 Nucleic Acids Research, 1999, Vol. 27, No. 19 **3851–3858**

Fold-back structures at the distal end influence DNA slippage at the proximal end during mononucleotide repeat expansions

G. Karthikeyan, Kandala V. R. Chary[1] and Basuthkar J. Rao*

Department of Biological Sciences and [1]Department of Chemical Sciences, Tata Institute of Fundamental Research, Homi Bhabha Road, Bombay 400005, India

Received June 2, 1999; Revised and Accepted August 12, 1999

ABSTRACT

Polymerase slippage during DNA synthesis by the Klenow fragment of DNA polymerase across A, C, G and T repeats (30 bases) has been studied. Within minutes, duplexes that contain only repeats (30 bp) expand dramatically to several hundred base pairs long. Rate comparisons in a repeat duplex when one strand was expanded as against that when both strands were expanded suggest a model of migrating hairpin loops which in the latter case coalesce into a duplex. Moreover, slippage (at the proximal or 3'-end) is subject to positive and negative effects from the 5'-end (distal) of the same strand. Growing T and G strands generate T.A:T and G-G:C motif fold-back structures at the distal end that hamper slippage at the proximal end. On the other hand, growing tails at the distal end upon annealing with excess complementary template accentuates proximal slippage several-fold.

INTRODUCTION

Monotonous repeats of short nucleotide motifs occur interspersed in eukaryotic genomes (1–4). These so-called 'minisatellites' are highly polymorphic and have been used extensively as probes in RFLP analyses of DNA from individuals and in genome mapping (5,6). Instability of such repeats, the source of DNA polymorphism, may involve multiple mechanisms such as unequal crossing-over between repeats, complex gene conversions and strand slippage during replication (7–9). Evidence from different studies mechanistically connects strand slippage with DNA replication which mediates DNA expansion. The instability of repeat tracts is highly dependent on the orientation with respect to replication (10). Although lagging strand synthesis shows a higher fidelity than that of the leading strand for a normal mixed sequence DNA (11), repeat sequences are more prone to instability in the lagging strand than in the leading strand (12,13). Deletion of Rad27, which encodes a nuclease that processes 5'-overhangs in Okazaki fragments in yeast, or mutation in its homologue in humans (FEN1) increases repeat expansion (14,15). In *Escherichia coli*, mutations that induce the SOS response and/or a defect in 5'→3' exonuclease activity of Pol I (flap processing activity) leads to a marked increase in repeat expansions (16). It is pertinent to mention here that mutations in mismatch repair genes and not those affecting the proof-reading function of DNA polymerase increase destabilisation of simple repeats in yeast (17,18). This strongly suggests that strand slippage involving short DNA loops may frequently escape polymerase proof-reading and be corrected by mismatch repair. Thus it seems likely that the larger expansions of hundreds of nucleotides that characterise fragile-X syndrome could involve multiple slippages during replication (19).

In vitro studies involving simple repeats in oligonucleotide substrates have demonstrated that repetitious di- and trinucleotide motifs show slippage synthesis whose rate depends on the types of sequence and polymerase involved and not on the length of the substrate fragments (20,21). Moreover, triplet repeats in the different strands of a duplex do not show the same level of expansion (21). Such a strand bias seems to arise due to a higher propensity of one strand to fold into hairpins, which mediate DNA slippage (22). The importance of strand foldability into hairpins is further underscored by the observation that long hairpins have long lifetimes even in the presence of their complementary strand, thus facilitating large scale expansion (23). Moreover, slippage does not depend on the whole stretch of duplex DNA but depends on the 3'-end of the slippage strand, which generates a DNA loop at every round of polymerase slippage. Here, in a case study involving mononucleotide repeats, we demonstrate that 3'-end (proximal) slippage is influenced positively as well as negatively by the 5'-end (distal) of the slippage strand.

MATERIALS AND METHODS

T4 polynucleotide kinase, Klenow large fragment, exonuclease (3'→5')-deficient Klenow polymerase (exo$^-$) and dNTPs were purchased from Amersham Life Science.

DNA substrates

Oligonucleotides were synthesised in an Applied Biosystems DNA synthesiser at the Keck Biotechnology Resource Laboratory at Yale University. These were purified by electrophoresis in 10% polyacrylamide gels containing 6 M urea as described (24). The sample of oligonucleotide was subsequently desalted by passing through a Sep-pak C18 cartridge (25). The purity of

*To whom correspondence should be addressed. Tel: +1 91 22 2152971; Fax: +1 91 22 2152110; Email: bjrao@tifr.res.in

oligomers was judged by ^{32}P-labelling of a small portion by T4 polynucleotide kinase, followed by analysis on a 12% polyacrylamide sequencing gel.

End-labelling of oligonucleotides

A standard protocol was used to phosphorylate 5'-ends with [γ-^{32}P]ATP (10 μCi) in 5 μl reactions containing 100 μM (total nucleotide concentration) oligomer. Subsequently the sample was diluted to 50 μl and heated at 70°C for 10 min to heat inactivate T4 polynucleotide kinase.

Standard reaction conditions

Concentrations of oligonucleotides are expressed as total nucleotides. Unless stated otherwise in the figure legends, the following standard conditions were used. To generate DNA duplexes with a near 1:1 ratio of strands, thermal annealing was done by incubating the template (3 μM) and the labelled primer (2.5 μM) in 33 mM Tris–HCl (pH 7.5) and 10 mM magnesium acetate at 90°C for 3 min followed by slow cooling to room temperature. Extensions were carried out in 20 μl at 37°C following the addition of required dNTPs at 500 μM each and Klenow/exo$^-$ Klenow (5 U). Polymerase extensions were arrested by adding 1 μl each of EDTA (500 mM) and SDS (10%). A 5 μl aliquot of this was then added to an equal volume of formamide loading buffer and the entire sample was loaded on a 10% polyacrylamide sequencing gel. Following electrophoresis, the gel was dried and autoradiographed.

