and kinetics of cruciform formation

Inverted repeat	Deletion frequency	K_c (rate of cruciform formation)	σ_c (superhelical density required for cruciform formation)[c]
S1	1.15×10^{-10}	$6.5 \pm 3.0 \times 10^{-4a}$	−0.046
S2	1.77×10^{-9}	$1.7 \pm 0.3 \times 10^{-3a}$	−0.043
S3B	1.63×10^{-8}	$2.6 \pm 0.3 \times 10^{-3a}$	−0.039
F1	6.2×10^{-7}	$6.5 \pm 0.4 \times 10^{-4b}$	−0.046
F2	1.45×10^{-6}	$2.3 \pm 0.4 \times 10^{-2b}$	−0.043

[a] Cruciform transition rates for the S series inverted repeats were for DNA with $\sigma = -0.035$ and at 25° (from ZHENG and SINDEN 1988).

[b] Cruciform transition rates for the F series inverted repeats were for DNA with $\sigma = -0.065$ and at 41° (from ZHENG and SINDEN 1988).

[c] σ_c values are from ZHENG and SINDEN (1988).

hairpin stems are very similar among the sequences in this series (Table 3).

Within the S series inverted repeats, which were very active in cruciform formation *in vitro*, there was a factor of four difference in the rate of cruciform formation comparing S1 with S3B (Table 4). The ratio of the rates of cruciform formation were 1:1.5:4 for inverted repeats S1, S2 and S3B, respectively (at the same superhelical density). The reversion frequencies showed a difference of greater than a factor of 100 between S3B and S1. The ratio of the reversion frequencies were 1:15:142 for S1, S2 and S3B, respectively. Because of the great difference in the thermal stability of the hairpin stems for the S series inverted repeats (which also correlates with the deletion frequencies) it may not be possible to attribute the difference in deletion frequencies exclusively to the rate of cruciform formation or the stability of the hairpin stem. In fact the big difference between the frequencies of S1 and S3B is likely a combination of the two (or other factors as mentioned below).

The frequency of deletion of inverted repeats is dependent on the superhelical density dependence of cruciform formation and the existence of cruciforms *in vivo*: DNA is maintained at a precise level of supercoiling *in vivo* (GOLDSTEIN and DRLICA 1984; ESPOSITO and SINDEN 1987). This level appears to be about half that observed for DNA purified from cells (SINDEN, BROYLES and PETTIJOHN 1983; SINDEN and KOCHEL 1987; ZHENG *et al.* 1991). At a defined level of superhelical energy *in vivo*, the level of cruciforms will reflect an equilibrium between the formation and loss of cruciforms (discussed in more detail under SUMMARY). The superhelical density dependent formation of cruciforms will be dependent on σ_c, the superhelical density at which cruciforms form, and k_c, the rate of cruciform formation.

There appears to be a relationship between the

TABLE 5

Existence of cruciforms *in vivo*, σ_c, and Cam^r reversion frequencies for the npF series inverted repeats under various growth conditions

		P_c[b]		Overnight in LB			Late log phase in LB		
				Strain			Strain		
Insert	σ_c[a]	JTT1	RS2	JTT1	RS2 *topA10*	Ratio: RS2/JTT1	JTT1	RS2 *topA10*	Ratio: RS2/JTT1
F14C	−0.038	0.18	0.55	1.44×10^{-7}	1.07×10^{-4}	743	2.13×10^{-7}	1.43×10^{-4}	671
F6S	≈−0.043[c]			5.0×10^{-8}	4.86×10^{-6}	97.2	ND	1.98×10^{-5}	
F8S	≈−0.047[c]			4.13×10^{-8}	1.39×10^{-7}	3.4	5.0×10^{-8}	6.93×10^{-7}	13.8
F10S	−0.049	0.033	0.096						
F14S	−0.065	0.037	0.04	3.1×10^{-8}	4.62×10^{-7}	14.9	3.1×10^{-8}	ND	

[a] σ_c is the critical superhelical density required for cruciform formation. Values from ZHENG *et al.* (1991).
[b] P_c is the fraction of the inverted repeat existing as cruciforms. Values from ZHENG *et al.* (1991).
[c] σ_c values have not been precisely determined on two dimensional agarose gels. The values for F6S and F8S represent the best estimate from preliminary experiments. The values are between those for F14C and F10S.

