Nucleic Acids Research, 1998, Vol. 26, No. 12 **2843–2848**

Non-homologous recombination mediated by *Thermus aquaticus* DNA polymerase I. Evidence supporting a copy choice mechanism

Peter G. Zaphiropoulos*

Department of Bioscience, Center for Nutrition and Toxicology, Karolinska Institute, Novum, 141 57 Huddinge, Sweden

Received March 31, 1998; Revised and Accepted May 4, 1998

ABSTRACT

RT–PCR amplification of P450 2C6 from rat liver, using primers in opposite orientations of exon 6, resulted in PCR products containing segments of exons joined at non-consensus splice sites. Moreover, many of the PCR products identified were composed of not only a single region containing exonic segments joined at non-consensus splice sites but, instead, of several repeats of the non-canonically joined region. To investigate whether these PCR products represent pre-existing molecules or are generated during the amplification process, the liver cDNA template was replaced by a plasmid containing the P450 2C6 cDNA. Surprisingly, PCR products containing repeats of non-canonically joined exonic segments were again revealed. In some cases the position of this non-canonical joining was a sequence of one or two identical nucleotides; however, there were also a number of products lacking any nucleotide identity at the position of joining. DNA nicking and/or DNA damage is thought to favour recombination during PCR, probably by misalignment of incomplete DNA strands; however, the presence of multiple repeats of the recombined region in the PCR products identified suggests a certain repetitiveness of the underlying mechanism. It is therefore proposed that these products result from a template switching event that occurs several times during a single polymerization step, following a rolling circle model of DNA synthesis.

INTRODUCTION

The splicing process has a unique capability of accurately removing intronic sequences from pre-mRNAs. There is a general sequence conservation at the position of intron removal with the most pronounced feature being the terminal intronic dinucleotides. Nearly all introns start with GT and end with AG, with the exception being a small number of introns that follow an AT to AC dinucleotide rule (1,2). Moreover, by mutagenesis analysis it has been shown that splicing can also occur between additional pairs of terminal dinucleotides GC to AG, GT to AT, AT to AA and AT to AG (3–5). Interestingly, the sequence of the terminal dinucleotides does not dictate which spliceosomal system is being used. Both the U2- and the U12-dependent splicing systems can be active on GT to AG or AT to AC introns (5).

Additionally, it has been found that the order of the exons that are spliced together may not always follow the same order as in genomic DNA. Transcripts containing 3′ exons joined together with 5′ exons have been found in several cases, as well as single exons joined head-to-tail (6–12). It has been proposed that these transcripts may represent circular RNA molecules, although the possibility of some being linear *trans*-spliced species has not been formally excluded. *In vitro*, with the use of splicing extracts, the detection of exonic RNA molecules with aberrant electrophoretic patterns, indicative of their circular nature, has been accomplished (13,14). Most recently, ribozymes have been engineered in a way that allows production of circular exons (15).

In this report evidence is presented that, during the efforts to detect RNA transcripts that might not follow the canonical rules for exon joining, RT–PCR products containing segments of exons joined at non-consensus splice sites were revealed. However, these products do not represent pre-existing RNA molecules, but are rather generated during the PCR amplification process. Moreover, the mechanism that produces these species appears to be an inherent template switching ability of the polymerase itself.

MATERIALS AND METHODS

PCR analysis

Total RNA from rat liver was isolated by the guanidinium isothiocyanate method (16) and subjected to reverse transcription using random hexamers or oligo dT primers as described before (12). PCR amplifications were performed using Perkin Elmer's model 2400 or 480 thermocyclers, for 1 min at 94°C, 1 min at the desired annealing temperature and 1 min at 72°C, in a total volume of 50 µl. The annealing temperature was identical to or at the most 3°C lower than the melting temperature of the oligos. Five units of *Taq* polymerase were used in a buffer containing 50 mM KCl, 10 mM Tris, pH 9.0, 0.1% Triton X-100, 2 mM $MgCl_2$ with 0.2 mM of each dNTP. For the nested amplifications, 1 µl from the initial amplification reaction was directly used. The PCR products were cloned after ligation with the pGEM T vector (Promega). Sequencing analysis of individual clones, that were randomly picked, was performed with dye-dideoxy nucleotides

*Tel: +46 8 6089151; Fax: +46 8 6081501; E-mail: peza@cnt.ki.se

using the facilities of Cybergene AB (Huddinge, Sweden). The inserts of the clones were fully sequenced ensuring that the expected PCR primers were present at the ends. Sequencing comparisons were performed using the GCG program of the University of Wisconsin.

