The Development and Use of a DNA Polymerase Arrest Assay for the Evaluation of Parameters Affecting Intrastand Tetraplex Formation*

(Received for publication, March 5, 1996, and in revised form, May 16, 1996)

M. Neale Weitzmann, Kerry J. Woodford, and Karen Usdin‡

From the Section on Genomic Structure and Function, Laboratory of Biochemical Pharmacology, NIDDK, National Institutes of Health, Bethesda, Maryland 20892-0830

We show here that a K⁺-dependent block to DNA synthesis is a sensitive and specific indicator of intrastand tetraplex formation that can be used, both to identify sequences with tetraplex-forming potential and to examine parameters that affect tetraplex formation. We show that tetraplex formation is determined by a complex combination of factors including the size and base composition of its constituent loops and stems. In the process of carrying out this study we have found that the number of sequences with the ability to form tetraplexes is larger than previously thought, and that such sequences are ubiquitous in eukaryote genomes.

A number of recent studies have shown that certain guanine-rich DNA sequences form interstrand and intrastand DNA tetraplexes. Interstrand tetraplexes are formed between G-rich tracts located on 2 or 4 different DNA strands. Intrastand tetraplexes are formed by a single strand of DNA that contains at least 4 G tracts (Fig. 1). Gs from each of the G tracts interact to form a planar array of 4 guanines (a tetrad) held together by Hoogsteen bonds (1). Recently it has been shown that (G + C)-rich tracts can also form tetrads of the form G:C (2, 3). A series of successive tetrads forms a hollow stem. In the case of intrastand tetraplexes, this stem is bounded by three loops formed by the bases between the G tracts.

Interstrand tetraplexes have been implicated in the pairing of homologous chromosomes during meiosis (4) and switch recombination in immunoglobulin genes (4), and they may play important roles in the functioning of telomeres (4, 5). Recently, proteins that bind specifically to tetraplexes (4–6), or that catalyze tetraplex formation have been described (6, 7). These results, together with the demonstration of intrastand tetraplex formation in vivo (8), suggest that tetraplexes may have a biological function. While a large number of studies have focused on interstrand tetraplexes formed by telomere or telomere-derived sequences (9–14), relatively little is known about the sequence constraints governing the formation and stability of intrastand tetraplexes.

We have recently shown that a G-rich sequence from the chicken β-globin promoter provides a K⁺-dependent block to DNA synthesis, whose properties are consistent with the formation of an intrastand tetrahedral structure (15). Analysis of the K⁺-dependent structure showed it to be an unusual "cinched" tetraplex containing a number of non-guanine bases in the tetraplex stem (2 Ts and a C) and stabilized by interactions between one of the tetraplex loops and bases in the flanking sequence (16). (CGG), tracts also form K⁺-dependent blocks to DNA synthesis consistent with tetraplex formation (17). This tetraplex may contain a mixture of G tetrads and G-C-G-C tetrads (2, 3), or G-tetrads and C-C base pairs (17, 18).

Using a series of known tetraplex-forming model sequences, we show that the ability to block DNA synthesis in a K⁺-dependent manner is not a peculiar property of these unusual tetraplexes, but rather a diagnostic of intrastand tetraplexes in general. This property provides a rapid and simple assay for these structures and for studying the parameters that affect tetraplex formation. This assay can be carried out under physiological conditions and can detect tetraplexes at concentrations that are orders of magnitude less than those required by classical analytical techniques such as CD spectroscopy, x-ray crystallography, and NMR, even when tetraplexes constitute only a small fraction of DNA molecules in the population. We have used this assay to identify a number of tetraplex-forming sequences in eukaryote genomes.

MATERIALS AND METHODS

Clone Construction—Oligonucleotides were synthesized on an ABI 381A oligonucleotide synthesizer and cloned into either the plasmid pMS1891 or M13mp18 as described previously (15, 17). Plasmids were replicated in Escherichia coli MBM7070, isolated by alkaline lysis, and purified by CsCl gradient centrifugation according to standard procedures. Single-stranded phage DNA was purified as described elsewhere (17). The plasmid (C), was a gift of R. Howell.

Methylation of Plasmids—DNA was methylated with the SsI methylase (New England Biolabs), using the buffers and reaction conditions recommended by the supplier. After phenol extraction, the methylated DNA was ethanol-precipitated and resuspended in 10 mM Tris-HCl, pH 8.0, 1 mM Na₂EDTA (TE) at a final concentration of approximately 1 μg/μl. Methylation was confirmed by testing for protection from HphI digestion.