level of cruciforms *in vivo* (P_c) and the frequency of deletion, suggesting that cruciform formation *in vivo* does contribute to the probability of deletion. Within a set of inverted repeats, sequences that form cruciforms most readily are the most prone to deletion. The npF series represents a family "torsionally tuned" sequences in which a central palindromic sequence is replaced with increasing lengths of nonpalindromic sequences. F14C is a perfect palindrome and is the most active cruciform forming sequence we have made to date. Sequences F6S to F14S are a series of inverted repeats in which the central 6, 8, 10, 12 and 14 bp are nonpalindromic (but have mirror repeat symmetry). As the length of the non-palindromic center increases, greater superhelical energy is required to unwind a larger region of DNA at the center to allow nucleation of intrastrand base pairing which is required for cruciform formation. The σ_c values for some of the npF series inverted repeats are shown in Table 5. When compared at the same superhelical densities, the rates of cruciform formation are fastest for the inverted repeats with the highest (least negative) σ_c values (*i.e.*, F14C with $\sigma_c = -0.038$ forms cruciforms most rapidly). We have measured the levels of cruciforms for F14C, F10S and F14S *in vivo* in plasmid pUC8 and the average P_c values are shown in Table 5 (ZHENG *et al.* 1991). Table 5 also shows reversion frequencies for F14C, F6S, F8S and F14S grown under the conditions used for the cruciform assay. There is a good correlation between the level of cruciforms measured *in vivo* and the deletion frequency. F14C existed at the highest levels *in vivo* (a level of about 5×10^{-3} in strain RS2) and had the highest deletion frequency (about 10^{-4} per cell). Inverted repeat F14S does not form cruciforms readily *in vivo* and had a deletion frequency of 10^{-7} to 10^{-8}, about 10^{-3} lower than that for F14C. The correlation was also observed with JTT1 (Table 5) but to a lower extent than with RS2. This is presumably due to the lower fraction of the inverted repeats that existed as cruciforms *in vivo*.

The fraction of the inverted repeat existing as cruciforms was measured in the *Eco*RI site of plasmid pUC8. We have shown that the levels of inverted repeats within transcription units, in plasmids pUC8 and pBR325, were low, *e.g.*, ≤0.6%. Outside transcription units, up to 50% of the F14C inverted repeat can exist as cruciforms. We have not yet precisely quantitated the levels of cruciforms in the *Eco*RI site of pBR325 except to confirm that the level is not appreciably higher than observed in pUC8, *e.g.*, $P_c <$ 1% (ZHENG *et al.* 1991). It is possible that the fraction of inverted repeats existing as cruciforms at the *Eco*RI site of pBR325 may be slightly different than that in pUC8 since the location of an inverted repeat may affect the ease of cruciform formation (SULLIVAN and LILLEY 1986). DASGUPTA, WESTON-HAFER and BERG (1987) have reported that the deletion frequency of inverted repeats can vary significantly when the inverted repeats were placed in different locations (sequence contexts) of plasmid pBR322. However, it is unlikely that the level of the npF series cruciforms is significantly different in pBR325 than in pUC8 since these sequences were designed to minimize any effects of flanking DNA sequence. This was accomplished by insulating the center AT rich region (which dominantly affects the cruciform transition) with GC rich blocks that will prevent transmission of flanking sequence effects (SULLIVAN, MURCHIE and LILLEY 1988). A more detailed analysis of deletion frequencies and levels of cruciforms existing *in vivo* is in progress and will be published separately.

In vivo assays have not detected cruciforms in the S and F series inverted repeats using the assay of SINDEN, BROYLES and PETTIJOHN (1983) which has a sensitivity of 1–2% (G. ZHENG and R. R. SINDEN, unpublished results). Cruciforms were not detected in the 66-bp *lac* operator inverted repeat (SINDEN,

BROYLES and PETTIJOHN 1983). It is entirely possible that cruciforms can exist in these inverted repeats but that the level is below the level of resolution of the assay used. As discussed below for the kinetic model, these inverted repeats would have a low rate of cruciform formation. Consequently, for inverted repeats that do not form cruciforms readily, DNA secondary structure substrates for deletion may accumulate primarily during replication.

Analysis of the results of deletion frequencies in wild-type, *topA10*, and *gyrB225 E. coli* suggests that the formation of cruciforms promotes deletion mutation. The rationale behind this experiment is that the topoisomerase deficient strain (*topA10*) has a higher level of DNA supercoiling than in wild type cells (PRUSS, MANES and DRLICA 1982) which will result in an increase in the level of cruciforms *in vivo*. Data supporting this are shown in Table 5. There may also be a corresponding decrease in the level of cruciforms in a strain with a mutation in DNA gyrase (*gyrB225*) resulting in a lower level of DNA supercoiling *in vivo*. For inverted repeat F14C, of the npF series, there was an increase in deletion frequencies by a factor of 700–1,400 in the *topA10* strain compared to wild type (Table 1 and Table 5). The deletion frequency for inverted repeat F6S was increased by a factor of 67–97. In other npF inverted repeats deletion frequencies in the *topA10* strain were increased by factors of five to ten. It is not clear why the SF series inverted repeats did not show this effect (Table 1). It is possible that a higher rate of cruciform formation for S1F2 is offset by the higher thermal stability of S1F1 (Table 4).

