Docket No. 62239-BZA9/JPW/AC

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant : The Trustees of Columbia University in the City of New York

Inventors : Jingyue Ju et al.

Serial No.: 15/380,284 Examiner: Layla D. Berry

Filed : December 15, 2016 Art Unit: 1673

For : MASSIVE PARALLEL METHOD FOR DECODING DNA AND RNA

30 Rockefeller Plaza
20th Floor
New York, New York 10112

BY EFS
Commissioner for Patents
Alexandria, VA 22313-1450

DECLARATION UNDER 37 C.F.R. §1.132 OF JINGYUE JU, PH.D.

I, Jingyue Ju, Ph.D., declare as follows:

1. I am the first named inventor on the above-identified patent application.

2. I am a Professor of Chemical Engineering and Pharmacology and the Director of the Center for Genome Technology and Biomolecular Engineering at Columbia University in the City of New York ("Columbia").

3. Columbia is the owner (assignee) of rights in the above-identified patent application and has granted a license to the patent application and any patent issued from it to Qiagen.

4. I have been a professor at Columbia since 1999 and my research at Columbia has focused on developing new molecules and methods for DNA sequencing.

5. Prior to 1999, I worked at Incyte Genomics, Inc. and my work there focused on developing and improving DNA sequencing methods for the discovery and sequence identification of genes within the human genome.

6. A copy of my curriculum vitae is attached hereto as **Exhibit A**. As indicated therein, I am a named inventor on 39 issued U.S. patents and an author on 96 scientific publications.

7. Based on my extensive experience and expertise in the research and development of DNA sequencing and related technologies, including the design and synthesis of nucleotide analogues and the preparation and use of labeled nucleotide analogues, I am very well familiar with the level of education, knowledge and experience of persons working in these areas.

8. Based on legal advice concerning the meaning of the phrase a "person of ordinary skill in the art" or "POSA", I think I am well qualified to comment on the appropriate definition of a POSA in the field of DNA sequencing and on issues in this field from the perspective of a POSA as of October 6, 2000. In this regard, my opinion is that a POSA would have been a person with a Ph.D. degree in chemistry, chemical biology, or a related discipline and at least two years postdoctoral experience in the area of DNA sequencing, particularly the design and synthesis of nucleotide analogues for use in DNA sequencing.

9. This Declaration sets forth my opinions as to what a POSA's understanding would have been as of October 2000 concerning various issues raised by the Examiner with respect to the single claim pending in the above-identified patent application.

I. Indefiniteness

A. Meaning of "small"

10. I understand that the Examiner has raised a concern that the term "small", which appears in the claim, is indefinite and that a POSA would not have had an understanding of the meaning of the term in the context of the claim because the

specification does not provide a standard for assessing whether a 3'-O capping group R on the sugar of the claimed nucleotide analogue is "small".

11. In my opinion, a POSA reading the application, particularly paragraphs 6-8 of the published version of the application and Fig. 1 of the application including the brief description of Fig. 1 in paragraph 35, would readily have understood that the application indicates that the standard for assessing whether a 3'-O capping group R is "small" is the ability of the group to fit into the active site of the DNA polymerase shown in Fig. 1 of the patent application.

12. The POSA would further have understood that the 3D structure shown in Fig. 1 is the same as Fig. 6 previously published by Pelletier et al. (*Science*, Vol. 264, June 24, 1994, 1891-1903) and referred to in paragraph 6 of the published version of this patent application. A copy of Pelletier et al. is attached hereto as **Exhibit B.**

13. The POSA would also have known the precise coordinates of the polymerase structure based on the publication of Pelletier et al. (see "References and Notes" 101, page 1903) and the distances between the sugar of the nucleotide analogue and the key amino acids located in the active site of the polymerase (see Table 3, page 1897 of Pelletier et al.).

14. With the information in the patent application and the information available in Pelletier et al. and software available in October 2000, such as Chem3D Pro, the POSA could have readily calculated the space available around the 3' position of a deoxyribose of a nucleotide analogue in the active site of the polymerase.

15. I have performed an analysis of the space available within the active site of the polymerase to accommodate a 3'-O capped

dNTP and prepared a written summary of the result of this analysis attached hereto as **Exhibit C.** As explained in greater detail in the attached Analysis, the available space is approximately 3.7Å in diameter. Given this available space, the POSA would have understood that the term "small", in the context of the claimed nucleotide analogue, assessed by the standard provided in the patent application meant a 3'-O capping group with a diameter less than 3.7Å.

16. Using this standard the POSA could have readily determined that the two examples of 3'-O capping groups listed in the application, Allyl and Methoxymethyl, had diameters of 3.0Å and 2.1Å, respectively, and therefore were "small." (see attached Analysis, Exhibit C)

17. Using this standard, the POSA could also have readily determined which of the many known 3'-O capping groups were "small". Thus, the POSA would have determined that Methylthiomethyl and Azidomethyl, with diameters of 2.4Å and 2.1Å, respectively, were "small" while 2-Nitrobenzyl with a diameter of 5Å was <u>not</u> small (see attached Analysis, Exhibit C).

18. In summary, a POSA reading the above-identified application and relying on information publicly known as of October 2000 would have known that the standard for assessing whether any specific 3'-O capping group in a nucleotide analogue was "small" was whether it had a diameter less than 3.7Å so that it could fit into the active site of the polymerase. Therefore, the meaning of "small" would not have been indefinite.

B. <u>Definition of R is clear</u>

19. With the meaning of "small" well defined, the POSA looking at the structure shown in the pending claim and the definition of

R in the claim would have been able to readily know whether any given 3'-O capping group in a nucleotide analogue came within the scope of the claim. In this regard, prior art references as of October 2000, such as Tsien (WO 91/06678, May 16, 1991) and Stemple (WO 00/53805, September 14, 2000), identify numerous, chemically cleavable 3'-O capping groups which could be readily evaluated to determine whether they were "small" and also meet other conditions of the claim such as the structural features "OR is not a methoxy or an ester group" and R "does not contain a ketone group". Thus, the POSA would have understood with reasonable certainty the definition of R in the context of the patent application as a whole.

C. The scope of Y

20. Y is defined as a chemically cleavable, chemical linker and as shown in the structure shown in the pending claim, Y is attached by covalent bonds at one end to the base of a nucleotide analogue at a specific position and at the other end to a detectable fluorescent moiety. A POSA would have been familiar with many such chemical linkers from the prior art as of October 2000 including such linkers described by Tsien and Stemple. Therefore, a POSA would have readily understood the meaning of Y in the context of the pending claim as a whole read in light of the patent application.