RESULTS

DNA expansion was studied with two repetitive duplexes, $(dA)_{30}$:$(T)_{30}$ and $(dC)_{30}$:$(dG)_{30}$. At a near equimolar ratio, the annealed products migrated as 30mer duplexes and no other products were seen (data not shown). The expansion of each of the strands in these repetitive duplexes was monitored separately by labelling only one strand at a time. The labelled strand is referred to as the primer and the unlabelled as the template. Time courses were recorded under four different experimental conditions. These were one dNTP (corresponding to the primer strand) versus two dNTPs (corresponding to primer and template strands) either at a near equimolar ratio of template and primer strands or with a molar excess of template strands.

Expansions in A:T duplexes

An illustrative example of dramatic expansion of the A and T strands in $(dA)_{30}$:$(T)_{30}$ is shown in Figure 1A. When both dTTPs and dATPs were provided, the 30mer strands were expanded simultaneously to products that were >400 bp within 5 min. Further incubation yielded products that were perhaps much larger and so could not be resolved in the gel (Fig. 1A). This DNA synthesis must involve the property of strand slippage that Klenow is known to facilitate in repeat regions (20,21). To explain such slippage products, two models of extreme versions can be considered. In one scenario, Klenow slides an entire strand past the other, involving complete breakage and reformation of AT base pairs. Such a process generates single-stranded template of a certain non-specific length at each 5'-end (left side path in Fig. 1B). In the second scenario, Klenow slides on the 3'-ends locally to generate short single-stranded templates and loops *de novo* (right side path in Fig. 1B). Such loops on both strands of a duplex should freely migrate towards the middle of the DNA and eventually coalesce into a duplex.

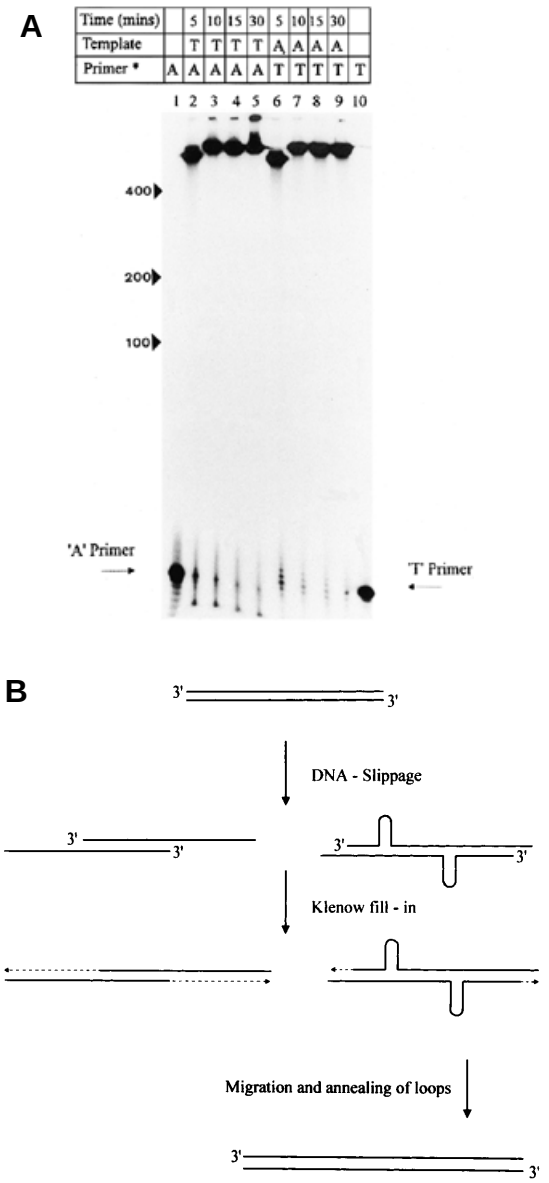

Figure 1. (**A**) Extensions in the 30mer A:T duplex using both dATP and TTP. The asterisk (in all legends) indicates the ^{32}P-labelled strand. The migration position of standard markers is shown at the side by either arrows or arrowheads in most figures. (**B**) Models depicting two extreme versions of strand slippage.

The models mentioned above make two simple testable predictions. In the first model, sliding of individual strands ought to be easier when only one strand is expanding since the duplex length to be undone remains constant. On the other hand, when both strands are expanding, the duplex that needs to be undone keeps on increasing (Fig. 1B). Therefore, the rate of expansion of a strand ought to be less when both strands are expanding than when only one strand is expanding. Conversely, in the hairpin loop model, the main component that drives expansion is coalescence of two migratory complementary loops into a duplex. In such a scenario, the rate of expansion of a strand will

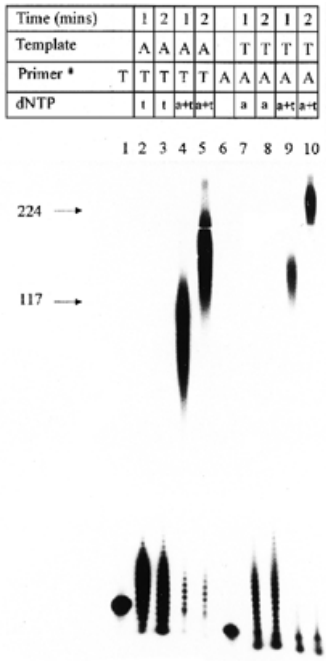

Figure 2. Extensions in the 30mer A:T duplex using either one or both dNTPs. Comparison of one strand versus both strand expansions.

be higher when both strands are expanding than when only one is expanding.

In the present study, the rate of strand expansion was significantly higher when both strands were expanding (Fig. 2). When both dNTPs were provided, the product of expansion was longer than a 100mer within 1 min, which in the next minute further expanded to more than a 200mer. On the other hand, when either TTP or dATP alone was provided, expansion of both the T and A primers was very little. This result clearly favours the loop model over that of the strand sliding model (Fig. 1B).

