Genetic Instability Within Monotonous Runs of CpG Sequences in *Escherichia coli*

Marc Bichara, Sylvie Schumacher and Robert P. P. Fuchs

Cancérogénèse et Mutagénèse Moléculaire et Structurale, CNRS, 67400 Strasbourg-Illkirch, France

Manuscript received May 4, 1994
Accepted for publication March 17, 1995

ABSTRACT

Genetic information can be altered by base substitutions, frameshift mutations, and addition or deletion of nucleotides. Deletions represent an important class of genetic aberration occurring at DNA sequences where it is often possible to predict the existence of intermediates of mutation. Instability within tracts of repetitive sequence have recently been associated with several genetic disorders, including the so-called triplet repeat diseases and certain forms of colorectal cancers. In *Escherichia coli*, (GpC)\textsubscript{n} repetitive sequences have been shown to be deletion prone, but the precise mechanism of this mutagenic pathway is still unknown. We show here that interrupting the monotony of the (GpC)\textsubscript{n} run with an ApT or a GpT dinucleotide decreases the rate of deletions within these sequences. On the other hand, introducing purine-pyrimidine alternating sequences beside the GpC insert results in an increased rate of deletion. Two pathways can be envisioned: (1) (GpC)\textsubscript{n} tracts can be seen as potential Z-forming DNA sequences, and this unusual DNA structure can be processed by an unknown cellular mechanism to give rise to the observed deletions and (2) (GpC)\textsubscript{n} monotonous runs can be considered as a succession of direct or palindromic repeats, allowing formation of DNA structures that are known to participate to frameshift mutagenesis. The results presented in this article are discussed in the light of these two alternative pathways.

SPONTANEOUS frameshift mutations have been widely studied and are thought to be the consequences of perturbed DNA metabolism (for a review, see Kunkel 1990; Ripley 1990). These mutations often occur at sequence sites where it is possible to predict the existence of mutation intermediates. Models for spontaneous frameshift mutation can sometimes be deduced from the knowledge of the DNA sequence context. This deductive approach led to a misalignment model that can either explain the deletion or the addition of one unit of repetition within repeated DNA sequences (Streisinger et al. 1966). Mutations of this type are frequently characteristic of spontaneous mutational hotspots (Streisinger et al. 1966) and of some induced hot spots (Lambert et al. 1992). Duplications or deletions within simple or tandem repeats (i.e., monotonous repeat of one or two base pairs) in both *Escherichia coli* (Streisinger et al. 1966; Levinson and Gutman 1987) and yeast (Henderson and Petes 1992) are most frequently additions or deletions of one or two units of the repeat. The mutagenic intermediates expected (i.e., bulges of one or a few bases) are known to be a substrate for the mismatch repair system (Levinson and Gutman 1987; Parker and Marinus 1991; Strand et al. 1993). The DNA context can influence misalignment mutagenesis. For example, the presence of palindromic DNA structure between two direct repeats may facilitate an interstrand misalignment, thus increasing the rate of deletion events (Glückman and Ripley 1984; Singer and Westley 1988; Weston-Hafer and Berg 1989; Pierce et al. 1991; Sindén et al. 1991; Trinh and Sindén 1991).

Tandem repeats are a succession of direct repeats and thus they allow interstrand misalignments; however, depending on the nucleotide sequence of the repeat, secondary structures may also form. It is known that unusual structures like cruciforms can be recognized by specific enzymatic pathways both in vivo and in vitro. For example, cruciform structures are processed in *E. coli* by the ruvC gene product (Iwasaki et al. 1991), which is part of the ruvABC Holliday junction resolving (for a review, see West and Conolly 1992). In a previous work (Freund et al. 1989), we have shown that plasmid-borne DNA sequences that can undergo B to Z transition are prone to deletions.

To get further insights on the biological mechanisms giving rise to these deletions, we analyzed the mutation patterns of a new set of plasmids designed to test the influence of the monotony of the repeats and of the neighboring sequences on the frameshift mutagenesis.

MATERIAL AND METHODS

**Bacterial strain and medium**: The *E. coli* strain TB1:F-ara Δ(lac proAB) rpsL (Str') (f 80 dΔ(lacZ)M15) was used in all experiments. Transformed bacteria were plated on LB plates...
containing ampicillin (100 µg/ml), X-Gal (25 µg/ml), isopropyl thigalactoside (30 µg/ml), and lactose (0.04%).