The *lac* operator inverted repeats form cruciforms with slower kinetics and at higher superhelical densities than the npF series. Consequently, the *lac* sequences may exist as cruciforms at very low levels *in vivo*. Deletion frequencies for the *lac* operator inverted repeats (pOCE15, pOCE44 and pOCE55) were increased about a factor of 10–25 in the topoisomerase mutant and decreased a factor of two to ten in the gyrase mutant (Table 2). Inserts that are nonpalindromic would not be expected to show the increase in deletion frequency observed in the *topA10* mutant nor the decrease observed in *gryB225* mutant. As shown in Table 2, this expected result was observed for three nonpalindromic fragments: a nonpalindromic *lac* operator in pOCE165 and two plasmids containing a 64-bp nonpalindromic fragment (pRS-1 and pRS-4). Even in pRS-4 where multiple *Eco*RI linkers significantly increase the overall frequency of deletion between direct repeats there was no effect of changes in superhelical tension on the deletion frequencies.

Effect of different growth conditions of the deletion frequency: Many of the experiments were done over a long period of time as we developed the families of inverted repeats. Consequently, the media used and state of cell growth varied with the particular rationale of different experiments. We noticed that while individual protocols were quite reproducible, the absolute deletion frequencies varied as a function of the media and growth phase of the cells. To demonstrate this in a controlled fashion we measured the reversion frequency of the npF series in K media in late log phase cells (Table 1) and cells grown to late log phase or to stationary phase in Luria broth (Table 5). Although the results are similar, some differences are observed. The deletion frequency appears highest in stationary phase cells. This is especially noticeable in F14C in JTT1. This could result from the accumulation of cruciforms during the incubation in media. Our previous results suggest that transcription and replication remove cruciforms, reconverting them into the linear form of the inverted repeat (ZHENG *et al.* 1991). Thus, in log phase cells the equilibrium level of cruciforms may be lower than that in stationary phase cells. For example, in log phase cells in K media, little difference was observed in the deletion frequencies of the npF inverted repeats in JTT1 (Table 1). However, in the late log phase cells or stationary phase cells the frequency of deletion for F14C in JTT1 was higher than for the other inverted repeats of that series. Experiments designed to test this hypothesis further are in progress and will be published separately.

The frequency of deletion of inverted repeats is dependent on the length of the inverted repeat: In general, the deletion frequency increased with increasing length of the inverted repeat. This result, which is consistent with the results of DASGUPTA, WESTON-HAFER and BERG (1987), WILLIAMS and MÜLLER (1987) and WESTON-HAFER and BERG (1989), was suggested from the higher frequency of deletion of the 100-bp F series inverted repeats (10^{-6} –10^{-7}) compared to the rate for the 42-bp S series (10^{-8} – 10^{-10}) analyzed under identical conditions (Table 1). In addition, comparison of the 60-bp SF series and the 106-bp npF series under identical conditions showed that the longer npF series inverted repeats were deleted at a higher frequency (Table 1), in this case by a factor of greater than five to ten. The S series inverted repeats were much more active in cruciform formation *in vitro* than the F series, as shown in Table 3. However, the F1 and F2 inverted repeats were genetically much more unstable. This argues that length, at least in this case, can be more important in determining deletion frequency than rate of cruciform formation *in vitro*. This conclusion must be tempered by the fact that the S and SF series inverted repeats are flanked by 6 bp direct repeats and the F and npF series flanked by 8-bp direct repeats, as discussed below.

Results for pOCE15 and pOCE17 may provide a dramatic example of the effect of length of the inverted repeat on the deletion frequency. In HB101, a deletion frequency of about 5.5×10^{-7} was observed for plasmid pOCE15 containing a 66-bp perfect palindromic *lac* operator. Plasmid pOCE17 contains two copies of the 66-bp inverted repeat creating a 132-bp palindromic insert. The deletion frequency of pOCE17 was about 5×10^{-3}, or 10,000 times greater than that for pOCE15. Although the length of the direct repeat in pOCE17 is twice that in pOCE15, the symmetry of the 132-bp insert in pOCE17 is complicated. Two copies of the 66-bp sequence form one large inverted repeat. However, the two copies of the 66-bp sequence also represent two direct repeats, each of which itself is palindromic. This highly symmetrical and direct repeated sequence has complex possibilities for DNA secondary structure formation or misalignment through slipped mispairing. Therefore, pOCE17 may not be directly comparable to inserts with simple inverted repeated symmetry. The length of the direct repeats flanking both the 66- and 132-bp inverted repeats is 8 bp.

