

# A photocleavable fluorescent nucleotide for DNA sequencing and analysis

Zengmin Li<sup>†</sup>, Xiaopeng Bai<sup>\*†‡</sup>, Hameer Ruparel<sup>\*†</sup>, Sobin Kim<sup>\*†</sup>, Nicholas J. Turro<sup>†‡</sup>, and Jingyue Ju<sup>\*†‡</sup>

<sup>\*</sup>Columbia Genome Center, Columbia University College of Physicians and Surgeons, New York, NY 10032; and Departments of <sup>†</sup>Chemical Engineering and <sup>‡</sup>Chemistry, Columbia University, New York, NY 10027

Contributed by Nicholas J. Turro, November 28, 2002

**DNA sequencing by synthesis during a polymerase reaction using laser-induced fluorescence detection is an approach that has a great potential to increase the throughput and data quality of DNA sequencing. We report the design and synthesis of a photocleavable fluorescent nucleoside triphosphate, one of the essential molecules required for the sequencing-by-synthesis approach. We synthesized this nucleoside triphosphate by attaching a fluorophore, 4,4-difluoro-5,7-dimethyl-4-bora-3 $\alpha$ ,4 $\alpha$ -diazas-indacene propionic acid (BODIPY), to the 5 position of 2'-deoxyuridine triphosphate via a photocleavable 2-nitrobenzyl linker. We demonstrate that the nucleotide analogue can be faithfully incorporated by a DNA polymerase Thermo Sequenase into the growing DNA strand in a DNA-sequencing reaction and that its incorporation does not hinder the addition of the subsequent nucleotide. These results indicate that the nucleotide analogue is an excellent substrate for Thermo Sequenase. We also systematically studied the photocleavage of the fluorescent dye from a DNA molecule that contained the nucleotide analogue. UV irradiation at 340 nm of the DNA molecule led to the efficient release of the fluorescent dye, ensuring that a previous fluorescence signal did not leave any residue that could interfere with the detection of the next nucleotide. Thus, our results indicate that it should be feasible to use four different fluorescent dyes with distinct fluorescence emissions as unique tags to label the four nucleotides (A, C, G, and T) through the photocleavable 2-nitrobenzyl linker. These fluorescent tags can be removed easily by photocleavage after the identification of each nucleotide in the DNA sequencing-by-synthesis approach.**

nucleoside triphosphate | 2-nitrobenzyl linker | photocleavage | fluorescence | DNA sequencing by synthesis

The ability to sequence DNA accurately and rapidly has revolutionized biology and medicine. The development of high-throughput genetic-analysis technologies involving chemistry, engineering, biology, and computer science has made it possible to move from studying single genes at a time to analyzing and comparing entire genomes. The current state-of-the-art technology for high-throughput DNA sequencing utilizes capillary array DNA sequencers with laser-induced fluorescence detection (1–4). Although capillary array DNA-sequencing technology addresses the throughput and read length requirements of large-scale DNA-sequencing projects to some extent, the sequencing throughput and accuracy required for disease-gene discovery and personalized medicine needs to be improved significantly. For example, electrophoresis-based DNA-sequencing methods have difficulty detecting heterozygotes unambiguously and are not 100% accurate in regions rich in nucleotides comprising guanine or cytosine because of compression (5, 6). In addition, the first few bases after the priming site are often masked by the high fluorescence signal from excess dye-labeled primers or dye-labeled terminators and are difficult to identify. Thus, the requirement of electrophoresis for DNA sequencing is still a significant limitation for high-throughput DNA-sequencing and mutation-detection projects. A variety of alternative methods have been exploited for DNA sequencing including mass spectrometry sequencing (7–9), pyrosequencing

(10), sequencing by hybridization (11), and massively parallel signature sequencing with enzymatic cleavage and ligation (12).

The concept of sequencing DNA by synthesis was revealed in 1988 (13). This approach involves the detection of the identity of each nucleotide immediately after its incorporation into a growing strand of DNA in polymerase reaction. As yet, no complete success has been reported by using such a system to sequence DNA unambiguously. The pyrosequencing approach that employs four natural nucleotides (A, C, G, and T) and several other enzymes for sequencing DNA by synthesis is now widely used for mutation detection (10). In this approach, the detection is based on a pyrophosphate (PPi) released during the DNA polymerase reaction, the quantitative conversion of pyrophosphate to ATP by sulfurylase and the subsequent production of visible light by firefly luciferase. This procedure can sequence routinely up to 30 bp of nucleotide sequences. However, in this method each of the four nucleotides needs to be added and detected separately. Additionally, it is difficult to accurately identify long stretches of the same base with this method. Another approach in this direction has been focused mostly on designing and synthesizing a photocleavable (PC) chemical moiety that is also linked to a fluorescent dye to cap the 3'-OH group of dNTPs (14, 15). The rationale behind this work is to use the PC fluorescent moiety as an identifying tag for the incorporated nucleotide while simultaneously capping the 3'-OH group, thus preventing further polymerase activity until the tag is removed by using photocleavage. However, only limited success has been reported for the incorporation of such a 3'-modified nucleotide by DNA polymerase into a growing DNA strand. This is because the 3' position on the deoxyribose is very close to the amino acid residues in the active site of the polymerase (16), and the polymerase is sensitive to modification with bulky groups in this area of the ribose ring. On the other hand, it is known that a modified DNA polymerase such as Thermo Sequenase (17, 18) is highly tolerable to extensive modifications with large groups such as energy transfer dyes at the 5 position of pyrimidines (T and C) and 7 position of purines (G and A) (19–21). We therefore reasoned that if a unique fluorescent dye is linked to the 5 position of the pyrimidines (T and C) or 7 position of purines (G and A) through a PC linker, and a small chemical group is used to cap the 3'-OH, the resulting nucleotide analogue should be a good substrate of the DNA polymerase Thermo Sequenase. Fluorescence detection coupled with the subsequent photocleavage of the fluorophore and selective removal of the 3'-capping group from the nucleotide analogues will allow unambiguous decoding of the DNA sequence in a polymerase reaction. The advantage of using the photon as a reagent for initiating photoreactions to cleave the fluorophore is that no additional chemical reagents are required to be introduced into

