

Termination of DNA synthesis by N^6 -alkylated, not 3'-O-alkylated, photocleavable 2'-deoxyadenosine triphosphates

Weidong Wu¹, Brian P. Stupi¹, Vladislav A. Litosh¹, Dena Mansouri²,
Demetra Farley³, Sidney Morris^{1,3}, Sherry Metzker¹ and Michael L. Metzker^{1,2,3,*}

¹LaserGen, Inc., Houston, TX 77054, ²Department of Molecular & Human Genetics and ³Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA

Received July 25, 2007; Revised August 17, 2007; Accepted August 21, 2007

ABSTRACT

The Human Genome Project has facilitated the sequencing of many species, yet the current Sanger method is too expensive, labor intensive and time consuming to accomplish medical resequencing of human genomes *en masse*. Of the 'next-generation' technologies, cyclic reversible termination (CRT) is a promising method with the goal of producing accurate sequence information at a fraction of the cost and effort. The foundation of this approach is the reversible terminator (RT), its chemical and biological properties of which directly impact the performance of the sequencing technology. Here, we have discovered a novel paradigm in RT chemistry, the attachment of a photocleavable, 2-nitrobenzyl group to the N^6 -position of 2'-deoxyadenosine triphosphate (dATP), which, upon incorporation, terminates DNA synthesis. The 3'-OH group of the N^6 -(2-nitrobenzyl)-dATP remains unblocked, providing favorable incorporation and termination properties for several commercially available DNA polymerases while maintaining good discrimination against mismatch incorporations. Upon removal of the 2-nitrobenzyl group with UV light, the natural nucleotide is restored without molecular scarring. A five-base experiment, illustrating the exquisite, stepwise addition through a homopolymer repeat, demonstrates the applicability of the N^6 -(2-nitrobenzyl)-dATP as an ideal RT for CRT sequencing.

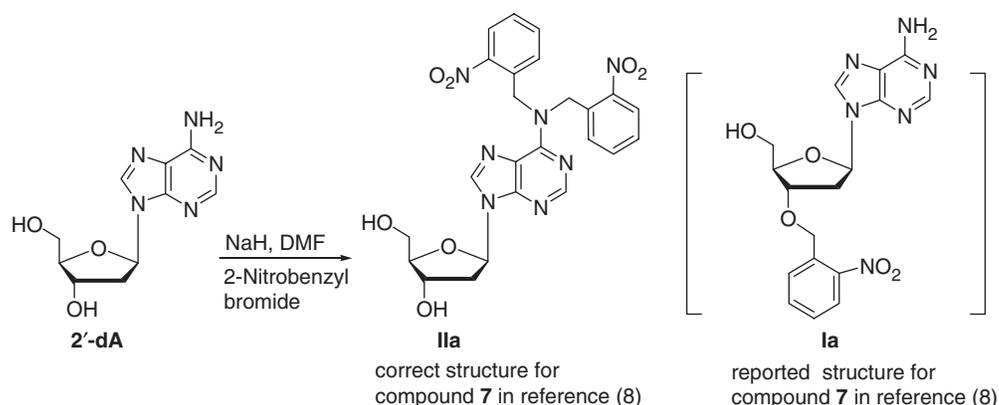
INTRODUCTION

Next-generation technologies are being developed to advance sequencing to the \$100 000, and eventually the

\$1000 genome. A number of strategies, albeit at different stages of development, have been proposed including pyrosequencing, sequencing-by-ligation (SBL), cyclic reversible termination (CRT), real-time sequencing and nanopore sequencing (1–6). CRT is a promising approach, which is comprised of a three-step process of incorporating modified nucleotides, fluorescence imaging and deprotecting after which the cycle begins again (5,6). CRT reactions can be performed in a high-density format using single-molecule arrays (3) or oligonucleotide arrays (5,7), eliminating the requirement for gel electrophoresis while significantly increasing sequence throughput. At the center of this technology is the reversible terminator (RT), whereby DNA polymerases exhibit specific and efficient incorporation of the modified nucleotide into the growing primer strand, with deprotection chemistries resulting in the efficient removal of the terminating group.

To date, known RTs have contained labile blocking groups at the 3'-OH of the ribose sugar resulting in termination of synthesis (7–11). In 1994, Metzker *et al.* reported the synthesis of 3'-O-(2-nitrobenzyl)-2'-deoxyadenosine and incorporation of its triphosphate by several DNA polymerases (8). The 2-nitrobenzyl group and its derivatives are widely used as photocleavable, 'caging' functionalities for altering normal biomolecular processes (12). Recently, we discovered that the reported synthesis of 3'-O-(2-nitrobenzyl)-2'-deoxyadenosine **Ia** [designated as compound 7 in Metzker *et al.* (8)] was incorrect, and that the actual product obtained from the reaction of 2'-deoxyadenosine with 2-nitrobenzyl bromide using NaH in DMF is N^6,N^6 -bis-(2-nitrobenzyl)-2'-deoxyadenosine **Ia** (Scheme 1). To investigate whether the 3'-O-alkylated compound could act as a terminator of DNA synthesis and confirm the identity of the active triphosphate reported by Metzker *et al.* (8), we synthesized and characterized 3'-O-(2-nitrobenzyl)-2'-deoxyadenosine analogs **Ia–Ic** and N^6,N^6 -bis-(2-nitrobenzyl)-2'-deoxyadenosine analogs **Ia–Ic** (Figure 1), both of which **Ic** and **Ic** triphosphates proved inactive

*To whom correspondence should be addressed. Tel: +1 713 798 7565; Fax: +1 713 798 5741; Email: mmetzker@bcm.edu



Scheme 1. Correct structure of the product obtained from the reaction of 2'-deoxyadenosine with 2-nitrobenzyl bromide using NaH/DMF conditions. The reported compound 7 in reference (8) was misassigned as the structure **Ia**.

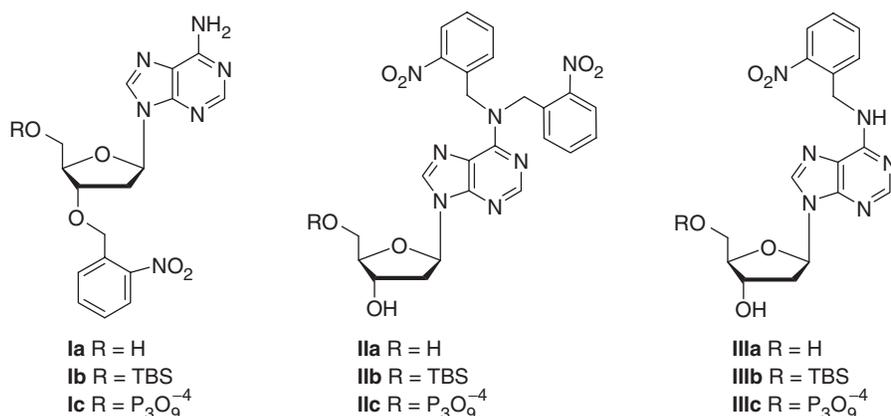


Figure 1. Structures of 3'-O-alkylated and N⁶-alkylated 2'-deoxyadenosine analogs.

by DNA polymerase incorporation assays. This finding led to the synthesis and characterization of a third set of N⁶-(2-nitrobenzyl)-2'-deoxyadenosine analogs **IIIa–IIIc** (Figure 1), of which compound **IIIb** was originally identified by us as an intermediate during the ultraviolet (UV) light-induced deprotection of compound **IIIb** to its corresponding parent analog. Here, we report that the active triphosphate described by Metzker *et al.* (8) is actually the N⁶-(2-nitrobenzyl)-dATP **IIIc**, representing a novel, 3'-unblocked terminator of DNA synthesis and an ideal candidate as a RT for the CRT approach.

MATERIALS AND METHODS

Reagents and materials

Chemical reagents and solvents were purchased from Alfa Aesar, Sigma-Aldrich, or EM Sciences. Oligonucleotides were purchased from Integrated DNA Technologies. 2'-Deoxyadenosine triphosphate (dATP), 2',3'-dideoxyadenosine triphosphate (ddATP) and Q Sepharose Fast Flow anion-exchange resin were purchased from GE Healthcare Life Sciences. All DNA polymerases and apyrase were purchased from New England Biolabs, with the exception of *AmpliTag*FS being purchased

from Applied Biosystems (AB). Analytical silica gel 60 F₂₅₄ TLC plates were purchased from Whatman, and silica gel 60 (230–400 mesh) was purchased from EM Sciences.

