Gene, 164 (1995) 59–63

GENE 09101

Template secondary structure can increase the error frequency of the DNA polymerase from *Thermus aquaticus*

(*Taq* polymerase; PCR; fidelity; stem-loop)

Peter C. Loewen and Jacek Switala

Department of Microbiology, University of Manitoba, Winnipeg, Manitoba R3T2N2, Canada

Received by J. Wild: 17 February 1995; Revised/Accepted: 23 May/25 May 1995; Received at publishers: 29 May 1995

SUMMARY

Amplification of portions of the intergenic spacer between the *katE* gene and cryptic *cel* operon of *Escherichia coli* was accomplished by the polymerase chain reaction using the DNA polymerase from *Thermus aquaticus*. Nine different segments were amplified and cloned without error, but one 83-bp fragment was amplified with a high error rate such that 32 of 34 selected clones had three or more nucleotide changes from the expected sequence. The changes were all located in two 9-bp segments immediately adjacent to the 3'-ends of the two primers. Moving the end points of the primers to increase the spacing between them resulted in the isolation of significantly fewer error-containing products. It is proposed that stem-loop structures in the template immediately downstream from the primers interfere with an early stage of elongation and cause misincorporation. This is supported by the observation that destabilisation of one of the stem-loop structures reduced the frequency of errors.

INTRODUCTION

The heat-stable *Taq* DNA polymerase, from *Thermus aquaticus*, has facilitated the polymerase chain reaction (PCR) and its myriad of variants. One concern that has arisen from the use of this enzyme is the potential for mutations in amplified products arising with the higher than normal frequency which has been estimated to be between 10^{-5} and 10^{-4} errors per bp, depending on reaction conditions and target sequences (Eckert et al., 1990; Cariello et al., 1991; Sandhu and Keohavong, 1994) as compared to 3×10^{-6} for T4 DNA polymerase, 3.4×10^{-5} for modified T7 DNA polymerase, and 1.3×10^{-4} for PolIk. Even at the high end of the error frequency spectrum, *Taq* polymerase should cause less than one mutation per 100 bp after 10^6 amplification, which for many applications will pose no problem, but will necessitate a rigorous sequence confirmation of PCR amplified and cloned fragments, and the sequencing of multiple clones where the sequence is not already known.

The *katE* gene and the cryptic *cel* operon of *Escherichia coli* are oriented in a convergent manner and are separated by a spacer of 254 bp (von Ossowski et al., 1991; Guzzo and DuBow, 1994). As part of a study of transcription termination in this spacer region, we have cloned segments varying in length from 80 to 200 bp using PCR amplification. The characterization of PCR-amplified products revealed that nine of ten products from different primer combinations were produced without error. However, one set of primers designed to produce an 83-bp duplex generated a mixture of products of which

Correspondence to: Dr. P.C. Loewen, Department of Microbiology, University of Manitoba, Winnipeg, Manitoba R3T 2N2, Canada. Tel. (1-204) 474-8334; Fax (1-204) 275-7615;
e-mail: peter_loewen@umanitoba.ca

Abbreviations: bp, base pair(s); *cel*, cryptic genes encoding a cellulase; *katE*, gene encoding catalase HPII; kb, kilobase(s) or 1000 bp; nt, nucleotide(s); PCR, polymerase chain reaction; PolIk, Klenow (large) fragment of *E. coli* DNA polymerase I.

SSDI 0378-1119(95)00383-5

94% contained three or more nt changed from the expected sequence. The aim of this paper is to describe the extent of and possible reasons for the high error frequency in this particular PCR product.

EXPERIMENTAL AND DISCUSSION

(a) Identification of error-containing PCR products

In the study of transcription termination in the *katE-cel* intergenic spacer, a total of ten different primer combinations were employed for the PCR amplification of different parts of the spacer using *Taq* polymerase. All of the products were characterised for correct sequence and orientation, revealing that nine of the primer combinations yielded the desired products with no errors, but one combination generated error-containing products at a very high frequency. The primers in question are shown in Fig. 1 as R1 and L1. From three separate amplification reactions, 23 separate clones were chosen for characterisation, but only one contained the correct sequence (Table I).