The results of Tables 1 and 2 are consistent with the general observation that inverted repeats above 130–150 bp are difficult, if not impossible, to maintain in plasmid DNA in *E. coli* (LILLEY 1981; BETZ and SADLER 1981a; WELLER *et al.* 1985; LOCKSHON and GALLOWAY 1986). Above a certain length, sequences may be deleted through intramolecular recombination (WARREN and GREEN 1985) even in a RecA$^-$ background. The *in vitro* persistence length of DNA of 140–150 bp (HAGERMAN 1988) would suggest that sequences less than 140–150 bp could not undergo intramolecular recombination. This is about the size of inverted repeats that can be stably maintained in *E. coli*. This may be a seemingly fortuitous correlation since the interaction of DNA with the bacterial histone-like HU protein can reduce the persistence length considerably, probably to less than 80 bp (HODGES-Garcia, Hagerman and PETTIJOHN 1989). A reduction in persistence length could facilitate intramolecular recombination in shorter regions of DNA. Alternatively, DNA in an HU:DNA complex may not be available for intramolecular genetic recombination. Were this the case, the instability of inverted repeats longer than the persistence length of DNA may, in fact, result from intramolecular recombination. It is important to point out that the sequences studied here are shorter than the persistence length of DNA, and therefore, may not undergo deletion by a mechanism involving intramolecular recombination.

The reversion frequencies for the several palindromic and nonpalindromic sequences were measured in both *recA*$^-$ and *recA*$^+$ backgrounds. Data presented in Table 2 show little effect of *RecA*$^+$ on the deletion frequencies for the *lac* operator inverted repeats or the nonpalindromic inserts. The independence of deletion on *RecA*$^+$ is consistent with the analyses of COLLINS, VOLCKAERT, and NEVERS (1982), DASGUPTA, WESTON-HAFER and BERG (1987), WILLIAMS and MÜLLER (1987), and BALBINDER, MACVEAN and WILLIAMS (1989).

The length of the direct repeats flanking the inverted repeats varies in some of these constructs. It is known from the work of ALBERTINI *et al.* (1982), DASGUPTA, WESTON-HAFER and BERG (1987), WILLIAMS and MÜLLER (1987), SINGER and WESTLYE (1988), and BALBINDER, MACVEAN and WILLIAMS (1989) that the length, spacing and base composition of directs repeats can influence the frequency of deletion between direct repeats. In this study we have not specifically addressed these questions. As shown in Figure 1 and 2, and Table 1 the S and SF series inverted repeats and pOCE55 are flanked by the 6-bp direct repeat sequence 5′GAATTC3′. The F and npF series sequences, pOCE15, pOCE17, pOCE44, pOE165 and pRS-1 are flanked by the 8-bp direct repeat 5′GGAATTCC3′. The pRS-4 insert is flanked by a 11-bp direct repeat (5′TCCGGAATTCC3′). The higher deletion frequency for pRS-4 compared to pRS-1 may illustrate the importance of the length of the direct repeats (although the symmetry within the direct repeats is complicated). A detailed discussion of alternate potential misalignment events occurring in pRS-4 has been presented (BALBINDER, MACVEAN and WILLIAMS 1989). While an increase in the length of the direct repeat from 6 to 8 bp might result in an increase in the deletion frequency, it is not the critical, sole factor responsible for the effects discussed above. For example, there is little correlation between the deletion frequencies of pRS-1, pOCE15, pOCE17 and pOCE44 which vary considerably and all have 8-bp direct repeats. Similarly, the frequencies of pOCE55 and the similarly sized SF series inverted repeats are different and they all have 6-bp direct repeats.

In addition to the perfect direct repeats at the *Eco*RI sites, the sequence flanking the *Eco*RI site in the CAT gene provides extended sequence homology for some of these sequences. For example there is an extended region of homology of 10/14 bp in pOCE15 compared to a 14/18-bp extended homology for pOCE44. Similarly S1 and S2 have an extended homology of 9/11 bp, whereas S3B contains an extended homology of 12/19 bp. There are several alternate mispairing schemes that can be drawn for some of these sequences. It is difficult to predict the effect of extended homologies with a significant number of mismatched bases. Because of numerous mispairs these alternate intermediates may be much less stable than the one shown in Figure 5. However, it is also possible that

