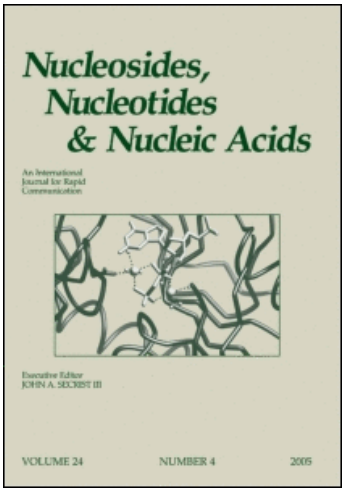

Nucleosides, Nucleotides and Nucleic Acids
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597286

Factors Influencing Dna Expansion in the Course of Polymerase Chain Reaction

Jitka Vondrušková[a]; Nad'a Pařízková[a]; Jaroslav Kypr[a]
[a] Institute of Biophysics, Academy of Sciences of the Czech Republic, Brno, Czech Republic

To cite this Article Vondrušková, Jitka , Pařízková, Nad'a and Kypr, Jaroslav(2007) 'Factors Influencing Dna Expansion in the Course of Polymerase Chain Reaction', Nucleosides, Nucleotides and Nucleic Acids, 26: 1, 65 — 82
To link to this Article: DOI: 10.1080/15257770601052299
URL: http://dx.doi.org/10.1080/15257770601052299

Nucleosides, Nucleotides, and Nucleic Acids, 26:65–82, 2007

ISSN: 1525-7770 print / 1532-2335 online
DOI: 10.1080/15257770601052299

Taylor & Francis
Taylor & Francis Group

FACTORS INFLUENCING DNA EXPANSION IN THE COURSE OF POLYMERASE CHAIN REACTION

Jitka Vondrušková, Nad'a Pařízková, and Jaroslav Kypr □ *Institute of Biophysics, Academy of Sciences of the Czech Republic, Brno, Czech Republic*

□ *We performed more than 3,500 polymerase chain reactions (PCRs) under various conditions with more than 400 DNA fragments of 4–150 nucleotides in length. Some of the PCRs provided expanded DNA molecules of kilobase lengths whereas others led to no expansion. Repetitiveness of the primary structure was mostly found to be necessary but not sufficient for the expansion. (A+T)-rich fragments expand better than (G+C)-rich ones and pyrimidine-rich fragments expand better than purine-rich fragments. Terminal nucleotides and the fragment length also are important for the expansion. Examples are presented when relatively small alterations of the DNA primary structure caused a dramatic change in the expansion. For example, A_8T_8 expanded a lot whereas T_8A_8 did not expand at all. The present work has implications for pathological expansions of microsatellites in the human genome as well as regarding the genome evolution in general.*

Keywords DNA expansion; PCR; Microsatellites

INTRODUCTION

DNA is not an invariant molecule in several respects. For example it undergoes changes in length. These changes are interesting in three respects. Firstly, we can imagine how short DNA pieces could arise in the prebiotic conditions[1] but we are aware of no justified idea of how these short pieces were used to generate even the shortest biological meaningful molecules of DNA that are more than two orders of magnitude larger. Perhaps the only idea in this direction are the Trifonov's circles.[2] Secondly, the genomic molecules of DNA undergo insertions and deletions which change the genomic DNA lengths. However, so called dynamic mutations when microsatellites lose or acquire one or more of its repeats are much more frequent.[3] These changes, for example, may influence expression of neighboring genes[4] and they are advantageously used in population genetics to discriminate among individuals of the same organism.[5] Third,

This project was supported by grant NM/7634–3 from the Ministry of Health of the Czech Republic.
Address correspondence to Jaroslav Kypr, Institute of Biophysics, Academy of Sciences of the Czech Republic, Královopolská 135, CZ-612 65, Brno, Czech Republic. E-mail: kypr@ibp.cz

expansions leading to DNA lengthening by even kilobases, accompany many neurological diseases[6] as well as some kinds of cancer.[7] Thus, there are sufficient reasons to study mechanisms of DNA expansions in great detail.

Short DNA fragments were found to undergo expansions to kilobase lengths even in the course of replication or polymeras chain reaction (PCR).[8–22] However, these studies were performed with small numbers of fragments and under limited conditions of PCR or the replication. Therefore, these studies did not permit arriving at conclusions that generally would allow the prediction of which fragments will expand and which will not. A possibility of predictions in this direction would be of a great value because it would help to understand factors standing behind the appearance of the first long DNA molecules on the Earth, microsatellite length polymorphism in genomes as well as pathological length alterations of microsatellites accompanying many neurological diseases as well as cancer. Here, we performed more than 3500 polymerase chain reactions under various conditions with more than 400 DNA fragments of 4–150 nucleotides in length with various primary structures. Results of this study are analyzed in terms of factors which influence DNA expansion during PCR.

MATERIAL AND METHODS

The oligonucleotide samples used in this work were synthetized and purified in various laboratories and companies. $(GCC)_8$, $(CGC)_8$, $(CCG)_8$, $(CGG)_8$, and $(GGC)_8$ were produced at the Faculty of Science of the Masaryk University in Brno. A_{16}, A_8T_8, T_8A_8, and $T_4(TA)_4A_4$ were provided by Dr. Arnold of the Institute of Organic Chemistry and Biochemistry of the Academy of Science of the Czech Republic in Prague. All other fragments were bought from VBC-Genomics Bioscience Research (Wien, Austria). All of the samples were obtained as lyophylized material and were originally dissolved in 1 mM sodium phosphate, 0.1 mM EDTA, pH 7.

The PCR reaction mixtures were prepared at about 4°C in thin-wall Eppendorf tubes. 2 μg of DNA were mixed with 2 μl 10 × Taq buffer. The 10 × Taq buffer contains 100 mM TrisHCl, 500 mM KCl, 1% Triton X-100, and 15 mM $MgCl_2$, pH 8.8. The 10 × Taq buffer was bought from Top-Bio, Prague. Further, 0.5 ml of the 10 mM triphosphate solution (MBI Fermentas, Ontario, Canada) and 2μl of Taq Purple DNA polymerase (concentration 1U/μl, Top-Bio, Prague, Czech Republic) were added. The polymerase contained an inert stain (Top-Bio). The reaction mixture was complemented by deionized sterilized water (PCR H_2O, Top-Bio) to get a 25 μl total volume of the sample.

The PCR experiments were performed using a Progene (Techne, Cambridge, Great Britain) thermocycler in the following temperature regimes. The first cycle-denaturation 94°C (2 minutes), the annealing temperature

(45 seconds), the extension temperature (45 seconds). The second to the last but one cycles: 94°C (30 seconds), the annealing temperature (45 seconds), 72°C (45 seconds). The last cycle: 94°C (30 seconds), the annealing temperature (45 seconds), 72°C (5 minutes). Then the samples were cooled in the PCR cycler to 4°C at a maximum speed and kept in the refrigerator until electrophoretic analysis. The annealing temperature was variable within 37–72°C in various experiments. The number of PCR cycles was mostly 35 or 20.

Besides the DNA primary structure, the expansion also depended on a number of other factors. One of the most important factors is the number of PCR cycles. Some of the fragments provided kilobase fragments even after few PCR cycles,[22] others required many more cycles to provide detectable amounts of the expanded material. In order to detect even less easily expanding fragments, we mostly used 35 or 20 cycles in the present experiments. PCR is further characterized by two temperatures. The first is the temperature at which the PCR is extended. We tested its influence and found negligible effects. That is why the extension temperature was fixed at 72°C in all present experiments. In contrast, the other temperature, i.e. annealing temperature Ta, was found to influence the results significantly. That is why we performed most of the present experiments at various annealing temperatures within 37–72°C.

