Triazole linking for preparation of a next-generation sequencing library from single-stranded DNA

Fumihito Miura[1,2,*], Tomoko Fujino[3], Kanako Kogashi[4], Yukiko Shibata[1], Miki Miura[1], Hiroyuki Isobe[3] and Takashi Ito[1,*]

[1]Department of Biochemistry, Kyushu University Graduate School of Medical Sciences, 3-1-1 Maidashi, Higashi-ku, Fukuoka 812-8582, Japan, [2]Precursory Research for Embryonic Science and Technology (PRESTO), Japan Science and Technology Agency (JST), 3-1-1 Maidashi, Higashi-ku, Fukuoka 812-8582, Japan, [3]Department of Chemistry, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan and [4]Department of Chemistry, Tohoku University, 6-3 Aramaki-aza-aoba, Aoba-ku, Sendai 980-8578, Japan

ABSTRACT

Next-generation sequencing of single-stranded DNA (ssDNA) is attracting increased attention from a wide variety of research fields. Accordingly, various methods are actively being tested for the efficient adaptor-tagging of ssDNA. We conceived a novel chemo-enzymatic method termed terminal deoxynucleotidyl transferase (TdT)-assisted, copper-catalyzed azide-alkyne cycloaddition (CuAAC)-mediated ssDNA ligation (TCS ligation). In this method, TdT is used to incorporate a single 3′-azide-modified dideoxyribonucleotide onto the 3′-end of target ssDNA, followed by CuAAC-mediated click ligation of the azide-incorporated 3′-end to a 5′-ethynylated synthetic adaptor. This report presents the first proof-of-principle application of TCS ligation with its use in the preparation of a next-generation sequencing library.

INTRODUCTION

Next-generation sequencing (NGS) has revolutionized the analysis of nucleic acids, thus contributing to a plethora of important biological findings (1). To make sample DNA readable with current NGS technologies, both ends of the DNA must be tagged with adaptors or synthetic oligodeoxyribonucleotides (ODNs) with defined sequences. Efficient methods are available for adaptor tagging of double-stranded DNA (dsDNA). For example, T4 DNA ligase is a classical enzyme used for ligating two ends of dsDNA and has been applied in a wide range of protocols. Tagmentation, a transposase-mediated adaptor tagging method, is a recently developed alternative for T4 DNA ligase-mediated ligation (2). Although the tagmentation reaction has a certain level of sequence preference or

bias, the simultaneous execution of fragmentation and insertion of the adaptor sequence into the target DNA is accomplished with unsurpassed efficiency (2).

By contrast, no *de facto* standard method has been established for adaptor tagging of single stranded DNA (ssDNA). For example, RNA ligases are able to join a 5′-phosphorylated adaptor to the 3′-hydroxyl end of ssDNA (3), and although it is known that RNA ligases are efficient at ligating 5′-phosphorylated ODNs to the 3′-end of RNA, their activities are significantly compromised when targeted to the 3′-end of DNA (4). In addition, RNA ligases have substantial preference to some terminal nucleotide bases, leading to biased ligation (5,6).

Homopolymeric tailing with terminal deoxynucleotidyl transferase (TdT) is another method of choice. Once attached to the 3′-end of ssDNA, the homopolymer can be used as a target sequence for an anchoring adaptor (7). Despite the highly efficient attachment of the homopolymeric tail by TdT, the homopolymer itself may be an obstacle in the downstream sequencing steps. Additionally, random priming-based methods can also be used for adaptor tagging of ssDNA. While random priming-based methods enable the highly efficient preparation of a sequencing library (8), they have a drawback of sampling bias against AT-rich sequences and are able to barely achieve an end-to-end coverage of the target DNA. Hence, all of the currently available methods for adaptor tagging to ssDNA have limitations in their efficiency, product structure, or sampling bias.

Various chemical reactions have been developed for the ligation of two molecules. Among them, copper-catalyzed azide alkyne cycloaddition (CuAAC), or so called 'click chemistry' (9), has become widely used in various fields of molecular biology (10,11). If a molecule is modified with an azido moiety, it can easily and with high specificity be connected to another molecule that has a terminal alkyne group in the presence of a Cu(I) catalyst. The reaction between

---

the two functional groups yields a 1,2,3-triazole (triazole linking) and importantly proceeds under conditions that are mild enough to preserve fragile biopolymers such as DNA. The use of CuAAC for ligating DNA molecules has been actively investigated (12–18). Intriguingly, DNA containing a triazole linking can serve as a template for DNA/RNA polymerases (12,14,16,18), and can be replicated and function as genetic material in bacterial cells (16).

Several types of DNA analogs with triazole linking have been synthesized that have different structures of ribose-to-ribose connections, and different biophysical properties (13,18). Spacing of the ribose-to-ribose connection is known to affect the stability of the duplex formed between analogs and natural DNA (13,18–20). Although natural DNA has five chemical bonds between the C3′ and C4′ atoms of adjacent sugar rings, analogs with four to seven bonds have been synthesized (14,16,18–20). Interestingly, the heteroduplex between natural DNA and analog DNAs, in which every phosphodiester bond is replaced with a five-bond triazole linkage, is extremely stable, in contrast to less-stable duplexes containing other analogs (19), demonstrating the impact of DNA backbone-periodicity on physical characteristics of DNA. Consistent with this notion, DNA analogs with five-bond triazole linking can be used as a primer for reverse transcription of mRNA (21), and have recently been shown to serve as a more efficient template for DNA polymerase than other DNA analogs (22). Of note, DNA containing a single triazole linkage in the middle of its sequence forms a less stable duplex with natural DNA than fully triazole-linked DNA does (22).

To establish an efficient method for the preparation of an NGS library from ssDNA, we investigated whether CuAAC generation of triazole linking could be employed as a basic ligation strategy for adaptor tagging of the 3′-end of ss-DNA. We demonstrated that TdT-mediated 3′-end incorporation of an azido moiety followed by CuAAC ligation with a synthetic 5′-ethynylated oligonucleotide provided a general strategy for adaptor tagging of ssDNA, which was applicable for NGS library preparation.

MATERIALS AND METHODS

Reagents

Nucleotide analogs 3′-azido-2′,3′-dideoxyribonucleotide-5′-triphosphate with adenine (Az-ddATP), cytosine (Az-ddCTP), guanine (Az-ddGTP), thymine (Az-ddTTP), and uracil (Az-ddUTP) were purchased from TriLink Biotechnologies, LLC (San Diego, CA). Az-ddATP was also purchased from Carbosynth (Compton, Berkshire, UK).

ODNs

The nucleotide sequences and modifications of the ODNs used in this study are summarized in Supplementary Table S1. Details regarding the chemical synthesis of the 5′-ethynylated ODN and ODNs containing a triazole linking are described in the Supplementary Procedures S1 and S2, respectively. All other ODNs were synthesized by Eurofins Genomics (Tokyo, Japan) and purified with oligonucleotide purification cartridges.

Double Sera-Mag purification

Removal of unreacted nucleotide analogs from ssDNA was performed based on solid-phase reversible immobilization (SPRI) (23) with several modifications as follows. The volume of solution that contains modified ssDNA was adjusted, as needed, with 10 mM Tris–HCl (pH 8.0) to a starting volume of 50 µl each. To the 50 µl solution, 2 µl of Sera-Mag Magnetic Carboxylate-modified Particles (GE Healthcare, Buckinghamshire, UK), 1 µl of 1 M MgCl$_2$, and 50 µl of 100% ethanol were added. After incubation at room temperature (approximately 21°C) for 10 min, the beads were collected on a magnetic stand to remove the supernatant, and rinsed with 70% ethanol. The DNA was eluted by suspending the beads in 50 µl of 10 mM Tris-acetate (pH 8.0). The second SPRI purification was performed by adding 1 µl of 1 M MgCl$_2$ and 50 µl of 100% ethanol to the bead suspension, and incubating the mixture for 10 min at room temperature. The supernatant was removed on a magnetic stand, and the beads were rinsed with 70% ethanol. The DNA was eluted by suspending the beads in an appropriate volume of 10 mM Tris-acetate (pH 8.0), and the supernatant was transferred to a new tube. The volume for each elution is specified for each experiment.