P450 2C6 plasmid construct

The P450 2C6 cDNA clone in pGEM T vector that was used in the experiments of Figure 2 was generated by PCR amplification of rat testicular cDNA with primers 2C6 1F and 2C6 9R. Sequencing established that there was a nucleotide substitution (T→C) in exon 7, at position 32 of Figure 2. This represents a normal mistake made by DNA polymerases, and could easily have occurred during the 40 cycles of PCR amplification that were performed in order to produce detectable amounts of that cDNA. This clone was digested with restriction enzyme *Pvu*II, producing one fragment containing mainly insert sequences and another containing the remaining of the vector and then purified through a Wizard DNA Clean up kit (Promega), prior to the amplifications.

Table 1. Primers used in the PCR amplification experiments

2C6, 6f: 5′ ACA GAG ACA ACA AGC ACA A
2C6, 6F (nested): 5′ ACT GAG ATA TGC TCT GTT G
2C6, 6r: 5′ ACA AAT CAG TCA CAG TGA TT
2C6, 6R (nested): 5′ ATA AGT TTT CAA GTG TAA ATT C
2C6, 1F: 5′ TGT TGC TGG TGC TCA CTC TCA
2C6, 9R: 5′ AGG CAA ATC CAT TGA AAA CTG GA

RESULTS

Identification of RT–PCR products containing segments of P450 2C6 exons joined at non-consensus splice sites

It has previously been proposed that the process of exon skipping might be related with the process of joining exons with an order different to that of genomic DNA (10,12). For that purpose the pattern of expression of P450 2C6 was investigated in rat liver, the tissue that is known to express the highest amounts of that P450. Using RT–PCR analysis, with an exon 1 sense and an exon 9 antisense primers (Table 1), the major product revealed was the canonically spliced mRNA of nine exons. Moreover, several exon skipped mRNAs were also detected: form A had skipped exons 2, 3, 4 and 5, form B had skipped exons 6 and 7 and form C had skipped all of these six exons (Fig. 1A). Using sets of primers in both orientations of exon 6 (Table 1), efforts were initiated in order to identify RNAs that might be composed of the skipped exons. Indeed a large number of PCR products was generated with the use of these primers (Fig. 1B). However, none of these products was exclusively composed of exons that were spliced at their expected splice sites. There were always exonic segments joined at positions bearing no similarity to a canonical splice site consensus. Moreover, in many cases, these products were not composed of a single non-canonically joined exonic region but of several repeats of that region (see Table 2 for a schematic representation). In one case these repeats reached the number of four. All of the cloned and sequenced RT–PCR products from the liver amplification are shown in Figure 1C.

When the same exon 6 primers were used with cDNA from rat testis, a tissue that expresses much lower amounts of P450 2C6 than the liver, only a single PCR product was detected. This product contained a segment of exon 8 joined to a segment of exon 3 at a position with no similarity to a splice consensus. Moreover, PCR products containing incomplete exons joined together had previously been detected during amplification of P450 2C18 from human epidermis with primers in opposite orientation of a single exon (12). In addition, when amplification of P450 2C18 was performed using exon 1 sense and exon 9 antisense primers, PCR products containing non-canonically joined exonic segments were also identified (data not shown).

