Nucleosides, Nucleotides and Nucleic Acids, 29:879–895, 2010

ISSN: 1525-7770 print / 1532-2335 online
DOI: 10.1080/15257770.2010.536191

LABELED NUCLEOSIDE TRIPHOSPHATES WITH REVERSIBLY TERMINATING AMINOALKOXYL GROUPS

Daniel Hutter,[1,2] Myong-Jung Kim,[1] Nilesh Karalkar,[1] Nicole A. Leal,[1,2] Fei Chen,[1] Evan Guggenheim,[3] Visa Visalakshi,[3] Jerzy Olejnik,[3] Steven Gordon,[3] and Steven A. Benner[1,2]

[1]*Foundation for Applied Molecular Evolution, Gainesville, Florida, USA*
[2]*Firebird Biomolecular Sciences, LLC, Gainesville, Florida, USA*
[3]*Intelligent Bio-Systems, Waltham, Massachusetts, USA*

□ *Nucleoside triphosphates having a 3′-ONH_2 blocking group have been prepared with and without fluorescent tags on their nucleobases. DNA polymerases were identified that accepted these, adding a single nucleotide to the 3′-end of a primer in a template-directed extension reaction that then stops. Nitrite chemistry was developed to cleave the 3′-ONH_2 group under mild conditions to allow continued primer extension. Extension-cleavage-extension cycles in solution were demonstrated with untagged nucleotides and mixtures of tagged and untagged nucleotides. Multiple extension-cleavage-extension cycles were demonstrated on an Intelligent Bio-Systems Sequencer, showing the potential of the 3′-ONH_2 blocking group in "next generation sequencing."*

Supplemental materials are available for this article. Go to the publisher's online edition of *Nucleosides, Nucleotides* and *Nucleic Acids* to view the free supplemental file.

Keywords triphosphate; aminooxy; mutant polymerases; sequencing technology; fluorescent nucleotide; oligonucleotide microarray; nitrous acid; primer elongation; alpha effect; synthesis; oxime

INTRODUCTION

Sequencing using cyclic reversible termination,[1] also known as "sequencing by synthesis,"[2] is a next generation sequencing architecture that infers the sequence of a template by stepwise primer elongation using

Received 8 April 2010; accepted 27 October 2010.
The Foundation and Firebird were supported by grants from the NIGMS (2R42HG003668-02A) and NHGRI (R41HG004589), as well as by Nucleic Acids Licensing, LLC.
Supplemental material is available that includes the synthetic procedures and characterization of all synthesized compounds, various primer extension experiments and sequencing cycles in solution using these triphosphates and various polymerases, and more detailed procedures describing the sequencing runs performed on the Intelligent Bio-Systems Instrument.
Address correspondence to Steven A. Benner, Foundation for Applied Molecular Evolution, P.O. Box 13174, Gainesville, FL 32604. E-mail: sbenner@ffame.org

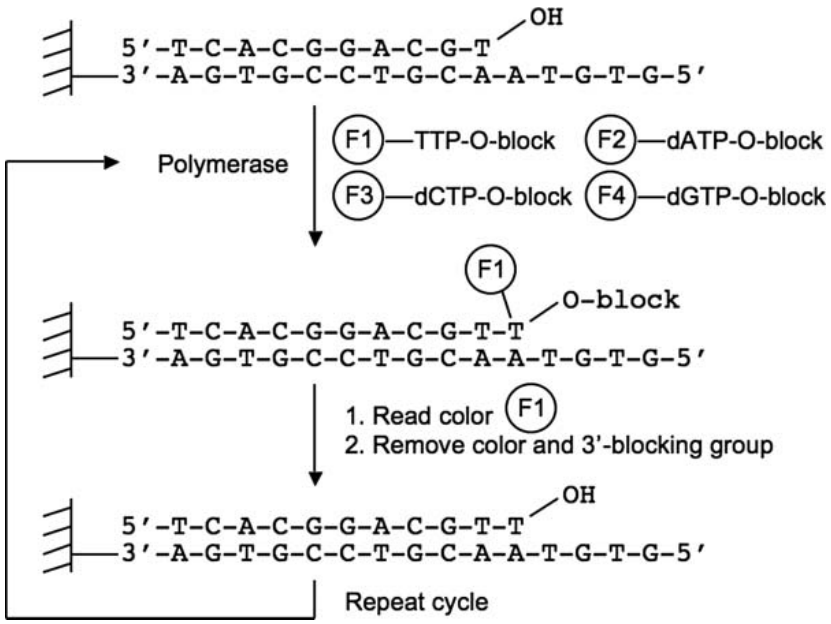


FIGURE 1 Schematic showing sequencing using cyclic reversible termination in the general case. Each triphosphate has its 3′-OH group blocked and a fluorescent tag (F1-F4) attached to its nucleobase via a cleavable linker.

fluorescently tagged triphosphates that have their 3′-OH group blocked (Figure 1). In this architecture, primer extension catalyzed by a DNA polymerase terminates because of the 3′-block, allowing the color of the fluorophore to be read to infer the nucleotide incorporated and to deduce its complement in the template being copied. Then, in subsequent steps that include the removal of the fluorescent tag and cleavage of the 3′-O blocking group, synthesis is resumed through the addition of another tagged nucleotide. The detection and cleavage cycle is then repeated.

Creating triphosphate reagents to deliver these specifications is especially challenging when the capabilities of standard polymerases are considered. Polymerases have evolved for billions of years to examine closely the 3′-ends of an incoming triphosphate, in part to distinguish between ribonucleoside and 2′-deoxyribonucleoside triphosphates that are both present inside of cells. Accordingly, many chemical functionalities appended to the 3′-oxygen of a nucleoside triphosphate cause that triphosphate to be rejected by many polymerases.

Accordingly, a "rule of thumb" was proposed that suggests that any blocking moiety attached to the 3′-OH is preferentially small and hydrophilic, in analogy to the OH group itself. Further, the group must be removable under conditions that neither destroy the duplex nor damage the DNA.

Unfortunately, realities of the periodic table constrain the size of such moieties. The bonds between oxygen and most of the elements in the first and second row of the periodic table are not stable. Carbon is an exception, but attachment of a carbon in its smallest implementation (e.g., as a -CH_3

group) yields a derivative that is too stable for this sequencing architecture; it cannot be removed under any conditions that leave DNA intact.

