scientific reports

OPEN

Clamping-mediated incorporation of single-stranded DNA with concomitant DNA synthesis by Taq polymerase involves nick-translation

Yoshiyuki Ohtsubo✉, Syoutaro Kawahara & Yuji Nagata

The development and characterization of a new enzyme reaction contribute to advancements in modern biotechnology. Here, we report a novel CIS (clamping-mediated incorporation of single-stranded DNA with concomitant DNA synthesis) reaction catalyzed by Taq polymerase. In the reaction, a single-stranded DNA (ssDNA) with 3′ Cs is attached with a preformed 3′ G-tail of double-stranded DNA (dsDNA); DNA syntheses starting from both 3′ ends result in the incorporation of ssDNA. A 3′ G-tail length of 3 nucleotides adequately supports this reaction, indicating that Taq polymerase can clump short Watson–Crick base pairs as short as three pairs and use them to initiate DNA polymerization. The reverse transcriptase from Molony murine leukemia virus catalyzes strand displacement synthesis and produces flapped-end DNA, whereas the reaction by Taq polymerase involves the nick translation. These new reaction properties may be beneficial for the development of new molecular tools applicable in various fields. Apart from its CIS reaction activity, we also report that Taq polymerase has the undesirable characteristic of removing 5′ fluorescent labels from dsDNA. This characteristic may have compromised various experiments involving the preparation of fluorescently-labeled dsDNA by PCR for a long time.

Several enzymatic reactions are often used in various biological experiments. Enzymes that recognize and catalyze DNA substrates are essential molecular tools frequently used in modern biotechnology.

Several DNA polymerases are used in the tailing reaction that adds one or more nucleotides to the 3′ end of double-stranded DNA or DNA-RNA hybrids, forming a protruding end. The well-known Taq polymerase developed from *Thermus aquaticus*[1] mediates this reaction and creates a 3′ protruding end consisting of a single adenosine monophosphate. In contrast, terminal deoxynucleotidyltransferase adds a large number of nucleotides in an uncontrolled manner[2]. We previously reported that Molony murine leukemia virus reverse transcriptase (MMLV-RT)[3] is substantially efficient in tailing double-stranded DNA (dsDNA). MMLV-RT can add several As, Cs, Gs, or Ts to the 3′ end of DNA molecules[4]. Moreover, specific tailing enhancers for C, G, and T-tailing reactions, such as deoxyguanosine, deoxycytidine, and adenosine, respectively, which facilitate the addition of 3–5 nucleotides (As, Cs, Gs, or Ts), were also reported (see reference[5] for proposed mechanisms of how the enhancers work). The protruding tails are utilized for various purposes, including TA cloning[6], CG cloning[7], DNA labeling[8,9], and the CIS reaction.

Besides the strong tailing capacity, the unique effectiveness of MMLV-RT includes its catalytic activity in the template-switching reaction. It appends a few Cs to the 3′ end of the newly synthesized DNA at the end of the RNA template. In the presence of G-tailed single-strand DNA (ssDNA), MMLV-RT can switch the template and polymerize the complementary strand of incoming ssDNA[10,11]. This template-switching reaction has been widely used in various studies to analyze RNA-seq[12].

Some reverse transcriptases, including MMLV-RT, can incorporate single-stranded DNA (ssDNA) into double-stranded DNA (dsDNA) based on complementarity as short as one nucleotide. These reverse transcriptases are considered to have clumping activity, an ability to stabilize the very short complementarity and aid the synthesis of the complementary strand of ssDNA[13–15]. However, the reaction was rarely used owing to

Department of Molecular and Chemical Life Sciences, Graduate School of Life Sciences, Tohoku University, 2-1-1 Katahira, Sendai 980-8577, Japan. ✉email: yoshiyuki.ohtsubo.a6@tohoku.ac.jp

its limited efficiency until we developed a method to create long protruding 3′ ends[5]. The long tail enabled the formation of thermodynamically stable Watson–Crick base pairings and improved the reaction efficiency up to approximately 100%. We marked this reaction as the CIS (clamping-mediated incorporation of single-stranded DNA with concomitant DNA synthesis) reaction[16].

The MMLV-RT-mediated CIS reaction using dsDNA with four Gs at the tail and ssDNA with multiple Cs at the 3′ end was highly effective, and all substrate DNAs were converted into products[16,17]. Notably, the ssDNA can carry a stretch of Ns, which can be used as a unique molecular identification tag[18–20]. These tailing and CIS reactions are highly efficient and can incorporate any desired sequence (up to 70 nucleotides long) into the ends of dsDNA with nearly 100% efficiency[16]. A potential disadvantage of the use of MMLV-RT in CIS reaction is that it catalyzes strand-displacement synthesis producing flapped-end DNA (Fig. 1)[16].

The addition of four Gs to the 3′ end of double-stranded DNA represents a novel reaction[5]. Methods for utilizing such ends, for example, in the labeling of DNA ends with UMIs, have not been fully explored. To pioneer new applications and to investigate whether clumping activity is unique to some reverse transcriptases, we evaluated the efficacies of different DNA polymerases in catalyzing CIS reactions. In this study, we investigated the ability of Taq polymerase to catalyze a different type of CIS reaction that develops nick-translated products[21] and described the optimum conditions required for such reactions.

Materials and methods

Reagents

Taq polymerase with the reaction buffer was obtained from New England Biolabs (Ipswich, Massachusetts). DNA ligation mix, Exonuclease I, and T4 DNA polymerase were purchased from TAKARA (Shiga, Japan).