Double AMPure purification

Removal of short DNA fragments was performed by repeating AMPure XP purification as follows. The volume for the solution that contains DNA was adjusted, as needed, with 10 mM Tris–HCl (pH 8.0) to a starting volume of 50 µl. To the 50-µl samples, 90 µl of AMPure XP reagent (Beckman Coulter, Brea, CA, USA) was added. Following incubation at room temperature for 10 min, the beads were collected on a magnetic stand to remove the supernatant, and rinsed with 70% ethanol. The DNA was eluted by suspending the beads in 45 µl of 10 mM Tris-acetate (pH 8.0), and the supernatant was transferred to a new tube. To the eluent, 5 µl of 10 × TdT Buffer and 90 µl of AMPure XP reagent were added and incubated at room temperature for 10 min. Following the removal of the supernatant on a magnetic stand, the beads were rinsed with 70% ethanol and suspended in an appropriate volume of 10 mM Tris-acetate (pH 8.0), and the supernatant was transferred to a new tube. The volume for each elution is specified for each experiment.

Exchange of storage buffer for commercially available TdT

Initially, 1 × TdT buffer was prepared by diluting 10 × TdT buffer (500 mM HEPES-KOH, pH 7.5; 100 mM MgCl$_2$; and 5% Triton X100). TdT enzyme (300 units) purchased from Takara Bio Inc (Otsu, Japan) was added to 400 µl of 1 × TdT buffer, and filtered using an Amicon Ultra 0.5 ml centrifugal filter device with a nominal molecular weight limit of 3,000. The filter device was centrifuged at 15 000 × *g* for 30 min. After removal of the filtrate from the lower chamber, 400 µl of 1 × TdT buffer was added to the upper chamber, mixed by pipetting, and re-centrifuged as before. The solution in the upper chamber containing the TdT enzyme was collected and stored at 4°C until use.

3′-Azide modification of ssDNA

To 50 μl of solution that contained 1 × TdT Buffer, 20 μM Az-ddNTP, and ssDNA indicated in each experiment, 60 units of buffer-exchanged TdT enzyme were added, and the mixture was incubated at 37°C for 1 h. After heat inactivation of the TdT by incubation at 70°C for 10 min, 1 μl of 20 mg/ml Protease K (Qiagen, Hilden, Germany) was added and incubated at 50°C for 10 min. After heat inactivation of the protease at 95°C for 5 min, the DNA was purified using the double Sera-Mag procedure detailed above with an elution volume of 7 μl.

Click ligation of 5′-ethynyl adaptor to 3′-azide-modified ss-DNA

The Cu(I) catalyst solution was prepared fresh just prior to use by combining equal volumes of a solution of 50 mM CuSO$_4$ (Wako Pure Chemicals Co., Osaka, Japan) and 500 mM tris(3-hydroxypropyltriazolyl-methyl)amine (THPTA) (Sigma-Aldrich Co. LLC, St Louis, MO, USA) with a solution of 100 mM sodium ascorbate (Sigma-Aldrich) and 100 mM aminoguanidine (Tokyo Chemical Industry Co., Ltd., Tokyo, Japan). To the solution containing purified 3′-azide-modified DNA (7 μl), 1 μl of 100 μM ET-anti-PEA2 (Supplementary Table S1), and 40 μl of tertiary butyl alcohol (*tert*-BuOH) (Nacalai Tesque, Kyoto, Japan) were added. The coupling reaction was started by adding 2 μl of the Cu(I) catalyst solution to the mixture, and incubating the tube at 25°C for 15 min. The ligated DNA product was directly captured with magnetic beads by adding 2 μl of Sera-Mag magnetic particles to the reaction and incubating at room temperature for 10 min. Following removal of the supernatant and rinsing the beads with 70% ethanol, the product DNA was eluted in 20 μl of 10 mM Tris-acetate (pH 8.0).

Synthesis of complementary strand for adaptor-tagged DNA

Purified adaptor-tagged DNA (20 μl) was combined with 5 μl of 10 × NEBuffer 2 (New England BioLabs, Ipswich, MA), 5 μl of 2.5 mM dNTPs (Takara Bio Inc), and 2 μl of 100 μM PEA2 (Supplementary Table S1). The total reaction volume was adjusted to 50 μl with water and sequentially incubated at 94°C for 3 min, 50°C for 3 min, and 37°C for 3 min. The reaction was supplemented with 1 μl (20 units) of Exonuclease I (New England BioLabs) and incubated at 37°C for 15 min. Then, the reaction was supplemented with 1 μl (50 units) of Klenow fragment exo (−) (New England BioLabs) and further incubated for 15 min at 37°C. After heat inactivation of the enzymes at 70°C for 10 min, the DNA was purified by the double AMPure procedure detailed above and recovered as a 10-μl eluate.

NGS library preparation from synthetic DNA

The 2nd adaptor solution containing 1 × PCR Buffer (10 mM Tris–HCl, pH 8.3; 50 mM KCl; and 2 mM MgCl$_2$), 45 μM P-anti-PEA1-P (Supplementary Table S1) and 45 μM PEA1T (Supplementary Table S1) was prepared beforehand by incubating it at 94°C for 3 min, followed by 50°C for 5 min.

Random hectamer (N$_{100}$) (5 pmol) spiked with or without a mixture of five specific 100-mers (10 fmol each, Supplementary Figure S9) was subjected to 3′-azido modification and adaptor ligation as described above. Following synthesis of the complementary DNA (cDNA) to the adaptor-tagged N$_{100}$ as described above, a second adaptor-ligation using T4 DNA ligase was performed. For the T4 DNA ligation, 10 μl of the solution containing the adaptor-tagged double-stranded DNA, 12 μl of 2 × Quick Ligase Reaction Buffer (New England BioLabs), and 1 μl of second adaptor solution were combined. The ligation reaction was started by adding 1 μl of Quick Ligase (New England BioLabs) and proceeded by incubation for 15 min at room temperature. After heat inactivation of the enzyme at 70°C for 10 min, the DNA was purified using the double AMPure procedure as described above and recovered in a 20-μl volume of eluate.

Library preparation for MNase-seq

The budding yeast *Saccharomyces cerevisiae* strain S288C, obtained from the Biological Resource Center at the National Institute of Technology and Evaluation (NBRC1136), was inoculated into 3 ml of YPD medium (1% (w/v) yeast extract, 2% (w/v) Bacto Peptone, 2% (w/v) glucose) and incubated overnight at 30°C with shaking. The overnight culture was diluted in 15 ml of fresh YPD and the optical density at 600 nm was adjusted to 0.5. The diluted cells were grown at 30°C for 2 h with shaking. The cultured yeast cells were collected with centrifugation at 2,000 × *g* for 5 min, and re-suspended in 1 ml of spheroplast buffer that contained 1 M sorbitol; 50 mM Tris–HCl, pH 7.5; and 1 mM β-mercaptoethanol. The cell suspension was supplemented with 20 μl of 10 mg/ml Zymolyase-100T (Seikagaku Kogyo, Tokyo, Japan) and incubated at 30°C for 10 min to make spheroplasts. The spheroplasts were collected by centrifugation at 2000 × *g* for 5 min and resuspended in 1 ml of MNase digestion buffer (50 mM Tris–HCl, pH 7.6; 1 mM CaCl$_2$; 0.2% (v/v) Triton X-100) containing 1 × protease inhibitor (Nacalai Tesque). The lysate was digested with 50 units of micrococcal nuclease (Takara Bio Inc.). After incubation at 37°C for 10 min, the digestion was stopped by adding 100 μl of 500 mM EDTA. The reaction mixture was then supplemented with 100 μl of 10% (w/v) SDS and 25 μl of 20 mg/ml Protease K (Qiagen), and incubated at 50°C for 1 h. Following phenol and phenol/chloroform extraction, nucleic acids were collected by isopropanol precipitation. Any contaminating RNA was digested with 250 units of RNase If (New England BioLabs) and the 3′-phosphate of the MNase-treated DNA was removed with 5 units of shrimp alkaline phosphatase (Takara Bio Inc.). The final DNA purification was accomplished using a QIAQuick PCR purification kit (Qiagen) according to the manufacturer's instructions. The library was prepared from 350 ng of the purified DNA using the same protocol as described above for the library preparation from the synthetic DNA. A conventional dsDNA library was prepared using a ThruPlex DNA-seq 6S Kit from Rubicon Genomics according to the manufacturer's instructions.

