The EMBO Journal Vol.17 No.17 pp.5095–5102, 1998

Transcription through a simple DNA repeat blocks replication elongation

Maria M.Krasilnikova[1], George M.Samadashwily, Andrey S.Krasilnikov[1] and Sergei M.Mirkin[2]

Department of Molecular Genetics, University of Illinois at Chicago, Chicago, IL 60607, USA

[1]On leave from the Institute of Molecular Genetics, Russian Academy of Sciences, Moscow

[2]Corresponding author
e-mail: mirkin@uic.edu

The influence of $d(G)_n \cdot d(C)_n$ repeats on plasmid replication in *Escherichia coli* cells was analyzed using electrophoretic analysis of replication intermediates. These repeats impeded the replication fork in a length- and orientation-dependent manner. Unexpectedly, the replication arrest relied primarily on the repeats' transcription. When the $d(C)_n$ sequence served as the transcriptional template, both transcription and replication were blocked. This was true for transcription driven by either bacterial or phage RNA polymerases. We hypothesize that the replication fork halts after it encounters a stalled ternary complex of the RNA polymerase, the DNA template and the $r(G)_n$ transcript. This constitutes a novel mechanism for the regulation of replication elongation. The effects of this mechanism on repeat length polymorphism and genome rearrangements are discussed.

Keywords: elongation/replication/simple sequence repeats/termination/transcription

Introduction

Regulation of replication elongation *in vivo* is poorly understood. While the replication fork constitutes an extremely fast and powerful machine (reviewed in Marians, 1992), its progression can be altered or even completely blocked. The best-studied example of such blockage is the termination of bacterial replication which is caused by specific proteins bound to the terminator sites in bacterial DNA (reviewed in Hill, 1992).

Another factor which affects replication elongation is the structure of the DNA template. DNA polymerization *in vitro* is inhibited by stable secondary structures, including hairpins, triplexes and quadruplexes, formed in the DNA template by specific repeats (Abbotts *et al*., 1988; Bedinger *et al*., 1989; Baran *et al*., 1991; Dayn *et al*., 1992; Hacker and Alberts, 1994; Woodford *et al*., 1994; Usdin and Woodford, 1995; Weitzmann *et al*., 1996; Krasilnikov *et al*., 1997). Some of these repeats inhibit DNA replication *in vivo* as well, though the mechanisms of this inhibition are unknown (Rao *et al*., 1988; Brinton *et al*., 1991; Rao, 1994). Recently, we have found that expandable trinucleotide repeats stall the bacterial replication fork *in vivo* (Samadashwily *et al*., 1997). This effect depended on the repeats' length and their orientation relative to the replication origin (*ori*) and was probably due to an unusual structure of the lagging strand template in the region of repeated DNA.

The third possible mechanism affecting replication elongation is based on the fact that replication is approximately an order of magnitude faster than transcription, which makes occasional collisions between the replication fork and the RNA polymerases almost inevitable. Several *in vitro* and *in vivo* studies showed that the replication fork pauses in transcribed regions (French, 1992; Liu and Alberts, 1995; Deshpande and Newlon, 1996), particularly when the directions of replication and transcription are opposite.

Here, we studied the effects of a simple repeat, $d(G)_n \cdot d(C)_n$, on plasmid replication *in vivo*, using electrophoretic analysis of replication intermediates. We found that, like trinucleotide repeats, the $d(G)_n \cdot d(C)_n$ repeat impeded the replication fork progression in a length- and orientation-dependent manner. Unlike the trinucleotide repeats, however, replication blockage was only evident when the $d(G)_n \cdot d(C)_n$ repeat was actively transcribed. Specifically, when the $d(C)_n$ strand served as a template for RNA polymerase, both transcription and replication were blocked. We conclude that the replication block is caused by the collision between the replication fork and an RNA polymerase stalled at the $d(G)_n \cdot d(C)_n$ repeat. This constitutes a novel mechanism for controlling replication elongation. The effects of this mechanism on repeat length polymorphism and genome rearrangements are discussed.

Results

$d(G)_n \cdot d(C)_n$ repeats block replication fork progression in vivo

$d(G)_n \cdot (C)_n$ inserts of varying lengths were cloned into derivatives of the pTrc99A plasmid (Aman *et al*., 1988) which we called pTrc99Δ (Samadashwily *et al*., 1997) (Figure 1). In these vectors, a 1.5 kb fragment containing the *lacI*^{q} coding sequence, the Ptrc transcription unit and the *amp* gene promoter were removed and substituted with a multiple cloning site in two possible orientations. As a result, the *amp* gene came under the control of the PlacI^q, while the multiple cloning sites were located downstream of this promoter but upstream of the translated part of the *amp* gene. The two orientations of the polylinker allowed us to clone $d(G)_n \cdot (C)_n$ inserts in both orientations relative to the replication *ori* and PlacI^q.

We used two-dimensional neutral/neutral gel electrophoresis of replication intermediates (Brewer and Fangman, 1987) to visualize the pattern of the replication fork movement. Since our plasmid replicates unidirec-

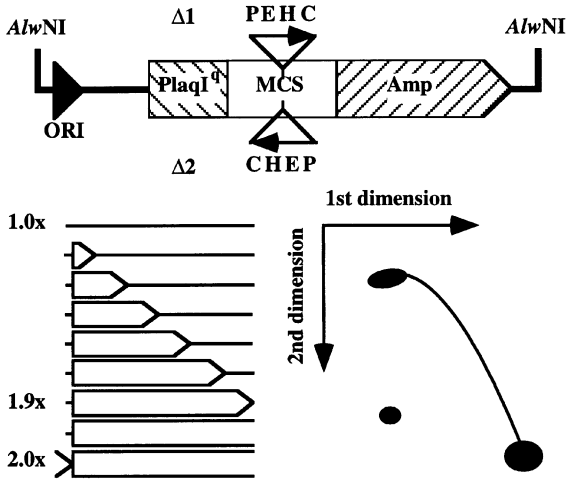

Fig. 1. Schematic representation of two-dimensional neutral/neutral gel electrophoresis for the pTrc99Δ plasmids. The structure of the pTrc99Δ plasmids is shown in the upper panel. The multiple cloning site (containing *Pst*I, *Eco*RI, *Hin*dIII and *Cla*I restriction sites) was inserted in two orientations relative to the *ori*, generating two vectors, pTrc99Δ1 and pTrc99Δ2. The restriction enzyme *Alw*NI cleaves this plasmid at a unique site located upstream of the *ori*. The lower left panel shows that upon *Alw*NI digestion, most of the replicative intermediates are bubble-like, and the size of a bubble reflects the extent of replication. The lower right panel diagrams the separation of bubble-like replicative intermediates by two-dimensional gel electrophoresis.

tionally, digestion of replicating DNA upstream of the *ori* leads to the appearance of bubble-like structures (Figure 1). The size of these bubbles increases with the duration of replication, and they can be separated from the non-replicating DNA by size (first dimension) and size plus shape (second dimension) using agarose gel electrophoresis. This results in a characteristic bubble arc (Figure 1). Stalling of the replication fork at a specific DNA sequence should lead to the accumulation of intermediates of a given length and shape, generating a bulge on a bubble arc.