Accordingly, nearly every 3′-OH blocking group that has been proposed for sequencing using cyclic reversible termination has been larger, and had difficulties managing the compromise between stability and reactivity. The 3′-O acyl blocking groups introduced when the architecture was initially proposed[3] turned out to be too big, too unstable, and possibly cleaved by the polymerases themselves. Accordingly, 3′-O ethers were examined, where atoms are added to the 3′-O-C carbon atom to generate the reactivity needed for cleavage. These include the 3′-O-allyl group,[4, 5] which is cleaved by transition metal catalysis, the 3′-O-methoxymethyl group, which is cleaved by acid, the 3′-O-nitrobenzyl group,[6] which is cleaved by light, and the 3′-O-azidomethylene group,[7, 8] which is cleaved by phosphines. In a creative approach to avoid the 3′-blocking problem entirely, Metzker recently suggested that reversible termination can also be achieved by groups appended not to the 3′-carbon but rather to the nucleobases.[9]

Some time ago, we proposed that nitrogen might be productively used to create a removable 3′-O blocking group.[10] For example, O-alkyl hydroxylamines are stable, hydrophilic species that add just two atoms (a nitrogen and a hydrogen) to the 3′-end of a nucleoside triphosphate. The 3′-ONH_2 group is as small as the 3′-OCH_3 group, and we speculated that it might be small enough to be accepted by polymerases. Indeed, evolutionary analyses gave variants of the commonly used *Taq* DNA polymerase that accept nucleoside triphosphates that have a 3′-ONH_2 blocking group (Figure 2).[11]

Although the native nucleosides with a 3′- or 5′-ONH_2 blocking group have been known for some time,[12,13] they have, to the best of our knowledge, never been transformed into triphosphates or appended with a fluorescent tag. Instead, they were applied as antiviral agents[12] or incorporated into oligonucleotide backbones for antisense applications.[13]

The 5′-triphosphates having a 3′-ONH_2 blocking group that are also fluorescently tagged and useful in this sequencing architecture are challenging synthetic targets. The side chain that holds the fluorescent tag must be reactive, so that the tag can be removed. This reactivity must not, however, be manifested throughout the organic synthesis. The triphosphate group is itself a delicate entity, sensitive to alkaline hydrolysis. Essentially impossible to protect and too polar to be introduced early in a classical chemical synthesis, triphosphate groups are generally added last in a synthetic sequence generating modified nucleosides as substrates for polymerases.

Further, fluorescent tags can themselves be reactive, and frequently confer physical properties on molecules that make them difficult to isolate and purify. Therefore, adding the fluorescent species is also a step that might appropriately be done last in a synthetic sequence.

Of course, the 3′-ONH_2 group is itself a potent nucleophile and a weak base ($pK_a \approx 6$ for the conjugate acid). The reactivity of this group must, therefore, also be managed, and the synthetic strategy must not trigger the

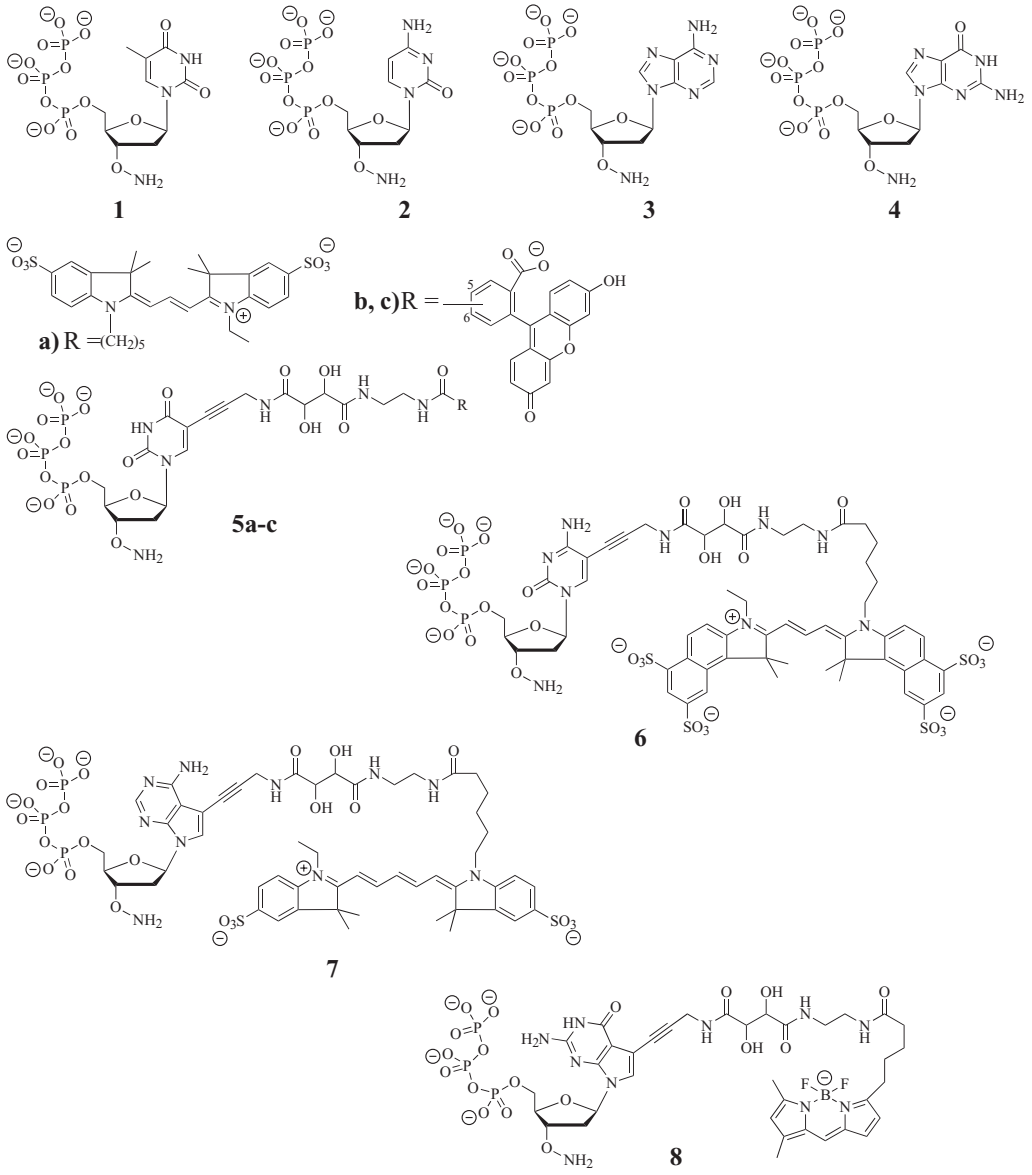


FIGURE 2 Reversible terminators discussed in this manuscript. **1-4** untagged, **5-8** tagged. Fluorophores: 5a) Cy3; 5b) 5-FAM; 5c) 6-FAM; 6) Cy3.5; 7) Cy5; 8) BODIPY-FL-C5.

cleavage reaction that makes it useful in sequence architectures as the other units are added.

To manage the synthesis of a compound that contains all four of these reactive units, we developed a strategy that exploits the special reactivity of each, including reactivity that enables them to be used in sequencing architectures. Once these triphosphates were in hand, we examined polymerases for their ability to accept, as substrates, fluorescently tagged 3′-ONH_2-blocked triphosphates, including those that accept untagged 3′-ONH_2-blocked triphosphates. We then developed a reagent to cleave the

O-NH_2 bond, and demonstrated that this cleavage reagent worked in part of a sequencing cycle.