PCR amplification of cloned P450 2C6 also results in products containing repeats of a recombined sequence

To investigate whether the identified RT–PCR products represent pre-existing RNA molecules or were generated during the RT–PCR amplification process, a cloned P450 2C6 cDNA was used for PCR with the same sets of exon 6 sense and antisense primers, as before. Surprisingly, the cDNA plasmid allowed the production of PCR products that had the same characteristics as the ones detected during RT–PCR, namely that several repeats of a non-canonically joined exonic region were present (Fig. 2A). However, none of these products was identical to the ones previously detected during the RT–PCR. When low amounts of plasmid cDNA were used for this amplification, recombination at only a single position was observed; however products containing varying numbers of repeats of this recombined region were identified. When higher amounts of cDNA were used, recombination at several different positions was detected (Fig. 2B). These findings suggest that the more starting material, the higher the probability for such an event to occur. Moreover, this observation is also consistent with the fact that cDNA from testis (a tissue that expresses minimal amounts of 2C6) produced only a single type of recombined product, while cDNA from liver (a tissue that expresses much higher amounts of 2C6) produced a variety of products that had recombined at different positions.

The finding that plasmid DNA allows the production of PCR products containing recombined sequences strongly disfavors the possibility that the non-canonically joined products, previously detected during RT–PCR, could represent pre-existing RNA molecules or might have been generated during the reverse transcription step. In addition, the absence of any consensus sequence or of extensive sequence identity at the positions of joining suggests that there may not be a critical sequence preference for these events to occur. In line with this apparent ‘random selection’ of the recombination junction is the identification of distinct PCR products in each of the amplification experiments that involved the sets of the exon 6 primers.

DISCUSSION

The observation that during the PCR process, products can be generated containing segments that are not contiguous in the unamplified molecules has been noted before (17–19). This *in vitro* recombination activity of the polymerase has been attributed to misalignment of prematurely terminated strands with homologous regions in the other strand. The misaligned strand would thus serve as a primer for the generation of a chimeric molecule during the subsequent elongation step. Moreover, DNA nicks would favor this *in vitro* recombination because they would allow the production of a population of incomplete strands that would tend to anneal at regions of partial homology (20). More recently, this

