3′-O-modified nucleotides as reversible terminators for pyrosequencing

Jian Wu**,†, Shenglong Zhang**,†, Qinglin Meng**,†, Huanyan Cao**,†, Zengmin Li**,†, Xiaoxu Li**,†, Shundi Shi*, Dae Hyun Kim**,§, Lanrong Bi**,†, Nicholas J. Turro†‡¶ and Jingyue Ju**†§

* Columbia Genome Center, Columbia University College of Physicians and Surgeons, New York, NY 10032; and Departments of †Chemical Engineering, §Biomedical Engineering, Columbia University, New York, NY 10027

Contributed by Nicholas J. Turro, August 10, 2007 (sent for review July 3, 2007)

Pyrosequencing is a method used to sequence DNA by detecting the pyrophosphate (PPi) group that is generated when a nucleotide is incorporated into the growing DNA strand in polymerase reaction. However, this method has an inherent difficulty in accurately deciphering the homopolymeric regions of the DNA templates. We report here the development of a method to solve this problem by using nucleotide reversible terminators. These nucleotide analogues are modified with a reversible chemical moiety capping the 3′-OH group to temporarily terminate the polymerase reaction. In this way, only one nucleotide is incorporated into the growing DNA strand even in homopolymeric regions. After detection of the PPI for sequence determination, the 3′-OH of the primer extension products is regenerated through different deprotection methods. Using an allyl or a 2-nitrobenzyl group as the reversible moiety to cap the 3′-OH of the four nucleotides, we have synthesized two sets of 3′-O-modified nucleotides, 3′-O-allyl-dNTPs and 3′-O-(2-nitrobenzyl)-dNTPs as reversible terminators for pyrosequencing. The capping moiety on the 3′-OH of the DNA extension product is efficiently removed after PPI detection by either a chemical method or photolysis. To sequence DNA, templates containing homopolymeric regions are immobilized on Sepharose beads, and then extension–signal detection–deprotection cycles are conducted by using the nucleotide reversible terminators on the DNA beads to unambiguously decipher the sequence of DNA templates. Our results establish that this reversible-terminator-pyrosequencing approach can be potentially developed into a powerful methodology to accurately determine DNA sequences.


The authors declare no conflict of interest.

Frequently available online through the PNAS open access option.

Abbreviations: NRT, nucleotide reversible terminator; PPi, pyrophosphate.

To whom correspondence may be addressed: Room 405A, Russ Berrie Medical Science Pavilion, Columbia Genome Center, Columbia University College of Physicians and Surgeons, New York, NY 10032. E-mail: jhw2@cb2.columbia.edu or njt3@columbia.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0707495104DC1.

© 2007 by The National Academy of Sciences of the USA
NH₄OH at room temperature for 1.5 h to produce compound 2 in Fig. 2 as an example. Treatment of 9-\(\text{O}-(2\text{-nitrobenzyl})\)-6-chloropurine 2, which was deprotected and converted to 2-deoxyadenosine derivative 3-\(\text{O}-(2\text{-nitrobenzyl})\)-2-\text{deoxyadenosine} 3 in a one-pot reaction. The precursor 3 was then transformed to the target molecule 3-\(\text{O}-(2\text{-nitrobenzyl})\)-dATP 4 with established triphosphorylation procedures (23, 29, 30).