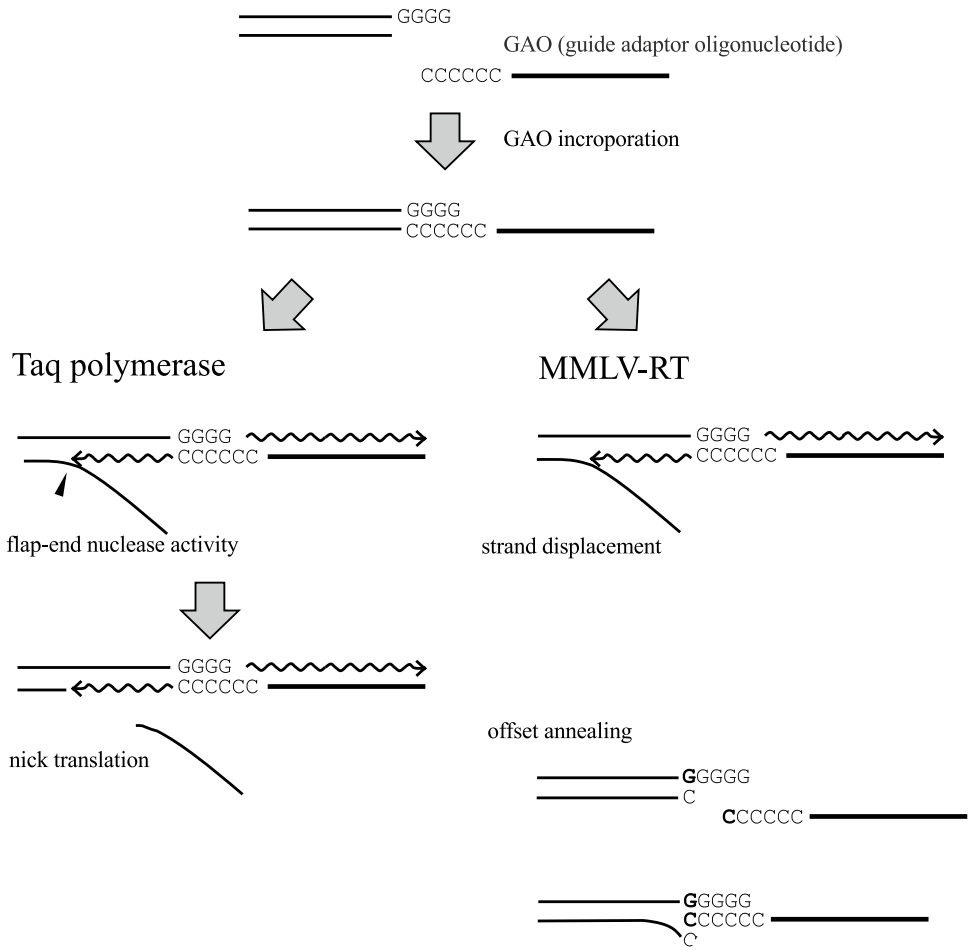

Figure 1. Schematic representation of CIS reactions mediated by Taq polymerase and MMLV-RT. To a dsDNA substrate carrying a 3′ G-tail, ssDNA with 3′ Cs (GAO; guide adaptor oligonucleotide) was attached via Watson–Crick base pairings. Taq polymerase and MMLV-RT catalyze DNA polymerization starting from the 3′ end of the G-tail, as well as the 3′ end of GAO, resulting in the incorporation of GAO. While Taq polymerase catalyzes nick translation (left), MMLV-RT causes strand displacement synthesis (right). In this study, MMLV-RT was postulated to have undergone offset annealing and extension (bottom right).

DNA length analysis using a capillary sequencer

Length of DNA was analyzed as described previously[4]. To prepare HiDi-LIZ1200, 1 ml of HiDi formamide was mixed with 10 µL of GeneScan-1200 LIZ Size Standard (Thermo Fisher Scientific, USA) containing 68 fragments of known sizes. Subsequently, 0.2–1.0 µl of the reaction mixture or diluted purified DNA samples not exceeding 0.5 fmol were added to 12.5 µl of HiDi-LIZ1200, as appropriate. The samples were analyzed using a 3130xl Genetic Analyzer (Thermo Fisher Scientific, Massachusetts, USA) with a 50-cm capillary array and POP7 polymer. The data obtained were analyzed using the TraceViewerForMolecularBiology software (our product available at App Store, Apple Inc., California)[4]; two LIZ bands were chosen to calibrate the electropherogram. Peak areas were determined using the software.

Preparation of DNA substrates

DNA substrate 1 (33G1, 33G2, 33G3, and 33G4) was adopted from a previous report[16]. To prepare substrate 2, SA680 (FAM-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGGG) and SA760 (5′-GAGTTCAGA CGTGTGCTCTTCCGATCTCCCCCCCTGTCTCTTATACACATCTGACGCTGCCGACGA-3′) were annealed. Substrate 3 was prepared by annealing SA761 (5′-CTGTCTCTTATACACATCTGACGCTGCCGACGA-3′-FAM) and SA763 (5′-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGGG-3′) and purified using polyacrylamide gel electrophoresis.

Taq and MMLV-RT CIS reactions

A total volume of 10 µL CIS reaction mixture contained 10 fmol DNA substrate, 0.25 mM each of dNTPs (TAKARA, Kyoto, Japan), and 2 pmol guide adaptor oligonucleotide (GAO). For the Taq CIS reaction, we used 0.5–1.25 units of Taq polymerase (0.1–0.25 µl). For the MMLV-RT CIS reaction, 50 units of MMLV-RT (0.25 µl) (Nippon Gene, Tokyo, Japan) were used with the buffer containing 50 mM Tris–HCl pH 8.3, 75 mM, KCl, 6 mM $MgCl_2$, and 2 mM DTT. To initiate the reactions, after incubation at 37 °C or temperatures as indicated, components of the reaction mixture were added to the premixed GAO and substrate DNA.