We report here the process by which these reagents, polymerases, and procedures were developed to support a sequencing cycle. While we recognize that the specific solutions used to address specific problems in this process are (in many aspects) specific for these exact molecules, this process is generally illustrative of the co-development of three parts of a tool in nucleic acid analysis, where the structure of the substrates, the structure of the polymerase, and the conditions, are all interdependent and co-optimizable.

RESULTS

Synthetic Strategy

The synthetic strategy for the tagged triphosphates (Figure 2, **5–8**) has these features, illustrated for the thymidine analog in Scheme 1:

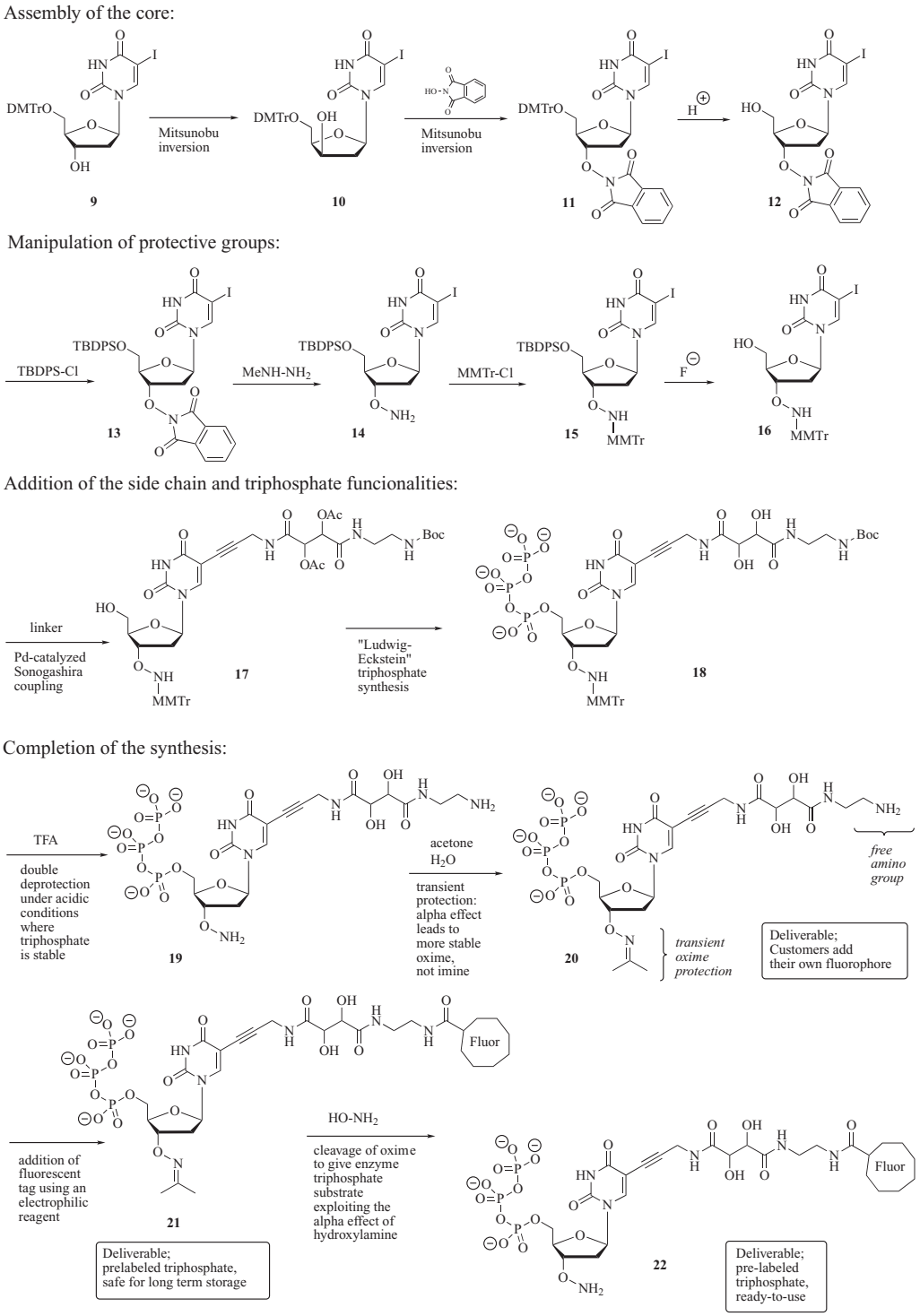
(a) 5-Iodinated natural pyrimidine nucleosides and 7-deaza-7-iodopurine nucleosides are the key starting materials.

(b) The 3′-ONH_2 group is introduced with the correct stereochemistry via two consecutive Mitsunobu reactions, the second using N-hydroxyphthalimide as the nucleophile.

(c) Following adjustment of the protecting groups, the 3′-ONH_2 group is carried through the majority of the remaining synthesis protected by the acid-labile monomethoxytrityl group.

(d) The taggable, cleavable side chains are introduced by palladium-catalyzed coupling of a functionalized acetylene with an iodinated pyrimidine or an iodinated 7-deazapurine. Introduction of this side chain relatively late in the synthesis permits flexibility in the selection of the cleavable functionality of the side chain. Scheme 1 illustrates this where the side chain contains a 1,2-diol unit, which is cleavable by periodate. A second set of syntheses (not shown) were done where the side chain contains a disulfide unit, which is cleavable by disulfide exchange. In both cases, the side chain nucleophilic group is protected as an acid-labile tert-butyloxycarbonyl (BOC) amide, which is removed in the same step as the MMT protection for the 3′-ONH_2 group.

(e) The triphosphate is introduced in the next-to-last step using the procedure developed by Ludwig and Eckstein.[14] Mild alkaline conditions selectively remove the side chain diol protecting groups; the corresponding disulfide-containing side chain needs no analogous deprotection. The triphosphate group as well as the glycosidic bond



SCHEME 1 Synthetic strategy for managing four reactive units in a taggable nucleoside triphosphate (3′-ONH2, 5′-triphosphate, 1,2-diol cleavable linker, and fluorophore) illustrated for the thymidine analog. Analogous syntheses were done with 5-iodo-2′-deoxycytidine, 7-deaza-7-iodo-2′-deoxyadenosine, and 7-deaza-7-iodo-2′-deoxyguanosine (see supplemental material).(Conditions: See the supplemental material.)

are stable under the anhydrous conditions where the acid labile MMT and BOC groups are removed.

(f) The 3′-ONH_2 group is selectively protected as an alkoxime when the fluorophore is appended to the side chain. This permits the fluorescent species to be added as an electrophile to the amino group of the linker under conditions where the triphosphate is unreactive.