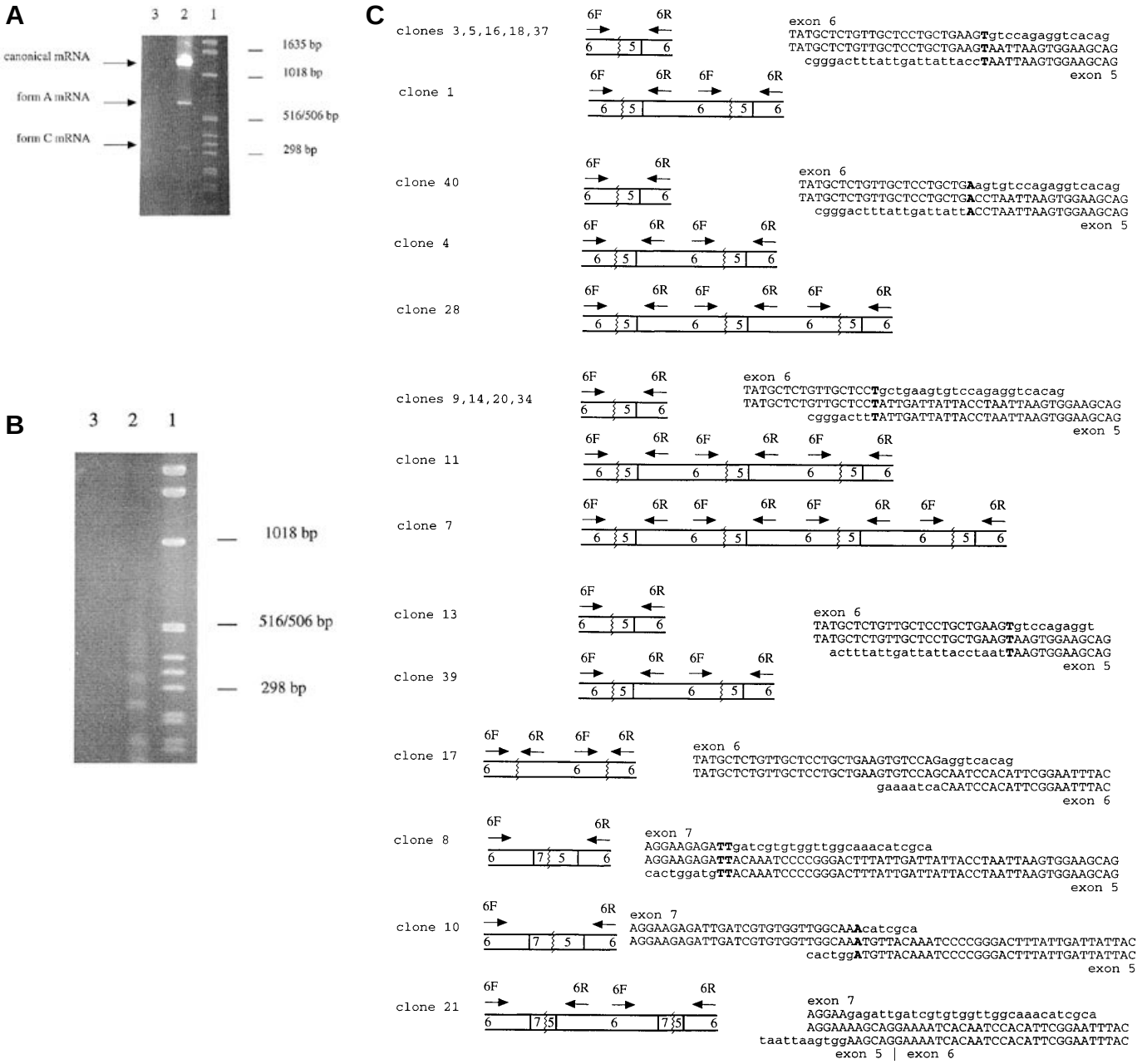
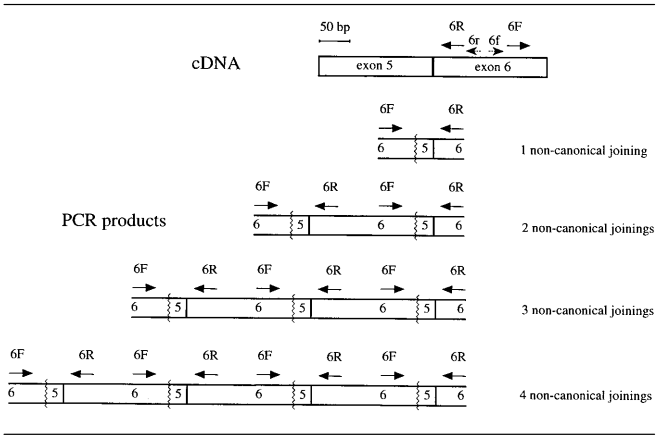


Figure 1. (**A**) RT–PCR analysis using primers 1F and 9R on rat liver RNA. The individual lanes on the agarose gel electrophoresis are as follows: 1, molecular weight markers; 2, RT–PCR products after 30 cycles of amplification with added liver RNA; 3, same as in lane 2 but without addition of RNA. Form B mRNA (skipped exons 6 and 7) cannot be clearly seen as a distinct band on the gel; however, it has been identified as a cloned product. (**B**) RT–PCR analysis using primers 6f and 6r and subsequently primers 6F and 6R on rat liver RNA. The individual lanes on the agarose gel electrophoresis are as follows: 1, molecular weight markers; 2, RT–PCR products after 30 cycles of initial and, subsequently, 15 cycles of nested amplification with added liver RNA; 3, same as in lane 2 but without addition of RNA. (**C**) Schematic representation of the structure of the cloned and sequenced RT–PCR products, and sequence of the non-canonically joined regions. The exons preceding and following the non-canonically joined regions (middle lines) are shown above and below, respectively. Bold letters indicate nucleotides that are identical in the junctions and small letters indicate nucleotides that are absent in the non-canonically joined molecules.

model of *in vitro* recombination during PCR has been extended by the observation that recombination at homologous regions can also take place during a single primer extension step (21). Thus, during the elongation step, the polymerase along with the extending strand has the capability to switch to another template, at a region of partial homology, therefore generating a recombined molecule.

