Syntheses of Nucleosides Designed for Combinatorial DNA Sequencing

Mike B. Welch,[a] Carlos I. Martinez,[b] Alex J. Zhang,[a] Song Jin,[a] Richard Gibbs,[b] and Kevin Burgess*[a]

Abstract: Nucleoside triphosphates **I** with 3′-O-blocking groups that are both photolabile and fluorescent were required to investigate the viability of a strategy for sequencing DNA in a combinatorial fashion (see Figure 1). Four compounds were prepared to realize this goal. Two of them, **14a** and **14t**, had dansyl-functionalized, 3′-*O*-(2″-nitrobenzyl) ether groups, while the other two, **18a** and **18t**, had similar pendant carbonate groups. Tests for incorporation of these analogues were performed by using five different DNA replicating enzymes, but the analogues were not incorporated. These results were surprising in view of the fact that previous studies had shown that 3′-*O*-(2″-nitrobenzyl)adenosine triphosphate **II** was incorporated by *Bst* DNA polymerase I. However, molecular simulations with the coordinates of a T7 polymerase crystal structure as a model demonstrates that analogues **14a**, **14t**, **18a** and **18t** are too large to fit into the enzyme active site, whereas accommodation of the unsubstituted 2-nitrobenzyl compound **II** is much less demanding. We conclude that both the nucleoside triphosphates and the DNA polymerase enzyme must be modified if the proposed DNA sequencing scheme is to be viable.

Keywords: DNA sequencing · enzyme catalysis · nucleotides

Introduction

Complete DNA sequence analysis of the human genome is a costly and time-consuming project. Accelerated methods for sequencing large DNA strands are therefore highly desirable. Most of the current efforts to improve sequencing are technological improvements of the Sanger[1, 2] or Maxam–Gilbert[3] schemes.[4] These include adaptation of robotic systems for processing fluorescent dideoxy-terminated nucleotides on commercially available DNA sequencing machines[5, 6] coupled with ultra thin,[7] or capillary gel,[8–11] electrophoresis to improve efficiency. Conventional gel electrophoresis is not required for some of these approaches, nevertheless they are unlikely to reduce the cost and time factors to acceptable levels. Other modifications of conventional sequencing schemes focus on the primer, but still require gel electrophoresis.[6, 12–14] For instance, contiguous hexamer strings may be used in primer walking methods wherein the appropriate primers are drawn from an oligonucleotide library,[15, 16] but the feasibility of this methodology remains to be proven for large-scale projects.

[a] Prof. Dr. K. Burgess, M. B. Welch, A. J. Zhang, S. Jin
Texas A & M University, Chemistry Department
P. O. Box 300012, College Station, Texas 77842 (USA)

[b] C. I. Martinez, R. Gibbs
Department of Molecular and Human Genetics
Baylor College of Medicine, One Baylor Plaza
Houston, TX 77030 (USA)

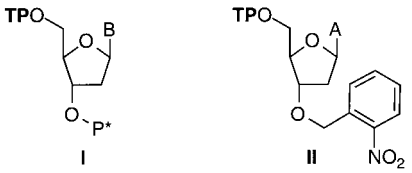
Other advances in sequencing focus on simultaneous processing of data. The simplest and most widely applied form of such multiplexing is separations of combinations of Sanger sequencing reactions in single gel lanes.[17] Schemes involving combination of two sets of Sanger sequencing reactions have also been proposed,[17] but are not used frequently. Multichannel capillary electrophoresis has also been explored, and shows considerable promise.[8, 18–20] Other forms of multiplex sequencing involve oligonucleotide probes to visualize fragments after they have been transferred to a nylon membrane.[21–23] However, efficiency enhancements from any one of these methods is unlikely to raise the throughput of sequence data by more than one or two orders of magnitude.

