

Full Paper

3' Branch ligation: a novel method to ligate non-complementary DNA to recessed or internal 3'OH ends in DNA or RNA

Lin Wang^{2†}, Yang Xi^{1,3†}, Wenwei Zhang^{1,3†}, Weimao Wang^{1,3}, Hanjie Shen^{1,3}, Xiaojue Wang^{1,3}, Xia Zhao^{1,3}, Andrei Alexeev², Brock A. Peters^{1,2,3}, Alayna Albert², Xu Xu^{1,3}, Han Ren^{1,3}, Ou Wang^{1,3}, Killeen Kirkconnell², Helena Perazich², Sonya Clark², Evan Hurowitz², Ao Chen^{1,3}, Xun Xu^{1,3}, Radoje Drmanac^{1,2,3,4*}, and Yuan Jiang ^{2*}

¹Institute of Biochemistry, BGI-Shenzhen, Shenzhen 518083, China, ²Advanced Genomics Technology Lab, Complete Genomics Inc., 2904 Orchard Pkwy, San Jose, California 95134, USA, ³China National GeneBank, BGI-Shenzhen, Shenzhen 518120, China, and ⁴Department of R&D, MGI, BGI-Shenzhen, Shenzhen 518083, China

*To whom correspondence should be addressed. Tel. +1 4086482560 3079; +1 4086482549.
Email: yjiang@completegenomics.com (Y.J.); Tel. +1 4088389539. Fax. +1 4086482549.
Email: rdrmanac@completegenomics.com (R.D.)

†Contributed equally to this work.

Edited by Dr. Yuji Kohara

Received 20 July 2018; Editorial decision 28 September 2018; Accepted 4 October 2018

Abstract

Nucleic acid ligases are crucial enzymes that repair breaks in DNA or RNA during synthesis, repair and recombination. Various genomic tools have been developed using the diverse activities of DNA/RNA ligases. Herein, we demonstrate a non-conventional ability of T4 DNA ligase to insert 5' phosphorylated blunt-end double-stranded DNA to DNA breaks at 3'-recessive ends, gaps, or nicks to form a Y-shaped 3'-branch structure. Therefore, this base pairing-independent ligation is termed 3'-branch ligation (3'BL). In an extensive study of optimal ligation conditions, the presence of 10% PEG-8000 in the ligation buffer significantly increased ligation efficiency to more than 80%. Ligation efficiency was slightly varied between different donor and acceptor sequences. More interestingly, we discovered that T4 DNA ligase efficiently ligated DNA to the 3'-recessed end of RNA, not to that of DNA, in a DNA/RNA hybrid, suggesting a ternary complex formation preference of T4 DNA ligase. These novel properties of T4 DNA ligase can be utilized as a broad molecular technique in many important genomic applications, such as 3'-end labelling by adding a universal sequence; directional tagmentation for NGS library construction that achieve theoretical 100% template usage; and targeted RNA NGS libraries with mitigated structure-based bias and adapter dimer problems.

Key words: novel ligation, T4 DNA ligase, NGS, molecular tool

1. Introduction

Ligases repair breaks in nucleic acids, and this activity is essential for cell viability and vitality. DNA ligases catalyse the formation of a phosphodiester bond between DNA ends and play crucial roles in DNA repair, recombination, and replication *in vivo*.^{1–3} RNA ligases join the 5'-phosphoryl (5'PO₄) and 3'-hydroxyl (3'OH) RNA termini via phosphodiester bonds and are involved in RNA repair, splicing, and editing.⁴ Ligases from all three kingdoms of organisms (bacteria, archaeobacteria, and eukaryotes) can be utilized *in vitro* as important molecular tools for applications such as cloning, ligase-based amplification or detection, and synthetic biology.^{5–7}

One of the most widely used ligases *in vitro* is bacteriophage T4 DNA ligase, which is a single 55-kDa polypeptide that requires ATP as an energy source.⁸ T4 DNA ligase typically joins the adjacent 5'PO₄ and 3'OH termini of duplexed DNA. In addition to sealing nicks and ligating cohesive ends, T4 DNA ligase can also efficiently catalyse blunt-end joining, which has not been observed for any other DNA ligases⁹ (<http://www.epibio.com/enzymes/ligases-kinases-phosphatases/dna-ligases/ampligase-thermostable-dna-ligase?details> (15 October 2018, date last accessed)). Some unusual catalytic properties of this ligase were reported previously, such as sealing single-stranded gaps in duplex DNA, sealing nicks adjacent to abasic sites in double-stranded DNA (dsDNA), promoting intramolecular loop formation with partially double-stranded DNA, and joining DNA strands containing 3' branch extensions.^{10–12} Researchers also observed template-independent ligation mediated by T4 ligase, such as mis-paired nick sealing in dsDNA¹³ or even single-stranded DNA (ssDNA) ligation, albeit at very low efficiency.¹⁴ These results suggest that perfect complementary base pairing at or adjacent to the ligation junction is not critically needed for some unconventional T4 DNA ligase activities. T4 RNA ligases 1 and 2 are the products of genes 63 and 24, respectively, of T4 phage. Both require an adjacent 5'PO₄ and 3'OH end for a successful ligation with the concurrent hydrolysis of ATP to AMP and PPi. The substrates for T4 RNA ligase 1 include single-stranded RNA and DNA, whereas T4 RNA ligase 2 preferentially seals nicks on dsRNA rather than ligating the ends of ssRNA.^{15,16}

Here, we demonstrate a non-conventional end-joining event mediated by T4 DNA ligase that we call 3'-branch ligation (3'BL). This method can join DNA or DNA/RNA fragments at nicks, single-stranded gaps, or 3'-recessive ends to form a branch structure. This report includes extensive study of a wide variety of ligation cofactors and activators and the optimization of the ligation conditions for this type of novel ligation. With our 3'BL protocol, no base pairing was required, and the ligation can reach 70–90% completion in most cases, including for a 1-nt gap. One application of this method is the attachment of adapters to DNA or RNA during NGS library preparation. Several genomic structures that were previously considered unligatable can now be used as substrates for 3'BL, resulting in a high conversion rate of input DNA into adapter-ligated molecules while avoiding chimeras. We demonstrate that 3'BL can be coupled with transposon tagmentation to increase library yield. The directional tagmentation strategy we propose will theoretically produce templates 100% of which can be utilized for sequencing. Our study demonstrated the value of this novel technique for NGS library preparation and the potential to advance many other molecular applications.