This type of template switching recombination, also termed copy choice recombination, represents one of the two major models proposed to explain genetic exchange, the other being the breaking and rejoining model. Although there has been overwhelming support for this latter model as the major mechanism of genetic recombination (22,23), there is still evidence for a role of copy

Table 2. Schematic representation of the positions of the exon 6 primers in the P450 2C6 cDNA and general structure of the PCR products identified



The non-canonical joinings are indicated by zigzag vertical lines. Note that, although PCR products with different non-canonical joinings were identified, the non-canonical joinings present in each of the multi-repeat containing molecules were always identical.

choice recombination in some biological processes. For example, several polymerases including reverse transcriptase and *Escherichia coli* polymerases PolI and PolIII have been shown to either switch templates at regions of extensive homology or to slip between small direct repeats *in vitro* (24–28). *In vivo*, copy choice recombination events between small repeats has been proposed to occur in yeast and *E.coli* (29,30). Moreover, template switching is considered to be the prevailing mechanism of recombination in RNA viruses (31).

The recombined PCR products that have been identified in this report contain, in some cases, one or two identical nucleotides (in fact, only A or T) at their recombination junction. However, there were also a number of products with no sequence identity at all in this position. Therefore, partial homology appears not to be a prerequisite for these events to take place, although it is conceivable that it might enhance the frequency of occurrence of such an event. What then would be a plausible mechanism that could rationalize the production of such molecules? If, for example the polymerase pauses at certain sites during one PCR cycle, then during the subsequent cycle the prematurely terminated strands might, by misalignment, serve as primers to generate recombined molecules. However, in order to account for the identification of PCR products that contained several identical repeats of the recombined sequence, one would have to assume

