DNA synthesis arrest sites at the right terminus of rat long interspersed repeated (LINE or LIRn) DNA family members

Ettore d'Ambrosio and Anthony V. Furano

Section on Genomic Structure and Function, Laboratory of Biochemical Pharmacology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Building 8, Room 203, Bethesda, MD 20892, USA

Received October 2, 1986; Revised and Accepted February 28, 1987

ABSTRACT

An ~150-bp GC-rich (~60%) region is at the right end of rat long interspersed repeated DNA (LINE or LIRn) family members (1). We report here that one of the DNA strands from this region contains several non-palindromic sites that strongly arrest DNA synthesis in vitro by the prokaryotic Klenow and T4 DNA polymerases, the eukaryotic a polymerase, and AMV reverse transcriptase. The strongest arrest sites are G-rich (~70%) homopurine stretches of 18 or more residues. Shorter homopurine stretches (12 residues or fewer) did not arrest DNA synthesis even if the stretch contains 11/12 G residues. Arrest of the prokaryotic polymerases was not affected by their respective single strand binding proteins or polymerase accessory proteins. The region of duplex DNA which contains DNA synthesis arrest sites reacts with bromoacetaldehyde when present in negatively supercoiled molecules. By contrast, homopurine stretches that do not arrest DNA synthesis do not react with bromoacetaldehyde. The presence of bromoacetaldehyde-reactive bases in a G-rich homopurine-containing duplex under torsional stress is thought to be caused by base stacking in the homopurine strand (2). Therefore, we suggest that base-stacked regions of the template arrest DNA synthesis.

INTRODUCTION

The long interspersed repeated DNA (LINE or LIRn) families are a dominant and consistent feature of mammalian genomes [(1, 3-15) and (16, 17) for recent reviews]. The members of such families are 6-7 kb long and interspersed throughout much of the genome. In the rat, where most of the 40,000 or so members are full length (approximately 6.7 kb) and very similar to each other, this family alone accounts for about 10% of the genome (1).

Although the function of LINE DNA is not known, these sequences were either unusually susceptible to, or capable of (or both) amplification and transposition. Furthermore, amplification and dispersal of LINE DNA must have recurred often during mammalian evolution to account for the presence of the related but clearly distinct LINE families in even relatively closely related genera such as Rattus (rats) and Mus (mice) (1, 18).

Recent evidence suggests that the rat LINE family has been amplified quite recently: First, in contrast to the heterogeneous families in mice and
primates, the rat LINE family is very homogeneous (1). Second, the presence or absence of LINE members is the cause of allelic variation of at least three single copy loci in R. norvegicus (19-21). This result suggests that LINE transposition may still be active in the rat population.

The mechanism of LINE amplification and transposition is unknown. Although it is possible that this occurred by recurring integration of DNA copies (retrotranscripts) of LINE RNA into the genome (8, 10, 16-18, 22, 23), rat LINE members lack several important hallmarks of retroposons and, in addition, contain a structural feature that suggests another mechanism (1). We previously showed that the ~150-bp GC-rich region that is the right end of rat LINE members contains sequences that strongly arrest DNA synthesis by the Klenow polymerase [(1) and our unpublished observations]. Only this region of the LINE contains these sequences, and since this region does not contain palindromes, we assumed (1) that arrest was caused by a type II DNA synthesis arrest site (24). In contrast to the palindromic type I sites (24), type II sites do not form any obvious physical barrier (e.g., hairpins or stems) to DNA synthesis in vitro. Therefore, the mechanism whereby these sites arrest DNA synthesis is unknown. Furthermore, type II sites, but not type I sites, are recognized in vivo (24-26) and therefore are biologically relevant structures. For example, they could be involved in the amplification and excision of genomic DNA sequences by a mechanism which involves the generation of a multi-stranded "onion skin" replicative intermediate (27-29).

For these reasons we analyzed in detail the arrest of DNA synthesis by the righthand LINE GC-rich region. We found that arrest is caused by several type II DNA arrest sites that strongly arrest DNA synthesis by both eukaryotic and prokaryotic polymerases. These sites consist of G-rich (~70%) homopurine stretches, and by making deleted and base-substituted versions of the right GC region we found that arrest is sensitive to the base composition of the site and is limited to homopurine stretches that are long enough to form stable base-stacked structures. The implication of these findings for the interaction of DNA polymerase with template and the possible effect that these homopurine stretches might have on the amplification and recombination of LINE sequences is discussed.