DNA Synthesis Arrest Assays—Arrest of DNA synthesis was monitored in a modified linear amplification sequencing reaction. Primers for these reactions were end-labeled with [γ-32P]ATP (DuPont NEN, 3000–6000 Ci/mmol) using T4 polynucleotide kinase (Epicentre Technologies, Inc.) and one of the following dideoxynucleotides at the concentration indicated in parenthesis: ddATP, ddCTP, ddTTP, and 7-deaza-dGTP, 50 mM Tris-HCl (pH 9.3), 2.5 mM MgCl₂, 5 units of Taq DNA polymerase (Life Technologies, Inc.) and one of the following dideoxynucleotides at the concentration indicated in parenthesis: ddATP (0.3 mM), ddGTP (0.017 mM), ddCTP (0.2 mM), ddTTP (0.6 mM), in a total volume of 1 μl. Cations were added to this mix at the indicated concentrations. Reactions were subjected to 30 cycles of heating and cooling (30 s at 95 °C, 30 s at 55 °C, and 30 s at 72 °C). One half volume (3 μl) of stop buffer (95% (w/v) formamide, 10 mM EDTA (pH 9.5), 10 mM NaOH, 0.1% xylene cyanid, 0.1% bromphenol blue) was then added and reactions heated at 90 °C.

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence should be addressed: Bldg. 8, Rm. 202, National Institutes of Health, 8 Center Dr., MSC 0830, Bethesda, MD 20892-0830. Tel.: 301-496-2189; Fax: 301-402-0240; E-mail: ku@helix.nih.gov.
**RESULTS**

K⁺-dependent Blocks to DNA Synthesis Are a General Property of Tetraplex-forming Sequences—When the known tetraplex-forming sequence (T₅G₄)₄ (5) is used as a template in in vitro DNA synthesis reactions, a set of premature polynucleotide chain termination products is observed at the beginning of the tetraplex-forming sequence (Fig. 2A). The major chain termination product comigrates with the base immediately preceding the 3' most G in the (T₅G₄)₄ tract (indicated by an arrow, Fig. 2A). The premature chain termination products or arrests are seen only when KCl is included in the reaction. No arrest is observed in the absence of monovalent cation (Fig. 2A), or when LiCl, NaCl, RbCl, CsCl, or NH₄Cl are substituted for KCl (data not shown). Potassium glutamate also produces the same premature chain termination products, confirming that it is the cation rather than anion that is responsible for the formation of the block to DNA synthesis (data not shown). K⁺-dependent arrest of DNA synthesis occurs with all DNA polymerases tested including Tα polymerase, T4 DNA polymerase, Sequenase™ (a modified T7 DNA polymerase), avian myeloblastosis virus reverse transcriptase, and SequiTHERM™ polymerase (data not shown), indicating that arrest of DNA synthesis is not polymerase-specific. Since DNA synthesis in the region preceding the tetraplex-forming sequence is similar in the presence or absence of K⁺, it is apparent that the arrest of DNA synthesis is not due to an adverse effect of K⁺ on the polymerase. The effect of K⁺ is not ameliorated by the use of polymerase accessory proteins such as E. coli single-stranded binding protein or gene 32 protein, or the inclusion of glycerol, formamide, or dimethyl sulfoxide (data not shown), reagents which have been shown to reduce premature chain termination (19).

The K⁺-dependent block to DNA synthesis is strand-specific since no block to DNA synthesis is observed when the pyrimidine-rich strand is used as a template (Fig. 2B). Furthermore, these blocks to DNA synthesis are seen even when single-stranded phage DNA containing these sequences is used as the template (data not shown). This rules out incomplete melting of the duplex or the formation of an interstrand structure involving the complementary strand, e.g., a triplex. Interstrand triplexes have been previously shown to block DNA synthesis, but blockage is independent of K⁺, requires the presence of both DNA strands, and is seen only when the pyrimidine-rich strand is used as a template (20). The same blocks to DNA synthesis are seen when the (T₅G₄)₄ sequence is cloned in different origins or in different vectors, indicating that the (T₅G₄)₄ sequence itself is sufficient to block DNA synthesis (data not shown).

Formation of these blocks to DNA synthesis is rapid since they are seen even when templates are incubated for less than 10 min in the presence of K⁺, and the amount of DNA synthesis arrest does not increase with longer incubations before polymerase addition (data not shown). These blocks are also very

---

**A Simple Assay for Tetraplex Formation**

---

**Fig. 1.** Diagrammatic representation of the basic structure of intrastrand tetraplexes. Panel A, generic tetraplex-forming sequence. GSs from four G-rich tracts hydrogen bond to form tetrads which stack to form the tetraplex stem. The spacer regions, labeled 1, 2, and 3, form the three loops, 1', 2', and 3'. Two basic intrastrand tetraplex conformations are known that vary in the orientation of loop 2. Panel B, diagonally looped tetraplex in which loop 2 joins diagonal strands within the stem and the guanine strand nearest neighbors are not equivalent, one being parallel and one antiparallel. Panel C, adjacent looped tetraplex in which loop 2 joins adjacent strands in the stem and each strand is antiparallel to both its nearest neighbors.