Other enzyme reactions

For the Exonuclease I reaction, 0.1 µl of the enzyme solution was added to the post-CIS reaction mixture and incubated for 3–5 min at 37 °C. For the blunting reaction using T4 DNA polymerase, the DNA in the post-CIS reaction mixture was column-purified and then used for the reaction.

DNA column purification

For DNA fragment purification, we used PCR Purification Kit (HiYield™ Gel/PCR DNA Fragments Extraction Kit) from RBC Bioscience.

Results

Taq DNA polymerase catalyzes the CIS reaction

We evaluated the efficiencies of several polymerases, including Taq polymerase, KOD DNA polymerase[22], DNA polymerase I, Bst DNA polymerase, and a mutant Klenow fragment that lacks 3′ → 5′ exonuclease activity in the CIS reaction. Notably, only Taq polymerase and Klenow fragment (3′ → 5′ exo-) were sufficiently effective in conducting the CIS reaction. Taq polymerase conducts a unique CIS reaction as described in the text; hence, we focused on this enzyme. Figure 1 (top to left) displays the overall features of the CIS reaction mediated by Taq polymerase. In this reaction, a single-stranded DNA called a guide adaptor oligonucleotide (GAO) anneals to the 3′ protruding G-tail of a double-stranded DNA. Subsequently, Taq polymerase catalyzes the synthesis of the complementary strand, starting from the 3′ OH end of the protruding G and the 3′ end of GAO. As the GAO strand extends, the strand ahead of the polymerase is degraded by the flap-end nuclease activity of Taq polymerase[21], leading to the nick translation[23]. In contrast, displacement synthesis occurs in the previously known reaction mediated by MMLV-RT, which lacks the 5′ → 3′ exonuclease activity (top to right)[16].

The following data demonstrate these features and optimal conditions for the reaction with Taq polymerase.

Top strand extension along with the GAO by Taq polymerase

To prepare the CIS reaction substrate (substrate 1), we annealed a 5′-FAM-labeled oligonucleotide SA680 (37 nucleotides) with SA659 (33 nucleotides). Substrate 1 was a 33 bp double-stranded DNA with four 3′-extruding Gs, which is identical to the previously reported 33G4[16]. We annealed SA680 with SA760 (66 nucleotides) to prepare substrate 2 for the Taq extension reaction.

GAO (SA727 with six Cs added at its 3′ end), Taq polymerase, and dNTPs were added to substrate 1 and incubated. The change in FAM-labeled strand length was analyzed using a capillary sequencer under denaturing conditions. As compared to the initial substrate (Fig. 2a), the Taq treatment extended the FAM-labeled strand (Fig. 2b). However, the peak pattern observed when the reaction solution was directly sampled was unexpected and showed two separated peaks (Fig. 2b). This unexpected pattern was likely due to the interference of the residual GAO with the FAM-labeled product DNA during capillary electrophoresis. To support this, degradation of the residual GAO by Exonuclease I (Fig. 2c) or removal by column purification (Fig. 2d) resulted in a single major peak. This interference was also verified by the reaction using substrate 2 (Fig. 2f) and Taq polymerase (Fig. 2g) and the subsequent addition of SA727 (Fig. 2h). The sizes of the two reaction products were indistinguishable (Fig. 2d and i), and the subsequent blunting reaction shifted the peak to the left by one nucleotide, indicating their identical sizes (Fig. 2e and j). The shift reflects the removal of a single A overhang appended by Taq polymerase.