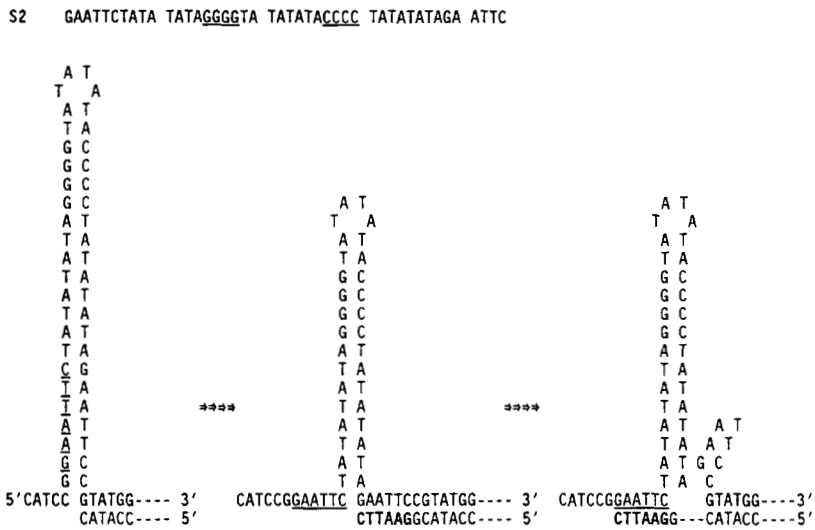

FIGURE 5.—Model for the slipped misalignment of direct repeats stabilized by a hairpin arm at the replication fork. Possible events are shown for the S2 inverted repeat. First, replication could occur up to the base of the S2 hairpin arm and pause (WEAVER and DEPAMPHILIS 1984). Second, replication into the structure would synthesize the first copy of the direct repeat (3′ CTTAAG 5′, shown in bold type). This brings the second copy of the direct repeat (underlined) within close proximity to the newly synthesized copy of the first direct repeat. This close proximity might increase the probability of misalignment. Third, misalignment occurs resulting in a hairpin structure with a small bubble at the base of the hairpin stem. Alternatively, an unpaired loop of DNA could form in the hairpin stem if the small bubble disrupted base pairing within the stem (structure not shown).

the frequency of deletion could be potentially increased by a greater number of alternate slipped mispaired intermediates. We would expect that the stability or suitability of these mismatched intermediates as extendable substrates for polymerase (in terms of a rate k_6 as described below) might be low compared to a structure with no mismatched base pairs. Examples of the types of mismatched intermediates that can be drawn for pRS-4 have been described by BALBINDER, MACVEAN and WILLIAMS (1989).

SUMMARY

Deletion mutations may be promoted by the formation of a stable hairpin arm within single stranded DNA or formation of a cruciform in supercoiled DNA. During replication, the hairpin arm leads to slipped mispairing between direct repeats resulting in a deletion. As discussed above, the deletion mutations analyzed here are not due predominantly to *RecA*-dependent recombination. The molecular mechanism involved in the deletion of the inverted repeats likely involves deletion between direct repeats stabilized by a hairpin arm as originally suggested by GLICKMAN and RIPLEY (1984). The length and base composition of the inverted repeat appear to influence the mutation frequency. The overall thermal stability of the DNA helix within the hairpin arm is not the sole determining factor. Thermal stability within a certain length of the stem may influence deletion. The rate and superhelical density dependence of cruciform formation are important in determining deletion frequency. A significant fraction of the mutations probably arise from resolution of supercoil-induced cruciforms (albeit presumably still through misreplication events).

Kinetic Model for the deletion of inverted repeats: The molecular events of misalignment during replication that might lead to the deletion of palindromic DNA between direct repeats are summarized in Figure 5. This model is consistent with the conclusions of WILLIAMS and MÜLLER (1987), SINGER and WESTLYE (1988), BALBINDER, MACVEAN and WILLIAMS (1989), and WESTON-HAFER and BERG (1989). Figure 6 shows an overall kinetic model for the deletion of inverted repeats. In Figure 6, the misalignment event (of Figure 5) is indicated as the step in which slipped misalignment occurs (at a rate k_6) producing structure (*f*). This analysis does not consider repair of this structure which might occur by some mechanism, possibly via a mismatch-like a methyl-directed strand specific repair pathway. The rate of slipped misalignment and continued replication, k_6, will occur at a finite probability when the replication machinery approaches a hairpin arm. As summarized in Figure 6, the formation of the DNA secondary structure substrate for replication-dependent mutation (*i.e.*, a hairpin arm in replicating DNA) could occur by two pathways with rates of formation k_1 and k_2. k_1 represents the rate of formation of cruciforms in supercoiled DNA (structure *b*). k_1 is superhelical density dependent and will be dependent on the base composition at the center of symmetry (ZHENG and SINDEN 1988; ZHENG *et al.* 1991). k_2 is the rate of hairpin formation in single stranded DNA that occurs as a consequence of DNA replication (structure *d*). Presumably, hairpins are most likely to form in the lagging strand which always contains an Okazaki size region of single stranded DNA (MCHENRY 1988). These hairpins may form rarely before the *E. coli* single strand binding protein (ssb) can prevent intrastrand base pairing. Although k_2 is superhelical density independent, it will be dependent on replication. The probability of hairpin formation, via the k_2 pathway, may be dependent on the thermal stability of the inverted repeat. Therefore, k_2 may be a function of length and base composition. Hairpin arms, once

