

Replication of simple DNA repeats

Review Article

Maria M. Krasilnikova, George M. Samadashwily and Sergei M. Mirkin*

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

Corresponding Author: Phone:(312)996-9610; Fax: (312)413-0353; E-mail:mirkin@uic.edu

Key Words: DNA repeats, inverted repeats, DNA replication, repeat length polymorphism, replication attenuation

Abbreviations: WC, Watson-Crick; IR, inverted repeats; MR, mirror repeats; DTR, direct tandem repeats; S-DNA, slipped-stranded DNA

Received: 7 October 1998; accepted: 16 October 1998

Summary

This chapter presents an overview of studies on the replication of simple DNA repeats conducted in our laboratory during the last seven years. The recent massive increase in available DNA sequences has led to the clear understanding that natural DNAs, particularly in eukaryotes, are extraordinarily enriched in different repeats (Schroth and Ho, 1995; Cox and Mirkin, 1997). This leads to an obvious question: what are the biological functions (if any) of these repeated elements? This problem is currently the subject of very intense studies in many laboratories all over the world. We came to this question after we realized that many repeated DNA sequences constitute a major obstacle to DNA polymerization *in vitro* (Dayn et. al., 1992; Samadashwily et. al., 1993; Samadashwily and Mirkin, 1994; Krasilnikov et. al., 1997). Subsequently, we found that several such repeats attenuate DNA replication *in vivo* as well (Samadashwily et. al., 1997; Krasilnikova et. al., 1998). Based on our data, we conclude that there are at least three mechanisms by which different repeats inhibit replication. We believe that this may reflect a potentially important role of repeated DNA as punctuation marks for major genetic processes in DNA texts. Repeat-caused replication attenuation might also contribute to the mechanisms of repeat length polymorphism seen in many human diseases.

I. Repeat types, structures and frequencies

Based on sequence arrangement and symmetry, three major types of simple DNA repeats are usually considered (Fig. 1): inverted repeats, mirror repeats, and direct tandem repeats. Inverted repeats (IR) are DNA sequences in which DNA bases that are equidistant from the symmetry center in a DNA strand are Watson-Crick (WC) complements to each other. Mirror repeats (MR) are also symmetrical, but here equidistant DNA bases are identical to each other. Finally, direct tandem repeats (DTR) are simple, uninterrupted iterations of a core repeat unit along the DNA strand. The distinction between the different repeat types is not absolute. There are DNA sequences that meet criteria for all three repeat types. A well studied example is the $d(A-T)_n \cdot d(T-A)_n$ sequence which is simultaneously an inverted, mirror and direct tandem repeat.

The secondary structure of repeated DNA often differs dramatically from canonical B-DNA. The exact conformation of a repeated DNA depends on its symmetry, base composition, DNA supercoiling, ambient conditions, etc. Below, we will briefly summarize the best characterized structures formed by different DNA repeats.

Inverted repeats are capable of forming cruciform structures in double stranded DNA or hairpins in single stranded DNA (Fig. 2). While hairpin formation in single-stranded DNA is generally energetically favorable, cruciform formation in a double-stranded DNA is only favorable under the influence of negative supercoiling (Lilley, 1980; Panayotatos and Wells, 1981; Mizuuchi, et. al., 1982). Indeed, to convert a duplex DNA segment into a cruciform state, one would need to at least partially unwind it in order to allow for self-pairing by each DNA strand. Since initial unwinding of a DNA duplex is energy consuming, this stage represents an energetic barrier for cruciform formation (reviewed in Sinden, 1994). In

addition, the resultant cruciform structure contains single-stranded bases in central loops and energetically costly

Fig. 1. Different repeat types. **A:** inverted repeats; **B:** mirror repeats; **C:** direct tandem repeats. Arrows of the same color represent symmetrical sequences. Complementary sequences differ in color.

Fig. 2. Cruciform and hairpin structures formed by inverted repeats in double- and single-stranded DNA, respectively. Complementary halves of an inverted repeat are **red** and **green**, single-stranded segments are **orange**, and the surrounding DNA is **purple**.

junctions between the cruciform and the adjacent duplex DNA (4-way junctions). Altogether, this leads to a high nucleation energy for cruciform formation, approaching 20 kcal mol⁻¹ (reviewed in Vologodskii, 1992). Since a cruciform is topologically equivalent to unwound DNA, its formation under torsional stress release negative supercoils, compensating for the high nucleation cost (Vologodskii and Frank-Kamenetskii, 1982). In fact, simple energetics calculations show that the probability of cruciform extrusion must increase exponentially with inverted repeat length (Benham, 1982; Vologodskii and Frank-Kamenetskii, 1982). *In vivo* formation of cruciform structures was observed in several studies. The most direct evidence was obtained by chemical probing of plasmids in *E. coli* cells. AT-rich inverted repeats were shown to adopt cruciform conformation when intracellular supercoiling increased due to certain changes in environmental conditions (McClellan, et. al., 1990; Dayn, et. al., 1991; Zheng, et. al., 1991) or as a consequence of transcriptional activation (Dayn, et. al., 1992). Formation of a cruciform-like structure in the enhancer of the enkephalin gene was suggested based on chemical probing of human intracellular DNA followed by ligation-mediated PCR (Spiro et. al., 1995).

While mirror repeats of arbitrary composition are common in natural DNA (Schroth and Ho, 1995; Cox and Mirkin, 1997), only one type of them, i.e. homopurine-homopyrimidine mirror repeats (H-palindromes), has been shown to adopt a non-B conformation. These sequences can adopt an intramolecular triplex called H DNA (Lyamichev, et. al., 1986). To form this structure (**Fig. 3**), a DNA strand from one half of the repeat folds back, forming a triplex with the duplex half of the repeat, while its complement remains single-stranded (Mirkin et. al., 1987). Depending on the chemical nature of the strand donated to the triplex, either pyrimidine- or purine-rich, the resultant structures are called H-y or H-r, respectively (**Fig. 3**). The H-y form is built from TA*T and CG*C⁺ triads (**Fig. 4A**), where pyrimidines from the third strand are situated in the major groove, forming Hoogsteen hydrogen bonds with the purines of the duplex (Hoogsteen, 1963). The extingency for cytosine protonation makes this structure preferred under acidic pH (Mirkin and Frank-Kamenetskii, 1994). The H-r form can be built of CG*G, TA*A and, unexpectedly, TA*T triads (Kohwi and Kohwi-Shigematsu, 1988; Beal and Dervan, 1991; Dayn, et. al., 1992). In this case, DNA bases of the third strand form reverse Hoogsteen hydrogen bonds with the purines of the duplex (**Fig. 4B**) (Hoogsteen, 1963). These triads are stable at physiological pH, but are greatly stabilized in the presence of divalent cations (Kohwi, 1989; Bernues, et. al., 1990; Beltran, et. al., 1993; Malkov, et. al., 1993; Martinez-Balbas and Azorin, 1993). Like cruciform structures, a H DNA is topologically equivalent to the completely unwound DNA and its formation requires substantial duplex unwinding (Lyamichev, et. al.,

1985). Thus it is favored in negatively supercoiled DNA, and its formation depends exponentially on repeat length (reviewed in Mirkin and Frank-Kamenetskii, 1994). Cloned homopurine-homopyrimidine repeats were shown

to adopt H conformation in *E. coli* cells when the intracellular supercoiling increased due to mutations in topoisomerase

Fig. 3. H DNA structure. Both H-y and H-r forms are shown. **Red line:** homopurine and **green line:** homopyrimidine strands of an H palindrome, respectively. **Purple lines:** adjacent DNA. **Black lines:** Watson-Crick hydrogen bonds; **stars:** Hoogsteen hydrogen bonds.

Fig. 4. Triplex forming triads. **A:** Hoogsteen triads forming H-y DNA. **B:** reverse Hoogsteen triads forming H-r DNA.

genes, chloramphenicol treatment, or transcriptional activation (Kohwi, et. al., 1992; Kohwi and Panchenko,

1993; Ussery and Sinden, 1993). Antibodies against triplex DNA were also shown to interact with

chromosomes of permeabilized mammalian cells (Agazie, et. al., 1996).