D. Other functional characteristics

21. In the context of the pending claim as a whole read in light of the patent application, a POSA would have readily understood the meaning of the functional characteristics recited in the claim with reasonable certainty. To the POSA who had read the patent application, the scope of the structures encompassed by the claimed nucleotide analogue would have been clear and not indefinite.

II. Written Description

22. Contrary to the Examiner's assertion a POSA would have understood that the inventors of the claimed invention had possession of the invention for at least the following reasons:

a. Only a limited number of 3'-O capping groups meet the standard of "small" along with the other structural and functional features recited in the claim. I estimate the number of such groups would be less than 10 and 2 examples of such groups were provided.

b. The application contains a detailed description, including figures and examples, which shows that the invention was complete.

c. There is literal support in the application for the claim language.

d. The invention was ready for patenting as evidenced by the previous grant of patents from the same specification cited in the rejection for obviousness-type double patenting.

23. Based on the application, including the drawings and examples, a POSA would have immediately envisioned the scope of compounds encompassed by the specific chemical structure recited in the claim, most of which is fixed and with only a few variables which are well defined and not indefinite, for the reasons set forth above.

Stemple's 2-Nitrobenzyl Group And Obviousness

24. As discussed in paragraph 17 above and in the attached Analysis (Exhibit C), the 2-Nitrobenzyl 3'-O capping group referred to in Stemple has a diameter of 5Å and therefore is

not "small" by the standard provided in the patent application.

25. Without the insight into the importance of the 3'-O capping group being small, a POSA, reading Stemple's hypothetical examples and the Metzker 1994 paper (Metzker et al., *Nucleic Acids Research*, Vol. 22, October 11, 1994, 4259-4267) relied upon by Stemple on pages 13, 29 and 31, would have expected that a nucleotide analogue with a 2-Nitrobenzyl 3'-O capping group would be incorporated by a polymerase and therefore must fit in the active site of the polymerase. A copy of the Metzker 1994 paper is attached hereto as **Exhibit D.**

26. Therefore, from Stemple it would not have been obvious to a POSA that the 3'-O capping group must be smaller than a 2-Nitrobenzyl group for a nucleotide analogue containing the group to be incorporated by a polymerase. The subject patent application teaches for the first time that Stemple's prediction based on the Metzker 1994 paper that a nucleotide analogue with a 2-Nitrobenzyl 3'-O capping group would be incorporated by a polymerase was incorrect because the 2-Nitrobenzyl group is not "small". Subsequently Metzker, the senior author of Wu et al., (*Nucleic Acids Research*, Vol. 35, September 18, 2007, 6339-6349) acknowledged in 2007 that the compound actually made and reported in the Metzker 1994 paper was not, in fact, a 3'-O 2-Nitrobenzyl capped nucleotide analogue. Wu et al. further indicated that they had actually made the 3'-O 2-Nitrobenzyl capped nucleotide analogue described in the Metzker 1994 paper (but not, in fact, made in 1994) and had tested it for incorporation by polymerase. The result was that the compound was inactive in DNA polymerase incorporation assays. (see page 6339, lower right, and page 6340, left column, of Wu et al. attached hereto as **Exhibit E.**)

27. Thus, the insight into the critical importance of using a "small" 3'-O capping group for a nucleotide analogue where

"small" means that the 3'-O capping group must fit in the active site of the polymerase and has a diameter less than 3.7Å, would not have been appreciated by a POSA prior to having this patent application available. It would not have been obvious to a POSA in October 2000 from the prior art, such as Metzker and Stemple, which mistakenly teach that a nucleotide analogue with a larger 3'-O capping group, i.e. a 2-Nitrobenzyl group, could be incorporated by polymerase, that it was required to use a nucleotide analogue with a "small" 3'-O capping group R having the additional features recited in the sole pending claim.

I hereby declare that all statements made herein of our own knowledge are true and that all statements made herein on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the subject application or any patent issuing thereon.

Dated: 5/2/2017

Jingyue Ju, Ph.D.

Jingyue Ju, Ph.D.
Samuel Ruben-Peter G. Viele Professor of Engineering
Professor of Chemical Engineering and Pharmacology
Director, Center for Genome Technology and Biomolecular Engineering
Columbia University, Room 806, Northwest Corner Building
550 West 120th Street, New York, NY 10027
Phone: (212) 851-9191; Fax: (212) 851-9330
Email: dj222@columbia.edu

BIOGRAPHICAL SKETCH

NAME	POSITION TITLE
Jingyue Ju	**Professor of Chemical Engineering and Pharmacology, Director, Center for Genome Technology and Biomolecular Engineering, Columbia University**

EDUCATION/TRAINING

INSTITUTION AND LOCATION	DEGREE	YEAR(s)	FIELD OF STUDY
University of California at Berkeley	Postdoctoral	1994-95	Genomics Research
University of Southern California	Ph.D.	1994	Bioorganic Chemistry
Institute of Chemical Physics, Chinese Academy of Sciences	M.S.	1988	Organic Chemistry
Inner Mongolia University, P. R. China	B.S	1985	Chemistry

PROFESSIONAL POSITIONS

2011- Present	Samuel Ruben-Peter G. Viele Professor of Engineering, Columbia University
2010- Present	Professor of Chemical Engineering and Pharmacology, Columbia University
2003-Present	Director, Center for Genome Technology and Biomolecular Engineering, Columbia University
2005-Present	Professor of Chemical Engineering and Head of DNA Sequencing and Chemical Biology, Columbia Genome Center, Columbia University
1999-2005	Associate Professor of Chemical Engineering and Head of DNA Sequencing and Chemical Biology, Columbia Genome Center, Columbia University
1995-1999	Senior Scientist & Director, Chemistry & Assay Development, Incyte Genomics, Inc.

AWARDS

- Samuel Ruben-Peter G. Viele Professor of Engineering, 2011
- Outstanding Chinese Scholar Achievement Award, Columbia University Chinese Students and Scholars Association, 2004
- Visiting Professor of Beijing Capital University of Medical Sciences, 2004
- Packard Fellowship in Science and Engineering, 2001-2006
- Invited Attendee, National Academy of Engineering *8th Annual Frontiers of Engineering Symposium*, September, 2002.
- DOE Human Genome Distinguished Postdoctoral Fellowship, 1994-1995

PROFESSIONAL ACTIVITIES

- Session Chair, *Next Generation Sequencing Conference*, July 7-8, 2011, San Francisco.
- Organizer and Chair, 2008 International Conference on Genomics, Hong Kong, China, 11/2-11/5, 2008.
- Organizer and Chair, 2007 International Conference on Genomics, Hong Kong, China, 10/30-11/2, 2007.