MATERIALS AND METHODS

Enzymes
The following enzymes and proteins were purchased from commercial sources: calf thymus DNA polymerase, Cooper Biomedical; AMV reverse transcriptase,
Boehringer-Mannheim; Klenow fragment of *E. coli* polymerase I, Bethesda Research Laboratories; and *E. coli* single strand binding protein, Pharmacia. The T4 DNA polymerase (gene 43 protein), the gene 32 single strand DNA binding protein, and the gene 45 and genes 44/62 DNA polymerase accessory proteins were provided by Dr. Nancy G. Nossal.

**DNA Templates**

The complementary strands of the right GC-rich region of a full length rat LINE member were cloned in the appropriate M13 vectors. Deletion derivatives were made with *ExoIII* basically as described by Henikoff (30). *In vitro* mutagenesis was performed by carrying out DNA synthesis *in vitro* on the C19 template (see Fig. 2 and its legend) primed with both the 17 mer M13 sequencing primer [New England Biolabs (No. 1211)] and p(dC)_{10}. The synthetic product was digested with restriction enzymes that flanked the region of interest, and the resulting 126-bp fragment was purified by electrophoresis on a polyacrylamide gel. The purified fragment was cloned in the appropriate M13 vector. Of the ten clones selected at random, two were mutant. One, CV1, was selected for further study and found to have two base substitutions (see Fig. 2). We also prepared two deleted derivatives (VD3b and VD2d) of this clone (see Fig. 2). The "synthetic" clone, Stl (see Fig. 5), was prepared by synthesis *in vitro* of the appropriate complementary oligonucleotides which were then annealed and cloned into the appropriate sites of the M13 vectors. The DNA sequence of all of the templates was verified by DNA sequence determination using the dideoxy procedure (31) as described earlier (1).

**Preparation of Radioactive Primer-Template**

New England BioLabs 17 mer sequencing primer (No. 1211) was used throughout and radiolabeled at the 5' end with $[^{32}P]P_4$ using T4 polynucleotide kinase in a 50 µl reaction of the following composition: 250 ng of primer, 0.1 M Tris-Cl (pH 8.0), 0.01 M MgCl$_2$, 0.01 M dithiothreitol, 0.2 mM spermidine, 0.4 mM NaPO$_4$ (pH 7.0), 200 µCi γ-$[^{32}P]ATP$ (3000 mCi/µmol), and 5 units of T4 polynucleotide kinase (Pharmacia). After 30 minutes at 37°C the reaction was adjusted to 0.02 M Na EDTA. The radioactive primer was purified from the reaction mixture using a "NENSORB" column (New England Nuclear) as described by the supplier except that the primer was eluted with 40% ethanol. Lower concentrations did not efficiently elute the primer. The specific activity of the primer was generally about 0.4 x 10$^6$ cpm (Cerenkov)/ng. The primer was annealed to template as follows: 200-400 ng of template was boiled in a siliconized capillary tube for 5 minutes with 8-10 ng of primer in a 15 µl reaction that contained 50 mM NaCl, 10 mM Tris-Cl (pH 7.5), 10 mM MgCl$_2$, and 1
mM dithiothreitol. After boiling, the reaction was slowly cooled to room temperature.

DNA Synthesis

In all cases the reactions were carried out in a volume of 25 μl with 50-100 ng of primer-template and 200 μM of each dNTP at 37°C for 30 minutes. By this time the reactions with all of the polymerases had gone to completion. The composition of the various reactions were as follows: (i) α DNA polymerase: The conditions are based on those of Weaver and DePamphilis (24): 50 mM Tris-Cl (pH 8.0), 10 mM MgCl₂, 12.5 to 50 mM KCl (depending on the volume of α polymerase added), 100 μg/ml bovine serum albumin (nuclease- and protease-free, Bethesda Research Laboratories), and 0.25 or 1 unit of the α polymerase. One unit is that amount of enzyme that catalyzes the incorporation of 1 nmol of dNMP into acid-precipitable product in 15 minutes at 37°C using DNAse I-activated DNA as template primer. Although the supplier stated the activity to be 0.11 U/μl, our assay showed it to contain 0.14 U/μl. We used the supplier’s value for calculating the units added. (ii) Klenow polymerase: 10 mM Tris-Cl (pH 7.5), 10 mM MgCl₂, 50 mM NaCl, 10 mM dithiothreitol, and 1 unit of the Klenow polymerase. One unit incorporates 10 nmol of dNMP into acid-precipitable product in 30 minutes at 37°C. In those cases where it was added, 360 ng of the E. coli single strand binding protein was used per reaction. We determined that this was in excess of any of the templates used by titrating each with the binding protein. In these reactions, template binding was readily monitored by agarose gel electrophoresis since saturating levels of binding protein dramatically alters the mobility of the template [e.g., see (32)]. (iii) T4 DNA polymerase: These reactions were carried out as previously described (33) and contained 25 mM Tris-Cl (pH 7.5), 6 mM Mg acetate, 60 mM potassium acetate, 5 mM dithiothreitol, bovine serum albumin, 100 μg/ml, and 2 μg/ml of purified T4 DNA polymerase, and when indicated, 60 μg/ml of the gene 32 protein, 40 μg/ml of the genes 44/62 proteins, and 6 μg/ml of the gene 45 protein. (iv) AMV reverse transcriptase: 100 mM Tris-Cl (pH 8.0), 10 mM MgCl₂, 50 mM dithiothreitol, and 25 units of the enzyme. One unit incorporates 1 nmol of TMP into acid-precipitable product in 10 minutes at 37°C with poly(A)°p(T)₁₅.