In order to check whether DNA expansion goes through any stable single-stranded intermediate, we probed the expansion reaction with mung bean nuclease, a single strand-specific enzyme. Mung bean nuclease is known to digest any single-stranded (ss)DNA intermediate in the reaction. Controls with ssDNA and a perfect duplex indicate the expected single strand specificity of mung bean nuclease action. Our results show that in an expansion reaction neither the size nor the yield of expansion products is reduced in the presence of mung bean nuclease (data not shown). This suggests that there is no stable ssDNA intermediate in this reaction and is consistent with the hypothesis that transient loops mediate Klenow expansion.

In the expansions where both the strands are growing, we found dissimilarity in the growth rates of individual strands. The T primer grows slightly more slowly and is more dispersed compared to the A primer (Fig. 2). We studied this more carefully in reactions where either the A or T strand alone was growing. We believe that the Klenow fragment that lacks 3'→5' exonuclease activity would yield a cleaner comparison of A strand versus T strand growth by eliminating the marginal effects of 3'→5' exonuclease action. Therefore, in all the experiments described below we used exo$^-$ Klenow. The

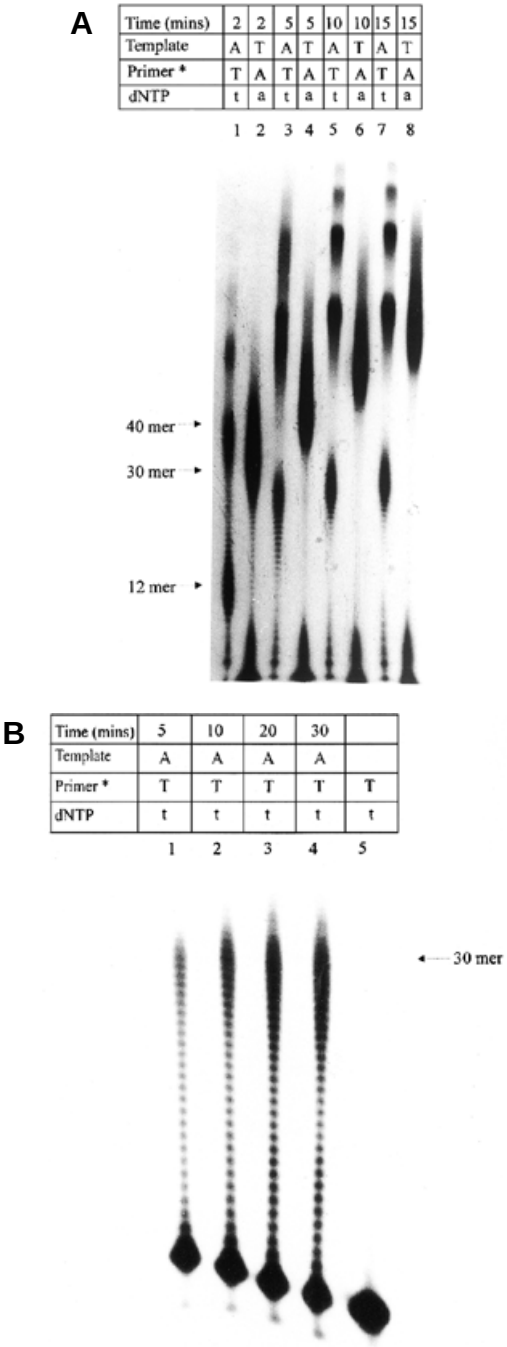

Figure 3. (**A**) Exonuclease-minus Klenow extensions of the 30mer A:T duplex generated by annealing 3 μM A strand with 2.5 μM T strand. Even numbered lanes are from extensions of the A strand whereas the odd numbered lanes are from that of the T strand. Positions marked on the side are +12mer, +30mer and +40mer extensions determined by counting the ladder in the autoradiogram, representing the 30+12mer, 30+30mer and 30+40mer positions, respectively. (**B**) A longer time course of exo$^-$ Klenow extension of the T strand in the 30mer A:T duplex generated by annealing 3 μM A strand with labelled 5 μM T strand [the reaction described in (A)]. The position marked at the side is the +30mer extension (i.e. 60mer) determined by counting the ladder in the autoradiogram.

products reveal a remarkable dissimilarity between the two reactions (Fig. 3A). The A primer shows monotonous growth while the T primer shows a non-monotonous and complex

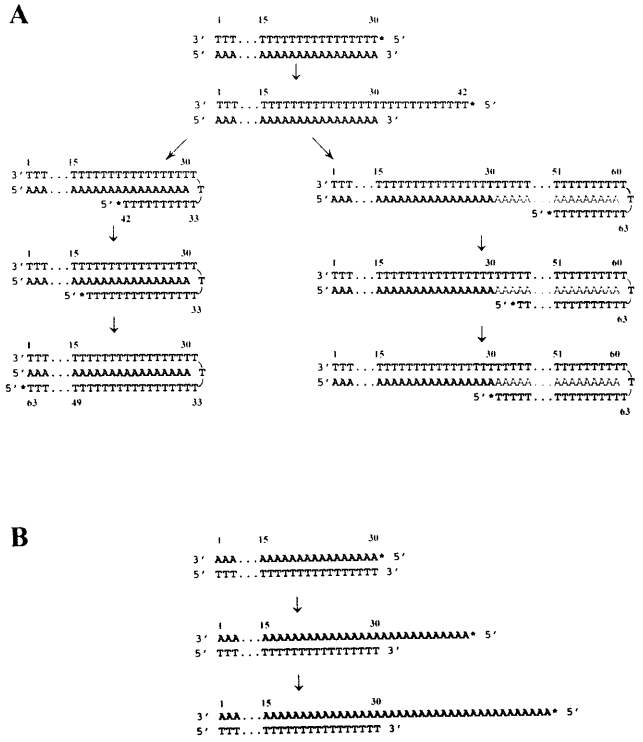

Figure 4. A model of T versus A strand growth. (**A**) T strand growth. The right diagram shows additional 30mer A strand (thin letters) annealing at the distal end prior to folding of the growing T strand. Nucleotides are counted 3' to 5' and indicated numerically above the labelled strand. (**B**) A strand growth.