2. Materials and methods

2.1. 3'-Branch ligation for duplex DNA

The substrates for 3'BL were composed of 2 pmol of ON1 or ON9 mixed with 4 pmol each of one or two additional oligos in pH

8 Tris-EDTA (TE) buffer (Life Technologies) as follows: substrates 1 and 5 (nick), ON-1/2/3 and ON-9/10/11; substrates 2 and 6 (1-nt gap), ON1/2/4 and ON9/10/12; substrate 3 (8-nt gap), ON1/4/5; substrates 4 and 9 (3'-recessive end), ON1/2 and ON9/10; substrate 7 (2-nt gap), ON9/10/13; substrate 8 (3-nt gap), ON9/10/14; blunt-end control, ON1 and ON6 (Fig. 1, Supplementary Table S1). The template was ligated to 1.5–5 μM of adapter (Ad-G: ON7/8, Ad-T: ON15/16, Ad-A: ON17/8, Ad-GA: ON19/20) using 20 units/μl of T4 ligase (Enzymatics, Inc.) in 3'BL buffer [0.05 mg/ml BSA (New England Biolabs), 50 mM Tris-Cl pH 7.8 (Amresco), 10 mM MgCl₂ (EMD Millipore), 0.5 mM DTT (VWR Scientific), 10% PEG-8000 (Sigma Aldrich), and 1 mM ATP (Sigma Aldrich)]. The reaction volume was 30, 40, or 60 μl. The optimization tests were conducted by altering the ATP concentration from 1 μM to 1 mM, the Mg²⁺ concentration from 3 to 10 mM, the pH value from 3 to 9, temperature from 12 to 42°C, and adjusting additives such as PEG-8000 from 2.5% to 10% and SSB from 2.5 to 20 ng/μl. The ligation mixture was prepared on ice and incubated at 37°C for 1–12 h before heat inactivation at 65°C for 15 min. Ligation products were assayed on TBE gels without further purification (Fig. 2). The samples run on denaturing polyacrylamide gels were purified using either home-made PEG32 beads [Axygen beads (Corning) mixed with 32% PEG] or Oligo Clean & Concentrator kit (Zymo research). All gels were visualized on an Alpha Imager (Alpha Innotech). An input control was loaded at either an equal quantity or half the quantity of the template used for ligation. A ligation efficiency rate was estimated by dividing the intensity of ligated products by the total intensity of ligated and unligated products using ImageJ Software (NIH). Optimized reaction conditions were used unless otherwise noted.

2.2. 3'-Branch ligation for DNA/RNA hybrid

The substrates for 3'BL were composed of 10 pmol ON-21 RNA oligo mixed with 2 pmol of ON-21 or ON-23 DNA oligo. For T4 DNA ligase-mediated 3'BL, the substrate was incubated with Ad-T (ON15/16) in 3'BL buffer as described above and incubated at 37°C for 1 h. 3'BL using T4 RNA ligase 1 or 2 was performed in 1× RNA ligase buffer (NEB) with either 20% DMSO or 25% PEG. All ligation products were assayed on 6% denaturing polyacrylamide gels.

2.3. Directional tagmentation library construction

The transposon oligonucleotides used in this experiment were synthesized by Sangon Biotech. For the 2 transposon experiments using TnA/TnB, oligos for TnA (ON24), TnB (ON25), and MRev (ON26) were annealed at a 1:1:2 ratio. For the single transposon experiment with TnA, ON24 and ON26 were annealed at a 1:1 ratio. For the Y (TnA&TnB) transposon experiment, ON24 and ON27 were annealed at a 1:1 ratio.

Transposon assembly was performed by mixing 100 pmol of pre-annealed adapters, 7 μl of Tn5 transposase, and sufficient glycerol to obtain a total 20-μl reaction, which was incubated at 30°C for 1 h. Tagmentation of genomic DNA (Coriell 12878) was performed in 20-μl reactions containing 100 ng of gDNA, TAG buffer (home-made), and 1 μl of the assembled transposon. The reaction was incubated at 55°C for 10 min; 40 μl of 6 M guanidine hydrochloride (Sigma) was then added to remove the transposon complex from tagged DNA, and DNA was purified using Agencourt AMPure XP beads (Beckman Coulter). The gap ligation of AdB (ON28 and ON29) to the tagged DNA was performed at 25°C for 1 h in reactions containing 100 pmol of the adapter, 600 U of T4 DNA ligase (Enzymatics, Inc.), and 3'BL buffer. Reactions were purified