Since the directionality of both replication and transcription through the $d(G)_n \cdot d(C)_n$ inserts in the pTrc99Δ plasmid was well established, we knew for every recombinant plasmid whether $(G)_n$ or $(C)_n$ clusters were situated in the lagging strand template which is equivalent to the sense strand for transcription. All of our plasmids were named accordingly. The effects of different $d(G)_n \cdot d(C)_n$ inserts on plasmid replication are presented in Figure 2. One can see that with the increase of repeat length, bulges appear on the replication arcs (marked by arrows), which reflect replication fork stalling. It is also evident from this figure that replication inhibition depends on the repeat's orientation in the plasmid. When the $(G)_n$ stretch is located in the lagging strand template and the sense strand for transcription, the stop is evident but weak at $n = 20$ and is very prominent at $n = 32$, while, in the opposite orientation, a weak stop appears only at $n = 32$.

The results of several such experiments were quantitated using the Betascope Blot Analyzer as described in Materials and methods. The relative stop strength (RSS) was calculated as the ratio of a bulge's radioactivity to the radioactivity of a smooth replication arc at this point. This estimates the replication fork retardation at the pause site. Our quantitation show that RSS for pG20 is 3.7 ± 1.0, for pG32 is 27 ± 6, for pC20 is 2.2 ± 1.0 and for pC32 is 3.2 ± 1.0. Thus, in the strongest case, the $(G)_{32}$ stretch in the lagging strand template/transcription sense strand slows down the replication fork ~27-fold.

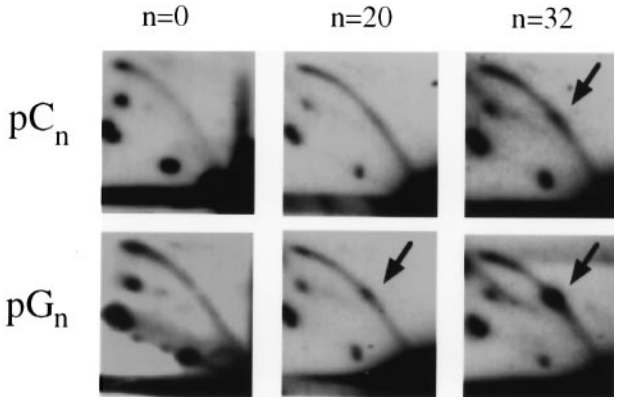

Fig. 2. Electrophoretic analysis of replication intermediates of plasmids with different $d(G)_n \cdot d(C)_n$ inserts. Plasmids pC_n and pG_n have $d(C)_n$ or $d(G)_n$ sequences, respectively, in the lagging strand template and the sense strand for transcription. Arrows show replication stop sites.

To understand whether the above replication stops coincide with the $d(G)_n \cdot d(C)_n$ inserts in our plasmids, we used a modified version of the two-dimensional electrophoretic analysis of replication intermediates (Friedman and Brewer, 1995) as presented in Figure 3A. Specifically, after the first dimension of electrophoresis, replication intermediates were digested with a restriction enzyme in the gel. As a result, a fraction of bubbled intermediates converted into identical y-shaped intermediates. In the second dimension of the electrophoresis, these intermediates migrate similarly and can be detected as a horizontal line upon hybridization with the probe adjacent to the replication *ori*. If the replication fork is stalled within a repeat (replicative intermediate shown in bold), restriction cleavage downstream of it (relative to the *ori*) would leave the bulge on the bubble arc, while cleavage upstream of it would shift the bulge onto the horizontal line.

Our results for the pG32 plasmid are presented in Figure 3B. One can see that cleavage of the replication intermediates with *Eco*RI (located downstream from the insert) leaves the bulge on the bubble arc. By contrast, cleavage by *Hin*dIII shifts the bulge away from the bubble arc. We conclude, therefore, that the replication fork is stalled within the $(G)_{32} \cdot (C)_{32}$ stretch. Note, however, that after *Hin*dIII cleavage the bulge does not co-migrate with the horizontal line, but migrates to a point in between the bubble arc and the horizontal line. Thus, the shape of this intermediate is less compact than the y-shape but more compact than the bubble. To explain this migration pattern, we assume that the portion of the lagging strand around the *Hin*dIII site in stalled replication intermediates was not yet synthesized. This will lead to an incomplete *Hin*dIII digestion and the appearance of butterfly-like DNA molecules (shown in the diagram). If this assumption is correct, our data show the under-replication of the lagging strand within the $d(G)_n \cdot d(C)_n$ sequences.

Replication blockage caused by $d(G)_n \cdot d(C)_n$ repeats is not due to cooperative protein binding

As discussed above, replication pausing or termination is caused by DNA–protein binding in many cases. Although

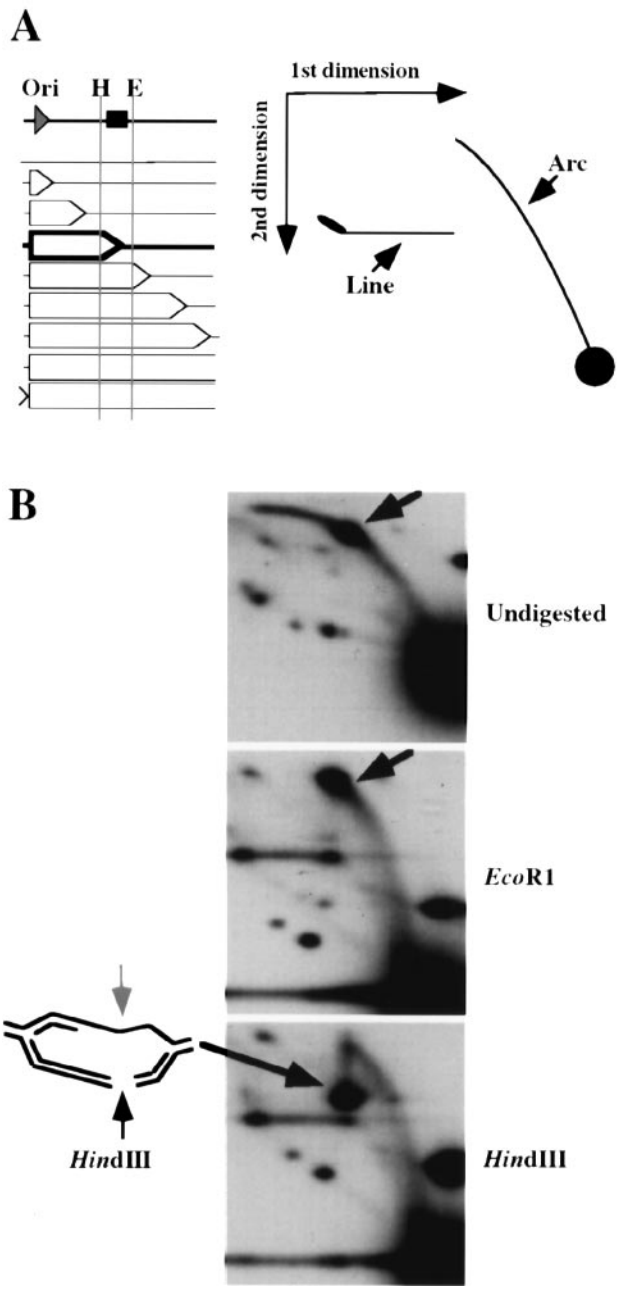

Fig. 3. Mapping of the replication stop sites in the $p(G)_{32}$ plasmid. (**A**) Schematic representation of two-dimensional gel electrophoresis upon restriction cleavage after the first dimension. The filled square shows the $d(G)_{32} \cdot d(C)_{32}$ insert. The dotted vertical lines show the restriction sites upstream and downstream from the insert. The stalled replicative intermediate is in bold. The right panel shows an arc-to-line transition reflecting the conversion of bubble-shaped intermediates into y-shaped intermediates. (**B**) Electrophoretic separation of replication intermediates. Upper panel, no digestion; middle panel, *Eco*RI digestion; bottom panel, *Hin*dIII digestion. Small arrows show the stop sites on the bubble arc. Long arrows show stalled intermediates moving towards the horizontal line. A schematic representation of this intermediate is presented (see text for details).

the length and orientation dependence of the $d(G)_n \cdot d(C)_n$-caused replication blockage makes this explanation unlikely, it does not rule it out. Indeed, the existence of a cellular protein which binds strongly and cooperatively to these repeats might explain the length dependence. Protein-associated termination of DNA replication is also known to depend on orientation (Hill, 1992).