- NIH Review Panel (*Genomics Computational Biology and Technology Study Section*) 2005, 2006
- Session Chair, "*Symposium on New DNA Sequencing Technologies*", International Conference on Genomics, Hangzhou, China, October 23-24, 2006.
- Organizer and Chair, "*New Technologies & Genome Sequencing*" BioArrays-2004-New York Conference, July 26-27, 2004.
- Organizer and Chair, "*New Technology and ToxicoGenomics*" BioArrays-2003-New York Conference, October 1-2, 2003.
- Organizer and Chair, Symposium on Genomics and Chemical Biology, Post 15th International Conference on Phosphorus Chemistry, Beijing, China, August 6-8, 2001.
- Chair, DNA Sequencing Technology Session, Human Genome Meeting, Human Genome Organization (HUGO), 1997.
- NIH Review Panel (*Genetic and Genomic Approaches to Nervous System Function and Dysfunction*) 2004
- NIH Review Panel (*Biophysical and Chemical Sciences*) SBIR/STTR 1997-2001
- NSF Review Panel (*Biochemical Engineering and Biotechnology*) 2001
- DOE Human Genome Program Review Panel (*Advanced DNA Sequencing Technology*) 1998
- Reviewer for the Journal *Proceedings of the National Academy of Sciences, Nature Materials, Nucleic Acids Research, Analytical Chemistry, JACS, Organic Letters, Bioconjugate Chemistry* and *Biotechniques*.

ISSUED US PATENTS

1. United States Patent 9,624,539 (2017) J. Ju, J. Wu, Z. Li. "*DNA Sequencing by Synthesis Using Raman and Infrared Spectroscopy Detection*".
2. United States Patent 9,528,151 (2016) J. Ju, D.H. Kim, L. Bi, Q. Meng, X. Li. "*Four-color DNA Sequencing by Synthesis Using Cleavable Fluorescent Nucleotide Reversible Terminators*".
3. United States Patent 9,297,042 (2016) J. Ju, L. Bi, D.H. Kim, Q. Meng. "*Chemically Cleavable 3'-O-Allyl-dNTP-Allyl-Fluorophore Fluorescent Nucleotide Analogues and Related Methods*".
4. United States Patent 9,255,292 (2016) J. Ju, Q. Meng, D.H. Kim, L. Bi, X. Bai, N. Turro. "*Synthesis of Four-color 3'-O-allyl Modified Photocleavable Fluorescent Nucleotides and Related Methods*".
5. United States Patent 9,250,169 (2016) J. Ju, D.W. Landry, Q. Lin, T. Nguyen, R. Pei, C. Oiu, M.N. Stojanovic "*Selective Capture and Release of Analytes*".
6. United States Patent 9,175,342 (2015) J. Ju, H. Cao, Z. Li, Q. Meng, J. Guo, S. Zhang "*Synthesis of Cleavable Fluorescent Nucleotides as Reversible Terminators for DNA Sequencing by Synthesis*".
7. United States Patent 9,169,510 (2015) J. Ju, J. Wu, D.H. Kim "*Pyrosequencing Methods and Related Compositions*".
8. United States Patent 9,133,511 (2015) J. Ju, Z. Li, J.R. Edwards and Y. Itagaki "*Massive Parallel Method for Decoding DNA and RNA*".
9. United States Patent 9,115,163 (2015) J. Ju, D.H. Kim, J. Guo, Q. Meng, Z. Li, H. Cao "*DNA Sequencing with Non-fluorescent Nucleotide Reversible Terminators and Cleavable Label Modified Nucleotide Terminators*".
10. United States Patent 8,889,348 (2014) J. Ju "*DNA Sequencing by Nanopore Using Modified Nucleotides*".
11. United States Patent 8,796,432 (2014) J. Ju, L. Bi, D.H. Kim, Q. Meng "*Chemically Cleavable 3'-O-Allyl-dNTP-Allyl-Fluorophore Fluorescent Nucleotide Analogues and Related Methods*".
12. United States Patent 8,298,792 (2012) J. Ju, D.H. Kim, L. Bi, Q. Meng and X. Li "*Four-Color DNA Sequencing By Synthesis Using Cleavable Fluorescent Nucleotide Reversible Terminators*".
13. United States Patent 8,088,575 (2012) J. Ju, Z. Li, J.R. Edwards and Y. Itagaki "*Massive Parallel Method for Decoding DNA and RNA*".

14. United States Patent 7,982,029 (2011) J. Ju, Q. Meng, D.H. Kim, L. Bi, X. Bai and N.J. Turro "*Synthesis of Four Color 3'O-allyl, Modified Photocleavable Fluorescent Nucleotides and Related Methods*".
15. United States Patent 7,883,869 (2011) J. Ju, D.H. Kim, L. Bi, Q. Meng and X. Li "*Four-Color DNA Sequencing By Synthesis Using Cleavable Fluorescent Nucleotide Reversible Terminators*".
16. United States Patent 7,790,869 (2010) J. Ju, Z. Li, J.R. Edwards and Y. Itagaki "*Massive Parallel Method for Decoding DNA and RNA*".
17. United States Patent 7,713,698 (2010) J. Ju, Z. Li, J.R. Edwards and Y. Itagaki "*Massive Parallel Method for Decoding DNA and RNA*".
18. United States Patent 7,635,578 (2009) J. Ju, Z. Li, J.R. Edwards and Y. Itagaki "*Massive Parallel Method for Decoding DNA and RNA*".
19. United States Patent 7,622,279 (2009) J. Ju, "*Photocleavable Fluorescent Nucleotides for DNA Sequencing on Chip Constructed by Site-Specific Coupling Chemistry*".
20. United States Patent 7,345,159 (2008) J. Ju, Z. Li, J.R. Edwards and Y. Itagaki "*Massive Parallel Method for Decoding DNA and RNA*".
21. United States Patent 7,074,597 (2006) J. Ju *"Multiplex Genotyping Using Solid Phase Capturable Dideoxynucleotides and Mass Spectrometry*".
22. United States Patent 7,015,000 (2006) R.A. Mathies, A.N. Glazer and J. Ju *"Probes Labeled with Energy Transfer Coupled Dyes*".
23. United States Patent 6,664,079 (2003) J. Ju, Z. Li, J.R. Edwards and Y. Itagaki "*Massive Parallel Method for Decoding DNA and RNA*".
24. United States Patent 6,627,748 (2003) J. Ju, Z. Li, A. Tong and J. Russo "*Combinatorial Fluorescence Energy Transfer Tags and their Applications for Multiplex Genetic Analyses*".
25. United States Patent 6,544,744 (2003) R.A. Mathies, A.N. Glazer and J. Ju *"Probes Labeled with Energy Transfer Coupled Dyes*".
26. United States Patent 6,177,247 (2001) R.A. Mathies, A.N. Glazer and J. Ju "*Detection Methods Using Probes Labeled with Energy Transfer Coupled Dyes for DNA Fragment Analysis*".
27. United States Patent 6,046,005 (2000) J. Ju, K, Konrad. "*Nucleic Acid Sequencing with Solid Phase Capturable Terminators Comprising a Cleavable Linking Group*".
28. United States Patent 6,150,107 (2000) A.N. Glazer, R.A. Mathies, S-C. Hung, and J. Ju "*Methods of Sequencing and Detection Using Energy Transfer Labels with Cyanine Dyes as Donor Chromophores*".
29. United States Patent 6,028,190 (2000) R.A. Mathies, A.N. Glazer and J. Ju "*Probes Labeled with Energy Transfer Coupled Dyes*".
30. United States Patent 5,876,936 (1999) J. Ju "*Nucleic Acid Sequencing with Solid Phase Capturable Terminators*".
31. United States Patent 5,952,180 (1999) J. Ju *"Sets of Energy Transfer Fluorescent Tags and Their Use in Multi-Component Analysis* ".
32. United States Patent 5,869,255 (1999) R.A. Mathies, A.N. Glazer and J. Ju "*Probes Labeled With Energy Transfer Coupled Dyes Exemplified with DNA Fragment Analysis*".
33. United States Patent 5,804,386 (1998) J. Ju *"Sets of Energy Transfer Fluorescent Tags and Their Use in Multi-Component Analysis* ".
34. United States Patent 5,707,804 (1998) R.A. Mathies, A.N. Glazer and J. Ju "*Primers Labeled with Energy Transfer Coupled Dyes for DNA Sequencing*".
35. United States Patent 5,728,528 (1998) R.A. Mathies, A.N. Glazer and J. Ju "*Universal Spacer/Energy Transfer Dyes*".
36. United States Patent 5,853,992 (1998) A.N. Glazer, S-C. Hung, R.A. Mathies, and J. Ju "*Cyanine Dyes with High Absorption Cross Section as Donor Chromophores in Energy Transfer Primers*".
37. United States Patent 5,814,454 (1998) J. Ju *"Sets of Energy Transfer Fluorescent Tags and Their Use in Multi-Component Analysis* ".
38. United States Patent 5,654,419 (1997) R.A. Mathies, A.N. Glazer and J. Ju "*Fluorescent Labels and Their Use in Separations*".
39. United States Patent 5,688,648 (1997) R.A. Mathies, A.N. Glazer and J. Ju "*Probes Labeled with Energy Transfer Coupled Dyes*".