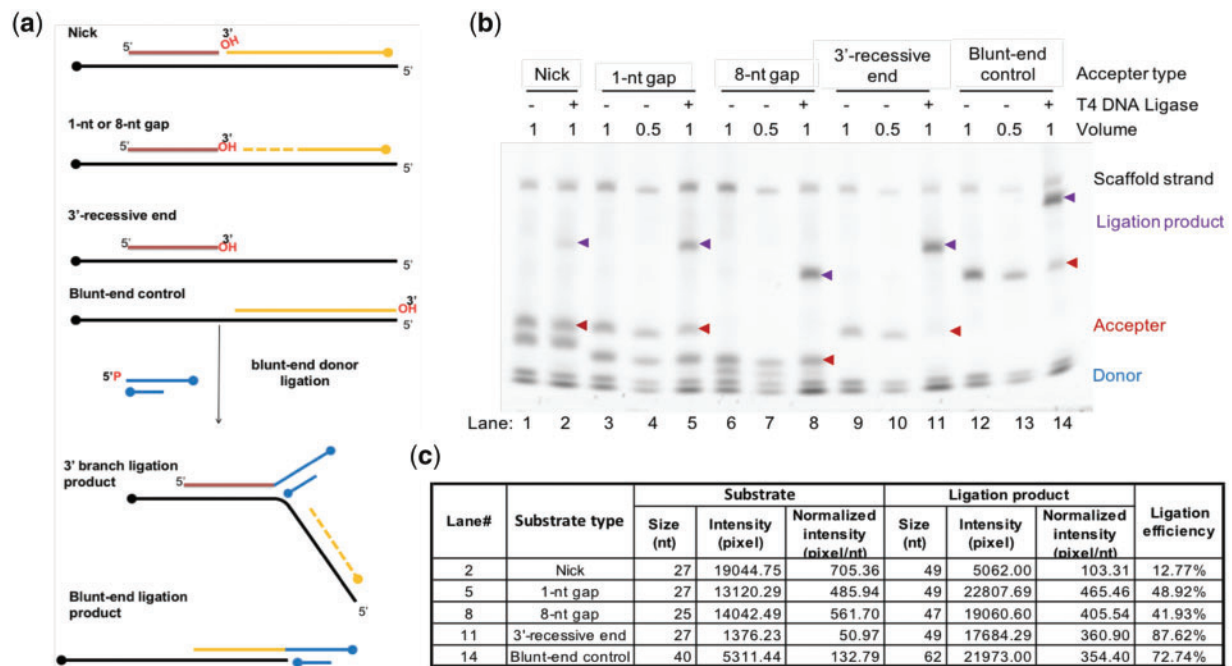


Figure 1. 3' Branch ligation by T4 DNA ligase at non-conventional DNA ends formed by nicks, gaps, and overhangs. (a) Schematic representation of ligation assay with different DNA acceptor types. The blunt-end DNA donor (blue) is a synthetic, partially dsDNA molecule with dideoxy 3'-termini (filled circles) to prevent DNA donor self-ligation. The long arm of the donor is 5'-phosphorylated. The DNA acceptors were assembled using 2 or 3 oligos (black, red, and orange lines) to form a nick (without phosphates), a gap (1 or 8 nt), or a 36-nt 3'-recessive end. All strands of the substrates are unphosphorylated, and the scaffold strand is 3' dideoxy protected. (b) Analysis of the size shift of ligated products of substrates 1, 2, 3, and 4, respectively, using a 6% denaturing polyacrylamide gel. Reactions were performed according to the optimized condition. The negative no-ligase controls (lanes 1, 3, 4, 6, 7, 9, 10, 12, and 13) were loaded at 1 or 0.5× volume of corresponding experimental assays. If ligation occurs, the substrate size is shifted up by 22 nt. Red arrowheads correspond to the substrate, and purple arrowheads correspond to donor-ligated substrates. Donor and substrate sequences in [Supplementary Table S1](#). (c) Expected sizes of substrate and ligation product and approximate ligation efficiency in each experimental group. The intensity of each band was estimated using ImageJ and normalized by its expected size. Ligation efficiency was estimated by dividing the normalized intensity of ligated products by the normalized total intensity of ligated and unligated products.

using AMPure XP beads. PCR amplification of tagmented and gap-ligated DNA was performed in 50- μ l reactions containing 2 μ l of the tagmented or gap-ligated DNA, TAB buffer, 1 μ l TruePrep Amplify Enzyme (Vazyme), 200 mM dNTPs (Enzymatics, Inc.), and 400 mM each of primers Pr-A and Pr-B. Tagmented reactions were incubated as follows: 72°C for 3 min; 98°C for 30 s; 8 cycles of 98°C for 10 s, 58°C for 30 s, and 72°C for 2 min; and 72°C for a 10-min extension. 3'BL-ligated reactions were run using the same program without the initial 3 min extension at 72°C. PCR reactions using either prA (ON30) or both prA and prB (ON31) were purified using AMPure XP beads. Purified products were quantified using the Qubit High-Sensitivity DNA kit (Invitrogen).

3. Results

3.1. 3' Branch ligation: a novel method to ligate non-complementary DNA ends

Conventionally, DNA ligation involves the joining of the 5'PO₄ and 3'OH DNA ends of cohesive or blunt-ended fragments. Cohesive-end ligation is generally faster and less dependent on enzyme concentration compared with blunt-end joining. Both processes can be catalysed by bacteriophage T4 DNA ligase, which uses ATP as an energy-yielding cofactor and requires Mg²⁺.⁸ T4 DNA ligase was also reported to ligate specific or degenerate single-stranded oligos to partially single-stranded substrates through hybridization.^{17,18} Here,

we demonstrated a non-conventional T4 DNA ligase-mediated ligation that does not require complementary base pairing and can ligate a blunt-end DNA donor to the 3'OH end of a duplex DNA acceptor at 3'-recessed strands, gaps, or nicks (Fig. 1a). Therefore, we use the term 3'-branch ligation (3'BL) to describe these ligations. The synthetic donor DNA we used contained a 5' blunt duplex end and a 3' ssDNA end. The acceptor substrates contained one of the following structures: a dephosphorylated nick, a 1- or 8-nt gap, or a 3' 36-nt 3'-recessive end ([Supplementary Table S1](#)). T4 ligase helps to join the 5'PO₄ of the donor strand to the sole ligatable 3'OH of the acceptor strand to form a branch-shaped ligation product.