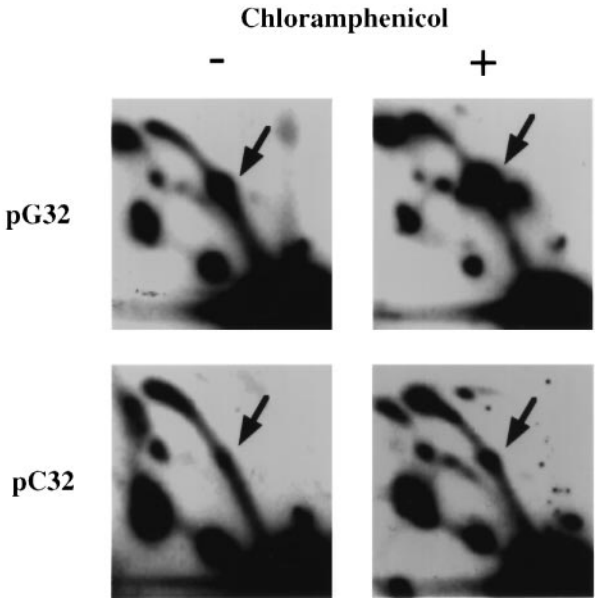

Fig. 4. The effects of chloramphenicol on plasmid replication. Plasmids are designated as in Figure 2. Cells were grown without chloramphenicol (–) followed by a 7 h incubation with chloramphenicol (+). Arrows show replication stop sites.

To address this concern, we utilized a long-known phenomenon of plasmid amplification (Clewell, 1972). While protein synthesis *de novo* is required for the initiation of bacterial DNA synthesis, it is not necessary for ColE1-type plasmids. Consequently, in the presence of the protein synthesis inhibitor, chloramphenicol, plasmids amplify rapidly. This gross increase in the plasmid copy number occurs in cells where the protein content is at best stagnant. Thus, if cooperative protein binding is responsible for the replication blockage, one would expect that the stops would decrease in intensity upon chloramphenicol treatment.

Figure 4 shows our experimental data for the replication of the pG32 and pC32 plasmids in cells that underwent prolonged (7 h) chloramphenicol treatment where the plasmid copy number increased ~10-fold. One can clearly see that chloramphenicol not only fails to abolish replication stops, it significantly enhances them. The quantitation of these results, as described above, showed that, in the presence of this antibiotic, the relative stop strength for both the pG32 and pC32 plasmids increased ~6-fold. To explain this enhancement of replication blockage by protein binding, one must assume that the DNA-binding constant of such a protein increases ~60-fold under chloramphenicol treatment, which is highly unlikely. We conclude, therefore, that the cooperative protein binding to the $d(G)_{32} \cdot d(C)_{32}$ target is not responsible for the replication arrest. In fact, we believe that the opposite is true: depletion of some protein(s) involved in DNA metabolism after prolonged chloramphenicol treatment leads to a stronger replication blockage. This notion is supported by our observations that the relative stop strength for our plasmids increases with chloramphenicol incubation time (data not shown).

Replication blockage relies on transcription of $d(G)_n \cdot d(C)_n$ repeats

The profound dependence of the $d(G)_n \cdot d(C)_n$-caused replication arrest on the repeats' orientation in a plasmid

could be explained in two ways. First, since our plasmids replicate unidirectionally, either the $d(G)_n$ or the $d(C)_n$ strand serves as the lagging strand template, depending on the repeat's orientation relative to the *ori*. Thus, orientation dependence may arise from differences between the modes of leading and lagging strand synthesis through repeated DNA. This seems to be the case for trinucleotide repeats, where the strength of the replication blockage correlates with the structural potential of the repeat in the lagging strand template (Samadashwily *et al*., 1997). Second, our $d(G)_n \cdot d(C)_n$ inserts are located in a transcribed area. Since transcription is asymmetric, it is possible that the orientation dependence is mediated by transcription.

In order to distinguish between these opportunities, we modified the original pG32 and pC32 plasmids as shown in Figure 5A. In ΔPlac plasmids, the *lac*I^q promoter was deleted, while the *amp* gene was put under the control of its own promoter P3 (Balbas *et al*., 1986). Thus, $d(G)_{32} \cdot d(C)_{32}$ repeats became positioned upstream of the *amp* promoter in the non-transcribed area, while their orientation relative to the *ori* remained unchanged. InvOri plasmids differ from the original pG32 and pC32 plasmids by the inversion of the *ori*. Therefore, the repeat's position in the transcription unit remained unchanged, while the leading and lagging strand templates for replication switched.

The results of the electrophoretic analysis of replication intermediates for these plasmids are shown in Figure 5B. One can see that the efficiencies of replication blockage for pG32InvOri and pC32InvOri are indistinguishable from those for pG32 and pC32, respectively. Thus, the efficiency of the replication blockage is not determined by the repeats' orientation with regard to the leading/lagging strand synthesis. On the other hand, we observed a dramatic decrease in the replication blockage for pG32ΔPlac in comparison with the pG32 plasmid. In fact, the replication stop for the pG32ΔPlac plasmid is as weak as the stops observed for pC32 or pC32ΔPlac plasmids. Thus, abolishing transcription through the $d(G)_{32} \cdot d(C)_{32}$ repeat made replication blockage inefficient.

These unexpected data could be explained by assuming that transcription stalls within or immediately after the $d(G)_n \cdot d(C)_n$ repeats, and the stalled transcription complex represents a barrier for replication fork progression. The length dependence of the effect could then be explained by increased transcriptional stalling for longer repeats, while the orientation dependence might be due to the higher incidence of RNA polymerase stalling and/or stability of the stalled complex when the RNA transcript is $r(G)_n$.

If the RNA polymerase is indeed stalled at a $d(G)_n \cdot d(C)_n$ repeat, this should lead to the accumulation of a truncated RNA transcript of a defined length. The appearance of such a transcript can be analyzed directly using Northern hybridization. The plasmids chosen for this analysis are presented in Figure 6A. They contain both the PlacI^q promoter driving transcription through the $d(G)_{32}.d(C)_{32}$ repeat, and the P3 promoter (Balbas *et al*., 1986) directing the *amp* gene transcription. Transcription which originated at the PlacI^q stops at the T1 terminator from the *Escherichia coli rrn*B gene (Brosius *et al*., 1981). Addition of this terminator (including a portion of the 5S RNA) to the 3′ end of our RNA transcript led to a substantial decrease in its degradation, which made analysis of truncated transcripts possible. In the resultant plasmids, the full-length transcripts which originated from the PlacI^q promoter are either 1110 or 1150 nucleotides long (depending on the presence of an insert), while transcripts abrogated by the $d(G)_{32} \cdot d(C)_{32}$ repeat should be ~910 nucleotides long (depending on the exact position of the polymerase blockage).