**Plasmid construction:** For the pUC (GC) plasmid family, constructions were made starting from plasmid pUC8 or a pUC8 derivative having a deletion of the two consecutive guanines located between the Pst and the HindIII site of the polynucleotide. These plasmids were used to generate the different families of (GpC)ₙ, containing plasmids (pUC(GpC)₁₅, pUC(GpC)₁₀pT (GpC₉); or pUC(GpC)₉pT (GpC₉)) by insertion of oligonucleotides in the Sacl site of the polynucleotide. The plasmids are described as being of the frameshift genetic sign (-1) or (-2) when they are out of frame by deletions of 1 or 2 bp, respectively. Plasmids pUC (GC)₉ (PuPy), pUC (PuPy) (GC)₉ (PuPy) and pUC (PuPy) were constructed by inserting oligos within pUC18 plasmid digested by EcoRI and SalI to give the values described in Table 2. All plasmids constructed for this study have a lacZ gene and produce white colonies on adequate medium. Phenotypic (blue colony) deletion mutants are detected by their ability to express a LacZ phenotype on adequate medium (FREUND et al. 1989). The sequencing of the GpG-rich plasmids was done following Sanger dyeoxy protocol except that the reactions were carried out at 50° in presence of 15% dimethylsulfoxide. They were then run at 50% on 20% sequencing gels for 4 hr.

**Determination of the spontaneous mutation rate:** Spontaneous mutation rates in plasmids were estimated using the method of LEA and COULSON (1949). TB1 competent cells were transformed with plasmid DNA, and ~100–1000 transformants were plated on indicator medium. On the next day, white ampicillin-resistant colonies were pooled and diluted to produce inocula of ~100 cells for each of 20 cultures. The small inoculum assuère that cultures are not initiated with preexisting mutants. Cultures were grown with vigorous shaking for 16 hr at 37°C. Samples from each culture were titrated for total cells. This number was used to calculate the total number of cell generations.

In addition, the ratio of blue-white colonies was determined on selective indicator plates. The number of blue colonies counted for each stock was in the range of several hundred. The blue-white colony ratios were used to determine the median value (the average of the two most central ratios for the range of ratios in the 20 stocks. The value was used with the table of LEA and COULSON to estimate the mean number of mutation events per culture and its variance. The mutation rate Fm was calculated from μ using the number of generations of cells in the cultures. The Fm rate is the "phenotypic mutation rate" because only mutants exhibiting a blue color were detectable in this assay.

**Determination of the absolute mutation rate:** When the mutation rate is high enough (>0.5 × 10⁻⁸胞/cell/generation), the absolute mutation, which includes all length change mutants independent of their phenotype, can be determined biochemically. The plasmid DNA is prepared from the two median cultures, cut with EcoRI, 5' ³²P-labeled, and cut with HindIII. The resulting digest is loaded on a 20% sequencing gel and run together with size markers at 50° for 4 hr. The bands corresponding to the length of the starting plasmid and to the different deletion events are identified and quantified using a Fuji Bioimager (BAS 2000). Intensities of bands corresponding to the phenotypic mutations whose rate is known are compared with the other bands that have length that do not produce phenotypically detectable mutants. The absolute mutation rate of the plasmidic sequence is the sum of the absolute mutation rates of all the observed bands.

**Two-dimensional agarose gel electrophoresis:** The gels were made as described in FREUND et al. (1989).

**RESULTS**

When studying spontaneous mutations that by definition occur during the entire growth of a bacterial culture, it is essential to determine accurately the mutation rate, that is, the frequency of mutation per cell per generation. This problem was experimentally addressed by LURIA and DELBRUCK (1943) and by LEA and COULSON (1949), who developed a statistical method to estimate the actual number of mutation events per generation in a bacterial culture containing a given number of observed mutants. This method (see MATERIALS AND METHODS) has been extensively used in the present work and is efficient and reliable.