The PCR results were furthermore influenced by the reagents contained in the PCR samples. We performed numerous control experiments when either the polymerase or the oligonucleotide was missing in the PCR solution and we had never detected any amplified material as expected. We also tested the effect of the oligonucleotide concentration but detected no qualitative effect within wide limits (0.125–2.0 μg). The concentration of KCl influenced the results. The maximum amplification was observed at 50 mM KCl, which is the concentration recommended by the Taq polymerase producer. Similarly, the optimum effect was observed at 1.5 mM and 2.0 mM $MgCl_2$, which also is the value recommended by the Taq polymerase producer. The value of pH of the PCR buffer had a strong effect on the amplification, being optimum at slightly alkaline values (~8.8) as recommended by the producer again. Hence, we performed all following experiments in the PCR buffer, pH 8,8, containing 50 mM KCl, 1.5 mM $MgCl_2$, and 2.0 μg of the oligonucleotide.

Electrophoresis was run in 1% agarose gels (Serva, Electrophoresis, Heidelberg, Germany) stained with ethidium (Top-Bio). The electrophoresis lasted 1 hour at 80 V and 50 mA. 15 μl of the PCR sample were loaded on the gels. λDNA/HindIII (MBI Fermentas, Ontario, Canada) was used as a marker. The digest contains fragments whose length is 564–23130 bp. Then the DNA was visualized under UV light, the gel was photographed, and the negative was scanned using the Personal Densitometer SI, Model 375 A (Molecular Dynamics, Sunnyvale, CA).

The amplified material was digested by the DNase I, P1, and S1 nucleases. The digestions were performed at small steps of the nuclease concentrations within 0.0025–80 U/ml. The reaction mixture contained 4 μl of the PCR product, the required amount of the nuclease and the samples were complemented by the reaction buffer to the volume of 15 μl. The reaction buffer was specific for each nuclease. The digestion was carried 10 minutes at 37°C. The reaction was stopped by addition of 3.2 μl of 0.1 M EDTA in an ice-bath. The digestion was followed by electrophoresis in a 1% agarose gel and ethidium staining.

RESULTS

Expanding and Nonexpanding Fragments

We performed over 3500 PCR reactions under various conditions with more than 400 DNA fragments of 4–150 nucleotide residues in length. Some of the fragments were found to expand into kilobasepair lengths during PCR whereas others did not expand at all. The expanding fragments included e.g. $(TA)_8$, $(TG)_8$, $(CA)_8$, and $(AAG)_6$ and the nonexpanding fragments included, for example, (CTGTTATGCCTACTT), (ATAGCAGTGCAGCTT), (TCAGTTCCATAGGTT), (AACTACCTACAGAGA), and (GCTTTAGAATGTGGT) (Figure 1). The fragments that did not expand under any conditions tested in this work (see below) are summarized in Table 1. They constitute 14.8% of the tested fragments (61 of 411) and their primary structure at first sight differs from the expanding fragments by its non-repetitive nature. However, not all repetitive DNA expands.

(G+C) Fragments

Figure 1 shows that many repetitive DNA fragments expand in the course of PCR whereas most non-repetitive fragments do not expand. However, not every repetitive fragment expands. (G+C) fragments $(GC)_5$, $(GC)_3$, and $(CG)_3$ (not shown) serve as examples. Many other (G+C) fragments also do not expand. These include (C_6G_6), $C_4(GC)_2G_4$, $C_2(GC)_4G_2$, $G_4(CG)_2C_4$, C_4G_4, and C_7G_7. G_{20} also does not expand. It presumably generates highly stable guanine tetraplexes in the course of PCR. The tetraplexes hinder replication[23] and, hence, the expansion as well. On the other hand, $(C_2G_2)_3$, $(C_3G_3)_2$, $(C_3G_3)_4$, and especially $(C_2G_2)_6$ expand in the course of PCR[22] so that the primary structure and not the very high (G+C) content prohibits the expansion. We also examined $(CCG)_n$ and related oligonucleotides (Figure 2) relevant to the expansion causing fragility of the chromosome X[24]. $(CCG)_8$ provided high amounts of about a 5 kbp fragment even after 20 PCR cycles ($T_a = 60°C$). $(CGC)_8$ expanded to only 0.5 kbp under these conditions whereas $(GCC)_8$ was intermediate (2

TABLE 1 Primary structures of DNA fragments that did not expand even at 35 cycles. The annealing temperatures are given in parentheses

A_{16} (37–72°C)	A_{20} (37–72°C)	G_{20} (37–72°C)
T_8A_8 (37, 50, 60, 72°C)	A_8T_4 (37, 50, 60, 72°C)	$T_4(TA)_4A_4$ (37, 50, 60, 72°C)
$(CG_4)_4$ (37–72°C)	$G_4T_4G_4$ (37–72°C)	$G_4A_4G_4$ (37–72°C)
$(GAA)_3$ (37, 50, 60, 72°C)	$(GAA)_4$ (37, 50, 60, 72°C)	$(GAA)_{20}$ (37, 50, 60, 72°C)
$(GAA)_{25}$ (37, 50, 60, 72°C)	$(GAA)_{40}$ (37, 50, 60, 72°C)	$(GAA)_{50}$ (37, 50, 60, 72°C)
$(GAA)_3G$ (37, 50, 60, 72°C)	$(GAA)_4G$ (37, 50, 60, 72°C)	$(GAA)_3GA$ (37, 50, 60,72°C)
$(GAA)_4GA$ (37, 50, 60, 72°C)	$(AAG)_3$ (37, 50, 60, 72°C)	$(AAG)_4$ (37, 50, 60, 72°C)
$(AAG)_3A$ (37, 50, 60, 72°C)	$(AAG)_4A$ (37, 50, 60, 72°C)	$(AAG)_3AA$ (37, 50, 60, 72°C)
$(AAG)_4AA$ (37, 50, 60, 72°C)	$(AGA)_3$ (37, 50, 60, 72°C)	$(AGA)_4$ (37, 50, 60, 72°C)
$(AGA)_4A$ (37, 50, 60, 72°C)	$(AGA)_4AG$ (37, 50, 60, 72°C)	$(TTC)_3$ (37, 50, 60,72°C)
$(TTC)_{10}$ (37, 50, 60, 72°C)	$(TTC)_3T$ (37, 50, 60, 72°C)	$(TTC)_3TT$ (37, 50, 60, 72°C)
$(TTC)_6TT$ (37, 50, 60, 72°C)	$(TTC)_{10}TT$ (37, 50, 60, 72°C)	$(TCT)_3$ (37, 50, 60, 72°C)
$(TCT)_3T$ (37, 50, 60, 72°C)	$(CTT)_3$ (37, 50, 60, 72°C)	$(CTT)_3CT$ (37, 50, 60, 72°C)
$(CAG)_3$ (37, 50, 60, 72°C)	$(AGC)_3A$ (37, 50, 60, 72°C)	$(AGC)_3AA$ (37, 50, 60, 72°C)
$(GCA)_3$ (37, 50, 60, 72°C)	$(GCA)_3G$ (37, 50, 60, 72°C)	$(GCA)_3GC$ (37, 50, 60, 72°C)
$CGA(GCC)_3$ (37–72°C)	$CGA(CGC)_3$ (37–72°C)	$(CTG)_3$ (37, 50, 60, 72°C)
$(CTG)_3C$ (37, 50, 60, 72°C)	$(CTG)_3CT$ (37, 50, 60, 72°C)	$TGC(AGC)_9AG$ (37, 50, 60, 72°C)
$(CGAA)_2CG$ (37–72°C)	TATAGTAC (37–72°C)	CGAGACG (37–72°C)
TCGATCGA (37–72°C)	$(TTAGGG)_2$ (62, 72°C)	$(ACC)_3CGC_{16}$ (60, 72°C)
TTGCGAAGCAAAACACACACACAGCGAAGC (60°C)		
TTAAGGGGAATGTTAGGGTCAGT (4–99°C)		
CCAGTGGTAAAGCAGACGCT (4–99°C)		
TTCCCCCTAGGAGTGGAATTA (4–99°C)		