**Polymerase Extension Using 3′-O-modified Nucleotides and Characterization by MALDI-TOF MS.** 3′-O-modified nucleotides pose a great challenge for incorporation by natural polymerase, especially when the 3′-O-labeling group is a bulky one (31, 32). To verify that the NRTs can be recognized by polymerase as substrates in a polymerase reaction, we performed extension reactions with four different primers corresponding to different regions of a DNA template whose next complementary base was either A, C, G, or T. A 9°N polymerase (exo–)A485L/Y409V, which has been shown previously to incorporate the 3′-O-modified nucleotides (21, 23), was used in the polymerase extension reaction. After the reaction, the eight different primer extension products [four for 3′-O-allyl-dNTPs and four for 3′-O-(2-nitrobenzyl)-dNTPs] were analyzed by MALDI-TOF MS, and the results are shown in Fig. 3. Single clear mass peaks at 6,437, 7,702, 6,500, and 8,310 (\(m/z\)) for each primer extension product was produced by using 3′-O-allyl-dNTPs with no left-over primer peak (Fig. 3 A–D). Similarly, the 3′-O-(2-nitrobenzyl)-dNTPs also produced complete primer extension DNA products at 8,414, 8,390, 8,430, and 5,602 (\(m/z\)) (Fig. 3 E–H). The small peaks at 8,279, 8,255, 8,295, and 5,467 (\(m/z\)) in the mass spectra for the 3′-O-(2-nitrobenzyl)-dNTP extension products correspond to the photolysis products that were generated by the partial photolysis of the DNA extension products induced by the nitrogen laser (337 nm) used for ionization of the analyte in MALDI-TOF MS. These results indicate that the primers were quantitatively extended by the 3′-O-modified-dNTPs in polymerase reaction and that the modified nucleotides are excellent substrates for the 9°N polymerase.

To further verify the utility of the NRTs in determining the homopolymeric regions of DNA sequences, we performed a continuous polymerase extension reaction in solution. This procedure allows the isolation of the DNA product at each step for detailed molecular characterization by MALDI-TOF MS. First, a polymerase extension reaction using 3′-O-(2-nitrobenzyl)-dTTP as a reversible terminator along with a primer and synthetic 100-mer DNA template corresponding to a portion of exon 7 in the human p53 gene was performed to yield a single-base extension product (product 2) (Fig. 4B Left). After the reaction, a small portion of the extension product was characterized by MALDI-TOF MS. The rest of the product was site-specific introduction of the 2-nitrobenzyl group to the 3′-oxygen, with 2-nitrobenzyl bromide under basic conditions furnished 2-nitrobenzylationated compound 9-\(\beta\)-d-5′-\(\text{O}-(\text{tert-butylidemethylsilyl})\)-3′-\(\text{O}-(2\text{-nitrobenzyl})\)-2′-deoxyribofuranosyl]-6-chloropurine 2, which was deprotected and converted to 2-deoxyadenosine derivative 3-\(\text{O}-(2\text{-nitrobenzyl})\)-2-\text{deoxyadenosine} 3 in a one-pot reaction. The precursor 3 was then transformed to the target molecule 3-\(\text{O}-(2\text{-nitrobenzyl})\)-dATP 4 with established triphosphorylation procedures (23, 29, 30).

**Fig. 1.** Structures of NRTs 3′-O-allyl-dNTP and 3′-O-(2-nitrobenzyl)-dNTP.

It is particularly challenging to synthesize 3′-O-(2-nitrobenzyl)-dNTPs because the nucleophilic nitrogen on the base preferentially reacts with the 2-nitrobenzyl group. Using a previously reported method (27) for the synthesis of a 3′-O-(2-nitrobenzyl)-dATP actually led to the final nucleotide analogue with the 3′-O-(2-nitrobenzyl)-dATP. After the removal of the capping moiety, the polymerase reaction will resume. Based on this rationale, we synthesized and evaluated two sets of nucleotide analogues as NRTs for pyrosequencing: 3′-O-allyl-dNTPs and 3′-O-(2-nitrobenzyl)-dNTPs (Fig. 1). The allyl group can be efficiently removed by Pd-catalyzed deallylation, and the removal of the 2-nitrobenzyl moiety is readily accomplished by laser irradiation at 355 nm. The design and synthesis of the 3′-O-allyl-dNTPs has been described previously (23).

First, a polymerase extension reaction using 3′-O-allyl-dNTPs and 3′-O-(2-nitrobenzyl)-dNTPs has been shown previously to incorporate the 3′-O-modified nucleotides (21, 23), which both the sugar and base were modified to allow the continuous polymerase extension reaction in solution. This procedure allows the isolation of the DNA product at each step for detailed molecular characterization by MALDI-TOF MS. The small peaks at 8,279, 8,255, 8,295, and 5,467 (\(m/z\)) in the mass spectra for the 3′-O-(2-nitrobenzyl)-dNTP extension products correspond to the photolysis products that were generated by the partial photolysis of the DNA extension products induced by the nitrogen laser (337 nm) used for ionization of the analyte in MALDI-TOF MS. These results indicate that the primers were quantitatively extended by the 3′-O-modified-dNTPs in polymerase reaction and that the modified nucleotides are excellent substrates for the 9°N polymerase.