Library amplification

The purified sequencing library prepared as described above was combined with 25 μl of 2 × PrimeStar Max Premix (Takara Bio Inc.), and 0.4 μl each of 100 μM Primer-3 and 100 μM Index-1 (Supplementary Table S1), and the total volume was adjusted to 50 μl with water. PCR amplification was performed with the following conditions: 98°C for 1 min; 5 or 15 cycles of 98°C for 15 s, 55°C for 30 s and 72°C for 30 s; 72°C for 3 min. Following amplification, 1 μl (20 units) of exonuclease I was added to the PCR reactions and incubated at 37°C for 15 min followed by 70°C for 10 min. The library DNA was purified by the double AMPure procedure described above in a final 20 μl volume.

Next-generation sequencing

The molar concentration of the library DNA was determined using a Library Quantification Kit (Takara Bio Inc.) according to the manufacturer's instructions. Sequencing was performed with single-end mode (76 cycles, 151 cycles, and 300 cycles for budding yeast's MNase-Seq, random hectamer (N100), and N100 spiked with specific sequences, respectively) on an Illumina MiSeq benchtop sequencing system using a MiSeq Reagent Kit v3 (Illumina, San Diego, CA) according to the manufacturer's instructions. The sequencing reads for library prepared from synthetic DNA have been submitted to the NCBI Sequence Read Archive (SRA) under accession number SRP133409, whereas reads and processed files for budding yeast's MNase-seq have been submitted to the NCBI Gene Expression Omnibus (GEO) under accession number GSE110666.

Analysis of MNase-seq reads

The *S. cerevisiae* S288C genome sequence and annotations were downloaded from the Saccharomyces Genome Database as a single gff3 file (https://downloads.yeastgenome.org/curation/chromosomal_feature/saccharomyces_cerevisiae.gff; 11 November 2011). MNase-seq reads were mapped to the reference yeast genome using bowtie version 0.12.7 with default parameters except that the output format was specified in SAM format and the number of processors used was 12 [24]. The SAM files were sequentially processed using in-house programs. The source code and detailed procedures are available through GitHub (https://github.com/FumihitoMiura/Project-1).

RESULTS

TdT-mediated incorporation of an azide-functional group onto the 3′-end of ssDNA

We examined whether chemical reactions such as CuAAC ligation could replace enzymatic reactions for adaptor tagging of ssDNA originating from various biological samples. CuAAC ligation of two ssDNA molecules requires that one of the molecules be modified with either an azido or ethynyl moiety on its 3′-end, and the other molecule be modified with the corresponding functional group on its 5′-end. In order to introduce a functional group onto the 3′-end of ssDNA, we used TdT because it can readily incorporate various modified nucleotides into ssDNA. Notably, the use of an elongation terminator nucleotide can prevent the formation of homopolymeric tails. It is also important that the structure created by the chemical ligation can serve as an efficient template for DNA polymerases.

Based on these considerations, we chose to use Az-ddNTPs (Figure 1A). Consistent with a previous study (25), TdT successfully utilized Az-ddNTPs as a substrate to attach single nucleotides to the 3′-ends of a model ODN, leading to an electrophoretic mobility shift (Figure 1B and C). The modified ODN appeared to retain the azido moiety because it demonstrated a super mobility shift when mixed with dibenzocyclooctyne–PEG4–biotin conjugate (DBCO–PEG4–biotin), which is known to specifically react with azido moieties (Figure 1B and C). The incorporation efficiency of the nucleotide analog (Supplementary Figure S1) was dependent on not only the incorporated nucleotide bases as described previously (25), but also the base composition at the 3′-terminal position of the target ODN (Supplementary Figure S2). However, we found that the use of excess TdT improved the reaction efficiency to be near-quantitative and enhanced reproducibility (Supplementary Figure S3), thereby making the effects of the biased incorporation largely negligible (Figure 1C). In addition, we found that the presence of Triton X100 in the reaction contributes to better yields (Figure 1B and C).

It was evident that unreacted Az-ddNTPs were competitive inhibitors of azide-modified ssDNA in the following CuAAC ligations. Additionally, the amount of DNA from biological sources is usually limited. Accordingly, a method capable of separating a limited amount of ssDNA from free nucleotide analogs, along with high efficiencies in both recovery and purity of the end product, was desired. Recently, methods based on SPRI (23) have been frequently used in the preparation of NGS libraries. SPRI-based methods are not only sensitive enough to be applicable to sub-nanogram amounts of DNA but also reliable in terms of yield and reproducibility. We, therefore, attempted to establish an SPRI-based protocol for the removal of unreacted nucleotide analogs from the ssDNA product. We refer to the method as the double Sera-Mag procedure. We found that unidentified components of the storage solution for commercially available TdT caused poor recovery of the ssDNA product and reduced efficiency in the removal of unreacted nucleotide analogs (data not shown). We also discovered that the presence of TdT in the solution reduced the recovery of ssDNA in SPRI purification through some unknown mechanism (data not shown). Improved recovery was achieved by exchanging the storage solution for TdT prior to use, and by treatment with protease K following the TdT reaction. The double Sera-Mag procedure, including the buffer exchange and proteinase K treatment, reproducibly removed the unreacted nucleotide analogs to an undetectable level, and allowed 65% recovery of the ssDNA product (Supplementary Figure S4).
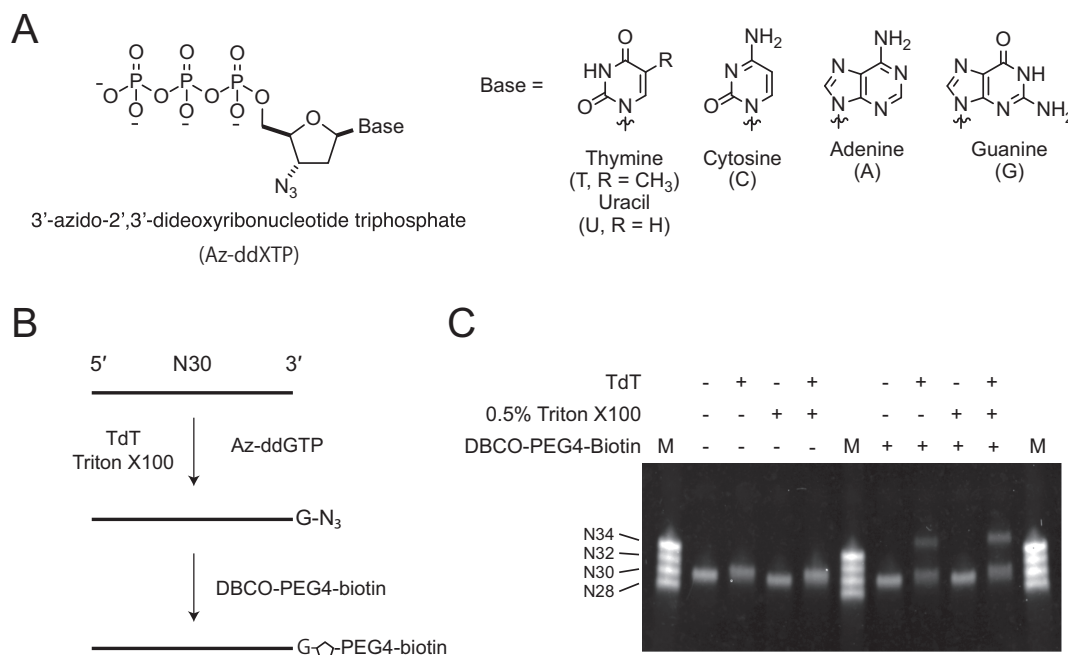
Figure 1. TdT-mediated addition of an azide-functional group to the 3′-end of ssDNA. (**A**) Chemical structure of the 3′-azide-modified dideoxyribonucleotide analogs. (**B**) Schematic overview of the assay for TdT-mediated incorporation of azide-modified nucleotide analogs. (**C**) Denatured polyacrylamide gel electrophoresis using 10% Novex TBE-Urea Gel (Invitrogen) analysis of incorporation of azide-modified nucleotide analogs onto the 3′-end of ssDNA. Incorporation of nucleotide analog induced electrophoretic mobility shifts (left four lanes). The incorporated azido-moiety was coupled with DBCO–PEG–biotin, leading to a super shift (right four lanes). After electrophoresis, the gel was stained with SYBR Gold Gel stain (Invitrogen) and image was taken with using a ChemiDoc system (Bio Rad Laboratories, Hercules, CA).