Direct tandem repeats (DTRs) can adopt a variety of conformations. As is obvious from the above discussion, some DTRs can form cruciforms or H DNA as long as they happened to be inverted repeats or H-palindromes, respectively. Another structure, called a G-quartet (**Fig. 5**) can be formed by DTRs containing tandemly arranged runs of guanines (Gellert, et. al., 1962; Zimmerman, et. al.,

It is built from stacked G₄ blocks that are additionally stabilized in the presence of monovalent ions (Pinnavaia, et. al., 1978; Williamson, et. al., 1989; Murchie and Lilley, 1994; Weitzmann, et. al., 1997). This structure is definitely formed by single-stranded G-rich DTRs, but there are also indications that it can exist in superhelical DNA (Ahmed, et. al., 1994). Formation of G-quartets *in vivo* has never been directly demonstrated.

DTRs consisting of regularly alternating purines and pyrimidines can adopt left-handed Z DNA conformation (Mitsui, et. al., 1970; Pohl and Jovin, 1972; Wang, et.

Fig. 6. Schematic representation of slipped-stranded DNA structure (S-DNA). Red and green lines represent complementary strands of a DTR. Purple lines - surrounding DNA.

al., 1979; reviewed in Rich, et. al., 1984). In linear DNA, this structure is only possible under rather exotic conditions (such as very high ionic strength) (Peck, et. al., 1982; Singleton, et. al., 1982). In superhelical DNA, by contrast, it is extremely favorable under physiological conditions, since it releases twice as many supercoils per DNA base as unwound DNA, cruciforms or H DNA (Singleton, et. al., 1982). Z-DNA was detected in bacterial cells after an increase in DNA supercoiling due to environmental changes or transcription (Haniford and Pulleyblank, 1983; Jaworski, et. al., 1989; Rahmouni and Wells, 1992). In permeabilized mammalian cells anti-Z antibodies specifically interact with chromosomes, targeting upstream parts of actively transcribed genes (Wittig, et. al., 1992; Wolf, et. al., 1996).

Finally, DTRs of various base compositions can adopt a structure called slipped-stranded DNA (S-DNA) (reviewed in Sinden, 1994). This structure (**Fig. 6**) utilizes the multiply repeated nature of the sequence: upon

Fig. 5. Quadruplex DNA structure. A. General overview. Black line - DNA strand, purple rectangles - stacked G-quartets. B - Chemical structure of a G-quartet.

1975; Sen and Gilbert, 1988; Sundquist and Klug, 1989).

denaturing and renaturing, the complementary repeats can mispair, resulting in a peculiar combination of double-helical stretches intervened by single-stranded loops. In linear DNA, this conformation is thermodynamically unfavorable but can be trapped kinetically. In superhelical DNA, it might become favorable given the release of substantial torsional tension. It is worth noting, that for core repeated units of certain base compositions, the loops can be additionally stabilized by hydrogen bonds of both WC and non-WC nature (Pearson and Sinden, 1996). This would certainly make S-DNA more favorable. For example, formation of S-DNA was suggested for expandable (CXG)_n trinucleotide repeats in linear or superhelical DNA upon denaturing/renaturing (Pearson and Sinden, 1996; Chen, et. al., 1998; Mariappan, et. al., 1998; Pearson, et. al., 1998; Pearson, et. al., 1998). In this case, the loops are likely to be stabilized by CG base pairs and some non-WC pairs such as GG. Although there is some indirect evidence for S-DNA *in vivo*, especially during DNA replication (reviewed in Pearson and Sinden, 1998), direct proof is still lacking.

The recent availability of large genomic texts of many different organisms has allowed their detailed computer analysis. This analysis (Trifonov, et. al., 1985; Karlin, 1986; Morris, et. al., 1986; Manor, et. al., 1988; Smillie and Bains, 1990; Lagercrantz, et. al., 1993; Han, et. al., 1994; Schroth and Ho, 1995; Karlin and Burge, 1996; Cox and Mirkin, 1997; Raghavan, et. al., 1997; Saunders, et. al., 1998) as well as numerous experimental approaches, including pattern matching (Galas, et. al., 1985), word frequency counting (Karlin and Burge, 1995) and basic linguistic techniques (Pevzner, et. al., 1989; Pevzner, et. al., 1989), has revealed that simple repeated sequences are remarkably abundant in natural DNAs, particularly in eukaryotic genomes.

We have recently carefully evaluated the representation of different repeat types in prokaryotes, eukaryotes and bacteria, and compared those values with the expected frequencies based on the local DNA base composition (Cox and Mirkin, 1997). This analysis led to several important conclusions. It became evident that simple DNA repeats of substantial length (>24 bp-long) occur in genomes with much higher frequency than it would be statistically predicted. However, genomes belonging to different kingdoms of life (prokarya, eukarya and archaea) are enriched in different types of repeats. Eukaryotic genomes showed the enrichment of all three types of simple repeats. Of all repeats, mirror repeats, and particularly H-palindromes, were the most overrepresented, reaching 10⁹ over the chance value. Bacterial genomes and organelles have a substantial overrepresentation of inverted repeats and sometimes direct tandem repeats. In contrast, in archae none of the repeats were abundant.

The enrichment in different repeats shows an interesting length dependence. The chance frequency,

calculated taking into account local GC-content, exponentially decreases with the length of repeats. The actual frequency of repeat occurrence also decreased exponentially but at a much slower rate. As a result, the normalized frequencies of overrepresented repeats showed almost perfect exponential increasing lengths (**Fig. 7**). This is particularly interesting since, as discussed above, the structure forming ability of a repeat also increases exponentially with length. One might speculate that the abundance of long repeats may indicate an evolutionary advantage conferred by unusual DNA structures.

While these questions are at the focus of numerous studies, we were specifically interested in the mechanisms of simple repeat replication. Those studies, which started from the effects of H DNA on DNA polymerization *in vitro* and expanded into the analysis of replication of different repeats *in vivo*, are outlined below.

Fig. 7. Ratio of observed to expected frequencies of H palindromes for pro- and eukaryotic genomes. **Red circles:** *H. sapiens* genome; **green squares:** *E. coli* genome.

II. Effects of simple DNA repeats on DNA polymerization *in vitro*

It has long been known that simple DNA repeats affect DNA polymerization *in vitro*, presumably via the unusual conformation of the DNA template. Many instances of inverted repeats slowing down different DNA polymerases, most likely due to hairpin formation, have been described (Sherman and Gefter, 1976; Chalberg and Englund, 1979; Huang and Hearst, 1980; Kaguni and Clayton, 1982; Weaver and DePamphilis, 1982; Bedinger, et. al., 1989). Tetraplex-forming repeats also inhibit DNA synthesis carried out by many different DNA polymerases

(Woodford, et. al., 1994; Usdin and Woodford, 1995; Weitzmann, et. al., 1997). The polymerization arrest was K^+ -dependent, strongly indicating the role of intrastranded G-quartets. Finally, numerous homopurine-homopyrimidine stretches are known to impede DNA polymerization, presumably due to triplex formation (Lapidot, et. al., 1989; Baran, et. al., 1991; Dayn, et. al., 1992; Samadashwily, et. al., 1993; Mikhailov and Bogenhagen, 1996; Krasilnikov, et. al., 1997). However, the detailed mechanisms of repeat-caused polymerization blockage were largely unknown until recently.

Our interest to the effects of repeated DNA on polymerization arose from the pioneering studies of Manor and colleagues (Lapidot, et. al., 1989; Baran, et. al., 1991). They found that polymerization by Klenow or Taq polymerases on single-stranded DNA templates was partially blocked within $d(GA)_n$ or $d(CT)_n$ tracts. Since polymerization halted in the middle of those stretches, they suggested that when the newly synthesized DNA strand reached the center of a stretch its remaining part folded

Fig. 8. Models of triplex-caused polymerization arrest.

A. polymerization on single-stranded template;
B. polymerization on linear double-stranded template;
C. polymerization on supercoiled template.

Red rectangles: homopurine halves of an H-palindrome; **green rectangles:** homopyrimidine halves of an H-palindrome; **purple lines:** growing DNA strands; **purple arrows:** polymerization direction; **black lines:** parent DNA strands; **purple stars:** polymerization stop sites.

back to form a triplex (**Fig. 8A**). As a result, the polymerase is trapped and is unable to continue elongation. This hypothesis was supported by the characteristic pH dependence for $d(CT)_n$ tracts, and the reversal of termination by substituting dGTP by deazaGTP which is incapable of forming Hoogsteen hydrogen bonds.

We first analyzed DNA polymerization on superhelical DNA templates containing different isoforms of H-r DNA (Dayn, et. al., 1992). We found that DNA polymerase terminates at specific sites on both DNA chains within supercoiled templates containing these structures. The location of the termination sites differed for various isoforms but always coincided with triplex boundaries as defined by chemical probing (**Fig. 8C**). We concluded, therefore, H DNA prevents DNA polymerization.