(g) The free 3′-ONH_2 group is restored by cleaving the alkoxime with hydroxylamine, a treatment that leaves the triphosphate and fluorophore intact.

Analogous syntheses are done with 5-iodo-2′-deoxycytidine, 7-iodo-2′-deoxyadenosine, and 7-iodo-2′-deoxyguanosine (refer to supplemental material).

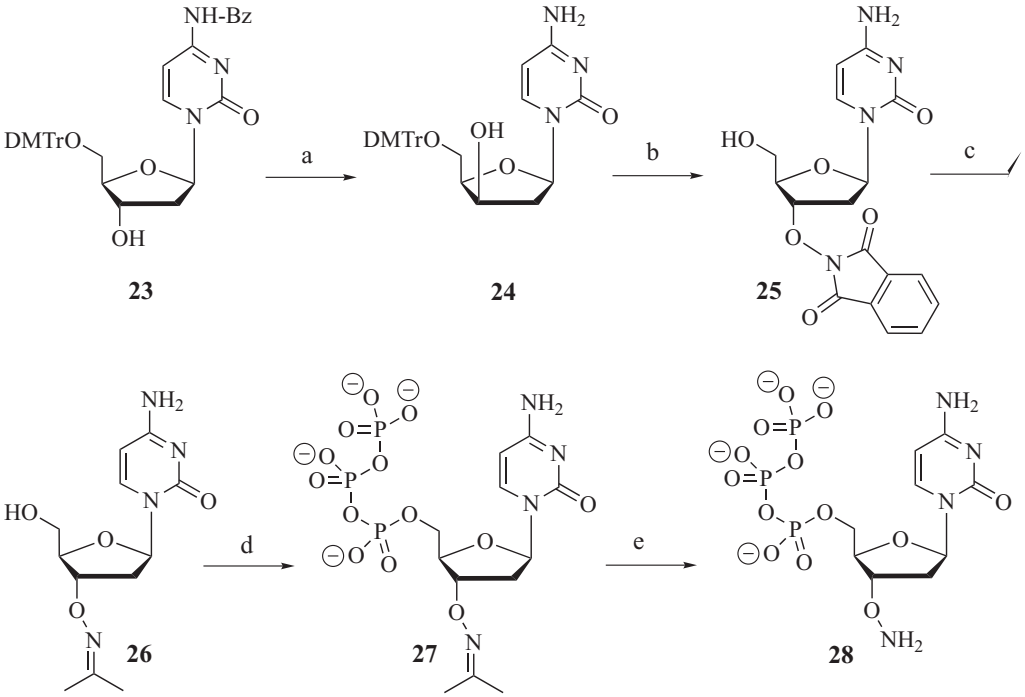
For commercialization, this strategy allows the oxime triphosphate having a free side chain amino group to be delivered (**20** in Scheme 1); users receiving this intermediate may then choose their own fluorescent derivatizing agents, many of which are available commercially. Their synthesis is then completed to give active substrate (**22**) by cleavage of the alkoxime. Alternative deliverables are the fluorescently labeled oxime (**21**) and the fluorescently labeled triphosphate ready for polymerase incorporation (**22**).

To develop this strategy, we began by preparing a set of untagged building blocks (Figure 2, **1–4**). Scheme 2 illustrates their preparation using the 2′-deoxycytidine analog. This model chemistry demonstrated the introduction of the 3′-ONH_2 group via two consecutive Mitsunobu reactions, and the transient protection of the 3′-ONH_2 group as an alkoxime during the preparation of the triphosphate. As discussed below, these untagged triphosphates are useful in cyclic reversible termination architectures, and for other purposes.[11]

Cleavage Chemistry

A variety of reagents were explored that might convert a 3′-ONH_2 group at the 3′-end of an oligonucleotide primer into a 3′-OH group under conditions sufficiently mild to leave a DNA duplex and its constituent nucleobases intact. Reagents examined included quinones, oxidants, and reductants. None of these proved to be satisfactory. Surprisingly, effective cleavage was obtained by a very simple reagent: sodium nitrite buffered to a pH of approximately 5.5, at room temperature.

To develop this reagent for the cleavage of a 3′-ONH_2 group at the 3′-end of a DNA primer, 3′-O-aminothymidine was used as a model. Measurement of the kinetics of the reaction as a function of pH in aqueous sodium acetate buffer (Table 1) showed that 700 mM HONO at pH 5.5 and room temperature was an excellent compromise between short reaction time (98%



SCHEME 2 Synthesis of the untagged 2′-deoxycytidine triphosphate having just two reactive units (3′-ONH_2 and 5′-triphosphate). Analogous syntheses were done with the other three natural nucleosides (see supplemental material). Conditions: (a) (1) BzOH, Ph_3P, DIAD, THF, 0°C to RT, 16 hours; (2) NaOMe, MeOH, RT, 2 hours (62% overall); (b) (1) *N*-hydroxyphthalimide, Ph_3P, DIAD, THF, 0°C to RT, 16 hours; (2) HCl, MeOH, RT, 10 minutes (63% overall); (c) (1) $MeNH_2$, H_2O, RT, 10 minutes; (2) acetone, H_2O, RT, 3 hours (71% overall after HPLC); (d) (1) 2-chloro-4H-1,3,2-benzodioxaphosphorin-4-one, pyridine/dioxane, RT, 15 minutes; (2) $(Bu_3N)_4P2O_7$, DMF, RT, 20 minutes; (3) iodine, pyridine/H_2O, 20 minutes (38% overall after double HPLC); (e) $HONH_2$, aq. NaOAc buffer (200 mM, final pH 4.5), RT, 2 hours (74% after HPLC).

TABLE 1 Cleavage of 3′-O-aminothymidine with aqueous buffered HONO

			cleavage of -ONH_2 (%)			
[$NaNO_2$] (M)	[NaOAc] (M)	pH (± 0.05)	t = 30 seconds	t = 1 minute	t = 2 minutes	t = 5 minutes
0.7	1.0	4.5	>99	>99		
0.7	1.0	5.0	>99	>99		
0.1	1.0	5.5			18	
0.35	1.0	5.5			91	
0.7	1.0	5.5	83	98	>99	
0.7	1.0	5.65		80	96	
0.7	1.0	6.0		29	51	83

3′-O-Aminothymidine (20 mM, 2 μL) was treated at room temperature with aqueous buffered nitrous acid (50 μL, prepared from sodium nitrite and 1 M NaOAc buffer).

TABLE 2 Oxidation of natural nucleosides with aqueous buffered HONO under cleavage conditions

Nucleoside	Degradation products after 72 hours (4,320 minutes) (quantified at 260 nm)
dG	20%
dA	13%
dC	15%

Nucleosides (20 mM, 30 μL) were treated at room temperature with aqueous buffered nitrous acid (1 mL, 0.7 M $NaNO_2$, 1.0 M NaOAc, pH 5.5).

cleavage after 1 min) and mild conditions that would leave the DNA duplex intact.