**Influence of the interruption of the mononucleotide of a GpC run on the rate of deletion:** The genetic instability of (GpC)ₙ sequences has been previously observed with constructions containing strictly monotonous repetition of the GpC dinucleotide (FREUND et al. 1989). Here the influence of an interruption of the strict mononucleotide of the (GpC)ₙ sequence was investigated by comparing plasmids with no interruption to plasmids with a central interruption of GpT or ApT dinucleotide. GpC sequences (with or without interruption) were inserted into the lacZ gene in two constructs. The insertion of each repeat into the first construct permits the detection of shifts to the (−1) reading frame. The second construct for each repeat bears a 2-bp deletion downstream from the mutation target region (see MATERIALS AND METHODS). This additional shift permits the detection of shifts to the opposite (−2) reading frame. The plasmids with interrupted repeats are referred to as pUC(GpC)₉ GpT (GpC)₀ (−1) or pUC(GpC)₁₀pT (GpC)₀ (−2) when they contain the GpT interruption and pUC(GpC)₀pT (GpC)₀ (−1) or pUC(GpC)₁₀pT (GpC)₀ (−2) when they contain the ApT interruption. The sum of the results from both constructs provides an estimate of the phenotypic Fm for each plasmid. Other length mutants (producing a shift of sign 0) having no phenotypic consequence were measured biochemically as described in MATERIALS AND METHODS. Mutation rates in interrupted plasmids were compared to those in plasmids without interruptions: pUC(GC)₁₅ (−1) and pUC(GC)₁₅ (−2).

Table 1 and Figure 2 show the phenotypic mutations for the six plasmid sequences. In each case, the biochemical measurement of rates of deletion has also been made with the DNA of the two median cultures, thus permitting the calculation of an absolute mutation rate, as described in MATERIALS AND METHODS. For a given sequence, slight differences in the analysis are often observed between the DNAs extracted from the two median cultures, reflecting the fact that these cultures were selected only as median values for the phenotypic mutation rate. To minimize this variation, the average of four analysis (two median values in two different reading frame contexts for each plasmid) was used to generate the mutation spectra presented in Figure 1.

None of the plasmids produced detectable levels of addition/duplication mutations. This result is provocative in terms of the slippage mutagenesis model for
Genetic Instability of CpG Runs

TABLE 1
Characterization of the deletions within pUC(GC)₉ derivatives

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Frame context</th>
<th>r (X10⁷/ml)</th>
<th>m (X10⁶/ml)</th>
<th>Fm (X10⁻³)</th>
<th>abs Fm (X10⁻³)</th>
<th>-σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>pUC(GC)₁₅</td>
<td>-2</td>
<td>26.5</td>
<td>14.9</td>
<td>13 ± 1*</td>
<td>31.6</td>
<td>0.033</td>
</tr>
<tr>
<td>pUC(GC)₁₅</td>
<td>-1</td>
<td>12.47</td>
<td>7.35</td>
<td>9 ± 0.8</td>
<td>10.2</td>
<td>0.036</td>
</tr>
<tr>
<td>pUC(GC) GT(CG)₇</td>
<td>-2</td>
<td>5.51</td>
<td>3.85</td>
<td>5.2 ± 0.5</td>
<td>10.2</td>
<td>0.036</td>
</tr>
<tr>
<td>pUC(GC) GT(CG)₇</td>
<td>-1</td>
<td>3.67</td>
<td>2.50</td>
<td>2 ± 0.2</td>
<td>3.9</td>
<td>0.042</td>
</tr>
<tr>
<td>pUC(GC) AT(CG)₇</td>
<td>-2</td>
<td>3.03</td>
<td>1.93</td>
<td>2.7 ± 0.2</td>
<td>3.9</td>
<td>0.042</td>
</tr>
<tr>
<td>pUC(GC)₃ AT(CG)₇</td>
<td>-1</td>
<td>0.45</td>
<td>3.24</td>
<td>0.27 ± 0.02</td>
<td>3.9</td>
<td>0.042</td>
</tr>
</tbody>
</table>

r is the mean of the two median values of the number of mutants per milliliter observed within 20 cultures; m is the deduced number of mutations per milliliter that has occurred in a culture having r mutants, following LEA and COULSON (1949); Fm is the phenotypic mutation rate expressed in mutations per generation (see text); abs Fm is the absolute mutation rate as derived from the two phenotypic one and the biochemical analysis of the mutations (see MATERIALS AND METHODS); and -σ is the superhelical density at which the B to Z transition occurs.