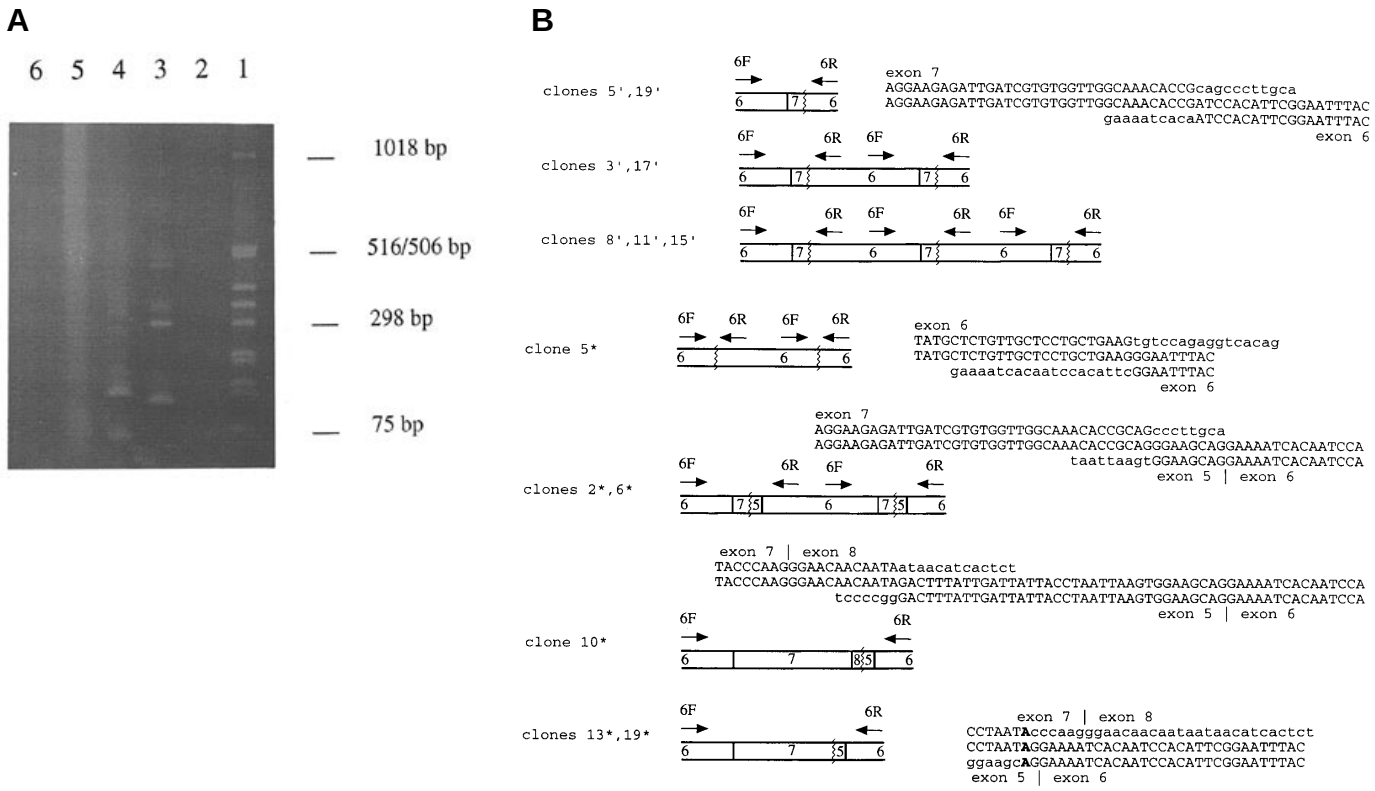


Figure 2. (**A**) PCR analysis using primers 6f and 6r and subsequently primers 6F and 6R on a cloned 2C6 cDNA plasmid. The individual lanes on the agarose gel electrophoresis are as follows: 1, molecular weight markers; 2, PCR products after 30 cycles of initial and, subsequently, 20 cycles of nested amplification with $10^{-6} \times 5$ ng of added 2C6 plasmid DNA; 3, same as in lane 2 but with $10^{-4} \times 5$ ng of added 2C6 plasmid DNA; 4, same as in lane 2 but with $10^{-2} \times 5$ ng of added 2C6 plasmid DNA; 5, same as in lane 2 but with 5 ng of added 2C6 plasmid DNA; 6, same as in lane 2 but with no addition of 2C6 plasmid DNA. (**B**) Schematic representation of the structure of the cloned and sequenced PCR products from lane 3 (labeled with ′) and lane 4 (labeled with *), and sequence of the non-canonically joined regions (same legend as in Fig. 1C).