Nitrates and HONO are widely present in the environment, as they are used as food preservatives. Prolonged treatment of DNA with HONO is known to cause the deamination of certain nucleobases. Accordingly, we asked how long DNA could be exposed to these cleavage conditions before it suffered significant amounts of damage. Accordingly, 2′-deoxyguanosine, 2′-deoxyadenosine and 2′-deoxycytidine were separately exposed to the cleavage conditions at room temperature; the products were assayed by HPLC. The results (Table 2) showed that only 10–20% degradation was observed after 72 hours (4,320 minutes), a period ca. 4,300 times longer than the time required to cleave 98% of the -ONH_2 group.

Enzymology of the Tagged Species

With syntheses in hand of both tagged and untagged nucleoside triphosphates having a 3′-ONH_2 block together with a reagent (buffered HONO) that allowed the block to be removed under mild conditions, we turned to polymerases. Here, we encountered several issues that must be addressed when using enzymes to add a nucleotide from a triphosphate that is modified in two ways, on its nucleobase and 3′-OH unit. Each of these modifications might cause a polymerase to reject such a triphosphate as "unnatural." Further, after the modified nucleotide is incorporated and the tag is removed, the "scar" from the linker that previously attached the tag also makes the 3′-end of the new primer unnatural. Preliminary data suggested that the ability of a polymerase to accept a tagged nucleoside triphosphate is influenced by the presence of a 3′-ONH_2 blocking group (and vice versa), while the presence of a scar on the nucleotide at the 3′-end of a primer influences the ability of a polymerase to accept an incoming tagged nucleoside triphosphate, where that influence is different for triphosphates that have or lack a 3′-O blocking moiety.

To dissect these issues, we examined first nucleoside triphosphates having only the 3′-ONH_2 group and lacking a linker, using primers that lacked

a scar. This immediately revealed two technical challenges. First, the alpha effect displayed by the 3′-ONH_2 unit makes it an aggressive scavenger for aldehydes and ketones, where the product is an oxime of the type that served as an intermediate in the synthesis of the triphosphates (**20** in Scheme 1 and **26** in Scheme 2). Aldehydes and ketones are ubiquitous in many laboratories; acetone, for example, is a common laboratory solvent. The oxime is, of course, much larger than a 3′-ONH_2 unit, which defeats the purpose of its use as a reversible terminator. Accordingly, conditions for polymerase extension were adjusted by incorporating hydroxylamine in primer extension buffers. This scavenged aldehydes and ketones without influencing the activity of the polymerase.

Second, unblocked nucleoside triphosphates are ubiquitous in molecular biology laboratories, and can contaminate polymerase preparations. Unblocked triphosphates can, in principle, compete with blocked triphosphates, causing a fraction of the extended primer to be extendable further. We, therefore, developed assays that detect small amounts of unblocked triphosphates in assay mixtures. These used test templates that called for the consecutive addition of two of the same nucleobases ("double addition test"). A preparation containing no unblocked triphosphate gave a single $n + 1$ primer extension band on the denaturing polyacrylamide gel. In contrast, a preparation containing some unblocked triphosphate gave both an $n + 1$ primer extension band and an $n + 2$ extension band. The ratio of the two provided a direct way to assess the relative extent of contamination by the unblocked triphosphate. Several applications of this test are shown in figures in the supplemental material.

With these technological issues addressed, we examined the commonly used thermostable DNA polymerase from *Thermus aquaticus* (*Taq*) as a catalyst for addition of our triphosphates.[11] Initial experiments showed that this polymerase accepted untagged 3′-ONH_2 blocked triphosphates **1–4** only poorly. We, therefore, applied a "Reconstructed Evolutionary Adaptive Path" (REAP) analysis, reported elsewhere,[11] to identify sites in the *Taq* protein sequence where amino acid replacements might improve the ability of the polymerase to accept this unnatural modification. REAP identified two sets of replacements (A597T, L616A, F667Y, E745H = Taq 442, and E520G, K540I, L616A = Taq 475) that, when incorporated into a polymerase, allowed the polymerase to accept the thymidine analog **1**.[11]

We, therefore, extended this work to include the three other nucleoside triphosphates **2–4** with similarly successful results, and with little evidence for mismatching. For example, dCTP-ONH_2 (**2**) was added by Taq 475 smoothly opposite dG in the template, but not opposite dA in the template (a mismatch that is believed to be facilitated by protonation of the nucleobase). (See Figure E1 in the supplemental material; this figure also illustrates the use of the double addition test for unblocked triphosphates).

Similarly, dATP- ONH_2 (**3**) was smoothly added by Taq 475 opposite dT in the template, but not opposite dG or dA (Figure E2 in the supplemental material, which also presents a double addition test showing the absence of detectable amounts of unblocked triphosphate in these preparations). Likewise, TTP-ONH_2 (**1**) was added by Taq 475 opposite dA in the template, but not opposite dG in the template, a mismatch that is believed to be facilitated by a wobble. Last, dGTP-ONH_2 (**4**) was added by Taq 475 opposite dC in the template, but not opposite dG (data not shown).

Extension-Termination-Cleavage-Extension Cycles on a Solid Support

Cycles of addition-with-termination, cleavage, and continued extension were done initially in solution using magnetic beads (Figure E3 in the supplemental material, adapted from Chen et al.[11]). Here, an addition of an untagged 3′-blocked triphosphate (**1–4**) was followed by treatment with buffered sodium nitrite, followed by a second primer extension with **1–4** triphosphates followed by a second treatment with buffered sodium nitrite, followed by addition of unblocked (natural) dNTPs to give a full length product.

The recently developed prototype of the PinPoint Sequencer (Figure E16 in the supplemental material) by Intelligent Bio-Systems (Waltham, MA, USA; IBS) made it possible to perform a similar cycling on a two-dimensional array (Figure 3; see also Figure E17 in the supplemental material). Here, self-priming hairpin templates immobilized on Codelink slides (SurModics, In Vitro Diagnostic Products, Eden Prairie, MN, USA) were created as low density arrays using amino-carboxyl chemistry (Figure 3A). Then, untagged 3′-blocked triphosphates (a mix of **1–4**) were added seven times using Taq 475 polymerase, each time followed by treatment with buffered sodium nitrite to convert the 3′-ONH_2 to a 3′-OH to restore extendability to the primer. All four untagged 3′-ONH_2 blocked triphosphates (**1–4**) were used with various templates. Extension steps were automatically performed within the flow cell of the Pinpoint instrument at 65°C for 15 minutes. After the seventh automated cycle, sequencing was completed by incubation with a mixture of fluorescently labeled dideoxynucleoside triphosphates and a proprietary IBS polymerase (5 minutes at 65°C). The flow cell was then washed and imaged using a CCD-based fluorescence imaging system integrated within the instrument. The images were corrected for background and color crosstalk between filter channels and quantified in the 4 filter channels.