Subsequently, we analyzed DNA polymerization through H-forming repeats in double-stranded open circular DNAs, where the triplex structure did not exist prior to polymerization (Samadashwily, et. al., 1993). DNA polymerization stopped almost completely at the center of those sequences but only when the homopyrimidine strand served as a template. Mutations that destroyed H-forming potential of a repeat abolished polymerization blockage, while compensatory mutations restoring H-forming potential, restored polymerization arrest as well. We concluded that the formation of H-r DNA during DNA polymerization was responsible for the observed polymerization arrest. During DNA synthesis on a double-stranded template, the DNA polymerase must displace the non-template DNA strand. When the displaced segment contains the purine-rich half of an H-motif, it can fold back to form an intramolecular triplex downstream of the polymerase, which, in turn, blocks polymerase progression (**Fig. 8B**).

The severity of the triplex-caused polymerization blockage led us to wonder about the mechanisms of their inhibitory effects. Several possibilities should be considered. First, under polymerization conditions, triplexes may be so much more stable than duplexes that the triplex blockage of polymerization is a simple reflection of their persistence. Second, the kinetics of polymerase passage through triplexes may be much slower

than through duplexes, simulating polymerization blockage. Finally, DNA polymerases, while capable of dismantling duplexes, may be unable to do so with triplexes.

Our recent study (Krasilnikov, et. al., 1997) distinguished between these possibilities. We used single-stranded DNA templates containing intrastranded H-r triplexes or control duplexes and studied the efficiency of Vent DNA polymerase passage at different temperatures and time intervals. In parallel, the stability of different triplex and duplex structures was determined in DNA melting experiments. At physiological temperatures, we found that triplexes completely block polymerization, but duplexes just slow it down several fold. Melting temperature curves showed that triplexes were only slightly more stable than the corresponding duplexes. Such small differences are unlikely to account for the dramatic differences in temperatures (up to 40°C) at which the polymerase traverses these structures. Projection of polymerase passage temperatures onto melting curves for different structures revealed that the polymerase passes triplex barriers at temperatures where they start to dissociate, whereas duplexes are overcome far below their dissociation temperatures (**Fig. 9**). This shows that DNA polymerase can slowly untangle duplexes in DNA templates, but not triplex structures.

Fig. 9. Comparison of melting curves for a triplex and duplex with the temperatures of DNA polymerase passage through these structures. **Blue circles:** triplex melting data; **red circles:** duplex melting data; **arrows:** temperatures of polymerase passage (**blue:** triplex template; **red:** duplex

template).

Based on these results, it is plausible to speculate that the elongating DNA polymerase is equipped to sense the structure of the DNA template ahead of it. Single-stranded DNA is an optimal template, double-helical segments represent an obstacle which can be slowly unwound by DNA polymerase, and unusual template conformations, such as triplexes or quadruplexes, represent steady roadblocks.

It is highly likely that other enzymes of DNA metabolism might experience similar problems while tracking along DNA. One important question is whether different RNA polymerases are similarly sensitive to the conformation of a DNA template. This question is less studied, but there are some provocative data suggesting that this is the case. RNA polymerase was shown to stall within or immediately after several H-palindromes (Reaban and Griffin, 1990; Reaban, et. al., 1994; Grabczyk and Fishman, 1995; Kiyama and Oishi, 1996). This stalling profoundly depended on the repeat's orientation in the transcription unit. In most cases, it occurred when the transcript carried an oligopurine stretch, though for $d(A)_n \cdot d(T)_n$ repeat, it happened for the oligo(U) transcript (Kiyama and Oishi, 1996). It was suggested that stalling occurs upon formation of a three-stranded complex between RNA and DNA strands corresponding to the H-palindrome. In this complex, RNA is resistant to RNase A but cleaved by RNase H, and DNA is unwound as evident from the release of supercoils (Reaban, et. al., 1994; Grabczyk and Fishman, 1995).

The exact nature of this complex remains unknown, and several possibilities are currently considered. One idea is that RNA polymerization generates negative supercoiling upstream of the enzyme, provoking transient H DNA formation. This structure might become kinetically trapped if an RNA transcript binds to its single-stranded portion (Grabczyk and Fishman, 1995). Formation of such a trapped complex immediately upstream of the RNA polymerase might attenuate its propagation. Another hypothesis is that transcription through a homopurine-homopyrimidine sequence could create an unusually long and stable R-loop (Reaban, et. al., 1994). The non-template DNA strand could collapse onto this R-loop, forming some hydrogen bonds with either the DNA or RNA strand as possible (collapsed R-loop) (Reaban, et. al., 1994). Future studies are needed to understand the structure of this RNA/DNA complex and how it affects RNA polymerization.

III. Effects of simple DNA repeats on DNA replication *in vivo*.

Because DNA polymerases cannot efficiently pass structured parts of DNA templates, one might envision problems during DNA replication *in vivo*. It is well documented that a portion of the lagging strand DNA template (of an Okazaki fragment size) must be single-stranded in order to pursue coordinated synthesis of both DNA strands (reviewed in Kornberg and Baker, 1992). This several hundred bp-long single stranded piece can adopt a plethora of different conformations, potentially serving as roadblocks for DNA polymerase, unless accessory replication proteins, including single-stranded DNA binding proteins and DNA helicases, helped to remove them. In some cases, however, even they may not be sufficient, as indicated by observations that the replication fork as a whole stalls within some simple DNA repeats *in vivo* (Rao, et. al., 1988; Brinton, et. al., 1991; Rao, 1994).

This consideration encouraged us to study the mode of replication fork progression through simple DNA repeats *in vivo* (Samadashwily, et. al., 1997; Krasilnikova, et. al., 1998). As discussed above, these repeats are widespread in natural DNAs, and can be cloned and maintained in many model systems including bacteria and yeast. This clearly shows that they are able to replicate *in vivo*. However, one might expect that the rate of replication fork progression through the repeated DNA is slower. Unfortunately, this is a difficult problem to study, since the normal replication rate is very fast, ranging from 1000 bp/sec in bacteria to several hundreds bp/sec in eukaryotes (reviewed in Kornberg and Baker, 1992). For example, given a 100 bp-long repeat in pBR322 slows the replication fork progres-

Fig. 10. Detection of repeat-caused replication blocks by 2-dimensional gel-electrophoresis. **A.** Schematic representation of our approach. Upper panel shows the structure of the linearized plasmid DNA. The green triangle corresponds to the replication origin, the red box corresponds to cloned repeated DNA. The lower left panel shows the shapes of different replication intermediates. The red intermediate corresponds to the one which preferentially accumulates due to repeat-caused replication blockage. The lower right panel shows the replication arc. The red circle corresponds to the replication stop site. **B.** Actual electrophoregram of replication intermediates of a plasmid containing a $d(G)_{32} \cdot d(C)_{32}$ repeat. The red arrow points to the replication stop site.

sion 10-fold, the overall plasmid replication would only be slowed from 5 sec to 6 sec. Therefore, most conventional methods of DNA replication analysis are not applicable to this problem.

To solve this problem we decided to analyze the effects of different DNA repeats on the replication of bacterial plasmids *in vivo* using an approach called 2-dimensional neutral/neutral electrophoresis of replication intermediates. This technique was developed for mapping of the replication origins (Brewer and Fangman, 1987; Huberman, et. al., 1987) but lately has become instrumental in defining replication termination sites as well (MacAllister, et. al., 1990; Zhu, et. al., 1992; Little, et. al., 1993). Bacterial plasmids were chosen for two reasons: (i) they replicate unidirectionally which unequivocally determines leading and lagging strands during DNA replication; (ii) they replicate very efficiently which allows an easy isolation and analysis of replication intermediates.

The idea of electrophoretic analysis of replication intermediates applied to unidirectional replication is presented in **Fig. 10A**. Intermediate products of plasmid replication are Θ -shaped. Upon cleaving these intermediates with a restriction enzyme upstream of the replication origin, they convert into bubble-shaped molecules, where the size of the bubble correlates with the duration of replication. Bubble intermediates differ in their molecular mass (ranging from 1 to 2 plasmid masses) and shape. They are separated in two dimensions: first by mass (low percentage agarose) and second by mass and shape

(high percentage agarose with ethidium bromide). Southern blotting hybridization with the radioactive plasmid probe reveals a so-called bubble arc. If there are no roadblocks during DNA replication, this arc is smooth. Stalling of the replication fork at a specific DNA repeat, however, leads to the accumulation of an intermediate of a given size and shape, generating a bulge on the arc. The ratio of the signal of this bulge to the signal of the corresponding area of a smooth replication arc (relative stop strength, RSS) is an index of replication fork retardation by the repeat.