The standard deviation on the phenotypic mutation rate is calculated as described in LEA and COULSON (1949) (see MATERIALS AND METHODS).

mutagenesis in base runs, which would suggest that both deletions and duplications are expected to occur by related misalignment mechanisms. The absence of the addition mutation is nevertheless reminiscent of in vitro results, showing that slippage during in vitro polymerization reactions produced primarily deletions (PAPANICOLAOU and RIPLEY 1991).

There are striking differences in the spectra of mutants produced by the different plasmids (Figure 1). For example, mutants of pUC(GC)₁₅ (Figure 1A) include many small deletions of <10 bp. These small deletions are infrequent in the interrupted plasmids. Moreover, the interrupted plasmid spectra are dominated by large deletions of 16 bp (Figure 1, B and C) that occur at rates higher than mutations of the same size in pUC (GC)₁₅. Clearly, the interruptions strongly influence mutagenesis and appear to influence different mutational outcomes in different ways.

A small but significant fraction of small deletion mutants arise in the plasmid with the GT interruption. These were primarily detected in the biochemical assay, suggesting that many of these may not produce a phenotypically detectable phenotype. Unfortunately, their unusual migration patterns of the DNAs in the gel precluded a precise determination of their size. The single example in this class recovered as a phenotypic mutant; the (GC)₇ GT (GC)₉ sequence (Figure 1B) shows an unusual migration pattern as well. Thus, we suspect that the unusual migration is due to the presence of the GT interruption, and we predict that the remaining bands are also from mutants that retain the GT interruption. In all cases the deletions observed were multiple of 2 bp within the insert. Some of these mutants were then used as migration markers during the biochemical analysis of the median cultures on sequencing gels.

As can be seen in Table 1, there is good agreement between the absolute mutation rate of the different sequences and their ease to undergo a B to Z structural change as monitored by the supercoiling density necessary for this change. The supercoiling density of transition (-σ) determined from two-dimensional agarose gel electrophoresis (data not shown) is directly related to the energy necessary to induce the B to Z structural transition. Indeed, the 3- and 10-fold decrease between the deletion rates within (GpC)₁₅ and (GpC)₇ GT (GpC)₇ or (GpC)₇ ApT (GpC)₇ sequences correlates with the values of -σ (0.033, 0.036 and 0.042) found for these plasmids, respectively.

Influence of the sequence context around the (GpC)₉ monotonous run on the deletion rate: To investigate the influence of the neighboring bases on the deletion induced within a (GpC)₉ monotonous run, we designed plasmids containing a simple (GpC)₉ monotonous run embedded within different sequence contexts. These contexts were chosen so that the B to Z transition of the (GpC)₉ insert was energetically favored by the presence of run(s) of alternating purine-pyrimidine tracts on one or on both sides of the sequence. As a control, we designed a plasmid containing a purine-pyrimidine alternating sequence of the same length (i.e., 44 bp). These sequences are described in Table 2.

The phenotypic mutation rate was determined for these sequences as described in MATERIALS AND METHODS. We observed a significant threefold increase of the mutation rates for the plasmids containing purine-pyrimidine tracts neighboring the (GpC)₉ sequence as compared with the one with a simple (GpC)₉ run (Table 3). The ability of these sequences to undergo a B to Z transition was monitored by two-dimensional agarose gel electrophoresis (Figure 2). These gels unambiguously show that the presence of the surrounding sequences reduces the supercoiling density needed to promote the transition from the regular B conformation to the unusual Z structure. Indeed, the superhelical density (-σ) needed for the B to Z transition of the (GpC)₉ insert alone is 0.043, whereas it is 0.037 for pUC
In this work, we wanted to study the mechanistic aspects of the genetic instability leading to deletions within (GpC)$_n$ monotonous run. Two main pathways can be evoked to account for the experimental results.

The first one implies the involvement of left-handed DNA structure (for a review, see Rich et al. 1984; Palacek 1991; Yagil 1991) in frameshift mutagenesis. Z-DNA structure has been shown to exist in crystals (Wang et al. 1979), and the B to Z transition has been demonstrated on natural plasmidic molecules both in vitro (Wang et al. 1983) and in vivo (Jaworski et al. 1987; Palacek et al. 1987) in E. coli. Putative Z-DNA structures have been found in higher organisms (Naylor and Clark 1990; Sarkar et al. 1991), but no clear role has been attributed to this unusual DNA structure even though a hypothetical role in the regulation of transcription has recently been suggested (Wittig et al. 1991, 1992). Alternatively, GpC runs permit several other types of DNA structures that create misalignments between the template and newly synthesized DNA strands that have been widely hypothesized to participate in deletion and frameshift mutagenesis (for a review, see Ripley 1990). It is possible that deletions in all types of repeats result from misalignment mechanisms. For self-complementary repeats, two general classes of misalignment mechanisms can be envisioned because the tandem repeats can be seen either as direct
Genetic Instability of CpG Runs

TABLE 2
Detailed sequence of the insert of plasmid pUC18, pUC (GC), pUC (PuPy), pUC (PuPy), and pUC (PuPy).