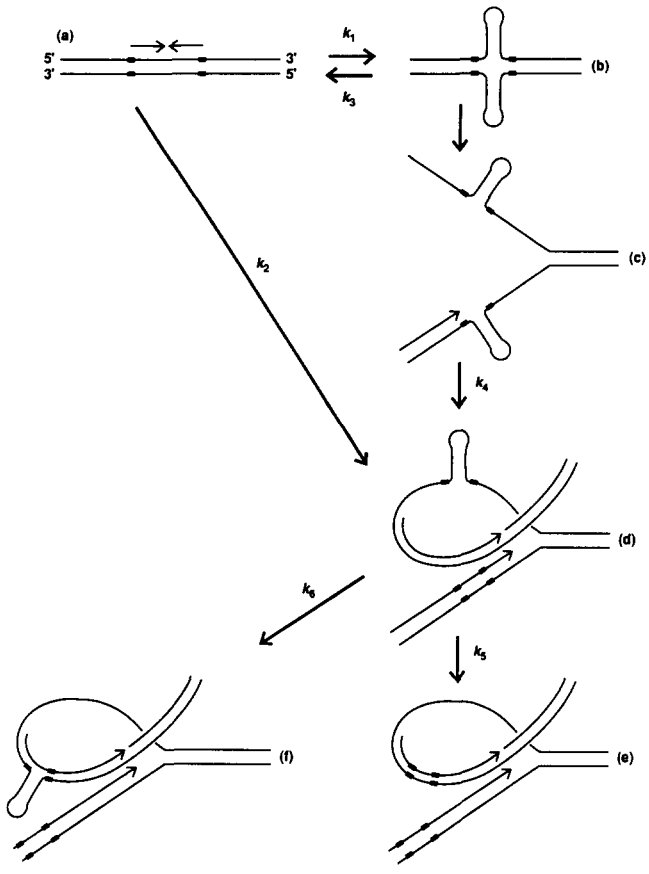

FIGURE 6.—A kinetic model to explain the relationship between the physical-chemical properties of inverted repeats and deletion in living cells. The model is explained in detail in the text. Briefly, (*a*) shows the linear form of the inverted repeat within a region of DNA. The inverted repeat is indicated by arrows above the sequence and the direct repeats by the thick lines. The formation of a cruciform, with rate k_1, is shown in (*b*). Cruciforms, although extremely stable in supercoiled DNA, can be reconverted into the linear form (*a*) by transcription through the inverted repeat. The rate for transcriptional loss of cruciforms in shown as rate k_3. Cruciforms may also be reconverted into the linear form by replication as shown in structures (*c*), (*d*) and (*e*). The rate for replication through a hairpin arm in the leading strand is defined as k_4. The rate for replication through a hairpin arm in the lagging strand is defined as k_5. A hairpin arm may also form during the process of DNA replication as indicated by rate k_2. It is expected that hairpins are most likely to form in the lagging strand where a region of single stranded DNA the size of an Okazaki fragment is maintained during replication according to the model involving concurrent replication of the leading and lagging strands by a asymmetric dimeric polymerase III holoenzyme complex (MCHENRY 1988). At a rate k_6, the slipped misalignment (shown in Figure 5) will occur in the lagging strand producing structure (*f*). At a rate k_7, the slipped misalignment shown in Figure 5 might occur in the leading strand (this structure is not shown). The rate k_7 might be different than k_6 due to the asymmetric nature of the polymerase complex.

formed by processes k_1 or k_2, will undergo slipped mispairing and deletion at the rate k_6, to form structure (*f*). k_6, like k_2, may be dependent on the thermodynamic stability of the hairpin arm or a defined length of the stem. k_6 may also be dependent on the length of the direct repeats and the secondary structures formed. A long length of direct repeats and a high G + C content should increase k_6. Multiple extended regions of homology may increase k_6. Utilization of these as extendable substrates may be dependent on the stability and structure of the mispaired intermediate. If the intermediate is less stable than the structure shown in Figure 5, for example if a region of extended homology contains many mismatched base pairs, the contribution to k_6 might be low.

Once formed, cruciforms can be lost by a variety of mechanisms. k_3 represents transcription dependent conversion of cruciforms back into the linear form of the inverted repeat. This has been demonstrated *in vitro* (MORALES, COBURN and MÜLLER 1990) and likely happens *in vivo* (ZHENG *et al.* 1991). Although the removal of cruciforms by transcription has not been extensively studied, k_3 may be relatively independent of base composition or length of the inverted repeat. Replication of structure (*b*) may lead to a situation in which the replication machinery is presented with a hairpin arm at a high frequency. Replication may stop at a hairpin arm (WEAVER and DEPAMPHILIS 1984) and eventually replicate through the inverted repeat in the leading strand at a rate k_4, and no mutation will occur, as shown in structure (*d*). Replication through the hairpin arm in the lagging strand forming structure (e) is represented by rate k_5. Rate k_5 may be different than k_4 since an asymmetric polIII complex is involved in DNA polymerization (MCHENRY 1988). A step in which misalignment and replication occurs in the leading strand (with rate k_7) is not shown in Figure 6. Because of the asymmetric polymerase complex, k_7 may not equal k_6.

The results for the npF series inverted repeats suggest that k_1 is very important in contributing to the frequency of deletion, since rates k_2-k_6 should be very similar if not identical for this series. k_1 may also be the determining factor in the deletion frequency of the F series inverted repeats because of the length of the hairpin stem. Although the overall T_m values of the stems are slightly different, the first 43 bp are identical (and the same as for the npF series).