PUBLICATIONS

1. "*Design and Characterization of a Nanopore-Coupled Polymerase for Single-Molecule DNA Sequencing by Synthesis on An Electrode Array*". P. B. Stranges, M. Palla, S. Kalachikov, J. Nivala, M. Dorwart, A. Trans, S. Kumar, M. Porel, M. Chien, C. Tao, I. Morozova, Z. Li, S. Shi, A. Aberra, C. Arnold, A. Yang, A. Aguirre, E. T. Harada, D. Korenblum, J. Pollard, A. Bhat, D. Gremyachinskiy, A. Bibillo, R. Chen, R. Davis, J. J. Russo, C. W. Fuller, S. Roever, J. Ju, G. M. Church. *Proceedings of the National Academy of Sciences USA.* 2016, **113,** E6749–E6756.
2. "*Real-Time Single-Molecule Electronic DNA Sequencing by Synthesis Using Polymer-Tagged Nucleotides on a Nanopore Array*". C. W. Fuller, S. Kumar, M. Porel, M. Chien, A. Bibillo, P. B. Stranges, M. Dorwart, C. Tao, Z. Li, W. Guo, S. Shi, D. Korenblum, A. Trans, A. Aguirre, E. Liu, E. Harada, J. Pollard, A. Bhat, C. Cech, A. Yang, C. Arnold, M. Palla, J. S. Hovis, R. Chen, I. Morozova, S. Kalachikov, J. J. Russo, J. Kasianowicz, R. Davis, S. Roever, G. M. Church, and J. Ju. *Proceedings of the National Academy of Sciences USA.* 2016, **113**, 5233-5238.
3. "*Mathematical Model for Biomolecular Quantification Using Large-Area Surface-Enhanced Raman Spectroscopy Mapping*". F. Bosco, J. Yang, T. Rindzevicius, T.S. Alstrøm, M.S. Schmidt, Q. Lin, J. Ju, A. Boisen. *Royal Society of Chemistry Advances*, 2015, **5**, 85845-85853.
4. "*DNA Sequencing by Synthesis Using 3'-O-azidomethyl Nucleotide Reversible Terminators and Surface-Enhanced Raman Spectroscopic Detection*". M. Palla, W. Guo, S. Shi, Z. Li, J. Wu, S. Jockusch, C. Guo, J.J. Russo, N.J. Turro and J. Ju. *Royal Society of Chemistry Advances*, 2014, **4**, 49342-49346.
5. "*A Microfluidic Device for Multiplex Single-Nucleotide Polymorphism Genotyping*". J. Zhu, C. Qiu, M. Palla, T. Nguyen, J.J. Russo, J. Ju, Q. Lin. *Royal Society of Chemistry Advances*, 2014, **4**, 4269-4277.
6. "*A Strategy to Capture and Characterize the Synaptic Transcriptome*". S.V. Puthanveettil, I. Antonov, S. Kalachikov, P. Rajasethupathy, Y.B. Choi, A.B. Kohn, M. Citarella, F. Yu, K.A. Karl, M. Kinet, I. Morozova, J.J. Russo, J. Ju, L.L. Moroz, E.R. Kandel. *Proceedings of the National Academy of Sciences USA.* 2013, 110, 7464-7469.
7. "*Surface-Enhanced Raman Spectroscopy Based Quantitative Bioassay on Aptamer-Functionalized Nanopillars Using Large-Area Raman Mapping*". J. Yang, M. Palla, F. Bosco, M. Schmidt, T. Rindzevicius, T. Alstrøm, A. Boisen, J. Ju, Q. Lin. *ACS Nano*, 2013, **7**, 5350-5359.
8. "*Mechanism of Flexibility Control for ATP Access of Hepatitis C Virus NS3 Helicase*". M. Palla, C.P. Chen, Y. Zhang, J. Li, J. Ju , J.C. Liao. *J. Biomol. Struct. Dyn.* 2013, **31**, 129-141.
9. "*PEG-labeled Nucleotides and Nanopore Detection for Single Molecule DNA Sequencing by Synthesis*". S. Kumar, C. Tao, M. Chien, B. Hellner, A. Balijepalli, J. W. F. Robertson, Z. Li, J. J. Russo, J. E. Reiner, J. J. Kasianowicz, and J. Ju. *Scientific Reports*. 2012, **2**, 684, 1-8.
10. "*Mitochondrial SNP Genotyping by MALDI-TOF Mass Spectrometry Using Cleavable Biotinylated Dideoxynucleotides*". C. Qiu, S. Kumar, J. Guo, J. Lu, S. Shi, S.M. Kalachikov, J.J. Russo, A.B. Naini, E.A. Schon and J. Ju. *Analytical Biochemistry*, 2012, **427**, 202-210.
11. "*Design and Synthesis of Cleavable Biotinylated Dideoxynucleotides for DNA Sequencing by MALDI-TOF Mass Spectrometry*". C. Qiu, S. Kumar, J. Guo, L. Yu, W. Guo, S. Shi, J. J. Russo, J. Ju. *Analytical Biochemistry*, 2012, **427**, 193-201.
12. "*Fluorescent Hybridization Probes for Nucleic Acid Detection*". J. Guo, J. Ju, N.J. Turro. *Anal. Bioanal. Chem.* 2012, **402**, 3115-3125.
13. "*CdSe/ZnS Core Shell Quantum Dot-based FRET Binary Oligonucleotide Probes for Detection of Nucleic Acids*". Y. Peng, C. Qiu, S. Jockusch, A.M. Scott, Z. Li, N.J. Turro, NJ, J. Ju. *Photochem. Photobiol. Sci.* 2012, **11**, 881-884.
14. "*Quantitative Evaluation of All Hexamers as Exonic Splicing Elements*". S. Ke, S. Shang, S. M. Kalachikov, I. Morozova, L. Yu, J. J. Russo, J. Ju, and L. A. Chasin. *Genome Research.* 2011, **21**, 1360-1374.
15. "*First Generation" Automated DNA Sequencing Technology.* B.E. Slatko, J. Kieleczawa, J. Ju, A.F. Gardner, C.L. Hendrickson, F.M. Ausubel. *Curr. Protoc. Mol. Biol.* 2011, Chapter 7:Unit7.2.