Figure 5. Effect of deaza-adenosines in the A strand of the 30mer A:T duplex on T strand growth. Extensions were done under conditions identical to those described in Figure 3B. The only change was that in the 30mer A strand 15 adenosines at the 5'-end were replaced by 7-deaza-adenosines. A and B mark the positions of the +45 and +75 additions, respectively.

pattern of growth. The average growth rate of the A strand (Fig. 3A, lanes 2, 4 and 6) is ~2–3 nt/min. In T strand expansion, the product distribution shows four size classes that seem to slowly mature into four different stable sizes. For instance, at the 2 min time point the smallest of the four classes shows a distribution maximal at +12mer (i.e. addition of 12 nt to the 30mer primer) that matures to a limit size of +33mer at subsequent time points (Fig. 3A). This peculiar limit size of +33mer was further confirmed by a longer time course of Klenow extension in a higher percentage gel (Fig. 3B). In this experiment we used excess T strand compared to A template to drive the products exclusively to the +33mer limit size and thereby reduce the yield of higher sizes that form due to secondary annealing with leftover template (cf. Fig. 3A). The higher size classes of products seen at 2 min (in Fig. 3A) also followed a trend where they matured into discrete limit sizes during the slow phase of growth. Accurate size analysis of these longer products was, however, not possible due to poor resolution. However, based on the mobility of standard size markers, we believe that the higher size classes of products in the slow phase of the reaction are spaced by 30 nt size differences.

We believe that the dissimilarity in T strand versus A strand growth observed in this experiment reflects underlying differences in the nature of DNA intermediates involved in expansion reactions. While it is easy to understand the monotonous and uninterrupted growth of the A strand, the growth of the T strand, which reached a limit size, is not easy to rationalise (Fig. 3A and B). One way to explain how T strand growth might reach a limit size is described in the model shown in Figure 4. The conclusion is based on the fact that the limit size matches perfectly with a folded back T.A:T triplex (26) (Watson–Crick base pairs are represented as R:Y and Hoogsteen base pairs as Y.R) product length. We tested this prediction in the following experiment.

We replaced the first 15 adenosines of the A strand at the 5'-end with 7-deaza-adenosines to prevent Hoogsteen pairing in the major groove of the A:T duplex (27). This prevents formation of the classical (T.A:T) triplex structures (28). Thus, in the DNA expansion reaction with the (7-deaza-A)$_{15}$-(dA)$_{15}$:(dT)$_{30}$ duplex, two different situations are likely to occur according to the model envisaged in Figure 4. (i) The T strand should not reach the limit size of a folded triplex if the 15 dA residues at the 3'-end of the A strand are too short to sustain T.A:T fold-back triplexes. (ii) On the other hand, if 15 dA residues can sustain T.A:T triplexes, the limit sizes of T strand growth should now change to about +18, +48, +78, etc. from the +33, +63 and +93 seen earlier (see Fig. 3A and B). Indeed, the latter was the case when a (7-deaza-A)$_{15}$-(dA)$_{15}$ strand duplexed with a (dT)$_{30}$ strand was studied (Fig. 5). An extended time course in this reaction yielded products whose distribution showed a new limit size of +48 and +78 (positions marked A and B, respectively, in Fig. 5) (these sizes were estimated by carefully counting the bands in the autoradiogram). Perhaps by the second time point chosen (5 min) the distribution of products went past the +18 limit size or the first round of template annealing at the distal end quickly competed out +18 fold-back triplexes.

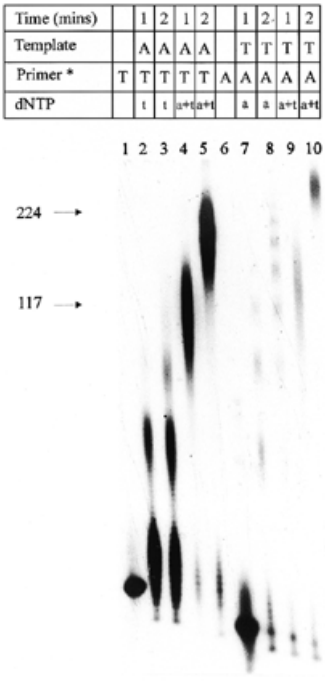

Figure 6. Effect of excess template on extension of the 30mer A:T duplex. Duplexes were formed by annealing 2.5 μM primer and 25 μM template strands.

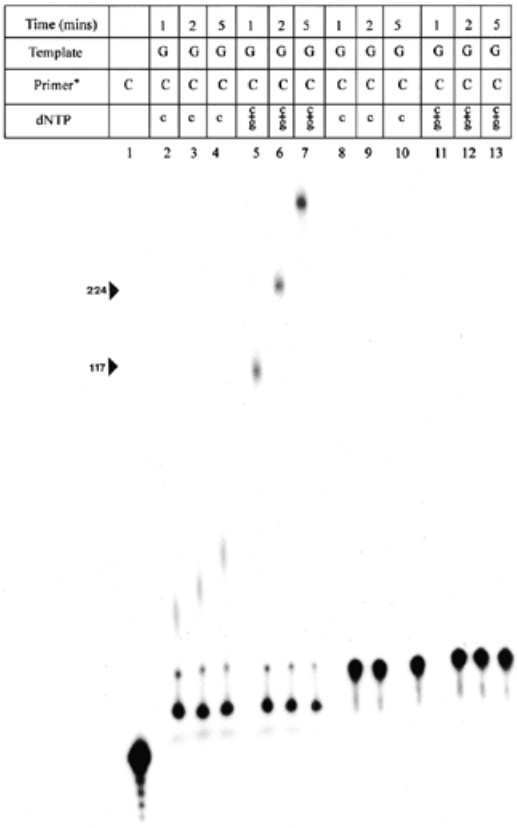

Figure 7. C strand growth in the 30mer C:G duplex. Annealing was done between a G template (3 μM in lanes 2–7; 30 μM for lanes 8–13) and a labelled C primer (2.5 μM) followed by extension by exo⁻ Klenow.