Abbreviations: PC, photocleavable; MALDI, matrix-assisted laser desorption ionization; TOF, time of flight; BODIPY, 4,4-difluoro-5,7-dimethyl-4-bora-3 $\alpha$ ,4 $\alpha$ -diazas-indacene propionic acid; NHS, *N*-hydroxysuccinimide; dd, dideoxy-; SPC, solid-phase capture.

<sup>‡</sup>To whom correspondence should be addressed at: Russ Berrie Medical Science Pavilion, Room 405A, Columbia Genome Center, Columbia University College of Physicians and Surgeons, New York, NY 10032. E-mail: dj222@columbia.edu.

the system and clean products can be generated with no requirement for subsequent purification.

We have designed a massive parallel DNA-sequencing chip system (22) based on the above-described approach. This sequencing system includes the construction of a chip with immobilized single-stranded DNA templates that can self-prime for the generation of the complementary DNA strand in the polymerase reaction and four unique fluorescently labeled PC nucleotide analogues with the 3'-OH capped by a small chemical moiety to allow efficient incorporation into the growing strand of DNA in the polymerase reaction. A four-color fluorescence imager is then used to identify the sequence of the incorporated nucleotide on each spot of the chip. After removing the fluorophore photochemically (UV irradiation at 340 nm) and the 3'-OH-capping group, the polymerase reaction will proceed to incorporate the next nucleotide analogue and detect the next base.

To this end we synthesized PC dye-labeled oligonucleotides and confirmed that the 2-nitrobenzyl moiety can be used as a PC linker, connecting the fluorophore with the oligonucleotide, and can be photocleaved efficiently by near-UV irradiation (23). We report here the design and synthesis of a nucleotide analogue with a fluorescent dye attached to the 5 position of 2'-deoxyuridine triphosphate via the PC 2-nitrobenzyl linker. We have demonstrated that this nucleotide analogue can be incorporated faithfully by DNA polymerase Thermo Sequenase into a growing DNA strand in a base-specific manner, and the fluorescent dye can be photocleaved efficiently after incorporation by near-UV irradiation.

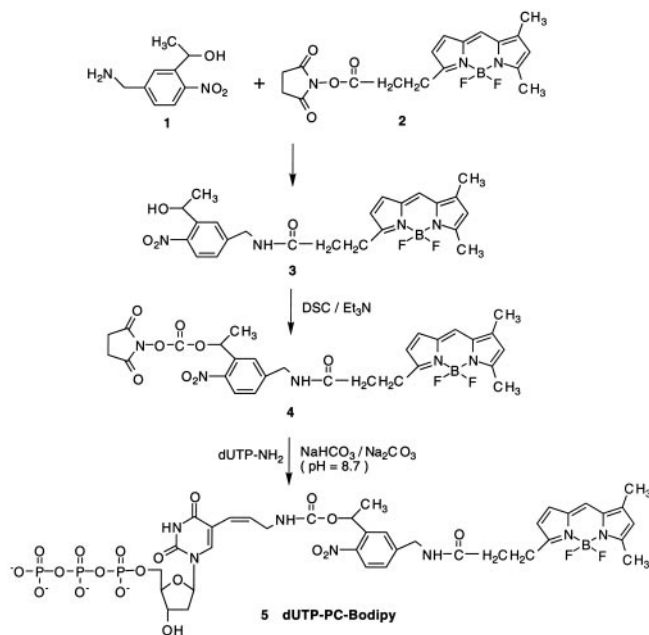
## Materials and Methods

All chemicals were purchased from Sigma–Aldrich unless otherwise indicated.  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra were recorded on a Bruker 300 spectrometer. High-resolution MS data were obtained by using a JEOL (Tokyo) JMS HX 110A mass spectrometer. Mass measurements were made on a Voyager DE matrix-assisted laser desorption ionization (MALDI)/time-of-flight (TOF) mass spectrometer (Applied Biosystems). Electrospray ionization MS was recorded on a Micromass (Manchester, U.K.) quadrupole-TOF mass spectrometer. HPLC was performed on a Waters system consisting of a Rheodyne 7725i injector, 600 controller, and a 996 photodiode array detector.

**Synthesis of dUTP-PC-4,4-difluoro-5,7-dimethyl-4-bora-3 $\alpha$ ,4 $\alpha$ -diazas-indacene Propionic Amide (BODIPY).** dUTP-PC-BODIPY was synthesized as shown in Scheme 1, involving the following steps.