The reaction products were analyzed on DNA sequencing gels. Prior to gel electrophoresis, each sample was treated as follows: After adjustment to 50 mM Na EDTA and 0.2% sodium dodecyl sarcosinate, they were incubated with 100 μg/ml proteinase K. They were then adjusted to 0.3 M sodium acetate and 10 μg/ml of tRNA, and the nucleic acids were precipitated with 2 volumes of
The pellets were dried in vacuo and dissolved in 6 μl of 94% formamide containing 12.5 mM Na EDTA and 0.025% xylene cyanol. They were denatured for 3 minutes at 90°C before a portion of which was applied to the gel.

In all of the experiments reported here, the DNA polymerases were in considerable excess, i.e., the amount of DNA synthesis was linearly dependent on the amount of template primer added. However, whereas increasing the Klenow polymerase concentration 10 times (i.e., to 10 U per 25 μl reaction) did not affect the results, the higher concentration of the α polymerase resulted in the synthesis of longer products, usually at the expense of the shorter ones synthesized by the lower concentration (see Results).

**Bromoacetaldehyde Reactions**

The preparation of and reactions with bromoacetaldehyde were carried out essentially as described by Kohwi-Shigematsu and Kohwi (2). Approximately 2 μg of supercoiled DNAs were subjected to one of 3 reactions: (i) The DNA was digested with Scal, reacted with 0.11 M bromoacetaldehyde, and then digested with nuclease S1. (ii) The DNA was reacted with bromoacetaldehyde, digested with Scal, and then digested with nuclease S1. (iii) The DNA was digested with Scal, treated with bromoacetaldehyde, and then with EcoRI. The bromoacetaldehyde reactions were carried out in 50 mM Na acetate, pH 5.0, 2 mM MgCl₂, for 1 hour at 37°C, and the S1 reactions were carried out for 1 hour at 37°C with 50 U/ml and the reaction buffer described by Vogt (34) containing 50 mM NaCl.

**RESULTS**

**Type II DNA Arrest Sites in the Right GC-Rich Region of a Rat LINE Member**

By using the complementary DNA strands of SV40 or parovirus DNA as templates for DNA synthesis in vitro, Weaver and DePamphilis (24) distinguished two classes of DNA synthesis arrest sites: type I, those that form intrastrand duplexes, and type II, those that do not. As would be expected, a pause or arrest is observed with only one of the complementary strands of DNA that contains type II sites, but with both strands of a type I site.

Fig. 1 shows the DNA synthesized off of each complementary strand of the right end of a previously sequenced LINE member (hereafter referred to as L3). The diagram in the center of Fig. 1 depicts the righthand 285 bp of L3, including all of its a 150-bp GC-rich region. As the autoradiogram on the right shows, this sequence contains two closely placed sites (I and II) that not only strongly arrest the α polymerase but also the Klenow polymerase. By
contrast, little or no pausing was observed with either polymerase when the complementary strand was used as template (left autoradiogram, Fig. 1). Although template 5AB contains only the 100 bp or so of GC-rich region that is 3' to the AvaiI site (see Fig. 1), identical results were obtained using the forerunner of template 5AB, which includes the GC-rich region 5' to the AvaiI site and an additional 1.3 kb or so of 5' L3 DNA as well (results not shown).

The sequences corresponding to the two strong arrest sites [designated I (closest to the primer) and II, see below for our determination of this] are shown in Fig. 2 (clone 5AB). A stretch of G's is 5' to both sites. Since DNA synthesis proceeds from right to left (3' to 5') on the template strand, these G's are encountered by the polymerase after it reaches the A indicated by arrow I or II. Therefore, in the remainder of the paper, terms such as "followed by" and "after" or "preceded by" and "before" are used in reference to the direction of DNA synthesis.