The latter possibility might indeed be true because as a result of a second round of template annealing at the distal end, the +48 limit size population decreased with a concomitant increase in the +78 population. This point is clearer from a comparison of the fall in distribution at A and the rise at B in lane 6 versus lane 5 (Fig. 5).

Coming back to the model described in Figure 4, the periodicity of higher products seen for T strand expansion (Fig. 3A, lanes 1, 3, 5 and 7) is explained as arising due to secondary annealing of the growing primer with the leftover template. This is because the A strand concentration (3 μM) was in slight molar excess to that of the T strand (2.5 μM). Moreover, the secondary annealing seems to compete with folding back of the T strand (Fig. 4). Even with slightly excess A strand (by 0.5 μM), the distribution of T strand expansions was significantly skewed in favour of higher period products (Fig. 3A). The relatively large yield of higher products such as +63mer, +93mer, etc. *vis-à-vis* that of the +33mer was incommensurate with the slight excess of the A strand. On the contrary, at the same concentrations the A primer grew monotonously as neither limit sizes nor periodic higher products were observed (Fig. 3A). Secondary annealing by the T template was not possible as the A primer was in excess, which led to continuous monotonous growth of the A strand. It is pertinent to mention here that the extent of annealing assessed by leftover A primer was variable. To overcome this variation and to study the source of periodicity in strand expansion, we repeated the experiment with a 10-fold molar excess of the template strand over the labelled primer. Even when the template strand was in excess, the annealed product migrated as a 30mer duplex and no other products were seen (data not shown).

When the template strand is in molar excess, expansion of the primer strand is periodic only when one strand (i.e. the primer strand) is growing (Fig. 6, compare lanes 2 and 3 with 4 and 5 and lanes 7 and 8 with 9 and 10, respectively). A strand expansion, which was monotonous due to the absence of secondary annealing in the earlier experiment (Fig. 3A), showed a clear periodic growth of higher products here due to secondary annealing with excess T template (Fig. 6). As described earlier (see Fig. 3A), the growth rate of the A strand was slightly faster than that of the T strand when both strands were growing (Fig. 6, compare lanes 4 and 5 with 9 and 10, respectively). In line with that, A strand growth was faster than that of the T strand even when only one strand was allowed to grow (Fig. 6, compare lanes 2 and 3 with 7 and 8, respectively). Comparison of the mobility with that of standard DNA markers suggested that the products in these lanes were periodically spaced at 30 nt intervals. This experiment revealed a mechanistic component of expansion where distal annealing of excess template with the growing primer in *trans* seems to accelerate primer growth at the proximal 3'-end. This results in a periodic distribution of products (Fig. 6).

Expansions in G:C duplexes

In order to determine the effects of distal ends on proximal slippage in C and G repeats, we also studied DNA expansion with $(dC)_{30}$:$(dG)_{30}$ duplexes. Almost all the experiments discussed above were repeated with a $(dC)_{30}$:$(dG)_{30}$ system.

Figure 8. Native gel analysis of annealed complexes formed between 30mer G strands (30, 3 and 0 μM, respectively, for lanes 1–3) and labelled C strands (2.5 μM) (as in Fig. 7). Following annealing, a 20 μl sample of each was electrophoresed in a 10% polyacrylamide non-denaturing gel, dried and autoradiographed. Arrowheads mark the positions of complexes containing multiple G strands per single C strand.

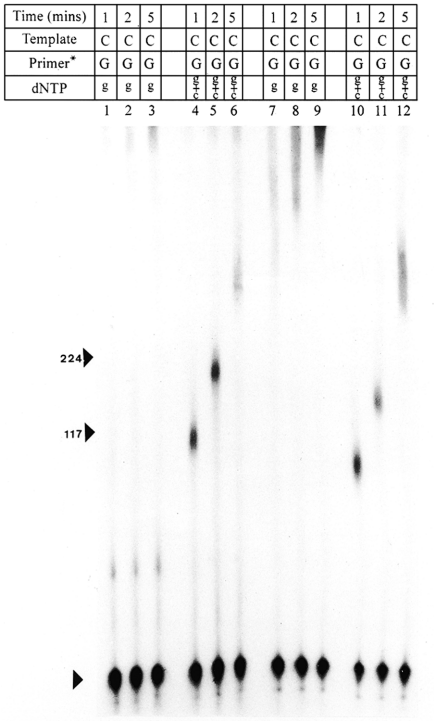

Figure 9. G strand growth in the 30mer C:G duplex. Annealing was done between a C template (3 μM in lanes 1–6; 30 μM for lanes 7–12) and a labelled G primer (2.5 μM) followed by extension by exo$^-$ Klenow. The arrowhead marks the position of the labelled 30mer G strand.