**PC-BODIPY 3.** 1-[5-(Aminomethyl)-2-nitrophenyl]ethanol **1** (5 mg, 0.026 mmol) prepared according to the literature (24) was dissolved in 550  $\mu\text{l}$  of acetonitrile and 100  $\mu\text{l}$  of 1 M  $\text{NaHCO}_3$  aqueous solution. A solution of BODIPY *N*-hydroxysuccinimidyl (NHS) ester **2** (Molecular Probes) (5 mg, 0.013 mmol) in 400  $\mu\text{l}$  of acetonitrile was added slowly to the above reaction mixture and then stirred for 5 h at room temperature. The resulting reaction mixture was purified on a preparative silica-gel TLC plate ( $\text{CHCl}_3/\text{CH}_3\text{OH} = 95:5$ ) to give pure PC-BODIPY **3** (5.81 mg, 95%).  $^1\text{H}$  NMR (300 MHz,  $\text{CD}_3\text{OD}$ )  $\delta$  7.83 (d, 1H), 7.78 (s, 1H), 7.45 (s, 1H), 7.32 (d, 1H), 7.01 (d, 1H), 6.33 (d, 1H), 6.23 (s, 1H), 5.37 (q, 1H), 4.47 (s, 2H), 3.30 (t, 2H), 2.75 (t, 2H), 2.52 (s, 3H), 2.30 (s, 3H), 1.49 (d, 3H).  $^{13}\text{C}$  NMR (75.2 MHz,  $\text{CD}_3\text{OD}$ )  $\delta$  173.7, 160.4, 157.3, 146.9, 145.2, 144.9, 142.3, 135.5, 133.9, 128.6, 126.7, 126.4, 124.8, 124.3, 120.4, 116.7, 65.0, 42.6, 34.9, 24.5, 24.1, 13.8, 10.2. High-resolution MS (fast atom bombardment-positive)  $m/z$ : anal. calcd for  $\text{C}_{23}\text{H}_{24}\text{O}_4\text{N}_4\text{FB}$  ( $\text{M} - \text{F} + \text{H}^+$ ), 450.1879; found, 450.1936.

**PC-BODIPY NHS ester 4.** *N,N'*-disuccinimidyl carbonate (4.27 mg, 0.017 mmol) and triethylamine (4.6  $\mu\text{l}$ , 0.033 mmol) were added to a solution of PC-BODIPY **3** (5 mg, 0.011 mmol) in 200  $\mu\text{l}$  of dry acetonitrile. The reaction mixture was stirred under argon at



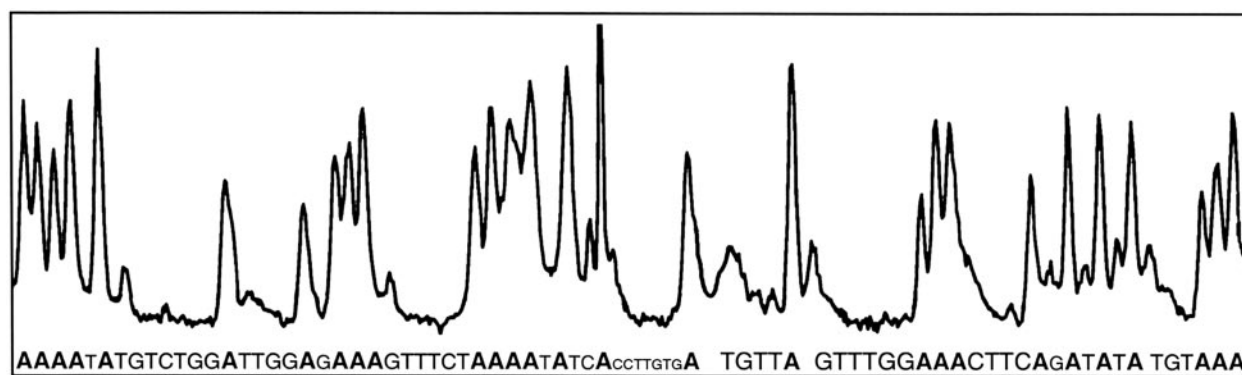
DSC : *N,N'*-disuccinimidyl carbonate

dUTP-NH<sub>2</sub> : 5-(3-aminoallyl)-2'-deoxyuridine-5'-triphosphate

**Scheme 1.** Synthesis of dUTP-PC-BODIPY.

room temperature for 6 h. The solvent was removed under vacuum, and 1 ml of 1 M  $\text{NaHCO}_3$  aqueous solution was added to the residual mixture. The crude product was extracted with ethyl acetate three times, and the combined extract was washed with saturated sodium-chloride solution, dried over  $\text{Na}_2\text{SO}_4$ , and then concentrated under vacuum. The concentrated product was purified on a preparative silica-gel TLC plate ( $\text{CHCl}_3/\text{CH}_3\text{OH} = 95:5$ ) to obtain pure PC-BODIPY NHS ester **4** (2.60 mg, 40% yield).  $^1\text{H}$  NMR (300 MHz,  $\text{CD}_3\text{OD}$ )  $\delta$  7.99 (d, 1H), 7.66 (s, 1H), 7.44 (m, 2H), 7.02 (d, 1H), 6.36 (m, 2H), 6.23 (s, 1H), 4.55 (d, 1H), 3.30 (t, 3H), 2.77 (s, 4H), 2.75 (t, 2H), 2.52 (s, 3H), 2.30 (s, 3H), 1.61 (d, 3H).  $^{13}\text{C}$  NMR (75.2 MHz,  $\text{CD}_3\text{OD}$ )  $\delta$  173.8, 170.2, 160.3, 157.4, 151.3, 146.7, 146.4, 145.1, 144.9, 136.2, 135.6, 133.9, 128.7, 128.5, 125.9, 125.1, 120.4, 116.8, 70.2, 42.5, 34.9, 25.3, 24.5, 20.9, 13.8, 10.2. High-resolution MS (fast atom bombardment-positive)  $m/z$ : anal. calcd for  $\text{C}_{28}\text{H}_{29}\text{O}_8\text{N}_5\text{F}_2\text{B}$  ( $\text{M} + \text{H}^+$ ), 612.2082; found, 612.2102.  $m/z$ : anal. calcd for  $\text{C}_{28}\text{H}_{29}\text{O}_8\text{N}_5\text{FB}$  ( $\text{M} - \text{F} + \text{H}^+$ ), 593.2098; found, 593.2122.