Click ligation between 3′-azido and 5′-ethynyl ODNs

Having established a general method for modification of the 3′-end of ssDNA with an azido moiety, we examined the use of a synthetic ODN with a terminal alkyne moiety on the 5′-end that could be click-ligated to the 3′-azide modified ssDNA. The number of bonds between C3′ and C4′ atoms of adjacent sugar rings significantly affects the biophysical properties of DNA that contains triazole linking (13,18–20). Because DNA with triazole linking containing five bonds forms the most stable duplex with natural DNA molecules, we chose the same linkage structure as the one used in the triazole-linked analogue of DNA (TLDNA) (19). We synthesized an ODN modified with an ethynyl functional group at the C5′ atom of the most 5′-terminal nucleotide (Figure 2A and Supplementary Table S1). Following the synthesis of a 33-mer ODN on controlled pore glass beads using an automated DNA synthesizer, 5′-ethynyl-thymidine (19) was loaded using a standard phosphoramidite method (Supplementary Information 1). The efficiency of elongation of the nucleotide was comparable to that of natural ODNs, with the target 34-mer ODN containing the 5′-ethynyl group generated with good yield. The structure of the 34-mer ODN was confirmed by high-resolution mass spectrometry analysis.

We next investigated the conditions for click ligation with CuAAC between the azido and the ethynyl ODNs. Because Cu(I) is unstable, and *in situ* generation of Cu(I) from Cu(II) using sodium ascorbate generates hydroxyl radicals that degrade DNA (26), ligands for both stabilizing Cu(I) and repressing the production of hydroxyl radicals are crucial

for CuAAC reactions with biopolymers (10,11,27–29). We chose THPTA (11) because it was the sole compound with the expected effects under the aqueous conditions of our system (data not shown).

There are two strategies for click ligation, namely templated ligation and non-templated ligation (17). Templated ligation is efficient but is not applicable to this study, which aims at a general protocol applicable to DNA molecules with unknown 3′-terminal sequences. Accordingly, we explored non-templated ligation, which is inefficient and requires a DNA concentration of more than tens of micromolar (17). Indeed, click ligation between 5 nM azido-modified ODN and 0.5 μM ethynylated ODN in aqueous solvent failed to yield a detectable amount of ligation product (Supplementary Figure S5). Ligation efficiency could have been improved by increasing the concentration of DNAs, but it is also true that such a high concentration is far from practical for most if not all NGS applications in biomedical samples. Thus, we intensively sought a condition for more efficient non-templated click ligation under practically relevant DNA concentrations. We found that addition of some organic solvents to the reaction greatly enhanced the efficiency of CuAAC ligation (Supplementary Figure S5). Among the solvents tested, *tert*-BuOH enhanced the ligation most effectively (Supplementary Figure S5). We found that ODNs were degraded under some buffer conditions, even in the presence of THPTA (Supplementary Figure S6), and that near-neutral conditions were suitable for both ligation and the preservation of DNA integrity (Supplementary Figure S6). The reactions appeared to terminate within 15 min at room temperature, as extending the
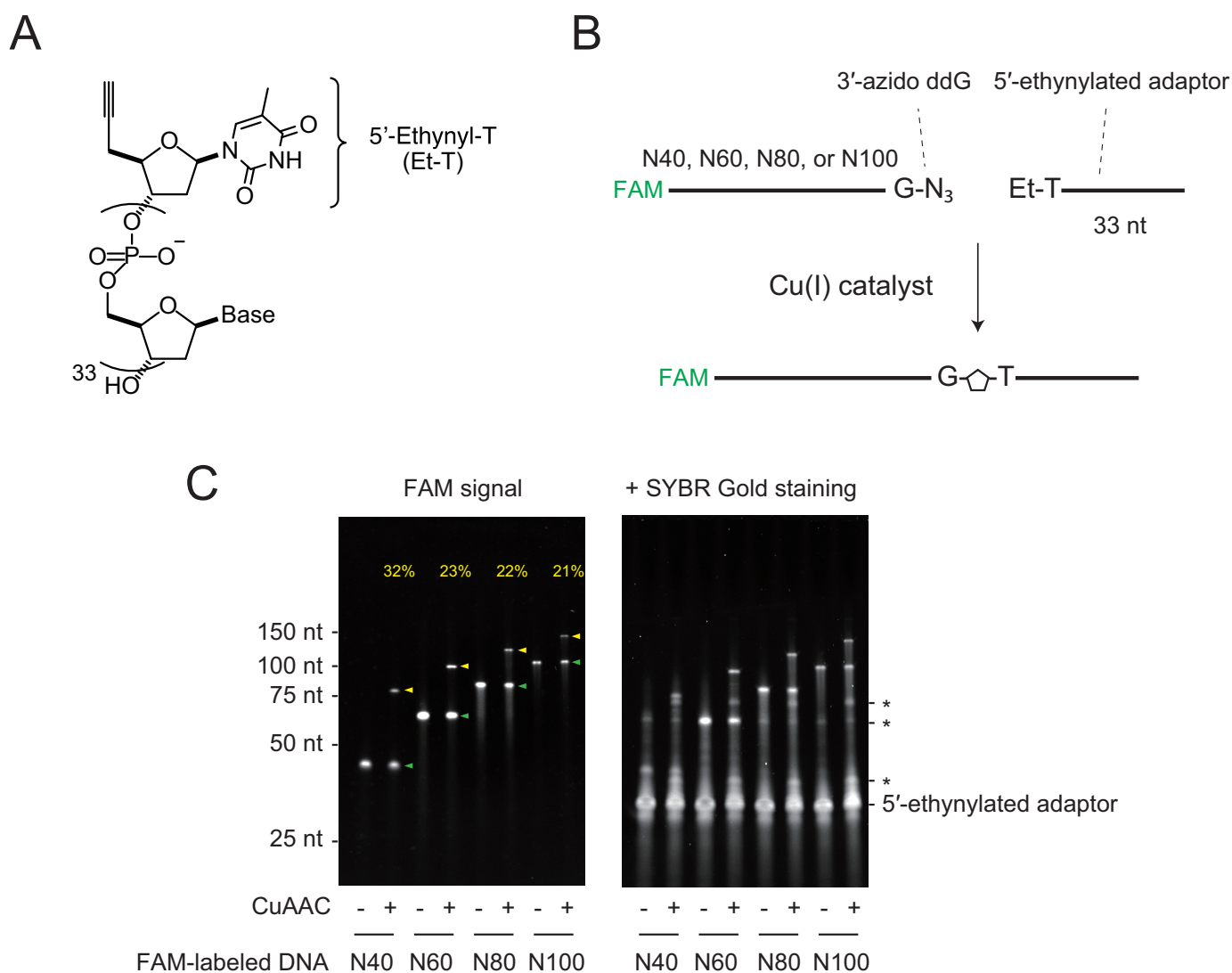
Figure 2. Ligation of the 5′-ethynylated adaptor with 3′-azide labeled ssDNA. (**A**) Chemical structure of the 5′-ethynylated adaptor. (**B**) Schematic overview of the assay for CuAAC ligation between the 3′-azido labeled ssDNA and the 5′-ethynylated adaptor. (**C**) CuAAC-mediated ligation. Four differently sized ODNs were used for the experiment. Left and right panels indicate the images for the 5′-FAM signal and SYBR Gold staining of the same gel, respectively. The green and yellow arrowheads indicate the 5′-FAM-labeled, 3′-azide-modified target DNA and the ligation products, respectively. Ligation efficiency is denoted for each reaction. Each reaction contained a 20-fold molar excess of 5′-ethynylated adaptor over target DNA. Asterisks on the right panel indicate byproducts of unknown structure derived from the 5′-ethynylated adaptor.

reaction time and increasing the temperature to 50°C had minimal enhancement effects on the coupling reaction (data not shown). Optimization of parameters led to the establishment of the most efficient reaction conditions with 20–30% efficiency (Figure 2C). Of note, DNA was readily recovered from the aqueous-organic solvent using Sera-Mag beads (see Methods section). This newly developed ligation strategy is referred to as TdT-assisted CuAAC-mediated ss-DNA ligation or TCS ligation.