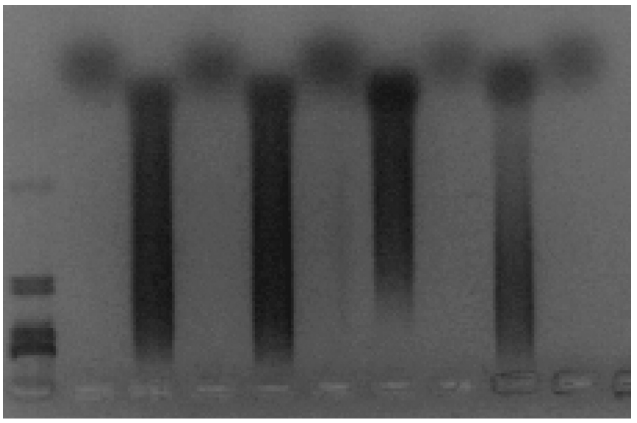

FIGURE 1 Products of PCR reactions of DNA fragments containing repetitive and non-repetitive primary structures. (lane 1) λDNA/Hind III marker, (lane 2) CTGTTATGCCTACTT, (lane 3) $(TA)_8$, (lane 4) ATAGCAGTGCAGCTT, (lane 5) $(TG)_8$, (lane 6) TCAGTTCCATAGGTT, (lane 7) $(CA)_8$, (lane 8) AACTACCTACAGAGA, (lane 9) $(AAG)_6$, (lane 10) GCTTTAGAATGTGGT. All of the PCRs were run at 35 cycles, annealing temperature 60°C.

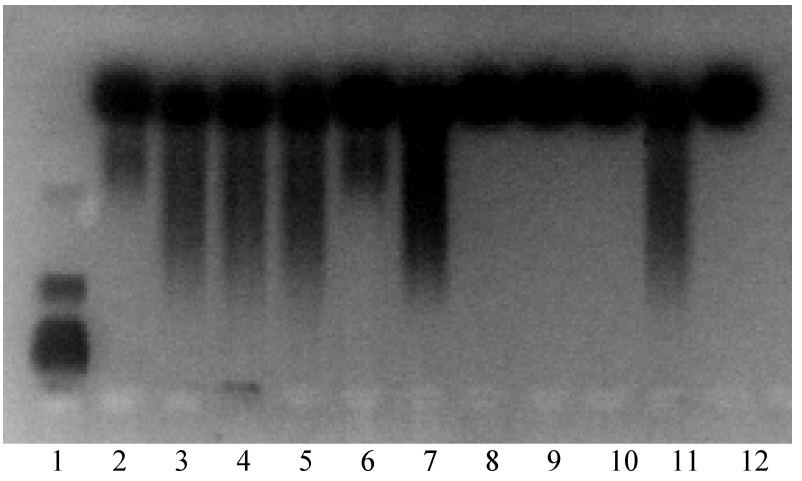

FIGURE 2 Products of PCR reactions of (G+C) fragments of DNA. (lane 1) λDNA/Hind III marker, (lane 2) $(CCG)_4$, (lane 3) $(CGC)_4$, (lane 4) $(GCC)_4$, (lane 5) $(CCG)_8$, (lane 6) $(CGC)_8$, (lane 7) $(GCC)_8$, (lane 8) $(GGC)_8$, (lane 9) $(GCG)_8$, (lane 10) $(CGG)_8$, (lane 11) $(CCCG)_5$, (lane 12) $(CGGG)_5$. 20 PCR cycles, annealing temperature 60°C.

kbp). $(CCG)_4$, $(CGC)_4$, and $(GCC)_4$ provided 0.7, 5, and 4 kbp fragments under the same conditions respectively. $(CGG)_8$, $(GCG)_8$, and $(GGC)_8$ did not expand to indicate that pyrimidine-rich strands expand better than purine-rich strands (see below). The same conclusion was reached with another complementary (G+C) repeat couple, that is, $(CCCG)_5$ and $(CGGG)_5$, where the former 20-mer expanded whereas the latter did not.

(A+T)-Rich Fragments

(A+T)-rich fragments expanded much better than the (G+C)-rich ones. But fragments containing only A or T or (A+T) also behaved various ways (Figure 3). T_{16} expanded dramatically whereas A_{16} and A_{20} did not expand. A_8T_8 expanded but T_8A_8 did not. The selfcomplementary $T_4(TA)_4A_4$ sequence was refractory to the expansion but the fragments containing T_4 tails, that is, $(TA)_8T_4$ and $T_4(TA)_8$ both expanded. T_{14}, T_{18}, T_{20} and T_{22} all expanded. $(TA)_8T_2$ also expanded. A_4T_8 expanded like A_8T_8 but only at some (37, 60°C) but not other (50, 72°C) annealing temperatures. This nonmonotonic temperature dependence seems to result from mechanistic complexity of the expansion, such as competition between sequence annealing and polymerase activity. $(TTG)_7TT$ expanded but much less than $(GTT)_7GT$. These two fragments were chosen for the present experiments because their occurrence in the human genome differs a lot, the former occuring 635 times in the human genome whereas the latter only exhibits 3 occurrences.[25] Their complement, that is, $AA(CAA)_7$, expanded less. $(TA)_{64}$ provided a lot of the expanded material but the

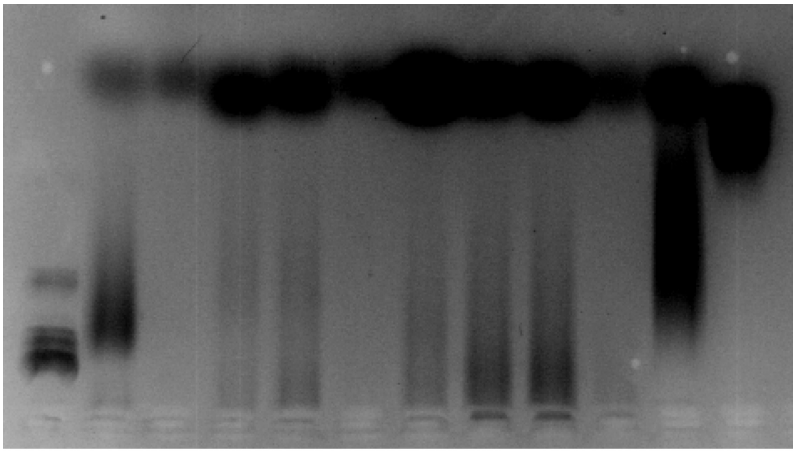

FIGURE 3 PCR products of (A+T) fragments of DNA. (lane 2) T_{16}, (lane 3) A_{16}, (lane 4) A_4T_8, (lane 5) A_8T_8, (lane 6) T_8A_8, (lane 7) $(TA)_8T_2$, (lane 8) $(TA)_8T_4$, (lane 9) $T_4(TA)_8$, (lane 10) $T_4(TA)_4A_4$, (lane 11) $(TA)_8$, (lane 12) $(TA)_{64}$. Marker (lane 1) is the λDNA/Hind III digest. 35 cycles, annealing temperature 60°C.

expanded material length was much shorter than with, for example, $(TA)_8$. It follows from the experiments with the $(TA)_n$ series of fragments that very short fragments, for example, $(TA)_4$ or $(TA)_5$, are not amplified at all, $(TA)_7$ or $(TA)_8$ amplify best, whereas $(TA)_{16}$ or $(TA)_{64}$ amplify much less or not at all. This behavior may indicate that the amplified oligonucleotide conformers should neither be too thermolabile nor too thermostable.