Other sites followed by G stretches only weakly arrest DNA synthesis (site III, see below) or do not arrest it all (e.g., the boxed nucleotides in Fig. 2). Therefore, G richness per se is not sufficient to arrest DNA synthesis. To characterize more precisely this type of DNA arrest site, we used deleted or base-substituted derivatives of the various G-rich sequences in the right GC region of L3, and these are shown diagramatically above the sequence of clone 5AB (Fig. 2). In all but one of the clones (15f), LINE DNA begins at the left at an AvaiI site. Sequences 5' of this are derived from one or another of the vectors used, and the bases in lower case, gag, are the first three bases of the SacI site of M13mpl9. In all cases except C19, DNA synthesis starts 9 bases to the right of the 3' g from the 17 mer M13 sequencing primer (see legend to Fig. 2 for more details, and Materials and Methods).

FIG. 1. Type II DNA synthesis arrest sites in the right GC-rich region of L3. The central bar depicts the right end of L3; the stippled part corresponds to the GC-rich region and the crosshatched part corresponds to the A-rich right end of L3, which is unusually long (1). The right line of the bar corresponds to template 5AB, and the products synthesized from it are shown in the right-hand autoradiogram, which is oriented in the direction of DNA synthesis (indicated by the arrow). The shorter products are at the bottom of the photograph, and the number on the arrow gives the length (in nucleotides) of the shortest product shown. DNA synthesized off of the complementary strand (template A62,26e) is shown in the lefthand autoradiogram (short products at the top of the photograph). T, 1, and 2 indicate the following: T, T-track, using the Klenow polymerase and the dideoxy T-containing sequencing mixtures, and 1 and 2, reactions with 0.25 and 1.0 unit of the a DNA polymerase, respectively. The strong arrest sites, I and II, and the AvaiI site are indicated on the corresponding regions of the autoradiograms.
DNA synthesis using these templates is shown in Fig. 3. The results show the following: First, as was found with the dideoxy sequencing reactions [Fig. 1, (1)], the Klenow polymerase is also strongly arrested on template 5AB in a reaction containing all four dNTPs. Some of the chains that extend beyond site II are arrested at a third site, site III (see Fig. 2 and below). Although the precise termination of nascent DNA chains at sites I and II cannot be determined from the results shown in Fig. 3, other experiments in which sequencing reactions were carried out at 50°C indicated that most of the chains terminated at the A's, indicated by arrows I and II (results not shown, but see also below).

Second, as the result with template C19 shows, the second strong arrest site [referred to as II' because it is missing one of the G's present in site II (see Fig. 2)] still strongly arrests DNA synthesis even though site I is absent. However, as might be expected with only one strong arrest site present, more of the chains are extended to site III and larger products (top of
Templates 15AB and 15f (Figs. 3A and 3B, respectively) contain site II' only 9 bp from the primer, and the results with the Klenow polymerase (T-track, 15AB, and 15f; complete reaction, 15AB) show that most of the chains arrested at site II' terminate at the A before the first stretch of G's. Likewise, all of the chains that are arrested at site III terminate at the A before the first stretch of G's at this site (see arrows II' and III, Fig. 2). Furthermore, the pattern of termination by the α polymerase at sites II' and III were similar to the Klenow except that termination began 1–3 nucleotides sooner (cf., tracks 1 or 2 to T, 15f, Fig. 3B).

Third, conversion of the penta G stretch that occurs after the first A in site II' to a deca G stretch (see clone CV1, Figs. 2 and 3A) significantly altered the pattern of chain termination at site II'. Although the A remained a prominent termination site, much more termination occurred at each of the four succeeding nucleotides in the altered site II' (CV1) than in the unaltered site II' (15AB).

Fourth, when the template, VD3b, which contains only the weak arrest site III, was used, two other pause sites increased in prominence (a and b, cf. lanes 15AB and VD3b, Fig. 3A; Fig. 2). However, all three are quite weak as judged by the considerable amounts of high molecular weight chains synthesized from this template. With all of the templates used, little or no arrest was seen at any of the G-rich sequences that are 5' to the just mentioned weak sites, III, a and b. This includes the deca G stretch present in templates CV1 and VD3b (see Fig. 2). The lack of arrest by this stretch was verified when it alone, preceded by gag, was used as a template (clone VD2d). Predominantly high molecular weight products were synthesized from this template (lane VD2d, Fig. 3A).