---

**Fig. 2.** K⁺-dependent arrest of DNA synthesis. Modified linear amplification sequencing reactions were carried out as described under "Materials and Methods." In the presence or absence of 50 mM KCl, and using (T₅G₄)₄ (panel A), the complementary strand (C₄A₄)₄ (panel B), and the hairpin forming sequence (G₁₃C₁₃) (panel C) as templates. The lane markers T, C, G, and A represent bases on the template strand. Arrows indicate the start of the tetraplex (panels A and B) or hairpin (panel C) forming region.
stable since they are still observed when templates are incubated at 85°C before addition of the polymerase. These data suggest that the extent of premature chain termination observed reflects the amount of underlying structure formed under any given set of circumstances, rather than their stability or rate of formation.

The sequence G13C13, which is 2 bases longer than (T2G4)4 and is able to form a fully Watson-Crick base paired hairpin, produces very little DNA synthesis arrest and the small amount of arrest seen is \( K^+ \)-independent (Fig. 2). These data suggest that a hairpin is not responsible for the \( K^+ \)-dependent arrest of DNA synthesis seen on the (T2G4)4 templates.

Both the number of different premature chain termination products at the beginning of the (T2G\(n\))4 sequence (\( S_1 \) in Fig. 3A), and the amount of each product increases as the number of Gs in each repeat tract increases (Fig. 3A). This is consistent with both an increase in the number of potential structures and an increase in the proportion of molecules that form such structures as the number of Gs increases. Some \( K^+ \)-dependent polymerase blockage is seen at the beginning of the second G-rich repeat (\( S_2 \)), and the third G-rich repeat (\( S_3 \)).

[\( K^+ \)], or a small amount of tetraplex formed in the absence of \( K^+ \). Some monovalent cation-independent DNA synthesis arrest is seen in the middle of the (T2G\(n\))4 sequence where \( n \approx 6 \) (\( S_3 \) in Fig. 3, A and B) and is characteristic of triplexes formed between the template and the nascent strand (21).

\( K^+ \)-dependent arrest of DNA synthesis is seen over the template concentration range 0.02–40 mM, with a greater proportion of plasmid template molecules blocking DNA synthesis as the template concentration decreases (data not shown). This suggests that arrest of DNA synthesis involves the formation of intrastrand rather than interstrand structures, and that their formation is favored when reannealing of the template strand with its complement is reduced.

These data are consistent with the underlying physical basis of the \( K^+ \)-dependent arrest of DNA synthesis on (T2G\(n\))4 templates being intrastrand tetraplex formation. To check that these are simple tetraplexes rather than "cinched tetraplexes" (16) or some other unusual tetraplex, we probed an oligonu...
A Simple Assay for Tetraplex Formation

The N-7 position of guanines in tetraplexes is involved in hydrogen bonding and is thus protected from modification by DMS, a reagent which reacts specifically with guanines at the N-7 position. As Fig. 4A shows, the sequence (T₂G₇)₄ has distinctly different patterns of DMS reactivity in the presence or absence of K⁺. In the absence of K⁺, strong uniform DMS modification is seen throughout the sequence; however, in its presence the extent of DMS modification is considerably reduced. The bases at the ends of the G tracts are modified about 50% less than they are in the absence of K⁺ as judged by densitometric analysis, while bases in the middle of the G tracts are modified only 10–25% as much. The pattern and extent of DMS reactivity is most consistent with the formation of a number of simple tetraplexes in which the inner bases in each G tract are more often involved in tetrads than the outer Gs which are more likely to reside at least part of the time within the loops.

Parameters Affecting Tetraplex Formation—These experiments demonstrate that K⁺-dependent arrest of DNA synthe-

FIG. 5. Comparison of extent of tetraplex formation by different templates. A modified linear amplification sequencing reaction of the plasmid templates (T₂G₇)₄, (T₃G₇)₄, (T₄G₇)₄, and (T₅G₇)₄, was carried out in the presence of 50 mM KCl as described under “Materials and Methods.” The lane markers T, C, G, and A indicate the bases on the template strand, and the arrow marks the beginning of the first G-rich repeat.

FIG. 6. Effect of the length and base composition of spacer region 2 on tetraplex stability. Modified linear amplification sequencing reactions using various plasmid templates with the sequence (T₂G₇T₂G₇XₙG₄T₂G₇), was carried out in the presence of 50 mM KCl as described under “Materials and Methods.” Panel A shows the effect of the base composition of spacer region 2 as indicated. Panel B shows the effect of the length of spacer region 2 as indicated. The lane markers T, C, G, and A represent bases on the template strand.

FIG. 7. Effect of the length of spacer regions 1, 2, and 3 on tetraplex formation. Modified linear amplification sequencing reactions on various plasmid templates were carried out as described under “Materials and Methods.” The number of Ts in each spacer region is indicated by Xₙ, Yₙ, and Zₙ for spacer regions 1, 2 and 3, respectively. The lane markers T, C, G, and A represent bases on the template strand.