Nucleosides and nucleotides synthesis

Complete experimental procedures describing synthesis of the compounds used in this work are available in the Supplementary Data.

UV deprotection studies for compounds **Ib**, **IIIb** and **IIIb**

5'-O-*tert*-butyldimethylsilyl (TBS) derivatives **Ib**, **IIIb** and **IIIb** in methanol (0.2 ml, 1 mM) were transferred to a Wheaton scintillation vial and irradiated with a 0.5 mW transilluminator light source at either 302 or 365 nm. Aliquots of the irradiated solution were taken at different time intervals and analyzed for the loss of starting material and appearance of the deprotected product by reverse-phase (RP) HPLC. Deprotection half-life times (DT_{1/2}) were determined from kinetic plots at which 50% of the compound was deprotected (i.e. loss of the 2-nitrobenzyl group). UV deprotection experiments were performed in triplicate for each compound.

Oligonucleotide template assay

As described for the polymerase end-point (PEP) assays, compounds **Ic**, **Ic** and **IIIc** were tested for incorporation with *Bst* DNA polymerase at concentrations of 200 nM and 100 μ M using the BODIPY-R6G labeled primer-1 (5'-TTGTA~~AAACGACGGCCAGT~~) (13) and oligoTemplate-1 (5'-TACGGAGCAGTTTTTACTGGC CGTCGTTTTACA, interrogation base is underlined and bolded) complex. Reactions were quenched with 10 μ l of stop solution and analyzed using an AB model 377 DNA sequencer.

PEP assay

All DNA polymerases (see Supplementary Data for definitions) were assayed in 1 \times ThermoPol buffer (20 mM Tris-HCl, pH 8.8; 10 mM (NH₄)₂SO₄; 10 mM KCl; 2 mM MgSO₄; 0.1% Triton X-100, New England BioLabs). We found that the addition of Triton X-100 stimulated the activity of many of the enzymes tested (data not shown), which is consistent with other reports (14,15). For all polymerases evaluated in this study, 5 nM BODIPY-FL labeled primer-1 was annealed with 40 nM of oligoTemplate-2 (5'-TACGGAGCAGTACTGGCC GTCGTTTTACA, interrogation base is underlined and bolded) in 1 \times ThermoPol buffer at 80°C for 30 s, 57°C for 30 s and then cooled to 4°C. The primer/template complex was then diluted in half (i.e. its final concentration was 2.5 nM in a volume of 10 μ l) by the addition of DNA polymerase, nucleotide analog and ThermoPol buffer. This defines the lower limit of the IC₅₀ value for nucleotide titrations to 1.25 nM (i.e. [primer] = [primer plus incorporated nucleotide]). Polymerase reactions were incubated at their appropriate temperature (Supplementary Table 1) for 10 min, then cooled to 4°C and quenched with 10 μ l of stop solution (98% deionized formamide; 10 mM Na₂EDTA, pH 8.0; 25 mg/ml Blue Dextran, MW 2000000). Stopped reactions were heated to 90°C for 30 s and then placed on ice. The extension products were analyzed on a 10% Long Ranger (Cambrex) polyacrylamide gel using an AB model 377 DNA sequencer, the quantitative data of which are displayed as a linear-log plot of product formation versus compound concentration. PEP assays were performed in triplicate, for each DNA polymerase/nucleotide analog combination, to calculate the average IC₅₀ \pm 1 SD.

Compound **IIIc** and ddATP were then titrated using the PEP assay with the eight DNA polymerases (unit activities defined in Supplementary Table 1) in the concentration range of either 0.1 nM to 100 nM, 1 nM to 1 μ M, 10 nM to 10 μ M or 100 nM to 100 μ M. Average IC₅₀ \pm 1 SD values were calculated for compound **IIIc** and ddATP using oligoTemplate-2 as described above.

PEP termination assays for compound IIIc

OligoTemplate-2 was substituted with 40 nM of oligoTemplate-3 (5'-CCGTTTTTTTTTTACTGGCCGT CGTTTTACAGCCGCCGCCGAACCGAGAC-Biotin, interrogation bases are underlined and bolded), annealed to 5 nM BODIPY-FL labeled primer-1 and

assayed as described above. The PEP titrations were then performed at 1 \times , 5 \times and 25 \times IC₅₀ values for compound **IIIc**, using the eight DNA polymerases, and reported as % primer product, % first-base product and % second-base product.

PEP mismatch assays for compound IIIc

OligoTemplate-2 was substituted with either 40 nM of oligoTemplate-4 (5'-TACGGAGCTGAACTGGCCGTC GTTTTACA), 40 nM of oligoTemplate-5 (5'-TACG GAGCAGCACTGGCCGTCGTTTTACA) or 40 nM of oligoTemplate-6 (5'-TACGGAGCAGGACTGGCCGTC GTTTTACA, interrogation bases are underlined and bolded), annealed to 5 nM BODIPY-FL labeled primer-1, and assayed as described above. Compound **IIIc** and dATP were assayed in the concentration range of 100 nM to 100 μ M, and ddATP in the range of 500 nM to 500 μ M. Average IC₅₀ \pm 1 SD values were calculated for dATP and compound **IIIc** using oligoTemplate-4, -5 and -6, as described above.

UV deprotection studies for compound IIIc incorporated into the primer/template complex

As described for the PEP assays, compound **IIIc** was incorporated at a concentration of 100 nM, using the BODIPY-FL labeled primer-1/oligoTemplate-2 complex, and quenched with 10 μ l of stop solution. The stopped reactions were exposed to 365 nm light for 0, 10, 20, 30, 45 or 60 s, using our custom-designed UV deprotector (Supplementary Figure 1), then analyzed using an AB model 377 DNA sequencer. The quantitative data are displayed as a linear-log plot of product formation versus time. Deprotection assays were performed in triplicate to calculate the average DT_{1/2} \pm 1 SD.

Five-base sequencing experiment using compound IIIc as reversible terminator

An 80 nM solution of oligoTemplate-3 in 1 M NaCl and 1 \times ThermoPol buffer (final volume: 12.5 μ l) was incubated for 15 min at room temperature with 5 μ l of streptavidin-coated M-270 magnetic Dynabeads (Invitrogen), which had been previously washed three times with 5 μ l 1 \times ThermoPol buffer. The oligoTemplate-3 bound beads were then washed an additional three times with 5 μ l 1 \times ThermoPol buffer and annealed with 5 μ l 20 nM BODIPY-FL labeled primer-2 (5'-GGCGGCGGCG GCTGTA~~AAACGACGGCCAGT~~) in 1 \times ThermoPol buffer at 80°C for 30 s, 57°C for 30 s and then cooled to 4°C. The beads were then washed twice with 5 μ l 1 \times ThermoPol buffer.

Incorporation step. The BODIPY-FL labeled primer-2/oligoTemplate-3 complex bound beads were incubated with four units of *Bst* DNA polymerase and 250 nM of compound **IIIc** in 1 \times ThermoPol buffer (reaction volume: 20 μ l) at 65°C for 6 min, then placed on ice. Compound **IIIc**-incorporated beads were washed four times with 50 μ l W10 washing solution (10 mM Tris-HCl, pH 8.0; 10 mM Na₂EDTA; 0.1% Triton X-100), then once with 20 μ l W10 washing solution.

Deprotection step. The beads were resuspended in 20 μ l deprotection solution (20% aqueous deionized formamide; 10 mM Na₂EDTA, pH 8.0; 16.6 mg/ml Blue Dextran, MW 2 000 000), exposed to 365 nm light for 9 min (i.e. 3 \times 3 min exposures interrupted with a 15-s mixing step to ensure good resuspension of the beads) using the customized UV deprotector (Supplementary Figure 1), then washed four times with 50 μ l 1 \times apyrase buffer (100 U/l apyrase in 1 \times ThermoPol), three times with 50 μ l W10 washing solution, twice with 50 μ l 1 \times ThermoPol buffer and then once with 20 μ l 1 \times ThermoPol buffer.

The entire cycle was then repeated from the incorporation step. Final reactions were washed twice with 50 μ l W10 washing solution, once with 20 μ l W10 washing solution, quenched with 10 μ l of stop solution, heated to 50°C for 30 s and placed on ice. The extension products were analyzed on a 10% Long Ranger polyacrylamide gel using an AB model 377 DNA sequencer.