To further verify the utility of the NRTs in determining the homopolymeric regions of DNA sequences, we performed a continuous polymerase extension reaction in solution. This procedure allows the isolation of the DNA product at each step for detailed molecular characterization by MALDI-TOF MS. First, a polymerase extension reaction using 3′-O-(2-nitrobenzyl)-dTTP as a reversible terminator along with a primer and synthetic 100-mer DNA template corresponding to a portion of exon 7 in the human p53 gene was performed to yield a single-base extension product (product 2) (Fig. 4B Left). After the reaction, a small portion of the extension product was characterized by MALDI-TOF MS. The rest of the product was

**Fig. 2.** Synthesis of 3′-O-(2-nitrobenzyl)-dATP. (Step a) 2-nitrobenzyl bromide, tetrabutylammonium bromide, NaOH, in CH₃Cl at room temperature for 1 h to produce compound 2 with a 95% yield. (Step b) Tetrabutylammonium fluoride in THF at room temperature for 1 h; methanolic ammonia and dioxane at 85–90°C for 12 h to produce compound 3 with a 56% yield. (Step c) POCl₃, PO(OMe)₃ at 0°C for 2 h; (Bu₃NH)₄P₂O₇, Bu₃N, triethylammonium bicarbonate, and NH₃·OH at room temperature for 1.5 h to produce compound 4 with a 30% yield.
irradiated with a laser at 355 nm for 30 s to cleave the 3′-O-(2-nitrobenzyl) group from the DNA to yield photocleaved product (product 3) (Fig. 4C Left), which was characterized by MALDI-TOF MS. The photocleaved DNA product (product 3) with a free 3′-OH group regenerated was then used as a primer for the next nucleotide extension reaction. Fig. 4A Right–E Right shows the sequential mass spectrum at each step of continuous DNA extension reaction using 3′-O-(2-nitrobenzyl)-dGTP as a reversible terminator. The primer alone produces a peak at 6,131 \((m/z)\) (Fig. 4A). The mass peak at 6,594 \((m/z)\) in Fig. 4B corresponds to the first extension product with a single modified nucleotide G incorporated in this homopolymeric region. The small peak at 6,459 \((m/z)\) in Fig. 4B corresponds to the photocleavage product that was generated by the nitrogen laser (337 nm) used for ionization of the analyte in MALDI-TOF MS. Fig. 4C shows the photocleavage result after irradiation of the extension product (product 2) at 355 nm. It can be seen from the data that the peak at 6,594 \((m/z)\) has completely vanished, and only a single peak corresponding to the DNA product (product 3) remains at 6,459 \((m/z)\), which indicates that the 2-nitrobenzyl moiety was efficiently removed to regenerate the 3′-OH group. Fig. 4D shows the MALDI-TOF MS data for the extension product obtained by using the photocleaved DNA product (compound 3) as a primer to incorporate another 3′-O-(2-nitrobenzyl)-dGTP. A dominant peak is seen at 6,922 \((m/z)\) corresponding to the extension product (product 4). The small peak at 6,787 \((m/z)\) corresponds to the photocleavage product that was generated by the nitrogen laser (337 nm) used for ionization of the analyte in MALDI-TOF MS. Upon further photolysis at 355 nm, the 2-nitrobenzyl moiety was removed to yield DNA product (product 5) at 6,787 \((m/z)\) with a free 3′-OH group (Fig. 4E). Similar data were obtained for 3′-O-(2-nitrobenzyl)-dTTP (SI Fig. 8). The other two nucleotides, 3′-O-(2-nitrobenzyl)-dATP, and 3′-O-(2-nitrobenzyl)-dCTP also were verified to be excellent reversible terminators for the 9°N polymerase.