There are few fundamentally new approaches to sequencing. Novel methodologies include those involving scanning tunneling microscopy,[24] single molecule detection,[25, 26] methods based on detection of the pyrophosphate liberated in each addition step,[27, 28] mass spectrometry,[29] and hybridization (SBH-techniques).[30–32] These protocols may offer significant increases in efficiency over the established procedures. They generally do not require gel electrophoresis, therefore some can potentially process larger numbers of samples without concomitant increases in equipment, reagents, or time. However, at this stage these procedures are largely unproven, and some have obvious disadvantages. Scanning tunneling microscopy and other single molecule detection methods, for instance, have not evolved to the level required for reliable

  0947-6539/99/0503-0951 $ 17.50+.50/0

sequencing, and the techniques based on pyrophosphate liberation and hybridization cannot be used to characterize repeated sequences. The main applications of SBH-techniques appear to be for detecting mutations.[33–35] Numerous other schemes have been proposed but work on these has apparently ceased.[36, 37]

Our group and others have been exploring another way to facilitate parallel analyses of multiple samples in an array without gel electrophoresis.[38–43] We call this the Base Addition Sequencing Scheme or BASS. Central to this approach is a set of four nucleoside triphosphates **I** that have 3′-O-blocking groups that are both labile and fluorescent. The



TP = triphosphate, $P_3O_9H_4$

B = adenine, A; thymine, T; guanine, G; cytosine, C

P* = photolabile blocking group that is also fluorescent and encodes for the base B (B = A, T, G, or C)

fluorescence of the protecting group should enable the parent base on the ribose skeleton (A, T, G, or C) to be identified. A cycle in the proposed sequencing scheme would consist of the following steps (Figure 1): 1) incorporation of the appropriate nucleotide triphosphate analogue **I** by a DNA replicating enzyme(s); 2) spectroscopic identification of the base analogue incorporated; and, 3) removal of the blocking group P* to regenerate a 3′-hydroxy terminus on the (now elongated) polynucleotide chain.

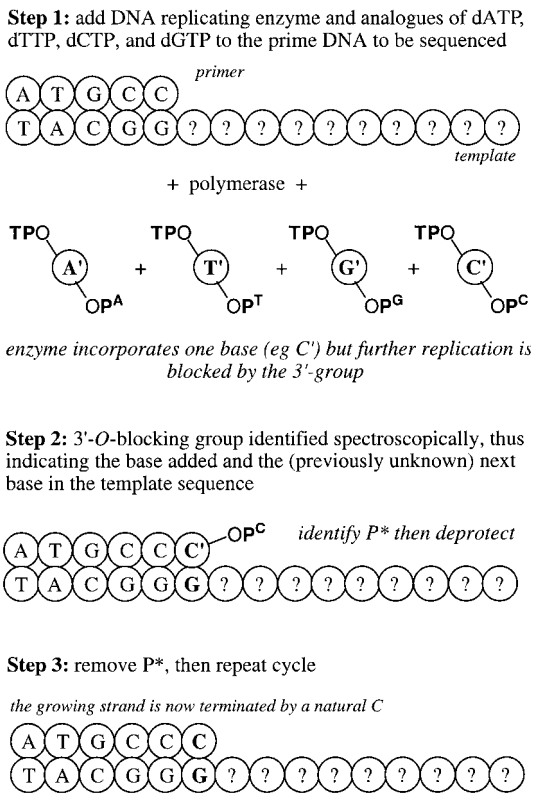


Figure 1. The base addition sequencing scheme (BASS).