We also tested the α polymerase on templates CV1, VD3b, and VD2b (results not shown). Although it paused at sites that did not significantly affect the Klenow polymerase, the strength of the Klenow polymerase arrest sites in these templates was generally the same for the two polymerases as was shown with template 15f (Fig. 3B). Furthermore, the altered pattern of stopping at the CV1 site was also seen with the α polymerase.

Effect of DNA Binding Proteins and DNA Polymerase Accessory Proteins on the Arrest of DNA Synthesis

Previous studies showed that pausing by various DNA polymerases in vitro could be reduced or eliminated by the E. coli single strand DNA binding protein (SSB) (35, 36), or the T4 DNA binding protein encoded by gene 32 and the T4 DNA polymerase accessory proteins encoded by gene 45 and genes 44/62 (33, 37,
FIG. 3. DNA synthesis in vitro with various templates. The templates used are indicated at the top of the autoradiograph, and T indicates T-track using Klenow polymerase. In part A, Klenow polymerase was used, and in part B the α DNA polymerase: 1 and 2, 0.25 U and 1 U per reaction, respectively. The inserts near the 5AB and C19 lanes are short exposures to show the arrest with
these templates in better detail. I, II, II', and III, the lower case a and b
and the * designate the chains which terminate at the nucleotides indicated by
the corresponding symbols in Fig. 2. The curly bracket in lane 5AB corres-
ponds to that region of the 5AB template that is underlined and italicized in
Fig. 2.

38). As Fig. 4 shows, the saturating amounts of the binding and accessory
proteins used in these experiments (see Materials and Methods) are completely
without effect on the arrest of either the Klenow or the T4 DNA polymerase
(gene 43 product) with templates 5AB and 15AB. These results are quite
striking since under the same conditions the T4 proteins completely copy the
ϕX174 genome (5.4 kb) in less than 5 minutes at 30°C [(33) and N. G. Nossal,
personal communication].

While this work was in progress, Charette et al. (39) also showed that
the T4 DNA binding and accessory proteins did not affect the pausing of the T4
DNA polymerase at both type I and type II DNA arrest sites. However, in this
study the concentration of both the T4 DNA polymerase and its accessory
proteins (i.e., the proteins encoded by genes 44/62 and 45) were considerably
lower than those used here. In the experiment shown in Fig. 4, the molar
ratio of T4 polymerase to template primer was 20 : 1, whereas in the experi-
ments reported by Charette et al. (39) it ranged from 0.005 to 0.05 : 1.
These low ratios were used to prevent reinitiation of DNA synthesis at DNA
arrest sites. Therefore, our results indicate that even with an excess of the
polymerase and the binding and accessory proteins, little reinitiation occurs
at LINE arrest sites.

The results of Fig. 4 also show a clear difference between the pattern of
stopping by the Klenow and T4 DNA polymerases. Whereas the Klenow polymerase
pauses about equally at sites I and II, the T4 DNA polymerase pauses much more
strongly at site I than site II (inserts, Fig. 4). Also, the T4 polymerase
pauses more strongly than the Klenow polymerase at the two weak Klenow arrest
sites that precede sites I and II (see the double-underlined nucleotides, Fig.
2). Furthermore, the T4 polymerase does not pause strongly at the A's,
indicated by arrows II' or III, but at several nucleotides beyond. These
results indicate that although the polymerases are sensitive to the same
regions of sequence, their interactions with these regions are distinct (see
Discussion).

Finally, Fig. 4 shows that the AMV reverse transcriptase is also strongly
arrested at sites I and II (lane RT). Therefore, even though the normal
template for reverse transcriptase is RNA, the enzyme is sensitive to the same
regions of template that affect the DNA polymerases. However, the reverse
transcriptase is also affected by other regions of the template, notably the (A), (A), and (A) stretches that occur at the end of the A-rich right end of L3 (see Fig. 4, lane RT, and Fig. 2).

DNA Sequence Requirements to Arrest DNA Polymerase

As mentioned above, G richness per se is not sufficient to cause DNA arrest. However, inspection of Fig. 2 shows that a G-rich, purine stretch of at least 18 residues is common to the two sites (I, II) that strongly arrest all three DNA polymerases. Furthermore, arrest is weakened but not eliminated by the disruption of a purine stretch by pyrimidines (e.g., site III). These conclusions are summarized in Fig. 5, which shows the sequences of the salient purine-rich stretches in Fig. 2 as well as another that we tested and one which we have taken from the results of Weaver and DePamphilis (24). Fig. 5 also shows the G content and a rough estimate of the strength of each arrest site.