**3′-O-Modified dATP Is Not a Substrate of Luciferase.** In pyrosequencing, luciferase converts luciferin to oxyluciferin by using the energy provided by ATP, yielding a chemiluminescence light signal. However, the natural nucleotide dATP also is a substrate for luciferase, which can produce a false positive signal to seriously interfere with the pyrosequencing result. To solve this problem, a sulfur-modified nucleotide, 3′-O-(2-nitrobenzyl)-dATP, was shown not to be a substrate of luciferase as indicated by the data in Fig. 5. 3′-O-modified-dATP and dATP were separately added to the luciferase and luciferin mixtures and the corresponding light intensities were measured and compared. dATP (0.5 nmol) produced a light signal intensity of 80, whereas 0.5 nmol and 1.5 nmol of 3′-O-modified-dATP only led to
light intensities near background level. These results confirmed that 3'-O-modified-ddATP is not a substrate of luciferase. Fig. 5 also shows that ddATP is not a substrate to luciferase. These results indicate that the 3'-OH group may play a significant role in luciferase-catalyzed reaction. This hydroxyl group may interact with the active catalytic site of the luciferase. This interaction is interrupted when the 3'-OH group is modified with an allyl group or a 2-nitrobenzyl group (or without a 3'-hydroxyl group as in ddATP), thereby preventing luciferase from using 3'-O-modified-ddATP and ddATP as a substrate. Thus, 3'-O-modified-ddATP can be directly used in pyrosequencing without any further modification.

**Pyrosequencing with 3'-O-Modified NRTs.** To verify that the NRTs can be successfully used in pyrosequencing, we carried out a sequencing reaction on a self-priming DNA template, which contained multiple homopolymeric regions, immobilized on Sepharose beads (SI Fig. 9). The pyrosequencing reaction was initiated by extending the DNA template using a polymerase extension reaction mixture containing the NRTs. The extension of the primer by only the complementary NRT was confirmed by subsequent enzymatic cascade reactions to convert the released PPi into a light signal. For the 3'-O-allyl-dNTPs, after detection of the light signal, the DNA beads were immersed in a Pd deallylation solution and incubated for 2 min to cleave the 3'-O-allyl group to regenerate a free 3'-OH for further extension. In the case of 3'-O-(2-nitrobenzyl)-dNTPs, after NRT incorporation, the DNA beads were irradiated with a laser at 355 nm to remove the 2-nitrobenzyl group for further extension. After washing the beads, the next extension cycle was initiated. Extension–signal detection–deprotection cycles were performed multiple times to decipher unambiguously the homopolymeric sequences in the DNA template.

The pyrosequencing data generated by 3'-O-allyl-dNTPs are shown in Fig. 6A. The 11 bases in the homopolymeric regions (five T, two A, two C, and two G bases) were sequenced by using 3'-O-allyl-dNTPs. The homopolymeric regions are clearly identified, with each peak corresponding to the identity of each base in the DNA template. (b) Pyrosequencing data using natural nucleotides. The homopolymeric regions produced one large peak corresponding to the stretch of T bases and three smaller peaks for stretches of A, C, and G bases. However, it is very difficult to decipher the exact sequence from the data.

**Fig. 4.** The polymerase extension scheme using 3'-O-(2-nitrobenzyl)-dGTP (A Left–E Left) and MALDI-TOF MS spectra of the two consecutive extension products and their photocleavage products (A Right–E Right). (A) Primer for the polymerase extension reaction. (B) Primer extended with 3'-O-(2-nitrobenzyl)-dGTP to yield DNA extension product 2. (C) Product 2 photocleaved to yield photocleavage product 3. (D) Product 3 extended with another 3'-O-(2-nitrobenzyl)-dGTP to yield product 4. (E) Product 4 photocleaved to yield photocleavage product 5. After 30 s of irradiation with a laser at 355 nm, photocleavage is complete with all of the 3'-O-(2-nitrobenzyl)-group cleaved from the DNA extension products.