Ab initio calculations

Adenine–thymine and *N*⁶-(2-nitrobenzyl)-adenine–thymine base pairs were created with the nucleobases planar to each other, using Watson–Crick hydrogen-bond distances of 2.82 Å (N¹...H–N³) and 2.91 Å (N⁶–H...O⁴) (16), utilizing ChemDraw and Chem3D Ultra 9.0 software packages (CambridgeSoft). A series of 3D, *N*⁶-(2-nitrobenzyl)-adenine–thymine base pairs were then created by rotating the 2-nitrobenzyl group, pivoted on the *N*⁶-position of adenine, 360° at 5° intervals, with the nitro group at 0°, 30°, 45°, 60° and 90° intervals relative to the plane of the phenyl group. Chem3D structures were then further optimized using the GAMESS program (17). Restricted Hartree–Fock (RHF) energy calculations were initially determined using the STO-3G atomic orbital/shell data set, and each nitro group conformation (i.e. 0°, 30°, 45°, 60° or 90° intervals) was plotted as RHF energy versus degrees of rotation of the 2-nitrobenzyl group. Our calculations revealed that a 45° rotation of the nitro group, relative to the plane of the phenyl group, gave the lowest RHF energy calculations. The *N*⁶-(2-nitrobenzyl)-adenine–thymine base-pair conformations were further characterized using the more stringent 6-31G* atomic orbital/shell calculations and plotted as RHF energy versus degrees of rotation of the 2-nitrobenzyl group (data not shown). To evaluate the efficacy of our existing software tools, we modeled the natural adenine–thymine base pair and compared our results to those reported by Šponer *et al.* (18,19). Here, we used the 6-31G** atomic orbital/shell set, with the X, Y, Z coordinates described in Ref. (19) to perform the calculations. Showing good agreement with these reports (Supplementary Table 3), this served as an independent validation of our method.

RESULTS

Nucleosides and nucleotides synthesis

To synthesize the 3'-*O*-(2-nitrobenzyl)-2'-deoxyadenosine analog **Ia**, the 5'-hydroxyl and 6-amino groups of

2'-deoxyadenosine were protected with *tert*-butyldimethylsilyl (TBS) and *tert*-butyloxycarbonyl (Boc) groups, respectively, to yield intermediate **1**, according to Furrer and Giese (20). Transformation of this precursor into intermediate **2** occurred via deprotection and selective reprotection procedures, as outlined in Scheme 2. Alkylation of intermediate **2** with 2-nitrobenzyl bromide, using phase transfer catalysis under basic conditions, gave the desired 3'-*O* alkylated intermediate **3** in 91% yield. The *bis*-Boc groups were removed by heating on silica gel under vacuum (20) to give compound **Ib** in 91% yield, followed by the removal of the 5'-*O*-TBS group with tetra(*n*-butyl)ammonium fluoride to give compound **Ia** in 23% yield. Synthesis of the triphosphate was performed using the 'one-pot' procedure described by Ludwig (21), followed by purification using Q Sepharose FF anion-exchange chromatography to yield compound **Ic** as an ammonium salt.

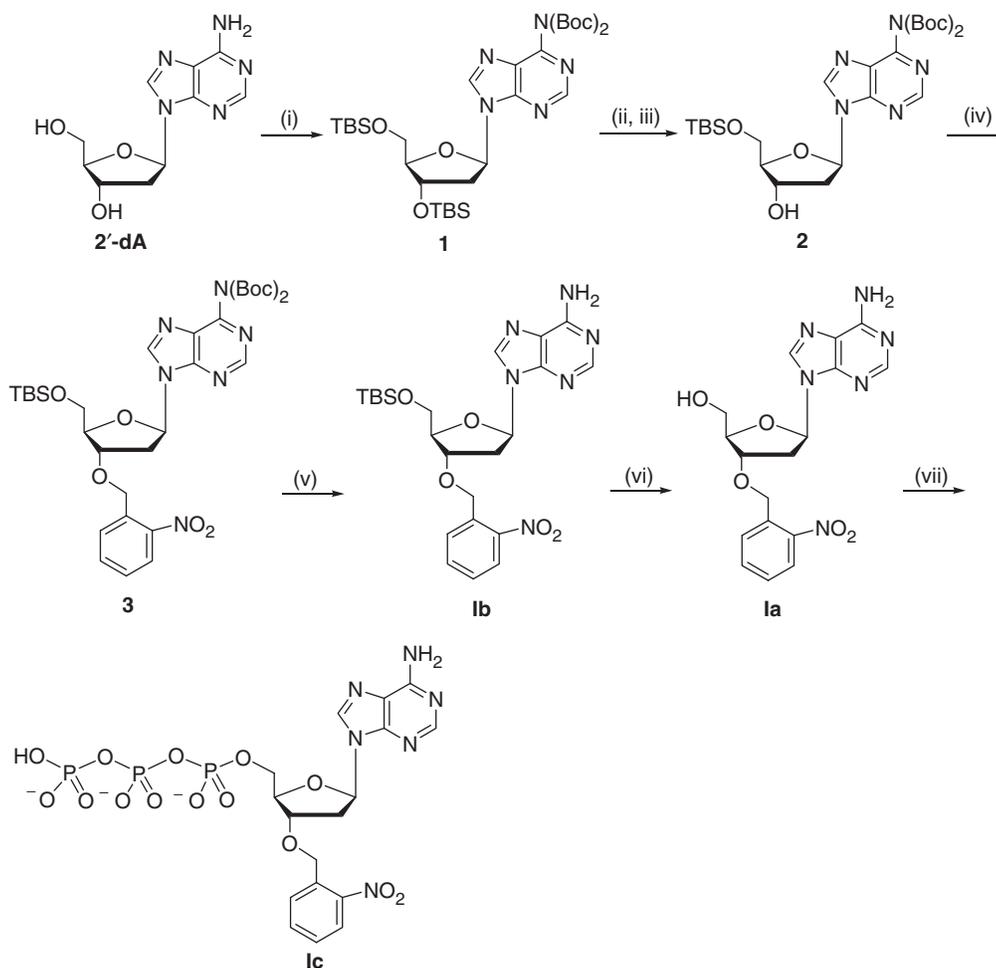
Treatment of 2'-deoxyadenosine with NaH in DMF at 0°C followed by 2-nitrobenzyl bromide gave *bis*-*N*⁶, *N*⁶-(2-nitrobenzyl)-2'-deoxyadenosine **IIa** in 22% yield (Scheme 3). The assignment of the **IIa** structure was based on the ¹H NMR spectra (in DMSO-*d*₆), which showed a total of eight aromatic hydrogens derived from the two 2-nitrobenzyl moieties and two D₂O-exchangeable signals from 3'- and 5'-hydroxyl groups of the 2'-deoxyribose. Selective 5'-*O*-TBS protection gave compound **IIb** in 57% yield. Phosphorylation of compound **IIa** was performed using the same procedure described for compound **Ia**, with the exception that purification of triphosphate **IIc** was achieved by preparative HPLC.

*N*⁶-(2-Nitrobenzyl)-2'-deoxyadenosine **IIIa** was prepared based on the work of Wan *et al.* (22). Treatment of 2'-deoxyinosine with 2-nitrobenzylamine in the presence of 1-*H*-benzotriazol-1-yloxy-*tris*(dimethylamino) phosphonium hexafluorophosphate (BOP) and *N*, *N*-diisopropylethylamine (DIPEA) in anhydrous DMF gave compound **IIIa** in 98% yield. Selective 5'-*O*-TBS protection gave compound **IIIb** in 63% yield (Scheme 4). Triphosphate **IIIc** was prepared from compound **IIIa** using the one pot procedure (21) and purified in a manner similar to that for triphosphate **Ic**.

Triphosphates **Ic**, **IIc** and **IIIc** (Figure 1) were further purified by preparative RP-HPLC, without UV detection, to provide modified triphosphates free from contamination by dATP, resulting from the deprotection of the 2-nitrobenzyl group under ordinary laboratory light conditions during synthesis and purification processes.