Synthesis of complementary DNA for the triazole-linked DNA

The TCS ligation product contained a triazole linkage called Tz3 (22), which is the same linkage structure as ^{TL}DNA (19). A recent study demonstrated that Tz3-

containing DNA serves as a better PCR substrate than DNA containing other types of triazole linking (22). However, primer extension assays performed in the study showed that the efficiency of read-through over Tz3 was low for both Taq DNA polymerase and Phusion DNA polymerase (22). Only the Klenow fragment harbors a detectable read-through activity for Tz3 (22). Independently, we examined 16 DNA polymerases for their abilities to read through Tz3 in primer extension assays. We designed two ODNs as Tz3-containing templates: one with thymidine and the other with uridine on the 3′-side of the triazole linking (Figure 3A, Supplementary Information 2). Although 14 of the polymerases failed to produce extension products, two exhibited adequate read-through activity on the Tz3-containing templates (Supplementary Figure S7). One was the Klenow fragment, consistent with a previous report
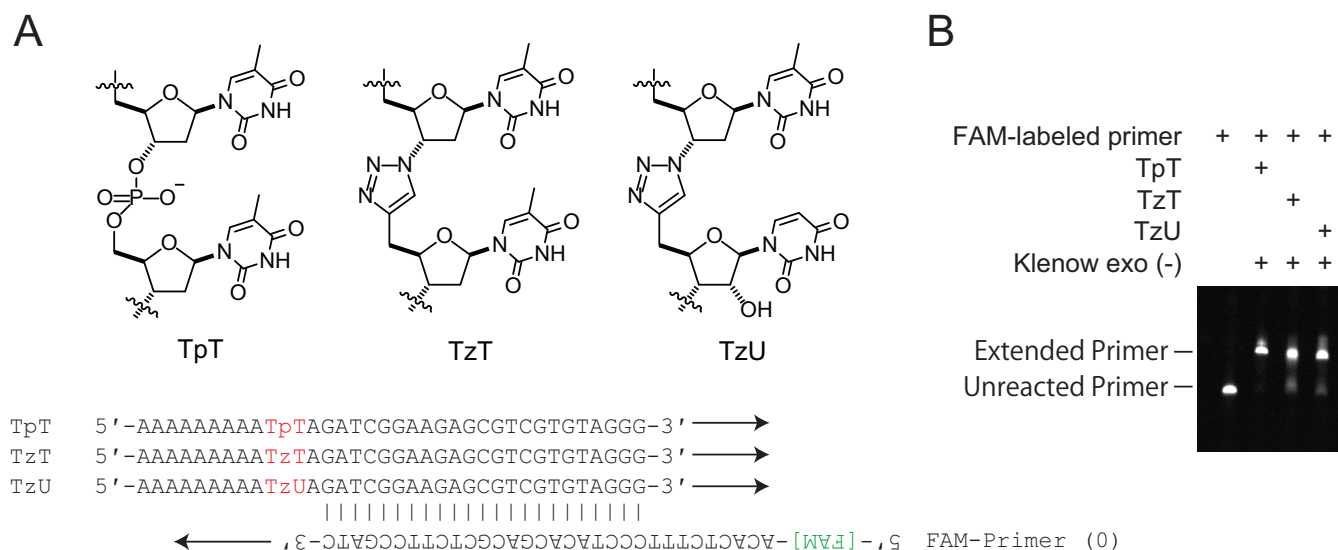
Figure 3. Enzymatic synthesis of DNA complementary to triazole-containing DNA. (**A**) Chemical structures and sequences of the template DNA. TpT contains no triazole linking (left), whereas TzT and TzU contain a single triazole linking (middle and right). Note that the triazole linkage is followed by thymidine and uridine (i.e., ribonucleoside) in TzT and TzU, respectively. Schematic representation of the primer extension assay (bottom). (**B**) Primer extension by the Klenow fragment exo ($-$).

(22), and the other was the mutant Klenow fragment exo ($-$) that lacks 3′-to-5′ exonuclease activity (Supplementary Figure S7). Both Klenow fragments did not show any apparent preference between the downstream thymidine and uridine nucleosides (Supplementary Figure S7). The use of excess Klenow fragment exo ($-$) resulted in effective DNA synthesis on the Tz3-containing ODN, leading to an efficiency exceeding 80% under optimized conditions (Figure 3B).

Preparation of a next-generation sequencing library from synthetic DNA

Once our novel procedure was established to attach an adaptor to the 3′-end of ssDNA followed by its conversion into dsDNA, the conventional T4 DNA ligase-based procedure was expected to be applicable for the attachment of the other adaptor molecule to the opposite end of the dsDNA. The low reaction temperature of T4 DNA ligase would be preferable for duplex DNA containing a single triazole-linkage because its thermal stability might be lower than that of natural DNA (22). Our goal was to prepare an NGS library from ssDNA by using TCS ligation, followed by complementary strand synthesis and then dsDNA ligation with T4 DNA ligase. Using chemically synthesized random hectamer ODNs (N_{100}), we attempted to determine the overall efficiency of adaptor tagging. We chose Az-ddGTP for the 3′-azido modification (Figure 4A) because TdT incorporated Az-ddGTP most efficiently (Supplementary Figure S1). Following the removal of unreacted Az-ddGTP using the double SPRI procedure combined with Protease K treatment, the 5′-ethynyl adaptor was conjugated to the 3′-azide-modified N_{100} with CuAAC (Figure 4A). An ODN complementary to the adaptor was then annealed to the adaptor-tagged ssDNAs to prime DNA synthesis by the Klenow fragment exo ($-$) (Figure 4A). The

enzyme has an optimum temperature at which the stringency of primer annealing may be compromised and hence may generate non-specific extension products. To minimize the non-specific extension products, we treated the primer-annealed DNAs with *Escherichia coli* exonuclease I prior to addition of the Klenow fragment. This treatment degraded the primers that were either free in solution or incompletely annealed to the DNAs, but not the primers correctly annealed to the adaptor sequence. Following SPRI purification, we used T4 DNA ligase to ligate a dsDNA adaptor to the opposite end of N_{100} that had been converted to its double-stranded form (Figure 4A). Finally, the adaptor-tagged products were amplified with five cycles of PCR, which also made the adaptor structure complete for Illumina sequencing (Figure 4A). The protocol included 12 operational steps and yielded 0.78 pmol of library DNA, starting with five pmol of input DNA (Figure 4A and Supplementary Figure S8). This value indicated that the overall efficiency for adaptor tagging of both ends of the model ODNs was 0.6%, assuming a 100% efficiency of each PCR cycle.