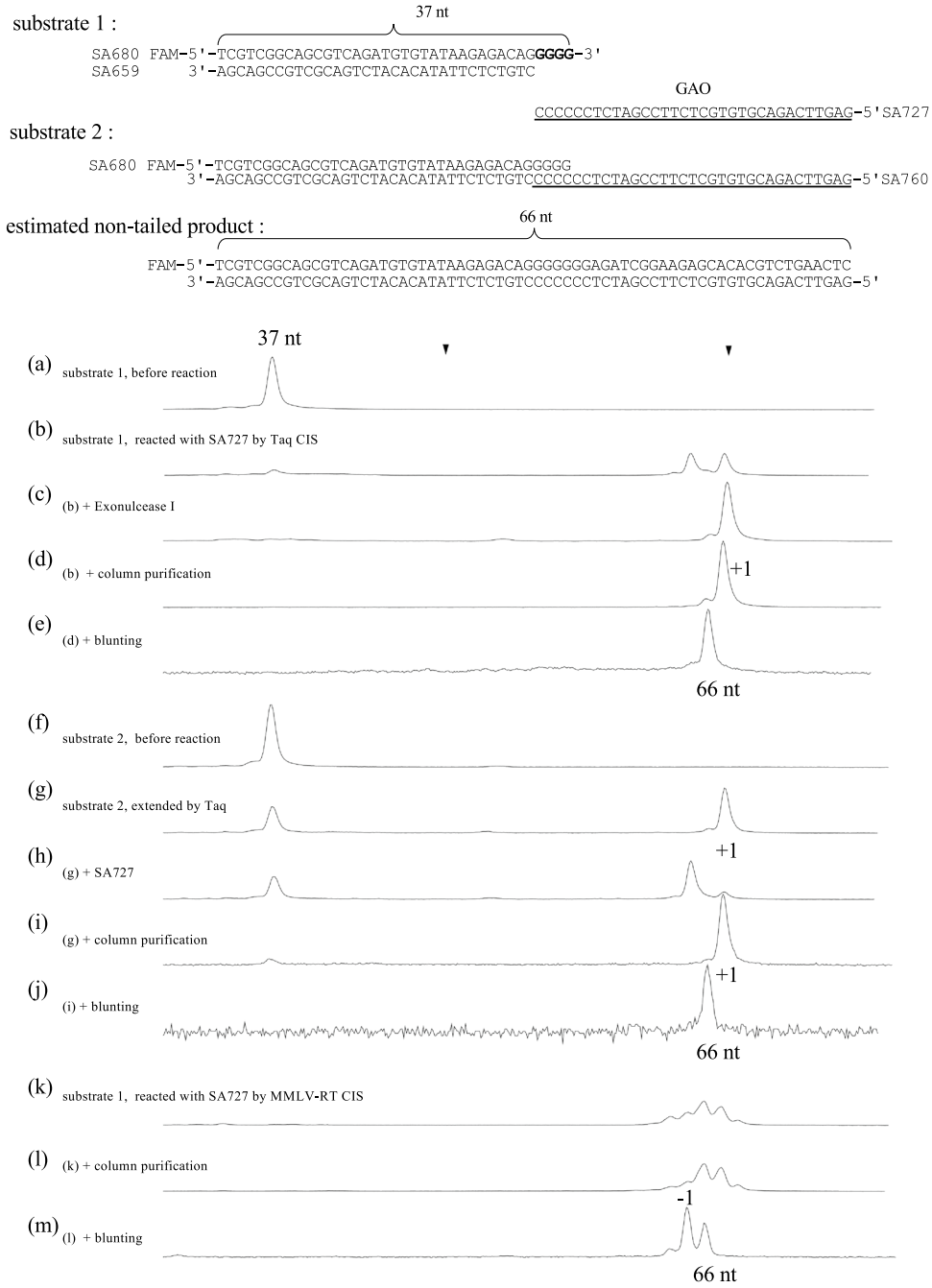

Figure 2. Taq polymerase conducts the CIS reaction. To a mixture of substrate 1 (top) and GAO (SA727), dNTP and Taq polymerase were added to initiate the CIS reaction. A fraction of the reaction mixture was sampled for the HiDi LIZ1200 before (a) and after the CIS reaction (b). DNAs in the post-reaction mixture were treated with Exonuclease I (c) or purified by column purification (d). The purified sample was used for the DNA blunting reaction with T4 DNA polymerase (e). Substrate 2 was used as a control for the canonical Taq extension reaction. Samples were collected before (f) and after (g) the Taq extension. The addition of SA727 to the post-reaction mixture (h) gave an electropherogram similar to that in (b). The DNAs in the post-reaction mixture were purified by column purification (i) and subjected to blunting reaction (j). Panels k, l, and m show the MMLV-RT CIS reaction products immediately after reaction (k), post-reaction after column purification (l), and after the blunting reaction (m). Two triangles indicate the LIZ size marker peaks (not shown) used to calibrate the electropherograms. An estimated non-tailed product of 66 nucleotides is indicated.

As a reference, the MMLV-RT CIS reaction products before (Fig. 2k) and after (Fig. 2l) the column purification are shown. Notably, the blunting reaction with T4 DNA polymerase presented a major peak representing a strand size one nucleotide shorter than expected (labeled as -1), which is discussed later in the text.

GAO strand elongation and nick translation

Taq polymerase has an intrinsic structure-dependent endonuclease activity that cleaves the displaced strand (also referred to as flapped-end DNA), exhibiting a phenomenon known as nick translation. We examined whether nick translation occurs during GAO strand elongation. We prepared a substrate DNA (substrate 3), which was FAM-labeled at the 3′ end of the bottom strand (Fig. 3a), and conducted a CIS reaction using Taq polymerase. Consequently, a group of short fragments was produced (Fig. 3b), indicating that the FAM-labeled strand was shortened by nick translation. The short fragments were sensitive to Exonuclease I (Fig. 3c), indicating that they are single-stranded DNAs that dissociate from double-stranded DNA before cleavage by the flapped-end nuclease activity of the Taq polymerase. Such short fragments were not generated if the substrate DNA was incubated in the absence of GAO (Fig. 3d) or MMLV-RT was used (Fig. 3e). Based on these results and previously known characteristics of Taq polymerase, we conclude that nick translation occurs at the bottom strand in the Taq polymerase-mediated CIS reaction.

Effect of reaction temperature

To establish the CIS reaction with Taq polymerase and enhance its usefulness as a DNA-manipulating tool, we optimized the reaction temperature. The CIS reaction product was effectively generated at a temperature ranging between 37 and 42 °C (Fig. 4). At higher temperatures (42 °C and 50 °C), a broad peak possibly representing FAM plus one nucleotide was observed (see Fig. 6a,b and relevant discussion).

3′ G-tailing length dependency

We analyzed whether the length of the 3′ G-tail influences the efficiency of the reaction (Fig. 5). We used four substrate DNAs with G-tails ranging from one to four nucleotides in length and conducted the Taq CIS reaction at 30 °C. G-tails of 3 and 4 nucleotides facilitated the generation of a product, and four-nucleotide-long G-tailed