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>r (x10^3/ml)</th>
<th>m (x10^3/ml)</th>
<th>Fm (x10^-6)</th>
<th>-σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>pUC (GC)</td>
<td>8.65</td>
<td>7</td>
<td>0.76 ± 0.09</td>
<td>0.043</td>
</tr>
<tr>
<td>pUC (GC) (PuPy)</td>
<td>28</td>
<td>21.1</td>
<td>2.04 ± 0.22</td>
<td>0.037</td>
</tr>
<tr>
<td>pUC (PuPy) (GC) (PuPy)</td>
<td>38.75</td>
<td>28.2</td>
<td>2.31 ± 0.24</td>
<td>0.037</td>
</tr>
<tr>
<td>pUC (PuPy)</td>
<td>3</td>
<td>3.68</td>
<td>0.26 ± 0.03</td>
<td>0.038</td>
</tr>
</tbody>
</table>

r is the mean of the two median values of the number of mutants per milliliter observed within 20 cultures; m is the deduced number of mutations per milliliter that has occurred in a culture having r mutants, following LEA and COULSON (1949); Fm is the phenotypic mutation rate expressed in mutations per generation; and -σ is the superhelical density at which the B to Z transition occurs.

*Values are means ± SD; SD on the phenotypic mutation rate is calculated as described in LEA and COULSON (1949) (see MATERIALS AND METHODS).
FIGURE 2.—Two-dimensional agarose gel electrophoresis of the plasmids pUC (GC)$_9$, pUC (GC)$_9$ (PuPy)$_{15}$, pUC (PuPy)$_9$ (GC)$_9$ (PuPy)$_9$ and pUC (PuPy)$_{22}$. The first dimension (indicated by the vertical arrow) was run in absence of intercalating dye and the second one (horizontal arrow) in presence of 2.5 mg/l chloroquine. A, pUC (GC)$_9$; B, pUC (GC)$_9$ (PuPy)$_{15}$; C, pUC (PuPy)$_9$ (GC)$_9$ (PuPy)$_9$; D and D', respective picture and drawing of pUC (PuPy)$_{22}$.

Table 3, Figure 1). The absolute frequency of deletion of $30 \times 10^{-5}$ mutation per generation measured for pUC (GC)$_{15}$ plasmids is decreased to $10 \times 10^{-5}$ and $4 \times 10^{-5}$ for pUC (GC)$_7$ GT (GC)$_7$ and pUC (GC)$_7$ AT (GC)$_7$, respectively (see Table 1). This result can be interpreted in different ways.
table 4

Deletion mutations induced within pUC (GC)6 (PuPy)6 and pUC (PuPy)7 (GC)6 (PuPy)5 plasmids

<table>
<thead>
<tr>
<th>Class</th>
<th>Mutant sequence</th>
<th>Occurrence (%)</th>
<th>Deleted sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>(CG)6 (PuPy)6: (CG)6TGCATGGCTGCGCGCATGCCAGCG3'</td>
<td>16</td>
<td>(CG)2</td>
</tr>
<tr>
<td>II</td>
<td>(CG)5 (PuPy)5: (CG)5GTGCATGGCGCATGCCAGCG3'</td>
<td>5</td>
<td>(GC)5 GT</td>
</tr>
<tr>
<td>II</td>
<td>(CG)6 (PuPy)6: (CG)6GTGCATGGCGCATGCCAGCG3'</td>
<td>68</td>
<td>(GC)6 GT, (GC)2 ATGGCGT</td>
</tr>
<tr>
<td>II</td>
<td>(CG)5 (PuPy)5: (CG)5GTGCATGGCGCATGCCAGCG3'</td>
<td>10</td>
<td>(GC)5 GT</td>
</tr>
</tbody>
</table>

These data result from the sequencing of 19 independent mutants from pUC (GC)9 (PuPy)12 plasmid and of 12 mutants from pUC (PuPy)6 (GC)9 (PuPy)5 plasmid. The mutant and the deleted sequences are represented in the 3' to 5' orientation following the lagging strand hypothesis (see text). The deleted sequences correspond to the "lagging strand model" as represented in Figure 3. The nucleotides originated from the (GC)6 sequence appear in bold characters both in the mutant sequences and in the deleted sequences.