Identification of active nucleoside triphosphate IIIc

Triphosphates **Ic** and **IIc** were initially tested for base-specific termination of DNA synthesis using a fluorescent-based oligonucleotide template assay. All compounds were handled in low light conditions to minimize 2-nitrobenzyl deprotection. A five-base poly-thymidine template was employed to test for specific incorporation of natural and modified dATP analogs. As shown in Figure 2A, compounds **Ic** and **IIc** did not show significant incorporation using *Bst* DNA polymerase, even at high concentrations (100 μ M), although compound **Ic** did exhibit natural nucleotide contamination, even after



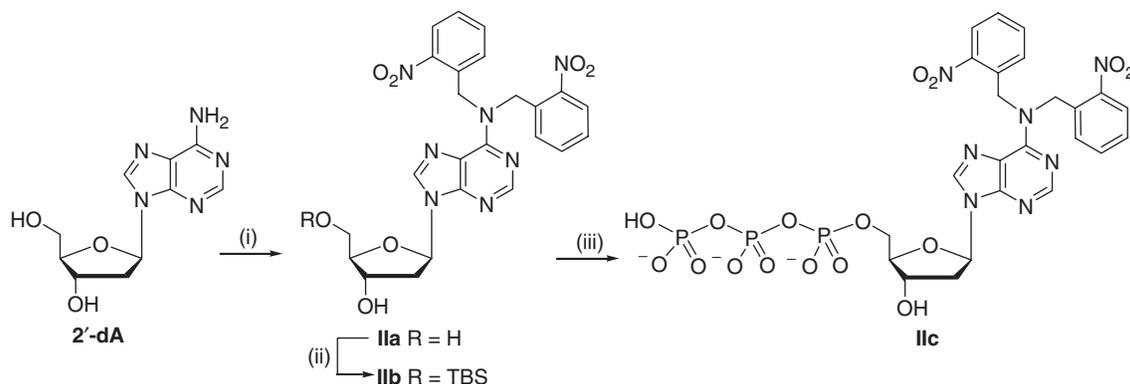
Scheme 2. Synthesis of 3'-O-(2-nitrobenzyl)-2'-deoxyadenosine analogs **Ia-Ic**. (i) TBSCl, imidazole, DMF, room temperature, overnight; Boc_2O , DMAP, DMF, room temperature, overnight, 83%; (ii) $n\text{-Bu}_4\text{NF}$, THF, 0°C , then gradually warmed to room temperature, 96%; (iii) TBSCl, imidazole, DMF, 83%; (iv) $n\text{-Bu}_4\text{NOH}$, NaI, NaOH, $\text{CH}_2\text{Cl}_2/\text{H}_2\text{O}$, 2-nitrobenzyl bromide, room temperature, 91%; (v) SiO_2 , high vacuum, $70\text{--}80^\circ\text{C}$, 24 h, 91%; (vi) $n\text{-Bu}_4\text{NF}$, THF, 0°C , then gradually warmed to room temperature, 23%; (vii) POCl_3 , $(\text{MeO})_3\text{PO}$, minus 20°C ; $(n\text{-Bu}_3\text{NH})_2\text{H}_2\text{P}_2\text{O}_7$, $n\text{-Bu}_3\text{N}$, DMF; 1 M $\text{HNEt}_3\text{HCO}_3$, 31%.

a second round of RP-HPLC purification was performed. The dATP contamination could, however, be substantially reduced using the Mop-Up assay (23). Next, we examined the rates of UV deprotection for 5'-O-TBS analogs **Ib** and **IIIb**, conducted at wavelengths of 302 and 365 nm. Compound **Ib** exhibited the expected first-order profile with deprotection half-life times ($\text{DT}_{1/2}$) of 60 and 152 s at 302 and 365 nm, respectively. The deprotection rate for compound **IIIb** was approximately 3-fold faster than that of compound **Ib** at both deprotection wavelengths (Figure 2B). UV deprotection of compound **IIIb**, however, revealed a transient intermediate before the appearance of the 5'-O-TBS-2'-deoxyadenosine product (Figure 2C). We suspected, and later confirmed, the identity of the intermediate to be the mono N^6 -(2-nitrobenzyl)-2'-deoxyadenosine analog.

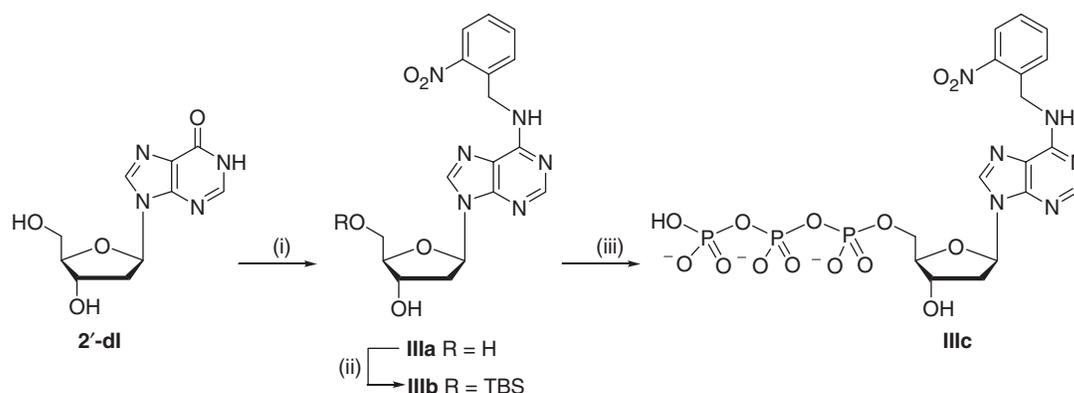
Following synthesis of N^6 -(2-nitrobenzyl)-dATP **IIIc**, we examined its incorporation using *Bst* DNA polymerase, which showed efficient, base-specific termination of DNA synthesis at a final concentration of 200 nM (Figure 2A).

The 2-nitrobenzyl group was efficiently removed from the DNA duplex using a custom built UV deprotector (Supplementary Figure 1), evidenced by the band shift of the extended dye-primer to the termination position of the first thymidine of the oligonucleotide template. Examination of the UV deprotection data for compound **IIIb** revealed a first-order reaction with $\text{DT}_{1/2}$ values of 46 and 144 s at 302 and 365 nm, respectively (Figure 2B). The UV deprotection data suggest that the attachment of a single 2-nitrobenzyl group to either the 3'-O or the N^6 -aromatic amine position does not significantly alter the rate of reaction for UV light-induced cleavage.

These data help to explain the structure misassignment and positive incorporation data presented in the Metzker *et al.* paper (8). Alkylation of adenosine with 2-nitrobenzyl bromide using NaH in DMF has been reported to occur on either the 2'- or 3'-hydroxyl group of the ribose ring (24), but not on the exo-cyclic amino group of the adenine base. Alkylation of 2'-deoxyadenosine under the same conditions, however, exclusively gave *bis*- N^6 , N^6 -alkylated



Scheme 3. Synthesis of N^6,N^6 -bis(2-nitrobenzyl)-2'-deoxyadenosine analogs **IIa–IIIc**. (i) NaH, 2-nitrobenzyl bromide, DMF, 0°C, 22%; (ii) TBSCl, imidazole, DMF, 57%; (iii) POCl₃, (MeO)₃PO, minus 20°C; (*n*-Bu₃NH)₂H₂P₂O₇, *n*-Bu₃N, DMF; 1 M HNET₃HCO₃, yield not determined.



Scheme 4. Synthesis of N^6 -(2-nitrobenzyl)-2'-deoxyadenosine analogs **IIIa–IIIc**. (i) BOP, DIPEA, 2-nitrobenzyl amine, DMF, room temperature, 20 h, then 50°C 3 h, 98%; (ii) TBSCl, imidazole, DMF, 63%; (iii) POCl₃, (MeO)₃PO, minus 20°C; (*n*-Bu₃NH)₂H₂P₂O₇, *n*-Bu₃N, DMF; 1 M HNET₃HCO₃, 60%.

compound **IIa**, albeit in lower yield. Based on the data presented here, we now conclude that the structure of the alkylation product reported by Metzker *et al.* (8) is the *bis*- N^6,N^6 -(2-nitrobenzyl)-2'-deoxyadenosine analog (Scheme 1). The reported termination of the corresponding triphosphate, occurring at a concentration of 250 μ M [Figure 5 in Metzker *et al.* (8)], is most likely the result of the contamination of minute quantities of triphosphate **IIIc**, derived from triphosphate **IIc** during its handling under ordinary laboratory light conditions. One of the minor termination bands observed for compound **IIc** incorporation with *Bst* DNA polymerase (Figure 2A, 100 μ M lane) reveals the presence of **IIIc** triphosphate. Our data in Figure 2C further support our supposition that during UV light-directed deprotection, compound **IIb** is transformed into intermediate **IIIb** before undergoing loss of the second 2-nitrobenzyl group to yield the natural nucleotide (Figure 2D). From this investigation, we have discovered that the N^6 -(2-nitrobenzyl)-dATP **IIIc** is the active species of the three triphosphates examined here.