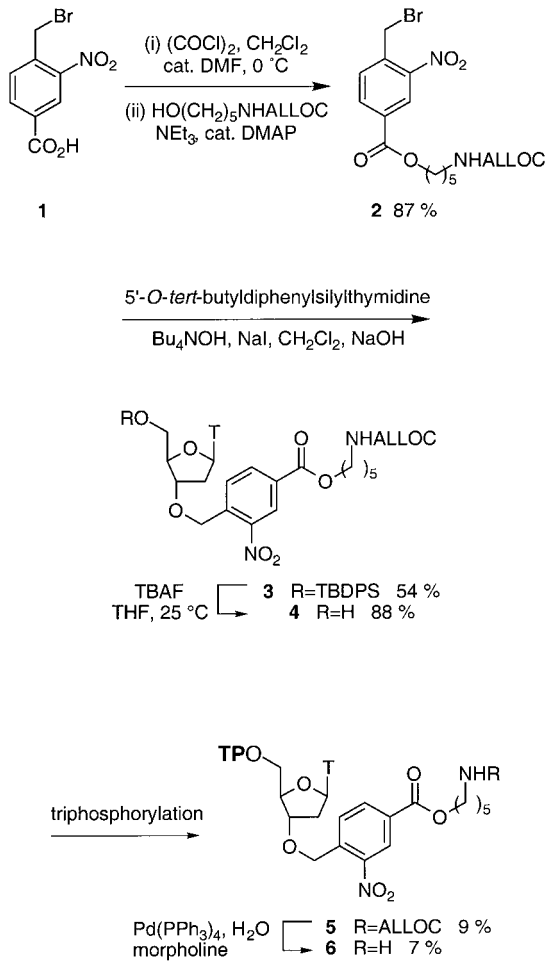
If realized, base addition sequencing would have several significant advantages over the methodologies currently used for sequencing DNA. First, results from each DNA sample will be distinguished by direct analysis of the array, so addition of more samples would not proportionately increase the amount of effort or materials required. This compares favorably with procedures wherein each DNA template must be handled separately, and for which every additional sample requires a new gel lane. Consequently, the proposed scheme potentially has a much greater capacity than conventional sequencing methods. Second, if the experiment could be arranged in such a way that millions of primed DNA fragments were analyzed simultaneously, then it would not be necessary to characterize each primer. Instead the primers could be generated by a combinatorial method, and the arrangement of sequences would be deduced from overlaps in the data. It is possible that the extent of multiplexing would be such that the method would be viable even if a relatively short read of DNA sequence was obtained from each individual experiment (for example 10 – 50 bases). Moreover, the ease of primer synthesis would represent a highly significant cost/time saving advantage. Finally, the method could have incidental benefits like circumvention of artifacts due to gel compressions (frequently associated with G,C-rich strings in the sequence).

In preliminary work we found that 3′-*O*-(2″-nitrobenzyl)-adenosine triphosphate **II** could be incorporated by a DNA polymerase, and that photodeprotection of the 3′-hydroxy was possible facilitating DNA replication.[38] This encouraging result indicated that 3′-modifications could be tolerated by DNA replicating enzymes. Herein we describe syntheses of nucleosides protected with photolabile 3′-O-blocking groups that are also fluorescent, and report preliminary tests for incorporation by DNA polymerase enzymes.

Results and Discussion

Syntheses of nucleoside triphosphates with 3′-ether linkages: Initial attempts to prepare 3′-O-protected nucleosides focused on the use of nitrobenzoic acid derivatives as illustrated in Scheme 1. The readily available starting material **1**[44] was coupled with (*N*-allyloxycarbonyl)pentan-5-ol amine, via the acid chloride, to give the ester **2**. A phase transfer catalyst was used to form the critical ether linkage; development of conditions for this step required considerable experimentation. Removal of the 5′-silyl protecting group, triphosphorylation,[45] and removal of the allyloxycarbonyl group then gave the nucleotide amine **6**. A potentially attractive feature of this route was that addition of fluorescent labels at the very end of the synthesis would allow one advanced intermediate to be transformed into several compounds. Unfortunately, labeling of the triphosphate with BODIPY-SE 503/512 was unsuccessful, due to the small

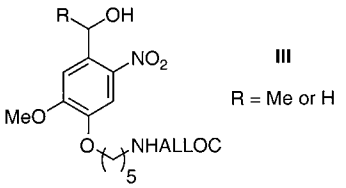
0947-6539/99/0503-0952 $ 17.50+.50/0 *Chem. Eur. J.* **1999**, *5*, No. 3