To optimize the ligation efficiency, we extensively tested a number of factors that affect general ligation efficiency, including the adapter: DNA substrate ratio, T4 ligase quantity, final ATP concentration, Mg²⁺ concentration, pH, incubation time, and different additives, such as polyethyleneglycol-8000 (PEG-8000) and single-stranded binding protein (SSB) ([Supplementary Fig. S2](#), data not shown). Adding PEG-8000 to a final concentration of 10% substantially increased the ligation efficiency by 2- to 10-fold ([Supplementary Fig. S1](#)). A large range of ATP concentrations (from 1 μ M to 1 mM in [Supplementary Fig. S2a](#)) and Mg²⁺ concentrations (3–10 mM, data not shown) were compatible with the 3'BL. The ligase quantity needed for 3'BL was comparable with that for blunt-end ligation. In our optimized conditions, we used donor: substrate DNA molar ratios of 30–100, and we performed the reactions at 37°C for 1 h at pH 7.8 with 1 mM ATP, 10 mM MgCl₂, and 10%

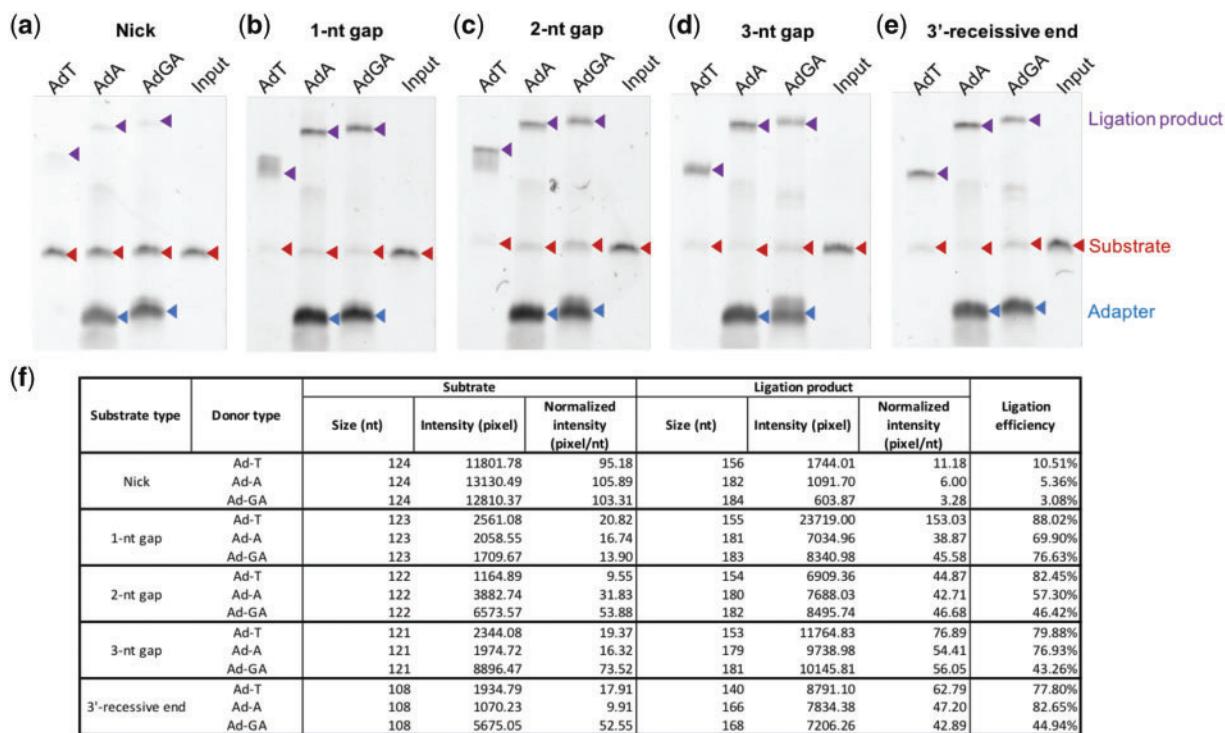


Figure 2. Gel analysis of size shift of ligated products using 6% TBE polyacrylamide gel. Red arrowheads correspond to the substrate, and purple arrowheads correspond to donor-ligated substrates: substrate 5 (nick) (a), substrate 6 (1-nt gap) (b), substrate 7 (2-nt gap) (c), substrate 8 (3-nt gap), (d) and substrate 9 (3'-recessive end) (e). Three DNA donors with different bases at the 5'-end of the ligation junction (T, A, or GA) were examined. Donor and substrate sequences are summarized in [Supplementary Table S1](#). (f) Table of ligation efficiency calculated based on normalized band intensity using ImageJ.

PEG-8000. The ligation of the same donors to blunt-end substrates and ligase-free reactions were used as positive and negative controls, respectively. The ligation donor (Ad-G) is double-stranded on one end (5' phosphorylated and 3' dideoxy protected) and single-stranded (3' dideoxy protected) on the other end (Fig. 1a, [Supplementary Table S1](#)). The ligation substrates are composed of the same bottom strand (ON1) with different top strands to compose nick, gap, and overhang structures. To quantify ligation product yields, the reaction products were separated on 6% denaturing polyacrylamide gels (Fig. 1b). Ligation efficiency was calculated as the ratio of product to substrate intensity using ImageJ (Fig. 1b and c). The experiments were repeated multiple times with very similar ligation efficiency (data not shown). The 3'-recessive ligation (lane 11 in Fig. 1b) appeared ~90% complete, which is even higher than the blunt-end ligation control (lane 14, 72.74%) and suggests remarkably high ligation efficiency with 3'-recessive DNA ends. The 1- or 8-nt gap substrates (lanes 5 and 8) showed good ligation efficiency of ~45%. Nick ligation (lane 2) efficiency was the lowest at ~13%. However, this ligation yield was improved when the nick ligation reaction was incubated longer (data not shown), suggesting slower kinetics for the nick ligation reaction.