Compound **IIIc** is active with a variety of DNA polymerases

Numerous groups have employed qualitative, Sanger-based assays to estimate incorporation efficiencies,

although these methods are not feasible for assaying modified analogs in the absence of natural nucleotides (8,25,26). This led us to develop a quantitative, PEP assay, which could be utilized for high-throughput screening of modified nucleotides against commercially available DNA polymerases. The PEP assay is designed with a polymerase concentration in excess of the primer/template complex, thereby limiting the reaction to nucleotide binding and coupling steps. The desired nucleotide is then titrated across the appropriate concentration range to observe extension of a dye-primer by gel electrophoresis. The end-point concentration, or IC₅₀ value, is the point at which the number of moles of substrate equals that of the product. The primer/template complex concentration (2.5 nM) defines the lower limit of the IC₅₀ value for nucleotide titrations to 1.25 nM (i.e. [primer] = [primer plus incorporated nucleotide]). The number of activity units for eight commercially available DNA polymerases was determined by titration with dATP (concentration range from 0.1 to 100 nM), with the goal of reaching the PEP IC₅₀ limit of 1.25 nM. In general, increasing the number of units reduced the IC₅₀ value for dATP towards this limit, the exceptions being Terminator and Terminator II (see Supplementary Data for polymerase definitions).

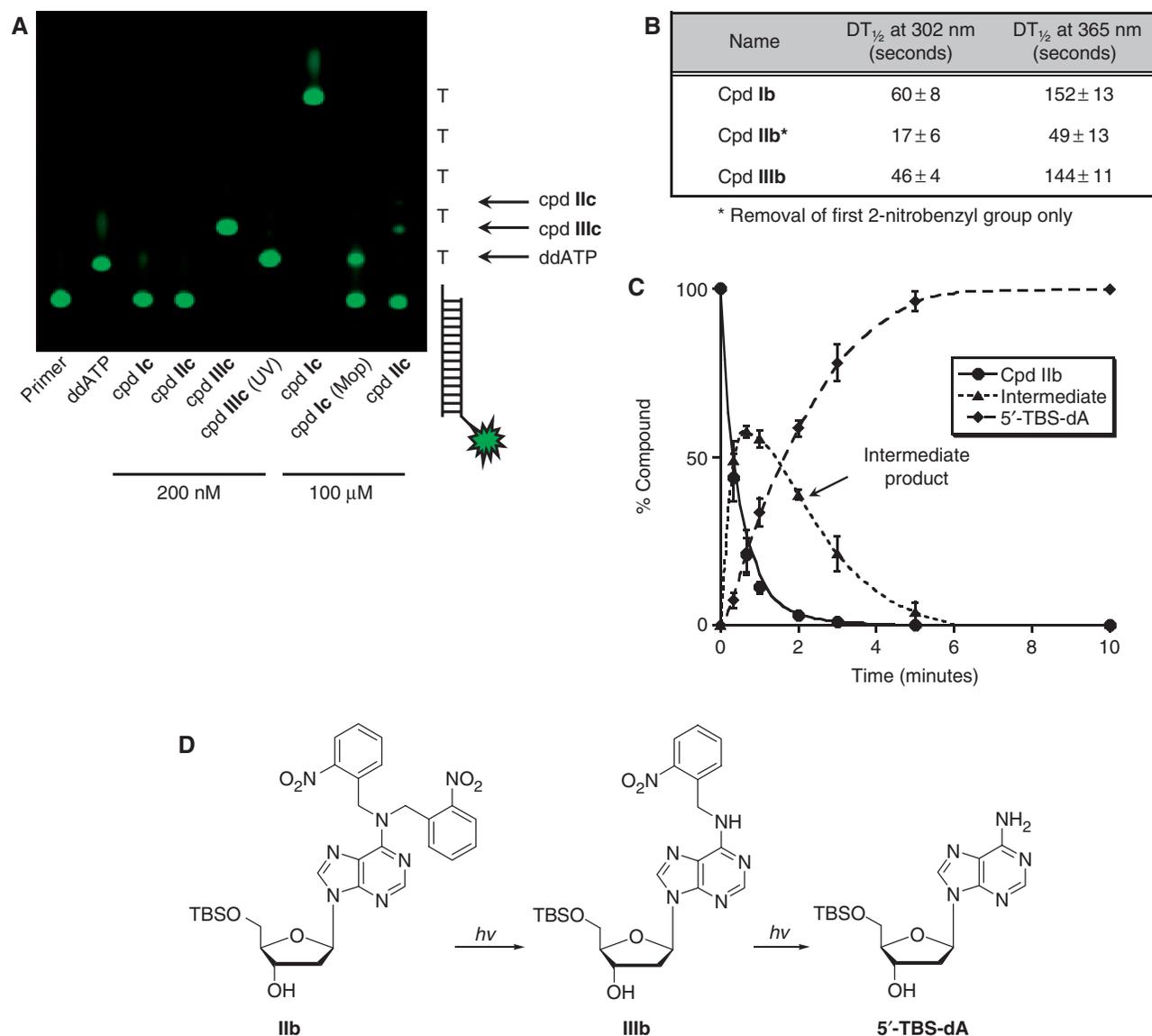


Figure 2. Identification of *N*⁶-(2-nitrobenzyl)-2'-deoxyadenosine triphosphate **IIIc** as the active analogue. (A) OligoTemplate assay for compounds **Ic**, **IIc** and **IIIc**, analyzed on a 377 DNA sequencer. The final concentration for ddATP was 50 μM. (B) Summary of the UV deprotection experiments for compounds **Ib**, **IIb** and **IIIb**. (C) Time plot of the UV light-induced deprotection of compound **IIb** at 302 nm. (D) Proposed stepwise photocleavage of *N*⁶,*N*⁶-bis-(2-nitrobenzyl)-2'-deoxyadenosine **IIb**.

For these enzymes, increasing the number of units yielded an increase in IC₅₀ values for dATP. This observation was not investigated further. In these cases, the number of units used for subsequent PEP assays were those yielding the lowest IC₅₀ value for dATP (Supplementary Table 1, highlighted blue boxes).

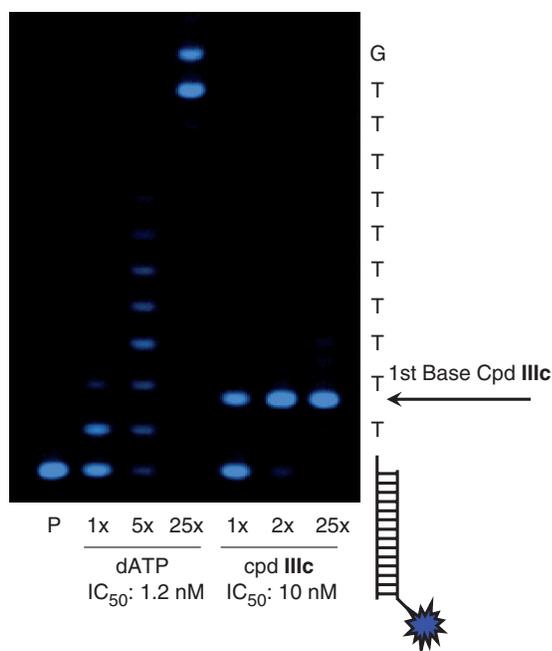
*N*⁶-(2-Nitrobenzyl)-dATP **IIIc** was tested for base-specific incorporation using eight polymerases with the PEP assay and compared with assay data for dATP and ddATP. Compound **IIIc** was incorporated by all polymerases examined, with IC₅₀ values ranging from 2.1 nM to 2.1 μM (Table 1). Five of the eight polymerases revealed a less than 4-fold preference for dATP over compound **IIIc**, with Terminator and Vent(exo-) showing the least bias of 1:1.3, providing evidence that compound **IIIc** is

incorporated almost as efficiently as dATP itself. In all cases except AmpliTaqFS, compound **IIIc** was preferred over ddATP with an incorporation bias range of 3.8 × 10⁻³:1 to 0.32:1, respectively. The F667Y mutation in AmpliTaqFS has been shown to prefer ddATP over dATP, with a ratio of 0.59:1 (25), which is in good agreement with the ratio of 0.62:1 reported in Table 1.