The results for four different DNA templates are presented in Figure 3B. Calls were correctly made regardless of the nucleotide present at position 8, indicating that the preceding seven cycles of incorporation and cleavage

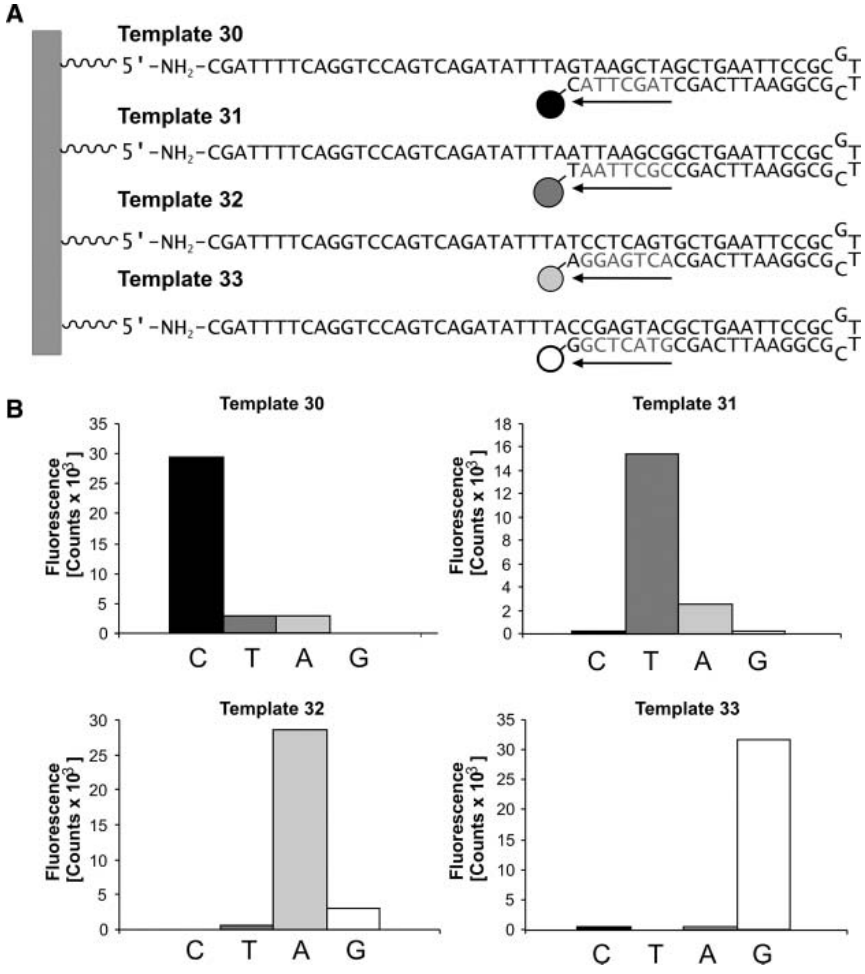


FIGURE 3 Cycling with untagged 3′-ONH2 triphosphates on two-dimensional array on the PinPoint Sequencer (Intelligent Bio-Systems, Waltham , MA, USA). A) Self-priming templates (3033) used in sequencing experiments. Black underlined sequence indicates 7 sequencing cycles in which a mixture of all four untagged 3'-ONH2 triphosphates (1-4) were used. Shaded circle on last base (cycle 8) indicates fluorescence readout for which a mixture of labeled ddNTPs was used. B) Bar graphs showing 4-channel fluorescence intensity observed for templates 3033. The calls (the channel with the highest signal) correctly detected the nucleotide present at position 8 for all four templates and indicate that the preceding 7 cycles of incorporation and cleavage of the correct 3′-ONH2 were complete and in phase. The residual signals observed in other channels are most likely due to a combination of incomplete background/crosstalk correction and/or some dephasing. (For view of instrument and more experimental details, see text and Figures E16 and E17 in the supplemental material.)

of the correct 3′-ONH_2-blocked nucleotides were complete and in phase (implying that the failure of the cycle to complete for a significant fraction of the oligonucleotides being sequenced did not cause a significant number of nucleotides to report at site n while the rest reported at site $n + 1$). The residual signals observed in other channels are most likely due to a combination of incomplete background/crosstalk correction and/or some dephasing.

Polymerases that Accept Nucleoside Triphosphates 5–8, Having Both a 3′-ONH_2 Group and a Fluorescent Tag Appended to the Nucleobase

We then turned to develop polymerases that accept triphosphates having both a 3′-ONH_2 group and an attachment to its nucleobase (**5–8**). Taq 475 was found to accept thymidine triphosphates having a 3′-ONH_2 moiety and either a diol linker with a free amino group (chosen because the diol can be cleaved by periodate in the sequencing step that removes the fluor; Figure E4 in the supplemental material) or that diol linker to which had been attached a fluor, here 5-FAM (**5b**) (Figure E5 in the supplemental material). In both modifications, incorporation occurred with lower efficiency. Accordingly, a series of experiments were run to seek optimal solutions to the "three part optimization" problem presented by the presence of a 3′-blocking group on the incoming triphosphate, the presence of a linker on the nucleobase of the incoming triphosphate, the presence of a fluorescent species appended via that linker, and the possible presence of a "scar" on the 3′-teminal nucleotide of the primer, a scar that remains after cleavage of the tag from the previous incorporation.

Interestingly, these modifications were found to interact to determine the efficiency of incorporation of the triphosphate. For example, Taq 475 could incorporate an untagged triphosphate carrying a 3′-ONH_2 group with an unscarred primer, a tagged triphosphate with a 3′-OH with an unscarred primer, a tagged triphosphate with a 3′-OH with a scarred primer if there is no 3′-ONH_2, and an untagged triphosphate with a 3′-OH with a scarred primer, all better than a tagged triphosphate having a 3′-ONH_2 and the primer having a scar (see supplemental material).

To understand better the interaction between a tag on the incoming triphosphate and a tag on the primer independent of the impact of a 3′-block, a series of experiments were run with various mutants with tagged, unblocked triphosphates and several Taq variants (Figure E6 in the supplemental material). Generally, the polymerases were able to add a tagged triphosphate to a tagged primer. The different performance of the different variants of Taq (native, Taq 475 and Taq 442) suggested that mutations in sites selected in our REAP analysis[11] have an impact on the ability of polymerases to accept tagged triphosphates and, therefore, involve appropriately chosen sites.

Likewise, these Taq variants were also able to add to an *unscarred* primer a triphosphate carrying both a tag and a 3′-ONH_2 block (Figure E7 in the supplemental material). Interestingly, the nature of the fluorophore was found to have subtle but noticeable impact on the performance of the system, even though the fluorophore was many atoms removed from the reacting atoms. For example, 5-FAM and 6-FAM fluorophores differed in their ability to be handled by various Taq variants, even though these fluorophores are simply regioisomers of each other.