16. "*Cultivation of Enterobacter Hormaechei From Human Atherosclerotic Tissue*". B. Rafferty, S. Dolgilevich, S. Kalachikov, I. Morozova, J. Ju, S. Whittier, R. Nowygrod, E. Kozarov. *J. Atheroscler Thromb.* 2011, **18**, 72-81.
17. "*Translational Control Analysis by Translationally Active RNA Capture/Microarray Analysis*". K. Kudo, Y. Xi, Y. Wang, B. Song, E. Chu, J. Ju, J. J. Russo, and J. Ju. *Nucleic Acids Research.* 2010, 38, e104.
18. "*An Integrated System for DNA Sequencing by Synthesis Using Novel Nucleotide Analogues*". J. Guo, L. Yu, N.J. Turro, J. Ju. *Acc. Chem. Research.* 2010, **43**, 551-563.
19. "*DNA Sequencing by Synthesis Using Novel Nucleotide Analogues*". L. Yu, J. Guo, N. Xu, Z. Li and J. Ju. Handbook of Mutation Detection (Eds. K. Meksem and G. Kahl) 2010, 319-336.
20. "*ReproArrayGTS: A cDNA microarray for identification of reproduction-related genes in the giant tiger shrimp Penaeus monodon and characterization of a novel nuclear autoantigenic sperm protein (NASP) gene*". N. Karoonuthaisiri, K. Sittikankeaw, R. Preechaphol, S. Kalachikov, T. Wongsurawat, U. Uawisetwathana, J. J. Russo, J. Ju, S. Klinbunga, K. Kirtikara. *Comparative Biochemistry and Physiology*, Part D4, 2009, 90–99.
21. "*Centrotemporal sharp wave EEG trait in rolandic epilepsy maps to Elongator Protein Complex 4 (ELP4)*". L.J. Strug, T. Clarke, T. Chiang, M. Chien, Z. Baskurt, W. Li, R. Dorfman, B. Bali, E. Wirrell, S. L. Kugler, D. E. Mandelbaum, S. M. Wolf, P. McGoldrick, H. Hardison, E. J. Novotny, J. Ju, D. A. Greenberg, J. J. Russo, D. K. Pal. *Eur. J. Hum. Genet.* 2009, 17, 1171-1181.
22. "*Four-color DNA Sequencing with 3'-O-modified Nucleotide Reversible Terminators and Chemically Cleavable Fluorescent Dideoxynucleotides*". J. Guo, N. Xu, Z. Li, S. Zhang, J. Wu, D. Kim, M. S. Marma, Q. Meng, H. Cao, X. Li, S. Shi, L. Yu, S. Kalachikov, J. Russo, N.J. Turro, J. Ju. *Proceedings of the National Academy of Sciences USA.* 2008, **105**, 9145–9150.
23. "*Click-functional Block Copolymers Provide Precise Surface Functionality via Spin Coating*". H.R. Rengifo, L. Chen, C. Grigoras, J. Ju, J.T. Koberstein. *Langmuir.* 2008, **24**, 7450-7456.
24. "*Spin-On End-Functional Diblock Copolymers for Quantitative DNA Immobilization*". L. Chen, H.R. Rengifo, C. Grigoras, X. Li, Z. Li, J. Ju, J.T. Koberstein. *Biomacromolecules.* 2008, **9**, 2345-2352
25. "*Genetic Architecture of the Human Tryptophan Hydroxylase 2 Gene: Existence of Neural Isoforms and Relevance for Major Depression*". F. Haghighi, H. Bach-Mizrachi, YY. Huang, V. Arango, S. Shi, AJ. Dwork, G. Rosoklija, HT. Sheng, I. Morozova, J. Ju, JJ. Russo, JJ. Mann. *Mol Psychiatry.* 2008, **13**, 813-820.
26. "*3'-O-Modified Nucleotides as Reversible Terminators for Pyrosequencing*". J. Wu, S. Zhang, Q. Meng, H. Cao, Z. Li, X. Li, S. Shi, D. Kim, N.J. Turro, J. Ju. *Proceedings of the National Academy of Sciences USA.* 2007, **104**, 16462-16467.
27. "*Design and Characterization of Two-dye and Three-dye Binary Fluorescent Probes for mRNA Detection*". A. A. Martí, X. Li, S. Jockusch, N. Stevens, Z. Li, B. Raveendra, S. Kalachikov, I. Morozova, J. J. Russo, D. L. Akins, J. Ju, N. J. Turro. *Tetrahedron*, 2007, **63**, 3591-3600.
28. "*Inorganic-Organic Hybrid Luminescent Binary Probe for DNA Detection Based on Spin-Forbidden Resonance Energy Transfer*" A. A. Martí, C. Puckett, J. Dyer, N. Stevens, S. Jockusch, J. Ju, J. K. Barton, N. J. Turro. *J. Am. Chem. Soc.* 2007, **129**, 8680-8681.
29. "*An Integrated System for DNA Sequencing by Synthesis*" in New High Throughput Technologies for DNA Sequencing and Genomics, *Perspectives in Bioanalysis.* J. Edwards, D. Kim, J. Ju. (*Editor. Keith Mitchelson*) Elsevier. 2007, **2**, 187-205.
30. "*A Mammalian microRNA Expression Atlas Based on Small RNA Library Sequencing*". P. Landgraf, M. Rusu , R. Sheridan , et al., J. Ju, et al., M. Zavolan, T. Tuschl, *Cell.* 2007, **129**, 1401-1414.
31. "*Cellular Cofactors Affecting Hepatitis C Virus Infection and Replication*". G. Randall, M. Panis, J. D. Cooper, et al., J. Ju et al. C. M. Rice. *Proceedings of the National Academy of Sciences USA.* 2007, **104**, 12884-12889.
32. "*Fluorescent Hybridization Probes for Sensitive and Selective DNA and RNA Detection*". A. A. Martí, S. Jockusch, N. Stevens, J. Ju, N. J. Turro. *Acc. Chem. Res.* 2007, **40**, 402-409.