Next, we examined the ability of *N*⁶-(2-nitrobenzyl)-dATP **IIIc** to terminate DNA synthesis using an oligonucleotide template containing a stretch of 10 thymidine bases. As expected, *Bst* DNA polymerase extended the growing primer utilizing dATP as a substrate in a concentration-dependent manner. At a concentration of 25× its IC₅₀ value, the enzyme completely extended the 10 thymidine template and partially misincorporated

Table 1. PEP assay results for dATP, compound **IIIc** and ddATP

DNA polymerase	Average IC ₅₀ ± 1 SD			Incorporation bias		
	dATP	Cpd IIIc	ddATP	Cpd IIIc /dATP	Cpd IIIc /ddATP	ddATP/dATP
<i>Bst</i>	1.2 ± 0.1 nM	21 ± 3 nM	0.37 ± 0.03 μM	18	0.057	310
Klenow(exo-)	1.6 ± 0.1 nM	4.3 ± 0.2 nM	29 ± 5 nM	2.7	0.15	18
<i>Taq</i>	5.5 ± 0.5 nM	2.1 ± 0.2 μM	12.6 ± 0.9 μM	380	0.17	2300
Ampli <i>Taq</i> FS	5.3 ± 0.1 nM	0.89 ± 0.06 μM	3.3 ± 0.1 nM	170	270	0.62
Therminator	2.3 ± 0.3 nM	3.1 ± 0.4 nM	9.7 ± 1.1 nM	1.3	0.32	4.2
Therminator II	4.4 ± 0.6 nM	7.8 ± 0.7 nM	0.23 ± 0.03 μM	1.8	0.034	52
Vent(exo-)	1.6 ± 0.2 nM	2.1 ± 0.2 nM	0.55 ± 0.04 μM	1.3	0.0038	340
DeepVent(exo-)	2.8 ± 0.2 nM	11.0 ± 0.6 nM	3.4 ± 0.4 μM	3.9	0.0032	1200

**Figure 3.** PEP termination assay comparing dATP with compound **IIIc** using *Bst* DNA polymerase.

dATP against a ‘G’ template base at the 11th base position (Figure 3). In contrast, the *N*⁶-(2-nitrobenzyl)-dATP **IIIc** efficiently incorporated and terminated *Bst* DNA synthesis at the first-base position up to 25× its IC₅₀ value. The difference in electrophoretic mobility of the first-base product for dATP and that of compound **IIIc** is due to the *N*⁶-attached 2-nitrobenzyl group.

IC₅₀ values for compound **IIIc** were retitrated using the poly(dT) template. In some cases, IC₅₀ values differed from those in Table 1, reflecting DNA sequence context effects. PEP termination assays were performed for the eight polymerases at 1×, 5× and 25× the IC₅₀ values (Table 2). Three of the four Family A DNA polymerases resulted in efficient termination of DNA synthesis at the first-base position using compound **IIIc**, while all Family B DNA polymerases showed significant, but variable, levels of second-base product. Therminator provided the most extreme example with ~98.5% of the growing primer extended as a second-base product. The majority

Table 2. PEP termination results for compound **IIIc**

Polymerase	Adjusted IC ₅₀ value (nM)	25× IC ₅₀		
		% Primer	% 1st Base	% 2nd Base
<i>Bst</i>	10	0.9 ± 0.3	99.1 ± 0.3	<0.1
Klenow(exo-)	4.3	17.7 ± 2.2	76.1 ± 1.8	6.2 ± 0.5
<i>Taq</i>	300	1.3 ± 0.4	98.7 ± 0.4	<0.1
Ampli <i>Taq</i> FS	350	4.8 ± 0.9	95.2 ± 0.9	<0.1
Therminator	1.8	0.2 ± 0.4	1.3 ± 1.0	98.5 ± 1.3
Therminator II	30	1.1 ± 0.1	85.9 ± 0.9	13.0 ± 1.0
Vent(exo-)	2.1	0.8 ± 0.3	68.3 ± 1.6	30.9 ± 1.5
DeepVent(exo-)	11	1.5 ± 0.7	84.4 ± 0.4	14.1 ± 1.0

of polymerases extended the primer with an efficiency of ~99%. Based upon the desired properties of termination at the first-base position and efficient primer extension, *Bst* emerged as a promising CRT polymerase in combination with compound **IIIc**.

Compound **IIIc** shows good mismatch discrimination with *Bst* DNA polymerase

PEP discrimination assays were then performed to evaluate the specificity of *N*⁶-(2-nitrobenzyl)-dATP **IIIc** against mismatched template bases (i.e. A, C or G). Comparing IC₅₀ values of mismatched versus matched bases for compound **IIIc** revealed nucleotide discrimination of greater than two orders of magnitude (Figure 4A). Surprisingly, the cytosine–adenine mismatch revealed the highest discrimination ratio of 1100-fold over the complement thymidine base, and only slightly less than that for dATP (Supplementary Table 2). The ddATP analog did not show mismatch incorporation at a final concentration of 500 μM (data not shown). These data suggest that compound **IIIc** incorporates as a base-specific terminator and reveals nucleotide characteristics more similar to dATP than those of ddATP, which we attribute to the presence of the 3'-OH group.

The *N*⁶-proton of 2-nitrobenzyl-adenine base, still capable of base pairing, may aid in the specificity observed in Figure 4A. To examine this further, we performed *ab initio* calculations for a thymine–*N*⁶-(2-nitrobenzyl)-adenine base pair, using the Hartree–Fock method coupled with the 6-31G* atomic shell set (17). The optimal

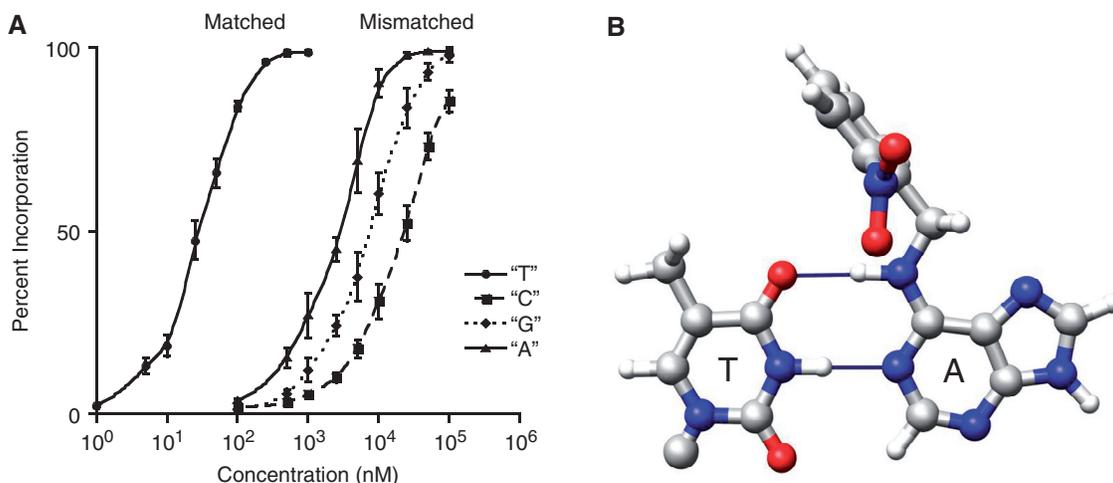


Figure 4. (A) PEP discrimination curves for compound **IIIc**. (B) Proposed thymine– N^6 -(2-nitrobenzyl)-adenine base pair.