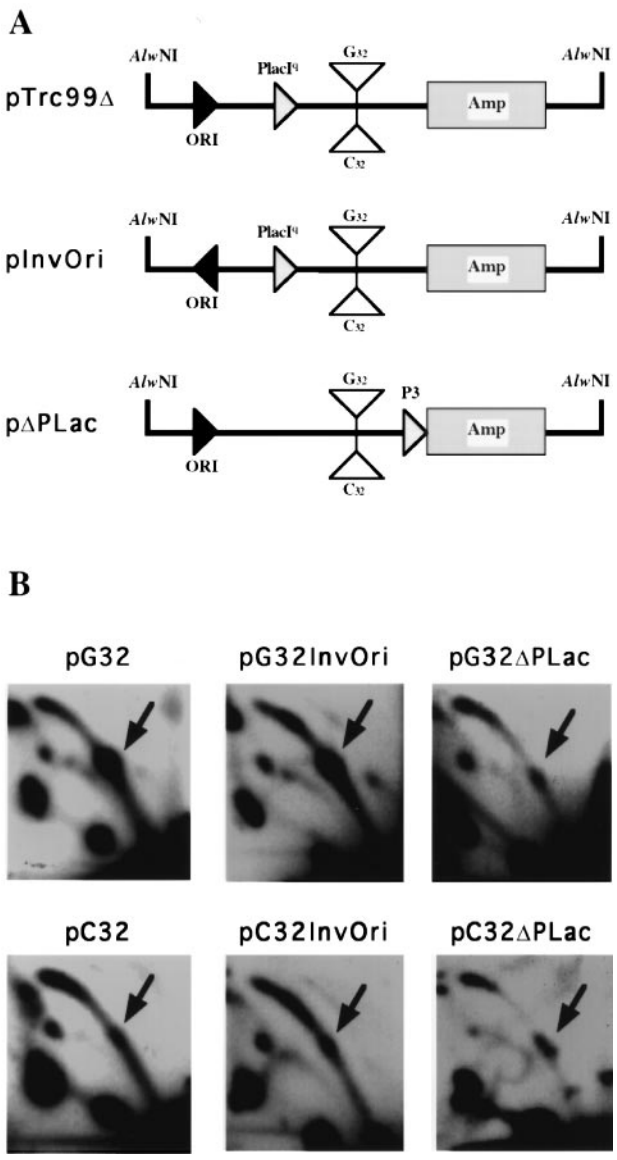

Fig. 5. Effects of the replication direction and transcription through the $d(G)_{32} \cdot d(C)_{32}$ insert on plasmid replication. (**A**) Schematic representation of pTrc99Δ and its pInvOri and pΔPlac derivatives carrying the $d(G)_{32} \cdot d(C)_{32}$ insert in two orientations (see text for details). (**B**) Electrophoretic separation of replication intermediates of the corresponding plasmids. Arrows show the replication stop sites.

RNAs isolated from cells bearing the above plasmids were analyzed by Northern hybridization with the probe corresponding to the 5′-part of the *lac*I^q transcription unit (Figure 6B). Transcription of a control plasmid with no insert results in a single full-length transcript. In contrast, transcription of the pG32* plasmid leads to the appearance of two distinct RNA bands: a full-length transcript and

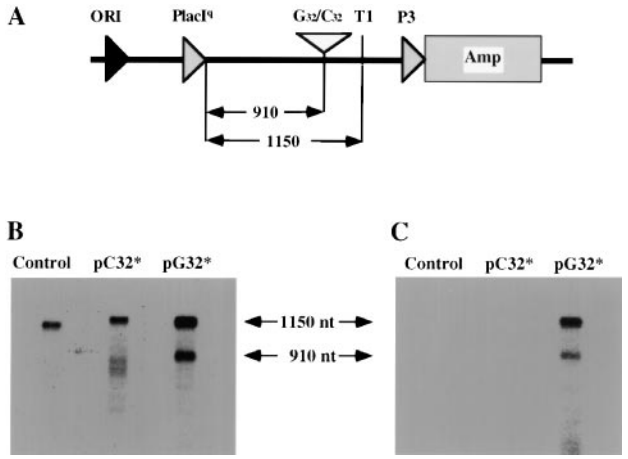

Fig. 6. Effects of the $d(G)_{32} \cdot d(C)_{32}$ insert on transcription *in vivo*. (**A**) Schematic representation of the plasmids used for the RNA analysis. They contain the $d(G)_{32} \cdot d(C)_{32}$ insert in two orientations downstream of the *lac*I^q promoter and upstream of the T1 terminator from the 5S RNA gene (see text for details). Plasmids pG32* and pC32* contain either the $d(G)_{32}$ or the $d(C)_{32}$ stretch, respectively, in the sense strand for transcription. (**B**) Northern blot hybridization of RNA isolated from cells bearing different plasmids with the probe corresponding to the 5′ part of the *lac*I^q gene. (**C**) Northern blot hybridization of the same RNA with the $d(C)_{14}$ probe.

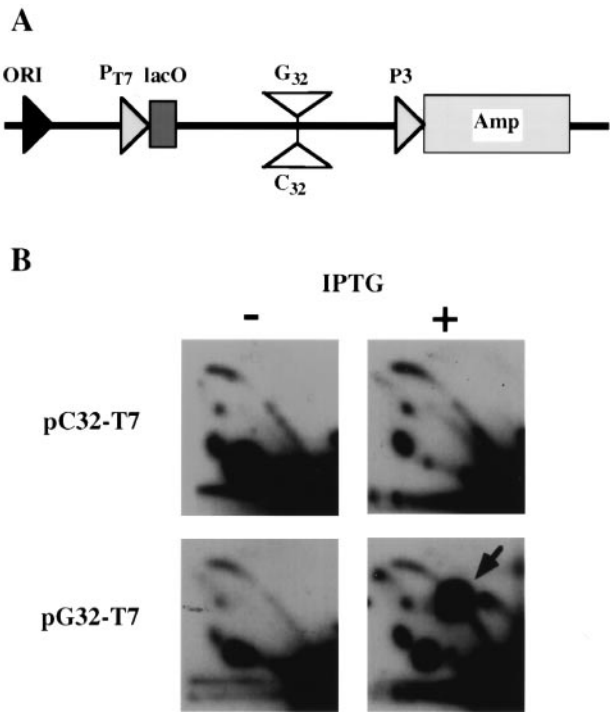

Fig. 7. Effects of transcription through the $d(G)_{32} \cdot d(C)_{32}$ insert mediated by the phage polymerase on plasmid replication. (**A**) Schematic representation of plasmids pC32-T7 and pG32-T7. They contain either the $d(C)_{32}$ or the $d(G)_{32}$ repeat, respectively, in the sense strand for transcription, driven by T7 RNA polymerase in the presence of IPTG (see text for details). (**B**) Electrophoretic separation of replication intermediates of the corresponding plasmids. The arrow shows a replication stop site which is evident only when $d(G)_{32}$ is in the sense strand for transcription.

an ~900 nucleotide long RNA, reflecting arrest at the $d(G)_{32} \cdot d(C)_{32}$ repeat. Quantitative analysis using a Betascope 603 analyzer confirmed that the efficiency of this transcription arrest is ~30%. The fact that transcription arrest is incomplete explains why cells bearing the original pG32 plasmid retained ampicillin resistance, notwithstanding the insert within the *amp* transcription unit. Finally, transcription of the pC32* plasmid results in a distinct full-length transcript and a mixture of truncated RNAs of varying lengths and relatively low intensity.