$(CTG)_n$ and $(CAG)_n$ Expansions

$(CTG)_n$ and $(CAG)_n$ repeat expansions accompany many neurological diseases[6] which is why we examined their expansions during PCR. $(CAG)_6$ C provides a continuous smear of long expanded products whereas $(CAG)_6$ and $(CAG)_6CA$ expand less and the products are discrete rather than continuous (Figure 4). The discrete products are also provided by $(AGC)_6$ A whereas $(AGC)_6$ and $(AGC)_6AG$ provide much larger amounts of continuous products. $(GCA)_6$ provides discrete products, $(GCA)_6G$ provides continuous products but $(GCA)_6GC$ does not expand at all (Figure 4). These data indicate that ends of the fragments are important for the expansion. Stability conferred by two strong bases at both ends of the fragments abolishes the expansion entirely (but see below). We also observed interesting effects of the expanding oligonucleotide length on the expansion. For example, $(AGC)_3AG$ and $(AGC)_4AG$ provided rather a small amount of the expanded material whereas $(AGC)_6AG$ and $(AGC)_{10}AG$ expanded much more

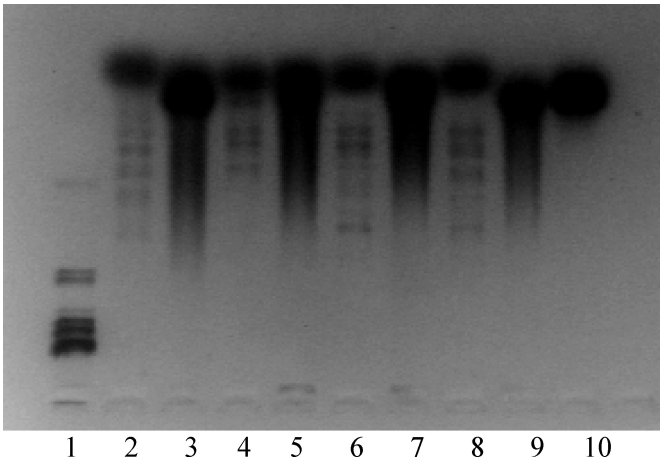

FIGURE 4 PCR (35 cycles, annealing temperature 50°C) products of $(CAG)_n$ fragments variants and permutations differing by the terminal nucleotides. (lane 1) the λDNA/Hind III marker, (lane 2) $(CAG)_6$, (lane 3) $(CAG)_6C$, (lane 4) $(CAG)_6CA$, (lane 5) $(AGC)_6$, (lane 6) $(AGC)_6A$, (lane 7) $(AGC)_6AG$, (lane 8) $(GCA)_6$, (lane 9) $(GCA)_6G$, (lane 10) $(GCA)_6GC$.

(Figure 5). However, molecules of the same length and nucleotide composition expanded differently depending on the primary structure. For example, $(AGC)_{10}AG$ expanded a lot, $(TGC)_{10}TG$ and $(CAG)_{10}CA$ expanded much less and $(GCA)_{10}$ GC did not expand at all. The latter effect seemed to be caused by the presence of two strong bases on both ends of the molecule again, but this was not true because $(AGC)_3$ AA did not expand as well.

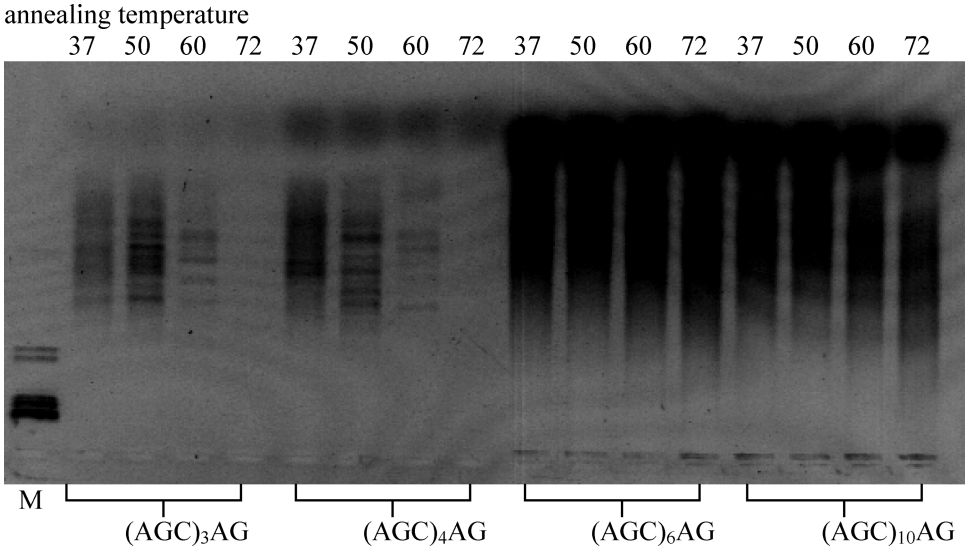

FIGURE 5 PCR (35 cycles) of $(AGC)_nAG$, n = 3, 4, 6, and 10 at the indicated annealing temperatures.

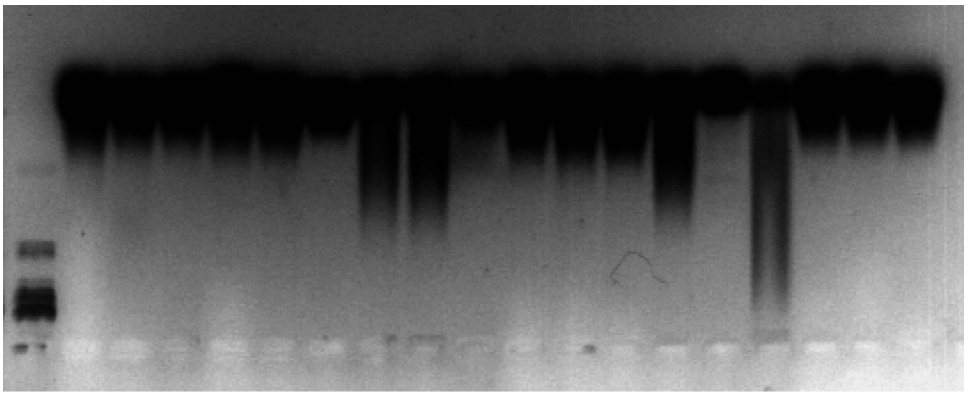

FIGURE 6 PCR (35 cycles, annealing temperature 60°C) products of (lane 2) $(CAG)_{10}$, (lane 3) $(CAG)_{10}C$, (lane 4) $(CAG)_{10}CA$, (lane 5) $(CTG)_{10}$, (lane 6) $(CTG)_{10}C$, (lane 7) $(CTG)_{10}CT$, (lane 8) $(GCA)_{10}$, (lane 9) $(GCA)_{10}G$, (lane 10) $(GCA)_{10}GC$, (lane 11) $(TGC)_{10}$, (lane 12) $(TGC)_{10}T$, (lane 13) $(TGC)_{10}TG$, (lane 14) $(AGC)_{10}$, (lane 15) $(AGC)_{10}A$, (lane 16) $(AGC)_{10}AG$, (lane 17) $(GCT)_{10}$, (lane 18) $(GCT)_{10}G$, (lane 19) $(GCT)_{10}GC$. (lane 1) the λDNA/Hind III marker.

On the other hand, $(AGC)_3AC$ did expand whereas $(AGC)_3AT$ displayed a much smaller tendency for the expansion. Finally, we studied 32-mers differing by only the 5'end nucleotide and found that $TGC(AGC)_9AG$ did not expand at all whereas $GGC(AGC)_9AG$ and $CGC(AGC)_9AG$ provided small amounts of expanded discrete length molecules.