FIG. 8. Effect of stem base composition on polymerase arrest. A modified linear amplification sequencing reaction of various plasmid templates in the presence of 50 mM KCl was carried out as described under “Materials and Methods.” Panel A shows the effect of a single base substitution in the stem forming region as judged by DNA synthesis arrest. The templates used were (A), which contains a single A substitution (template sequence: (T₂G₅T₂GGAGGT₂G₅T₂G₅)), and (C)₁, which contains a single C substitution (template sequence: (T₂G₅T₂GCGGT₂G₅T₂G₅)). (T₂G₅)₄ is shown for comparison. Panel B shows the effect of methylation on tetraplex formation by the (C)₁ template. Methylation status is indicated by the − and + signs. The lane markers T, C, G, and A indicate bases on the template strand and the arrows indicate the beginning of the first G-rich repeat. The number of Ts in each spacer region is indicated by Xₙ, Yₙ, and Zₙ for spacer regions 1, 2 and 3, respectively. The lane markers T, C, G, and A represent bases on the template strand.

Parameters Affecting Tetraplex Formation—These experiments demonstrate that K⁺-dependent arrest of DNA synthe-
A Simple Assay for Tetraplex Formation

TABLE I
List of potential tetraplex-forming sequences

| GenBank number | GenBank Description | Strand | Potential tetraplex forming sequence |
|----------------|---------------------|--------|--------------------------------------|
| X15316         | Heterodontus franscisci lambda immunoglobulin | light chain gene | GGGGGCGGGGAGGGGGGAGGGGAGGGG |
| M38191         | Human 5-s盱 hoxygenase gene | + | GGGGCGGGGAGGGGGGAGGGGAGGGG |
| M34901         | Human argininosuccinate synthetase gene | + | GGGGCGGGGAGGGGGGAGGGGAGGGG |
| X68969         | Human gene for beta 1 integrin | + | GGGGCGGGGAGGGGGGAGGGGAGGGG |
| X02158         | Human gene for erythropoietin | - | GGGGCGGGGAGGGGGGAGGGGAGGGG |
| M24228         | Human liver/bone/kidney-type alkaline phosphatase (ALP) gene | + | GGGGCGGGGAGGGGGGAGGGGAGGGG |
| M94250         | Human retinoid acid inducible factor (MK) gene | + | GGGGCGGGGAGGGGGGAGGGGAGGGG |
| X77900         | M. fascularis apolipoprotein gene C-III | + | GGGGCGGGGAGGGGGGAGGGGAGGGG |
| X62826         | M. musculus pva gene for parvalbumin | + | GGGGCGGGGAGGGGGGAGGGGAGGGG |
| Z2794          | M. musculus troponin I gene | + | GGGGCGGGGAGGGGGGAGGGGAGGGG |
| X54077         | Mouse intracisternal A-particle IAP DNA containing HOX-2.4 gene | - | GGGGCGGGGAGGGGGGAGGGGAGGGG |
| X56850         | Mouse retinoid acid receptor beta gene | + | GGGGCGGGGAGGGGGGAGGGGAGGGG |
| M34651         | Pseudorabies virus | + | GGGGCGGGGAGGGGGGAGGGGAGGGG |
| M10613         | Rat alpha-1 procollagen(I) gene | + | GGGGCGGGGAGGGGGGAGGGGAGGGG |

Introns

| GenBank number | GenBank Description | Strand | Potential tetraplex forming sequence |
|----------------|---------------------|--------|--------------------------------------|
| M29384         | Herpes simplex virus type 2 immediate-early (IE4) protein | + | GGGGCGGGGAGGGGGGAGGGGAGGGG |
| M95740         | Human alpha-L-iduronidase gene | + | GGGGCGGGGAGGGGGGAGGGGAGGGG |
| M26313         | Human CD8, genomic organization and nucleotide sequence | + | GGGGCGGGGAGGGGGGAGGGGAGGGG |
| X06079         | Human gene for HLA-DR alpha | + | GGGGCGGGGAGGGGGGAGGGGAGGGG |
| X13546         | Human HMG-17 gene for non-histone chromosomal protein HMG-17 | + | GGGGCGGGGAGGGGGGAGGGGAGGGG |
| M14949         | Human R-ras gene | + | GGGGCGGGGAGGGGGGAGGGGAGGGG |
| X63240         | Mouse mAmh gene for anti-Mullerian hormone | + | GGGGCGGGGAGGGGGGAGGGGAGGGG |
| D90360         | Mouse gene for cytokeratin endo A | + | GGGGCGGGGAGGGGGGAGGGGAGGGG |
| X13337         | Human platelet-derived growth factor A-chain gene | + | GGGGCGGGGAGGGGGGAGGGGAGGGG |
| X55068         | Herpes Simplex virus 1 infected-cell protein | + | GGGGCGGGGAGGGGGGAGGGGAGGGG |
| X15499         | Human 5-s盱 hoxygenase gene | + | GGGGCGGGGAGGGGGGAGGGGAGGGG |
| X76420         | M. musculus mAmh gene for anti-Mullerian hormone | + | GGGGCGGGGAGGGGGGAGGGGAGGGG |
| D90360         | Mouse gene for cytokeratin endo A | + | GGGGCGGGGAGGGGGGAGGGGAGGGG |
| X13337         | Human platelet-derived growth factor A-chain gene | + | GGGGCGGGGAGGGGGGAGGGGAGGGG |