We also extended our study to different adapter and substrate sequences using the optimized reaction condition (Fig. 2). The 5'PO₄ ends of three different adapters (Ad-T, Ad-A, or Ad-GA in [Supplementary Table S1](#)) contained either a single T or A or the dinucleotide GA at the ligation junction before a consensus CTGCTGA sequence. These 5'PO₄ ends were individually ligated to the 3'OH ends of acceptor templates with a T at the ligation junction. Overall, high ligation efficiency (70–90%) was seen in most cases except for nick ligations or 3'BL using Ad-GA (Fig. 2f), thus indicating some nucleotide

preference of T4 DNA ligase at the ligation junctions. Independent of the adapter and substrate sequences, the 3'-recessive end or gap ligations always showed better efficiencies (60–90%), whereas the nick ligation was fairly inefficient, in a 1-h incubation. We hypothesize that these discrepancies in ligation efficiency are due to the DNA bending where the nick/gap/overhang starts and exposes the 3'OH group for ligation. The longer ssDNA region likely makes the 3' termini more accessible in the ligation and results in higher ligation efficiency. We also tested whether a similar end-joining event could occur as 5' branch ligation. In contrast to 3'BL, no obvious ligation of a blunt-end adapter to the 5'PO₄ end at the gap or the 5'-recessive end was observed (data not shown). This result suggests greater steric hindrance of T4 DNA ligase at the donor's 5' termini compared with 3' termini.

3.2. 3' Branch ligation to ligate DNA to RNA

We further investigated 3'BL on DNA/RNA hybrids (ON-21/ON-23 in [Supplementary Table S1](#)) that form one DNA and one RNA 3'-recessive end (Fig. 3a). Ligation on DNA/DNA hybrids served as a positive control, whereas negative ligation controls included DNA/RNA hybrids, ssDNA, or ssRNA oligos incubated individually or with adapters (lanes 3, 4, and 5 in Fig. 3c and data not shown). Interestingly, when DNA/RNA hybrids were incubated with a blunt-end dsDNA donor, we observed a size change of the RNA oligo from the original 29–49 nt upon ligation. However, the DNA substrate remained unchanged (lanes 1 and 2 in Fig. 3c). This result suggests that the blunt-end dsDNA donor ligated to the 3'-end of RNA but not to the 3'-end of DNA. As a positive control, DNA/DNA hybrids with 3'-recessive ends on each side showed band shifts to larger species on both strands with nearly 100%

A previous study reported that for sealing nicks in DNA/RNA hybrids, T4 DNA ligase, and T4 RNA ligase 2, but not T4 RNA ligase 1, can effectively join a 5'PO₄ DNA end to a juxtaposed 3'OH DNA or RNA end when the complementary strand is RNA but not DNA.¹⁶ Therefore, we performed the same ligation test using T4 RNA ligases 1 and 2 either in 20% DMSO (Fig. 3d) or in 10% PEG (data not shown). In both tests, T4 RNA ligase 1 and T4 RNA ligase 2 slightly ligated the blunt end adapters to the 3' end of RNA in a DNA/RNA hybrid. Notably, in the RNA-only controls, T4 RNA ligase 2 could join a blunt-end dsDNA adapter to ssRNA. In conclusion, T4 DNA ligase, but not T4 RNA ligases, is competent to efficiently ligate blunt-end dsDNA to the 3' end of RNA through 3'BL.

3.3. Directional tagmentation library construction

Because 3'BL is useful for ligating adapters to several genomic structures with high efficiency, we explored its application in NGS workflows. Transposon-based library construction is rapid and consumes less input DNA compared with conventional NGS library preparation. However, using commercial transposon-based library preparation systems, only half of tagged molecules are flanked by two different adapter sequences (Fig. 4a), and tagged DNA is flanked by self-complementary regions that could form stable hairpin structures and compromise sequencing quality.¹⁹ In addition, the PCR-mediated incorporation of adapter sequences has not been adapted for whole-genome sequencing or PCR-free NGS library construction.

To overcome these limitations, we developed a new protocol for transposon-based NGS library construction by incorporating 3'BL. Both Tn5 and MuA transposons work through a 'cut and paste' mechanism, in which a transposon adapter sequence is end-joined to the 5'-end of target DNA to create a 9 or 5-bp gap, respectively, at the 3'-end of the genomic DNA (Fig. 4a). Subsequently, 3'BL can be used to add another adapter sequence to the 3'-end of genomic DNA

at the gap to complete the directional adapter ligation (Fig. 4c). We used Tn5 transposons in this manuscript to compare the efficiency of the single-tagmentation + 3'BL approach (Fig. 4c) to that of a double-tagmentation approach, which uses the two different Tn5-based adapters TnA and TnB (Fig. 4a), and to that of another directional single-tagmentation strategy using Y adapters that contain two different adapter sequences (Fig. 4b). Human genomic DNA was incubated with equimolar amounts of TnA and TnB transposome complexes, with the TnA transposome complex alone, or with the TnY (TnA/B) transposome complex. The product of TnA transposome-only fragmentation was further used as a template for 3'BL with the blunt-end adapter AdB, which shares a common adapter sequence with TnB. PCR amplification was performed using two primers, Pr-A and Pr-B, designed to recognize the TnA and AdB/TnB adapters, respectively. The quantification data suggested that TnA&AdB had the highest efficiency compared with TnA&TnB and TnY (TnA/B) (Fig. 4d). No significant amplification was observed when only one primer specific to TnA adapter was used (Fig. 4d). As expected due to PCR suppression, the TnA 3'BL approach, the double-tagmentation approach, and the TnY approach all showed significantly higher PCR efficiency than the tagmentation reaction with only the TnA or TnB transposome complex alone (Fig. 4d).