We next analyzed the produced library and determined its nucleotide sequences. As shown in Figure 4B, approximately a quarter (24.2%) of the sequenced reads were 102 nucleotides in length after trimming of the adaptor sequences, which corresponded to the exact size of the random hectamer plus the two specific bases incorporated by the adaptor tagging steps. In contrast, more than half of the reads (57.8%) were shorter than the expected 102 nucleotides, ranging from 101 to 90 nucleotides in length (Figure 4B). Because >77% of reads were 102 nucleotides in length for a library prepared only with automatic DNA synthesizer (Supplementary Figure S9), the lower ratio for reads of 102 nucleotide in the TCS-ligation-based method indicates the presence of deletion events in TCS-ligation. Therefore, we investigated the cause of the shortened read
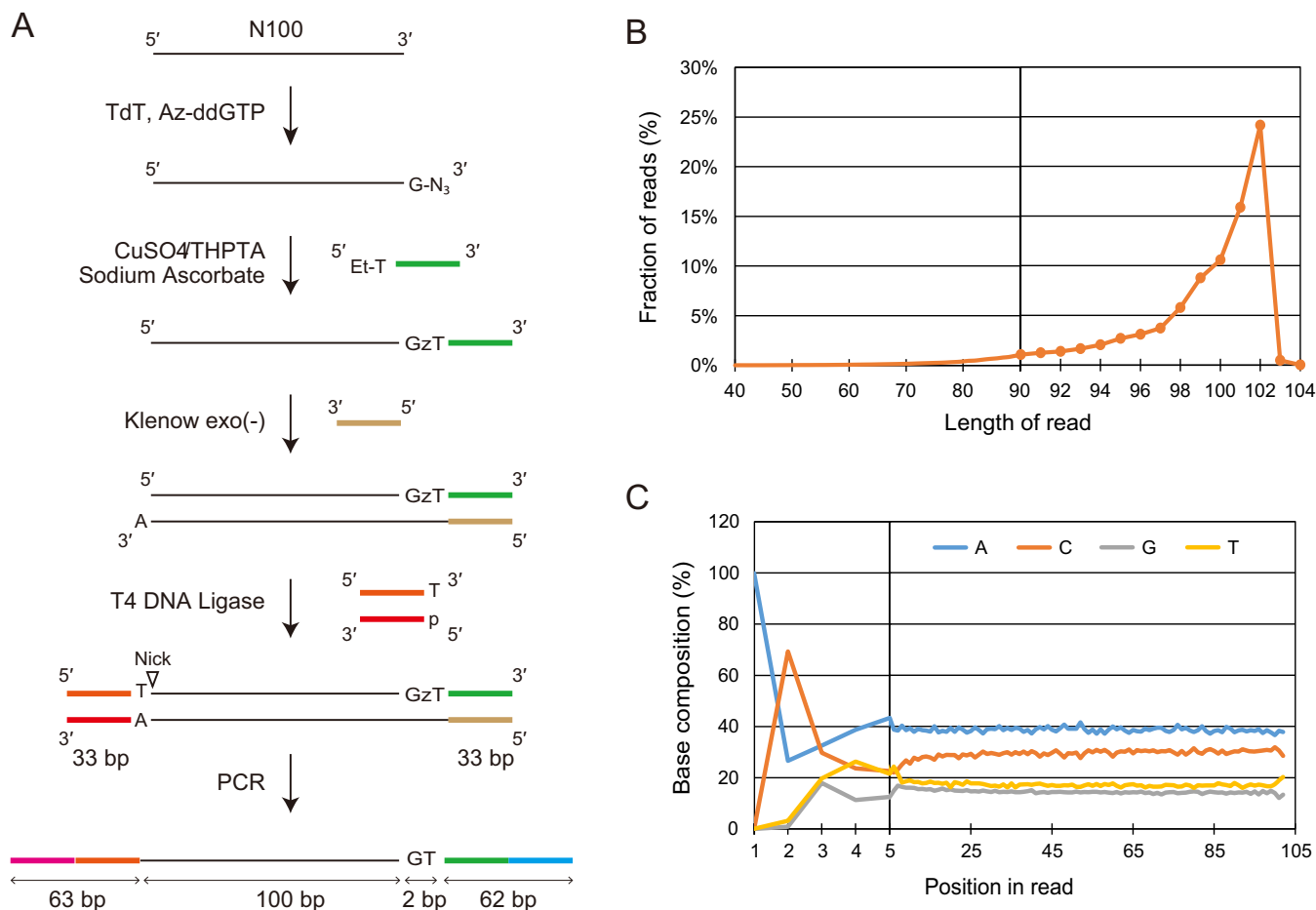
Figure 4. Library preparation from chemically synthesized DNA. (**A**) Schematic overview for the library preparation. The sequencing library was prepared from chemically synthesized random 100-mer ODNs. (**B**) Distribution of the length of reads after trimming of adaptor sequences. (**C**) Mean base composition of reads at each position. Sequencing was performed from right to left in panel A.

lengths. Although a previous study reported that the Tz3-containing template occasionally causes the Klenow fragment to induce a single-nucleotide interstitial deletion at the site complementary to the 3′-adjacent position of Tz3 in a local base-dependent manner (22), this type of point deletion could not fully explain the larger deletions of two or more nucleotides (Figure 4B). Because the random hectamer-based assay could not identify the exact positions of the deletions, we analyzed the deletion profiles of five synthetic 100-mer ODNs of defined sequences spiked in N$_{100}$. As shown in Supplementary Figure S10, >99% of the reads showed no evidence of deletions at the 3′-adjacent position of the Tz3. We concluded that the point deletion described in a previous study (22) was rare under the conditions used in our study. Conversely, deletions were more frequently observed at the distal positions from Tz3 or the 3′ end of DNA (Supplementary Figure S10). We suspect that these terminal deletions were attributable to degradation of the DNA during CuAAC ligation.

To examine whether the library was constructed in an unbiased fashion, we analyzed the nucleotide composition of the sequenced reads. As shown in Figure 4C, 99.9% of the first position of the reads were A, which was complementary to the 5′-ethynylated T (Figure 2A and Supple-

mentary Table S1). In contrast, only 69.4% of the reads had C for the second position, which was expected to be C because it is the complementary site of the azide-modified G (Figure 4C). The most common of the other nucleotide bases incorporated at this position was A (26.6%). Because the point deletion at 3′-adjacent position of triazole linking was a rare event in our system (Supplementary Figure S10), this observation indicated nucleotide substitutions. It thus seemed that Tz3 in DNA may be mutagenic during DNA synthesis by the Klenow fragment exo (−). The mean base composition of the region spanning from the 3rd to 102nd positions was 37.8 ± 2.0%, 28.0 ± 2.2%, 15.4 ± 1.1% and 18.8 ± 2.7% for A, C, G and T, respectively (Figure 4C). Although these values substantially differed from the ideal composition of random nucleotide bases (i.e., 25% for each), sequencing of a library prepared by only automated phosphoramidite-based chemical synthesis displayed an almost identical base composition, with 37.0 ± 0.5%, 25.4 ± 0.7%, 17.1 ± 0.4% and 20.4 ± 0.3% for A, C, G and T, respectively (Supplementary Figure S9). We concluded that the nucleotide composition that was apparently biased from the ideal random hexamers originated during the chemical synthesis of DNA rather than during the library preparation. We also noted that the nucleotide composition was

somewhat distorted from the average at the third, fourth, and fifth positions of the reads (Figure 4C). Because no similar deviations were evident in the control library (Supplementary Figure S9), the 3′-proximal region of the target DNA likely influenced the efficiency of the TCS ligation-based ssDNA library preparation.

Application of TCS ligation to MNase-seq library preparation

We applied the TCS ligation to the preparation of an NGS library from a biological sample. We focused on enzyme-mediated footprinting because nuclease-treated DNA may contain a substantial number of nicks that lead to the generation of short ssDNA. We treated budding yeast chromatin with micrococcal nuclease (MNase) (Supplementary Figure S11) and prepared two MNase-seq libraries, one using a ThruPlex kit (Rubicon Genetics) designed for dsDNA, and the other using the TCS ligation-based protocol identical to the one used for the synthetic ODNs (Figure 4A). Although both protocols resulted in the successful preparation of libraries, the ThruPlex kit outperformed the TCS ligation-based protocol for yield of library product. The ThruPlex kit required 4 cycles of PCR amplification and the TCS ligation required 15 cycles, even when starting with equal amounts of input DNA. We sequenced both libraries using the Illumina MiSeq system and obtained 24.0 and 31.2 million mapped reads from the libraries prepared by ThruPlex kit and TCS-ligation based protocol, respectively. Meta-analysis of the gene promoter regions revealed that the mean cut frequency patterns were consistent between the conventional and TCS ligation-based protocols (Figure 5). We concluded that TCS ligation could be used for NGS library preparation, although there is scope for significant optimization of its efficiency.