Exons/coding regions

| GenBank number | GenBank Description | Strand | Potential tetraplex forming sequence |
|----------------|---------------------|--------|--------------------------------------|
| M33699         | Herpes simplex virus type 1 infected-cell protein | + | GGGGCGGGGAGGGGGGAGGGGAGGGG |
| M77831         | Human acoresin gene | + | GGGGCGGGGAGGGGGGAGGGGAGGGG |
| X74628         | M. musculus mAmh gene for MUC18 glycoprotein | + | GGGGCGGGGAGGGGGGAGGGGAGGGG |

5'-Untranslated regions

| GenBank number | GenBank Description | Strand | Potential tetraplex forming sequence |
|----------------|---------------------|--------|--------------------------------------|
| X74028         | B. taurus TSPY gene | + | GGGGCGGGGAGGGGGGAGGGGAGGGG |
| M76558         | Human neuronal DHP-sensitive, voltage-dependent, calcium channel | + | GGGGCGGGGAGGGGGGAGGGGAGGGG |
| M20488         | Human platelet-derived growth factor A-chain gene | + | GGGGCGGGGAGGGGGGAGGGGAGGGG |

3'-Untranslated regions

| GenBank number | GenBank Description | Strand | Potential tetraplex forming sequence |
|----------------|---------------------|--------|--------------------------------------|
| X55068         | Human c-erbB-1 gene | + | GGGGCGGGGAGGGGGGAGGGGAGGGG |
| D26535         | Human gene for dihydrofolamide reductase | + | GGGGCGGGGAGGGGGGAGGGGAGGGG |

Structural RNA

| GenBank number | GenBank Description | Strand | Potential tetraplex forming sequence |
|----------------|---------------------|--------|--------------------------------------|
| X07201         | Mouse H19 gene | + | GGGGCGGGGAGGGGGGAGGGGAGGGG |
| M97168         | Human X (inactive)-specific transcript (XIST) | + | GGGGCGGGGAGGGGGGAGGGGAGGGG |
| M15703         | X. laevis 28S ribosomal RNA gene | + | GGGGCGGGGAGGGGGGAGGGGAGGGG |

Miscellaneous

| GenBank number | GenBank Description | Strand | Potential tetraplex forming sequence |
|----------------|---------------------|--------|--------------------------------------|
| M55173         | D. hydei Z-DNA flanking regions | + | GGGGCGGGGAGGGGGGAGGGGAGGGG |
| X58849         | H. annuus mitochondrial plasmid-like DNA (P1) | + | GGGGCGGGGAGGGGGGAGGGGAGGGG |
| S51329         | Human AAVS1 - adenovirus-associated virus integration site | + | GGGGCGGGGAGGGGGGAGGGGAGGGG |
| U09115         | Human retrotransposable L1 element LRE2 | + | GGGGCGGGGAGGGGGGAGGGGAGGGG |

As is not a unique feature of the previously described “cinched” tetraplex (16), but rather is a property of tetraplexes in general. We have used this property to examine the effects of the length and composition of the stem and spacer regions, as well as the effect of pH and methylation status on tetraplex formation.

Loop Parameters—The shortest sequence capable of producing a block to DNA synthesis in this assay was (T 2G 2) 4 (Fig. 3A). In experiments not shown here, neither single-stranded templates containing (T 2G 2) 4 or the thrombin-binding aptamer, G, T2G2TGTG2T2G2 (22) blocked DNA synthesis even after prolonged incubation in the presence of K+. The thrombin-binding aptamer has recently been shown by NMR to form an intrastand tetraplex in which loops 1 and 3 each contain 2 bases, while loop 2 has 3 bases (22). However, the NMR obser-
A Simple Assay for Tetraplex Formation

vaations of tetraplex formation were made at low temperature (20°C or below) or in the presence of more than 100 mM KCl. Our DMS modification data indicate that, for (TG5)_4, the most stable structures formed have at least 3 bases in each loop. The negative results with (TG3)_2 and the thrombin-binding aptamer suggest that at least 2 tetrads and 3 loop bases are essential for stable tetraplex formation under these conditions. This would be consistent with the observations of Balagurumoorthy and Brahmachari (23) for the human telomeric sequence (T2AG3)_4.