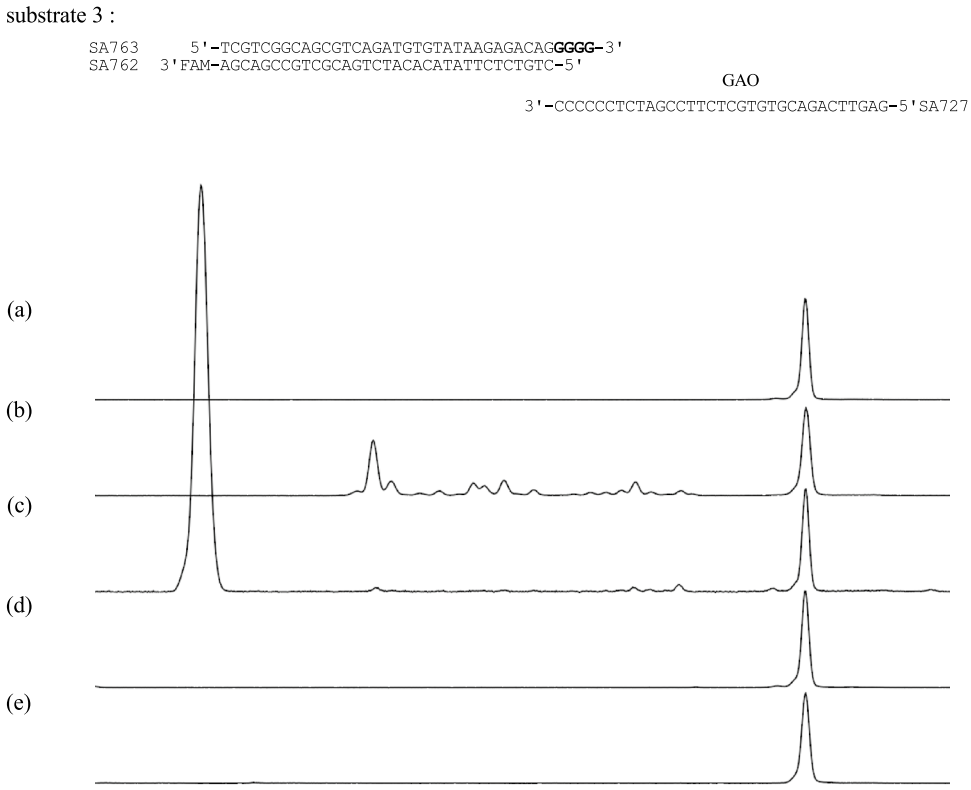

Figure 3. Taq polymerase-mediated CIS reaction causes nick translation. Substrate 3 carrying FAM at the 3′ end of the bottom strand (top) was analyzed before (a) and after the Taq CIS reaction, followed by phenol/chloroform/isoamyl alcohol treatment and ethanol precipitation (b). The sample shown in (b) was treated with exonuclease I to digest the single-stranded DNA (c). Short fragments were not observed when substrate 3 was incubated with Taq polymerase in the absence of GAO (d) or when MMLV-RT was used for the CIS reaction (e).

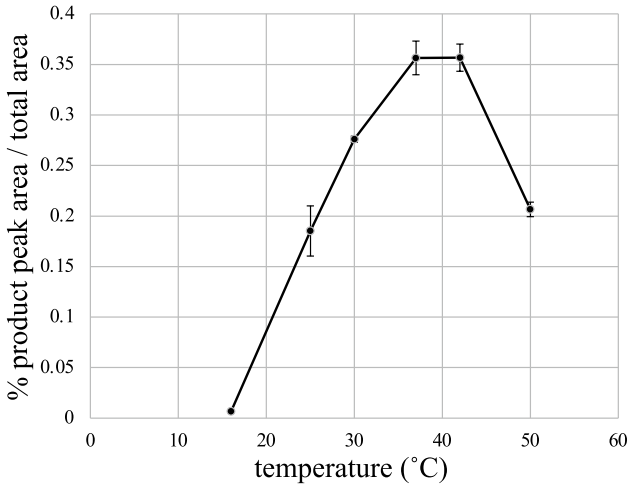

Figure 4. Optimal temperature for the Taq CIS reaction. Substrate 1 (10 fmol) was reacted with 2 pmol GAO (SA727) in a 10 μL reaction mixture using Taq polymerase (0.5 unit). The reactions were conducted at different temperatures for 10 min and analyzed by capillary electrophoresis using a 3130xl capillary sequencer. The ratio of the CIS reaction product peak area to the total peak area was calculated. The values represent the average of three independent reactions, and the error bars represent standard deviations.