The results for the S series inverted repeats are consistent with all implications of the kinetic parameters. The rate k_1 varies significantly, with a greater rate of cruciform formation for the S3B inverted repeat. In addition, differences in overall thermal stability of the hairpin arm, as well as regional T_m values will affect k_2 with the formation of a hairpin in S3B being most favored. These stability differences might also differentially affect rates k_4-k_7. The increased deletion frequency of S3B compared to S1 is consistent with faster rates of k_1, k_2, k_4 and k_6.

An increased deletion frequency for S1F2 compared to S1F1 would be consistent with the faster rate

k_1 for S1F2 (since more cruciforms would form in S1F2). However, rates k_2–k_7 (except k_3) might predict a greater deletion frequency for S1F1 compared to S1F2. This is because of the greater thermal stability of the S1F1 hairpin arm. The similar deletion frequencies observed for S1F1 and S1F2 are likely to be a compromise based on the influence of all expected rates (k_1–k_7) on the deletion frequencies. This may be especially true compared to the great difference observed for S1 and S3B where all kinetic predictions favored deletion in S3B.

This work was supported by grants NP 490, NP 490B, and NP 490C from the American Cancer Society and by Grant ES05508 from the National Institute of Environmental Health Sciences.

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 deletion. Cell **29:** 319–328.

BALBINDER, E., 1988 Use of predesigned plasmids to study deletions, pp. 378–391 in *DNA Replication and Mutagenesis*, edited by R. E. MOSES and W. C. SUMMERS. ASM Publications, Washington D.C.

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.

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.

BETZ J. L., H. M. SASMOR, F. BUCK, M. Y. INSLEY and M. H. CARUTHERS, 1986 Base substitution mutants of the *lac* operator: *in vivo* and *in vitro* affinities for *lac* repressor. Gene **50:** 123–132.

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

COLLINS, J., G. VOLCKAERT and P. NEVERS, 1982 Precise and nearly-precise excision of the symmetrical inverted repeats of Tn5; common features of *recA*-independent deletion events in *Escherichia coli*. Gene **19:** 139–146.

COUREY, A., and J. C. WANG, 1983 Cruciform formation in a negatively supercoiled DNA may be kinetically forbidden under physiological conditions. Cell **33:** 817–829.

COUREY, A., and J. C. WANG, 1988 Influence of DNA sequence and supercoiling on the process of cruciform formation. J. Mol. Biol. **202:** 35–43.

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.

DEMASSEY, B., R. A. WEISBERG and F. W. STUDIER, 1987 Gene 3 endonuclease of bacteriophage T4 resolves conformationally branched structures on double-stranded DNA. J. Mol. Biol. **193:** 359–376.

DINARDO, S., K. A. VOELKEL, R. STERNGLANZ, A. E. REYNOLDS and A. WRIGHT, 1982 *Escherichia coli* DNA topoisomerase I mutants have compensatory mutations in DNA gyrase genes. Cell **31:** 43–51.

DRAKE, J. W., L. S. RIPLEY and B. W. GLICKMAN, 1983 Updating the theory of mutation. Am. Sci. **71:** 621–630.

ESPOSITO, F., and R. R. SINDEN, 1987 Supercoiling in prokaryotic and eukaryotic DNA: changes in response to topological perturbation of plasmids in *E. coli* and SV40 *in vitro*, in nuclei, and in CV-1 cells. Nucleic Acids Res. **15:** 5105–5124.

GELLERT, M., M. H. O'DEA and K. MIZUUCHI, 1983 Slow cruciform transitions in palindromic DNA. Proc. Natl. Acad. Sci. USA **80:** 5545–5549.

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.

GOLDSTEIN, E., and K. DRLICA, 1984 Regulation of bacterial DNA supercoiling: plasmid-linking numbers vary with growth temperature. Proc. Natl. Acad. Sci. USA **81:** 4046–4050.

HAGAN, C. E., and G. J. WARREN, 1983 Viability of palindromic DNA is restored by deletions occurring at low but variable frequency in plasmids of *Escherichia coli*. Gene **24:** 317–326.

HAGERMAN, P. J., 1988 Flexibility of DNA. Annu. Rev. Biophys. Biophys. Chem. **17:** 265–286.

HODGES-GARCIA, Y., P. J. HAGERMAN and D. E. PETTIJOHN,, 1989 DNA ring closure mediated by protein HU. J. Biol. Chem. **264:** 14621–14623.

HOLMES, D. S., and M. QUIGLEY, 1981 A rapid boiling method for the preparation of bacterial plasmids. 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.

KOCHEL, T. J., and R. R. SINDEN, 1988 Analysis of trimethylpsoralen photoreactivity to Z-DNA provides a general *in vivo* assay for Z-DNA: Analysis of the hypersensitivity of $(GT)_n$ junctions. BioTechniques **6:** 532–543.