In the context of a mechanism involving a Z-DNA structure, the lower mutation rate of pUC (GC)7 AT (GC)7 is consistent with the lower B to Z transition in this plasmid (Figure 2, Table 1). This reflects the high energetic cost needed to force an ApT dinucleotide to adopt a left-handed conformation comparatively with a GpC or a GpT one (McLean et al. 1986). However, DNA sequence of the mutants are quite different for plasmids pUC (GC)15, pUC (GC)7 GT (GC)7, and pUC (GC)7 AT (GC)7; in the last two plasmids >80% of the deletions are longer than 14 bp, suggesting there are likely to be major differences in the details of the mechanism(s) that are responsible for mutagenesis. Perhaps the identity of the central dinucleotides plays a key role in the intrinsic mechanism of deletion. So, even though the correlation between the mutation rate and the ability to form a Z-DNA structure is good, the qualitative effects of the interruption of the monotony of the GpC run on the deletion spectra cannot be explained by characteristics of the B to Z transition alone as deduced from the two-dimensional agarose gel electrophoresis.

Alternatively, in the context of the slipped pairing between direct repeats, the presence of the dinucleotide interruption in the middle of the sequence is expected to strongly influence misalignment characteristics. The strikingly lower frequency of small deletions in the interrupted repeats in contrast to the perfect repeat led us to consider the number of slippage misalignments available to produce these mutations in all plasmids. Taking as an example the case of ~4 bp, the repeat present in plasmid pUC(GC)15 predicts that perfectly paired misalignments extending for distances as long as 26 bp could be intermediates of mutagenesis. In contrast, the longest such misalignments available to the GT and the AT containing repeat is 11 and 10 bp, respectively. Considered in this light, the low frequency of short deletions seems reasonable. The high frequency of longer deletion in the interrupted plasmid may be explained by the fact that the misaligned intermediates that could have produced small deletions by slipped pairing isomerized to another misaligned configuration producing the longer deletions. Alternatively, the interruption in itself may lead to the enhanced production of intermediates available for the misalignments leading to large deletions. The spectrum of detected deletions giving rise to mutants still containing the central dinucleotide occurred only within pUC(GC)7 GT (GC)7 plasmid (see Figure 1). If these deletions are the result of misalignments within one of the (GC)7 runs, it is unclear why they form more efficiently within (GpC)7 GpT (GpC)7 than within (GpC)7 ApT (GpC)7 sequences.

In the second part of this work, we looked at the influence of the adjacent sequence context on the frameshift mutagenesis within GpC tract. We showed that adding purine-pyrimidine alternating sequences on one or on both sides of a monotonous (GpC)9 run results in a significant (threelfold) increase of the deletion rate (see Table 3). We have sequenced mutations arising both from plasmid pUC (GC)9 (PuPy)15 and pUC (PuPy)5 (GC)9 (PuPy)5, and the resulting sequences are presented on Table 4. The deletions have been filed in two types: type I groups the deletions that occur exclusively within the (GpC)9 tract, whereas in type II mutations, the deletions also involve purine-pyrimidine alternating sequences neighboring the (GpC)9 sequence.

One possible interpretation of these results in view of the data presented in Figure 2 showing that the B to Z transition of the (GpC)9 sequence is favored by adding the (PuPy)5 sequences could suggest a relationship between increasing the length of the sequence that
FIGURE 3.—Possible misalignment explaining the type II deletions within pUC (GC)<sub>9</sub> (PuPy)<sub>13</sub> and pUC (PuPy)<sub>5</sub> (GC)<sub>9</sub> (PuPy)<sub>8</sub> plasmids. The sequences presented correspond to the lagging strand template shown in the 3' to 5' orientation. The sequence repeats giving rise to a deletion are on the same level, underlined with a bold line linked by a dashed line. The secondary structures shown were drawn following the lagging strand hypothesis (see text). The relative frequency of occurrence of each deletion event is indicated below the deletion intermediate.