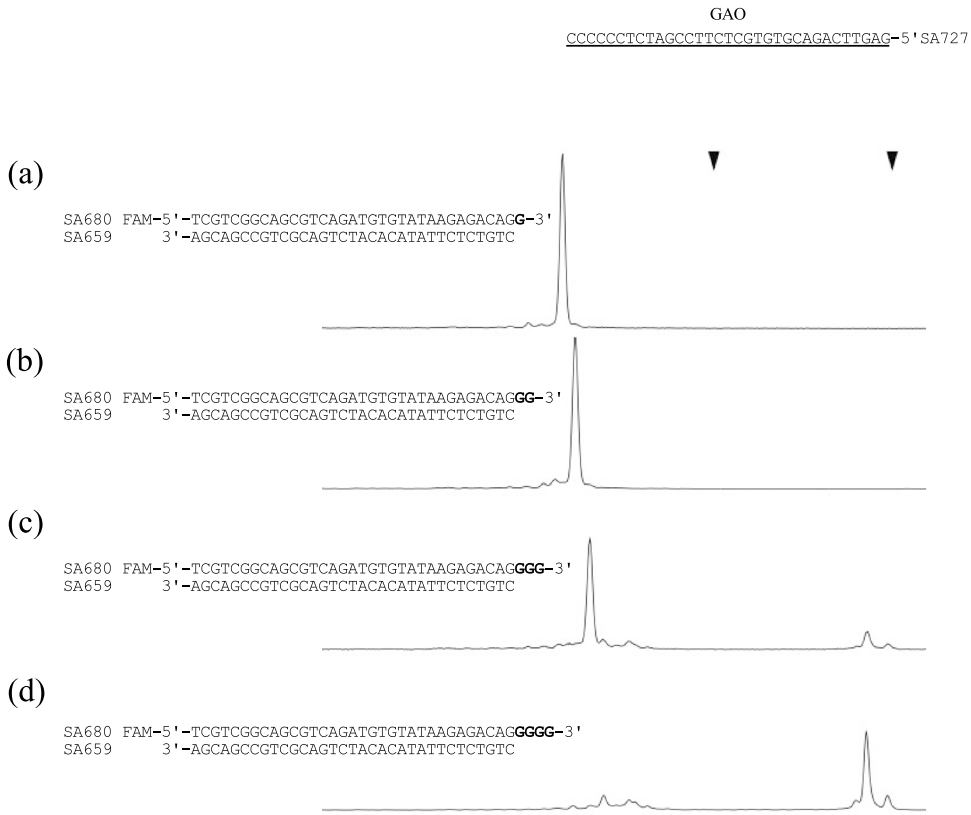

Figure 5. 3′-G tail length of three is the minimum requirement for the CIS Taq reaction. Four different substrates (a–d) with one to four extruding Gs were used for the CIS reaction with SA727 (GAO). After 10 min at 30 °C, the reaction mixture was analyzed by capillary electrophoresis. Two triangles indicate the LIZ size marker peaks (not shown) used to calibrate the electropherograms.

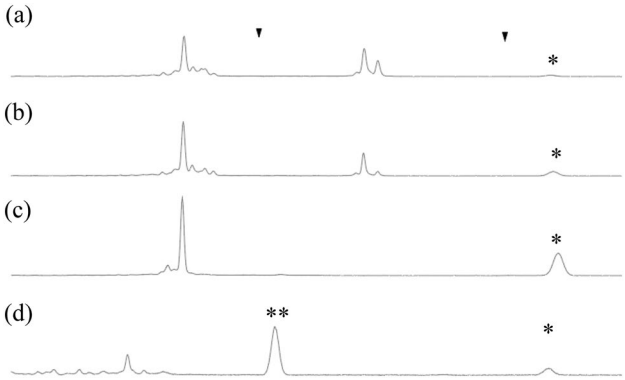

Figure 6. Taq polymerase cleaves off 5′ FAM independent of the CIS reaction. Taq CIS reaction of substrate 1 with GAO (SA727) produced an unknown broad peak (marked with "*") at elevated temperatures of 42 °C (a) and 50 °C (b). A peak was observed when substrate 1 was incubated with Taq polymerase in the absence of SA727 (c). Substrate 1 treated with T4 DNA polymerase in the absence of dNTP also produced the same peak (indicated by "*") and an additional peak (indicated by "**") (d). The peaks indicated by "*" and "**" represent FAM-5′-T-3′ and FAM-5′-TC-3′, respectively. Two triangles indicate the LIZ size marker peaks (not shown) used to calibrate the electropherograms.

substrate exhibited distinctly higher efficiency. In contrast, no CIS reaction products were obtained when the G-tails of nucleotides 1 and 2 were used.

Discussion

In this study, we demonstrated that well-known and widely used Taq DNA polymerase catalyzes the CIS reaction. This reaction is similar to but different from that catalyzed by MMLV-RT, as Taq polymerase and MMLV-RT produce nicked (Fig. 1) and flapped-end products, respectively.

Usage of CIS reaction in DNA labeling with unique molecular identifiers (UMIs)

Our development of the CIS reaction will benefit modern biotechnologies by providing a means to label DNA ends with unique molecular identifiers (UMIs). The CIS reaction can be used for the efficient labeling of DNA ends with desired DNA sequences carrying UMIs.

The UMI-labeling of DNA ends is key technology to the development of a method called qTnSeq for quantitative analysis of Tn mutant libraries. In qTnSeq, DNA obtained from a Tn mutant library was fragmented by sonication. The generated DNA ends were end-repaired and G-tailed using MMLV-RT in the presence of deoxycytidine, a tailing enhancer. Subsequently, for Illumina analysis, a CIS reaction was conducted to append the adapter sequence to the tailed ends; and then the NGS sequence library is prepared by PCR.

If a portion of the GAO contains a random sequence (e.g., N13), upon acquisition of NGS reads, sequences with the same UMI can be identified as originating from the same DNA molecule, thereby preventing overcounting of duplicate reads. Using UMI, Tn mutant libraries can be analyzed with high quantitative accuracy. The usefulness of UMI labeling of DNA ends in qTnSeq will be published elsewhere.

General guidance for preparation of CIS reaction substrate

A proper 3′ G-tailing of the substrate DNA is crucial for the CIS reaction. The G-tailing reaction conducted in the presence of the specific enhancer deoxycytidine typically appends 4–5 Gs. However, the tailing reaction requires a 3′ OH group in the substrate DNA. The DNA 3′ ends generated by sonication have long been misunderstood as primarily carrying a hydroxyl group. We previously reported that only approximately 20% of sonication-generated DNA 3′ ends carry hydroxyl groups, whereas approximately 50% and 30% of them carry 3′-phosphate groups or unknown structures, respectively[17]. The possible unknown structures may include a 2',3'-double bond, which could be formed upon breakage at an abasic site, although this remains to be fully elucidated[17]. To convert DNA termini into hydroxyl groups, we routinely use SB-repairing (scrap-and-build repair), in which 3′ → 5′ activity of exonuclease III is counter balanced by the DNA synthesis activity of T4 DNA polymerase[17].