All of the primers contained a double *Bam*HI site at their ends and, even though they did not give rise to errors in nine of the ten cases, we wanted to eliminate the possibility that this feature was a factor in the error generation with R1 and L1. Primer R2, which differs from R1 at three positions near the 5′-end eliminating the second *Bam*HI site and making it a better complement to the intergenic spacer (only 4 of 32 bp mismatched) was prepared. As shown in Table I, only one of eleven selected products from the amplification involving primers R2+L1 had the correct sequence, and sequencing revealed that the products were similar to those created in the R1+L1-primed amplification (including some identical products), leading to the conclusion that the tandem *Bam*HI sequence was not a factor in the high error frequency. Thus, from the two combinations of R1+L1 and R2+L1 used in four separate amplification reactions, 34 different product clones were selected and sequenced, of which 32 (94%) contained incorrect sequences and only two (6%) had the correct sequence (Table I).

(b) Comparison of error-containing sequences

A survey of the sequences in the 34 selected product clones revealed 19 different sequences which are shown in Fig. 1. The upper strand of the duplex with the expected sequence is oriented 5′ to 3′, left to right, as are the 19 sequences below. Correctly incorporated nt in the 19 sequences are shown as dots as in sequence No. 1 where there are no errors. The minimum number of errors

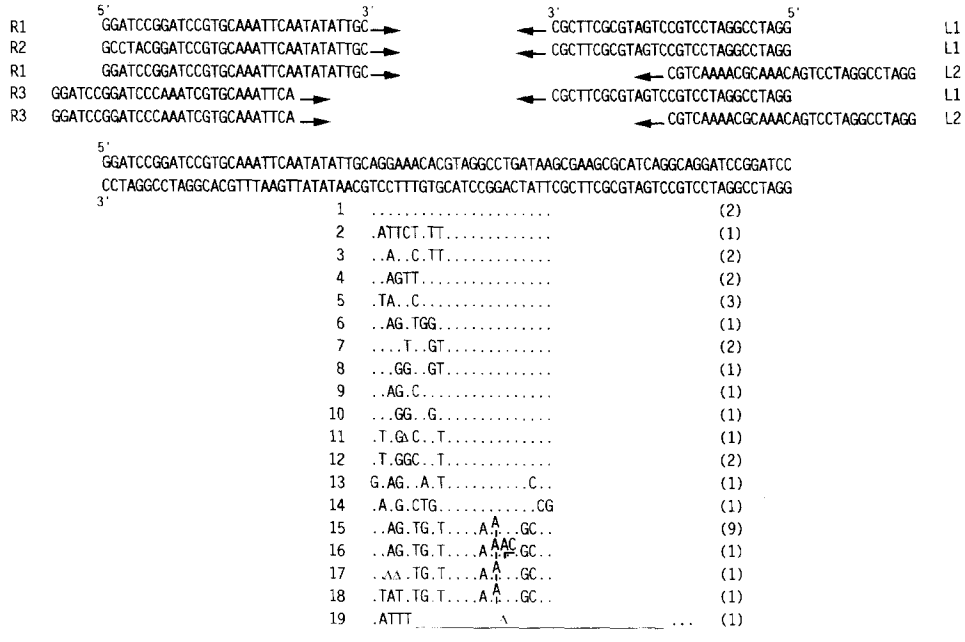

Fig. 1. Sequences of the primers used in the PCR amplification reactions and their relationship relative to the 83-bp product that would be formed using primers R1 and L1. Primers R1, R2 and R3 correspond to the upper strand in the product duplex, and primers L1 and L2 correspond to the lower strand of the product duplex. The arrows at the 3′ end of each primer indicate the direction of elongation by *Taq* polymerase. The sequences of the clones from the amplification with primer pairs R1+L1 and R2+L1 are indicated below the duplex and are numbered on the left. The numbers in parentheses on the right indicate the number of times a particular product was found in the pools sequenced. The sequences correspond to the upper strand of the duplex. Correctly incorporated nt are indicated by a dot and misincorporated nt are indicated with letters. Deletions in products 11 and 17 are indicated by a Δ and the 31-bp deletion in product 19 is indicated by a solid line with a Δ above. The sequences above the line in products 15 to 18 indicate inserted nt. The isolation and sequencing of the products are described in the footnotes to Table I.

TABLE I

Frequency of products with incorrect sequences

Amplification reaction[a]	Primer[b]	No. of products sequenced[c]	No. of products with errors[d]
1	R1+L1	6	6
2	R1+L1	9	9
3	R1+L1	8	7
4	R2+L1	11	10
5	R1+L2	13	2
6	R3+L1	12	0
7	R3+L2	6	0

[a] Samples were denatured for 1 min at 93°C, annealed at 53°C for 1 min and elongated at 72°C for 2 min. Reactions were carried out using enzyme and stock reagents supplied by Promega (Madison, WI, USA). The final reaction mixture contained 50 mM KCl/1.5 mM $MgCl_2$/10 mM Tris·HCl (pH 9.0 at 25°C)/0.1% Triton X-100/0.2 mM of each of dATP, dGTP, dTTP and dCTP/2 units of *Taq* polymerase per 100 µl.