turns into Z-DNA structure and the mutant sequence specificity. For a given supercoil density of \( \alpha = 0.046 \), we can estimate that 20 bp have turned into Z structure for (GpC)<sub>9</sub> sequence, whereas 34 bp adopt this conformation for either (GpC)<sub>9</sub> (PuPy)<sub>13</sub> or (PuPy)<sub>5</sub> (GpC)<sub>9</sub> (PuPy)<sub>8</sub> sequences. Thus, the extension of the left-handed structure to the neighboring bases should give rise by a putative mechanism involving the processing of the Z-DNA to an increase in the mutation rate and to larger deletions involving sequences 3' or 5' to the (GpC)<sub>9</sub> sequence (see Table 4).

Alternatively, the observed mutations can result from misalignment mechanisms of self-complementary repeats. For both pUC (GpC)<sub>9</sub> (PuPy)<sub>13</sub> or pUC (PuPy)<sub>5</sub> (GpC)<sub>9</sub> (PuPy)<sub>8</sub> plasmids, the type I mutations can be envisioned as the result of misalignments generated either by direct repeat or by inverted repeat slippage. The resulting misalignments lead to the formation of a 4-bp extrahelical loop corresponding to the deleted sequence. Our analysis of the type II mutation events reveals that all of the mutations observed with plasmids pUC (GO)<sub>9</sub> (PuPy)<sub>13</sub> and pUC (PuPy)<sub>5</sub> (GC)<sub>9</sub> (PuPy)<sub>8</sub> can be explained by misalignment mechanisms. Whether the misalignments result from direct repeat slippage or from palindromic misalignments can only be determined for mutant II-3 obtained with plasmid pUC(PuPy)<sub>5</sub> (GpC)<sub>9</sub> (PuPy)<sub>8</sub> (see below). Even though all other mutations can be formally explained by both
1. Pausing of the polymerase

\[
\text{5' CG CG CG CG TG CG CA TG CA CG AGATCT'3'}
\]

\[
\text{5' CCG GC GT AC GC GT GC GC GC GC GC GC AC GC GT AC GC GT GC TCTAGA'}
\]

2. Foldback of a short sequence

\[
\text{5' CC GG GC GT AC GC GT GC GC GC GC GC GC AC GC GT AC GC GT GC TCTAGA'}
\]

3. Polymerisation back and secondary structure formation

\[
\text{5' CC GG GC GT AC GC GT GC GC GC GC GC AC GC GT AC GC GT GC TCTAGA'}
\]

4. Unfolding and misalignment.

\[
\text{5' CGCA CG CG CG CG CG TG CG CA TG CA CG AGATCT'3'}
\]

\[
\text{5' CC GG GC GT AC GC GT GC GC GC GC GC AC GC GT AC GC GT GC TCTAGA'}
\]