The same membrane was also hybridized with an oligo $d(C)_n$ probe. The results presented in Figure 6C show that both the full-length and truncated transcript generated from pG32* contain the $r(G)_n$ stretch. We conclude, therefore, that transcription terminates within or immediately after the $d(G)_{32} \cdot d(C)_{32}$ sequence.

Though plasmids used for RNA analysis differed from those used in replication studies by the presence of the T1 terminator and the P3 promoter, these differences did not affect the strength and orientation dependence of the $d(G)_{32} \cdot d(C)_{32}$-caused replication stops (data not shown). Thus, there exists an obvious correlation between the transcription and replication blockage: smooth transcription and replication for the control plasmid, strong transcription and replication arrest when the $d(G)_n$ repeat is in a sense strand for transcription, and modest transcription and replication blockage for the $d(C)_n$ repeat in a sense strand for transcription. We believe, therefore, that the replication arrest at $d(G)_n \cdot d(C)_n$ repeats is caused by transcription blockage within these repeats.

To confirm this additionally, we studied the replication of plasmids carrying the $d(G)_n \cdot d(C)_n$ repeats under the control of the phage T7 RNA polymerase. For this purpose, we utilized a modified pET system (Studier *et al.*, 1990). We constructed mini-pET derivatives where the $d(G)_{32} \cdot d(C)_{32}$ repeat in two orientations was cloned upstream of the *amp* gene promoter but downstream of the T7 promoter–*lac*O cassette (Figure 7A). The resultant plasmids were introduced into *E.coli* NovaBlue (DE3) cells (Novagen). These cells contain a λDE3 prophage with the T7 polymerase gene under the control of the *lac*UV5 promoter, and the lacI^q repressor on the F′ episome. As a result, the T7 polymerase level is low in the absence of the transcriptional inducer isopropyl-β-D-thiogalactopyranoside (IPTG), and transcription from T7 promoters is inefficient. To achieve even stronger levels of T7 polymerase repression in these cells, we additionally introduced the pLysS plasmid (Novagen) encoding a natural inhibitor of the phage polymerase, T7 lysozyme (Studier *et al.*, 1990). Thus, in the absence of IPTG, transcription from the T7 promoter was virtually nonexistent, while the addition of the IPTG led to a rapid and powerful activation of the T7 promoter.

Electrophoretic analysis of replication intermediates of plasmids bearing $d(G)_{32} \cdot d(C)_{32}$ repeats under the control of the T7 promoter are shown in Figure 7B. There are no replication stops in the case of the plasmid carrying the $d(C)_{32}$ stretch in the sense strand for transcription with or without IPTG. In contrast, the addition of IPTG to cells with the plasmid carrying the $d(G)_{32}$ stretch in the sense strand for transcription led to the appearance of an enormously strong stop signal on the replication arc. We conclude that transcription through $d(G)_n \cdot d(C)_n$ repeats by the phage RNA polymerase causes profound replication blockage. This block depends on the repeat's orientation, being pronounced only when the $(G)_n$ stretch is within

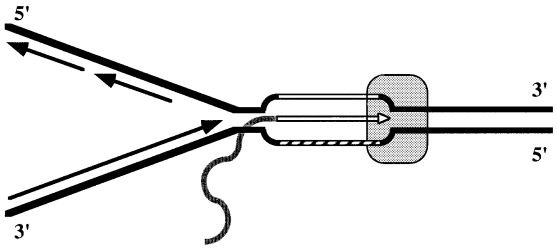

Fig. 8. Model for replication blockage caused by transcription through $d(G)_n \cdot d(C)_n$ repeats. Stalled RNA polymerase is shown by a gray oval. The $(G)_n$ stretches in DNA and RNA chains are depicted as open lines, while the $d(C)_n$ stretch is depicted as a striped line. Arrows show the 3′ ends of the newly synthesized DNA and RNA chains. Transcriptional stall is believed to be caused by the formation of a stable complex between the G-rich RNA chain and its DNA template. The exact structure of this three-stranded complex remains to be established.

the RNA transcript. These results are generally consistent with the data presented in Figure 5, but the effect is much stronger. The latter could be attributed to the higher selectivity and activity of the phage RNA polymerase, compared with its bacterial counterpart (Studier *et al*., 1990).

Discussion

We show that $d(G)_n \cdot d(C)_n$ repeats block the replication fork progression *in vivo*. The efficiency of the replication blockage depends on the repeats' length, their transcription and orientation within the transcription unit, but does not depend on the orientation relative to the replication *ori*. Transcription, in turn, is arrested by the $d(G)_n \cdot d(C)_n$ repeats as indicated by the accumulation of the specifically truncated transcripts. There is an obvious correlation between transcription and replication blockage. These results suggest the existence of a novel mechanism attenuating the replication elongation. We believe that it comprises a collision between the replication fork and an RNA polymerase stalled at a repeated DNA stretch (Figure 8).

RNA polymerase is stalled at $d(G)_n \cdot d(C)_n$ stretches specifically when the $d(C)_n$ stretch serves as a transcription template. Truncated transcripts hybridize with the $d(C)_n$ probe, indicating that they contain oligo $r(G)_n$. We believe, therefore, that transcription stops within or immediately after $d(G)_n \cdot d(C)_n$ repeats.

This hypothesis is in agreement with the published data on transcription elongation *in vitro*. It is generally acknowledged that it is not monotonous, being affected by structures of both the DNA template and the RNA transcript (reviewed in Uptain *et al*., 1997). More specifically, RNA polymerase was shown to stall at several homopurine–homopyrimidine repeats in an orientation-dependent manner (Reaban and Griffin, 1990; Reaban *et al*., 1994; Grabczyk and Fishman, 1995; Kiyama and Oishi, 1996). This happens when a transcript carries either an oligopurine stretch for G-rich repeats or an oligopyrimidine stretch for A-rich repeats. It was suggested that the stall could be due to the formation of a stable complex between a transcript and its DNA template. Though the exact nature of this complex remains unclear, several structures, such as parallel-stranded triplex (Reaban *et al*., 1994), transcript-stabilized H-DNA (Reaban and Griffin, 1990; Grabczyk and Fishman, 1995), collapsed R-loop (Reaban *et al*., 1994) etc., were discussed. Here we studied a 100% GC-rich homopurine–homopyrimidine repeat. Thus, it is plausible to speculate that transcription arrest *in vivo* observed by us depends on the formation of a complex between the $(G)_n$ stretch in the transcript and the $d(G)_n \cdot d(C)_n$ repeat in the DNA template.