Using this approach we found that different simple DNA repeats, including $d(\text{CGG})_n \cdot d(\text{CCG})_n$, $d(\text{CTG})_n \cdot d(\text{CAG})_n$, $d(\text{G})_n \cdot d(\text{C})_n$, $d(\text{G-A})_n \cdot d(\text{T-C})_n$, etc., block the replication fork progression. The typical picture of such repeat-caused blockage is presented in **Fig. 10B** for the $d(\text{G})_{32} \cdot d(\text{C})_{32}$ repeat. In this case the RSS is ≈ 30 , i.e. this repeat slows the replication down 30-fold. Notably, in all

cases, longer repeats caused more profound replication stops than the shorter ones.

To prove that replication stops coincides with those repeats, we used a modified version of the electrophoretic analysis of replication intermediates (Friedman and Brewer, 1995). After the first dimension of electrophoresis, replication intermediates were digested with a restriction enzyme in the gel. The enzymes selected for this analysis cut the plasmid either upstream or downstream of the repeat. As a result, a fraction of bubble-shaped intermediates converted into identical y-shaped intermediates (**Fig. 11A**). In the second dimension of electrophoresis, these intermediates migrate similarly and can be detected as a horizontal line upon hybridization with a probe adjacent to the replication *ori*. As is clear from **Fig. 11A**, restriction cleavage downstream of the repeat (relative to the *ori*) would leave the bulge on the bubble-

Fig. 11. Mapping of the replication stop sites. **A.** Schematic representation of 2-D gel-electrophoresis upon restriction cleavage after the first dimension. The red square shows the $d(\text{G})_{32} \cdot d(\text{C})_{32}$ insert. Purple vertical lines show *Hind*III and *Eco*RI restriction sites located upstream and downstream from the insert, respectively. The stalled replication intermediate is shown in red. Upon *Eco*RI digestion, this stalled intermediate should remain bubble-shaped and, thus, remain on the arc after the second dimension of

electrophoresis (right panel). In contrast, upon *Hind*III digestion, this intermediate should become y-shaped and move onto the line after the second dimension of electrophoresis (left panel). **B.** Actual figures of electrophoretic separation of replication intermediates. Left panel - *Hind*III digestion together with the hypothetical structure of an underreplicated stalled intermediate, right panel - *Eco*R1 digestion. Red arrows point to replication stop sites.

arc, while upstream cleavage shifts the bulge from the bubble-arc onto the horizontal line.

Fig. 11B shows a characteristic example of such mapping for the $(G)_{32} \cdot (C)_{32}$ repeat (Krasilnikova, et. al., 1998). One can see that cleavage of the replication intermediates downstream from the repeat leaves the bulge on the bubble arc. By contrast, cleavage upstream of the repeat shifts the bulge away from the bubble arc. Thus, the replication fork is indeed stalled within the $(G)_{32} \cdot (C)_{32}$ stretch. Note, however, that after cleavage upstream of the repeat, the bulge does not co-migrate with the horizontal line, but migrates to a point in between the bubble-arc and the horizontal line. Thus, the shape of this intermediate is less compact than the y-shape but more compact than the bubble. To explain this migration pattern, one must assume that a portion of the lagging strand around the *Hind*III site in stalled replication intermediates was not yet synthesized. This will lead to an incomplete *Hind*III digestion and the appearance of butterfly-like DNA molecules (shown in the diagram). If this assumption is correct, we detect the underreplication of the lagging strand within the $d(G)_n \cdot d(C)_n$ sequences.

Different DNA repeats mentioned above gave phenomenologically similar results in the electrophoretic analysis of the replication intermediates: (i) they caused replication blockage; (ii) the efficiency of replication blockage increased with repeat length; and (iii) the lagging strand at the repeated DNA segment was underreplicated. We have found, however, that there are at least three different mechanisms responsible for the replication fork blockage by different repeats.

The first mechanism applies to the expandable trinucleotide repeats such as $(CGG)_n \cdot (CCG)_n$, $(CTG)_n \cdot (CAG)_n$ (Samadashwily, et. al., 1997). These repeats attracted very broad attention, since more than a dozen human neurological disorders were attributed to their length expansion (reviewed in Ashley and Warren, 1995; McMurray, 1995; Wells, 1996). Trinucleotide repeats expand with a length-dependent probability. In normal individuals carrying 5-to-30 repeats, expansion is highly unlikely. Individuals with repeat numbers exceeding a threshold of $n \approx 30$ can transmit expanded repeats to their progeny. In the following generations, expansions become more frequent, and each subsequent expansion has a higher probability than the previous one. The latter phenomenon is likely to account for the anticipation in the inheritance of these disorders (Caskey, et. al., 1992; Bates and Lehrach, 1994)

The length dependence of repeat expansion suggests the involvement of an unusual DNA secondary structure(s) (Cox and Mirkin, 1997). Supporting this, it was demonstrated that these repeats in a single-stranded state fold into imperfect hairpins stabilized by both WC and non-WC base pairs (Chen, et. al., 1995; Gacy, et. al., 1995; Yu, et. al., 1995; Petruska, et. al., 1996; Zheng, et. al., 1996; Yu, et. al., 1997). Moreover, the threshold length for expansion is similar to the threshold energy of hairpin formation (Gacy, et. al., 1995).

The mechanisms of repeat expansion remain unknown, but most data implicate replication in this process. Trinucleotide repeats stall *in vitro* DNA polymerization (Kang, et. al., 1995; Usdin and Woodford, 1995; Ohshima and Wells, 1997). This blockage can facilitate a misalignment between the newly synthesized and the template DNA strand (Ohshima and Wells, 1997), potentially leading to expansion. *In vivo*, expansion of different trinucleotide repeats occurs preferentially on their 3'-ends, implying that it could be due to the miscoordination between the leading and lagging strand synthesis (Jodice, et. al., 1994; Kunst and Warren, 1994; Snow, et. al., 1994). This hypothesis is additionally supported by observations that in bacterial and yeast models, the equilibrium between the repeats' expansions or contractions depends on their positioning with regard to the replication origin (Kang, et. al., 1995; Kang, et. al., 1996; Shimizu, et. al., 1996; Freudenreich, et. al., 1997).

To obtain direct data on trinucleotide repeats replication *in vivo*, we analyzed the replication fork movement through these repeats within bacterial plasmids using the electrophoretic approach discussed above (Samadashwily, et. al., 1997). We found that $(CGG)_n \cdot (CCG)_n$ and $(CTG)_n \cdot (CAG)_n$ repeats blocked replication fork progression, and the efficiency of blockage increased with the repeat length so that the length responsible for signifi-

Fig. 12. Quantitative analysis of the replication stop strength (RSS). **Purple squares:** RSS for plasmids carrying the d(CGG)_n-insert in the lagging strand template; **red squares:** RSS for plasmids carrying the d(CCG)_n-insert in the lagging strand template.

Fig. 13. Model of replication blockage caused by trinucleotide repeats. The structure prone strand of a repeat is depicted by a red line, while its complementary strand is shown by a green line. The purple lines show the neighboring DNA. Arrows show the 3'-ends of the growing DNA strands. When the structure-prone DNA sequence is in the lagging strand template, it can form a hairpin-like structure which might prevent the lagging strand synthesis. Since synthesis of both DNA strands is coordinated during DNA replication, this results in stalling of the whole replication fork.

cant replication stalling (≈ 5 -fold) was similar to the threshold length for expansion (**Fig. 12**). The inhibitory effect didn't depend on whether the repeated segment was situated in the transcribed or non-transcribed part of the

Fig. 14. Model for replication blockage caused by transcription through d(G)_n·d(C)_n repeats. Stalled RNA polymerase is shown by an orange oval. The (G)_n stretches in DNA and RNA chains are depicted as red lines, while the d(C)_n stretch is depicted as green line. DNA adjacent to the repeat is shown in purple. The RNA chain is depicted by a gray line except for the r(G)_n stretch, shown in red. Arrows show the 3'-ends of the newly synthesized DNA and RNA chains. Transcriptional stall is believed to be caused by the formation of a stable complex between the G-rich RNA chain and its DNA template. The exact structure of this three-stranded complex remains to be established. The replication fork stops upon encountering the stalled transcription complex.

plasmid. However, it depended on the repeat's orientation relative to the replication origin. Specifically, when structure-prone strands of the repeated DNA, such as (CGG)_n or (CTG)_n, were in the lagging strand template, the replication blockage was the most prominent. We believe, therefore, that the unusual structure of repeated DNA in the lagging strand template is responsible for the replication blockage (**Fig. 13**).