Types of misalignments, we shall consider direct repeat slippage to be the mechanism of choice if there are no reasons to explain the deletions in the other way. Following the misalignment hypothesis, the increase of mutation rate within these sequences is due to an increase of the possibilities of misalignment generated by the presence of short GpC repeats neighboring the (GpC) central tract. Theoretically, these misalignments can occur during the replication of the leading or of the lagging strand DNA template. Nevertheless, it has been proposed that slippage can be favored by the presence of single-stranded DNA during the replication of the lagging strand template (Trinh and Sinden 1991). Although there is no evidence supporting this hypothesis in our results, the misalignments shown in Figure 3 result from modeling direct repeat slippage mechanisms during the replication of the lagging strand DNA. In the case of the pUC (GpC) plasmid, the different direct misalignments are drawn to explain the three different type II deletions (see Figure 3). Two of them involve a sequence repeat located 1 bp from the (GpC) tract (mutant II-1 and II-2) representing 73% of the total deletions, whereas the third one involves a sequence 6 bp distant from this tract. Another possible misalignment with a GpCpGpC sequence 5 bp from the end of the (PuPy) sequence was not found, possibly due to different sequence contexts on the 3' end of the considered repeat on the newly synthesized strand. Although the
number of sequences analyzed was small, most of the deletions found for the (PuPy)_5, (GpC)_9, (PuPy)_8 sequence can also be explained by direct repeat slippage mechanisms. The second type deletions II-1 and II-2 can be interpreted as the result of misalignments between sequences from the (GpC)_9 and sequences within (PuPy)_8 neighboring tract. It is interesting to note that within the (PuPy)_5, (GpC)_9, (PuPy)_6 sequence there are misalignment possibilities on both sides relative to the (GpC)_9 sequence but that only one mutant (mutant II-3, see Table 4) is found on the (PuPy)_5 side of this sequence. This mutation cannot be explained by a direct repeat misalignment model. Although this model can account for the deleted bases, the observed sequence change would require an unexplained base pair substitution. However, the observed mutation can be explained by a palindromic misalignment mechanism. Figure 4 shows how a fold-back misalignment mechanism mediated by a palindromic structure could produce the mutation during replication of the leading strand DNA template. In the figure, pausing of the polymerase is hypothesized to occur after copying 10 bp into the alternating GC sequence (copying 14 bp is the largest extent of synthesis compatible with the observed mutation). Misalignment of the free 3' OH end provides a template and a primer of 4 bp (shown in italic) for misaligned synthesis that produces the base substitution and the deletion found. Palindromic misalignments involving a strand-switch can produce the same sequence change and would have to occur during the replication of the leading strand DNA template. The rationale for the relative frequency of the deletion events in both pUC (GpC)_9 (PuPy)_13 and pUC (PuPy)_8, (GpC)_9, (PuPy)_9 plasmids is not clearly established. It has been shown that the number of base pair matches allowed by each precise misalignment play an important role in determining the deletion frequency, both "in vivo" (L. Wang and J. R. Ripley 1994) and "in vitro" (Albertini et al. 1982). We observed that the length of the repeats seems to roughly correlate with the relative occurrence of the mutation events in both plasmids (see Figure 3) (Albertini et al. 1982).

To have further insights into the mechanism of formation of deletion within potential Z-forming sequences, we designed a purine-pyrimidine alternating sequence referred to as (PuPy)_18 that contains no monotonous repeat longer than four nucleotides. The mutation rate of plasmids containing this sequence is only 0.26 × 10^-6 per cell per generation, even though their ability to undergo Z transition is the same as the plasmid bearing (GpC)_8 (PuPy)_13 or (PuPy)_8, (GpC)_9 (PuPy)_8 inserts (see Figure 2). This result unambiguously shows that the ability of the sequence to adopt a Z-DNA structure cannot be the only determinant of the deletion frequency. The low frequency of mutation within (PuPy)_18 sequences compared with GpC monotonous run could result from differences in the Z-DNA structure (Kladde et al. 1994) that allows frameshift mutations. Alternatively, if mutagenesis is produced by slipped mispairing, one can plausibly explain the low deletion rate by the poor misalignment possibilities within this sequence.

CONCLUSIONS

In this work, we propose the possible participation of alternative DNA structures in the genetic instability of (GpC)_9 monotonous runs in E. coli.

The first structure involves the Z-DNA in (GpC)_9-mediated frameshift mutagenesis. By interrupting the (GpC)_9 run or by modifying the sequence context around it, we show that the frequency of frameshift mutation within (GpC)_9 tandem repeats parallels the ability of the sequence to form Z-DNA structure. However, the low mutation frequency obtained with random purine-pyrimidine alternating sequence shows that the monotonous of the GpC repeat plays an important role in the putative Z-DNA structure that could give rise to frameshift mutation.

On the other hand, (GpC)_9 runs allow several DNA structures that are hypothesized to participate in misalignment-induced frameshift mutagenesis during DNA synthesis. (GpC)_9 self-complementary sequences can lead to both direct repeat misalignment and palindromic misalignment. Such mechanisms enabled us to explain at the sequence level the specificity of all the frameshift mutations found, either with interrupted (GpC)_9 sequences or with (GpC)_9 sequences where the neighboring context was modified. Following this pathway, the low mutation rate obtained with random purine-pyrimidine alternating sequence could be explained by reduced misalignment possibilities. Further clues to the proposed frameshift mechanisms will come from the identification of the gene(s) or of the cellular protein(s) involved in the actual mechanism.

We thank Professor Lynn Ripley for helpful discussion and critical reading of the manuscript. This work was supported in part by a grant from the Association pour la Recherche sur le Cancer (ARC grant 6610).

LITERATURE CITED


Communicating editor: N. R. Drinkwater