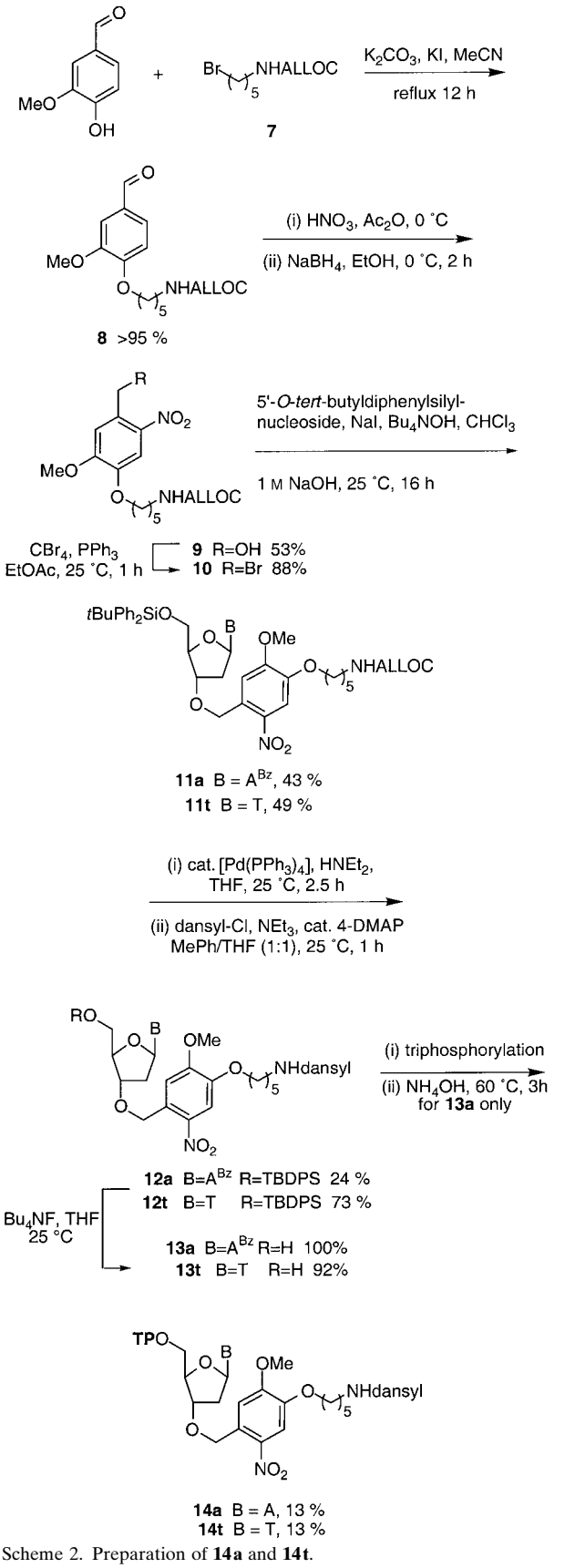
Scheme 1. Synthesis of nucleoside amine **4**.

amounts of material and difficulty in handling triphosphates. A disadvantage of the design **6** is that 2-nitrobenzyl groups substituted with carboxy functionalities cleave less readily under photolytic conditions than comparatively electron-rich systems.[46] This factor, combined with the experimental difficulties associated with the triphosphorylation and labeling steps, led us to investigate alternative routes featuring more photosensitive molecules.

The generic structure **III** represents the photolabile connection sought in the next phase of this work. Photodecomposition of the methyl substituted compounds **III**, R = Me, gives nitroso ketones, whereas the corresponding compounds without this methyl substituent (R = H) give nitroso aldehydes.[47] Nitroso ketones are less reactive by-products hence initial efforts focused on the methyl-substituted compounds. However, the secondary benzylic alcohol **III** where R = Me did not undergo coupling with the 3′-hydroxy of the nucleoside under a variety of conditions.



Coupling of derivatives **III** (R = H) to nucleosides was problematic, but eventually, suitable conditions were developed. Scheme 2 outlines syntheses of derivatives of adenosine