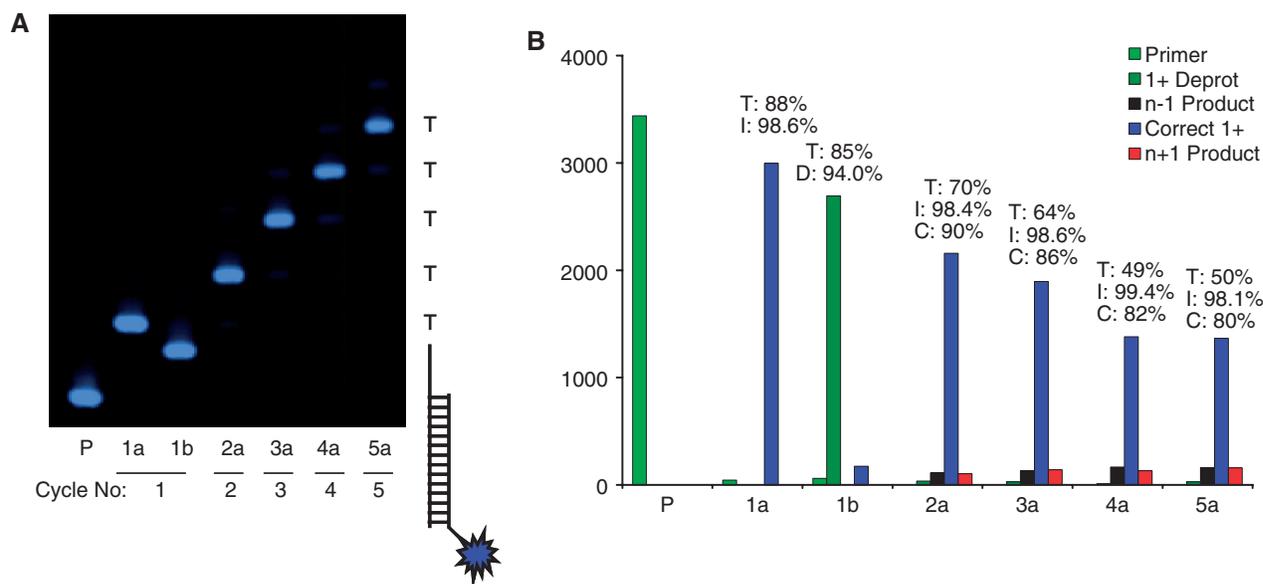


Figure 5. (A) Five-base CRT sequencing with N^6 -(2-nitrobenzyl)-dATP **IIIc**. Lanes: 'P' represents the dye-primer, '1a–5a' represents final incorporations of compound **IIIc** at different cycles, and '1b' represents the first base addition followed by UV deprotection. (B) Histogram plot of the gel image in Figure 5A represents the signal intensities of band products. 'T' is the percentage of total signal to that of the dye-primer signal (lane P), 'I' is the incorporation efficiency, 'D' is the deprotection efficiency and 'C' is the percentage of correct +1 product to that of the total signal.

molecular structure was a nitro group positioned 45° and a 2-nitrobenzyl group positioned 80° counterclockwise, relative to the aromatic amino group (Figure 4B). Hydrogen bond distances were determined to be 2.93 Å ($N^1 \dots H-N^3$) and 2.97 Å ($N^6-H \dots O^4$), which are longer than those reported by Watson and Crick (16). A ΔE of -2.68 kcal/mol for the modified base pair, however, suggests that hydrogen bonding is favorable. Active site tightness (27,28) of compound **IIIc**, involving hydrophobic interactions of the 2-nitrobenzyl with key amino acids, may also contribute to the observed enzymatic properties. Upon incorporation, these interactions may also be involved with misalignment of the 3'-OH group, preventing nucleophilic attack by the incoming nucleotide, thus terminating DNA synthesis.

Five-base CRT sequencing with compound **IIIc**

To test N^6 -(2-nitrobenzyl)-dATP **IIIc** as a RT in CRT sequencing, a five-base experiment was performed using a biotinylated template containing a poly(dT) stretch attached to streptavidin-coated magnetic beads (Figure 5A). For the first cycle, incorporation 'a' and deprotection 'b' products are shown in the gel, with subsequent cycles showing only incorporation products. The data illustrate an advantage over the pyrosequencing method (29), namely the stepwise addition through a homopolymer repeat. The gel image in Figure 5A was analyzed further by quantitating the fluorescent bands at different CRT cycles (Figure 5B). During the first cycle, the product of incorporation efficiency (I: 98.6%) and deprotection efficiency (D: 94.0%) resulted in a cycle

efficiency (C_{eff}) of 92.7% on a solid support. The estimated signal 'S', calculated from the equation $[S = (C_{\text{eff}})^{\text{RL}}]$ (5), is 68.4% for the five-base read. At cycle five, the signal for the correct +1 product is 40% (i.e. $T \times C$), the difference of which we attribute to photobleaching of the dye-primer with increasing exposure to the UV light (data not shown). Signal loss is also due to dephasing, observed as $n - 1$ (incomplete deprotection) and $n + 1$ products (natural nucleotide carryover from the previous cycle). While ongoing efforts are focused on reducing dephasing products, $\geq 80\%$ of the total signal is derived from the correct +1 product, with base-calling easily performed from the primary data in Figure 5A and B.

DISCUSSION

We have discovered that the attachment of a small, photocleavable 2-nitrobenzyl group to the N^6 -position of 2'-deoxyadenosine results in this triphosphate acting as a RT 'without blocking the 3'-end'. The novel RT, N^6 -(2-nitrobenzyl)-dATP **IIIc**, provides favorable enzymatic properties with a variety of wild-type and mutant DNA polymerases. This is unlike the situation for 3'-modified nucleotides, which typically act as poor substrates for DNA polymerases. For example, screening 3'-*O*-allyl-dATP with eight different DNA polymerases revealed limited activity at high micromolar concentrations with only Vent(exo-) DNA polymerase (8). The highly related 9°N(exo-) DNA polymerase (30), containing the A485L and Y409V amino acid variants (Therminator II), has been shown to incorporate 3'-*O*-allyl-dNTPs, however, efficient incorporation requires up to 50 min per single base addition, highlighting the difficulty of incorporating these analogs (7,10). The A485L and Y409V mutations are analogous to those described for Vent(exo-) DNA polymerase (26), with the Y409V residue acting as a 'steric' gate for incorporation of ribonucleotides (26,31–33). Little is known regarding the mechanism by which a 2'-steric gate residue alters the incorporation of 3'-*O*-allyl terminators.

Research efforts have been focused on optimizing the cycle efficiency and time, which determine read-length and throughput, respectively. Targeting a 50% loss in signal as an end-point (5), the cycle efficiency must be $\geq 97.3\%$ to achieve a 25 base read-length. Although we show a slightly smaller cycle efficiency of 92.7% with compound **IIIc**, primarily due to its deprotection efficiency of 94%, work is ongoing to improve this by substitution of the 2-nitrobenzyl group. We also anticipate further improvements in cycle efficiency with development of instrumentation supporting the CRT chemistry, allowing for integration and automation of the incorporation, imaging, deprotection and washing steps.

Although the cycle efficiency is primarily the product of incorporation and deprotection efficiencies of the RT, other factors including dephasing (natural nucleotide carryover) and accumulating 'molecular scars' (residual linker structures left over after deprotection) can also influence the efficiency in an adverse manner. The N^6 -attachment to the adenine nucleobase is also unique,

differing from that of the traditional 7-deaza position of BigDye terminators (34) and 3'-*O*-allyl terminators (7,10). Upon chemical deprotection of the 3'-*O*-allyl terminators, a residual propargyl amino group remains on the nucleobase, resulting in an accumulating molecular scar with subsequent CRT cycles. The primary advantage of N^6 -alkylation is that, upon directed photocleavage with 365 nm UV light, the modified nucleotide is transformed back into its natural state without molecular scarring (Figure 2D). The enhanced enzymatic properties and the N^6 -cleavage site, transforming the efficiently incorporated RT back into natural DNA, are anticipated to improve the cycle efficiency and read-length of the CRT method.

These observations suggest that the N^6 -(2-nitrobenzyl)-dATP **IIIc** represents an ideal candidate as a RT for CRT sequencing, illustrated by stepwise, single base addition through a homopolymer repeat. Of the polymerases tested here, however, not all exhibited this property, with the Family B polymerases revealing incorporation of a second modified nucleotide, albeit at varying levels. Recently, we have discovered that substitution of the 2-nitrobenzyl group can also 'tune' the termination properties of these polymerases to give exclusively single-base products (unpublished data). To exploit the application of 3'-unblocked RTs in CRT sequencing, efforts are underway to create fluorescently labeled analogs of compound **IIIc**, and extrapolate this nucleotide model to the remaining nucleobases for production of a novel, four-color RT set. We note that the adenine example as a RT may not be directly applicable to the remaining nucleobases, which present their own unique challenges. Nonetheless, a base modification strategy can still be employed with careful selection of the attachment site of the 2-nitrobenzyl group on each nucleobase structure, the triphosphates of which exhibit similar enzymatic properties described in this report and transform into its natural nucleobase structure upon UV deprotection (manuscript in preparation).