33. "*Quantitative Technologies Establish a Novel microRNA Profile of Chronic Lymphocytic Leukemia*". V. Fulci, S. Chiaretti, M. Goldoni, G. Azzalin, N. Carucci, S. Tavolaro, L. Castellano, A. Magrelli, F. Citarella, M. Messina, R. Maggio, N. Peragine, S. Santangelo, F.R. Mauro, P. Landgraf, T. Tuschl, D.B. Weir, M. Chien, J.J. Russo, J. Ju, R. Sheridan, C. Sander, M. Zavolan, A. Guarini, R. Foa, G. Macino. *Blood.* 2007, **109**, 4944-4951.
34. "*Four-Color DNA Sequencing by Synthesis Using Cleavable Fluorescent Nucleotide Reversible Terminators*". J. Ju, D. Kim, L. Bi, Q. Meng, X. Bai, Z. Li, X. Li, M.S. Marma, S. Shi, J. Wu, J.R. Edwards, A. Romu, N.J. Turro. *Proceedings of the National Academy of Sciences USA.* 2006, **103**, 19635-19640.
35. "*Neuronal Transcriptome of Aplysia: Neuronal Compartments and Circuitry*". L.L. Moroz, J.R. Edwards, S.V. Puthanveettil, A. Kohn , T. Ha, A. Heyland, B. Kudsen, A. Sahni , F. Yu, L. Liu, S. Jezzini, R. Sadreyev, P. Lovell, W. Iannucculli, M. Chen, T. Nguyen, H. Sheng, R. Shaw, S. Kalachikov, Y. Panchin, W. Farmerie, J.J. Russo, J. Ju, E.R. Kandel. *Cell.* 2006, **127**, 1453-1467.
36. "*Design and Synthesis of a Chemically Cleavable Fluorescent Nucleotide 3'-O-Allyl-dGTP-allyl-Bodipy-FL-510 as a Reversible Terminator for DNA Sequencing by Synthesis*". L. Bi, D. H. Kim, J. Ju. *J. Am. Chem. Soc.* 2006, **128**, 2542-2543.
37. "*Combinatorial Fluorescence Energy Transfer Molecular Beacons for Probing Nucleic Acid Sequences*". X. Li, Z. Li, A. Marti, S. Jockusch, N. Stevens, D. L. Akins, N. J. Turro and J. Ju. *Photochemical & Photobiological Sciences.* 2006, **5**, 896-902.
38. "*Design and Synthesis of a Photocleavable Fluorescent Nucleotide 3'-O-Allyl-dGTP-PC-Bodipy-FL-510 as a Reversible Terminator for DNA Sequencing by Synthesis*". Q. Meng, D. H. Kim, X. Bai, L. Bi, N.J. Turro, J. Ju. *J. Org. Chem.* 2006, **71**, 3248-3252.
39. "*Pyrene Binary Probes for Unambiguous Detection of mRNA Using Time-Resolved Fluorescence Spectroscopy*" A. Marti, X. Li, S. Jockusch, Z. Li, B. Raveendra, S. Kalachikov, J. Russo, I. Morozova, S. Puthanveettil, J. Ju, N. J. Turro. *Nucleic Acids Research.* 2006, **34**, 3161-3168.
40. "*Molecular Beacons with Intrinsically Fluorescent Nucleotides*". A. Marti, S. Jockusch, Z. Li, J. Ju, N. J. Turro. *Nucleic Acids Research.* 2006, **34**, e50.
41. "*The Large-Scale Structure of Genomic Methylation Patterns*" R.A. Rollins, F.G. Haghighi, J.R. Edwards, J. Ju & T.H. Bestor. *Genome Research.* 2006, **16**, 157-163.
42. "*Computational Prediction of DNA Methylationn Landscape in the Human Genome*". R. Das, N. Dimitrova, Z.Y. Xuan, R.A. Rollins, F.G. Haghighi, J.R. Edwards, J. Ju, T.H. Bestor, and M.Q. Zhang. *Proceedings of the National Academy of Sciences USA.* 2006, 103, 10713-10716.
43. "*A Novel Class of Small RNAs Bind to MILI Protein in Mouse Testes*". Aravin, D. Gaidatzis, S. Pfeffer, M. Quintana, P. Morris, S. Kuramochi-Miyagawa, T. Nakano, M. Chien, J. J. Russo, J. Ju, R. Sheridan, C. Sander, M. Zavolan & T. Tuschl. *Nature.* 2006, **442**, 203-207.
44. "*Spectroscopic Investigation of a FRET Molecular Beacon Containing Two Fluorophores for Probing DNA/RNA Sequences*". S. Jockusch, A. Marti, N. J. Turro, Z. Li, X. Li, J. Ju, N. Stevens, and D. L. Akins. *Photochemical & Photobiological Sciences.* 2006, **5**, 493-498.
45. "*MassTag Polymerase Chain Reaction Detection of Respiratory Pathogens, Including A New Rhinovirus Genotype, that Caused Influenza-Like Illness in New York State During 2004-2005*". D. Lamson, N. Renwick, V. Kapoor, Z. Liu, G. Palacios, J. Ju, A. Dean, K. St George, T. Briese, W. Ian Lipkin. *Journal of Infectious Diseases.* 2006, **194**, 1398-1402
46. "*MassTag Polymerase Chain Reaction for Differential Diagnosis of Viral Hemorrhagic Fevers*". G. Palacios, T. Briese, V. Kapoor, O. Jabado, Z. Liu, M. Venter, J. Zhai, N. Renwick, A. Grolla, T. W. Geisbert, C. Drosten, J. Towner, J. Ju, J. Paweska, S. Nichol, R. Swanepoel, H. Feldmann, P. Jahrling, W. I. Lipkin. *Emerging Infectious Diseases.* 2006, **12**, 692-695.
47. "*Multiplex Single Nucleotide Polymorphism Detection by Combinatorial Fluorescence Energy Transfer Tags and Molecular Affinity*" A. K. Tong, J. Ju. *Methods in Molecular Biology.* 2006, **335**, 201-214.
48. "*Prognostic Values of microRNAs in Colorectal Cancer*". X. Yaguang, A. Formentini, M. Chien, D.B. Weir, J.J. Russo, J. Ju, M. Kornmann and J. Ju. *Biomarker Insights.* 2006, **2**, 113–121.
49. "*Four-Color DNA Sequencing by Synthesis on Chip Using Photocleavable Fluorescent Nucleotide Analogues*". T. S. Seo, X. Bai, D. H. Kim, Q. Meng, S. Shi, H. Ruparel, Z. Li, N. Turro & J. Ju. *Proceedings of the National Academy of Sciences USA.* 2005, **102**, 5926-5931.

