Synthesis and Enzymatic Incorporation of Allyl-Based DNA Sequencing-By-Synthesis Probes for 3'-*O*-Mass Tag Analysis

Gwangji Oh, Yeonhee Nam, Dongyun Shin,* and Dae-Ro Ahn*

*Life Sciences Research Division, Korea Institute of Science and Technology, P.O. Box 131, Cheongryang, Seoul 130-650, Korea. *E-mail: dyshin21@kist.re.kr (DS); drahn@kist.re.kr (DA)*
Received June 5, 2009, Accepted September 11, 2009

Key Words: Sequencing-by-synthesis, Mass tags, Polymerases, Primer extension, Reversible terminators

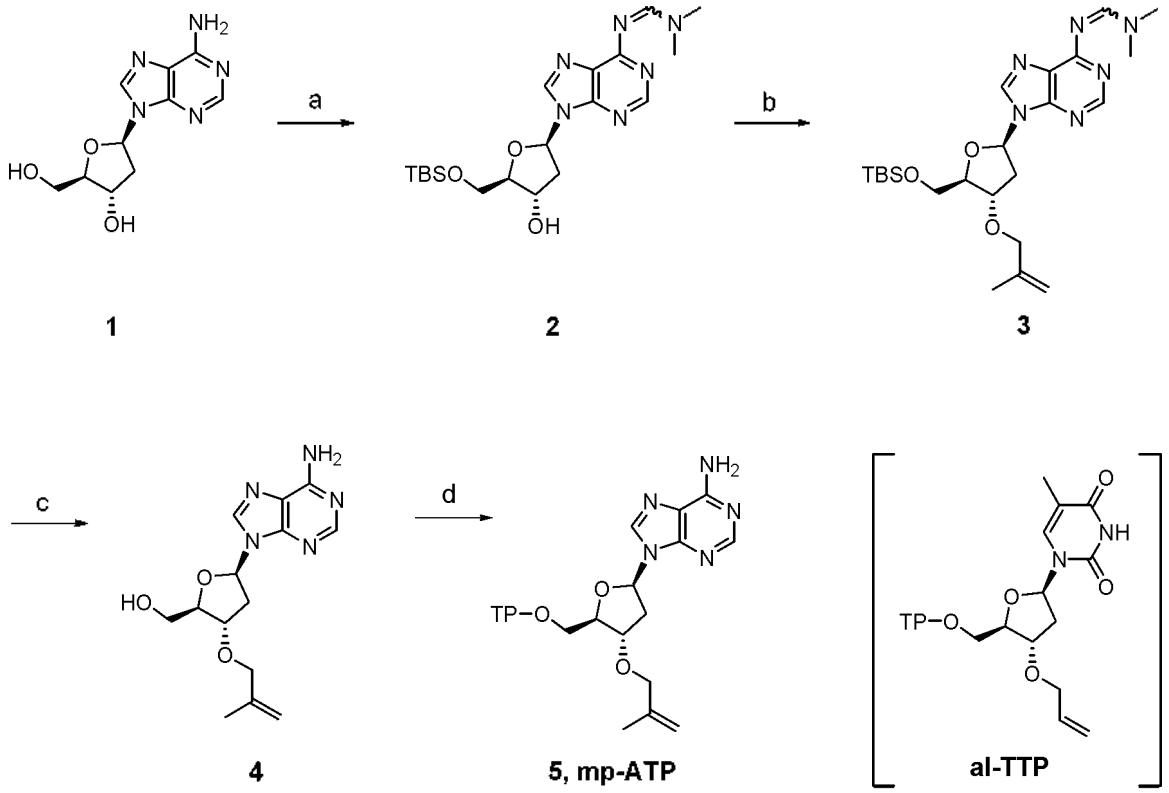
After completion of the Human Genome Project (HGP), the Personal Genome Project (PGP) to obtain individual genetic information is considered as the next mission for many genomic scientists. The conventional CE-based Sanger sequencing method, which has been used for HGP, has limitations in cost and throughputness to perform the new mission. Since 2004, the National Human Genome Research Institute (NHGRI) in USA has been awarding research funds more than $100 million to investigators for development of new DNA sequencing. The new methods should be able to sequence whole human genome (*ca.* 3 billion bases) either with a cost less than $100,000 for near-term development or with a cost less than $1,000 for the ultimate goal. In the same context, a number of research groups have been developing cost-effective 'next generation' sequencing methods including sequencing-by-synthesis (SBS),[1,2] sequencing-by-hybridization (SBH),[3-5] sequencing-by-ligation,[6] sequencing-by-degradation (SBD),[7] pyrosequencing,[8] and nanopore sequencing.[3] Among these many developing sequencing technologies, SBS method is an interesting topic for bioorganic chemists because the chemical modification of dNTP is a key issue to successfully perform SBS method.