Sequences of the form (TG5)_4 were used to examine the effect of loop size and base composition on the amount of tetraplex formed. In initial experiments, a template containing only a single T residue in each repeat, i.e., x = 1, and y = 5 was compared to templates where x = 2, and y = 4 or 5. In these experiments equally strong tetraplex formation was observed for all templates where y = 5 (Fig. 5). Assuming that 3 bases are required for each loop, (TG5)_4 must have 2 Gs and a T in each loop, with 3 tetrads in the stem, while (TG3)_2 must have 2 Ts and a G, with 4 tetrads in the stem. The fact that (TG5)_4 blocked DNA synthesis to a similar extent as (T2G5)_4 suggests that Gs in the loop stabilize tetraplexes perhaps by means of interloop interactions or stacking energy contributions. This is consistent with our previous observation that the sequence G26 forms a much stronger block to DNA synthesis than any of these templates, completely blocking chain extension (16).

In the next set of experiments, the size and base composition of the central spacer region (spacer 2 in Fig. 1A) of the sequence (TG5)x TGTx TGTx TGTx TGTx was altered. x represents the altered bases which would fall into loop 2 of the tetraplex. There was a small but reproducible increase in the amount of tetraplex formed when loop 2 contained a G residue (Fig. 6A). This is consistent with our earlier observation that Gs in the loop stabilize tetraplexes. The base composition of loop 2 had little effect for spacer lengths of 3 where guanine was not one of the loop bases (Fig. 6A). For spacer regions consisting of TTTT, TTTA, CCC, or AAAA, the pure C and A tracts were more destabilizing than pure T or T-rich tracts. The greater stability of T-rich loops may be the result of base stacking or interactions between loop bases. The potential for T:T base pairing in T-rich loops may be the result of base stacking or interactions between loop bases.

Tetraplex formation occurred even when spacer 2 consisted of 10 Ts (Fig. 6B). The optimal spacer length was found to be between 4 and 6 bases consistent with studies of DNA hairpins, where a loop size of 4–5 nucleotides has been found to be optimal (25).

Templates with 6 Ts in loop 1 (T2T2T6 in Fig. 7) formed a similar amount of tetraplex as did templates with 6 Ts in loop 2 (T2T6T2 in Fig. 7). Increasing the number of bases to 6 in both spacer regions 1 and 3, or in all 3 spacers had little effect on the amount of polymerase arrest seen. Two pathways for tetraplex folding are currently favored (24). The first involves the formation of two hairpins (i.e. between G tracts 1 and 2, and between G tracts 3 and 4, Fig. 1A) that then interact to form the tetraplex. In this model the loops of the two hairpins become loops 1 and 3 of the tetraplex, while the bases between the two hairpins form loop 2 of the tetraplex. The second model involves the formation of a single large hairpin that becomes folded over to form a tetraplex. The loop of this hairpin forms loop 2 of the tetraplex. The fact that tetraplex formation is not affected by the addition of extra bases to loops 1 or 2 is most consistent with the first model, since these bases would be expected to adversely affect the stability of the folding intermediate in the second model.

A Limited Number of Non-G Bases May Be Tolerated in the Stem—The cinched tetraplex formed in the chicken β-globin gene promoter (15, 16) is able to incorporate a number of non-G bases into the stem of the tetraplex. Much of the destabilizing effect expected of these bases we believe to be offset by stability provided by the “cinch” (16). To test whether non-G bases could be incorporated into a normal tetraplex, and to examine the effect of different non-G bases on tetraplex stability, we altered a model tetraplex-forming sequence so that it would contain a single non-G base in the stem.

The introduction of either a single A or C into the stem of the sequence (TG5)_4 greatly reduces the extent of polymerase arrest (Fig. 8A, panels (A1) and (C1)), consistent with the effect of disruption of a tetrad. The effect of the C residue is added proof that hairpin formation is not involved in K⁺-dependent DNA synthesis arrest since the substitution of a G-C pair for a G-G one should not reduce the stability of such a hairpin. Cytosine methylation had no effect on DNA synthesis arrest (Fig. 8B), nor did lowering the pH to 6.5 (data not shown).

Tetraplex Formation at Physiological Temperature and K⁺ Concentrations—To see whether tetraplex formation was possible at physiological temperatures, we carried out DNA synthesis arrest assays on templates containing the sequence (TG5)_4, where n = 2–4. Templates were heated and cooled, in the absence of K⁺, and then K⁺ added after the reaction had equilibrated at 37°C for 5 min. K⁺-dependent tetraplex formation was seen for templates in which n ≥ 3 even under these conditions (data not shown).

Arrest of DNA synthesis is seen with as little as 5 mM K⁺ and increases as the K⁺ concentration increases (data not shown). Since intracellular K⁺ concentrations range from 100 to 150 mM in mammalian cells (26), these results suggest that tetraplexes could be formed in vivo by relatively short G-rich sequences whenever DNA is single-stranded for short periods of time.

DISCUSSION

We have shown here that the formation of a K⁺-dependent block to DNA synthesis is a general diagnostic feature of intranstrand tetraplexes. The K⁺ ion specificity is probably related to the fact that K⁺ is unique among the monovalent cations in being both small enough to enter the stem cavity and large enough to stabilize tetraplexes by forming a coordination complex with Gs in adjacent tetrads (4). The K⁺ effect on polynucleotide chain extension makes a simple, rapid, and specific assay for sequences able to form such structures that is easy to visualize and very sensitive. Structures can be resolved at single base resolution on sequencing gels, and it is possible to detect multiple structures simultaneously even when only a small fraction of the molecules have formed a tetraplex. Tetraplex formation can be observed under conditions that might reasonably be expected to occur in cells, and the assay is amenable to studies of factors that affect tetraplex stability.