As a result, the replication fork might collide with the transcription complex stalled at $d(G)_n \cdot d(C)_n$ repeats. Collisions between the replication fork and RNA polymerase were analyzed in several studies. Alberts and co-authors found that when the T7 RNA polymerase is stalled at a specific DNA site *in vitro*, reconstituted T7 replication fork bypasses it, but after a pause of several seconds (Liu *et al*., 1993; Liu and Alberts, 1995). Several *in vivo* studies imply that replication pauses within heavily transcribed DNA segments (French, 1992; Deshpande and Newlon, 1996). Thus, when transcription is stalled by $d(G)_n \cdot d(C)_n$ repeats *in vivo*, one might expect replication delays. In the studies described in French (1992) and Liu and Alberts (1995), replication delays were most prominent when replication and transcription proceeded in opposite directions. Our data show, by contrast, that a replication fork is stalled within the transcribed repeat notwithstanding the direction of replication. We believe, therefore, that it is not the stall *per se*, but the overall structure of a stalled ternary complex, consisting of RNA polymerase, the $r(G)_n$ transcript and the $d(G)_n \cdot d(C)_n$ DNA template, that halts DNA replication.

How can this hypothesis explain an enhancement of the replication blockage in the presence of chloramphenicol? First, the $d(G)_n \cdot d(C)_n$-caused transcriptional stall might be stronger in the presence of chloramphenicol due to the depletion of some factors essential for transcriptional elongation or to the uncoupling of transcription and translation. Indeed, our RNA analysis indicates that the relative amount of the truncated transcript is 80% in the presence of this antibiotic compared with 30% in untreated cells (data not shown). However, this 2.5-fold increase in transcription stalling cannot fully account for the 6-fold increase in replication arrest. Second, roadblocks on the replication path are counteracted by replication helicases and other accessory proteins (reviewed in Kornberg and Baker, 1992). In the presence of chloramphenicol, where plasmid DNA replicates rapidly while new proteins are not produced, these accessory replication proteins might be in deficit, resulting in an inability to bypass barriers in DNA templates.

It has been reported previously that transcription of certain repeats leads to their instability, and several possible mechanisms of this instability were discussed (Jaworski *et al*., 1989; Kiyama and Oishi, 1994; Bowater *et al*., 1997). Our data prove that one such mechanism is the blockage of the repeats' replication, mediated by their transcription. It is clear that the $d(G)_n \cdot d(C)_n$ sequence represents only one of many possible repeats implicated in this mechanism. Many other homopurine–homopyrimidine stretches, which are enormously overrepresented in the genomic DNA (Cox and Mirkin, 1997), might lead to similar replication attenuation. One particularly interesting candidate is the sequence $(AGGAG)_{28}$ from the murine IgA switch region (Stanton and Marcu, 1982). Transcription of this repeat depends on its orientation, much like $d(G)_n$

(Reaban *et al*., 1994), and we expect it to cause replication arrest as well. Another interesting repeat is $(CGCGGGGC-GGGG)_n$, expansion of which leads to the progressive myoclonus epilepsy (Lalioti *et al*., 1997). Our recent results show that transcription through this repeat blocks DNA replication (data not shown). We believe, therefore, that attenuation of replication elongation mediated by transcription of repeated DNA might be rather common.

What might be the biological outcomes of this mechanism? First, it can contribute significantly to the length polymorphism of DNA repeats situated in transcribed areas (Jaworski *et al*., 1989; Kiyama and Oishi, 1994; Bowater *et al*., 1997). Another provocative opportunity is that stalling of the replication fork caused by transcription of repeated DNA generates DNA ends that potentially are highly recombinogenic. This may contribute to the well-documented stimulation of genetic recombination by transcription (reviewed in Gangloff *et al*., 1994), as well as to the recombinational hot-spot activity of some DNA repeats (Wahls *et al*., 1990; Rooney and Moore, 1995, and references therein). This may be particularly relevant in the case of immunoglobulin switch recombination discussed above, which is driven by the transcription of repeated regions (Stavnezer *et al*., 1988). Finally, in the notoriously long eukaryotic genes, collision of the replication and transcription machinery is almost inevitable. Our mechanism might prevent the replication of genes that are undergoing active transcription.

Materials and methods

Plasmids

Plasmids pTrc99Δ1 and pTrc99Δ2 were obtained by substituting an *Eco*RV–*Ssp*I fragment of pTrc99A (Aman *et al*., 1988) with a *Sma*I–*Hin*cII portion of the pBluescript SK(–) (Stratagene) multiple cloning site in two orientations (Figure 1). Plasmid pCATCG32 containing the $d(G)_{32} \cdot d(C)_{32}$ insert between the *Sac*I and *Sma*I sites of the pUC13 polylinker (Kohwi *et al*., 1992) was kindly provided by Dr Sergei Malkhosyan. Plasmids pG32 and pC32 were obtained by cloning the *Eco*RI–*Hin*dIII fragment of the plasmid pCATCG32 into the *Eco*RI–*Hin*dIII sites of pTRC99Δ2 and pTRC99Δ1, respectively. Plasmids pG20 and pC20 were obtained by substituting the *Eco*RI–*Xba*I fragments of plasmids pG32 and pC32, respectively, with the synthetic duplex:

5′-AATTCGAGCTC$_{20}$GGGGATCCT-----3′
3′-----GCTCGAG$_{20}$CCCCTAGGAGATC-5′

Plasmids pG32InvOri and pC32InvOri were obtained as follows. pG32 and pC32, respectively, were digested by *Bsp*HI and *Nsi*I, blunt-ended by the Klenow fragment and then re-ligated. Clones with the opposite orientation of the origin were detected by restriction analysis.

Plasmid pG32ΔPLac was constructed in two steps. First, a P3 promoter was inserted upstream of the ampicillin resistance gene. To this end, the *Alw*NI–*Eco*RI fragment of pG32 was replaced with the P3-containing *Alw*NI–*Eco*RI fragment from the mini-pBR described by us in Samadashwily *et al*. (1997). In the resultant plasmid, we substituted the *Nsi*I–*Mlu*I fragment (bearing the *lac*I^q promoter) with the PCR-obtained fragment which was identical to the *Nsi*I–*Mlu*I except for the 70 bp containing PlacI^q.

Plasmids pG32* and pC32*, used for RNA analysis, were obtained as follows. pG32* was obtained by substituting the *Sac*I–*Eco*O109I fragment of pG32-P3 by the *Bsr*BI–*Sac*I fragment from pTrc99A, which contains the T1 terminator region from the *rrn*B gene. pC32* was obtained by substituting the *Cla*I–*Bam*HI fragment of pG32* with the *Eco*RI–*Cla*I fragment of pC32. The control plasmid was obtained by deleting the *Pst*I–*Ecl*136II fragment, containing the poly(G) insert, from pG32*.

Plasmids pG32-T7 and pC32-T7 were obtained by cloning the *Eco*RI–*Hin*dIII fragment from pG32 between the *Eco*RI and *Hin*dIII sites of our vectors pT7-1 and pT7-2, respectively. The latter vectors were obtained by inserting the *Xba*I–*Bgl*II fragment from pET-15b, containing the T7 promoter and *lac* operator, into the *Ehe*I site of the mini-pBR derivatives described earlier (Samadashwily *et al*., 1997).

Bacteria

All plasmids were maintained in either the XL1-Blue (Stratagene) or NovaBlue (DE3) (pLysS) (Novagen) *E.coli* strain. For the isolation of replication intermediates, bacteria were grown in LB medium with 100 μg/ml of ampicillin at 37°C until the mid-logarithmic stage (A_{600} ~0.6). For the experiments with protein synthesis inhibition, cells were grown in the same conditions until early logarithmic stage (A_{600} ~0.2) followed by the addition of chloramphenicol up to 170 μg/ml. For transcription studies, bacteria were grown until A_{600} ~0.2, IPTG (Sigma) was added up to 2 mM, and cells were incubated for another 1.5 h.