C strand growth was monotonous at an average rate of ~1–2 nt/min when only the C strand was allowed to expand (Fig. 7, lanes 2–4). Unlike T expansion, which generated T.A:T fold-back triplexes of defined limit size (Fig. 3A and B), C expansion showed no such pattern of equivalent fold-back triplexes containing a C$^+$.G:C motif even over a long time course (Fig. 7 and data not shown). The growth was entirely monotonous. This result was consistent with the reported findings that unlike the T.A:T triplex, which is stable at neutral pH, C$^+$.G:C triplexes are stable only at acid pH due to a requirement for protonation of C at the N3 position for its Hoogsteen pairing with G (26). The C strand growth rate increased several-fold (100 nt/min) when both the C and G strands were allowed to expand (lanes 5–7), which mirrored the trend observed for A and T (Fig. 2). However, the same reactions were completely dead when excess G template strands were present, a result that contrasted with that of A and T strand growth described above (compare Fig. 7, lanes 8–13 with Fig. 6, lanes 7–10). This may be due to the fact that G template strands are known to attain unusual secondary structures encompassing G strand dimers, quartets, etc. (29). We compared the electrophoretic mobilities of G strands with those of A, C and T strands in native polyacrylamide gels under our reaction conditions (data not shown). As expected, A, C and T strands migrated as single conformers while the G strand migrated as a smear in which three populations were more distinct. Such a distribution of multiple conformers of G strands is consistent with the notion that G strands aggregate (29). This aspect was corroborated further when we analysed C:G annealed complexes in a native gel (Fig. 8). These samples were derived from the experiment described earlier (Fig. 7). The mobility of a labelled C strand is compared following its annealing with the G strand at 1:1 as well as 1:10 (excess G template case in Fig. 7) molar ratios (Fig. 8). At a 1:1 ratio, as expected, one sees a unique species consistent with a normal C:G duplex population that is competent in C expansion (see Fig. 8, lane 2 and Fig. 7, lanes 2–7), whereas at a 1:10 ratio one sees multiple species of C:G complexes (see Fig. 8, lane 1), pointing to the high propensity for higher order structures in C:G duplexes (shown by arrowheads in Fig. 8). Such multistrand complexes might be resistant to C strand slippage as shown above (Fig. 7, lanes 8–13). Such a blockage of C strand slippage is not alleviated even with the inclusion of dGTP in the reaction, perhaps again due to the formation of higher order structure (for example G-G:C) (mismatch base pairs are represented as G-G and Watson–Crick base pairs are represented as G:C) in which the 3'-end of the Watson–Crick paired G strand is not available for slippage synthesis.

Finally, we studied G strand growth in G:C duplexes. When only the G strand was allowed to grow, the expansion seen was only ~5–10 nt, which was halted very early in the reaction (Fig. 9, lanes 1–3). It was as if the short G tail at the distal end inhibited its own further expansion. The same strand grew at a much greater rate (~100 nt/min) when both the G and C strands were allowed to grow (Fig. 9, lanes 4–6). In another set, inhibition of G strand growth was completely relieved when the C template was present in excess. In fact, the rate of G strand growth increased remarkably in the presence of excess C template (Fig. 9, lanes 7–9). When both the G and C strands were allowed to grow in the presence of excess C strand (Fig. 9, lanes 10–12), the

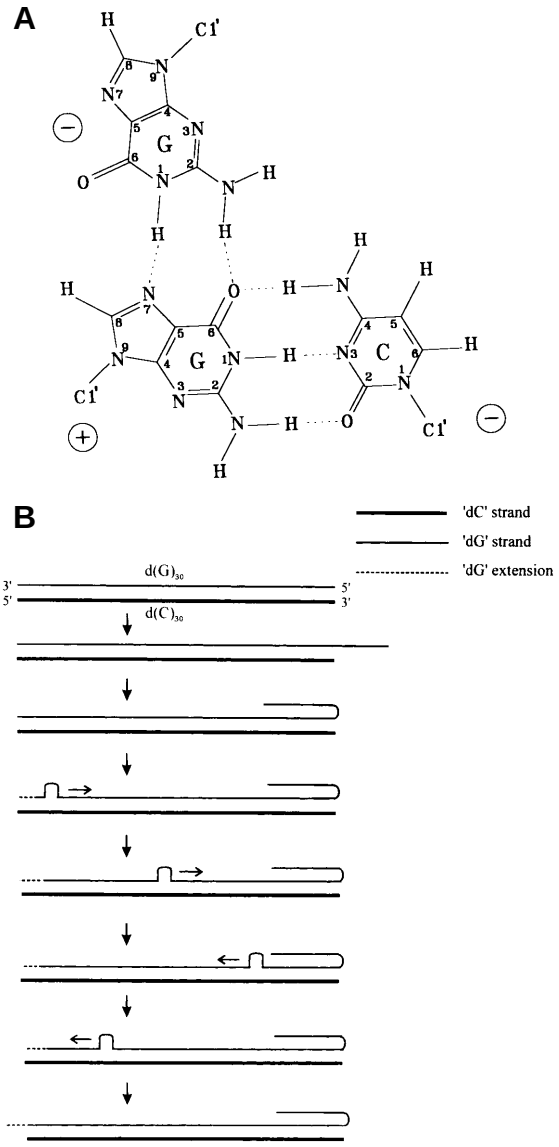

Figure 10. Model of G strand expansion. (**A**) The G-G:C triad motif that facilitates G-G:C triplex formation. (**B**) G strand growth generates a distal G tail that folds back on the G:C duplex leading to formation of a G-G:C triplex. Next, the G loop traverses along the G strand, only to be deflected back by the G-G:C stretch at the distal end. Residual DNA synthesis that might ensue during this step is shown by the dotted line. Return of the loop back to the 3'-end leads to strand protrusion that halts further expansion by Klenow, since it is deficient in 3'→5' exonuclease activity.

growth rate was again comparable to what was seen in the absence of excess C template (Fig. 9, lanes 4–6).

DISCUSSION

There have been numerous reports on polymerase slippage during DNA synthesis. Several years ago Fresco and Alberts had discussed the possibility that deletion and addition mutations arise via a DNA intermediate that involves 'helix with loops' (30). Kornberg and co-workers demonstrated a novel property of Pol I that expands dA:dT oligomers into long products which were believed to be generated by Pol I-mediated strand slippage (31). Several recent studies also reiterated that polymerases misalign simple repetitive sequences leading to DNA expansion or contraction (20,21,32,33). Direct evidence is now available to suggest that expansions arise by slippage of Okazaki fragments (34–36).