$(CAG)_n$, n = 17, 20, 25, 30, all expanded and the longest molecules were provided by $(CAG)_{25}$. Analogous $(CTG)_n$ fragments expanded less (not shown). $(CAG)_{10}$, $(CAG)_{10}$ C, and $(CAG)_{10}CA$ expanded similarly, but there were qualitative differences in the expansion of $(AGC)_{10}$ A (almost no expansion, Figure 6), $(AGC)_{10}AG$ (extensive expansion) and $(AGC)_{10}$ (small expansion). Hence, the 3′end terminal G in $(AGC)_{10}AG$ caused the expansion of this 32-mer. However $(GCA)_{10}$ GC did not almost expand, and $(GCA)_{10}$ and $(GCA)_{10}G$ did not expand as well. $(CTG)_{10}$, $(CTG)_{10}$ C, $(CTG)_{10}CT$, $(TGC)_{10}$, $(TGC)_{10}$ T, $(TGC)_{10}TG$, $(GCT)_{10}$, $(GCT)_{10}G$, and $(GCT)_{10}$ GC were all almost reluctant to the expansion.

Further experiments with a $(CTG)_n$ set of fragments showed that $(CTG)_3$, $(CTG)_3C$, and $(CTG)_3CT$ did not expand at any annealing temperature within 37–72°C whereas $(TGC)_3$ modestly expanded at 37, 50, and 60°C (not shown). $(TGC)_3T$ provided two major discrete bands and $(TGC)_3TG$ provided an average of the amplification products obtained with $(TGC)_3$ and $(TGC)_3TG$. The sequence permutation $(GCT)_3$, $(GCT)_3G$, and $(GCT)_3$ GC surprisingly provided more expanded material than the previous set. This was especially surprising with $(GCT)_3GC$ that has two strong bases (G or C) at both ends because $(GCA)_3GC$ does not expand at all. $(CTG)_4$ slightly expands, $(CTG)_4C$ expands more and $(CTG)_3CT$

provides a similar but stronger ladder of expanded fragments at 50°C like $(CTG)_4 \cdot (TGC)_4 \cdot (TGC)_4TG$ and especially $(TGC)_3TG$ expanded more than the previous set. $(GCT)_4$ provided a small amount of expanded material. $(GCT)_3G$ and $(GCT)_4GC$ provided much more expanded material but the molecules were short.

$(GAA)_n$ and Other Purine-Rich Nucleotide Fragments

Purine nucleotide fragments, thst is, AGGAA, $(AGGAA)_2$, $(AGGAA)_3$, and $(AGGAA)_3GAGAA$ did not expand (not shown) at any examined annealing temperature (37, 50, 60, 72°C) at 20 cycles. However $(TA)_8GCGAAGC$ and $GCGAAGC(TA)_8$ expanded at 60°C but not at 37, 50, 72°C. The GCGAAGC sequence folds into a superstable hairpin.[26] GCGAAAGCT did not expand but $(GCGAAAGCT)_2$ did though the expanded material was only about 200 base pairs in length. AGCTTTCGC also did not expand but $(AGCTTTCGC)_2$did at 50°C but not at 37, 60, and 72°C. Neither $(AGGAA)_4$ nor GGAGGGAG did expand at any examined annealing temperature (37, 50, 60, 72°C). $(TGGTT)_4$ did not expand but $(TGGTA)_4$ did. Expansion of $(GAA)_n$ repeats is related to Friedreich ataxia.[27] Initially, we performed PCR of $(GAA)_3$, $(GAA)_3G$, $(GAA)_3GA$, $(AAG)_3$, $(AAG)_3A$, $(AAG)_3AA$, $(AGA)_3$, $(AGA)_3A$, and $(AGA)_3AG$ and found that only $(AGA)_3A$ and $(AGA)_3AG$ were expanded at 37°C. The expansion of these two fragments vanished at 50, 60, and 72°C and the remaining seven fragments did not expand at any of the four annealing temperatures. Lengthening of the fragments by one GAA triplet totally eliminated the expansions because $(GAA)_4$, $(GAA)_4G$, $(GAA)_4GA$, $(AAG)_4$, $(AAG)_4A$, $(AAG)_4AA$, $(AGA)_4$, $(AGA)_4A$, and $(AGA)_4AG$ failed to expand at any of the four annealing temperatures. Obviously, the purine nucleotide strands of DNA are reluctant to the expansion. Lengthening of the purine nucleotide fragments by two more GAA triplets led to expansion of $(GAA)_6$ and $(GAA)_6G$ at 37°C, and mainly $(AAG)_6$ (Figure 7) at 37, 50, and 60°C. Expansions or formation of separate heavy electrophoretic bands were observed after PCR of $(AAG)_6A$, $(AAG)_6AA$, $(AGA)_6$, $(AGA)_6A$, and $(AGA)_6AG$ at 37°C. Lengthening of the fragments by four more GAA repeats led to a strong expansion of $(AGA)_{10}AG$ at all four annealing temperatures, whereas $(GAA)_{10}G$ slighly expanded at 37°C and 50°C, and $(AAG)_{10}A$, $(AAG)_{10}AA$, $(AGA)_{10}$, and $(AGA)_{10}A$ slightly expanded at 37°C. Very weak or no expansions were observed with $(GAA)_{10}$, $(GAA)_{10}GA$, and $(AAG)_{10}$. Further lengthening led to expansion of $(GAA)_{30}$ at all four annealing temperatures whereas $(GAA)_{17}$, $(GAA)_{20}$, and $(GAA)_{25}$ expanded very slightly or not at all (Figure 8). $(GAA)_{40}$ and $(GAA)_{50}$ did not expand at all. Treatment of the amplified product of $(GAA)_{30}$ by DNase I led to the product elimination whereas the single specific nucleases P1 and S1 exerted no significant effect on the PCR product. This indicates that the

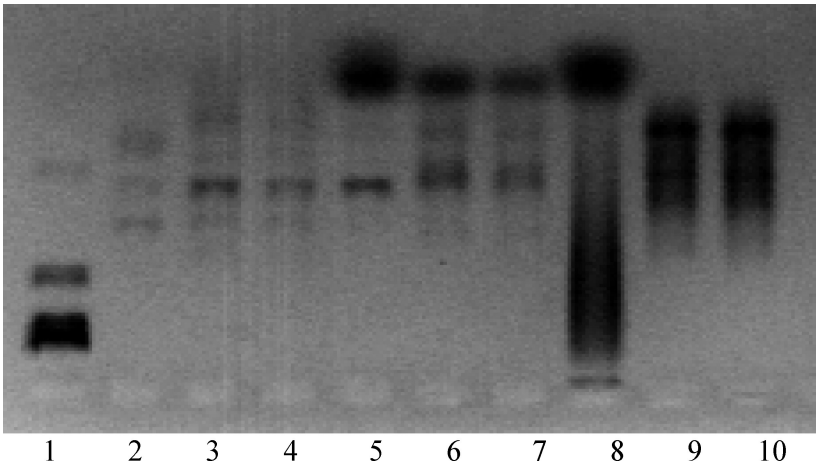

FIGURE 7 PCR products of (lane 2) $(GAA)_6$, (lane 3) $(GAA)_6G$, (lane 4) $(GAA)_6GA$, (lane 5) $(AGA)_6$, (lane 6) $(AGA)_6A$, (lane 7) $(AGA)_6AG$, (lane 8) $(AAG)_6$, (lane 9) $(AAG)_6A$ and (lane 10) $(AAG)_6AA$. (lane 1) marker λDNA/Hind III. 35cycles, annealing temperature 37°C.

PCR product is double-stranded DNA. This suggestion is corroborated by the fact that the product is stained by ethidium.