LEACH, D. R. F., and F. W. STAHL, 1983 Viability of lambda phages carrying a perfect palindrome in the absence of recombination nucleases. Nature **305:** 448–451.

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

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.

MCHENRY, C. S., 1988 DNA polymerase III holoenzyme of *Escherichia coli*. Annu. Rev. Biochem. **57:** 519–550.

MIZUUCHI, K., B. KEMPER, J. HAYS and R. A. WEISBERG, 1982 T4 endonuclease VII cleaves Holliday structures. Cell **29:** 357–365.

MORALES, N. M., S. D. COBOURN and U. E. MÜLLER, 1990 Effect of *in vitro* transcription on cruciform stability. Nucleic Acids Research **18:** 2777–2782.

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.

PRUSS, G. J., S. H. MANES and K. DRLICA, 1982 *Escherichia coli* DNA topoisomerase I mutants: increased supercoiling is corrected by mutations near gyrase genes. Cell **31:** 35–42.

RIPLEY, L. S., 1990 Frameshift mutation: determinants of specificity. Annu. Rev. Genet. **24:** 189–213.

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

SADLER, J. R., H. SASMOR and J. L. BETZ, 1983 A perfectly symmetric *lac* operator binds the *lac* repressor very tightly. Proc. Natl. Acad. Sci. USA **80:** 6785–6789.

SADLER, J. R., and M. TECKLENBERG, 1981 Cloning and characterization of the natural lactose operator. Gene **13:** 13–23.

SADLER, J. R., J. L. BETZ, M. TECKLENBERG, D. V. GOEDDEL, D. G. YAVSURA and M. H. CARUTHERS, 1978 Cloning of chemically synthesized lactose operators. II. *Eco*RI-linkered operators. Gene **3:** 211–232.

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

SINDEN, R. R., and T. J. KOCHEL,, 1987 Reduced 4,5′,8-trimethylpsoralen cross-linking of left handed Z-DNA stabilized by DNA supercoiling. Biochemistry **26:** 1343–1350.

SINDEN, R. R., and D. E. PETTIJOHN, 1984 Cruciform transitions in DNA. J. Biol. Chem. **259:** 6593–6600.

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

SINGLETON, C. K., 1983 Effects of salts, temperature, and stem length on supercoil-induced formation of cruciforms. J. Biol. Chem. **258:** 7661–7668.

STERNGLANZ, R., S. DINARDO, K. A. VOELKEL, Y. NISHIMURA, Y. HIROTA, A. K. BECHERER, L. ZUMSTEIN and J. C. WANG, 1981 Mutations in the gene encoding the *Escherichia coli* DNA topoisomerase I affect transcription and transposition. Proc. Natl. Acad. Sci. USA **78:** 2747–2751.

STREISINGER, G., Y. OKADA, J. EMRICH, J. NEWTON, A. TSUGITA, E. TERZAGHI and I. INOUYE, 1966 Frameshift mutation and the genetic code. Cold Spring Harbor Symp. Quant. Biol. **31:** 77–86.

SULLIVAN, K. M., and D. M. J. LILLEY, 1986 A dominant influence of flanking sequences on a local structural transition in DNA. Cell **47:** 817–827.

SULLIVAN, K. M., and D. M. J. LILLEY, 1987 Influence of cation size and charge on the extrusion of a salt-dependent cruciform. J. Mol. Biol. **193:** 397–404.

SULLIVAN, K. M., A. I. H. MURCHIE and D. M. J. LILLEY, 1988 Long range structural communication between sequences in supercoiled DNA. J. Biol. Chem. **263:** 13074–13082.

TAYLOR, A. F., and G. R. SMITH, 1990 Action of RecBCD enzyme on cruciform DNA. J. Mol. Biol. **211:** 117–134.

WARD, G. K., R. MCKENZIE, M. ZANNIS-HADJOPOULOS and G. B. PRICE, 1990 The dynamic distribution and quantification of DNA cruciforms in eukaryotic nuclei. Exp. Cell Res. **188:** 235–246.

WARREN, G. L., and R. L. GREEN,, 1985 Comparison of the physical and genetic properties of palindromic DNA sequences. J. Bacteriol. **161:** 1103–1111.

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

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.

WEST S. C., and A. KÖRNER, 1985 Cleavage of cruciform DNA structures by an activity from *Saccharomyces cervisae*. Proc. Natl. Acad. Sci. USA **82:** 6445–6449

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

WESTON-HAFER, K., and D. E. BERG, 1991 Limits to the role of palindromy in deletion formation. J. Bacteriol. **173:** 315–318.

WILLIAMS, W. L., and U. E. MÜLLER,, 1987 Effects of palindrome size and sequence on genetic stability in the bacteriophage Φx174 genome. J. Mol. Biol. **196,** 743–755.

ZHENG, G., and R. R. SINDEN,, 1988 Effect of base composition at the center of inverted repeated DNA sequences on cruciform transitions in DNA. J. Biol. Chem. **263:** 5356–5361.

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