**Fig. 5.** Signal intensity of luciferase catalyzed reactions using 0.5 nmol of dATP, 0.5 nmol of 3'-O-(2-nitrobenzyl)-dATP, 1.0 nmol of 3'-O-(2-nitrobenzyl)-dATP, 0.5 nmol of 3'-O-allyl-dATP, 1.5 nmol of 3'-O-allyl-dATP, and 1.5 nmol of ddATP. The results show that 3'-O-(2-nitrobenzyl)-dATP and 3'-O-allyl-dATP are not substrates of luciferase (NB, 2-nitrobenzyl).
in conventional pyrosequencing, the other advantage of using the NRTs is that higher efficiency can be achieved with multiple extensions or deprotections without any dephasing in the sequence determination or a reduction in the sequencing accuracy. Therefore, one can achieve >99% efficiency in each cycle to reach read lengths of at least several hundred. The signal reduction in our preliminary pyrosequencing data generated with the NRTs is mainly due to the loss of DNA beads during each washing step because the reaction was performed manually. Therefore, longer read lengths can be achieved when using single DNA-bead extension and automated washing systems, such as the 454 genome sequencer (18). It is well established that PCR templates can be generated on millions of beads through emulsion PCR (18, 34). Thus, future implementation of the reversible-terminator pyrosequencing on a high-density bead array platform will provide a high-throughput and accurate DNA sequencing system with wide applications in genome biology and biomedical research.

Materials and Methods

Synthesis of 3′-O-allyl-dNTPs and 3′-O-(2-nitrobenzyl)-dNTPs. 3′-O-allyl-dNTPs were synthesized according to the literature (23), and the synthesis of 3′-O-(2-nitrobenzyl)-dNTP is described in SI Appendix. An enzymatic method was used to yield ultrapure 3′-O-modified nucleotide analogues based on the literature (35) (also see SI Appendix).

Incorporation of 3′-O-Modified NRTs in Solution and Characterization by MALDI-TOF MS. Each polymerase reaction solution consists of 40 pmol of templates, 40 pmol of primers (the template and primer sequences are described in SI Table 1), 100 pmol of NRTs, 2 μl of 10× Thermopol II reaction buffer (New England Biolabs, Ipswich, MA), 2 μl of 20 mM MnCl₂, and 2 μl (4 units) of 9°N polymerase (exo-)A485LY409V in a total volume of 20 μl. After an initial incubation at 95°C for 5 min and 4°C for 5 min, the reaction was performed at 95°C for 15 seconds, 55°C for 15 seconds, and 65°C for 1 min for 20 cycles. The resulting DNA products were purified for MALDI-TOF MS analysis by using a previously reported procedure (23). We also characterized 3′-O-(2-nitrobenzyl)-dGTP by performing a continuous DNA extension reaction using a primer (5′-GTTGATGTACACATTGTCAA-3′) and a synthetic DNA template (SI Table 1). The detailed procedure is described in SI Appendix. The other 3′-O-(2-nitrobenzyl)-dNTPs were similarly characterized.

Pyrosequencing Using the NRTs. Each extension reaction consisted of Sepharose bead-immobilized DNA (the procedure to prepare the DNA beads is described in SI Appendix). 200 pmol of NRTs, 1.2 μl of 50 mM MnCl₂, 1 μl (2 units) of 9°N polymerase (exo-)A485LY409V, and 20 μl of annealing buffer (20 mM Tris-acetate/5 mM magnesium acetate, pH 7.6). Extension was conducted in a thermal cycler and incubated at 65°C for 20 min with occasional stirring to prevent the beads from settling. After the polymerase reaction, the beads were washed with 30 μl of deionized water. For 3′-O-allyl-dNTP extensions, deallylation was conducted under aqueous-Pd-catalyzed conditions (23). After deallylation, the beads were washed three times with 180 μl of deionized water. For 3′-O-(2-nitrobenzyl)-dNTP extensions, extended DNA beads were suspended in 1 ml of annealing buffer in a cuvette with stirring and irradiated with a laser at 355 nm (3 W/cm²) for 1 min. After photobleaching, the beads were washed two times with annealing...
buffer for the continuation of the subsequent extension reactions. Conventional pyrosequencing data shown in Figs. 6B and 7B were generated by using the same instrument in parallel to compare the data with those of pyrosequencing by using the NRTs.

We thank Dr. Steffen Jockusch, Dr. James J. Russo, and Mr. Liyong Deng for generous discussions and technical support. This work was supported by National Institutes of Health Grants P50 HG002806, R01 HG003582, and R21HG004404 and by the Packard Fellowship for Science and Engineering.