We have used this assay to explore some of the sequence parameters that affect tetraplex formation and stability. We show that at least 3 bases in each loop are needed, and that Gs in the loops have a stabilizing effect. Ts, Cs, and As all have similar effects on tetraplex stability for loop sizes of 3, but when the loop sizes were increased, Ts are slightly more stabilizing than either As or Cs. At least 10 Ts can be tolerated in the central loop, with 4–6 being optimal. Our data indicate that a folding mechanism that involves the interaction of two small hairpin intermediates to form the tetraplex is a significant pathway for the formation of these structures. We also show that even a single non-G base in the putative tetraplex stem severely destabilizes the structure, although some arrest is still observed. These structures are, however, relatively insensitive.
to differing numbers of bases in the constituent loops. This is interesting since it indicates that sequences lacking a recognizable 4 G-repeat motif may still be able to form tetraplexes. This vastly increases the number of sequences that can potentially form such structures.

From the experiments described in this report we were able to identify a number of sequences in GenBank™ that might form strong intranstrand tetraplexes (Table I). In addition to DNA tetraplexes, it is probable that transcripts containing these sequences also form tetraplexes, and an effect of these structures at the RNA level is thus also possible.

A number of the sequences listed in Table I are located in the 5′-regulatory region of a variety of eukaryote genes. It is possible that these structures have a regulatory role by affecting protein binding to the promoter. For example, the tetraplex in the chicken β-globin promoter overlaps with the binding site of an erythroid-specific DNA binding protein (27). We speculate that tetraplex formation may block binding of this protein and thus represents one level at which regulation of the β-globin gene could be accomplished. Tetraplex-forming sequences are also located in the 3′-untranslated regions of a number of interesting genes including the human genes encoding arginosuccinate lyase (28), erythropoietin (29), and acrosin (30). A number of nontranslated RNAs also have tetraplex-forming potential. Examples of this class of sequences include the 28 S ribosomal RNA gene in Xenopus (31), the human X (inactive)-specific transcript (32), and the product of the mouse H19 gene, a developmentally regulated gene under the control of raf and Rif (33). We have confirmed that at least one of the sequences shown in Table I, the one found in the 3′-untranslated region of the human retrotransposable element known as L1 (34), does indeed form a tetraplex. This structure may play a role in L1 repositioning because it has been conserved for the ~100 million years since the mammalian radiation in the absence of sequence conservation. These structures may act in other contexts as cell cycle check points by slowing replication through these regions. We would also expect that such structures would block transcription and translation. Transcriptional blocks may function to prevent unwanted transcriptional read through from an upstream promoter, while translational blocks may prevent inappropriate protein synthesis by blocking scanning by the 40 S ribosomal subunit.

The biological significance of intranstrand DNA tetraplexes is obviously contingent on their ability to form in vivo. In order for such structures to form, the DNA must be single-stranded. In the assay described here, single strands were obtained by heat denaturation. In eukaryotic cells it is thought that only relatively small regions of DNA are unpaired during replication, although it has been suggested that many hundreds of bases would be unpaired under some circumstances (35). However, even if this were so, much of this DNA would probably be bound by single-stranded binding proteins. On the other hand, tetraplex formation has been demonstrated in supercoiled plasmids in vitro (36, 37), suggesting that under some conditions these structures may be formed in otherwise duplex DNA. Direct evidence for tetraplex formation in vivo has been obtained for at least one sequence, that of the human insulin promoter (8), suggesting that formation of DNA tetraplexes by sequences such as those shown in Table I might in fact be possible.

Our exploration of sequence variations in tetraplex-forming sequences is clearly not exhaustive. It does, however, illustrate the flexibility of the requirements for tetraplex formation, and greatly expands the category of known tetraplex-forming sequences.

Acknowledgments—We thank Herbert Tabor and Deborah Hinton for critical reading of this manuscript, and Anthony V. Furano for many helpful discussions.