We also sequenced these libraries using BGI's DNA nanoball-based sequencing platform and compared the base-positional bias among the transposon-interfered end, the 3'BL end, and the regular TA ligation end (Fig. 5). It is obvious that the positional bias at the 3'BL end is less than that at the Tn5 end (Fig. 5a and b), which occurs because the 3'BL end is influenced by both transposon interruption and 3'BL. Because only the first 6 nt (positions 1–6) of the 3'BL end showed base bias and the bias was similar to but not exactly the same as that of its hybridized Tn5 end (positions 30–35, after the 9-nt overhang), we conclude that the positional bias re-

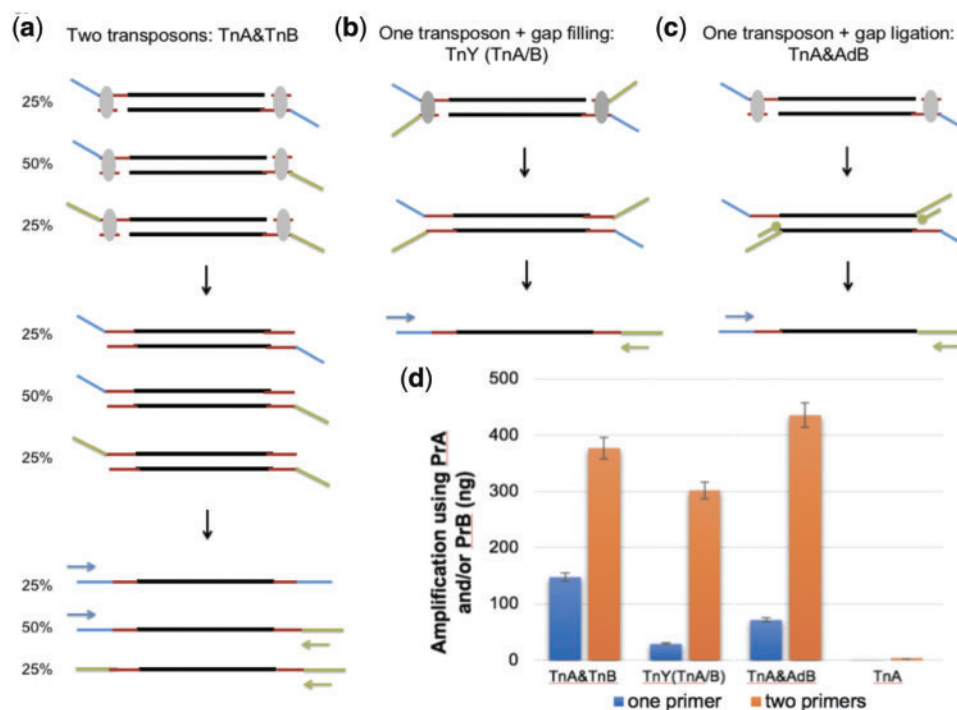


Figure 4. (a–c) Schematic representation of three transposon tagmentation methods followed by PCR amplification using Pr-A (blue arrow) and Pr-B (green arrow). Two-transposon method (a); one Y transposon tagmentation with 3'-gap filling (b); one-transposon method with adapter ligation at 3'-gap (c). (d) Graph of amplification signal after purification using pr-A or pr-A with pr-B after the various tagmentation and gap ligation conditions.

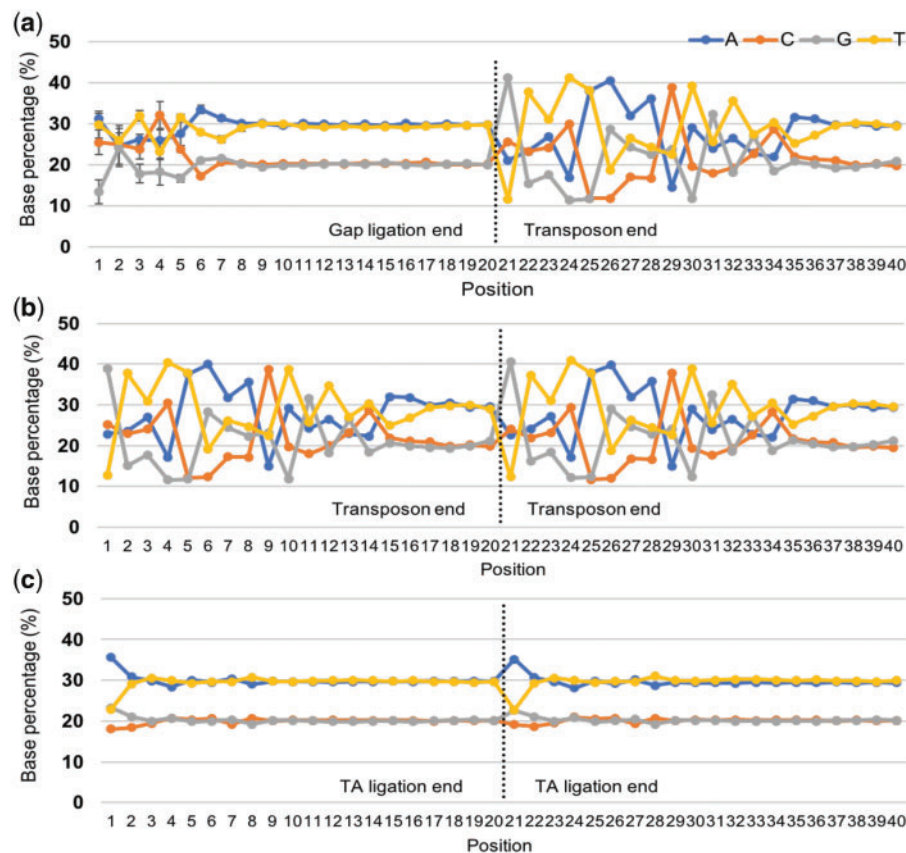


Figure 5. Base distribution bias of Tn5-gap ligation (a), two transposons (b), and regular TA ligation (c). Only the first 20 bases from each end of the ligation are presented; adenine, blue; cytosine, orange; guanine, grey; thymine, yellow; the average and standard deviation of five-independent libraries are presented.

observed at the 3'BL end is mainly caused by the Tn5 transposon. Therefore, 3'BL causes minimal bias and is similar to regular TA ligation (Fig. 5c).