Similar to the conventional Sanger sequencing method, SBS also takes advantage of termination of DNA polymerization by incorporation of 3'-OH-modified nucleotides. The Sanger method has the sequential signal decay due to the partial incorporation of dideoxynucleotides (ddNTPs). In contrast, the SBS method uses dual-labeled dNTPs in which the nucleobases are fluorescently labeled for detection of the incorporated base, and 3'-OH is modified with a blocking group for termination of DNA synthesis. The 3'-*O*-blocking group is removable to reversibly generate 3'-OH for the next polymerization reaction without partial loss of sequences. Typically, a SBS method works on immobilized primer sequences and consists of three basic steps: 1) incorporation of the reversible terminator dNTPs, 2) detection of fluorophores labeled on the nucleobases of the incorporated monomers, and 3) removal of reversible blocking groups of 3'-OH and fluorophores on bases to regenerate free 3'-OH group for the next incorporation cycle.

In the past decade, several reversible blocking groups for 3'-OH such as 3'-ester,[9] 3'-*O*-allyl,[10] 3'-*O*-hydrocarbyldithiomethyl,[11] 3'-*O*-dithioethylcarbonate,[12] and 3'-*O*-azidomethyl[13] have been developed, and some of them are currently utilized for SBS technology. The fluorophores for analysis of the incorporated dNTPs are usually labeled on bases *via* a propargyl amine-based linker. The linker part would still remain on each base after the cleavage step of fluorophores and blocking groups. The residual linkers would be accumulated, influence the efficiency of DNA synthesis, and eventually inhibit the polymerase after several cycles of SBS. Instead of the dual-labeled nucleotides, if SBS can utilize 3'-*O*-mono-labeled nucleotides in which the 3'-*O*-blocking group plays a role as a signal reporter as well as a reversible terminator, labeling on bases with fluorescent molecules would no longer be necessary. However, due to the bulky size of usual fluorescent molecules, previous attempts to incorporate a dNTP containing a fluorophore on the 3'-position by DNA polymerase were unsuccessful.[13,14] Development of a reporting system on 3'-hydroxyl group which uses signals other than emission from fluorophores for sequencing and acts as a 3'-*O*-blocking group would enable us to perform SBS without any label on bases. For example, small mass tags can work as both a signal reporter and a 3'-*O*-blocking group.

In this study, we report the synthesis of nucleotides with small allyl-based reversible 3'-*O*-blocking groups, which would work not only as a terminator of the polymerization but also as a mass tag to sequence the base by means of mass spectrometry, and our results present an possible extension of the bulkiness of allyl-based 3'-OH blocking group, which is one of the major issues in SBS.

As the allyl group was previously used as an efficient reversible blocking group for 3'-OH of dNTP,[10] we have designed nucleotides containing an allyl-based functional group at 3'-OH that can be used as a reversible blocking group and a mass tag in SBS. For a proof of the concept, we decided to prepare a purine (dTTP) and a pyrimidine (dATP) nucleotide in which 3'-OH was blocked by allyl and 2-methylpropenyl group, respectively. The 2-methylpropenyl group was expected to be removed under the same condition previously used for deallylation.[15] 3'-*O*-Allylated dTTP (**al-dTTP**) has been synthesized according to the literature.[1] The synthesis of 3'-*O*-(2-methylpropenyl)-dATP (**mp-dATP**) started with 2'-deoxyadenosine (**1**) (Scheme 1). 5'-OH and 4-NH_2 were protected by *tert*-butyldimethylsilyl chloride (TBS-Cl) and *N,N*-dimethylforamide dimethyl acetal, respectively, to yield protected alcohol **2**. 3'-Hydroxy group of **2** was subsequently blocked by 2-methylpropenyl bromide to produce alkylated ribose **3**.