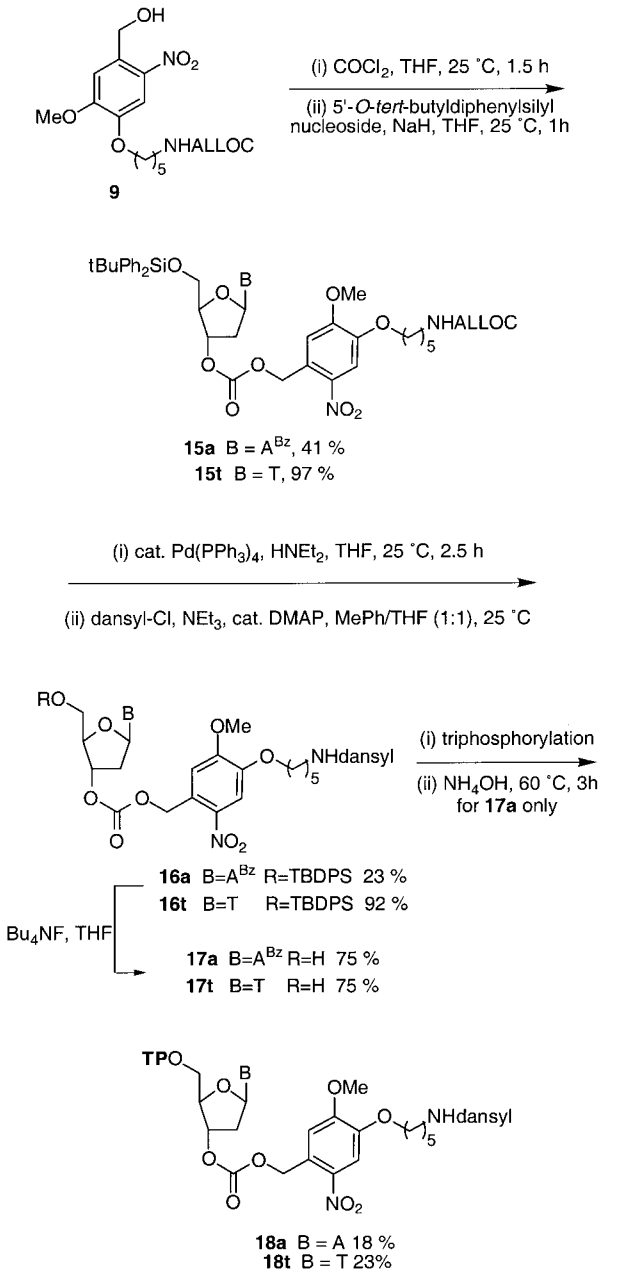
Scheme 2. Preparation of **14a** and **14t**.

  0947-6539/99/0503-0953 $ 17.50+.50/0

and thymidine. The synthesis began with alkylation of vanillin to install a linker for the fluorescent reporter. Nitration followed by reduction provided the primary benzylic alcohol **9**. The critical coupling step involved conversion of the benzylic alcohol **9** into the corresponding benzyl bromide, then phase transfer catalyzed reaction with a 5′-protected nucleoside under biphasic conditions. Numerous other approaches were attempted, but only phase transfer based methods gave positive results. In preparation for the conversion to the nucleotide, the amine was deprotected then dansylated; 5′-desilation then afforded nucleosides **13a** and **13t**.

Triphosphorylation of nucleosides **13** and **17**, and of many other unnatural nucleosides prepared in our laboratories, has proved to be experimentally difficult and tends to give poor yields. The protocol developed by Eckstein et al. was the best out of several approaches attempted,[45] although none of those were entirely satisfactory. Fortunately, only small amounts of the product are required for feasibility tests in bioassays. Debenzoylation of the adenosine derivative **13a** and **17a** was performed after the triphosphorylation sequence (NH_4OH, 60 °C, 3 h). Both pairs of final products, **14a**/**14t** and **18a**/**18t**, were purified by chromatography, first on diethylaminoethyl (DEAE) cellulose, then by RP HPLC.

Syntheses of nucleoside triphosphates with 3′-carbonate linkages: Difficulties encountered in the syntheses of the ether linked derivatives, as outlined above, led us to explore preparations of structurally similar, but hopefully more accessible compounds. Syntheses of structural variants would also help probe the tolerance of DNA replicating enzymes to unnatural nucleosides. Consequently, two carbonate-linked compounds were constructed as described in Scheme 3. Alternative routes to the same compounds were attempted, for instance, by forming a chlorocarbonate functionality from the nucleoside 3′-hydroxy, but none worked as well as that shown.