Our findings that DNA synthesis is arrested at or near the beginning of an 18-nucleotide G-rich homopurine stretch but not one that is 12 nucleotides long suggested that these stretches have different conformations. Others (2, 40, 41) have proposed that G-rich homopurine stretches that are about 15 bases or longer, but not ones that are about 10 bases long, form stable base-stacked structures. These are thought to deform duplex DNA to produce a region of what has been call heteronomous DNA (42). The evidence for this is that when present in negatively supercoiled molecules, the longer homopurine stretches, but not the shorter ones, are sensitive to the single strand-specific SI nuclease (41, 43-45). Although some unpairing of the duplex in the base-stacked region may account for the SI sensitivity (2, 40, 46), other perturbations of duplex DNA are also sensitive to SI nuclease (41, 47). Heteronomous DNA apparently is not induced by the mildly acidic condition of the SI nuclease assay, since it has been detected with a variety of agents from pH 4.6 to pH 9 (2, 40, 41, 44, 46).

FIG. 4. Effect of DNA binding proteins and DNA polymerase accessory proteins on the arrest of the Klenow or the T4 DNA polymerase. The templates used are indicated at the top of the autoradiograms. T4 indicates the reactions with the T4 proteins, all of which contained the T4 DNA polymerase (gene 43 protein): 1, polymerase alone; 2, genes 44/62 and gene 45 proteins added; 3, gene 32 protein added; 4, genes 44/62, gene 45, and gene 32 proteins added. K indicates the reactions with the Klenow polymerase: 1, polymerase alone, and 2, polymerase plus the E. coli single strand binding protein. Lane RT shows the reaction products using reverse transcriptase (25 U/reaction). Inserts are lower exposures of the contiguous lanes. Jagged lines indicate where photographs were cropped.
FIG. 5. G-Rich homopurine stretches and their ability to arrest DNA synthesis. N and the number in the parentheses give the number of nucleotides and pyrimidines, respectively. The StI template, which was synthesized in vitro (see Materials and Methods), did not arrest the Klenow polymerase (results not shown). The Rsab9C2 sequence was taken from Weaver and DePamphilis (24). We designate it a moderately strong arrest site (+) because like site III (template VD3b) the Klenow polymerase can progress through this site in the presence of the dideoxy sequencing mixtures (24). The strength of an arrest site can also be estimated by visual comparison of the relative amounts of high molecular weight and arrested DNA chains (see Fig. 3). DNA synthesis on templates containing any of these sites proceeds from right to left.

If base stacking in the template strand is related to the arrest of DNA synthesis, then one might expect that those homopurine stretches which arrest DNA synthesis should deform duplex DNA when present in negatively supercoiled molecules, whereas those that do not arrest DNA synthesis should not. To carry out these experiments, we cloned the LINE DNA in templates 15AB, VD3b, and VD2d between the EcoRI and AvaI sites of plasmid pUC19 (48). Attempts to clone the LINE sequences from clone 5AB into pUC19 or several other plasmid vectors produced various complicated rearrangements that included partial duplication of the vectors and deletions of part of the LINE DNA. These results are related in part to both the sequence of the insert and the size of the vector. Although the 5AB insert can be propagated in M13, the yield of
Nucleic Acids Research

phage containing the 5AB insert is considerably less than normal. This is also true of the 15AB-containing plC19 clone (results not shown).

The structure of the pUC19 clones of 15AB, VD3b, and VD2d are shown in Fig. 6A (cf. with Figs. 2 and 5). To determine whether the homopurine stretches of these DNAs are structurally deformed when negatively supercoiled, we examined them for unpaired bases. To do this, we treated the cloned DNAs and pUC19 with bromoacetaldehyde (2, 46) as either supercoiled or linear molecules (see Materials and Methods). Bromoacetaldehyde reacts specifically with the N₁ and N₆ of unpaired adenines and the N₃ and N₄ of unpaired cytidines (46). Although it is not known whether the reactive bases are present in a stable, unpaired region of duplex (i.e., a denaturation bubble) or in rapid equilibrium between the paired and unpaired state, the reactive region remains susceptible to the single-stranded nuclease SI, even after the bromoacetaldehyde-treated supercoiled molecule is linearized.

Fig. 6B shows that clones 15AB and VD3 (which contain the DNA arrest sites), but not clones VD2 or pUC19 (which do not contain arrest sites), are susceptible to SI after bromoacetaldehyde treatment of supercoiled molecules (lane 2) but not of linear molecules (lane 1). Comparison of the two major SI fragments with the Scal/EcoRI fragments (lane 3) shows that the bromoacetaldehyde-reactive region is slightly to the right of the EcoRI site, as would be expected from the location of the arrest sites (Fig. 6A). The faint single band in lane 2 of the VD2 and pUC19 reactions, which is also present in lane 2 of the 15AB and VD3 reactions just below the upper SI fragment, is not due to SI digestion since it is also present in these samples before SI treatment (see Fig. 6C).