[b] The primers are described in Fig. 1.

[c] Products from the PCR amplification were separated on a 2% agarose gel and the region of the gel containing the amplified fragment was excised. The DNA was isolated, treated with *Bam*HI (Maniatis et al., 1982) and separated on a 2% agarose gel. The isolated DNA was incorporated into a *Bam*HI site between the *katE* promoter and *lacZ* in the plasmid pET1 constructed by inserting a 3.4-kb *Eco*Ri-*Sst*I fragment from pRSkatE16 (containing a *katE-lacZ* fusion; Mulvey et al., 1990) into pSK+(Stratagene, La Jolla, CA, USA). The sequences were determined (Sanger et al., 1977) using the primer 5'-CTCTACCGTTCAGGGTGGT encompassing bp 1318–1336 of the *katE* sequence (von Ossowski et al., 1991), 27 bp upstream from the *Bam*HI site in pET1.

[d] The number of products with sequences that differed from the expected sequence in Fig. 1 (von Ossowski et al., 1991) are listed. The sequences of these products are contained in Fig. 1.

in the other 18 is three which is found in four different products (No. 5, 6, 9 and 10). Product No. 19 contains a deletion of 31 bp and product No. 18 contains nine substitutions and one insertion. Not including product No. 19, the average number of errors including changes, deletions and insertions, is 5.4 per 83-bp product. Several of the products were isolated from more than one amplification mixture and only one of the products (No. 15) was isolated more frequently than three times.

A survey of the 19 sequences reveals that the errors are grouped in two separate 9-bp stretches separated by 4 bp. The group of errors in the left-side 9-bp stretch (including nt 33 to 41 from the left end of the duplex in Fig. 1) are present in all of the sequences, but errors in the right-side 9-bp stretch (from nt 46 to 54) are present in only a subset of six sequences. Furthermore, there are obvious similarities or patterns in the errors that suggest a gradual incorporation of errors during sequential rounds of amplification. This is particularly obvious in the group including Nos. 15, 16, 17 and 18 and the group including Nos. 6, 8, 9 and 10.

(c) Possible role of stem-loop structures in misincorporation

Such precise localisation of the errors suggests that there is a unique structural feature in the primer-template duplex that is affecting the polymerase. The structures of the primer-template pairs are presented in Fig. 2 with primer R1 annealed to an 83-bp template strand in Fig. 2A, and primer L1 annealed to the complementary 83-bp sequence in Fig. 2B. Included in these diagrams are potential stem-loop structures immediately downstream from the primers. The calculated free energies of the loops are −9.0 kcal/mol and −7.7 kcal/mol for the loops in Fig. 2A and B, respectively, which suggests that both loops will be present in significant amounts at 72°C, the elongation temperature used for amplification. In both cases the error-prone regions are found in the first 9 bp after the primer as indicated by the sequence above or below the dashed lines. There was considerable variation in the sequence in Fig. 2A as indicated by the X's, but the misincorporation in Fig. 2B was unique at each location as shown.

We are proposing that the stem-loop structures in such close proximity to the initiation site interfere with normal elongation in the early rounds of nt incorporation. Two further observations are consistent with this proposal. First, increasing the separation between the starting points (primer ends) and the stem-loops by 9 or 14 bp, as with primer combinations R1+L2, R3+L1 and R3+L2 (Fig. 1), significantly decreases the accumulation of error-containing products (Table I). Second, there are more error-containing products selected in Fig. 2A where the stem-loop is 13 nt from the primer as compared to Fig. 2B where the separation is 19 nt. The simplest picture is that the stem-loop can prevent normal progress of the enzyme just after initiation resulting in prolonged residency by the enzyme in the first incorporation sites, while it waits for the stem-loop to melt. Why such pauses early in the elongation process should allow significant misincorporation is not clear. Once the enzyme is elongating normally, the stem-loop structures do not present such a barrier.

The long deletion in product No. 19 requires a slightly different explanation, but it also involves the stem-loop structures. This explanation is based on an earlier report of a stem-loop structure being implicated in a 54-bp deletion from the human 18S rRNA gene wherein the complete stem-loop was excised as *Taq* polymerase replicated across the bottom of the stem (Cariello et al., 1991). In product No. 19 the excision includes the complete stem-loop and also some of the sequence between the primer and the loop.