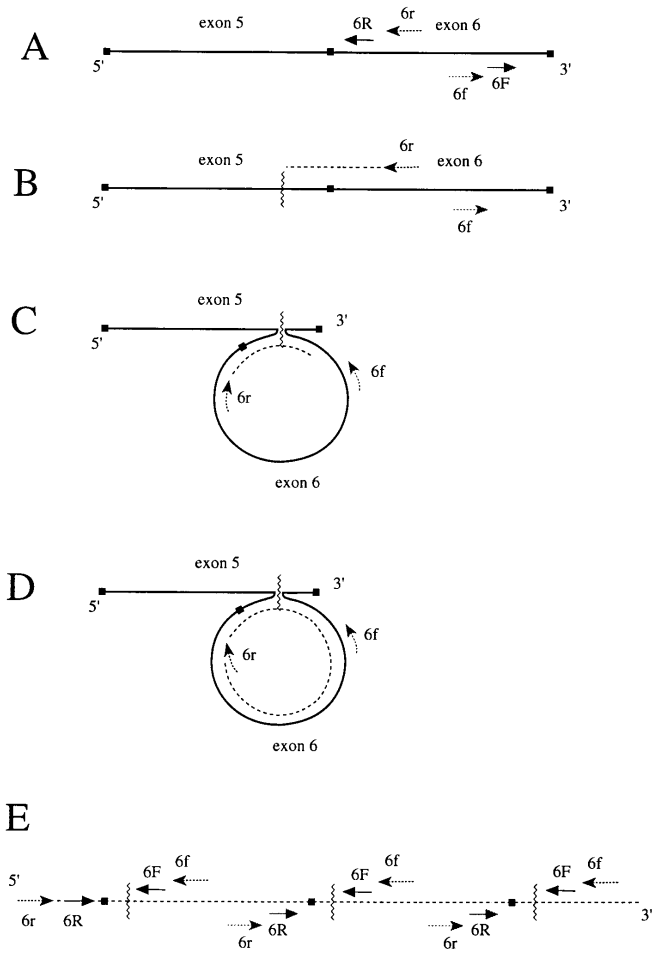


Figure 3. Proposed mechanism for the generation of PCR products containing several identical repeats of a recombined region. The polymerase initiates DNA synthesis at a specific priming sequence of the original template (**A**) and (**B**). At some nucleotide position during the elongation process the polymerase with the extending strand slips to a downstream nucleotide position of the same template and keeps the elongation process (**B**) and (**C**). The loop structure of the template molecule, that has been generated by the slippage event, is maintained by the complementarity with the newly synthesized strand (C). This growing strand displaces the previously synthesized strand during several rounds of DNA synthesis (**D**) resulting in a molecule, containing a large number of identical repeats (**E**). During the subsequent PCR amplification steps, there will be a tendency to amplify molecules having a single recombined region, since a number of internal priming sites will be present in the multi-repeat containing molecules (E). However if the rolling circle synthesized DNA contains a large number of repeats, PCR products with more than one repeat could easily be identified.

that premature termination and subsequent misalignment would keep occurring at the same positions during a number of cycles, using longer and longer misaligned molecules as primers, and this process could even take place at positions of no sequence identity. This interpretation becomes even more unlikely, since, at the same time, recombination events at different positions would also be occurring, resulting therefore in the generation of mixed molecules, i.e., molecules containing several DNA regions that had recombined at different positions. To account for the fact that repeats of the same recombined region were always present in the multi-repeat containing molecules, a more plausible scenario might be outlined as follows: during a single elongation step, the extending polymerase may switch template or, even more simply, just slip back to another position of the original template. This slippage event could occur several times during the elongation, because the newly synthesized strand, being complementary to the original template, would stabilize the loop structure that had been generated on the slipped template (Fig. 3). Therefore, following a rolling circle model of DNA synthesis, several identical repeats of the recombined region could be produced. In fact, this rolling circle model of template switching provides a reasonable interpretation that is consistent with the finding of the presence of the same type of repeat and not of combinations of different types of repeats in the multi-repeat containing PCR products.

Non-homologous recombination events, in a cell-free system, have also been observed with replicable RNA molecules (28). However, the mechanism proposed to rationalize these findings was suggested to be a splicing-like transesterification process that is guided by the secondary structure of the RNA molecule and not that of a template switching activity of the replicase enzyme. On the other hand, since these experiments involved amplification by replicase in the order of 10^{12}, followed by RT and then a two-step PCR, it is still conceivable that some of the recombined molecules detected may represent products of copy choice events.

In summary, evidence has been presented in this report that, *in vitro*, *Taq* polymerase has the capability to promote recombinational events that appear to follow a copy choice mechanism and, moreover, can take place at positions of no detectable homology. Given the accumulated evidence that template switching is a recombination process that may occur with a number of different polymerases, at homologous regions, it could be speculated that polymerases, in general, might have an inherent capability to slip, at a low frequency, to a different template or to a different position in the same template, even at non-homologous regions. Although the exponential increase of newly synthesized DNA during the PCR process facilitated the detection of such recombined molecules, the frequency of occurrence of these copy choice events is apparently quite low. If such phenomena do indeed occur *in vivo*, considerably more effort would be required to unambiguously identify them.

ACKNOWLEDGEMENTS

This study was supported by grants from the Swedish Natural Science Research Council, the foundations Lars Hiertas Minne and Åke Wiberg and the Karolinska Institute.