The second mechanisms appears to be responsible for the replication blockage caused by the d(G)_n·d(C)_n repeat (Krasilnikova, et. al., 1998). Similarly to the (CGG)_n·(CCG)_n repeats, d(G)_n·d(C)_n blocks replication in a length dependent manner, except much stronger: for n=30 replication is slowed down ≈ 30 -fold. Unlike the trinucleotide repeats, however, replication blockage relied exclusively on the repeats' transcription so that when the

$d(C)_n$ sequence served as the transcriptional template the replication was severely impaired.

This led us to study transcription through $d(G)_n \cdot d(C)_n$ repeat *in vivo*. We found that when the $d(C)_n$ sequence served as the transcriptional template, transcription was stalled, as was detected by the accumulation of a truncated transcript. This truncated transcript contained an oligo(G) stretch. We conclude, therefore, that transcription is stalled within or immediately after the $d(G)_n \cdot d(C)_n$ repeat, and this is likely caused by the formation of a multistranded complex between the G-rich transcript and its DNA template (Fig. 14). The replication fork, in turn, can not progress through this stalled ternary complex of the RNA polymerase, the DNA template, and the $r(G)_n$ transcript (Fig. 14).

The third mechanism applies to the $d(G-A)_n \cdot d(T-C)_n$ repeat. In this case, there is also a length-dependent replication blockage. However, it depends neither on repeat's transcription, nor on its orientation relative to the replication origin (Krasilnikova et al., unpublished results).

In order to clarify this situation, we studied the replication of this repeat in bacterial cells treated with chloramphenicol. This protein synthesis inhibitor is necessary for bacterial chromosome replication but is

dispensable for the replication of ColE1-type plasmids. Thus, in the presence of chloramphenicol, plasmid DNA becomes profoundly amplified, while the protein content of the cell remains at best stagnant (Clewell, 1972). We found that replication stops in plasmids containing $d(G-A)_n \cdot d(T-C)_n$ inserts completely disappeared under chloramphenicol treatment (Krasilnikova et al., unpublished results). It is plausible to speculate that the inhibitory effect of this repeat on replication is due to a protein binding to this repeat. The length-dependence of the $d(G-A)_n \cdot d(T-C)_n$ -caused replication blockage could be explained by assuming cooperative protein binding to the repeated DNA (Fig. 15).

This situation is markedly different from the trinucleotide repeats and the $d(G)_n \cdot d(C)_n$ repeat. In the latter cases, chloramphenicol treatment does not abolish replication blockage but rather enhances it.

IV. Conclusions

The fact that very different types of simple DNA repeats impede the replication fork progression *in vivo*

Fig. 15. Model for replication blockage caused by transcription through $d(G-A)_n \cdot d(T-C)_n$ repeats. DNA strands of the repeated segment are depicted by red and green lines. Purple lines show surrounding DNA. Gray ovals show cooperatively repeat-bound protein molecules. The replication fork stalls upon encountering a protein/DNA complex.

might have profound biological implications. First, and most important, this may explain the remarkable length polymorphism observed for simple DNA repeats in genomic DNAs. Indeed, to bypass a roadblock involving a DNA repeat, the replication fork could either jump past it

(which might cause contractions or deletions) or pull back and try again (which might cause expansions). Recently, several groups have suggested models detailing the above explanation for repeat expansions and contractions (Kang, et. al., 1995; McMurray, 1995; Gordenin, et. al., 1997;

Tishkoff, et. al., 1997). Note, however, that the current knowledge on the mechanisms of repeat replication is insufficient to choose between those models.

Second, the interplay between transcription and replication blockage may play a role in several biological processes. In notoriously long eukaryotic genes, the collision of the replication and transcription machinery is almost inevitable. Our proposed mechanism might prevent the replication of genes that undergo active transcription. Another provocative opportunity is that stalling of the replication fork caused by transcription of repeated DNA generates DNA ends that are potentially highly recombinogenic. This may contribute to the well-documented stimulation of genetic recombination by transcription (reviewed in Gangloff, et. al., 1994), as well as to the recombinational hotspot activity of some DNA repeats (Schon, et. al., 1989; Wahls, et. al., 1990; Weiller, et. al., 1991; Sumegi, et. al., 1997; Boan, et. al., 1998)

Future studies will undoubtedly contribute to a better understanding of both the replication of simple DNA repeats and the consequences of their replication peculiarities for different processes of DNA metabolism.

Acknowledgments

We thank the current and former members of our lab Andrey Dayn, Randal Cox, Andrey Krasilnikov and Gordana Raca for their invaluable contribution for studying the effects of simple sequence repeats on DNA replication and many helpful discussions, and Randal Cox for critical reading of this manuscript. Supported by grants from the National Institutes of Health (GM54247), the National Science Foundation (MCB-9723924) and the Council for Tobacco Research (CTR-4468) to S.M.M. M.M.K. was in part supported by the Office of International Affairs of the National Cancer Institute.

References

- Agazie, Y. M., Burkholder, G. D., and Lee, J. S. (1996). Triplex DNA in the nucleus: direct binding of triplex-specific antibodies and their effect on transcription, replication and cell growth. **Biochem. J.** 316, 461-466.
- Ahmed, S., Kintanar, A., and Henderson, E. (1994). Human telomeric C-strand tetraplexes. **Nature Struct. Biol.** 1, 83-88.
- Ashley, C., Jr., and Warren, S. T. (1995). Trinucleotide repeat expansion and human disease. **Annu. Rev. Genet.** 29, 703-728.
- Baran, N., Lapidot, A., and Manor, H. (1991). Formation of DNA triplexes accounts for arrests of DNA synthesis at d(TC)_n and d(GA)_n tracts. **Proc. Natl. Acad. Sci. USA** 88, 507-511.
- Bates, G., and Lehrach, H. (1994). Trinucleotide repeat expansions and human genetic disease. **Bioessays** 16, 277-284.
- Beal, P. A., and Dervan, P. B. (1991). Second structural motif for recognition of DNA by oligonucleotide-directed triple-helix formation. **Science** 251, 1360-1363.
- Bedinger, P., Munn, M., and Alberts, B. M. (1989). Sequence-specific pausing during in vitro DNA replication on double stranded DNA templates. **J. Biol. Chem.** 264, 16880-16886.
- Beltran, R., Martinez-Balbas, A., Bernues, J., Bowater, R., and Azorin, F. (1993). Characterization of the zinc-induced structural transition to *H-DNA at a d(GA*CT)₂₂ sequence. **J. Mol. Biol.** 230, 966-978.
- Benham, C. J. (1982). Stable cruciform formation at inverted repeat sequences in supercoiled DNA. **Biopolymers** 21, 679-696.
- Bernues, J., Beltran, R., Casanovas, J. M., and Azorin, F. (1990). DNA-sequence and metal-ion specificity of the formation of *H-DNA. **Nucleic Acids Res.** 18, 4067-4073.
- Boan, F., Rodriguez, J. M., and Gomez-Marquez, J. (1998). A non-hypervariable human minisatellite strongly stimulates in vitro intramolecular homologous recombination. **J. Mol. Biol.** 278, 499-505.
- Brewer, B. J., and Fangman, W. L. (1987). The localization of replication origins on ARS plasmids in *S. cerevisiae*. **Cell** 51, 463-471.
- Brinton, B. T., Caddle, M. S., and Heintz, N. H. (1991). Position and orientation-dependent effects of a eukaryotic Z-triplex DNA motif on episomal DNA replication in COS-7 cells. **J. Biol. Chem.** 266, 5153-5161.
- Caskey, C. T., Pizzuti, A., Fu, Y.-H., Fenwick Jr., R. G., and Nelson, D.L. (1992). Triplet repeat mutations in human disease. **Science** 256, 784-789.
- Chalberg, M. D., and Englund, P. T. (1979). The effect of template secondary structure on vaccinia DNA polymerase. **J. Biol. Chem.** 254, 7820-7826.
- Chen, X., Mariappan, S. V., Catasti, P., Ratliff, R., Moyzis, R. K., Laayoun, A., Smith, S. S., Bradbury, E. M., and Gupta, G. (1995). Hairpins are formed by the single DNA strands of the fragile X triplet repeats: structure and biological implications. **Proc. Natl. Acad. Sci. USA** 92, 5199-5203.
- Chen, X., Mariappan, S. V., Moyzis, R. K., Bradbury, E. M., and Gupta, G. (1998). Hairpin induced slippage and hypermethylation of the fragile X DNA triplets. **J. Biomol. Struct. Dyn.** 15, 745-756.
- Clewell, D. B. (1972). Nature of Col E1 plasmid replication in *Escherichia coli* in the presence of chloramphenicol. **J. Bacteriol.** 110, 667-676.
- Cox, R., and Mirkin, S. M. (1997). Characteristic enrichment of DNA repeats in different genomes. **Proc. Natl. Acad. Sci. USA** 94, 5237-5242.
- Dayn, A., Malkhosyan, S., Duzhy, D., Lyamichev, V., Panchenko, Y., and Mirkin, S. (1991). Formation of (dA-dT)_n cruciforms in *Escherichia coli* cells under different environmental conditions. **J. Bacteriol.** 173, 2658-2664.
- Dayn, A., Malkhosyan, S., and Mirkin, S. M. (1992). Transcriptionally driven cruciform formation in vivo. **Nucleic Acids Res.** 20, 5991-5997.
- Dayn, A., Samadashwily, G. M., and Mirkin, S. M. (1992). Intramolecular DNA triplexes: unusual sequence requirements and influence on DNA polymerization. **Proc.**