**dUTP-PC-BODIPY 5.** PC-BODIPY NHS ester **4** (7.2 mg, 0.012 mmol) in 300  $\mu\text{l}$  of acetonitrile was added to a solution of 5-(3-aminoallyl)-2'-deoxyuridine 5'-triphosphate (dUTP-NH<sub>2</sub>, Sigma) (2 mg, 0.004 mmol) in 300  $\mu\text{l}$  of  $\text{Na}_2\text{CO}_3$ - $\text{NaHCO}_3$  buffer (0.1 M, pH 8.7). The reaction mixture was stirred at room temperature for 3 h. A preparative silica-gel TLC plate was used to separate the unreacted PC-BODIPY NHS ester from the fractions containing dUTP-PC-BODIPY ( $\text{CHCl}_3/\text{CH}_3\text{OH} = 85:15$ ). The product was concentrated further under vacuum and purified with reverse-phase HPLC on a 150  $\times$  4.6-mm C18 column to obtain the pure product (retention time = 35 min). Mobile phase: A, 8.6 mM triethylammonium/100 mM hexafluoroisopropyl alcohol in water (pH 8.1); B, methanol. Elution was performed with 100% A isocratic over 10 min followed by a linear gradient of 0–50% B for 20 min and then 50% B isocratic over another 20 min. The compound dUTP-PC-BODIPY **5** was characterized by TOF MS electrospray+  $m/z$ : anal. calcd for  $\text{C}_{36}\text{H}_{44}\text{O}_{19}\text{N}_7\text{F}_2\text{BP}_3$  ( $\text{M} + \text{H}^+$ ), 1,020.50; found, 1,020.27.



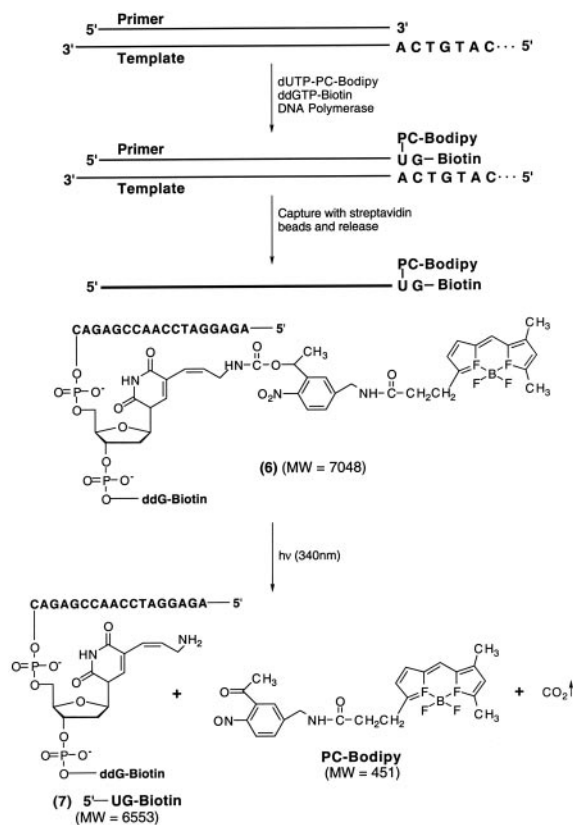
**Fig. 1.** DNA-sequencing electropherogram demonstrating the ability of the DNA polymerase to incorporate dUTP-PC-BODIPY into a growing DNA strand in a sequence-specific manner. Each peak corresponds to the fluorescence signal emitted by dUTP-PC-BODIPY incorporated into a Sanger DNA-sequencing fragment terminated with biotin-ddATP.

**DNA-Sequencing Reaction.** A Sanger DNA-sequencing reaction was conducted by using dUTP-PC-BODIPY **5**, all four dNTPs (A, C, G, and T), and biotinylated dideoxy-ATP (biotin-ddATP, Perkin-Elmer) as the only terminator. The 30- $\mu$ l reaction mixture consisted of 20 pmol of primer (5'-TTCACATAAG-GTTTTGCTGACA-3'), 300 ng of template (PCR amplicon of exon 11 of the *BRC42* gene), 60 pmol of biotin-ddATP, 4,500 pmol of dNTPs, 300 pmol of dUTP-PC-BODIPY **5**, and 10 units of DNA polymerase Thermo Sequenase (Amersham Pharmacia). The reaction consisted of 15 cycles of 94°C for 20 sec, 48°C for 30 sec, and 60°C for 90 sec. The resultant DNA-sequencing fragments were purified from unextended primer and other reaction components by solid-phase capture (SPC) using streptavidin-coated magnetic beads (Seradyn, Indianapolis) (25, 26). Only correctly terminated sequencing fragments had a biotin moiety at their 3' ends. Therefore only these fragments were captured on the streptavidin-coated beads, whereas all other components were washed away. The biotin-streptavidin bond then was cleaved by heating the beads with 98% formamide solution (containing 1 M EDTA) to obtain purified DNA-sequencing fragments. These fragments were separated on the basis of their size by a capillary array electrophoresis DNA sequencer (MegaBACE 1000, Amersham Pharmacia) to generate a fluorescent electropherogram (Fig. 1).