(d) Timing of error incorporation

It is not clear from the pattern of errors whether the errors all occurred in a few of the very early rounds of

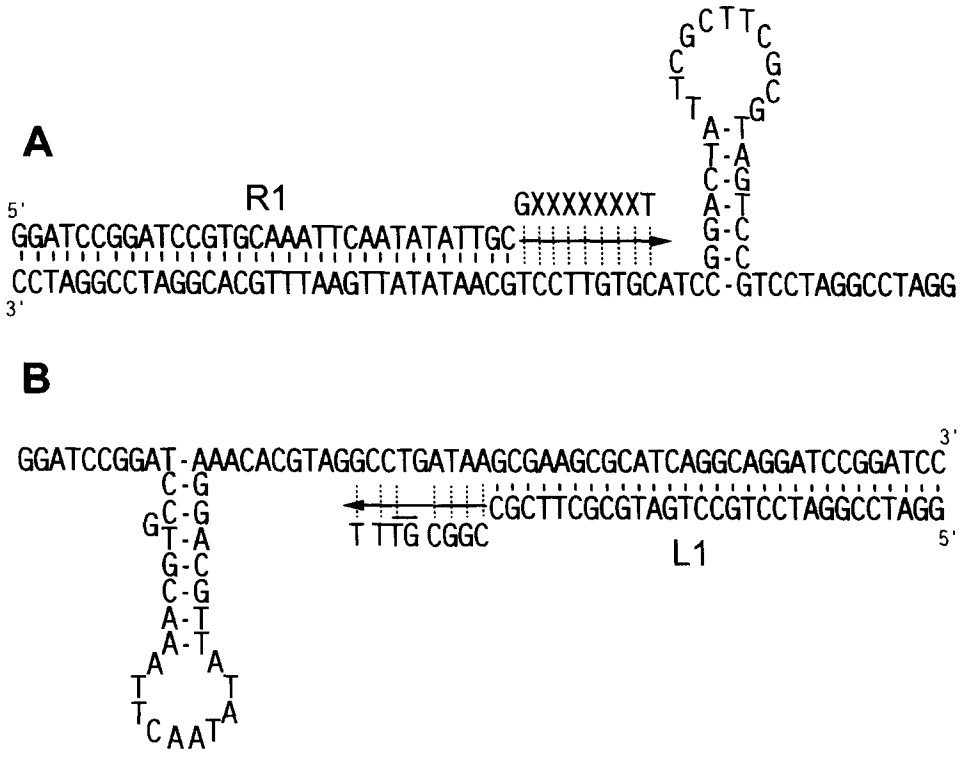

Fig. 2. Alignment of the primers R1 (**A**) and L1 (**B**) on their complementary single strand template with the potential stem-loop structures in the template shown. The location and identity of the misincorporated nt is indicated by the letters above (in **A**) or below (in **B**) the dashed lines at the 3′ ends of the primers. In panel **A** there was heterogeneity in the misincorporated nt every site but two and the mistakes are represented by **X**. In panel **B** the mistakes at each location were unique and are indicated as such. Also in panel **B** the dashed lines that point between the bases in the template are the locations of the inserted sequences shown.

amplification followed by later error free rounds of amplification, or they accumulated gradually over many rounds of amplification. In the first possibility, the incorporation of errors early in the process might eliminate or destabilise the element responsible for the errors, thereby facilitating error-free elongation in subsequent rounds of amplification. To test this possibility, we used DNA from products No. 3 and No. 7, which contained four and three errors respectively compared to the original sequence, as templates for amplification with primers R1 + L1. While the sample size is smaller, the accumulation of errors remains substantial with three of six products from template No. 3, and four of five products from template No. 7 containing additional errors (Table II). Clearly, these additional errors are being incorporated despite the 'errors' already present in templates, suggesting that the changes in Fig. 1 are incorporated gradually throughout the amplification process.

The results in Table II also provide additional support for the conclusion that stem-loops are involved in the generation of errors. While still substantial at 50%, the proportion of error-containing products is clearly lower with template No. 3 than with template No. 7. Furthermore, the sequence changes with template No. 3 are contained in just two different products whereas each product from template No. 7 contained a different sequence. These results are consistent with the G to A change in template No. 3 having destabilised the stem-loop in Fig. 2B resulting in fewer errors, and with the changes in template No. 7 having had no effect on the stem-loops.