4. Discussion

One important property of T4 DNA ligase is its efficient joining of blunt-end dsDNA,^{20,21} which has not been observed with other DNA ligases. This ligase was also reported to mediate some unusual catalytic events, such as ligating single-strand gaps or mismatched bases in duplex DNA,^{10,11} forming a stem-loop molecule from partially double-stranded DNA,¹² or inefficiently ligating ssDNA in a template-independent manner.¹⁴

Here, we demonstrated that T4 DNA ligase catalysed the joining of blunt-end dsDNA to the 3'OH end of dsDNA with a nick and the joining of partially single-strand duplex DNA with a gap or 3'-recessed end. In contrast, no ligation to the 5'PO₄ end at the 5'-recessed ends or in the gaps was observed (data not shown). This indicates that after binding to the 5'PO₄ end of the dsDNA adapter, T4 DNA ligase can access the recessed 3' end in the target DNA when DNA bends and exposes the 3'OH end. Unlike the mismatched or gapped ligation in previous reports,^{10–13} no base pairing of the donor DNA to the acceptor DNA was required with our 3'BL method. Moreover, even for a 1-nt gap, greater than 70% completion was accomplished using optimized conditions, while other template-independent ligation was of very low yield and was only detected by qPCR measurement.¹⁴

However, different 3'BL ligation efficiencies were observed for ligating 5' T, A, or GA to 3' T (Fig. 2), which indicates some sequence preference at the ligation junction. Despite the recognized ligation bias,²² T4 DNA ligase is commonly used in the adapter addition step during NGS library preparation. With its ability to perform 3'BL, T4 ligase can ligate adapters to several genomic structures that were previously considered unligatable, resulting in a higher template usage rate. For example, 3'BL can be coupled with single transposon tagmentation. The traditional double-transposon strategy has only 50% of the tagmented molecules that are amenable for the following amplification step. However, when DNA tagmentation is performed using one transposon with subsequent 3'BL, an increased yield of molecules with different adapters on each insert end can be acquired (Fig. 4). Furthermore, the tagmented 3'BL products can be directly loaded on Illumina's flow cell as PCR free WGS libraries, which was difficult to achieve using double-transposon strategy. Other directional transposon protocols have been proposed using a Y transposon composed of two different adapter sequences²³ or replacing the unlinked strand from a single transposon with a second adapter oligo followed by gap filling and ligation.^{24,25} However, these approaches continue to preserve the inverted adapter sequences and/or cannot insert sample barcodes adjacent to genomic DNA as the tagmented 3'BL protocol can. We performed a proof-of-concept study of a new directional tagmentation protocol for NGS library construction that eliminates inverted adapters and allows sample barcode insertion adjacent to genomic DNA. 3'BL after single transposon tagmentation can theoretically achieve 100% usable template,

and our empirical data demonstrate that the new approach produced higher yield compared with traditional double transposon or Y transposon tagmentation. There are several explanations that the actual efficiency in Fig. 4d did not achieve 100%: The products of TnA&TnB contained not only the correct PCR products from Primer A and B, but also the incorrect ones from single primer A or B (blue bars from Fig. 4d). Therefore, the process efficiency increase might be underestimated, if we directly compare the PCR yield of TnA&TnB to that of TnA&AdB. On top of that, The AdB used in this test was not the adapters (Ad-T or Ad-A) with highest ligation efficiency, because we had to match the sequences from TnA/TnB and TnY for direct comparison. Wang et al. later changed the adapter sequence to achieve better process efficiency and described it in their sLFR paper. Based on NGS data, the 3'BL ligated genomic ends also demonstrated fewer positions with positional base composition bias, and the first 6-nt bias was mild and mainly caused by transposon insertion, suggesting that 3'BL has minimal positional bias. Using this new library construction method, Wang et al. successfully achieved highly accurate and complete variant calling in WGS and near-perfect phasing of variants into long contigs with N50 size up to 23.4 mb for long fragment reading (BioRxiv, <https://doi.org/10.1101/324392>). 3'BL was also incorporated in a novel mate-pair NGS library workflow to effectively cut down the laboratory steps and costs by Dong et al. (BioRxiv, <http://dx.doi.org/10.1101/400689>).

In this study, we also investigated 3'BL using templates of a chimeric DNA/RNA duplex that forms a 5' DNA and a 5' RNA overhang (Fig. 3). Unexpectedly, blunt-ended dsDNA was efficiently ligated to the 3' termini of RNA, but not DNA, suggesting that T4 ligase has a ternary complex formation preference. The ligation efficiency was greatly reduced if T4 RNA ligase I or II was used to join the ends. Some preliminary yet important applications of 3'BL with T4 DNA ligase are 3' end labelling of RNA, the enrichment of mRNA or the construction of targeted RNA libraries, especially for miRNAs, the small regulatory RNAs whose uncontrolled expression leads to a number of diseases.^{26,27} Thus, our 3'BL technique can be readily applied to the detection of cancer or Alzheimer's disease by monitoring targeted miRNA. Hybridization with DNA probes targeting the Poly(A) tail or specific miRNA sequences can be used to create DNA-RNA hybrids with a RNA 3'-recessive end, which is followed by ligation of RNA to adapter sequences with sample and/or UID barcodes through 3'BL. These common sequences can then be reverse transcribed to produce the cDNA of targeted RNA sequences. Compared with current miRNA capture technologies, the use of 3'BL mediated by T4 DNA ligase could potentially provide several advantages for NGS RNA library construction. First, hybridization with a DNA strand would prevent secondary structure formation by the RNA strand and therefore mitigate the bias introduced by other protocols. Second, T4 DNA ligase enables high-efficiency adapter addition through 3'BL, which avoids intramolecular RNA interactions that can be promoted by RNA ligases. Third, adapter dimers can be effectively eliminated, possibly rendering undesirable gel purification unnecessary. This new method could lead to improved unbiased microRNA expression profiling with simple and scalable workflows, and thus, large-scale research studies would become more affordable.

The findings of this study add to the growing understanding of T4 DNA ligase activities. We envision 3' branch ligation becoming a general tool in molecular biology that will advance the development of new nucleic acid engineering methods beyond described NGS applications.

