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

Weidong Wu\(^1\), Brian P. Stupi\(^1\), Vladislav A. Litosh\(^1\), Dena Mansouri\(^2\), Demetra Farley\(^3\), Sidney Morris\(^1,3\), Sherry Metzker\(^1\) and Michael L. Metzker\(^1,2,3,\ast\)

\(^1\)LaserGen, Inc., Houston, TX 77054, \(^2\)Department of Molecular & Human Genetics and \(^3\)Human 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 sequencing 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 \( N^6-(2\text{-nitrobenzyl})\)-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 \( N^6-(2\text{-nitrobenzyl})\)-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 \( 3'-O-(2\text{-nitrobenzyl})\)-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 \( 3'-O-(2\text{-nitrobenzyl})\)-2'-deoxyadenosine \( \text{Ia} \) [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 \( \text{NaH} \) in DMF is \( N^6,N^6\text{-bis-(2-nitrobenzyl)}\)-2'-deoxyadenosine \( \text{IIa} \) (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 \( 3'-O-(2\text{-nitrobenzyl})\)-2'-deoxyadenosine analogs \( \text{Ia–Ic} \) and \( N^6,N^6\text{-bis-(2-nitrobenzyl)}\)-2'-deoxyadenosine analogs \( \text{IIa–IIc} \) (Figure 1), both of which \( \text{Ic} \) and \( \text{IIc} \) triphosphates proved inactive.
by DNA polymerase incorporation assays. This finding led to the synthesis and characterization of a third set of N<sup>6</sup>-(2-nitrobenzyl)-20-deoxyadenosine analogs III<sub>a</sub>–III<sub>c</sub> (Figure 1), of which compound III<sub>b</sub> was originally identified by us as an intermediate during the ultraviolet (UV) light-induced deprotection of compound II<sub>b</sub> to its corresponding parent analog. Here, we report that the active triphosphate described by Metzker et al. (8) is actually the N<sup>6</sup>-(2-nitrobenzyl)-dATP III<sub>c</sub>, representing a novel, 30-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 AmpliTaqFS being purchased from Applied Biosystems (AB). Analytical silica gel 60 F<sub>254</sub> 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, IIb and IIIb**

5'-O-tert-butylidimethylsilyl (TBS) derivatives Ib, IIb 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<sub>1/2</sub>) 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 1c, 1l and 1lc were tested for incorporation with Bst DNA polymerase at concentrations of 200 nM and 100 μM using the BODIPY-R6G labeled primer-1 (5'-TTGTAAAACGACGGCCAGT) (13) and oligoTemplate-1 (5'-TACGGAGCCTTTTTACTGGC GTCGTTTTACA, interrogation base is underlined and bolded) complex. Reactions were quenched with 10 μl 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× 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'-TACGGAGCCTTTTTACTGGCC GTCGTTTTACA, interrogation base is underlined and bolded) in 1× 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 μl) 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] = 5 nM, [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 μl of stop solution (98% deionized formamide; 10 mM Na₂EDTA, pH 8.0; 25 mg/ml Blue Dextran, MW 200000). 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 are displayed as a linear–log plot of product formation versus time. Deprotection assays were performed in triplicate to calculate the average DT₁/₂ ± 1 SD.

Five-base sequencing experiment using compound 1lc as reversible terminator

An 80 nM solution of oligoTemplate-3 in 1 M NaCl and 1× ThermoPol buffer (final volume: 12.5 μl) was incubated for 15 min at room temperature with 5 μl of streptavidin-coated M-270 magnetic Dynabeads (Invitrogen), which had been previously washed three times with 5 μl 1× ThermoPol buffer. The oligoTemplate-3 bound beads were then washed an additional three times with 5 μl 1× ThermoPol buffer and annealed with 5 μl 20 nM BODIPY-FL labeled primer-1 (5'-GGCGGCGCGCG GCGTTAAAAACGAGCGGCGGT) in 1× 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 μl 1× 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 1lc in 1× ThermoPol buffer (reaction volume: 20 μl) at 65 °C for 6 min, then placed on ice. Compound 1lc-incorporated beads were washed four times with 50 μl W10 washing solution (10 mM Tris–HCl, pH 8.0; 10 mM Na₂EDTA; 0.1% Triton X-100), then once with 20 μl W10 washing solution.
Deprotection step. The beads were resuspended in 20 μl deprotection solution (20% aqueous deionized formamide; 10 mM Na2EDTA, pH 8.0; 16.6 mg/ml Blue Dextran, MW 2000000), exposed to 365 nm light for 9 min (i.e. 3×3 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 μl l× apyrase buffer (100 U/l apyrase in l× ThermoPol), three times with 50 μl W10 washing solution, twice with 50 μl l× ThermoPol buffer and then once with 20 μl l× ThermoPol buffer.