Because it was not possible to obtain clear sequencing information from the mixture of products immediately following PCR amplification, a cloning step was necessary to isolate individual products. This gives rise to the possibility of a bias in the cloning step against the selection of the wild type sequence. While this cannot be excluded from the data presented, it seems unlikely for three reasons. First, extension of the PCR fragment at one end by only 9 bp (using primers R3 + L1) or the other by 14 bp (using primers R1 + L2) leaving the remainder of the sequence unchanged eliminated the selection of error-containing fragments. Second, the changes in the modified sequences represent only a small proportion of the PCR fragment (3 to 8 bp of an 83-bp duplex) and most do not affect the inverted repeat sequences which is the only feature that might affect cloning. Third, errors continue to be selected even when previously selected error-containing templates were used (see Table II).

TABLE II

Frequency of errors and sequence identity when products #3 and #7 from Fig. 1 were used as templates for amplification with primers R1+L1

Number screened[a]	Number with additional errors[b]	Sequences[c] (5′→3′)	Repeats[d]
		Template No. 3	
6	3	No. 3 CAGAAACCTTGTAGGCCTGATAA	3
		No. 3-1 ....G.TTGC.....	2
		No. 3-2 ....G.TGA.....^{A}...GC..	1
		Template No. 7	
5	4	No. 7 CAGGATACGTGTAGGCCTGATAA	1
		No. 7-1 .....C.....	1
		No. 7-2 ...T.C..T.....	1
		No. 7-3 ..A.T.T.....	1
		No. 7-4 .....A.....	1

[a] PCR reaction, isolation of amplified products and sequencing are described in the footnotes to Table I.

[b] The number of products with sequences that differed from the starting template is indicated.

[c] The sequences of the templates, either No. 3 or No. 7, are provided at the top and correctly incorporated nt are shown with a dot. Incorrectly incorporated nt are specified in capitals and the single A shown above the line in No. 3-2 indicates an inserted nt at that position.

[d] Indicates the number of times a given sequence appeared in the pool of characterized products.

(e) Conclusions

(*1*) PCR amplification using one particular primer combination produced a mixture containing a high proportion of error-containing products.

(*2*) Separation of the primer 3′-ends from a putative stem-loop structure reduced the appearance of error-containing products.

(*3*) Destabilization of the putative stem-loop structures by mutation also reduced the appearance of error-containing products.

(*4*) Errors accumulated gradually throughout the amplification process.

ACKNOWLEDGEMENT

This work was supported by grant OGP 9600 from the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

Cariello, N.F., Swenberg, J.A. and Skopek, T.R.: Fidelity of *Thermococcus litoralis* DNA polymerase (Vent™) in PCR determined by denaturing gradient gel electrophoresis. Nucleic Acids Res. 19 (1991) 4193–4198.

Cariello, N.F., Thilly, W.G., Swenberg, J.A. and Skopek, T.R.: Deletion mutagenesis during polymerase chain reaction: dependence on DNA polymerase. Gene 99 (1991) 105–108.

Eckert, K.A. and Kunkel, T.A.: High fidelity DNA synthesis by the *Thermus aquaticus* DNA polymerase. Nucleic Acids Res. 18 (1990) 3739–3744.

Guzzo, A. and DuBow, M.S.: A *luxAB* transcriptional fusion to the cryptic *celF* gene of *Escherichia coli* displays increased luminescence in the presence of nickel. Mol. Gen. Genet. 242 (1994) 455–460.

Maniatis, T., Fritsch, E.F. and Sambrook, J.: Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982.

Mead, D.A., Skorupa, E.S. and Kemper, B.: Single-stranded DNA SP6 promoter plasmids for engineering mutant RNAs and proteins: synthesis of a "stretched" preparathyroid hormone. Nucleic Acids Res. 13 (1985) 1103–1118.

Mulvey, M.R., Switala, J., Borys, A. and Loewen, P.C.: Regulation of transcription of *katE* and *katF* in *Escherichia coli*. J. Bacteriol. 172 (1990) 6713–6720.

Sandhu, D.K. and Keohavong, P.: Effects of the T4 bacteriophage gene 32 product on the efficiency and fidelity of DNA amplification in using T4 DNA polymerase. Gene 144 (1994) 53–58.

Sanger, F., Nicklen, S. and Coulson, A.R.: DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74 (1977) 5463–5467.

von Ossowski, I., Mulvey, M.R., Leco, P.A., Borys, A. and Loewen, P.C.: Nucleotide sequence of *Escherichia coli katE*, which encodes catalase HPII. J. Bacteriol. 173 (1991) 514–520.