The robust version of polymerase slippage described in this paper provides an insight into the mechanistic aspects of how slippage at the (proximal) 3'-end is affected by the DNA secondary structure at the distal 5'-end. First of all, T strand expansion interestingly leads to a fold-back structure that is consistent with a T.A:T triplex. This conclusion is based on two observations: (i) T strand growth is punctuated at lengths that match with T.A:T fold-back structures (Figs 3A and B and 4); and (ii) the periodicity of T strand growth changes appropriately when some of the adenosines in the A strand are replaced by 7-deaza-adenosines, which prevent Hoogsteen pairing (Fig. 5; 27,28). In T strand growth, the fold-back structure (T.A:T triplex) slows down loop migration through it, thereby limiting the rate of synthesis. The rate in this phase plummets to ~0.2–0.3 nt/min, with only ~3–5 nt being added in 15 min (Fig. 3B). Eventually, as the T.A:T triplex matures to its full length, the 3'-end of the T strand becomes completely refractory to slippage and expansion comes to a halt (Fig. 3A and B).

The monotonous growth of A and C tails (2–3 nt/min) (Figs 3A and 7) is much slower than that of branch migration in naked DNA, which is estimated to be ~200 nt/min in the presence of magnesium cations (37). The DNA loop migration described in Figure 1B is analogous to branch migration described earlier (37,38). Hence, polymerase slippage itself, rather than loop migration, may be rate limiting in DNA expansion reactions. In contrast to A, C and T strands, G strand expansion failed to take off under our experimental conditions (Fig. 9, lanes 1–3). G strands have a strong tendency to dimerise intramolecularly as well as intermolecularly (29). Comparative mobility analysis of A, C, G and T strands in native polyacrylamide gels revealed species that are consistent with intra- as well as intermolecular dimerisation only for G (data not shown). Furthermore, labelled C strands anneal with G strands into complexes that presumably contain more than one G strand per C strand (Fig. 8). These results are consistent with secondary structural motifs proposed for G-rich strands (29,39). G expansion generated a small G tail that probably attained secondary structures of the type cited (Fig. 10A and references cited above). This blocked expansion by preventing the release of G loops at the distal end and further strand slippage at the 3'-end. This simple but interesting mechanism is depicted in Figure 10B, where a traversing G loop is deflected back when the protruding G tail at the distal end is left to itself (as evidenced by lack of G strand growth beyond a few nucleotides in Fig. 9, lanes 1–3). On the contrary, in the same reaction this inhibitory effect of protruding G tails is efficiently nullified when the C strand is simultaneously allowed to grow (Fig. 9, lanes 4–6). It is interesting to visualise how a fold-back G-G pairing might function as a block to an incoming G loop (Fig. 10B). The simplest known chemical motif that explains the distal fold-back region (in Fig. 10B) is the classical G-G:C triad (Fig. 10A; 40). In this setting, the middle G base of the triad is so well caged between the outer bases that it cannot be exchanged with another G base of the incoming G loop. This explains how the migratory G loop is deflected back on its path (Fig. 10B). Since the migration rate of the deflected G loop is

likely to be similar to the branch migration rate (~200 nt/min), the G loop returns back to the proximal end before the next round of polymerase slippage. This results in an effective abrogation of new rounds of slippage.

The differential behaviour of single strand growth in A, C, G and T repeats described above disappears when expansion is allowed on both strands of the duplex (Figs 2, 7 and 9). In all cases the growth rates increase to ~100 nt/min from low rates of single strand growth. Within the time course of measurement, rates seem to be independent of the initial length, which is consistent with the model that slippage at ends mediates expansion rather than sliding of strands (Fig. 1B). The rate enhancement seen when both ends of the duplex promoted slippage is a direct consequence of stabilisation of migrating loops. A simple rotation along the long axis of the duplex can twist the two loops of complementary sequence into a normal duplex, thereby stabilising the looped-out strands. This enhanced rate (~100 nt/min) is in the same range as the branch migration rate (~200 nt/min) (37), thus suggesting that stabilisation of migrating loops by mutual annealing accentuates the slippage rate to the limit of the branch migration rate. A similar enhancement of rate is observed when excess template strand anneals with the growing primer at the distal end. However, it was not possible to make rate estimates in this situation because the product distribution was more complex (Fig. 6). The products were distributed at a spacing equal to the length of the template strand. Such a periodicity in product distribution suggests that slippage at the 3'-end was accentuated by the stabilisation caused by template strand annealing at the distal end. How does distal annealing accelerate proximal slippage? The explanation seems to hinge on a communication through DNA that relays the stabilising effects of distal end annealing on the slippage rate at the proximal end. Future studies would assess the relative roles of chemical continuity in the DNA backbone versus that in base stacking for such communication to function.

The *in vitro* experiments described here suggest that slippage occurs at a high rate within simple repeats. The *in vivo* rate estimate for slippage-driven mutation in *E.coli* is about one mutation per 100 replication events (8). In addition to observations in humans, trinucleotide repeat instabilities have also been demonstrated in *E.coli* (41). Mono- and dinucleotide repeats, including those of Z-DNA-forming sequences, promote multiple slippages (17,33,42,43). DNA expansions in mono- and dinucleotide repeats are more likely to be deleterious to the cell by causing not only addition mutations but also frameshift mutations. On mechanistic grounds, mono- and dinucleotide repeat expansion by polymerase slippage ought to be as preponderant as that of trinucleotides. However, most DNA expansion-related diseases reported in humans so far seem to be due to triplet repeat expansions (44). Is this because *in vivo* expansions from mono- and dinucleotide repeats are subject to more rigorous controls and corrections?

ACKNOWLEDGEMENTS

We thank Prof. Girjesh Govil of the Department of Chemical Sciences (TIFR) for critically reading the manuscript. We thank the Department of Atomic Energy (TIFR) for financial support.