We have also performed experiments with $(GA)_{10}$ and $(AG)_{10}$ (not shown). $(GA)_{10}$ provided short (about 2 kbp) molecules at lower annealing temperatures but long (about 20 kbp) at high annealing temperatures. $(AG)_{10}$ expanded significantly less providing about 2 kbp molecules, at most, even at high annealing temperatures. This is another illustration that the terminal nucleotides are very important for the expansion mechanism analyzed in this study. It is interesting that the very long molecules were obtained with $(GA)_{10}$ even in the absence of dCTP or dGTP whereas dATP and dTTP were absolutely necessary to obtain the long molecules

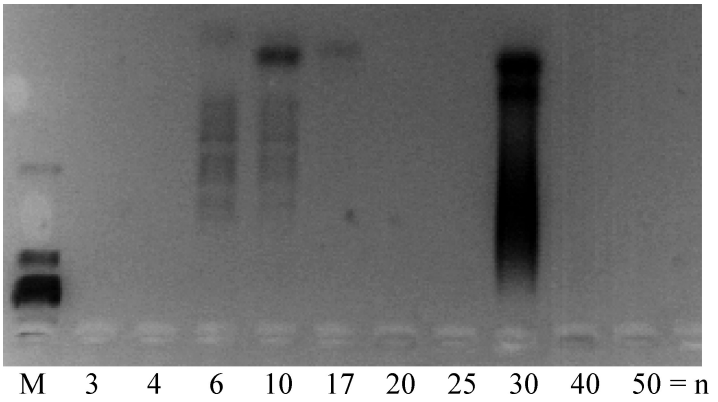

FIGURE 8 Dependence of the expansions of $(GAA)_n$ on the number n of the GAA repeats. 35 cycles, annealing temperature 37°C. M is the λDNA/Hind III marker.

(not shown). This indicates that the expanded molecules are duplexes of poly(dA-dT) rather than poly(dA-dG)·poly(dC-dT). On the other hand, $(AG)_{10}$ failed to expand in the absence of any of the four triphosphates. We also have tried PCR with $(GA)_3$ $(TA)_4$ $(GA)_3$ and CGAGACG but none of them was found to expand.

$(TTC)_n$ Strands

$(TTC)_n$ strands are complementary to the $(GAA)_n$ strands and, therefore, their expansions are also relevant to Friedreich ataxia.[27] We performed PCR at the usual four annealing temperatures (37, 50, 60, and 72°C) with $(TTC)_n$, n = 3, 4, 6, 10, 17, 20, 25, 30, 40, and 50, $(TTC)_nT$, n = 3, 4, 6, 10, $(TTC)_nTT$, n = 3, 4, 6, 10, $(TCT)_n$, n = 3, 4, 6, 10, $(TCT)_nT$, n = 3, 4, 6, 10, $(TCT)_nTC$, n = 3, 4, 6, 10, $(CTT)_n$, n = 3, 4, 6, 10, $(CTT)_3C$, n = 3, 4, 6, 10, $(CTT)_nCT$, n = 3, 4, 6, 10. Large amounts of the most expanded molecules were provided by $(TTC)_4$ at 37°C whereas 50°C and higher temperatures led to no expansion of $(TTC)_4$ (Figure 9). $(TTC)_{25}$ was another strongly expanding molecule (Figure 9). It provided large amounts of the most expanded molecules at all four annealing temperatures. Remarkably $(TTC)_{10}$ did not expand at all. $(TTC)_6$ expanded at lower temperatures whereas $(TTC)_{17}$ provided large amounts of relatively short molecules. $(TTC)_{20}$ provided a lot of relatively short PCR products at 37°C, but small amounts of the PCR products of a similar length at 50°C and 60°C, and no expanded molecules at 72°C. $(TTC)_{30}$ neither expanded at 72°C whereas it provided a lot of PCR product at 37°C and 50°C. $(TTC)_{40}$ and $(TTC)_{50}$ both failed to expand at 50, 60, and 72°C and they provided a little bit of the PCR product at 37°C.

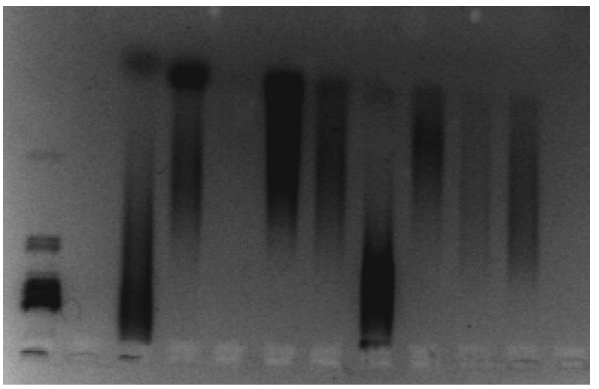

FIGURE 9 Dependence of the expansions of $(TTC)_n$ on the number n of the TTC repeats. 35 cycles, annealing temperature 37°C. M is the λDNA/Hind III marker.

$(TTC)_3T$ did not expand at all, $(TTC)_4T$ expanded at 37°C but not at 50, 60, and 72°C, $(TTC)_6T$ and $(TTC)_{10}T$ provided a lot of the PCR product at 37°C and 50°C whose length was, however, longer with $(TTC)_6T$. None of the two fragments expanded at 72°C. $(TTC)_3TT$, $(TTC)_6TT$, and $(TTC)_{10}TT$ expanded at any of the four annealing temperatures whereas $(TTC)_4TT$ provided a medium amount of very long molecules at 37°C and 50°C, but not at 60°C and 72°C.

$(TCT)_3$ did not expand whereas $(TCT)_4$ and $(TCT)_6$ generated mostly medium amounts of long molecules at the lower annealing temperatures but not at the higher annealing temperatures. Neither $(TCT)_{10}$ nor $(TCT)_3T$ expanded. In contrast, $(TCT)_4T$ and $(TCT)_6T$ belonged among the most expanding molecules at the lower temperatures whereas the higher annealing temperatures eliminated the expansion. $(TCT)_3TC$ also belonged among the most expanding molecules at the lower temperatures. $(TCT)_4TC$ expanded less. $(TCT)_6TC$ was still more reluctant to the expansion but, remarkably, $(TCT)_{10}TC$ generated a lot of the expanded material at 50°C and 60°C though the material was of a medium length.

$(CTT)_3$ did not expand, $(CTT)_4$ expanded a lot at 37°C, $(CTT)_6$ and $(CTT)_{10}$ expanded rather less. Unlike $(CTT)_3$, $(CTT)_3C$ expanded a lot at 37°C as did $(CTT)_4C$ at 50°C and $(CTT)_6C$ at 37°C and 50°C. In contrast, $(CTT)_{10}C$ did not expand almost at all. $(CTT)_3CT$ did not expand, $(CTT)_4CT$ did a lot at 37°C, and $(CTT)_6CT$ and $(CTT)_{10}CT$ expanded modestly at the lower temperatures. It is remarkable that $(TCT)_6$, $(TCT)_6T$ and $(CTT)_6CT$ expanded much better than $(TTC)_6TT$ and $(TCT)_6TC$ (Figure 10).

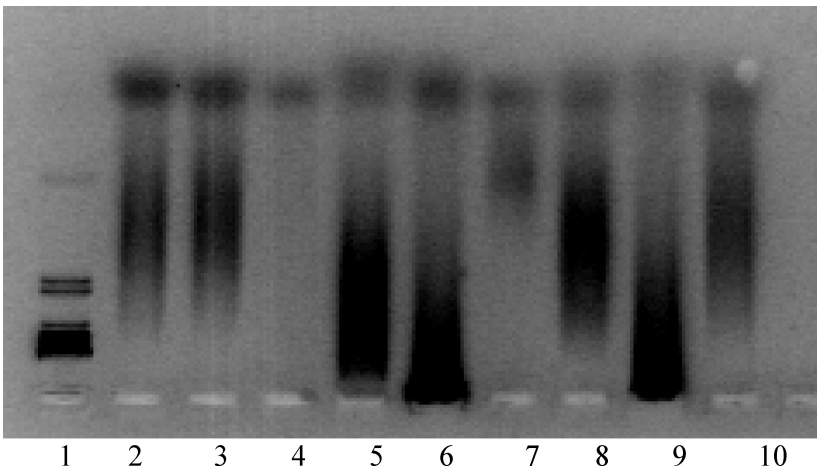

FIGURE 10 PCR products of (lane 2) $(TTC)_6$, (lane 3) $(TTC)_6T$, (lane 4) $(TTC)_6TT$, (lane 5) $(TCT)_6$, (lane 6) $(TCT)_6T$, (lane 7) $(TCT)_6TC$, (lane 8) $(CTT)_6$, (lane 9) $(CTT)_6C$ and (lane 10) $(CTT)_6CT$. (lane 1) marker λDNA/Hind III. 35cycles, annealing temperature 37°C.