**DNA-Extension Reaction and Photocleavage Study.** **DNA-extension reaction.** A DNA-extension reaction (Fig. 2) was conducted by using dUTP-PC-BODIPY **5**, ddGTP-biotin (Perkin-Elmer), a primer (5'-AGAGGATCCAACCGAGAC-3'), and a synthetic DNA template (100 bp) corresponding to a portion of exon 7 of the human *p53* gene. The two nucleotides in the template immediately adjacent to the annealing site of the primer were 5'-CA-3'. Thus, the extension reaction would be terminated after extension by only two bases (U-PC-BODIPY and then G) to generate DNA fragment **6**. The reaction mixture consisted of 20 pmol of template, 40 pmol of primer, 40 pmol of dUTP-PC-BODIPY, 40 pmol of ddGTP-biotin, 2  $\mu$ l of 10 $\times$  reaction buffer, and 2 units of DNA polymerase Thermo Sequenase in a total volume of 20  $\mu$ l. The reaction consisted of 25 cycles of 94°C for 20 sec, 48°C for 40 sec, and 72°C for 20 sec. Subsequently, the extension product was purified by SPC and reverse-phase HPLC. An Xterra MS C18 (4.6  $\times$  50-mm) column (Waters) was used for the HPLC analysis. Elution was performed over 90 min at a flow rate of 0.5 ml/min with the temperature set at 50°C by using a linear gradient (12–34.5%) of methanol in a buffer consisting of 8.6 mM aqueous triethylammonium and 100 mM hexafluoroisopropyl alcohol (pH 8.1). The fraction containing

the desired product was collected and freeze-dried for MALDI-TOF MS measurement and photolysis as described (23).

**Photolysis of extension product 6 and HPLC analysis.** The purified DNA-extension product was dissolved in 1:1 (vol/vol) acetonitrile/water to a final concentration of 2  $\mu$ M and irradiated in a quartz cell (1-cm path length) employing an LX300UV xenon lamp (ILC Technologies, Sunnyvale, CA) in conjunction with a monochromator at 340 nm (light intensity = 3 mW/cm<sup>2</sup>). Fractions were taken out at different irradiation times for HPLC and MALDI-TOF MS analysis.



**Fig. 2.** DNA-extension reaction to incorporate dUTP-PC-BODIPY, generating a DNA fragment **6** and the photocleavage reaction of **6** to produce DNA fragment **7** and PC-BODIPY. MW, molecular weight.



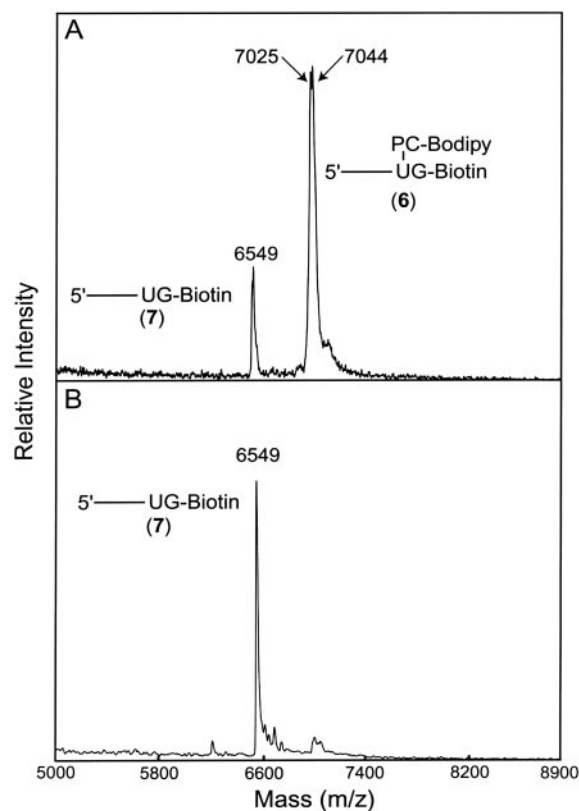
## Results and Discussion

The PC fluorescent nucleotide analogue, dUTP-PC-BODIPY **5**, synthesized as shown in Scheme 1, bears a fluorescent dye on its 5' position via a PC 2-nitrobenzyl linker. The 2-nitrobenzyl moiety has been shown to have a high photocleavage efficiency under UV irradiation ( $\lambda \approx 340$  nm) (24, 27–30) and thus was chosen as the PC linker in our study. BODIPY, which has been used in electrophoresis-based DNA sequencing and protein analysis (31, 32), was chosen as the fluorophore to couple with dUTP-NH<sub>2</sub>.