Scheme 3. Preparation of carbonate-linked compounds.

Tests for incorporation of 3′-blocked nucleoside triphosphates: Analogues **14a**, **14t**, **18a**, and **18t** were tested as substrates for a series of commercially available DNA replicating enzymes. The protocol used for these experiments was based on a procedure we have reported previously.[38] Briefly, 5′-fluorescein-labeled universal primer was annealed to a synthetic oligo template, 5′-*TACGGAGGTG-GACTGGCCGTCGTTTTACA* (italic sequence indicates the replication region). The reactions were carried out in the presence of a mixture containing the corresponding enzyme, some dNTPs, and no other added nucleotides (control), a ddNTP (positive control), or a sample of analogue **14a**, **14t**, **18a**, or **18t**. After incubation the reactions were stopped and loaded on a 20% acrylamide gel, subjected to gel electrophoresis, the gel was scanned, and the results were visualized with a fragment analysis software.

Table 1 shows the results for the incorporation assays. The enzymes tested were unable to recognize the nucleotide

Table 1. Tests of analogues **14a, 14t**, **18a**, **18t** as substrates for DNA polymerases.

Analogue	Polymerase				
	Klenow	rTth DNA Pol.	Vent (exo-) DNA Pol.	Ampli Taq DNA Pol.	Ampli Taq FS
14a	inhibition (100 μM)	inhibition (10 μM)	inhibition (100 μM)	nonselective inhibition (100 μM)	nonselective inhibition (100 μM)
14t	no incorporation	nonselective inhibition (6.5 μM)	no incorporation	no incorporation	no incorporation
18a	no incorporation	no incorporation	no incorporation	no incorporation	no incorporation
18t	no incorporation	no incorporation	no incorporation	inhibition (1.8 mM)	no incorporation

0947-6539/99/0503-0954 $ 17.50+.50/0 *Chem. Eur. J.* **1999**, *5*, No. 3

analogues as substrates for the termination of the DNA amplification under the conditions studied. No incorporation was observed in most of the experiments, but in some there was evidence that the analogue being tested inhibited the polymerase under study, that is no incorporation was observed and only a band corresponding to the unreacted primer could be seen. This was the case for Ampli Taq DNA polymerase for which a 1.18 mM final concentration of the thymidine carbonate **18t** caused complete inhibition. Klenow Fragment, rTth, and Vent (exo –) DNA polymerases exhibited the same behavior but at 100 μM of the adenosine ether **14a**. Nonspecific inhibition (termination of the amplification reaction at different positions along the template with no specificity) was observed for Ampli Taq DNA polymerase, FS and Ampli Taq DNA polymerase when a 100 μM of **14a** was used (Figure 2).

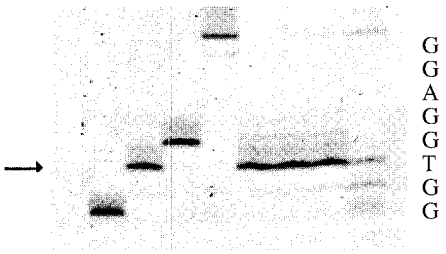


Figure 2. Illustrative data from incorporation assay. Attempted incorporation of **14a** by Ampli Taq DNA polymerase, FS. Arrow marks bands corresponding to no incorporation. Template amplification sequence is shown on the right of the gel. Fluorescein-labeled universal primer was annealed to a complementary oligo template (5′-*TACGGAGGTG-G*ACTGGCCGTCGTTTTACA). Lane 1 contained no dNTPs or ddNTPs. Lanes 2 – 8 contained 0.1 μM dCTP, in addition lanes 3 and 4 contained 0.5 μM ddATP and 0.1 μM dATP, 0.1 μM dTTP, respectively. Lanes 5 to 8 contained 0.1 μM, 2 μM, 10 μM and 100 μM D*ATP, respectively.

Molecular simulations of 3′-blocked nucleoside triphosphates in the active sites of DNA replicating enzymes: Molecular simulations were performed to rationalize the lack of incorporation of the analogues prepared in the course of the work described above. It was perplexing that compound 3′-*O*-(2″-nitrobenzyl)adenosine triphosphate **II** was previously incorporated by a DNA polymerase, whereas similar analogues prepared in the current study were not.