Given this result, we screened various polymerases, including Therminator and these Taq variants, for their ability to accept a TTP-ONH_2 block with a different fluorophore attached to the nucleobase. Screening of several polymerases showed that Therminator was found to support the addition of a tagged TTP- ONH_2 (**5**) better than the best Taq variant identified to date (Figure E8 in the supplemental material).

With these results in hand, we asked whether tagged purine triphosphates carrying a 3′-ONH_2 could also be incorporated by one or more of these polymerases. Wholly unexpectedly, wild-type Taq was able to incorporate dATP carrying both a 3′-ONH_2 group and a 7-position attachment (7-deazaadenine is the nucleobase) carrying a fluorophore (here Cy5) (**7**), although the Taq 475 and Taq 442 variants did better (Figure E9 in the supplemental material). A range of other Family A (Bst, Klenow) and Family B (Therminator and 9°N, Vent and Deep Vent) polymerases also worked, with a variety of fluorescent tags.

Evidence of an occasional contaminant triphosphate in various incubation mixtures (generating an $n + 2$ band in the double addition assay) encouraged us to develop a second assay to detect its presence. Here, instead of looking for a double addition of the same triphosphate to a primer, the triphosphate with a 3′-block was added with the three other unblocked triphosphates. Thus, if an unblocked triphosphate were present, the polymerase would use the remaining triphosphates to give a full length product. As the full length product would be quantifiable more easily than the $n + 2$ product (where spill over from the $n + 1$ product necessarily places a bound on the limits of detection of the contaminant, and where the different mobilities of various extension products with different fluorophores creates confusion), this should be a more sensitive test. This assay was applied in Figure E10 (in the supplemental material) for 7-deaza-dATP-ONH_2-Cy5 (**7**) and in Figuer E11 (in the supplemental material) for dCTP-ONH_2-Cy3.5 (**6**).

Extension-Termination-Cleavage Cycles with Mixtures of Untagged and tagged Triphosphates

These results are sufficient to establish a cyclic reversible termination sequencing cycle using a mixture of tagged (**5–8**) and untagged (**1–4**) nucleotides, similar to the strategy developed previously at Intelligent BioSystems for other reversibly terminating chemistries (data not shown). This strategy avoids the need to identify polymerases that add a tagged triphosphate to a primer whose 3′-nucleotide also contains a scar. It also avoids having extension products that are highly modified, something that may give intractable DNA after multiple cycles.[15] Therminator was chosen as the best compromise polymerase for this proof-of-concept experiment. Accordingly,

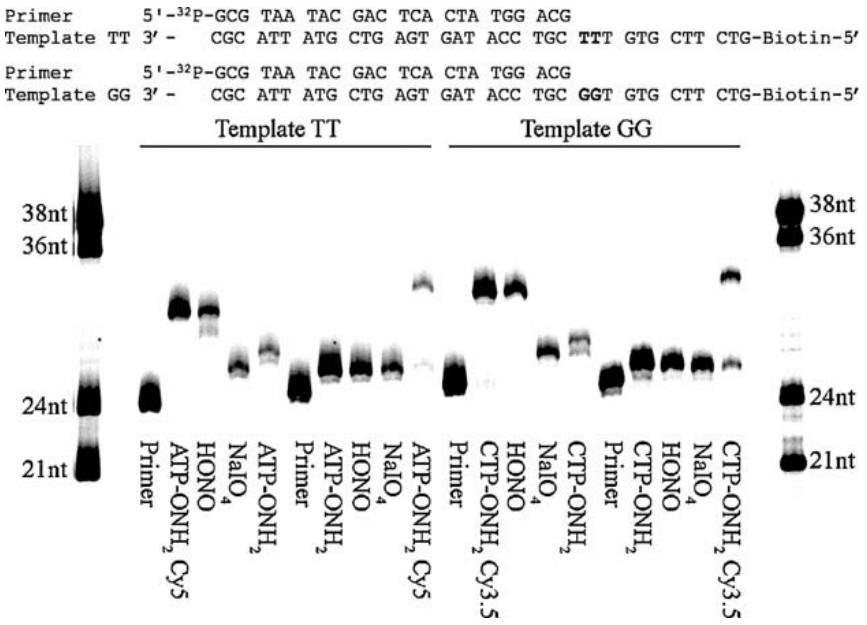


FIGURE 4 Reversible terminator sequencing cycle using 7-deaza-dATP-ONH2-Cy5 (7), dCTP-ONH2-Cy3.5 (6), dATP-ONH2 (3), dCTP-ONH2 (2), and Therminator (0.5 U/reaction). The cycles involve (a) adding reversibly terminated triphosphate to a primer annealed to a template attached via a biotin tag to a streptavidin bead (to give an n+1 band that migrates more slowly if the triphosphate is tagged); (b) treating the beads with buffered sodium nitrite (the active reagent is HONO) to generate a free extendable 3′-OH end; (c) cleaving the 1,2-diol side chain with aqueous sodium periodate (50 mM) to release the fluorophore; (d) resuming extension. The tagged and untagged triphosphates were alternated. If a tagged triphosphate was added first, the untagged triphosphate was added second. If an untagged triphosphate was added first, the tagged triphosphate was added second. (For more experimental details, see Figure E15 in the supplemental material.)

experiments were first run to determine the amount of Therminator polymerase needed for the sequential reactions (Figure E12 in the supplemental material). A half unit (0.5 U) of Therminator per reaction works well.

We then optimized the ratio of tagged and untagged dNTP-ONH_2 species by placing them in competition and using the different electrophoretic mobilities of the extended, terminated product with and without a fluorescent tag. These experiments are shown with the thymidine analogs **1** and **5b** in Figure E13 (in the supplemental material) and Figure E14 (in the supplemental material). When challenged with a mixture of TTP-ONH_2 (**1**) and TTP-ONH_2 tagged with 5-FAM (**5b**) at a ratio of 1:9, 81% of the $n + 1$ product arose via incorporation of **1**, while 19% arose from incorporation of the tagged derivative **5b**.

With these results in hand, a series of sequencing cycles were established (Figure 4; see also Figure E15 in the supplemental material). The cycle began by adding reversibly terminated triphosphate (shown are either **6** or **7**) to a primer annealed to a template attached via a biotin tag to a streptavidin magnetic bead. Incorporation with termination gave an $n + 1$ band that

migrates more slowly if the triphosphate is tagged. The beads were then treated with buffered sodium nitrite to generate a free extendable 3′-OH end. The 1,2-diol side chain was cleaved with periodate to release the fluor. The extension was then resumed.