To use the dUTP-PC-BODIPY for DNA sequencing, it is critical that the nucleotide analogue be incorporated faithfully and efficiently into a growing DNA strand in a polymerase reaction. To verify this, we conducted a Sanger DNA-sequencing reaction by using dUTP-PC-BODIPY **5**, all four dNTPs (A, C, G, and T), and biotin-ddATP. The basis of this experiment was that if dUTP-PC-BODIPY **5** was a good substrate of DNA polymerase, there would be a significant number of DNA-sequencing fragments with dUTP-PC-BODIPY incorporated in them, yielding a strong fluorescent signal for each DNA fragment ending with A, because each fragment bearing the fluorescence signal has been terminated with biotin-ddATP and isolated by streptavidin-coated magnetic beads (25). The experimental results matched well with the known template sequence, illustrating the highly efficient incorporation of dUTP-PC-BODIPY in a base-specific manner in the polymerase reaction.

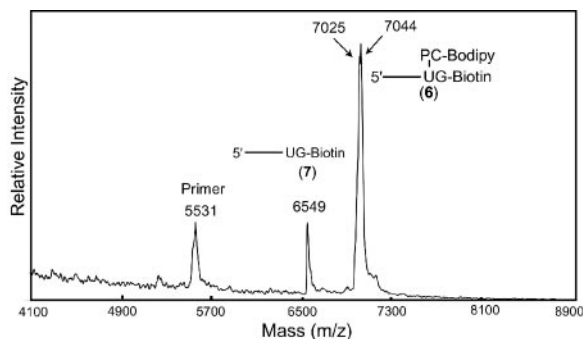
It was observed that the electropherogram in Fig. 1 consisted of some minor peaks accompanying the major A peaks. The appearance of these peaks can be explained by the variable incorporation of dUTP-PC-BODIPY into the growing DNA strand. Each DNA fragment has multiple copies present in the sequencing product mixture. Each such copy will have numerous possible sites of incorporation of dUTP-PC-BODIPY. Because dUTP-PC-BODIPY competes with dTTP for incorporation, which is an event that is governed by their relative amounts in the reaction mixture, it is quite possible that DNA fragments of the same length have different numbers of dUTP-PC-BODIPY incorporated in them. Such fragments therefore will have different mobilities during high-resolution capillary gel electrophoresis. Thus the multiple peaks in the sequence electropherogram will correspond to sequencing fragments of the same length bearing variable numbers of dUTP-PC-BODIPY incorporated in them. We conducted a series of experiments with varying amounts of dUTP-PC-BODIPY and dTTP to investigate the effect of their relative amounts on the appearance of these multiple peaks. The results shown in Fig. 1, were obtained by using a 15:1 ratio of dTTP to dUTP-PC-BODIPY. Higher amounts of dUTP-PC-BODIPY exacerbated the problem of the appearance of multiple peaks.

To further evaluate the feasibility of dUTP-PC-BODIPY as a PC substrate of DNA polymerase, we conducted a DNA-extension reaction as depicted schematically in Fig. 2. The extension reaction was performed by using dUTP-PC-BODIPY **5**, ddGTP-biotin, and a synthetic template (100 bp) corresponding to a portion of exon 7 of the human *p53* gene. The two nucleotides in the template immediately adjacent to the annealing site of the primer (5'-AGAGGATCCAACCGAGAC-3') were 5'-CA-3'. Consequently, the primer was extended in the presence of dUTP-PC-BODIPY and ddGTP-biotin such that the extension reaction would be terminated after extension by only two bases (U and then G) to generate DNA fragment **6**. The purified extension product **6** was analyzed by MALDI-TOF MS as shown in Fig. 3A, where a strong signal corresponding to **6**

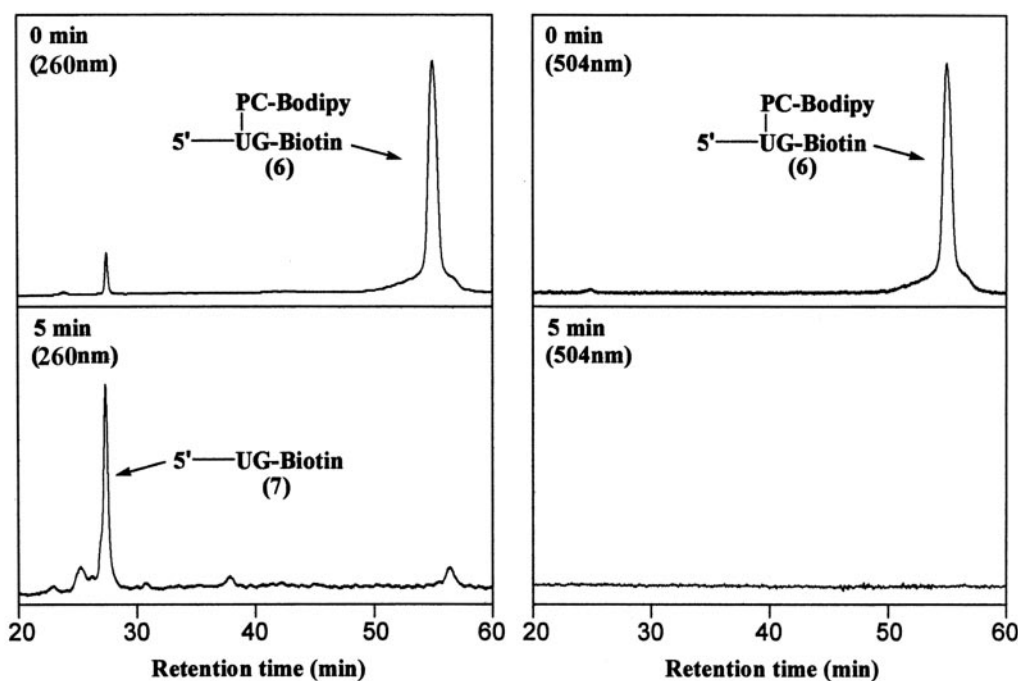


**Fig. 3.** MALDI-TOF MS spectra of the extension product **6** ( $m/z$ : found, 7,044; anal. calcd, 7,048) and its photocleavage fragment **7** ( $m/z$ : found, 6,549; anal. calcd, 6,553). (A) Without irradiation. (B) After 2 min of irradiation ( $\lambda_{\text{irr}} = 340$  nm).