- Natl. Acad. Sci. USA** 89, 11406-11410.
- Freudenreich, C. H., Stavenhagen, J. B., and Zakian, V. A. (1997). Stability of a CTG/CAG trinucleotide repeat in yeast is dependent on its orientation in the genome. **Mol. Cell. Biol.** 17, 2090-2098.
- Friedman, K. L., and Brewer, B. J. (1995). Analysis of replication intermediates by two-dimensional agarose gel electrophoresis. **Meth. Enzymol.** 262, 613-627.
- Gacy, A. M., Goellner, G., Juranic, N., Macura, S., and McMurray, C. T. (1995). Trinucleotide repeats that expand in human disease form hairpin structures in vitro. **Cell** 81, 533-540.
- Galas, D. J., Eggert, M., and Waterman, M. S. (1985). Rigorous pattern-recognition methods for DNA sequences. Analysis of promoter sequences from Escherichia coli. **J. Mol. Biol.** 186, 117-128.
- Gangloff, S., Lieber, M. R., and Rothstein, R. (1994). Transcription, topoisomerases and recombination. **Experientia** 50, 261-269.
- Gellert, M., Lipsett, M. N., and Davies, D. R. (1962). Helix formation by guanylic acid. **Proc. Natl. Acad. Sci. USA** 48, 2013-2018.
- Gordenin, D. A., Kunkel, T. A., and Resnick, M. A. (1997). Repeat expansion - all in a flap? **Nature Genet.** 16, 116-118.
- Grabczyk, E., and Fishman, M. C. (1995). A long purine-pyrimidine homopolymer acts as a transcriptional diode. **J. Biol. Chem.** 270, 1791-1797.
- Han, J., Hsu, C., Zhu, Z., Longshore, J. W., and Finley, W. H. (1994). Over-representation of the disease associated (CAG) and (CGG) repeats in the human genome. **Nucleic Acids Res.** 22, 1735-1740.
- Haniford, D. B., and Pulleyblank, D. E. (1983). Facile transition of poly[d(TG)*d(CA)] into a left-handed helix in physiological conditions. **J. Biomol. Struct. Dynam.** 1, 593-609.
- Hoogsteen, K. (1963). The crystal and molecular structure of a hydrogen-bonded complex between 1 methylthymine and 9 methyladenine. **Acta Cryst.** 16, 907-916.
- Huang, C. C., and Hearst, J. E. (1980). Pauses at positions of secondary structure during in vitro replication of single-stranded fd bacteriophage DNA by T4 DNA polymerase. **Anal. Biochem.** 103, 127-139.
- Huberman, J. A., Spotila, L. D., Nawotka, K. A., El-Assouli, S. M., and Davis, L. R. (1987). The in vivo replication origin of the yeast 2 microns plasmid. **Cell** 51, 473-481.
- Jaworski, A., Blaho, J. A., Larson, J. E., Shimizu, M., and Wells, R. D. (1989). Tetracycline promoter mutations decrease non-B DNA structural transitions, negative linking differences and deletions in recombinant plasmids in Escherichia coli. **J. Mol. Biol.** 207, 513-526.
- Jodice, C., Malaspina, P., Persichetti, F., Novelletto, A., Spadaro, M., Giunti, P., Morocutti, C., Terrenato, L., Harding, A. E., and Frontali, M. (1994). Effect of trinucleotide repeat length and parental sex on phenotypic variation in spinocerebellar ataxia I. **Am. J. Hum. Genet.** 54, 959-965.
- Kaguni, L. S., and Clayton, D. A. (1982). Template-directed pausing in in vitro DNA synthesis by DNA polymerase α from Drosophila melanogaster embryos. **Proc. Natl. Acad. Sci. USA** 79, 983-987.
- Kang, S., Jaworski, A., Ohshima, K., and Wells, R. D. (1995). Expansion and deletion of CTG repeats from human disease genes are determined by the direction of replication in E.coli. **Nature Genet.** 10, 213-218.
- Kang, S., Ohshima, K., Jaworski, A., and Wells, R. D. (1996). CTG triplet repeats from the myotonic dystrophy gene are expanded in Escherichia coli distal to the replication origin as a single large event. **J. Mol. Biol.** 258, 543-547.
- Karlin, S. (1986). Significant potential secondary structures in the Epstein-Barr virus genome. **Proc. Natl. Acad. Sci. USA** 83, 6915-6919.
- Karlin, S., and Burge, C. (1995). Dinucleotide relative abundance extremes: a genomic signature. **Trends Genet.** 11, 283-290.
- Karlin, S., and Burge, C. (1996). Trinucleotide repeats and long homopeptides in genes and proteins associated with nervous system disease and development. **Proc. Natl. Acad. Sci. USA** 93, 1560-1565.
- Kiyama, R., and Oishi, M. (1996). In vitro transcription of a poly(dA)*poly(dT)-containing sequence is inhibited by interaction between the template and its transcripts. **Nucleic Acids Res.** 24, 4577-4583.
- Kohwi, Y. (1989). Cationic metal-specific structures adopted by the poly(dG) region and the direct repeats in the chicken adult β A globin gene promoter. **Nucleic Acids Res.** 17, 4493-4502.
- Kohwi, Y., and Kohwi-Shigematsu, T. (1988). Magnesium ion-dependent triple-helix structure formed by homopurine-homopyrimidine sequences in supercoiled plasmid DNA. **Proc. Natl. Acad. Sci. USA** 85, 3781-3785.
- Kohwi, Y., Malkhosyan, S. R., and Kohwi-Shigematsu, T. (1992). Intramolecular dG*dGdC triplex detected in Escherichia coli cells. **J. Mol. Biol.** 223, 817-822.
- Kohwi, Y., and Panchenko, Y. (1993). Transcription-dependent recombination induced by triple-helix formation. **Genes Dev.** 7, 1766-1778.
- Kornberg, A., and Baker, T. (1992). DNA Replication - 2nd. ed. (New York: W. H. Freeman and Co).
- Krasilnikov, A. S., Panyutin, I. G., Samadashwily, G. M., Cox, R., Lazurkin, Y. S., and Mirkin, S. M. (1997). Mechanisms of triplex-caused polymerization arrest. **Nucleic Acids Res.** 25, 1339-1346.
- Krasilnikova, M. M., Samadashwily, G. M., Krasilnikov, A. S., and Mirkin, S. M. (1998). Transcription through a simple DNA repeat blocks replication elongation. **EMBO J.** 17, 5095-5102.
- Kunst, C. B., and Warren, S. T. (1994). Cryptic and polar variation of the fragile X repeat could result in predisposing normal alleles. **Cell** 77, 853-861.
- Lagercrantz, U., Ellegren, H., and Andersson, L. (1993). The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. **Nucleic Acids Res.** 21, 1111-1115.
- Lapidot, A., Baran, N., and Manor, H. (1989). (dT-dC)_n and