Advantages of CIS reaction conducted by Taq polymerase

We have identified several advantages of the use of Taq polymerase in CIS reactions compared to the use of MMLV-RT. First, when MMLV-RT is used, dsDNA with a displaced strand is obtained as the CIS reaction product. It would be difficult to estimate such DNA size by electrophoresis. In contrast, in Taq polymerase-based CIS reaction, the size can be reliably estimated by electrophoresis. Second, MMLV-RT appends GAO recursively and produces a concatenated end. Owing to the tailing activity of MMLV-RT, it appends some nucleotides to the 3′ end after synthesizing the complementary strand of GAO, and another GAO can be incorporated, resulting in a GAO-concatenated end. Although placing a biotin moiety at the 5′ end of GAO prevents concatemer formation by MMLV-RT[16], the use of Taq polymerase, which mostly appends a single A without concatemer formation, is favorable. Third, as shown in Fig. 2m, the CIS reaction by MMLV-RT yielded a shorter product than the

expected size. In the CIS reaction, we expected that the last 3′ cytosine base of GAO would anneal to the first guanine base of the tail. The unexpected short (-1) product peak of the MMLV-RT-mediated CIS reaction is possibly attributable to offset annealing of GAO, in which the last 5′ cytosine base of the dsDNA is displaced by the 3′ cytosine base of GAO (Fig. 1 right bottom). The Taq-polymerase-mediated CIS reaction product has less possibility to be affected by offset annealing, which could compromise downstream in silico analysis. Lastly, Taq polymerase is considered to be significantly more efficient DNA polymerase compared to MMLV-RT. Those are several advantages of using Taq polymerase in CIS reactions compared to the use of MMLV-RT. On the other hand, the use of MMLV-RT in CIS reactions also has its own unique advantages, such as its faster reaction speed and its strong catalytic activity towards short G-tails of 1 or 2 Gs. The insights provided in this discussion will guide researchers in selecting either one of the two polymerases for their CIS reactions.

Those advantages discussed above highlights the novelty of the CIS reaction catalyzed by Taq polymerase, which would be useful in future research activities and potential practical applications.

GAO Interfering with the FAM-labeled DNA in capillary electrophoresis

When the post-CIS reaction solution was directly sampled and analyzed using a capillary sequencer in HiDi formamide, one unknown peak was observed in addition to the expected product peak (Fig. 2b). We later found that the commercially available size standard obtained from Promega (CLS600), which is designed to be used under denaturing conditions, contains counteracting oligonucleotides that suppress the extra peaks[24]. In this study, we evaluated and confirmed the interference of GAO affecting the outcome of the capillary electrophoresis of FAM-labeled strands. DNA analysis under denaturing conditions using capillary sequencers is convenient and effectively used in various analyses[4,5,16,17,25–27]. This information regarding the interference of single-stranded DNA in electrophoresis would be crucial for its further applications in various analyses.

Nature of the broad peak observed at higher temperatures

A broad peak was observed when the Taq polymerase-mediated CIS reaction was conducted at 42 °C or 50 °C (Fig. 6a,b indicated by *). As it was also observed when 5′-FAM-labeled dsDNA was incubated with Taq polymerase in the absence of GAO at 50 °C (Fig. 6c), it must not be related to the CIS reaction. Furthermore, a peak at the same location was observed when substrate 1 was treated with T4 DNA polymerase in the absence of dNTPs (T4 DNA polymerase degraded dsDNA into mononucleotides in the absence of dNTPs) (Fig. 6d).

This broad peak may represent the FAM with a T nucleotide (FAM-5′-T-3′). It is possible that at elevated temperatures, two strands of DNA locally melt at the terminus and form a structure recognized and cleaved by the Taq polymerase with its flap-end nuclease activity[21]. The resulting FAM-5′-T-3′ may migrate slowly owing to its large mass compared to its electric charge. Although further investigation is needed to understand the mechanism by which Taq polymerase produces this peak, we recommend avoiding the use of Taq polymerase when preparing double-stranded DNA with 5′ fluorescent labels for downstream applications such as gel-shift assays or FISH experiments.

Clumping activity of Taq polymerase

The unusual activity of Taq polymerase demonstrated in this study indicates that a 3–4 nucleotide long G-tail was sufficient to accept C-tailed single-stranded DNA, and DNA polymerization was primed and initiated. Hence, Taq polymerase has a clumping activity. Since Watson–Crick base pairing of four G-C pairs is not stable at 42 °C or 50 °C, the base-stacking interaction between the base at the 3′ end of GAO and the 5′ end of the accepting dsDNA was possibly involved in this reaction.

In conclusion, we believe that our findings will aid the establishment of new molecular techniques applicable in various fields of research in the future. It is possible to design and obtain a mutant Taq polymerase that has increased clumping activity and can accept dsDNA with shorter G-tails as a substrate.

Data availability

The original data supporting the findings of this study are available upon reasonable request. To request the data, please contact Yoshiyuki Ohtsubo at yoshiyuki.ohtsubo.a6@tohoku.ac.jp. The data will be made available under the condition that it is used only for non-commercial research purposes and that any publications or presentations resulting from the use of the data cite this study as the source of the data.

Received: 27 October 2023; Accepted: 13 January 2024
Published online: 23 January 2024

References

1. Chien, A., Edgar, D. B. & Trela, J. M. Deoxyribonucleic acid polymerase from the extreme thermophile *Thermus aquaticus*. *J. Bacteriol.* **127**, 1550–1557 (1976).
2. Roychoudhury, R., Jay, E. & Wu, R. Terminal labeling and addition of homopolymer tracts to duplex DNA fragments by terminal deoxynucleotidyl transferase. *Nucleic Acids Res.* **3**, 863–877 (1976).
3. Shinnick, T. M., Lerner, R. A. & Sutcliffe, J. G. Nucleotide sequence of Moloney murine leukaemia virus. *Nature* **293**, 543–548 (1981).
4. Ohtsubo, Y., Nagata, Y. & Tsuda, M. Efficient N-tailing of blunt DNA ends by Moloney murine leukemia virus reverse transcriptase. *Sci. Rep.* **7**, 41769 (2017).
5. Ohtsubo, Y., Nagata, Y. & Tsuda, M. Compounds that enhance the tailing activity of Moloney murine leukemia virus reverse transcriptase. *Sci. Rep.* **7**, 6520 (2017).
6. Zhou, M. Y. & Gomez-Sanchez, C. E. Universal TA cloning. *Curr. Issues Mol. Biol.* **2**, 1–7 (2000).
7. Stevenson, J. & Brown, A. J. Universal CG cloning of polymerase chain reaction products. *Anal. Biochem.* **471**, 80–82 (2015).
8. Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53–59 (2008).