50. "*Design and Synthesis of a 3'-O-Allyl Photocleavable Fluorescent Nucleotide as a Reversible Terminator for DNA Sequencing By Synthesis*". H. Ruparel, L, Bi, Z. Li, X. Bai, D. H. Kim, N. Turro & J. Ju. *Proceedings of the National Academy of Sciences USA*. 2005, **102**, 5932-5937.
51. "*Mass Spectrometry DNA Sequencing*" J. R. Edwards, H. Ruparel, J. Ju. *Mutation Research*. 2005, **573**, 3-12.
52. "*Molecular Engineering Approaches for DNA Sequencing and Analysis*". X. Bai, J.R. Edwards, & J. Ju. *Expert Review of Molecular Diagnostics*. 2005, **5**, 797-808.
53. "*Two-Photon Excitation Induced Fluorescence of a Tri-fluorophore Labeled DNA*" S. Jockusch, Z. Li, J. Ju & N.J. Turro. *Photochemistry and Photobiology*. 2005, **81**, 238-241.
54. "*The Developmental miRNA Profiles of Zebrafish as Determined by Small RNA Cloning*". P.Y. Chen, H. Manninga, K. Slanchev, M. Chien, J.J. Russo, J. Ju, R. Sheridan, C. Sander, M. Zavolan & T. Tuschl. *Genes and Development*. 2005, **11**, 1288-1293.
55. "*Identification of microRNAs of the Herpesvirus Family*". S. Pfeffer, A. Sewer, M. Lagos-Quintana, R. Sheridan, C. Sander, F.A. Grasser, L. F. van Dyk, C. K. Ho, S. Shuman, M. Chien, J.J. Russo, J. Ju, G. Randall, B. D. Lindenbach, C. M. Rice, V. Simon, D. D. Ho, M. Zavolan & T. Tuschl. *Nature Methods*. 2005, **2**, 269-276.
56. "*Diagnostic System for Rapid and Sensitive Differential Detection of Pathogens*". T. Briese, G. Palacios, M. Kokoris, O, Jabado, Z. Liu, N. Renwick, V. Kapoor, I. Casas, F. Pozo, R. Limberger, P. Perez-Brena, J. Ju, W. I. Lipkin. *Emerging Infectious Diseases*. 2005, **11**, 310-313.
57. "*Identification of Virus-Encoded microRNAs*". S. Pfeffer, M. Zavolan, F.A. Grasser, M. Chien, J.J. Russo, J. Ju, B. John, A.J. Enright, D. Marks, C. Sander and T. Tuschl. *Science*. 2004, **304**, 734-736.
58. "*Photocleavable Fluorescent Nucleotides for DNA Sequencing on a Chip Constructed by Site-Specific Coupling Chemistry*". T. S. Seo, X. Bai, Z. Li, H. Ruparel, N. Turro & J. Ju. *Proceedings of the National Academy of Sciences USA*. 2004, **101**, 5488-5493.
59. "*Thirty Fold Multiplex Genotyping of p53 Gene Using Solid Phase Capturable Dideoxynucleotides and Mass Spectrometry*". S. Kim, M. Ulz, T. Nguyen, C. Li, B. Tycko, and J. Ju. *Genomics*. 2004, **83**, 924-931.
60. "*Digital Detection of Genetic Mutations Using SPC-Sequencing*" H. Ruparel, M. E. Ulz, S. Kim, and J. Ju. *Genome Research*. 2004, **14**, 296-300.
61. "*Design and Synthesis of a Photocleavable Biotinylated Nucleotide for DNA Analysis by Mass Spectrometry*" X. Bai, S. Kim, Z. Li, N. J. Turro & J. Ju. *Nucleic Acids Research*. 2004, **32**, 535-541
62. "*1,3-Dipolar Cycloaddition of Azides with Electron-deficient Alkynes Under Mild Condition in Water*". Z. Li, T. S. Seo, J. Ju. *Tetrahedron Letters*. 2004, **45**, 3143-3146.
63. "*The Genomic Sequence of the Accidental Pathogen Legionella Pneumophila*". M. Chien, I. Morozova, S. Shi, H. Sheng, J. Chen, S.M. Gomez, G. Asamani, K. Hill, J. Nuara, M. Feder, J. Rineer, J.J. Greenberg, V. Steshenko, S.H. Park, B. Zhao, E. Teplitskaya, J.R. Edwards, S. Pampou, A. Georghiou, I.C. Chou, W. Iannuccilli, M.E. Ulz, D.H. Kim, A. Geringer-Sameth, C. Goldsberry, P. Morozov, S.G. Fischer, G. Segal, X. Qu, A. Rzhetsky, P. Zhang, E. Cayanis, P.J. De Jong, J. Ju, S. Kalachikov, H.A. Shuman, J.J. Russo. *Science*. 2004, **305**, 1966-1968.
64. "*A Photocleavable Fluorescent Nucleotide for DNA Sequencing and Analysis*". Z. Li, X. Bai, H. Ruparel, S. Kim, N. Turro & J. Ju. *Proceedings of the National Academy of Sciences USA*. 2003, **100**, 414-419.
65. "*Digital Genotyping Using Molecular Affinity and Mass Spectrometry*" S. Kim, H. D. Ruparel, T. C. Gilliam, and J. Ju. *Nature Reviews Genetics*. 2003, **4**, 1001-1008.
66. "*Photocleavage of a 2-Nitrobenzyl Linker Bridging a Fluorophore to the 5' End of DNA*". X. Bai, Z. Li, S. Jockusch, N. Turro & J. Ju. *Proceedings of the National Academy of Sciences USA*. 2003, **100**, 409-413.
67. "*Multiplex Genotyping of the Human β2-Adrenergic Receptor Gene Using Solid Phase Capturable Dideoxynucleotides and Mass Spectrometry*". S. Kim, S. Shi, T. Bonome, J. R. Edwards, J. Russo, and J. Ju. *Analytical Biochemistry*. 2003, **316**, 251-258
68. "*Site-Specific Fluorescent Labeling of DNA Using Staudinger Ligation*". T. S. Seo, C. Wang, Z. Li, H. Ruparel, and J. Ju. *Bioconjugate Chemistry*. 2003, **14**, 697-701