- (dG-dA)_n tracts arrest single stranded DNA replication in vitro. **Nucleic Acids Res.** 17, 883-900.
- Lilley, D. M. (1980). The inverted repeat as a recognizable structural feature in supercoiled DNA molecules. **Proc. Natl. Acad. Sci. USA** 77, 6468-6472.
- Little, R. D., Platt, T. H., and Schildkraut, C. L. (1993). Initiation and termination of DNA replication in human rRNA. **Mol. Cell. Biol.** 13, 6600-6613.
- Lyamichev, V. I., Mirkin, S. M., and Frank-Kamenetskii, M. D. (1985). A pH-dependent structural transition in the homopurine-homopyrimidine tract in superhelical DNA. **J. Biomol. Struct. Dyn.** 3, 327-338.
- Lyamichev, V. I., Mirkin, S. M., and Frank-Kamenetskii, M. D. (1986). Structures of homopurine-homopyrimidine tract in superhelical DNA. **J. Biomol. Struct. Dyn.** 3, 667-669.
- MacAllister, T., Khatri, G. S., and Bastia, D. (1990). Sequence-specific and polarized replication termination in vitro: complementation of extracts of *tus*-*Escherichia coli* by purified Ter protein and analysis of termination intermediates. **Proc. Natl. Acad. Sci. USA** 87, 2828-2832.
- Malkov, V. A., Voloshin, O. N., Soyfer, V. N., and Frank-Kamenetskii, M. D. (1993). Cation and sequence effects on stability of intermolecular pyrimidine-purine-purine triplex. **Nucleic Acids Res.** 21, 585-591.
- Manor, H., Sridhara-Rao, B., and Martin, R. G. (1988). Abundance and degree of dispersion of genomic d(GA)_nd(TC)_n sequences. **J. Mol. Evol.** 27, 96-101.
- Mariappan, S. V., Silks, L. A. 3., Chen, X., Springer, P. A., Wu, R., Moyzis, R. K., Bradbury, E. M., Garcia, A. E., and Gupta, G. (1998). Solution structures of the Huntington's disease DNA triplets. (CAG)_n. **J. Biomol. Struct. Dyn.** 15, 723-744.
- Martinez-Balbas, A., and Azorin, F. (1993). The effect of zinc on the secondary structure of d(GA⁺TC)_n DNA sequences of different length: a model for the formation *H-DNA. **Nucleic Acids Res.** 21, 2557-2562.
- McClellan, J. A., Boublikova, P., Palecek, E., and Lilley, D. M. J. (1990). Superhelical torsion in cellular DNA responds directly to environmental and genetic factors. **Proc. Natl. Acad. Sci. USA** 87, 8373-8377.
- McMurray, C. T. (1995). Mechanisms of DNA expansion. **Chromosoma** 104, 2-13.
- Mikhailov, V. S., and Bogenhagen, D. F. (1996). Termination within oligo(dT) tracts in template DNA by DNA polymerase gamma occurs with formation of a DNA triplex structure and is relieved by mitochondrial single-stranded DNA-binding protein. **J. Biol. Chem.** 271, 30774-30780.
- Mirkin, S. M., and Frank-Kamenetskii, M. D. (1994). H-DNA and related structures. **Annu. Rev. Biophys. Biomol. Struct.** 23, 541-576.
- Mirkin, S. M., Lyamichev, V. I., Drushlyak, K. N., Dobrynin, V. N., Filippov, S. A., and Frank-Kamenetskii, M. D. (1987). DNA H form requires a homopurine-homopyrimidine mirror repeat. **Nature** 330, 495-497.
- Mitsui, Y., Langridge, R., Grant, R. C., Kodama, M., Wells, R. D., Shortle, B. E., and Cantor, C. R. (1970). Physical and enzymatic studies ion poly(dI-dC)•poly(dI-dC), an unusual double helical DNA. **Nature** 228, 1166-1169.
- Mizuuchi, K., Mizuuchi, M., and Gellert, M. (1982). Cruciform structures in palindromic DNA are favored by DNA supercoiling. **J. Mol. Biol.** 156, 229-243.
- Morris, J., Kushner, S. R., and Ivarie, R. (1986). The simple repeat poly(dT-dG)•poly(dC-dA) common to eukaryotes is absent from eubacteria and rare in protozoans. **Mol. Biol. Evol.** 3, 343-355.
- Murchie, A. I., and Lilley, D. M. (1994). Tetraplex folding of telomere sequences and the inclusion of adenine bases. **EMBO J.** 13, 993-1001.
- Ohshima, K., and Wells, R. D. (1997). Hairpin formation during DNA synthesis primer realignment in vitro in triplet repeat sequences from human hereditary disease genes. **J. Biol. Chem.** 272, 16798-16806.
- Panayotatos, N., and Wells, R. D. (1981). Cruciform structures in supercoiled DNA. **Nature** 289, 466-470.
- Pearson, C. E., Eichler, E. E., Lorenzetti, D., Kramer, S. F., Zoghbi, H. Y., Nelson, D. L., and Sinden, R. R. (1998). Interruptions in the triplet repeats of SCA1 and FRAXA reduce the propensity and complexity of slipped strand DNA (S-DNA) formation. **Biochemistry** 37, 2701-2708.
- Pearson, C. E., and Sinden, R. R. (1996). Alternative structures in duplex DNA formed within the trinucleotide repeats of the myotonic dystrophy and fragile X loci. **Biochemistry** 35, 5041-5053.
- Pearson, C. E., and Sinden, R. R. (1998). Trinucleotide repeat DNA structures: dynamic mutations from dynamic DNA. **Curr. Opin. Struct. Biol.** 8, 321-330.
- Pearson, C. E., Wang, Y. H., Griffith, J. D., and Sinden, R. R. (1998). Structural analysis of slipped-strand DNA (S-DNA) formed in (CTG)_n•(CAG)_n repeats from the myotonic dystrophy locus. **Nucleic Acids Res.** 26, 816-823.
- Peck, L. J., Nordheim, A., Rich, A., and Wang, J. C. (1982). Flipping of cloned d(pC-pG)_n•d(pG-pC)_n DNA sequences from right- to left handed helical structure by salt, Co(III), or negative supercoiling. **Proc. Natl. Acad. Sci. USA** 79, 4560-4564.
- Petruska, J., Arnheim, N., and Goodman, M. F. (1996). Stability of intrastrand hairpin structures formed by the CAG/CTG class of DNA triplet repeats associated with neurological diseases. **Nucleic Acids Res.** 24, 1992-1998.
- Pevzner, P. A., Borodovsky, M., and Mironov, A. A. (1989). Linguistics of nucleotide sequences. I: The significance of deviations from mean statistical characteristics and prediction of the frequencies of occurrence of words. **J. Biomol. Struct. Dyn.** 6, 1013-1026.
- Pevzner, P. A., Borodovsky, M., and Mironov, A. A. (1989). Linguistics of nucleotide sequences. II: Stationary words in genetic texts and the zonal structure of DNA. **J. Biomol. Struct. Dyn.** 6, 1027-1038.
- Pinnavaia, T. J., Marshall, C. L., Metterl, C. M., Fisk, C. L., Miles, T., and Becker, E. D. (1978). Alkali metal ion specificity in the solution ordering of a nucleotide, 5'-guanosine monophosphate. **J. Am. Chem. Soc.** 100, 3625-3627.
- Pohl, F. M., and Jovin, T. M. (1972). Salt induced co-operative