( $m/z$ : found, 7,044; anal. calcd, 7,048) is observed. The peak at  $m/z$  7,025 is most likely due to the loss of a fluorine atom from the parent molecule **6** during the MS analysis. The loss of a fluorine atom was also observed in the mass spectrum of PC-BODIPY **3**. A small peak at  $m/z$  6,549 that corresponds to the photocleaved fragment **7** ( $m/z$ : anal. calcd, 6,553) was observed also. This is due to the photocleavage induced by the nitrogen laser pulse (337 nm) used for ionization in MALDI-TOF. When the purified extension product **6** was injected into a capillary array electrophoresis DNA sequencer, only one strong peak was observed from the electropherogram, indicating that only DNA fragments terminated by ddGTP-biotin were present



**Fig. 4.** Mass spectrum of extension product mixture from the reaction shown in Fig. 2 before SPC purification. A primer peak, the extension product **6**, and the photocleaved fragment **7** were observed. No false stop peak caused by dUTP-PC-BODIPY was observed.



**Fig. 5.** HPLC analysis of the photolysis of the extension product **6** ( $2 \mu\text{M}$ ) in 1:1 (vol/vol) water/acetonitrile. Detection channels: 260 nm (Left) and 504 nm (Right). (Upper) Before irradiation. (Lower) After 5 min of irradiation ( $\lambda_{\text{irr}} = 340 \text{ nm}$ ).

after purification by SPC. The MALDI-TOF mass spectrum of the extension product mixture before any purification (Fig. 4) showed three peaks corresponding to a primer peak ( $m/z$  5,531), the DNA-extension product **6**, and the photocleaved DNA fragment **7**. No peak was observed corresponding to a DNA fragment that was stopped after the incorporation of a dUTP-PC-BODIPY alone. This indicates that there are no false stops caused by dUTP-PC-BODIPY in the enzymatic extension. All these results indicate that the dUTP-PC-BODIPY can be incorporated efficiently into the growing DNA strand by a DNA polymerase, and its incorporation does not hinder the addition of the subsequent nucleotide.

Complete photocleavage of the fluorescent dye, BODIPY, from the extension product **6** is essential for the successful application of dUTP-PC-BODIPY in the DNA sequencing-by-synthesis approach. We first investigated the photocleavage efficiency of the DNA-extension product by MALDI-TOF MS. Two minutes of UV irradiation at 340 nm of the solution containing the extension product **6** eliminated the corresponding peak in the MALDI-TOF mass spectrum but significantly increased the signal of the photocleaved fragment **7** ( $m/z$  6,549) as shown in Fig. 3B.

We also investigated the photocleavage of the DNA fragment **6** by HPLC as shown in Fig. 5. Two detection channels (260 and 504 nm) corresponding to the DNA and dye chromophores, respectively, were used to monitor the cleavage process. The purified extension product **6** gave a strong single peak at retention time 56.8 min in both channels before any UV light irradiation was applied. A small peak was also observed at retention time 27.4 min in the 260-nm channel with  $\approx 5\%$  of the integration value as that of the peak at 56.8 min. This small peak was found to be the photocleavage product **7**, and it may have resulted from the exposure of the sample to UV light during the detection process in HPLC. Five minutes of irradiation at 340 nm resulted in a significant reduction of the peaks from the extension product **6**, whereas the peak at the retention time of 27.4 min in the 260-nm channel increased significantly. These data confirm that the fluorescent dye is released almost completely from

the template in solution within 5 min of UV light irradiation at 340 nm. There was a small shoulder peak observed at a retention time of 56.8 min in the 260-nm channel both before and after UV irradiation. However, no such peak was observed in the 504-nm channel corresponding to the fluorophore after photocleavage. Thus, the presence of this shoulder peak does not interfere with removal of the fluorophore from DNA.

In conclusion, we have synthesized a PC fluorescent dye-labeled nucleotide analogue. This nucleotide analogue is shown to be an excellent substrate for DNA polymerase Thermo Sequenase in a DNA-sequencing reaction. The fluorescent dye can be photocleaved efficiently after incorporation of the nucleotide analogue into a growing DNA strand. The synthesis and evaluation of this PC fluorescent nucleotide in a DNA-sequencing reaction has paved the way for the future development of the sequencing-by-synthesis approach. Moreover, the high efficiency of photocleavage of the fluorescent dye ensures that there is no carryover of the previous fluorescence signal during detection of the next nucleotide and therefore prevents any ambiguity in sequencing results. Four different fluorescent dyes with distinct emissions can therefore be used as unique identification tags for all four nucleotides (A, C, G, and T), and these fluorescent nucleotides can be used to rapidly and accurately sequence a DNA template in a DNA polymerase reaction. Further work is under way to modify the 3'-OH of the nucleotide with a small chemical group that can cap the 3'-OH without hindrance to the incorporation of the nucleotide analogue by DNA polymerase. After fluorescence detection to identify the nucleotide, the 3'-OH-capping group then can be removed by specific chemical or enzymatic reactions. All these efforts will facilitate the development of a highly efficient and automated DNA-sequencing approach.