REFERENCES

1. Tautz,D. and Renz,M. (1984) *Nucleic Acids Res.*, **12**, 4127–4138.
2. Tautz,D., Trick,M. and Dover,G. (1986) *Nature*, **322**, 652–656.
3. Levinson,G. and Gutman,G.A. (1987) *Mol. Biol. Evol.*, **4**, 203–221.
4. Epplen,J.T. (1988) *J. Hered.*, **79**, 409–417.
5. Jeffrys,A.J., Wilson,V. and Thein,S.L. (1985) *Nature*, **314**, 67–73.
6. Jeffreys,A.J., Allen,M.J., Armour,J.A., Collick,A., Dubrova,Y., Fretwell,N., Guram,T., Jobling,M., May,C.A. and Neil,D.L. (1995) *Electrophoresis*, **16**, 1577–1585.
7. Streisinger,G., Okada,Y., Emrich,J. *et al*. (1966) *Cold Spring Harbor Symp. Quant. Biol.*, **31**, 77–84.
8. Levinson,G. and Gutman,G.A. (1987) *Nucleic Acids Res.*, **15**, 5323–5337.
9. Djian,P. (1998) *Cell*, **94**, 155–160.
10. Freudenreich,C.H., Stavenhagen,J.B. and Zakian,V.A. (1997) *Mol. Cell. Biol.*, **17**, 2090–2098.
11. Fijalkowska,I.J., Jonczyk,P., Tkaczyk,M.M., Bialoskorska,M. and Schaaper,R.M. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 10020–10025.
12. Trinh,T.Q. and Sinden,R.R. (1991) *Nature*, **352**, 544–547.
13. Veaute,X. and Fuchs,R.P.P. (1993) *Science*, **261**, 598–600.
14. Gordenin,D.A., Kunkel,T.A. and Resnick,M.A. (1997) *Nature Genet.*, **16**, 116–118.
15. Freudenrreich,C.H., Kantrow,S.M. and Zakian,V.A. (1998) *Science*, **279**, 853–856.
16. Morel,P., Reverdy,C., Michel,B., Ehrlich,S.D. and Cassuto,E. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 10003–10008.
17. Strand,M., Prolla,T.A., Liskay,R.M. and Petes,T.D. (1993) *Nature*, **365**, 274–276.
18. Strand,M.K., Early,M.C., Crouse,G.F. and Petes,T.D. (1995) *Proc. Natl Acad. Sci. USA*, **92**, 10418–10421.
19. Kang,S., Jaworski,A., Ohshima,K. and Wells,R.D. (1995) *Nature Genet.*, **10**, 213–218.
20. Schlotterer,C. and Tautz,D. (1992) *Nucleic Acids Res.*, **20**, 211–215.
21. Ji,J., Clegg,N.J., Peterson,K.R., Jackson,A.L., Laird,C.D. and Loeb,L.A. (1996) *Nucleic Acids Res.*, **24**, 2835–2840.
22. Chen,X., Mariappan,S.V.S., Catasti,P., Ratliff,R., Moyzis,R.K., Laayoun,A., Smith,S.S., Bradbury,E.M. and Gupta,G. (1995) *Proc. Natl Acad. Sci. USA*, **92**, 5119–5203.
23. Gacy,A.M. and McMurray,C.T. (1998) *Biochemistry*, **37**, 9426–9434.
24. Karthikeyan,G., Wagle,M.D. and Rao,B.J. (1998) *FEBS Lett.*, **425**, 45–51.
25. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
26. Bhaumik,S.R., Chary,K.V.R., Govil,G., Liu,K. and Miles,H.T. (1995) *Nucleic Acids Res.*, **23**, 4116–4121.
27. Sayers,E.W. and Waring,M.J. (1993) *Biochemistry*, **32**, 9094–9107.
28. Kim,M.G., Zhurkin,V.B., Jernigan,R.L. and Camerini-Otero,R.D. (1995) *J. Mol. Biol.*, **247**, 874–889.
29. Sundquist,W.I. and Klug,A. (1989) *Nature*, **342**, 825–829.
30. Fresco,J.R. and Alberts,B.M. (1960) *Proc. Natl Acad. Sci. USA*, **46**, 311–321.
31. Schachman,H.K., Adler,J., Radding,C.M., Lehman,I.R. and Kornberg,A. (1960) *J. Biol. Chem.*, **235**, 3242–3249.
32. Kunkel,T.A. (1990) *Biochemistry*, **29**, 8003–8011.
33. Ohshima,K. and Wells,R.D. (1997) *J. Biol. Chem.*, **272**, 16798–16806.
34. Eicher,E.E., Holdeu,J.J.A., Popvich,B.W., Reiss,A.L., Snow,K., Thibodeau,S.N., Richards,C.S., Ward,P.A. and Nelson,D.L. (1994) *Nature Genet.*, **8**, 88–94.
35. Richards,R.I. and Sutherland,G.R. (1994) *Nature Genet.*, **6**, 114–116.
36. Heale,S.M. and Petes,T.D. (1995) *Cell*, **83**, 539–545.
37. Panyutin,I.G. and Hsieh,P. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 2021–2025.
38. Hsieh,P. and Panyutin,I.G. (1995) *Nucleic Acids Mol. Biol.*, **9**, 42–65.
39. Smith,F.W. and Feigon,J. (1992) *Nature*, **356**, 164–168.
40. Radhakrishnan,I. and Patel,D.J. (1994) *Biochemistry*, **33**, 11405–11416.
41. Kang,S., Jaworski,A., Ohshima,K. and Wells,R.D. (1995) *Nature Genet.*, **10**, 11019–11023.
42. Ripley,L.S. (1990) *Annu. Rev. Genet.*, **24**, 159–213.
43. Tautz,D. and Schlotterer,C. (1994) *Curr. Opin. Genet. Dev.*, **4**, 832–837.
44. Ashley,C.T. and Warren,S.T. (1995) *Annu. Rev. Genet.*, **29**, 703–728.