REFERENCES

1 Hall, S.L. and Padgett, R.A. (1996) *Science*, **271**, 1716–1718.
2 Tarn, W.-Y. and Steitz, J.A. (1996) *Cell*, **84**, 801–811.
3 Parker, R. and Siliciano, P.G. (1993) *Nature*, **361**, 660–662.
4 Scadden, A.D.J. and Smith, C.W.J. (1995) *EMBO J.*, **14**, 3236–3246.
5 Dietrich, R.C., Incorvaia, R. and Padgett, R.A. (1997) *Mol. Cell*, **1**, 151–160.
6 Nigro, J.M., Cho, K.R., Fearon, E.R., Kern, S.E., Ruppert, J.M., Oliner, J.D., Kinzler, K.W. and Vogelstein, B. (1991) *Cell*, **64**, 607–613.
7 Cocquerelle, C., Daubersies, P., Majérus, M.-A., Kerckaert, J.-P. and Bailleul, B. (1992) *EMBO J.*, **11**, 1095–1098.
8 Cocquerelle, C., Mascrez, B., Hétuin, D. and Bailleul, B. (1993) *FASEB J.*, **7**, 155–160.
9 Capel, B., Swain, A., Nicolis, S., Hacker, A., Walter, M., Koopman, P., Goodfellow, P. and Lovell-Badge, R. (1993) *Cell*, **73**, 1019–1030.
10 Zaphiropoulos, P.G. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 6536–6541.
11 Bailleul, B. (1996) *Nucleic Acids Res.*, **24**, 1015–1019.
12 Zaphiropoulos, P.G. (1997) *Mol. Cell. Biol.*, **17**, 2985–2993.

13 Braun, S., Domdey, H. and Wiebauer, K. (1996) *Nucleic Acids Res.*, **24**, 4152–4157.
14 Pasman, Z., Been, M.D. and Garcia-Blanco, M.A. (1996) *RNA*, **2**, 603–610.
15 Mikheeva, S., Hakim-Zargar, M., Carlson, D. and Jarrell, K. (1997) *Nucleic Acids Res.*, **25**, 5085–5094.
16 Chomczynski, P. and Sacchi, N. (1987) *Anal. Biochem.*, **162**, 156–159.
17 Pääbo, S., Irwin, D.M. and Wilson, A. (1990) *J. Biol. Chem.*, **265**, 4718–4721.
18 Jansen, R. and Ledley, F.D. (1990) *Nucleic Acids Res.*, **18**, 5153–5156.
19 Meyerhans, A., Vartanian, J.-P. and Wain-Hobson, S. (1990) *Nucleic Acids Res.*, **18**, 1687–1691.
20 Marton, A., Delbecchi, L. and Bourgaux, P. (1991) *Nucleic Acids Res.*, **19**, 2423–2426.
21 Odelberg, S.J., Weiss, R.B., Hata, A. and White, R. (1995) *Nucleic Acids Res.*, **23**, 2049–2057.
22 Meselson, M. and Weigle, J.J. (1961) *Proc. Natl. Acad. Sci. USA*, **47**, 857–868.
23 Tease, C. (1978) *Nature*, **272**, 823–824.
24 Papanicolaou, C. and Ripley, L.S. (1989) *J. Mol. Biol.*, **207**, 335–353.
25 Madsen, C.S., Ghivizzani, S.C. and Hauswirth, W.W. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 7671–7675.
26 Negroni, M., Ricchetti, M., Nouvel, P. and Buc, H. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 6971–6975.
27 Canceill, D. and Ehrlich, S.D. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 6647–6652.
28 Chetverin, A.B., Chetverina, H.V., Demidenko, A.A. and Ugarov, V.I. (1997) *Cell*, **88**, 503–513.
29 Gordenin, D.A., Malkova, A.L., Peterzen, A., Kulikov, V.N., Pavlov, Y.I., Perkins, E. and Resnick, M.A. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 3785–3789.
30 D'Alencon, E., Petranovic, M., Michel, B., Noirot, P., Aucouturier, A., Uzest, M. and Ehrlich, S.D. (1994) *EMBO J.*, **13**, 2725–2734.
31 Lai, M.M.C (1992) *Microbiol. Rev.*, **56**, 61–79.