DISCUSSION

In this study, we investigated whether CuAAC could be used as a general ligation strategy in the preparation of an NGS library from ssDNA. The TCS ligation method was developed to overcome various limitations inherent to the currently available methods for adaptor tagging to ssDNA, including the low efficiency of RNA ligase-based methods, homopolymer-induced drawbacks of TdT-mediated methods, and GC-biased coverage of random priming-based methods. Because TCS ligation attached only a single nucleotide to the 3′-end of target ssDNA, it could readily circumvent the problems inherent to homopolymer tailing in conventional TdT-mediated methods. In addition, TCS ligation was less affected by GC content than random priming-based methods, although it showed a certain level of preference for some sequences in the proximal regions of the 3′-end of target DNA (Figure 4C and Supplementary Figure S9C). Efficiency of library preparation by TCS ligation remained rather limited, leaving room for improvement. Nevertheless, we successfully applied TCS ligation to prepare a yeast MNase-seq library that revealed nucleosome profiles comparable to those obtained by standard methods. These results demonstrated the potential of TCS ligation in NGS library preparation.

Several improvements are expected to make TCS ligation-based library preparation more efficient. First, the entire protocol should be simplified. As shown in Supplementary Figure S8, the current protocol is composed of 12 steps in total, including not only CuAAC ligation, but also five purification steps. Because the large number of steps inevitably reduced yields, reduction in these steps may be critical for improvement of the methodology. Second, degradation of DNA during CuAAC ligation should be further controlled. Although we used THPTA to suppress DNA degradation (11), the length distribution of the sequenced reads indicated that the suppression was not complete (Figure 4B). Longer template DNA is more difficult to preserve. Accordingly, further optimization of CuAAC ligation and/or development of a new ligand will be crucial for the preservation of DNA integrity. Third, DNA polymerase selection should be optimized. Our primer extension assays revealed that only two out of the 16 DNA polymerases tested could utilize the Tz3-containing DNA as a template for DNA synthesis (Figure 3 and Supplementary Figure S7). Because the two successful polymerases are mesophilic, it was difficult to keep the stringency of primer annealing conditions at or below their optimal temperatures. We successfully circumvented this problem by introducing an exonuclease I digestion step, which eliminated free and incompletely annealed primers, prior to DNA polymerization. However, this step made the protocol more complicated. Therefore, it would be ideal to identify, or even develop, thermophilic DNA polymerases that are able to efficiently use Tz3-containing DNA as a template.

A recent study (30) reported a protocol termed Click-Seq that uses CuAAC for RNA-seq library preparation. In ClickSeq, Az-ddNTP mixed in dNTPs serves as a chain terminator of cDNA synthesis, thereby eliminating the need for cDNA fragmentation. Because each truncated cDNA molecule bears an azido moiety at its 3′-end, CuAAC is able to conjugate it with an adaptor modified with a hexynyl functional group at its 5′-phosphate. Although ClickSeq can produce a largely unbiased RNA-seq library, the authors noted a low efficacy in amplification of the click-ligated products (30). Our results suggest that the combined use of 5′-ethynyl adaptor and the Klenow fragment may enhance the power of ClickSeq.

Different procedures are actively being investigated for efficient adaptor tagging of ssDNA. For instance, two methods were recently reported for the direct joining of an adaptor sequence to the 3′-end of ssDNA. One uses a thermostable RNA ligase (31), whereas the other uses T4 DNA ligase for ssDNA ligation mediated by a splinter ODN with a stretch of random bases hybridized to the 3′-end portion of target ssDNA (32). Although the RNA ligase-based method is limited in its applicability to long ssDNA fragments, it has a high efficiency for short ssDNA fragments (31,32). The T4 DNA ligase-based method, called ssDNA 2.0, was shown to be highly efficient and less biased (32). The TCS ligation reported in the current study is distinct from these and other fully enzymatic methods in that it is a hybrid approach, combining chemical and enzymatic reactions. TCS ligation provides a prototype for other chemo-enzymatic methods to be developed for NGS library preparation.
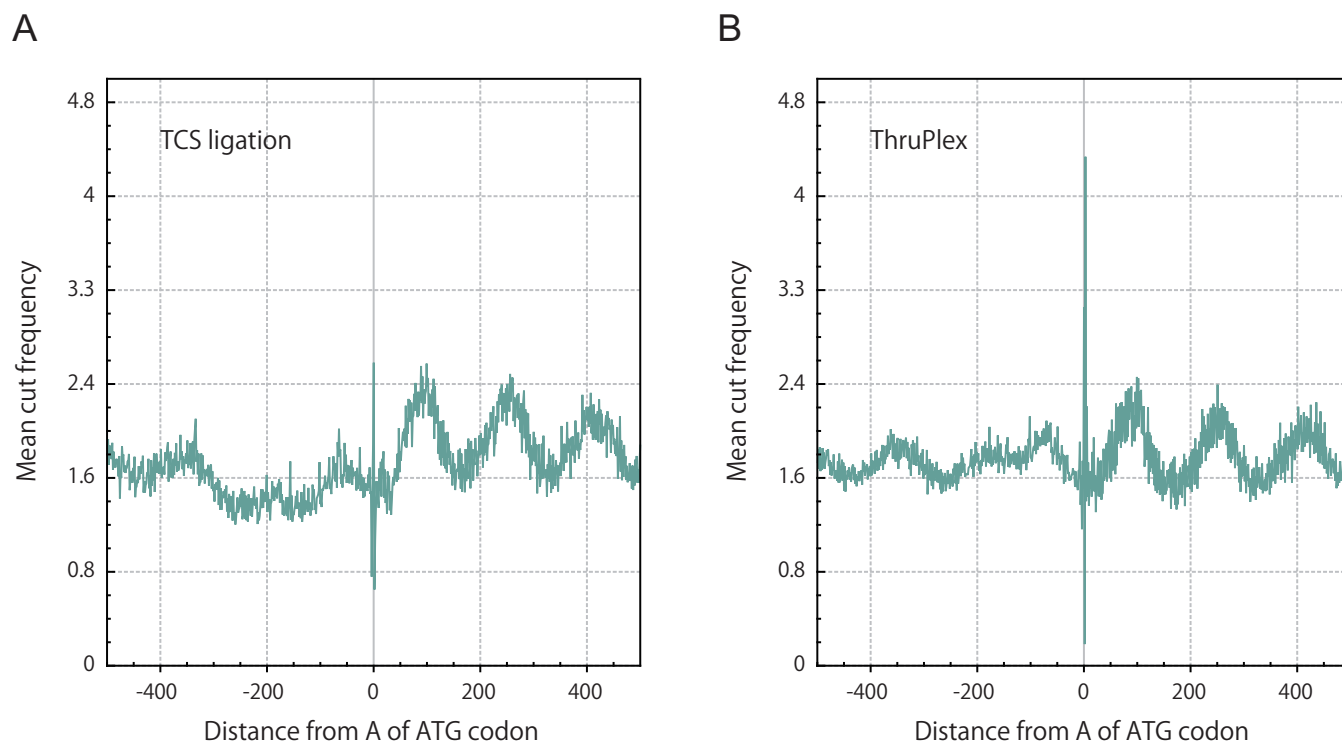
A

B

Figure 5. Comparison of yeast MNase-seq data obtained using TCS ligation (**A**) and a ThruPlex kit (**B**). Mean cut frequency is shown for gene promoters. Both graphs were drawn using 24 million mapped reads. The horizontal axis indicates the genomic coordinate relative to the A nucleotide of the ATG initiation codon. The vertical axis indicates the mean cut frequency.