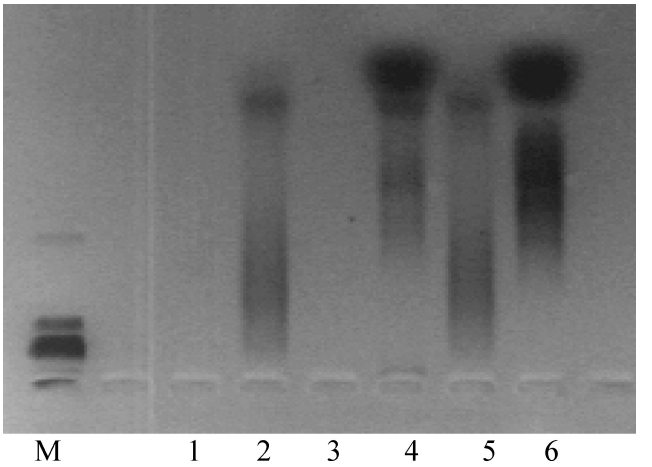

FIGURE 11 PCR products of (lane 1) $(GAA)_7$, (lane 2) equimolar mixture of $(GAA)_7$with $(TTC)_7$, (lane 3) $(TTC)_7$, (lane 4) $(CAG)_8$, (lane 5) equimolar mixture of $(CAG)_8$ with $(CTG)_8$, and (lane 6) $(CTG)_8$. 20cycles, annealing temperature 60°C. (lane M) the λDNA/Hind III marker.

Duplexes of Complementary Strands

We also have performed many experiments with mixtures of complementary strands. A general conclusion following from these experiments is that the mixtures provide more expanded material than the separate strands. An extreme example is provided by $(GAA)_7$ and $(TTC)_7$ none of which is expanded in the course of PCR whereas the equimolar mixture of $(GAA)_7$ with $(TTC)_7$ gives a lot of the expanded material (Figure 11). The equimolar mixture of $(CAG)_8$ and $(CTG)_8$ also provides longer molecules than either separate $(CAG)_8$ or $(CTG)_8$ (Figure 11). These experiments demonstrate that duplex formation promotes DNA expansion during PCR.

DISCUSSION

This work deals with the phenomenon of DNA expansion when short fragments of DNA are at the beginning of the reaction and multikilobase molecules of DNA are at the end of the reaction. The experiments are relevant to two biologically relevant issues. Firstly DNA expansions are connected with many diseases including neurological diseases[6] and cancer.[7] Second, microsatellites change their length within non-pathological limits, which is a means of genome evolution and modulation of its function.[4] Both of these issues are fundamental for biology and medicine and, therefore, understanding of the mechanisms of DNA expansion is of a great interest. The mechanism is what we have tried to understand in this work. We performed a bioinformatic analysis of the DNA length expansion in eukaryotic genomes leading to a conclusion that the length

expansion is a general property of simple sequence repeats in eukaryotic genomes,[28] with a possible exception of those simple sequence repeats where the CG dinucleotide is combined with TA pairs in the repeat, or the TA dinucleotide is combined with GC pairs in the repeat (J. Kypr and J. Mrázek, unpublished). On the other hand, the simple sequence repeats occur in prokaryotic genomes in similar amounts as in the correspomding randomized sequences (J. Kypr and J. Mrázek, unpublished). Hence, a eukaryote-specific property, most probably a protein or proteins, contribute to the expansion phenomenon.

Here, we studied the possible contribution of DNA structure of the expansion phenomenon by means of PCR. With some oligonucleotides, for example, $(TA)_8$, only a few PCR cycles sufficed to obtain expanded DNA molecules whereas 35 cycles were necessary to observe a similar effect with other oligonucleotides. Expanded molecules were even obtained with simple repeats such as $(CA)_n$, $(CTT)_n$ and others for which we have never observed homoduplex formation under the conditions of PCR, that is, at the alkaline pH. Hence, the duplex formation either is not necessary for the molecules to expand, which, therefore, would require an unprimed activity of the Taq polymerase, or the Taq polymerase somehow stabilizes the duplex formation. PCR is an exponential method so that very few duplexes suffice to generate the expanded molecules. Very short molecules, for example, $(TA)_4$, do not expand presumably because they do not generate stable duplexes, and relatively long molecules, for example, $(TA)_{16}$, also do not expand, presumably because they generate too stable conformers that also are not the best substrates for the expansion. We imagine that the optimum conformer is a slipped duplex or a slipped hairpin[10,11] where the slippage generates single-stranded ends. The slipped conformers may serve as a template for the polymerase. The polymerase adds the missing nucleotides and this way expands the molecule. Then the sample is denatured, other slipped conformers arise after annealing and the process is repeated. However, this proposal does not expain all observations presented in this article. At present, we cannot offer an exhaustive explanation of all our observations.

Yet, there are several rules that we were able to derive form the experiments reported in this article. First, the PCR result depends on the number of PCR cycles. Some fragments, for example, $(TA)_8$ are heavily expanded even after two cycles and the further cycles significantly change neither the length nor the amount of the expanded material. On the other hand, a majority of the other fragments provides detectable amounts of the expanded material only after many more (e.g., 35) PCR cycles. Second, the expansion depends on the polymerase. Taq polymerase provides the largert amounts of the expanded material among the polymerases we tested (unpublished data). Third, the expanded material mostly are the normal Watson-Crick duplexes as indicated by the dependence of the expansion

phenomenon on the presence of all of the four triphosphates, staining of the expanded molecules by ethidium, digestion by the DNase I nuclease and the resistance of the P1 and S1 nucleases. Fourth, the PCR product depends on the annealing temperature. The expanded material mostly is detected at some annealing temperatures but not at others. The annealing temperatures leading and not leading to the expanded PCR product depend on the original DNA fragment present in the PCR sample. On the other hand, the expansion does not significantly depend on the extension temperature. Fifth, most expanded molecules arise at the pH values and the $MgCl_2$ and KCl concentrations recommended by the polymerase producer. Sixth, the PCR result depends on the length of the DNA fragment. Too short as well as too long fragments expand much less than intermediate fragments whose length lies within approximately 10–25 nucleotides. Seventh, temperature melting experiments demonstrate[22] that the annealing temperatures leading to expansion cause total denaturation of the fragment duplexes or hairpins that are formed, for example, in the case of selfcomplementary sequences such as $(TA)_n$. Eighth, we observe massive expansion even with fragments such as T_n or $(CA)_n$ where we have never observed homoduplex formation at the alkaline pH values used for PCR. Ninth, some apparently small changes of the fragment primary structure cause qualitative changes in the expansion. For example, A_8T_8 belongs among the most expandable molecules whereas T_8A_8 does not expand at all. Some fragments composed only of G and C do not expand at all whereas others do. Hence, neither the (G+C) content is a decisive factor though (A+T)-rich molecules, on average, expand better than the (G+C)-rich ones. Tenth, pyrimidine-rich fragments expand better than purine-rich fragments. Eleventh, terminal nucleotides seem to influence expansion more than internal nucleotides. Specifically, terminal G and C nucleotides frequently, but not always, suppress the expansion.