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

Ab initio calculations

Adenine–thymine and N6-(2-nitrobenzyl)-adenine–thymine base pairs were created by rotating the 2-nitrobenzyl group, pivoted on the N1-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 N6-(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 I, 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-N6, N6-(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-d6), 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.

N6-(2-Nitrobenzyl)-2′-deoxyadenosine IIIa was prepared based on the work of Wan et al. (22). Treatment of 2′-deoxynosine with 2-nitrobenzylamine in the presence of 1H-benzotriazol-1-ylxylo-iritis(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 μM), although compound Ic did exhibit natural nucleotide contamination, even after
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 IIb, conducted at wavelengths of 302 and 365 nm. Compound Ib exhibited the expected first-order profile with deprotection half-life times (DT 1/2) of 60 and 152 s at 302 and 365 nm, respectively. The deprotection rate for compound IIb was approximately 3-fold faster than that of compound Ib at both deprotection wavelengths (Figure 2B). UV deprotection of compound IIb, 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 N6-(2-nitrobenzyl)-2′-deoxyadenosine analog.

Following synthesis of N6-(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 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 N6-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-N6,N6-alkylated...
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 \( \mu \text{M} \) [Figure 5 in Metzker et al. (8)], is most likely the result of the contamination of minute quantities of triphosphate IIc, 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 \( \mu \text{M} \) 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 \( N^6-(2\text{-nitrobenzyl})\)-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 IC\(_{50}\) 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 IC\(_{50}\) 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 IC\(_{50}\) limit of 1.25 nM. In general, increasing the number of units reduced the IC\(_{50}\) value for dATP towards this limit, the exceptions being Therminator and Therminator II (see Supplementary Data for polymerase definitions).
For these enzymes, increasing the number of units yielded an increase in IC\textsubscript{50} 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 IC\textsubscript{50} value for dATP (Supplementary Table 1, highlighted blue boxes).

\[
\text{N}^6-(2-\text{Nitrobenzyl})-\text{dATP} \quad \text{IIIc}
\]

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 IC\textsubscript{50} values ranging from 2.1 nM to 2.1 \textmu M (Table 1). Five of the eight polymerases revealed a less than 4-fold preference for dATP over compound IIIc, with Therminator and Vent(exo\textsuperscript{-}) 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 \times 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 \textit{N}^6-(2-nitrobenzyl)-dATP IIIc to terminate DNA synthesis using an oligonucleotide template containing a stretch of 10 thymidine bases. As expected, \textit{Bst} DNA polymerase extended the growing primer utilizing dATP as a substrate in a concentration-dependent manner. At a concentration of 25 \times its IC\textsubscript{50} value, the enzyme completely extended the 10 thymidine template and partially misincorporated
of polypeptides 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 \(\text{IIIc}\).

**Compound \(\text{IIIc}\) shows good mismatch discrimination with \(Bst\) DNA polymerase**

PEP discrimination assays were then performed to evaluate the specificity of \(N^6\)-(2-nitrobenzyl)-dATP \(\text{IIIc}\) against mismatched template bases (i.e. A, C or G). Comparing IC\(_{50}\) values of mismatched versus matched bases for compound \(\text{IIIc}\) 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 comple-
ment 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\,\mu M (data not shown). These data suggest that 
compound \(\text{IIIc}\) 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)- 
adene base pair, using the Hartree–Fock method coupled 
with the 6-31G* atomic shell set (17). The optimal
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 (N1...H–N3) and 2.97 Å (N6–H...O4), 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 N6-(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
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 $9^\text{N}(\text{exo}-)$ 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
REFERENCES