We anticipate that 3'-unblocked terminators will have utility beyond the application of CRT sequencing. For example, a complete set of non-fluorescent RTs could be used in pyrosequencing, with the advantage of improving accuracy through homopolymer repeat stretches. Reduced incorporation biases of 3'-unblocked terminators over natural nucleotides, exhibited by several polymerases, may also prove useful for more accurate heterozygote analysis in Sanger sequencing. With other applications envisioned, 3'-unblocked terminators may well find their way into the general arsenal of molecular biology tools used in genomic sciences today.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

National Institutes of Health (R01 HG003573, R41 HG003072, and R43 HG003443). Funding to pay the

Open Access publication charges for this article was provided by R01 HG003573.

Conflict of interest statement. We declare that LaserGen plans on commercializing this compound, along with its derivatives. No other conflicts have been declared.

REFERENCES

- Shendure, J., Mitra, R.D., Varma, C. and Church, G.M. (2004) Advanced sequencing technologies: methods and goals. *Nat. Rev. Genet.*, **5**, 335–344.
- Bai, X., Edwards, J. and Ju, J. (2005) Molecular engineering approaches for DNA sequencing and analysis. *Expert Rev. Mol. Diagn.*, **5**, 797–808.
- Bennett, S.T., Barnes, C., Cox, A., Davies, L. and Brown, C. (2005) Toward the \$1000 human genome. *Pharmacogenomics*, **6**, 373–382.
- Chan, E.Y. (2005) Advances in sequencing technology. *Mutat. Res.*, **573**, 13–40.
- Metzker, M.L. (2005) Emerging technologies in DNA sequencing. *Genome Res.*, **15**, 1767–1776.
- Fan, J.-B., Chee, M.S. and Gunderson, K.L. (2006) Highly parallel genomic assays. *Nat. Rev. Genet.*, **7**, 632–644.
- Ju, J., Kim, D.H., Bi, L., Meng, Q., Bai, X., Li, Z., Li, X., Marma, M.S., Shi, S. *et al.* (2006) Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators. *Proc. Natl Acad. Sci. USA*, **103**, 19635–19640.
- Metzker, M.L., Raghavachari, R., Richards, S., Jacutin, S.E., Civitello, A., Burgess, K. and Gibbs, R.A. (1994) Termination of DNA synthesis by novel 3'-modified deoxyribonucleoside triphosphates. *Nucleic Acids Res.*, **22**, 4259–4267.
- Canard, B. and Sarfati, R. (1994) DNA polymerase fluorescent substrates with reversible 3'-tags. *Gene*, **148**, 1–6.
- Ruparel, H., Bi, L., Li, Z., Bai, X., Kim, D.H., Turro, N.J. and Ju, J. (2005) Design and synthesis of a 3'-O-allyl photocleavable fluorescent nucleotide as a reversible terminator for DNA sequencing by synthesis. *Proc. Natl Acad. Sci. USA*, **102**, 5932–5937.
- Barnes, C., Balasubramanian, S., Liu, X., Swerdlow, H. and Milton, J. (2006) US patent 7,057,206 B2.
- Corrie, J.E.T. (2005) Photoremovable protecting groups used for caging of biomolecules. In Goeldner, M. and Givens, R. (eds), *Dynamics Studies in Biology*. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, pp. 1–28.
- Metzker, M.L., Lu, J. and Gibbs, R.A. (1996) Electrophoretically uniform fluorescent dyes for automated DNA sequencing. *Science*, **271**, 1420–1422.
- Innis, M.A., Myambo, K.B., Gelfand, D.H. and Brow, M.A.D. (1988) DNA sequencing with *Thermus aquaticus* DNA polymerase and direct sequencing of polymerase chain reaction-amplified DNA. *Proc. Natl Acad. Sci. USA*, **85**, 9436–9440.
- Kong, H., Kucera, R.B. and Jack, W.E. (1993) Characterization of a DNA polymerase from the hyperthermophile archaea *Thermococcus litoralis*. *J. Biol. Chem.*, **268**, 1965–1975.
- Watson, J.D. and Crick, F.H. (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, **171**, 737–738.
- Schmidt, M.W., Baldrige, K.K., Boatz, J.A., Elbert, S.T., Gordon, M.S., Jensen, J.H., Koseki, S., Matsunaga, N., Nguyen, K.A. *et al.* (1993) General atomic and molecular electronic structure system. *J. Comput. Chem.*, **14**, 1347–1363.
- Sponer, J., Leszczynski, J. and Hobza, P. (1996) Structures and energies of hydrogen-bonded DNA base pairs. A nonempirical study with inclusion of electron correlation. *J. Phys. Chem.*, **100**, 1965–1974.
- Sponer, J., Jurečka, P. and Hobza, P. (2004) Accurate interaction energies of hydrogen-bonded nucleic acid base pairs. *J. Am. Chem. Soc.*, **126**, 10142–10151.
- Furrer, E. and Giese, B. (2003) On the distance-independent hole transfer over long (A•T)_n-sequences in DNA. *Helv. Chim. Acta*, **86**, 3623–3632.
- Ludwig, J. (1981) A new route to nucleoside 5'-triphosphates. *Acta Biochem. Biophys. Acad. Sci. Hung.*, **16**, 131–133.
- Wan, Z.K., Binnun, E., Wilson, D.P. and Lee, J. (2005) A highly facile and efficient one-step synthesis of N⁶-adenosine and N⁶-2'-deoxyadenosine derivatives. *Org. Lett.*, **7**, 5877–5880.
- Metzker, M.L., Raghavachari, R., Burgess, K. and Gibbs, R.A. (1998) Elimination of residual natural nucleotides from 3'-O-modified-dNTP syntheses by enzymatic Mop-Up. *BioTechniques*, **25**, 814–817.
- Ohtsuka, E., Tanaka, S. and Ikehara, M. (1977) Studies on transfer ribonucleic acids and related compounds. XVI. Synthesis of ribooligonucleotides using a photosensitive o-nitrobenzyl protection for the 2'-hydroxyl group. *Chem. Pharm. Bull.*, **25**, 949–959.
- Tabor, S. and Richardson, C.C. (1995) A single residue in DNA polymerases of the *Escherichia coli* DNA polymerase I family is critical for distinguishing between deoxy- and dideoxyribonucleotides. *Proc. Natl Acad. Sci. USA*, **92**, 6339–6343.
- Gardner, A.F. and Jack, W.E. (1999) Determinants of nucleotide sugar recognition in an archaeon DNA polymerase. *Nucleic Acids Res.*, **27**, 2545–2553.
- Kool, E.T. (2002) Active site tightness and substrate fit in DNA replication. *Annu. Rev. Biochem.*, **71**, 191–219.
- Henry, A.A. and Romesberg, F.E. (2003) Beyond A, C, G and T: augmenting nature's alphabet. *Curr. Opin. Chem. Biol.*, **7**, 727–733.
- Ronaghi, M., Uhlen, M. and Nyren, P. (1998) A sequencing method based on real-time pyrophosphate. *Science*, **281**, 363–365.
- Southworth, M.W., Kong, H., Kucera, R.B., Jannasch, H.W. and Perler, F.B. (1996) Cloning of thermostable DNA polymerases from hyperthermophilic marine Archaea with emphasis on *Thermococcus* sp. 9 degrees N-7 and mutations affecting 3'-5' exonuclease activity. *Proc. Natl Acad. Sci. USA*, **93**, 5281–5285.
- Gao, G., Orlova, M., Georgiadis, M.M., Hendrickson, W.A. and Goff, S.P. (1997) Conferring RNA polymerase activity to a DNA polymerase: a single residue in reverse transcriptase controls substrate selection. *Proc. Natl Acad. Sci. USA*, **94**, 407–411.
- Astatke, M., Ng, K., Grindley, N.D.F. and Joyce, C.M. (1998) A single side chain prevents *Escherichia coli* DNA polymerase I (Klenow fragment) from incorporating ribonucleotides. *Proc. Natl Acad. Sci. USA*, **85**, 3402–3407.
- Fa, M., Radeghieri, A., Henry, A.A. and Romesberg, F.E. (2004) Expanding the substrate repertoire of a DNA polymerase by directed evolution. *J. Am. Chem. Soc.*, **126**, 1748–1754.
- Rosenblum, B., Lee, L., Spurgeon, S., Khan, S., Menchen, S., Heiner, C. and Chen, S. (1997) New dye-labeled terminators for improved DNA sequencing patterns. *Nucleic Acids Res.*, **25**, 4500–4504.