Scheme 1. Synthesis of **mp-dATP**. Reaction conditions: (a) imidazole, tert-butyldimethylsilyl chloride, DMF, 75% (borsm), (b) 3-bromo-2-methylpropene, dichloromethane, 10% NaOH, 98%, (c) i) tetrabutylammonium fluoride, THF 91% (borsm); ii) 2 M NH_3 in EtOH, (d) i) 2-chloro-4*H*-1,2,3-benzodioxaphosphorin-4-one, dioxane, pyridine; ii) $(Bu_3NH)_4P_2O_7$, DMF; iii) I_2, pyridine, H_2O.

Table 1. Oligonucleotide sequences.

Oligo	Sequence
Primer	Biotin-TTT(fluorescein)GCGTAATACGACTCACTATGGACG-5'[a]
Template-1	5'-CGCATTATGCTGAGTGATACCTGCTTTTT
Template-2	5'-CGCATTATGCTGAGTGATACCTGCAA

[a]Underlined bases denote regions of complementarity between the primer and templates.

Deprotection of the TBS group and dimethylformamide group yielded the precursor **4** for triphosphorylation on 5'-OH. Finally, the compound **4** was transformed into the corresponding triphosphate by Ludwig-Eckstein phosphorylation.[16] followed by a series of HPLC purification (reverse-phase/anion exchange/reverse-phase chromatography) to yield **mp-dATP**.

While incorporation of **al-dTTP** by polymerases was demonstrated in the literature.[17] **mp-dATP** has not been tested as a substrate for DNA polymerization. Thus, first of all, we screened various commercially available polymerases (T7 polymerase, *Therminator I*, *Therminator II*, *Taq*, 9°N_m, *Bst*, *Vent* (exo$^-$), and *Deep Vent* (exo$^-$)) using appropriate primers and templates as presented in Table 1 for incorporation of **mp-dATP** as well as **al-dTTP**. After polyacrylamide gel electrophoresis (PAGE) analysis, upon using **mp-dATP** for primer extension we could clearly observe its incorporation only by *Therminator II* (Figure 1a). Successful incorporation of **al-dTTP** by *Therminator II* and termination of the polymerization were also observed as expected (Figure 1b). The single nucleotide extended primers were obtained almost quantitatively using the templates containing a stretch of the complement bases, suggesting that the DNA synthesis was terminated after incorporation of our 3'-*O*-modified nucleotides. These results clearly demonstrated that not only the allyl moiety but also the 2-methylpropenyl moiety could be used for terminator of DNA polymerization. Because of the blocking group on 3'-OH, terminated DNA strands created by **mp-dATP** (lane 4 in Figure 1a) and **al-dTTP** (lane 4 in Figure 1b) migrated slightly slower than the control strand for single nucleotide extension with ddATP (lane 2 in Figure 1a) and ddTTP (lane 2 in Figure 1b), respectively. The other polymerases tested were unable to accept **mp-dATP** and **al-dTTP** as substrates. Unexpectedly, when the 9°N_m polymerase was used for the primer extension experiment, degradation of primers was observed. This might be due to exonuclease activity of the 9°N_m, which is possibly less suppressed compared with that of the other polymerases.