Coordinates for a crystal structure of T7 DNA polymerase encapsulating a primed template and ddGTP were downloaded from the Protein Data Bank and used as a model for this study. This particular set of coordinates was used since they include all the components (enzyme, primed template, and nucleoside triphosphate) and because the data set was recorded at high resolution (2.2 Å). An expansion of the active site of this enzyme complex is shown in Figure 3 top. Removal of the ddGTP entity gave a vacant active site, and several conformers of 3′-*O*-(2″-nitrobenzyl)adenosine triphosphate **II** were fitted in this void by visually docking to form reasonable contacts and avoid unfavorable interactions. A few orientations seemed reasonable, and one illustrative representation is shown in Figure 3 middle. Conversely,

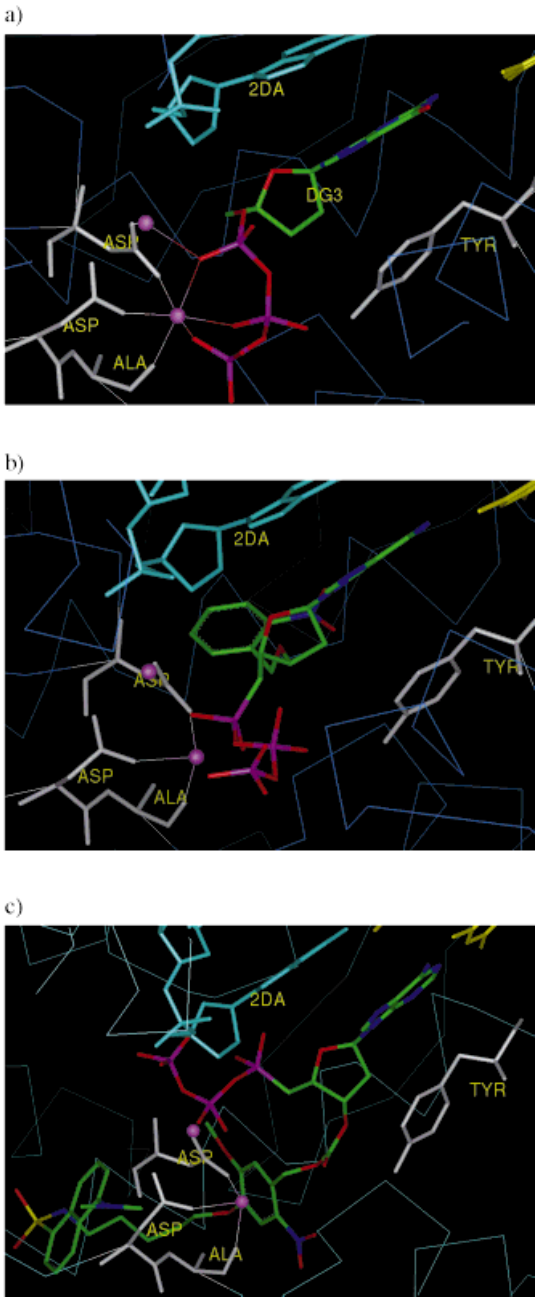


Figure 3. a) Active site of T7 polymerase highlighting two magnesium atoms coordinated to dideoxyguanosine triphosphate with critical side-chains of the protein (grey) and terminus of an encapsulated primer (blue) highlighted; b) as above but with 3′-*O*-(2″-nitrobenzyl)adenosine triphosphate encapsulated; c) as in a) but with the nucleoside triphosphate **18a** encapsulated.

attempts to fit the analogues with fluorescent groups and a methoxy-substituent attached to the aromatic ring were less successful. Figure 3 bottom shows a representation of **18a** in the active site; several interactions were involved that would not be permissible in reality. These docking experiments are too crude to allow detailed conclusions to be formulated, but it does seem clear that 3′-*O*-(2″-nitrobenzyl)adenosine triphosphate **II** is much more easily accommodated in this particular enzyme than any of the analogues prepared in this study.

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.