Acknowledgements

We would like to acknowledge the ongoing contributions and support of all Complete Genomics and BGI-Shenzhen employees, especially Benjamin Allred for commenting the manuscript.

Funding

This work was supported by Shenzhen Peacock Plan (No. KQTD20150330171505310).

Supplementary data

Supplementary data are available at DNARES online.

Conflict of interest

Employees of BGI and Complete Genomics have stock holdings in BGI.

References

- Lehman, I.R. 1974, DNA ligase: structure, mechanism, and function, *Science (80-)*, **186**, 790–7.
- Tomkinson, A.E. and Mackey, Z.B. 1998, Structure and function of mammalian DNA ligases, *Mutat. Res.*, **407**, 1–9.
- Timson, D.J., Singleton, M.R. and Wigley, D.B. 2000, DNA ligases in the repair and replication of DNA, *Mutat. Res.*, **460**, 301–18.
- Ho, C.K., Wang, L.K., Lima, C.D. and Shuman, S. 2004, Structure and mechanism of RNA ligase, *Structure*, **12**, 327–39.
- Tomkinson, A.E., Vijayakumar, S., Pascal, J.M. and Ellenberger, T. 2006, DNA ligases: structure, reaction mechanism, and function, *Chem. Rev.*, **106**, 687–99.
- Pascal, J.M. 2008, DNA and RNA ligases: structural variations and shared mechanisms, *Curr. Opin. Struct. Biol.*, **18**, 96–105.
- Shuman, S. 2009, DNA ligases: progress and prospects, *J. Biol. Chem.*, **284**, 17365–9.
- Dickson, K.S., Burns, C.M. and Richardson, J.P. 2000, Determination of the free-energy change for repair of a DNA phosphodiester bond, *J. Biol. Chem.*, **275**, 15828–31.
- Cai, L., Hu, C., Shen, S., Wang, W. and Huang, W. 2004, Characterization of bacteriophage T3 DNA ligase, *J. Biochem.*, **135**, 397–403.
- Nilsson, S.V. and Magnusson, G. 1982, Sealing of gaps in duplex DNA by T4 DNA ligase, *Nucleic Acids Res.*, **10**, 1425–37.
- Goffin, C., Bailly, V. and Verly, W.G. 1987, Nicks 3' or 5' to AP sites or to mispaired bases, and one-nucleotide gaps can be sealed by T4 DNA ligase, *Nucleic Acids Res.*, **15**, 8755–71.
- Mendel-Hartvig, M., Kumar, A. and Landegren, U. 2004, Ligase-mediated construction of branched DNA strands: a novel DNA joining activity catalyzed by T4 DNA ligase, *Nucleic Acids Res.*, **32**, e2.
- Alexander, R.C., Johnson, A.K., Thorpe, J.A., Gevedon, T. and Testa, S.M. 2003, Canonical nucleosides can be utilized by T4 DNA ligase as universal template bases at ligation junctions, *Nucleic Acids Res.*, **31**, 3208–16.
- Kuhn, H. and Frank-Kamenetskii, M.D. 2005, Template-independent ligation of single-stranded DNA by T4 DNA ligase, *FEBS J.*, **272**, 5991–6000.
- Ho, C.K. and Shuman, S. 2002, Bacteriophage T4 RNA ligase 2 (gp24. 1) exemplifies a family of RNA ligases found in all phylogenetic domains, *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 12709–14.
- Bullard, D.R. and Bowater, R.P. 2006, Direct comparison of nick-joining activity of the nucleic acid ligases from bacteriophage T4, *Biochem. J.*, **398**, 135–44.

17. Broude, N.E., Sano, T., Smith, C.L. and Cantor, C.R. 1994, Enhanced DNA sequencing by hybridization, *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 3072–6.
18. Gunderson, K.L., Huang, X.C., Morris, M.S., Lipshutz, R.J., Lockhart, D.J. and Chee, M.S. 1998, Mutation detection by ligation to complete n-mer DNA arrays, *Genome Res.*, **8**, 1142–53.
19. Gorbacheva, T., Quispe-Tintaya, W., Popov, V.N., Vijg, J. and Maslov, A.Y. 2015, Improved transposon-based library preparation for the Ion Torrent platform, *Biotechniques*, **58**, 200.
20. Sgaramella, V. and Khorana, H.G. 1972, CXII. Total synthesis of the structural gene for an alanine transfer RNA from yeast. Enzymic joining of the chemically synthesized polydeoxynucleotides to form the DNA duplex representing nucleotide sequence 1 to 20, *J. Mol. Biol.*, **72**, 427–44.
21. Sgaramella, V. and Ehrlich, S.D. 1978, Use of the T4 polynucleotide ligase in the joining of flush-ended DNA segments generated by restriction endonucleases, *Eur. J. Biochem.*, **86**, 531–7.
22. Seguin-Orlando, A., Schubert, M., Clary, J., et al. 2013, Ligation bias in illumina next-generation DNA libraries: implications for sequencing ancient genomes, *PLoS One*, **8**, e78575.
23. Zhang, F., Christiansen, L., Thomas, J., et al. 2017, Haplotype phasing of whole human genomes using bead-based barcode partitioning in a single tube, *Nat. Biotechnol.*, **35**, 852.
24. Goryshin, I., Baas, B., Vaidyanathan, R. and Maffitt, M. 2016, *Oligonucleotide replacement for di-tagged and directional libraries*. 2011, Google Patents. US20140031261A1.
25. Adey, A. and Shendure, J. 2012, Ultra-low-input, tagmentation-based whole-genome bisulfite sequencing, *Genome Res.* **6**, 1139–43.
26. Bushati, N. and Cohen, S.M. 2007, microRNA functions, *Annu. Rev. Cell Dev. Biol.*, **23**, 175–205.
27. Mallory, A.C. and Vaucheret, H. 2006, Functions of microRNAs and related small RNAs in plants, *Nat. Genet.*, **38**, S31.