Isolation of replication intermediates

The method used was according to Martin-Parras *et al*. (1991). A bacterial cell suspension was cooled rapidly with ice-cold physiological solution (0.9% NaCl) and collected by centrifugation at 6000 *g* for 7 min. The cell pellet was resuspended in 2.5 ml of cold Tris–sucrose buffer (25% w/v sucrose, 0.25 M Tris–HCl pH 8.0, 10 mg/ml lysozyme, 0.1 mg/ml RNase A) and incubated on ice for 5 min. Then 1 ml of 0.25 M EDTA was added and, after a brief, gentle shake, the cell suspension was incubated on ice for another 5 min. Cell lysis was achieved by the addition of 4 ml of lysis buffer (1% v/v Brij-58, 0.4% w/v sodium deoxycholate, 63 mM EDTA, 50 mM Tris–HCl pH 8.0). Lysed cells were incubated on ice for 10–20 min, and plasmid DNA was separated from the cell debris and chromosomal DNA by centrifugation at 36 000 *g* for 1 h at 4°C. Plasmid DNA was precipitated from the supernatant by 5 ml of precipitation solution (1.5 M NaCl, 25% PEG 8000). Co-precipitated proteins were then hydrolyzed in a deproteinization solution (1 M NaCl, 10 mM Tris–HCl pH 9.0, 1 mM EDTA, 0.1% w/v SDS, 100 mg/ml proteinase K) for 20 min at 65°C, followed by phenol/chloroform extraction. Plasmid DNA was then precipitated with ethanol.

Two-dimensional gel electrophoresis

Replication intermediates were digested by *Alw*NI, loaded on a 0.4% agarose gel in 1× TBE and run at 1 V/cm for 24 h at room temperature allowing the DNA to separate according to its molecular weight. The slice of the gel containing DNA fragments corresponding to 1–2 plasmid sizes was cut out and implanted into a 1% agarose gel in 1× TBE with 0.6 μg/ml of ethidium bromide. The second direction was run perpendicularly to the first one for 5–6 h at 4–5 V/cm at 4°C with continuous buffer recirculation. The agarose gel containing replication intermediates was transferred onto a positively charged nylon membrane (Zeta-Probe GT) and hybridization was carried out with a plasmid probe labeled by a random prime labeling kit (Gibco-BRL).

Mapping of the replication stop sites

After the first dimension of the electrophoresis, the slice of the 0.4% agarose gel (Sea Kem LE) containing replication intermediates was soaked twice in TE for 30 min and twice in 1× restriction buffer for 30 min at room temperature. In-gel digestion was carried out in 5 ml of 1× restriction buffer containing 3000 U of an appropriate restriction enzyme for 15 h at 37°C. Subsequently, the gel slice was washed twice with 15 ml of TE and imbedded in 1.5% agarose for the second dimension of the electrophoresis.

Quantitative analysis of replication stops

Replication arcs were quantitated using the Betascope 603 Blot Analyzer (Betagen Corp.). The RSS was determined as the number of counts in this stop divided by the number of counts in the corresponding area of a smooth replication arc.

RNA analysis

RNA was isolated by a modification of the method of Dennis and Nomura (1975). Approximately 5×10^8 cells were collected by centrifugation. The cell pellet was resuspended in 50 ml of ice-cold 100 mM NaCl, followed by the immediate addition of 350 μl of 100 mM NaCl, 20 mM Tris–HCl, pH 7.4, 8 mM EDTA, 1% SDS (pre-heated to boiling point) and 400 μl of water-saturated phenol, pH 4.3. After vigorous shaking, the phases were separated by centrifugation. The aqueous phase was re-extracted by phenol/chloroform and then chloroform, followed by precipitation with ethanol. The pellet was dissolved in RNase-free water.

RNA samples were denatured in 2.2 M formaldehyde, 1× MOPS, 50% (v/v) deionized formamide, followed by separation in 1.5% agarose gel with 2.2 M formaldehyde in a 1× MOPS buffer.

The gel was vacuum blotted onto a nylon membrane (Zeta-Probe GT) and hybridized with a plasmid probe labeled by a random prime labeling kit (Gibco-BRL).

Acknowledgements

We thank Jim Dahlberg, Cynthia McMurray, Robert Reeder, Mark Groudine, Igor Panyutin, Ed Grabczyk, Karen Usdin and Jim Maher for stimulating discussions, Sergei Malkhosyan for providing us with the plasmid pCATCG32, Victor Levenson for his helpful suggestions on electrophoretic separation of replication intermediates, Alexander Mankin for sharing his thoughts on RNA isolation and stability, and Randal Cox for critical reading of the manuscript. This work was supported by grants from the National Institutes of Health (GM54247) and the National Science Foundation (MCB-9723924) to S.M.M. M.M.K. was supported in part by the Office of International Affairs of the National Cancer Institute.

References

Abbotts,J., SenGupta,D.N., Zon,G. and Wilson,S.H. (1988) Studies on the mechanisms of *Escherichia coli* DNA polymerase I large fragment. Effect of template sequence and substrate variation on termination of synthesis. *J. Biol. Chem.*, **263**, 15094–15103.

Aman,E., Ochs,B. and Abel,K.-J. (1988) Tightly regulated tac promoter vectors useful for the expression of unfused and fused proteins in *Escherichia coli*. *Gene*, **69**, 301–315.

Balbas,P., Soberon,X., Merino,E., Zurita,M., Lomeli,H, Valle,F., Flores,N. and Bolivar,F. (1986) Plasmid vector pBR322 and its special-purpose derivatives—a review. *Gene*, **50**, 3–40.

Baran,N., Lapidot,A. and Manor,H. (1991) Formation of DNA triplexes accounts for arrests of DNA synthesis at $d(TC)_n$ and $d(GA)_n$ tracts. *Proc. Natl Acad. Sci. USA*, **88**, 507–511.

Bedinger,P., Munn,M. and Alberts,B.M. (1989) Sequence-specific pausing during *in vitro* DNA replication on double stranded DNA templates. *J. Biol. Chem.*, **264**, 16880–16886.

Bowater,R.K., Jaworski,A., Larson,J.E., Parniewski,P. and Wells,R.D. (1997) Transcription increases the deletion frequency of long CTG·CAG triplet repeats from plasmids in *Escherichia coli*. *Nucleic Acids Res.*, **25**, 2861–2868.

Brewer,B.J. and Fangman,W.L. (1987) The localization of replication origins on ARS plasmids in *S.cerevisiae*. *Cell*, **51**, 463–471.

Brinton,B.T., Caddle,M.S. and Heintz,N.H. (1991) Position and orientation-dependent effects of a eukaryotic Z-triplex DNA motif on episomal DNA replication in COS-7 cells. *J. Biol. Chem.*, **266**, 5153–5161.

Brosius,J., Dull,T.D., Sleeter,D.D. and Noller,H.F. (1981) Gene organization and primary structure of a ribosomal RNA operon from *Escherichia coli*. *J. Mol. Biol.*, **148**, 107–127.

Clewell,D.B. (1972) Nature of Col E1 plasmid replication in *Escherichia coli* in the presence of chloramphenicol. *J. Bacteriol.*, **110**, 667–676.