We thank Dr. Steffen Jockusch for the help in using the photolysis instrument. This research is supported by the National Science Foundation Sensing and Imaging Initiative Grant 97793, a Packard Fellowship for Science and Engineering (to J.J.), and the Columbia University Genomics Initiative.

1. Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R., Heiner, C., Kent, S. B. H. & Hood, L. E. (1986) *Nature* **321**, 674–679.
2. Ju, J., Ruan, C., Fuller, C. W., Glazer, A. N. & Mathies, R. A. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 4347–4351.
3. Ju, J., Glazer, A. N. & Mathies, R. A. (1996) *Nucleic Acids Res.* **24**, 1144–1148.
4. Kheterpal, I., Scherer, J., Clark, S. M., Radhakrishnan, A., Ju, J., Ginther, C. L., Sensabaugh, G. F. & Mathies, R. A. (1996) *Electrophoresis* **17**, 1852–1859.
5. Bowling, J. M., Bruner, K. L., Cmarik, J. L. & Tibbetts, C. (1991) *Nucleic Acids Res.* **19**, 3089–3097.
6. Yamakawa, H. & Ohara, O. (1997) *Nucleic Acids Res.* **25**, 1311–1312.
7. Fu, D. J., Tang, K., Braun, A., Reuter, D., Darnhofer-Demar, B., Little, D. P., O'Donnell, M. J., Cantor, C. R. & Koster, H. (1998) *Nat. Biotechnol.* **16**, 381–384.
8. Roskey, M. T., Juhasz, P., Smirnov, I. P., Takach, E. J., Martin, S. A. & Haff, L. A. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 4724–4729.
9. Edwards, J. R., Itagaki, Y. & Ju, J. (2001) *Nucleic Acids Res.* **29**, e104.
10. Ronaghi, M., Uhlen, M. & Nyren, P. (1998) *Science* **281**, 363–365.
11. Drmanac, S., Kita, D., Labat, I., Hauser, B., Schmidt, C., Burczak, J. D. & Drmanac, R. (1998) *Nat. Biotechnol.* **16**, 54–58.
12. Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D. H., Johnson, D., Luo, S., McCurdy, S., Foy, M., Ewan, M., *et al.* (2000) *Nat. Biotechnol.* **18**, 630–634.
13. Hyman, E. D. (1988) *Anal. Biochem.* **174**, 423–436.
14. Metzker, M. L., Raghavachari, R., Richards, S., Jacutin, S. E., Civitello, A., Burgess, K. & Gibbs, R. A. (1994) *Nucleic Acids Res.* **22**, 4259–4267.
15. Welch, M. B. & Burgess, K. (1999) *Nucleosides Nucleotides* **18**, 197–201.
16. Pelletier, H., Sawaya, M. R., Kumar, A., Wilson, S. H. & Kraut, J. (1994) *Science* **264**, 1891–1903.
17. Reeve, M. A. & Fuller, C. W. (1995) *Nature* **376**, 796–797.
18. Tabor, S. & Richardson, C. C. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 6339–6343.
19. Rosenblum, B. B., Lee, L. G., Spurgeon, S. L., Khan, S. H., Menchen, S. M., Heiner, C. R. & Chen, S. M. (1997) *Nucleic Acids Res.* **25**, 4500–4504.
20. Zhu, Z., Chao, J., Yu, H. & Waggoner, A. S. (1994) *Nucleic Acids Res.* **22**, 3418–3422.
21. Duthie, R. S., Kalve, I. M., Samols, S. B., Hamilton, S., Livshin, I., Khnt, M., Nampalli, S., Kumar, S. & Fuller, C. W. (2002) *Bioconjugate Chem.* **13**, 699–706.
22. Ju, J., Li, Z., Edwards, J. & Itagaki, Y. (2002) *PCT Intl. Patent Appl. WO* 0229003.
23. Bai, X., Li, Z., Jockusch, S., Turro, N. J. & Ju, J. (2003) *Proc. Natl. Acad. Sci. USA* **100**, 409–413.
24. Olejnik, J., Sonar, S., Krzymanska-Olejnik, E. & Rothschild, K. J. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 7590–7594.
25. Ju, J. (2002) *Anal. Biochem.* **309**, 35–39.
26. Kim, S., Edwards, J. R., Deng, L., Chung, W. & Ju, J. (2002) *Nucleic Acids Res.* **30**, e85.
27. Pillai, V. N. R. (1980) *Synthesis* 1–26.
28. Guillier, F., Orain, D. & Bradley, M. (2000) *Chem. Rev. (Washington, D.C.)* **100**, 2091–2157.
29. Bochet, C. G. (2002) *J. Chem. Soc. Perkin Trans. 1*, 125–142.
30. Kahl, J. D. & Greenberg, M. M. (1999) *J. Am. Chem. Soc.* **121**, 597–604.
31. Metzker, M. L., Lu, J. & Gibbs, R. A. (1996) *Science* **271**, 1420–1422.
32. Gite, S., Mamaev, S., Olejnik, J. & Rothschild, K. (2000) *Anal. Biochem.* **279**, 218–225.