This article presents a large body of experimental data, that is, more than 3500 polymerase chain reactions with more than 400 DNA oligonucleotides. The experiments permitted us to identify factors influencing results of the PCR, as described above. The amplification is found to be promoted by a repetitive primary structure of DNA, its length being within 10–25 nucleotides, an appropriate annealing temperature, (A+T)-richness, pyrimidine-richness and an absence of (G+C)-rich ends. Evidently, the process of DNA expansion is a much more complex phenomenon than we thought at the beginning of this project. In addition, proteins other than polymerases certainly influence the expansion in vivo. The expansion is worth of further detailed studies. However, the present, relatively very simple experimental design can be used to study DNA structure aspects contributing to microsatellite length polymorphism. This contribution is important because the mocrosatellite length polymorphism discriminates among individuals of the same organism as well as belongs to factors causing

pathological microsatellite expansion in many neurological diseases and cancer.

REFERENCES

1. Orgel, L.E. The origin of life—A review of facts and speculations. *TIBS* **1998**, 23, 491–495.
2. Trifonov, E.N. Segmented structure of protein sequences and early evolution of genome by combinatorial fusion of DNA elements. *J. Mol. Evol.* **1995**, 40, 337–342.
3. Cleary, J.D.; Pearson, C.E. Replication fork dynamics and dynamic mutations: The fork-shift model of repeat instability. *Trends in Genetics* **2005**, 21, 272–280.
4. Hancock, J.M. Simple sequences and the expanding genome. *BioEssays* **1996**, 18, 421–425.
5. Schäfer, R.; Zischler, H.; Epplen, J.T. (CAC)5 a very informative oligonucleotide probe for DNA fingerprinting. *Nucl. Acids Res.* **1988**, 16, 5196–5196.
6. Wells, R.D.; Dere, R.; Hebert, M.L.; Napierala, M.; Son, L.S. Advances in mechanisms of genetic instability related to hereditary neurological diseases. *Nucl. Acids Res.* **2005**, 33, 3785–3798.
7. Goellner, G.M.; Tester, D.; Thibodeau, S.; Almqvist, E.; Goldberg, Y.P.; Hayden, M.R.; McMurray, C.T. Different mechanisms underlie DNA instability in Huntington disease and colorectal cancer. *Am. J. Hum. Genet.* **1997**, 60, 879–890.
8. Schachman, H.K.; Adler, J.; Radding, C.M.; Lehman, I.R.; Kornberg, A. Enzymatic synthesis of deoxyribonucleic acid. VII. Synthesis of a polymer of deoxyadenylate and deoxythymidylate. *J. Biol. Chem.* **1960**, 235, 3242–3249.
9. Kornberg, A.; Bertsch, L.L.; Jackson, J.F.; Khorana, H.G. Enzymatic synthesis of deoxyribonucleic acid. XVI. Oligonucleotides as templates and the mechanism of their replication. *Proc. Natl. Acad. Sci. USA* **1964**, 51, 315–323.
10. Tautz, D.; Renz, M. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucl. Acids Res.* **1984**, 12, 4127–4138.
11. Schlötterer, C.; Tautz, D. Slippage synthesis of simple sequence DNA. *Nucl. Acids Res.* **1992**, 20, 211–215.
12. Behn-Krappa, A.; Doerfler, W. Enzymatic amplification of synthetic oligodeoxyribonucleotides: Implications for triplet repeat expansions in the human genome. Hum. Mutation **1994**, 3, 19–24.
13. Hanaki, K.; Odawara, T.; Muramatsu, T.; Kuchino, Y.; Masuda, M.; Yamamoto, K.; Nozaki, C.; Mizuno, K.; Yoshikura, H. Primer/template-independent synthesis of poly d(a-t) by Taq polymerase. *Biochem. Biophys. Res. Commun.* **1997**, 238, 113–118.
14. Ogata, N.; Miura, T. Genetic information 'created' by archaebacterial DNA polymerase. *Biochem. J.* **1997**, 324, 667–671.
15. Ogata, N.; Miura, T. Creation of genetic information by DNA polymerase of the thermophilic bacterium thermus thermophilus. *Nucl. Acids Res.* **1998**, 26, 4657–4661.
16. Ogata, N.; Miura, T. Creation of genetic information by DNA polymerase of the archaeon thermococcus litoralis: Influences of temperature and ionic strength. *Nucl. Acids Res.* **1998**, 26, 4652–4656.
17. Wu, M.-J.; Chow, L.-W.; Hsieh, M. Amplification of GAA/TTC triplet repeat in vitro: Preferential expansion of (TTC) *n* strand. *Biochim. Biophys. Acta* **1998**, 1407, 155–162.
18. Lyons-Darden, T.; Topal, M.D. Effects of temperature, Mg^{2+} concentration and mismatches on triplet-repeat expansion during DNA replication in vitro. *Nucl. Acids Res.* **1999**, 27, 2235–2240.
19. Ogata, N.; Morino, H. Elongation of repetitive DNA by DNA polymerase from a hyperthermophilic bacterium thermus thermophilus. *Nucl. Acids Res.* **2000**, 28, 3999–4004.
20. Tuntiwechapikul, W.; Salazar, M. Mechanism of in vitro expansion of long DNA repeats: Effect of temperature, repeat length, repeat sequence, and DNA polymerases. *Biochemistry* **2002**, 41, 854–860.
21. Heidenfelder, B.L.; Topal, M.D. Effects of Sequence on repeat expansion during dna replication. *Nucl. Acids Res.* **2003**, 31, 7159–7164.
22. Reichova, N.; Kypr, J. Expansion during PCR of short single-stranded DNA fragments carrying nonselfcomplementary dinucleotide or trinucleotide repeats. *Mol. Biol. Rep.* **2003**, 30, 155–163.
23. Usdin, K.; Woodford, K.J. CGG repeats associated with DNA instability and chromosome fragility form structures that block DNA synthesis in vitro. *Nucl. Acids Res.* **1995**, 23, 4202–4209.

24. Kremer, E.J.; Pritchard, M.; Lynch, M.; Yu, S.; Holman, K.; Baker, E.; Warren, S.T.; Schlessinger, D.; Sutherland, G.R.; Richards, R.I. Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)n. *Science* **1991**, 252, 1711–1714.
25. Hrabcova, I.; Kypr, J. Genomic occurrence of microsatellites containing integral and non-integral repeat numbers. *Biochem. Biophys. Res. Commun.* **2003**, 300, 824–831.
26. Hirao, I.; Nishimura, Y.; Tagawa, Y.-i.; Watanabe, K.; Miura, K.-i. Extraordinarily stable mini-hairpins: Electrophoretical and thermal properties of the various sequence variants of d(GCFAAAGC) and their effect on DNA sequencing. *Nucl. Acids Res.* **1992**, 20, 3891–3896.
27. Campuzano, V.; Montermini, L.; Molto, M.D.; Pianese, L.; Cossée, M.; Cavalcanti, F.; Monros, E.; Rodius, F.; Duclos, F.; Monticelli, A.; Zara, F.; Canizares, J.; Koutnikova, H.; Bidichandani, S.I.; Gellera, C.; Brice, A.; Trouillas, P.; De Michele, G.; Filla, A.; De Frutos, R.; Palau, F.; Patel, I.P.; Di Donato, S.; Mandel, J.; Cocozza, S.; Koenig, M.; Pandolfo, M. Friedreich's ataxia: Autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science* **1996**, 271, 1423–1427.
28. Mrázek, J.; Kypr, J. Length expansion is a general property of simple sequence repeats in eukaryotic genomes. *Miami Biotechnology Short Reports* **1994**, 5, 39–39.