- conformational changes of a synthetic DNA: Equilibrium and kinetic studies with poly (dG-dC). **J. Mol. Biol.** 67, 375-396.
- Raghavan, S., Burma, P. K., and Brahmachari, S. K. (1997). Positional preferences of polypurine/polypyrimidine tracts in *Saccharomyces cerevisiae* genome: implications for cis regulation of gene expression. **J. Mol. Evol.** 45, 485-498.
- Rahmouni, A. R., and Wells, R. D. (1992). Direct evidence for the effect of transcription on local DNA supercoiling in vivo. **J. Mol. Biol.** 223, 131-144.
- Rao, B. S. (1994). Pausing of simian virus 40 DNA replication fork movement in vivo by (dG-dA)_n(dT-dC)_n tracts. **Gene** 140, 233-237.
- Rao, S., Manor, H., and Martin, R. G. (1988). Pausing in simian virus 40 DNA replication by a sequence containing (dG-dA)₂₇(dT-dC)₂₇. **Nucleic Acids Res.** 16, 8077-8094.
- Reaban, M. E., and Griffin, J. A. (1990). Induction of RNA-stabilized DNA conformers by transcription of an immunoglobulin switch region. **Nature** 348, 342-344.
- Reaban, M. E., Lebowitz, J., and Griffin, J. A. (1994). Transcription induces the formation of a stable RNA-DNA hybrid in the immunoglobulin α switch region. **J. Biol. Chem.** 269, 21850-21857.
- Rich, A., Nordheim, A., and Wang, A. H. (1984). The chemistry and biology of left-handed Z-DNA. **Annu. Rev. Biochem.** 53, 791-846.
- Samadashwily, G. M., Dayn, A., and Mirkin, S. M. (1993). Suicidal nucleotide sequences for DNA polymerization. **EMBO J.** 12, 4975-4983.
- Samadashwily, G. M., and Mirkin, S. M. (1994). Trapping DNA polymerases using triplex-forming oligodeoxyribonucleotides. **Gene** 149, 127-136.
- Samadashwily, G. M., Raca, G., and Mirkin, S. M. (1997). Trinucleotide repeats affect DNA replication in vivo. **Nature Genet.** 17, 298-304.
- Saunders, N. J., Peden, J. F., Hood, D. W., and Moxon, E. R. (1998). Simple sequence repeats in the *Helicobacter pylori* genome. **Mol. Microbiol.** 27, 1091-1098.
- Schon, E. A., Rizzuto, R., Moraes, C. T., Nakase, H., Zeviani, M., and DiMauro, S. (1989). A direct repeat is a hotspot for large-scale deletion of human mitochondrial DNA. **Science** 244, 346-349.
- Schroth, G. P., and Ho, P. S. (1995). Occurrence of potential cruciform and H-DNA forming sequences in genomic DNA. **Nucleic Acids Res.** 23, 1977-1983.
- Sen, D., and Gilbert, W. (1988). Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. **Nature** 334, 364-366.
- Sherman, L. A., and Gefter, M. L. (1976). Studies of the mechanism of enzymatic DNA elongation of by *Escherichia coli* DNA polymerase II. **J. Mol. Biol.** 103, 61-76.
- Shimizu, M., Gellibolian, R., Oostra, B. A., and Wells, R. D. (1996). Cloning, characterization and properties of plasmids containing CGG triplet repeats from the FMR-1 gene. **J. Mol. Biol.** 258, 614-626.
- Sinden, R. R. (1994). DNA structure and function. (San Diego: Academic Press, Inc.).
- Singleton, C. K., Klysik, J., Stirdivant, S. M., and Wells, R. D. (1982). Left-handed Z DNA is induced by supercoiling in physiological ionic conditions. **Nature** 299, 312-316.
- Smillie, F., and Bains, W. (1990). Repetition structure of mammalian nuclear DNA. **J. Theor. Biol.** 142, 463-471.
- Snow, K., Tester, D. J., Kruckeberg, K. E., Schaid, D. J., and Thibodeau, S. N. (1994). Sequence analysis of the fragile X trinucleotide repeat: implications for the origin of the fragile X mutation. **Hum. Mol. Genet.** 3, 1543-1551.
- Spiro, C., Bazett-Jones, D. P., Wu, X., and McMurray, C. T. (1995). DNA structure determines protein binding and transcriptional efficiency of the proenkephalin cAMP-responsive enhancer. **J. Biol. Chem.** 270, 27702-27710.
- Sumegi, A., Birko, Z., Szeszak, F., Vitalis, S., and Biro, S. (1997). A short GC-rich sequence involved in deletion formation of cloned DNA in *E. coli*. **Acta Biol. Hung.** 48, 275-279.
- Sundquist, W. I., and Klug, A. (1989). Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. **Nature** 342, 825-829.
- Tishkoff, D. X., Filosi, N., Gaida, G. M., and Kolodner, R. D. (1997). A novel mutation avoidance mechanism dependent on *S. cerevisiae* RAD27 is distinct from DNA mismatch repair. **Cell** 88, 253-263.
- Trifonov, E. N., Konopka, A. K., and Jovin, T. M. (1985). Unusual frequencies of certain alternating purine-pyrimidine runs in natural DNA sequences: relation to Z-DNA. **FEBS Lett.** 185, 197-202.
- Usdin, K., and Woodford, K. J. (1995). CGG repeats associated with DNA instability and chromosome fragility from structures that block DNA synthesis in vitro. **Nucleic Acids Res.** 23, 4202-4209.
- Ussery, D. W., and Sinden, R. R. (1993). Environmental influences on the in vivo level of intramolecular triplex DNA in *Escherichia coli*. **Biochemistry** 32, 6206-6213.
- Vologodskii, A. (1992). Topology and physics of circular DNA. (Boca Raton: CRC Press).
- Vologodskii, A. V., and Frank-Kamenetskii, M. D. (1982). Theoretical study of cruciform states in superhelical DNA. **FEBS Lett.** 143, 257-260.
- Wahls, W. P., Wallace, L. J., and Moore, P. D. (1990). Hypervariable minisatellite DNA is a hotspot for homologous recombination in human cells. **Cell** 60, 95-103.
- Wang, A. H. J., Quigley, G. J., Kolpak, F. J., Crawford, J. L., van Boom, J. H., Van der Marel, G., and Rich, A. (1979). Molecular structure of a left-handed double helical DNA fragment at atomic resolution. **Nature** 282, 680-686.
- Weaver, D. T., and DePamphilis, M. L. (1982). Specific sequences in native DNA that arrest synthesis by DNA polymerase alpha. **J. Biol. Chem.** 257, 2075-2086.
- Weiller, G. F., Bruckner, H., Kim, S. H., Pratje, E., and Schweyen, R. J. (1991). A GC cluster repeat is a hotspot for mit- macro-deletions in yeast mitochondrial DNA. **Mol. Gen. Genet.** 226, 233-240.
- Weitzmann, M. N., Woodford, K. J., and Usdin, K. (1997). DNA secondary structures and the evolution of hypervariable

- tandem arrays. **J. Biol. Chem.** 272, 9517-9523.
- Wells, R. D. (1996). Molecular basis of genetic instability of triplet repeats. **J. Biol. Chem.** 271, 2875-2878.
- Williamson, J. R., Raghuraman, M. K., and Cech, T. R. (1989). Monovalent cation-induced structure of telomeric DNA: the G-quartet model. **Cell** 59, 871-880.
- Wittig, B., Wolfl, S., Dobric, T., Vahrson, W., and Rich, A. (1992). Transcription of human c-myc in permeabilized nuclei is associated with formation of Z-DNA in three discrete regions of the gene. **EMBO J.** 11, 4653-4663.
- Wolfl, S., Martinez, C., Rich, A., and Majzoub, J. A. (1996). Transcription of the human corticotrophin-releasing hormone gene in NPLC cells is correlated with Z-DNA formation. **Proc. Natl. Acad. Sci. USA** 93, 3664-3668.
- Woodford, K. J., Howell, R. M., and Usdin, K. (1994). A novel K(+)-dependent DNA synthesis arrest site in a commonly occurring sequence motif in eukaryotes. **J. Biol. Chem.** 269, 27029-27035.
- Yu, A., Barron, M. D., Romero, R. M., Christy, M., Gold, B., Dai, J., Gray, D. M., Haworth, I. S., and Mitas, M. (1997). At physiological pH, d(CCG)₁₅ forms a hairpin containing protonated cytosines and a distorted helix. **Biochemistry** 36, 3687-3699.
- Yu, A., Dill, J., Wirth, S. S., Huang, G., Lee, V. H., Haworth, I. S., and Mitas, M. (1995). The trinucleotide repeat sequence d(GTC)₁₅ adopts a hairpin conformation. **Nucleic Acids Res.** 23, 2706-2714.
- Zheng, G., Kochel, T., Hoepfner, R. W., Timmons, S. E., and Sinden, R. R. (1991). Torsionally tuned cruciform and Z-DNA probes for measuring unrestrained supercoiling at specific sites in DNA of living cells. **J. Mol. Biol.** 221, 107-129.
- Zheng, M., Huang, X., Smith, G. K., Yang, X., and Gao, X. (1996). Genetically unstable CXG repeats are structurally dynamic and have a high propensity for folding. An NMR and UV spectroscopic study. **J. Mol. Biol.** 264, 323-336.
- Zhu, I. J., Newlon, C. S., and Huberman, J. A. (1992). Localization of a DNA replication origin and termination zone on chromosome III of *Saccharomyces cerevisiae*. **Mol. Cell. Biol.** 12, 4733-4741.
- Zimmerman, S. B., Cohen, G. H., and Davies, D. R. (1975). X-ray fiber diffraction and model-building study of polyguanylic acid and polyinosinic acid. **J. Mol. Biol.** 92, 181-192.