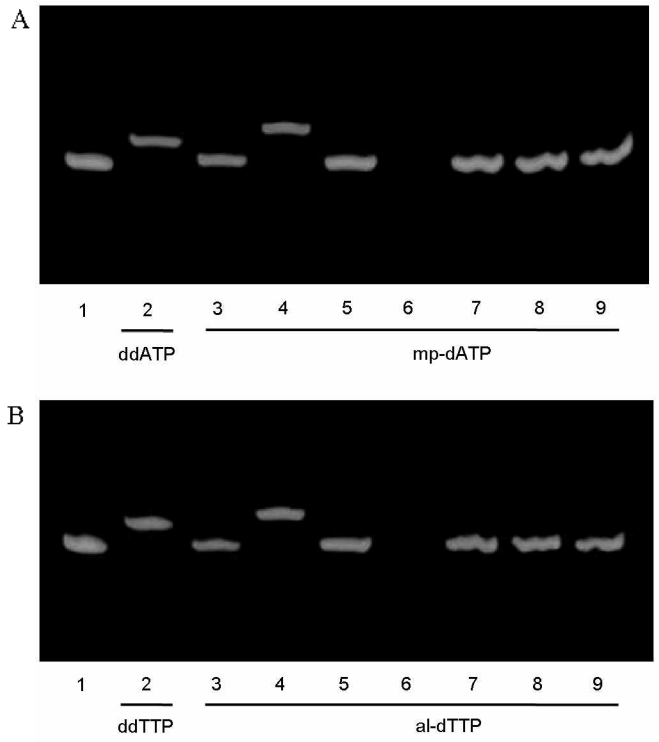


Figure 1. Primer extension assay for (A) **mp-dATP** and (B) **al-dTTP** using various DNA polymerases, analyzed on a fluorescent imager. Lane 1 - no enzyme, lane 2 - T7 polymerase, lane 3 - *Therminator I*, lane 4 - *Therminator II*, lane 5 - *Taq*, lane 6 - 9°N_m, lane 7 - *Bst*, lane 8 - *Vent* (exo$^-$), lane 9 - *Deep Vent* (exo$^-$).

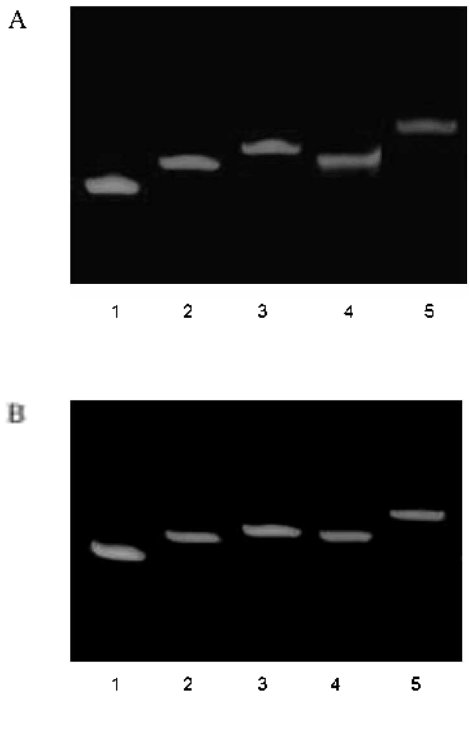


Figure 2. Sequencial two-base extension assay with (A) **mp-dATP** and (B) **al-dTTP** using homo-base stretched templates and the *Therminator II* DNA polymerase, analyzed on a fluorescent imager. Lane 1 - primer, lane 2 - extension with ddATP by T7 DNA polymerase, lane 3 - first extension, lane 4 - de-blocked, lane 5 - second extension.

To examine subsequent DNA synthesis after incorporation of the 3'-*O*-modified nucleotides and cleavage of the blocking groups, we attached the biotinylated primer containing a poly(dA) or poly(dT) stretch to streptavidin-coated magnetic beads. After obtaining the first extension product with **mp-dATP** (lane 3 in Figure 2a) or **al-dTTP** (lane 3 in Figure 2b), the primer bound on the magnetic beads was washed. According to the literature,[10] the cleavage reaction was carried out by incubation of the first extension product immobilized on magnetic beads with the Pd catalyst at 70 °C for 10 min, and analyzed on PAGE. The gel data illustrate that the product band from the cleavage reaction (lane 4 in Figure 2a, b) migrated at the same rate as the dideoxynucleotide control (lane 2 in Figure 2a, b), demonstrating successful de-blocking at 3'-OH. Additional extension on the de-blocked oligonucleotide with **mp-dATP** (lane 5 in Figure 2a) or **al-dTTP** (lane 5 in Figure 2b) afforded subsequent elongated products. This result suggests that different blocking groups such as methylpropenyl and allyl moiety can be used as reversible blocking group for SBS, and simultaneously removed under the same condition, which is important for the overall efficiency of SBS. Moreover, after removing the blocking group on 3'-OH the modified nucleotide is transformed back into its natural state, which should not disturb the subsequent incorporation of nucleotides, unlikely the case with the dual-labeled nucleotides.