DATA AVAILABILITY

The sequence data for libraries prepared from synthetic DNA was deposited to NCBI SRA with accession number SRP133409. The sequence data for MNase-seq can be found in NCBI GEO with accession number GSE110666. The programs and the budding yeast reference genome used in this study are deposited in GitHub (https://github.com/FumihitoMiura/Project-1).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Mardis,E.R. (2017) DNA sequencing technologies: 2006–2016. *Nat. Protoc.*, **12**, 213–218.

2. Adey,A., Morrison,H.G., Asan, Xun,X., Kitzman,J.O., Turner,E.H., Stackhouse,B., MacKenzie,A.P., Caruccio,N.C., Zhang,X. *et al.* (2010) Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome Biol.*, **11**, R119.
3. Edwards,J.B.D.M., Delort,J. and Mallet,J. (1991) Oligodeoxyribonucleotide ligation to single-stranded cDNAs: a new tool for cloning 5′ ends of mRNAs and for constructing cDNA libraries by in vitro amplification. *Nucleic Acids Res.*, **19**, 5227–5232.
4. 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**, 135–144.
5. Romaniuk,E., McLaughlin,L.W., Neilson,T. and Romaniuk,P.J. (1982) The effect of acceptor oligoribonucleotide sequence on the T4 RNA ligase reaction. *Eur. J. Biochem.*, **125**, 639–643.
6. Song,Y., Liu,K.J. and Wang,T.-H. (2015) Efficient synthesis of stably adenylated DNA and RNA adapters for microRNA capture using T4 RNA ligase 1. *Sci. Rep.*, **5**, 15620.
7. Peng,X., Wu,J., Brunmeir,R., Kim,S.Y., Zhang,Q., Ding,C., Han,W., Xie,W. and Xu,F. (2015) TELP, a sensitive and versatile library construction method for next-generation sequencing. *Nucleic Acids Res.*, **43**, e35.
8. Miura,F., Enomoto,Y., Dairiki,R. and Ito,T. (2012) Amplification-free whole-genome bisulfite sequencing by post-bisulfite adaptor tagging. *Nucleic Acids Res.*, **40**, e136.
9. Kolb,H.C., Finn,M.G. and Sharpless,K.B. (2001) Click chemistry: Diverse chemical function from a few good reactions. *Angew. Chem. Int. Ed. Engl.*, **40**, 2004–2021.
10. Presolski,S.I., Hong,V., Cho,S.-H. and Finn,M.G. (2010) Tailored ligand acceleration of the Cu-catalyzed azide−alkyne cycloaddition reaction: Practical and mechanistic implications. *J. Am. Chem. Soc.*, **132**, 14570–14576.
11. Presolski,S.I., Hong,V.P. and Finn,M.G. (2011) Copper-catalyzed azide-alkyne click chemistry for bioconjugation. *Curr. Protoc. Chem. Biol.*, **3**, 153–162.

12. Birts,C.N., Sanzone,A.P., El-Sagheer,A.H., Blaydes,J.P., Brown,T. and Tavassoli,A. (2014) Transcription of click-linked DNA in human cells. *Angew. Chem. Int. Ed. Engl.*, **53**, 2362–2365.

13. El-Sagheer,A.H. and Brown,T. (2012) Click nucleic acid ligation: Applications in biology and nanotechnology. *Acc. Chem. Res.*, **45**, 1258–1267.

14. El-Sagheer,A.H. and Brown,T. (2009) Synthesis and polymerase chain reaction amplification of DNA strands containing an unnatural triazole linkage. *J. Am. Chem. Soc.*, **131**, 3958–3964.

15. El-Sagheer,A.H. and Brown,T. (2014) Combined nucleobase and backbone modifications enhance DNA duplex stability and preserve biocompatibility. *Chem. Sci.*, **5**, 253–259.

16. El-Sagheer,A.H., Sanzone,A.P., Gao,R., Tavassoli,A. and Brown,T. (2011) Biocompatible artificial DNA linker that is read through by DNA polymerases and is functional in *Escherichia coli*. *Proc. Nat. Acad. Sci. U.S.A.*, **108**, 11338–11343.

17. Kumar,R., El-Sagheer,A., Tumpane,J., Lincoln,P., Wilhelmsson,L.M. and Brown,T. (2007) Template-directed oligonucleotide strand ligation, covalent intramolecular DNA circularization and catenation using click chemistry. *J. Am. Chem. Soc.*, **129**, 6859–6864.

18. Isobe,H. and Fujino,T. (2014) Triazole-linked analogues of DNA and RNA (TLDNA and TLRNA): Synthesis and functions. *Chem. Rec.*, **14**, 41–51.

19. Isobe,H., Fujino,T., Yamazaki,N., Guillot-Nieckowski,M. and Nakamura,E. (2008) Triazole-linked analogue of deoxyribonucleic acid (TLDNA): Design, synthesis, and double-strand formation with natural DNA. *Org. Lett.*, **10**, 3729–3732.

20. Varizhuk,A., Chizhov,A. and Florentiev,V. (2011) Synthesis and hybridization data of oligonucleotide analogs with triazole internucleotide linkages, potential antiviral and antitumor agents. *Bioorg. Chem.*, **39**, 127–131.

21. Fujino,T., Yasumoto,K., Yamazaki,N., Hasome,A., Sogawa,K. and Isobe,H. (2011) Triazole-linked DNA as a primer surrogate in the synthesis of first-strand cDNA. *Chem. Asian J.*, **6**, 2956–2960.

22. Shivalingam,A., Tyburn,A.E.S., El-Sagheer,A.H. and Brown,T. (2017) Molecular requirements of high-fidelity replication-competent DNA backbones for orthogonal chemical ligation. *J. Am. Chem. Soc.*, **139**, 1575–1583.

23. DeAngelis,M.M., Wang,D.G. and Hawkins,T.L. (1995) Solid-phase reversible immobilization for the isolation of PCR products. *Nucleic Acids Res.*, **23**, 4742–4743.

24. Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.

25. Winz,M.-L., Linder,E.C., André,T., Becker,J. and Jäschke,A. (2015) Nucleotidyl transferase assisted DNA labeling with different click chemistries. *Nucleic Acids Res.*, **43**, e110.

26. Drouin,R., Rodriguez,H., Gao,S.W., Gebreyes,Z., O'Connor,T.R., Holmquist,G.P. and Akman,S.A. (1996) Cupric ion/ascorbate/hydrogen peroxide-induced DNA damage: DNA-bound copper ion primarily induces base modifications. *Free Radic. Biol. Med.*, **21**, 261–273.

27. Chan,T.R., Hilgraf,R., Sharpless,K.B. and Fokin,V.V. (2004) Polytriazoles as copper(I)-stabilizing ligands in catalysis. *Org. Lett.*, **6**, 2853–2855.

28. Christen,E.H., Gubeli,R.J., Kaufmann,B., Merkel,L., Schoenmakers,R., Budisa,N., Fussenegger,M., Weber,W. and Wiltschi,B. (2012) Evaluation of bicinchoninic acid as a ligand for copper(I)-catalyzed azide-alkyne bioconjugations. *Org. Biomol. Chem.*, **10**, 6629–6632.

29. Hong,V., Presolski,S.I., Ma,C. and Finn,M.G. (2009) Analysis and optimization of copper-catalyzed azide–alkyne cycloaddition for bioconjugation. *Angew. Chem. Int. Ed. Engl.*, **48**, 9879–9883.

30. Routh,A., Head,S.R., Ordoukhanian,P. and Johnson,J.E. (2015) ClickSeq: Fragmentation-free next-generation sequencing via click ligation of adaptors to stochastically terminated 3'-azido cDNAs. *J. Mol. Biol.*, **427**, 2610–2616.

31. Gansauge,M.-T. and Meyer,M. (2013) Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat. Protoc.*, **8**, 737–748.

32. Gansauge,M.T., Gerber,T., Glocke,I., Korlevic,P., Lippik,L., Nagel,S., Riehl,L.M., Schmidt,A. and Meyer,M. (2017) Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. *Nucleic Acids Res.*, **45**, e79.