We also carried out SI nuclease reactions on both the supercoiled and linear molecules that had not been treated with bromoacetaldehyde. This experiment showed that the bromoacetaldehyde-reactive sites in clones 15AB and VD3 were also sensitive to SI nuclease treatment alone and that this sensitivity was dependent on supercoiling since the linear molecules were not cleaved by the SI nuclease (results not shown). This experiment also revealed a second significantly less reactive SI nuclease-sensitive site in the three supercoiled recombinant clones that was not seen in the SI digest of pUC19. This minor SI nuclease-sensitive site does not react with bromoacetaldehyde and is located outside of the LINE DNA inserts in vector DNA about 150 bp to the left (5') of the EcoRI site shown in Fig. 6.

Finally, subsequent to the submission of this paper we found a way to stably propagate the 5AB sequence (see Fig. 2). The arrest site-containing
region also reacts with bromoacetaldehyde under the conditions used here as well as at pH 7.0. Furthermore, preliminary fine mapping of the bromoacetaldehyde reactive sites in this clone showed that the major site of

FIG. 6. Bromoacetaldehyde-reactive sites in the right GC-rich region of L3. (A) Diagrammatic representation of the four plasmid DNAs used for the experiments shown in panel B. The orientation of the G-rich-containing strand is opposite to that shown in Fig. 2, and clones VD3 and VD2 contain the DNA from clones VD3b and VD2d, respectively. (B) Lane 1, the indicated plasmid treated with Scal, followed by bromoacetaldehyde and nuclease SI digestion; lane 2, plasmid treated with bromoacetaldehyde and then Scal and SI; lane 3, plasmid doubly digested with Scal and EcoRI. In each case lanes 1 and 2 contained 0.6 μg of DNA, and lane 3, 0.3 μg of DNA. (C) a, b, c, and d refer to plasmids 15AB, VD3, VD2, and PUC19, respectively, and show the DNA from reaction 2 before (-) and after (+) SI treatment co-electrophoresed with the Scal/EcoRI digest (reaction 3). The bands indicated by a white dot in panel C correspond to those indicated by a white dot in panel B and are not products of SI digestion (see text).
reactivity is within the homopurine stretch shown at the top line of Fig. 5 (Usdin and Furano, unpublished observations).

DISCUSSION

Characteristics of G-Rich Homopurine DNA Synthesis Arrest Sites

Our results show that G-rich purine stretches that are at least 18 nucleotides long strongly arrest DNA synthesis by the Klenow, phage T4, and calf thymus α DNA polymerases, and AMV reverse transcriptase. Interruption of the homopurine stretches, with as many as two pyrimidines, weakens but does not abolish the arrest. On the other hand, G-rich stretches that are only 12 nucleotides long do not arrest DNA synthesis (Fig. 5). As Fig. 6 shows, we found an excellent correlation between the ability of a homopurine stretch to arrest DNA synthesis and whether it reacts with bromoacetaldehyde when present in negatively supercoiled molecules. Since unpairing of the reactive bases is thought to be caused by strong intrastrand base stacking in the homopurine strand (2, 40, 42), we conclude by analogy with the results of others (2, 40, 41) that the LINE G-rich homopurine stretches that arrest DNA synthesis can form base-stacked structures. This implies that base-stacked regions of the template arrest DNA synthesis.

Although the detailed structure of stacked bases in a single strand is not known, stacked and unstacked regions of the template will undoubtedly have different conformations. Therefore, if the uncopied template that is just ahead of the nascent chain makes intimate contact with the polymerase, then stacked and unstacked regions could interact differently with the polymerase. The recent structural studies of the Klenow polymerase indicate that as many as five nucleotides of the uncopied template just beyond the nascent chain could be in close contact with the polymerase within its DNA binding cleft (49). Therefore, improper contact between polymerase and yet to be copied template could well arrest DNA synthesis by either impeding the progress of or causing dissociation of the polymerase.