Cox,R. and Mirkin,S.M. (1997) Characteristic enrichment of DNA repeats in different genomes. *Proc. Natl Acad. Sci. USA*, **94**, 5237–5242.

Dayn,A., Samadashwily,G.M. and Mirkin,S.M. (1992) Intramolecular DNA triplexes: unusual sequence requirements and influence on DNA polymerization. *Proc. Natl Acad. Sci. USA*, **89**, 11406–11410.

Dennis,P.P. and Nomura,M. (1975) Regulation of the expression of ribosomal protein genes in *Escherichia coli*. *J. Mol. Biol.*, **97**, 61–76.

Deshpande,A.M. and Newlon,C.S. (1996) DNA replication fork pause sites dependent on transcription. *Science*, **272**, 1030–1033.

French,S. (1992) Consequences of replication fork movement through transcription units *in vivo*. *Science*, **258**, 1362–1365.

Friedman,K.L. and Brewer,B.J. (1995) Analysis of replication intermediates by two-dimensional agarose gel electrophoresis. *Methods Enzymol.*, **262**, 613–627.

Gangloff,S., Lieber,M.R. and Rothstein,R. (1994) Transcription, topoisomerases and recombination. *Experientia*, **50**, 261–269.

Grabczyk,E. and Fishman,M.C. (1995) A long purine–pyrimidine homopolymer acts as a transcriptional diode. *J. Biol. Chem.*, **270**, 1791–1797.

Hacker,K.J. and Alberts,B.M. (1994) The rapid dissociation of the T4 DNA polymerase holoenzyme when stopped by a DNA hairpin helix. A model for polymerase release following the termination of each Okazaki fragment. *J. Biol. Chem.*, **269**, 24221–24228.

Hill,T.M. (1992) Arrest of bacterial DNA replication. *Annu. Rev. Microbiol.*, **46**, 603–633.

Jaworski,A., Blaho,J.A., Larson,J.E., Shimizu,M. and Wells,R.D. (1989) Tetracycline promoter mutations decrease non-B DNA structural transitions, negative linking differences and deletions in recombinant plasmids in *Escherichia coli*. *J. Mol. Biol.*, **207**, 513–526.

Kiyama,R. and Oishi,M. (1994) Instability of plasmid DNA maintenance caused by transcription of poly(dT)-containing sequences in *Escherichia coli*. *Gene*, **150**, 57–61.

Kiyama,R. and Oishi,M. (1996) *In vitro* transcription of a poly(dA)·poly(dT)-containing sequence is inhibited by interaction between the template and its transcripts. *Nucleic Acids Res.*, **24**, 4577–4583.

Kohwi,Y., Malkhosyan,S.R. and Kohwi-Shigematsu,T. (1992) Intramolecular dG·dCdC triplex detected in *Escherichia coli* cells. *J. Mol. Biol.*, **223**, 817–822.

Kornberg,A. and Baker,T. (1992) *DNA Replication*. 2nd edn. W.H.Freeman and Co, New York.

Krasilnikov,A.S., Panyutin,I.G., Samadashwily,G.M., Cox,R., Lazurkin,Y.S. and Mirkin,S.M. (1997) Mechanisms of triplex-caused polymerization arrest. *Nucleic Acids Res.*, **25**, 1339–1346.

Lalioti,M.D., Scott,H.S., Buresi,C., Rossier,C., Bottani,A., Morris,M.A., Malafosse,A. and Antonarakis,S.E. (1997) Dodecamer repeat expansion in cystatin B gene in progressive myoclonus epilepsy. *Nature*, **386**, 847–851.

Liu,B. and Alberts,B.M. (1995) Head-on collision between a DNA replication apparatus and RNA. *Science*, **267**, 1131–1137.

Liu,B., Wong,M.L., Tinker,R.L., Geiduschek,E.P. and Alberts,B.M. (1993) The DNA replication fork can pass RNA polymerase without displacing the nascent transcript. *Nature*, **366**, 33–39.

Marians,K.J. (1992) Prokaryotic DNA replication. *Annu. Rev. Biochem.*, **61**, 673–719.

Martin-Parras,L., Hernandez,P., Martinez-Robles,M. and Schvartzman,J.B. (1991) Unidirectional replication as visualised by two-dimensional agarose gel electrophoresis. *J. Mol. Biol.*, **220**, 843–855.

Rao,B.S. (1994) Pausing of simian virus 40 DNA replication fork movement *in vivo* by (dG–dA)n·(dT–dC)n tracts. *Gene*, **140**, 233–237.

Rao,S., Manor,H. and Martin,R.G. (1988) Pausing in simian virus 40 DNA replication by a sequence containing $(dG–dA)_{27}·(dT–dC)_{27}$. *Nucleic Acids Res.*, **16**, 8077–8094.

Reaban,M.E. and Griffin,J.A. (1990) Induction of RNA-stabilized DNA conformers by transcription of an immunoglobulin switch region. *Nature*, **348**, 342–344.

Reaban,M.E., Lebowitz,J. and Griffin,J.A. (1994) Transciption induces the formation of a stable RNA·DNA hybrid in the immunoglobulin A switch region. *J. Biol. Chem.*, **269**, 21850–21857.

Rooney,S.M. and Moore,P.D. (1995) Antiparallel, intramolecular triplex DNA stimulates homologous recombination in human cells. *Proc. Natl Acad. Sci. USA*, **92**, 2141–2144.

Samadashwily,G.M., Raca,G. and Mirkin,S.M. (1997) Trinucleotide repeats affect DNA replication *in vivo*. *Nature Genet.*, **17**, 298–304.

Stanton,L.W. and Marcu,K.B. (1982) Nucleotide sequence and properties of the murine gamma 3 immunoglobulin heavy chain gene switch region: implication for successive C gamma gene switching. *Nucleic Acids Res.*, **10**, 5993–6006.

Stavnezer,J., Radcliffe,G. Lin,Y.-C. Nietupski,J. Berggren,L., Sitia,R. and Severinson,E. (1988) Ig heavy chain switching may be directed by prior induction of transcripts from constant region genes. *Proc. Natl Acad. Sci. USA*, **85**, 7704–7708.

Studier,F.W., Rosenberg,A.H., Dunn,J.J. and Dubendorff,J.W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.*, **185**, 60–89.

Uptain,S.M., Kane,C.M. and Chamberlin,M.J. (1997) Basic mechanisms of transcript elongation and its regulation. *Annu. Rev. Biochem.*, **66**, 117–172.

Usdin,K. and Woodford,K.J. (1995) CGG repeats associated with DNA instability and chromosome fragility from structures that block DNA synthesis *in vitro*. *Nucleic Acids Res.*, **23**, 4202–4209.

Wahls,W.P., Wallace,L. and Moore,P.D. (1990) Hypervariable minisatellite DNA is a hotspot for homologous recombination in human cells. *Cell*, **60**, 95–103.

Weitzmann,M.N., Woodford,K.J. and Usdin,K. (1996) The development and use of a DNA polymerase arrest assay for the evaluation of parameters affecting intrastrand tetraplex formation. *J. Biol. Chem.*, **271**, 20958–20964.

Woodford,K.J., Howell,R.M. and Usdin,K. (1994) A novel K(+)-dependent DNA synthesis arrest site in a commonly occurring sequence motif in eukaryotes. *J. Biol. Chem.*, **269**, 27029–27035.

Received May 18, 1998; revised June 29, 1998; accepted July 2, 1998