DISCUSSION

Several features of the synthetic strategy reported here are worthy of comment. The first is the use of the "alpha effect" in this strategy. The alpha effect is defined by experimental observations that show that a nucleophilic center attached directly to an atom that carries an unshared pair of electrons has extraordinary nucleophilicity attacking a carbon electrophile when compared to its basicity. An alpha effect is displayed by the 3′-ONH_2 group itself, which permits it to be selectively protected in the presence of the side chain nucleophilic CH_2-NH_2 group. The alpha effects displayed by methylhydrazine and hydroxylamine are also exploited, respectively, in the manipulation of the protecting groups and in the very last step to create the enzymatic substrate from the oxime. This, coupled with the fact that phosphorus is oxophilic (but not aminophilic), allows deprotection of the 3′-ONH_2 group in the presence of the triphosphate.

The alpha effect is also exploited in the cleavage step in the cyclic reversible termination architecture. Here, the special reactivity of a nucleophilic center appended to an atom having unshared pairs of electrons was the basis for the selective cleavage of the 3′-ONH_2 group after it had been added to the 3′-end of a primer by template directed polymerization mediated by a DNA polymerase.

As has been discussed in the literature for nearly two decades,[16] polymerases are idiosyncratic in their performance with modified triphosphate substrates. We were especially surprised to see different variants behaving differently with tagged purine and pyrimidine analogs. This notwithstanding, the sites in *Taq* polymerase identified by the evolutionary analysis of the past three billion years of Family A polymerase history[11] were clearly variable to good effect. Both Taq 475 and Taq 442 are useful polymerases, even if Therminator is the preferred polymerase in the proof-of-concept experiments reported here.

These experiments show the feasibility of using 3′-ONH_2-blocked nucleoside triphosphates **1–8** in extension-termination-cleavage cycles, in solution and in a highly parallel automated sequencing instrument. More generally, this work illustrates how the development of reagent solutions to problems in the analysis of DNA can benefit from the long tradition in physical organic chemistry that provided many semi-empirical rules to relate function to molecular behavior. These rules, some a half century old, are valuable even in "next generation" architectures for sequencing DNA, and underscore the importance of a strong foundation in chemistry, even in the age

of the genome. Indeed, when combined with analysis of polymerases, they can advance the age of the genome, providing an interesting example of multidisciplinarity that combines the extremes of chemistry and biology.

REFERENCES

1. Metzker, M.L. Emerging technologies in DNA sequencing. *Genome Res.* **2005**, 15, 1767–1776.
2. Fuller, C.W.; Middendorf, L.R.; Benner, S.A.; Church, G.M.; Harris, T.; Huang, X.; Jovanovich, S.B.; Nelson, J.R.; Schloss, J.A.; Schwartz, D.C.; Vezenov, D.V. The challenges of sequencing by synthesis. *Nature Biotechnology* **2009**, 27, 1013–1023.
3. Hiatt, A.C. Compositions for enzyme catalyzed template-independent creation of phosphodiester bonds using protected nucleotides, 2001, U.S. Patent No. 6232465.
4. Ruparel, H.; Bi, L.R.; Li, Z.M.; Bai, X.P.; Kim, D.H.; Turro, N.J.; Ju, J.Y. 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* **2005**, 102, 5932–5937.
5. Ju, J.Y.; Kim, D.H.; Bi, L.R.; Meng, Q.L.., Bai, X.P.; Li, Z.M.; Li, X.X.; Marma, M.S.; Shi, S.; Wu, J.; Edwards, J.R.; Romu, A.; Turro, N.J. Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators. *Proc. Natl. Acad. Sci. USA* **2006**, 103, 19635–196340.
6. Metzker, M.L.; Raghavachari, R.; Richards, S.; Jacutin, S.E.; Civitello, A.; Burgess, K.; Gibbs, R.A. Termination of DNA synthesis by novel 3′-modified deoxyribonucleoside 5′-triphosphates. *Nucleic Acids Res.* **1994**, 22, 4259–4267.
7. Guo, J.; Xu, N.; Li, Z.M.; Zhang, S.L.; Wu, J.; Kim, D.H.; Marma, M.S.; Meng, Q.L.; Cao, H.Y.; Li, X.X.; Shi, S.D.; Yu, L.; Kalachikov, S.; Russo, J.J.; Turro, N.J.; Ju, J.Y. Four-color DNA sequencing with 3′-O-modified nucleotide reversible terminators and chemically cleavable fluorescent dideoxynucleotides. *Proc. Natl. Acad. Sci. USA* **2008**, 105, 9145–9150.
8. Ju, J.Y.; Kim, D.H.; Guo, J.; Meng, Q.L.; Li, Z.M.; Cao, H.Y. DNA sequencing with non-fluorescent nucleotide reversible terminators and cleavable label modified nucleotide terminators, 2008, PCT Int. Appl. Publ. WO2009054922.
9. Wu, W.; Stupi, B.P.; Litosh, V.A.; Mansouri, D.; Farley, D.; Morris, S.; Metzker, S.; Metzker, M.L. Termination of DNA synthesis by N6-alkylated, not 3′-O-alkylated, photocleavable 2′-deoxyadenosine triphosphates. *Nucleic Acids Res.* **2007**, 35, 6339–6449.
10. Benner, S.A. Method for sequencing DNA and RNA by synthesis, 2009, U.S. Patent No. 7544794.
11. Chen, F.; Gaucher, E.A.; Leal, N.A.; Hutter, D.; Havemann, S.A.; Govindarajan, S.; Ortlund, E.A.; Benner, S.A. Reconstructed Evolutionary Adaptive Paths (REAP) give polymerases accepting reversible terminators for sequencing and SNP detection. *Proc. Natl. Acad. Sci. USA* **2010**, 107, 1948–1953.
12. Kondo, K.; Ogiku, T.; Inoue, I. Synthesis of 5′(3′)-O-amino nucleosides. *Nucleic Acids Res. Symp. Ser.* **1985**, 16, 93–96.
13. Burgess, K.; Gibbs, R.A.; Metzker, M.L.; Raghavachari, R. Synthesis of an oxyamide linked nucleotide dimer and incorporation into antisense oligonucleotide sequences. *J. Chem. Soc. Chem. Commun.* **1994**, 8, 915–916.
14. Ludwig, J.; Eckstein, F. Rapid and efficient synthesis of nucleoside 5′-O-(1-thiotriphosphates), 5′-triphosphates and 2′,3′-cyclophosphorothioates using 2-chloro-4H-1,3,2-benzodioxaphosphorin-4-one. *J. Org. Chem.* **1989**, 54, 631–635.
15. Tasara, T.; Angerer, B.; Damond, M.; Winter, H.; Doerhoefer, S.; Huebscher, U.; Amacker, M. Incorporation of reporter molecule-labeled nucleotides by DNA polymerases. II. High-density labeling of natural DNA. *Nucl. Acids Res.* **2003**, 31, 2636–2646.
16. Horlacher, J.; Hottiger, M.; Podust, V.N.; Huebscher, U.; Benner, S.A. Recognition by viral and cellular DNA polymerases of nucleosides bearing bases with nonstandard hydrogen bonding patterns. *Proc. Natl. Acad. Sci. USA* **1995**, 92, 6329–6333.