Two results indicate that the Klenow polymerase is sensitive to the structure of the template that is just beyond the nascent chain. First, the major termination site occurs at or near the beginning of the homopurine stretch (see arrows I, II, II', and III, Fig. 2). Second, substitution of the second A of template 15AB (site II') by a G to produce template CV1 significantly altered the pattern of arrest at this site (Fig. 3A). The T4 DNA polymerase also was arrested near the beginning of the arrest sites. However, the T4 polymerase progressed two nucleotides further than the Klenow polymerase into sites II' and III before significant chain termination occurred.
These results indicate that the interactions of the two polymerases with the yet to be copied template differ, presumably reflecting a difference between their DNA binding sites. The fact that the arrest of both polymerases is unaffected by their respective DNA binding accessory proteins (Fig. 4) indicates that the region of the template responsible for the arrest is either not accessible to the accessory proteins (perhaps because it is in the DNA binding cleft of the polymerase) or that its conformation is not affected by them.

Two of the results discussed in the preceding paragraph also essentially rule out intermolecular aggregation of template molecules as a cause of DNA arrest. Polyriboguanylic acid aggregates to form a four-stranded structure (50). If the DNA templates aggregated via their G-rich homopurine stretches, then DNA synthesis would be arrested at the region of aggregation. However, this mechanism would not account for the distinct pattern of stopping found between the 15AB and singly-substituted CV1 templates or between the Klenow and T4 DNA polymerases. Furthermore, gel electrophoresis showed no evidence of intermolecular aggregation; each of the templates migrated as monomers (results not shown).

A number of other nonpalindromic sequences, in addition to the G-rich homopurine stretches, arrest DNA synthesis (24, 26, 36, 51). These other arrest sites do not share any common sequence elements and do not arrest DNA synthesis in vitro nearly as strongly as the G-rich homopurine sites described here. Nonetheless, pausing at many of the non-G sites in vitro correlates quite well with their ability to arrest DNA synthesis in vivo (24, 26). To our knowledge, pausing specifically at G-rich homopurine stretches in vivo has not yet been examined, and experiments to do so are now underway. It should be noted, however, that the 200-bp segment of the drug-resistant factor R6K, which provides a specific DNA replication termination signal recognized both in vivo and in vitro (25, 52), contains an 18-bp homopurine stretch that is 61% G and is interrupted just once by a pair of C's. By comparison to the sequences shown in Fig. 5, we would expect the R6K homopurine stretch to be a moderately strong DNA synthesis arrest site in vitro.

Possible Relevance of DNA Arrest Sites to the Biological Properties of LINE DNA

Arrest or pausing of DNA synthesis relative to its initiation could produce a multistranded ("onion skin") replication intermediate (27-29). Such an intermediate has been implicated in the amplification and excision of inte-
grated viral genomes (27, 28) and has also been suggested as a cause of the rapid amplification of genomic sequences that can be induced under certain experimental conditions (29). As we have discussed earlier (1), LINE DNA arrest sites could lead to the amplification and transposition of LINE DNA by this mechanism. However, the G-rich homopurine arrest sites could also be involved in the recombinational activity of LINE DNA by a mechanism that does not involve the arrest of DNA synthesis.

Recent work has shown that an unpaired homo G stretch in negatively supercoiled DNA induces the unpairing of a second, noncontiguous stretch of bases 3' to the homo G region (2). If this is also true of the LINE G-rich homopurine stretches, the second unpaired region would involve the A-rich right end of the LINE (see Fig. 2), and preliminary experiments indicate that this is so (Usdin and Furano, unpublished observations). Furthermore, recent work from our laboratory (21) suggests that the A-rich right end of the LINE may initiate recombination with the target site into which a LINE element transposes. If the unpaired A-rich sequences are susceptible to strand scission, the resulting free DNA end(s) most likely would be highly recombinogenic, as has been routinely found for free DNA ends (53-55). Recombination of LINE DNA with a target site could then induce transposition by a replicative mechanism as has been proposed for a variety of transposable elements (56). On the basis of their comparisons of cloned truncated versions of rat, mouse, and primate LINE family members, Soares et al. (14) also concluded that the right end of LINE elements is prone to recombination.

The right ends of both mouse and primate LINE family members also contain G-rich homopurine stretches (13, 18, 22) similar to those in the rat LINE. Although these sequence elements may be necessary for or predispose LINE sequences to amplification, or recombination, or both, it is most unlikely that these sequences alone are sufficient. For example, 70% or more of the rodent and primate LINE sequences consist of highly conserved ORFs that have evolved like bona fide protein-encoding sequences (1, 57, 58). This suggests that the ORFs were or still are under selective pressure, and one reason could be that their protein products are essential for LINE transposition and amplification.

ACKNOWLEDGMENTS

We thank Dr. Nancy G. Nossal (National Institutes of Health) for carrying out the assays containing the T4 proteins.
REFERENCES