In summary, this paper presents the first result, to our knowledge, that describes sequencing-by-synthesis using reversible 3'-*O*-blocking groups as mass tags for DNA sequencing. Our mono-labeled nucleotide-based SBS method can potentially contribute to address the read length problems currently existing in the SBS technology. Detailed mass analysis results will be also reported in near future

Spectral data of compound 5: ^{1}H-NMR (300 MHz, D_2O) δ 8.37 (s, 1H, 2-H), 8.09 (s, 1H, 8-H), 6.33 (brm, 1'-H), 4.93 (s, 1H, one of $CH_2(CH_3)C=CH_2$), 4.86 (s, 1H, one of $CH_2(CH_3)C=CH_2$), 4.42 (m, 1H, 3'-H), 4.30 (m, 1H, 4'-H), 4.02 (m, 2H, 5'-H), 3.95 (s, 2H, $CH_2(CH_3)C=CH_2$), 2.60-2.72 (m, 2H, 2'-H), 1.63 (s, 3H, $CH_2(CH_3)C=CH_2$); ^{31}P NMR (242.6 MHz, D_2O) δ −8.8 (d, J = 21.1 Hz, 1P, γ-P), −13.8 (d, J = 18.9 Hz, 1P, α-P), −24.9 (t, J = 20.6 Hz, 1P, β-P). MALDI-TOF MS (negative mode): calcd for $C_{14}H_{21}N_5O_{12}P_3$ (M-H): 544, found: 544.

Acknowledgments. This research has been supported by the Functional Proteomics Research Center of the 21st Century Frontier Research Program funded by the Korean Ministry of Science and Technology and by the KIST grants.

References

1. Ju, J.; Kim, D. H.; Bi, L.; Meng, Q.; Bai, X.; Li, X.; Li, X.; Marma, M. S.; Shi, S.; Wu, J.; Edwards, J.; Romu, A.; Turro, N. J. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 19635-19640.
2. Guo, J.; Xu, N.; Li, Z.; Zhang, S.; Wu, J.; Kim, D. H.; Marma, M. S.; Meng, Q.; Cao, H.; Li, X.; Shi, S.; Yu, L.; Kalachikov, S.; Russo, J.; Turro, N. J.; Ju, J. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 9145-9150.

3. Drmanac, S. *Nature Biotechnol.* **1998**, *16*, 54-58.
4. Lipshutz, R.; Morris, D.; Chee, M.; Hubbell, E.; Kozal, M.; Shah, N.; Shen, N.; Yang, R.; Fodor, S. *Biotechniques* **1995**, *19*, 442-447.
5. Patil, N.; Berno, A.; Hinds, D.; Barrett, W.; Doshi, J.; Hacker, C.; Kautzer, C.; Lee, D.; Marjoribanks, C.; McDonough, D.; Nguyen, B.; Norris, M.; Sheehan J.; Shen, N.; Stern, D.; Stokowski, R.; Thomas, D.; Trulson, M.; Vyas, R.; Frazer, K.; Fodor, S.; Cox, D. *Science* **2001**, *294*, 1719-1723.
6. Shendure, J.; Porreca, G.; Reppas, N.; Lin, X.; McCutcheon, J.; Rosenbaum, A.; Wang, M.; Zhang, K.; Mitra, R.; Church, G. *Science* **2005**, *309*, 1728-1732
7. Stephan, J.; Dörre, K.; Brakmann, S.; Winkler, T.; Wetzel, T.; Lapczyna, M.; Stuke, M.; Angerer, B.; Ankenbauer, W.; Földes-Papp, Z.; Rigler, R.; Eigen, M. *J. Biotechnol.* **2001**, *86*, 255-267.
8. Ronaghi, M. *Genome Res.* **2001**, *11*, 3-11.
9. Deamer, D. W.; Akeson, M. *Trends Biotechnol.* **2000**, *18*, 147-151.
10. Canard, B.; Cardona, B.; Sarfati, R. S. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 10859-10863.
11. Ruparel, H.; Bi, L.; Li, Z.; Bai, X.; Kim, D. H.; Turro, N. J.; Ju, J. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 5932-5937.
12. Kwiatkowski, M. **2007**, *US20077279563*
13. Turcatti, G.; Romieu, A.; Fedurco, M.; Tairi, A.-P. *Nucleic Acids Res.* **2008**, *36*, e25.
14. Canard, B.; Sarfati, R. S. *Gene* **1994**, *148*, 1-6.
15. Meng, Q.; Kim, D. H.; Bai, X.; Bi, L.; Turro, N. J.; Ju, J. *J. Org. Chem.* **2006**, *71*, 3248-3251.
16. Ludwig, J.; Eckstein, F. *J. Org. Chem.* **1989**, *54*, 631-645.
17. Wu, J.; Zhang, S.; Meng, Q.; Cao, H.; Li, Z.; Li, X.; Shi, S. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 19462-16467.