REFERENCES

1. Zimmerman, S. B., Cohen, G. H., and Davies, D. R. (1975) J. Mol. Biol. 92, 181–192
2. Leonard, G. A., Zhang, S., Peterson, M. R., Harrop, S. J., Hellwell, J. R., Cruse, W. B., d’Estainnot, B. L., Kennard, O., Brown, T., and Hunter, W. N. (1995) Structure 3, 335–340
3. Ketchum, A., Kumar, R. A., and Patel, D. J. (1995) J. Mol. Biol. 254, 688–656
4. Sen, D., and Gilbert, W. (1988) Nature 334, 364–366
5. Williamson, J. R., Raghuraman, M. K., and Cech, T. R. (1989) Cell 59, 871–880
6. Faeg, G., and Cech, T. R. (1993) Biochemistry 32, 11646–11657
7. Giraldo, R., Suzuki, M., Chapman, L., and Rhodes, D. (1994) Proc. Natl. Acad. Sci. U. S. A. 91, 7658–7662
8. Hammond-Kosack, M. C., Kilpatrick, M. W., and Dougherty, K. (1992) J. Mol. Endocrinol. 9, 221–227
9. Guo, Q., Xu, M., and Kallenbach, N. R. (1993) Biochemistry 32, 3596–3603
10. Hardin, C. C., Watson, T., Corregan, M., and Bailey, C. (1992) Biochemistry 31, 833–841
11. Hardin, C. C., Henderson, E., Watson, T., and Prosser, J. K. (1991) Biochemistry 30, 4460–4472
12. Kang, S. M., Wohlrab, F., and Wells, R. D. (1992) J. Biol. Chem. 267, 1259–1264
13. Oka, Y., and Thomas, C. A. J. (1987) Nucleic Acids Res. 15, 8877–8898
14. Sen, D., and Gilbert, W. (1990) Nature 344, 410–414
15. Woodford, K. J., Howell, R. M., and Usdin, K. (1994) J. Biol. Chem. 269, 27029–27035
16. Howell, R. M., Woodford, K. J., Wetzmann, M., and Usdin, K. (1996) J. Biol. Chem. 271, 5208–5214
17. Usdin, K., and Woodford, K. J. (1995) Nucleic Acids Res. 23, 4202–4209
18. Chen, F.-M. (1995) J. Biol. Chem. 270, 23090–23096
19. Kaspar, P., Zadrazil, S., and Fabry, M. (1989) Nucleic Acids Res. 17, 3616
20. Samadashwily, G. M., Dayn, A., and Mirkin, S. M. (1993) EnBio 1, 12, 4975–4983
21. Baran, N., Lapidot, A., and Manor, H. (1991) Proc. Natl. Acad. Sci. U. S. A. 88, 507–513
22. Macaya, R. F., Schultze, P., Smith, F. W., Roe, J. A., and Felgon, J. (1993) Proc. Natl. Acad. Sci. U. S. A. 90, 3745–3749
23. Balagurumoorthy, P., and Brahmachari, S. K. (1994) J. Biol. Chem. 269, 21858–21869
24. Balagurumoorthy, P., Brahmachari, S. K., Mohanty, D., Bansal, M., and Sadasivkaran, V. (1992) Nucleic Acids Res. 20, 4061–4067
25. Haas-Honeck, C. A. G., Hilbers, C. W., van der Mareel, G. A., van Boon, J. H., Singh, U. C., Pattabiraman, N., and Kirkman, P. A. (1986) J. Biomol. Struct. Dyn. 3, 843–857
26. Lai, Y. T., Yassin, R. R., and Horowitz, S. B. (1988) Science 240, 1321–1323
27. Lewis, C. D., Clark, S. P., Felsenfeld, G., and Gould, H. (1988) Genes Dev. 2, 963–973
28. Abranson, R. D., Barbosa, P., Kalumuk, K., and O’Brien, W. E. (1991) Genomics 10, 126–132
29. Lin, F.-K., Suggs, S., Lin, C.-H., Browne, J. K., Smalling, R., Egejir, C. J., Chen, K.-K., Fox, G. M., Martin, F., Stasiński, B., Badrawi, S. M., Lai, P.-H., and Goldwasser, E. (1995) Proc. Natl. Acad. Sci. U. S. A. 92, 7580–7584
30. Keime, S., Adham, I. M., and Engel, W. (1990) Eur. J. Biochem. 190, 195–200
31. Ware, V. C., Tague, B. W., Clark, G., Gourse, R. L., Brand, R. C., and Gerbi, S. A. (1983) Nucleic Acids Res. 11, 7765–7781
32. Brown, C. J., Ballabio, A., Rupert, J. L., Laferriere, R. G., Grompe, M., Tonioni, R., and Willard, H. F. (1991) Nature 349, 38–44
33. Pachnis, V., Brannan, C. I., and Tilghman, S. M. (1988) EMBO J. 7, 673–681
34. Dombrowski, B. A., Mathias, S. L., Nanthakumar, E., Scott, A. F., and Kazazian, H. H. (1991) Science 254, 1805–1808
35. Kunket, T. A. (1992) Bioessays 14, 303–308
36. Voloshin, O. N., Veselev, A. G., Belotserkovskii, B. P., Danilevskaya, O. N., Pavlova, M. N., Dobrynin, V. N., and Frank-Kamenetski, M. D. (1992) Biol. Struct. Dyn. 9, 643–652
37. Hammond-Kosack, M. C., Dobrinski, B., Lurz, R., Doeherty, K., and Kilpatrick, M. W. (1992) Nucleic Acids Res. 20, 231–236