9. Rhoads, A. & Au, K. F. PacBio sequencing and its applications. *Genomics Proteomics Bioinform.* **13**, 278–289 (2015).
10. Zhu, Y. Y., Machleder, E. M., Chenchik, A., Li, R. & Siebert, P. D. Reverse transcriptase template switching: A SMART approach for full-length cDNA library construction. *Biotechniques* **30**, 892–897 (2001).
11. Kapteyn, J., He, R. F., McDowell, E. T. & Gang, D. R. Incorporation of non-natural nucleotides into template-switching oligonucleotides reduces background and improves cDNA synthesis from very small RNA samples. *BMC Genomics* https://doi.org/10.1186/1471-2164-11-413 (2010).
12. Hrdlickova, R., Toloue, M. & Tian, B. RNA-Seq methods for transcriptome analysis. *WIREs RNA* **8**, e1364 (2017).
13. Oz-Gleenberg, I., Herschhorn, A. & Hizi, A. Reverse transcriptases can clamp together nucleic acids strands with two complementary bases at their 3′-termini for initiating DNA synthesis. *Nucleic Acids Res.* **39**, 1042–1053 (2011).
14. Oz-Gleenberg, I., Herzig, E., Voronin, N. & Hizi, A. Substrate variations that affect the nucleic acid clamp activity of reverse transcriptases. *FEBS J.* **279**, 1894–1903 (2012).
15. Oz-Gleenberg, I. & Hizi, A. Strand selections resulting from the combined template-independent DNA synthesis and clamp activities of HIV-1 reverse transcriptase. *Biochem. Biophys. Res. Commun.* **408**, 482–488 (2011).
16. Ohtsubo, Y., Sasaki, H., Nagata, Y. & Tsuda, M. Optimization of single strand DNA incorporation reaction by Moloney murine leukaemia virus reverse transcriptase. *DNA Res.* **25**, 477–487 (2018).
17. Ohtsubo, Y., Sakai, K., Nagata, Y. & Tsuda, M. Properties and efficient scrap-and-build repairing of mechanically sheared 3′ DNA ends. *Commun. Biol.* **2**, 409 (2019).
18. Lundberg, D. S., Yourstone, S., Mieczkowski, P., Jones, C. D. & Dangl, J. L. Practical innovations for high-throughput amplicon sequencing. *Nat. Methods* **10**, 999–1002 (2013).
19. Vvedenskaya, I. O. *et al.* Massively systematic transcript end readout, "MASTER": Transcription start site selection, transcriptional slippage, and transcript yields. *Mol. Cell* **60**, 953–965 (2015).
20. Girardot, C., Scholtalbers, J., Sauer, S., Su, S. Y. & Furlong, E. E. Je, a versatile suite to handle multiplexed NGS libraries with unique molecular identifiers. *BMC Bioinform.* **17**, 419 (2016).
21. Lyamichev, V., Brow, M. A. & Dahlberg, J. E. Structure-specific endonucleolytic cleavage of nucleic acids by eubacterial DNA polymerases. *Science* **260**, 778–783 (1993).
22. Takagi, M. *et al.* Characterization of DNA polymerase from *Pyrococcus* sp. strain KOD1 and its application to PCR. *Appl. Environ. Microbiol.* **63**, 4504–4510 (1997).
23. Rigby, P. W., Dieckmann, M., Rhodes, C. & Berg, P. Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. *J. Mol. Biol.* **113**, 237–251 (1977).
24. McLaren, R. S. *et al.* Post-injection hybridization of complementary DNA strands on capillary electrophoresis platforms: A novel solution for dsDNA artifacts. *Forensic Sci. Int. Genet.* **2**, 257–273 (2008).
25. Nonoyama, S. *et al.* A transcriptional regulator, IscR, of *Burkholderia multivorans* acts as both repressor and activator for transcription of iron-sulfur cluster-biosynthetic isc operon. *Res. Microbiol.* **171**, 319–330 (2020).
26. Kishida, K., Inoue, K., Ohtsubo, Y., Nagata, Y. & Tsuda, M. Host range of the conjugative transfer system of IncP-9 naphthalene-catabolic plasmid NAH7 and characterization of its oriT region and relaxase. *Appl. Environ. Microbiol.* https://doi.org/10.1128/AEM.02359-16 (2017).
27. Sato, T. *et al.* The small protein HemP is a transcriptional activator for the hemin uptake operon in *Burkholderia multivorans* ATCC 17616. *Appl. Environ. Microbiol.* https://doi.org/10.1128/AEM.00479-17 (2017).

Author contributions

Y.O. and S.K. performed the experiments, and Y.O. wrote the first draft of the manuscript, while Y.N. provided critical feedback and contributed to the final version of the manuscript.

Funding

This work was supported by Institute for Fermentation, Osaka (IFO) (Grant ID: K-2016-004), Grants-in-Aid for Scientific Research (B) from the Japan Society for the Promotion of Science (JSPS)(Grant ID: 19H02865 and 22H02233), and Grants-in-Aid for Challenging Exploratory Research from JSPS (Grant ID: 22K19124).

Competing interests

Y.O. and Y.N. are affiliated with Tohoku university that holds a related domestic patent. S. K. declares no competing interest.

Additional information

Correspondence and requests for materials should be addressed to Y.O.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2024