69. "*Identification of Critical Residues in a Class C β-Lactamase Using Combinatorial Scanning Mutagenesis*". S. D. Goldberg, W. Iannuccilli, T. Nguyen, J. Ju, and V. W. Cornish. *Protein Science*. 2003, **12**, 1633-1645.
70. "*Solid Phase Capturable Dideoxynucleotides for Multiplex Genotyping Using Mass Spectrometry*" S. Kim, J. R. Edwards, L. Deng, W. Chung, and J. Ju. *Nucleic Acids Research*. 2002, **30**, e78 (p1-6).
71. "*Click Chemistry to Construct Fluorescent Oligonucleotides for DNA Sequencing*" T. S. Seo, Z. Li, H. Ruparel, and J. Ju. *Journal of Organic Chemistry*. 2002, **68**, 609-612.
72. "*Single-nucleotide Polymorphism Detection by Combinatorial Fluorescence Energy Transfer Tags and Biotinylated Dideoxynucleotides*". A. K. Tong, & J. Ju. *Nucleic Acids Research*. 2002, **30**, e19 (p1-7).
73. "*Synthesis of Oligodeoxyribonucleoside Phosphorothioates Using Lawesson's Reagent for the Sulfur Transfer Step*" J. Ju. & C. McKenna, *Bioorganic & Medicinal Chemistry Letters*. 2002, **12**, 1643-1645.
74. "*DNA Sequencing with Solid Phase Capturable Dideoxynucleotides and Energy Transfer Primers*" J. Ju. *Analytical Biochemistry*. 2002, **309**, 35-39.
75. "*Combinatorial Fluorescence Energy Transfer Tags: New Molecular Tools for Genomics Applications*". A. K. Tong, Z. Li, J. Ju. *Journal of Quantum Electronics Special Issue for Biomedical Applications*. 2002, **38**, 110-121.
76. "*Fluorescence Energy Transfer Reagents for DNA Sequencing and Analysis*" in *Structure and Dynamics of Confined Spaces (NATO Science Series, High Technology)* J. Ju. 2002, **87**, 371-383. eds. J.J. Kasianowicz, M.S.Z. Kellermayer, and D.W. Deamer (Kluwer Academic Publishers, Netherlands).
77. "*Combinatorial Fluorescence Energy Transfer Tags for Multiplex Biological Assays*" A. K. Tong, Z. Li, G. S. Jones, J. J. Russo, J. Ju. *Nature Biotechnology*. 2001, **19**, 756-759.
78. "*DNA Sequencing Using Biotinylated Dideoxynucleotides and Mass Spectrometry*" J.R. Edwards, Y. Itagaki and J. Ju. *Nucleic Acids Research*. 2001, **29**, e104 (p1-6).
79. "*Triple Fluorescence Energy Transfer in Covalently Tri-Chromophore-Labeled DNA*" A. Tong, S. Jockusch, Z. Li, H-R. Zhu, D. Akins, N. J. Turro, and J. Ju. *J. Am. Chem. Soc.* 2001, **123**, 12923-12924.
80. "*DNA Sequencing*" in *Short Protocols in Molecular Biology, 4th Ed".* B.E. Slatko, L.M. Albright, S. Tabor and J. Ju. (1999) eds. Ausubel, F.M. *et al.* (John Wiley & Sons, Inc, New York and London), pp. 7-25 to 7-37. (Book Chapter).
81. "*Energy Transfer Fluorescent Primers: State-of-the-art in High-throughput DNA Sequencing*" J. Ju, A.N. Glazer and R.A. Mathies. *Genome Digest*, 1997, 8-9.
82. "*Cassette Labeling for Facile Construction of Energy Transfer Fluorescent Primers*" J. Ju, A.N. Glazer and R.A. Mathies. *Nucleic Acids Research*, 1996, **24**, 1144-1148.
83. "*Energy Transfer Primers: A New Fluorescence Labeling Paradigm for DNA Sequencing and Analysis*" J. Ju, A.N. Glazer and R.A. Mathies. *Nature Medicine*, 1996, **2**, 246-249.
84. "*DNA Sequencing Using a Four-Color Confocal Fluorescence Capillary Array Scanner*" I. Kheterpal, J. Scherer, S. M. Clark, A. Radhakrishnan, J. Ju, C. L. Ginther, G. F. Sensabaugh and R. A. Mathies. *Electrophoresis*, 1996, **17**, 1852-1859.
85. "*High-Resolution Capillary Array Electrophoretic Sizing of Multiplexed Short Tandem Repeat Loci Using Energy-Transfer Fluorescent Primers*" Y. Wang, J.M. Wallin, J, Ju, G.F. Sensabaugh and R.A. Mathies. *Electrophoresis*, 1996, **17**, 1485-1490.
86. "*Cyanine Dyes with High Absorption Cross Section as Donor Chromophores in Energy Transfer Primers*" S-C, Hung, J. Ju, A.N. Glazer and R.A. Mathies. *Analytical Biochemistry*, 1996, **243**, 15-27.
87. "*Energy Transfer Primers with 5- or 6-Carboxyrhodamine-6G as Acceptor Chromophores* " S-C, Hung, J. Ju, A.N. Glazer and R.A. Mathies. *Analytical Biochemistry*, 1996, **238**, 165-70.
88. "*Energy Transfer Fluorescent Dye-Labeled Primers for DNA Sequencing and Analysis*" J. Ju, C. Ruan, C.W. Fuller, A.N. Glazer and R.A. Mathies. *Proc. Natl. Acad. Sci. U.S.A.* 1995, **92**, 4347-4351.

89. "*Design and Synthesis of Energy Transfer Fluorescent Dye-Labeled Oligonucleotide Primers and Their Application for DNA Sequencing and Analysis* " J. Ju, I. Kheterpal, J. Scherer, C. Ruan, C. Fuller, A. N. Glazer and R. A. Mathies. *Analytical Biochemistry,* 1995, **231**, 131-140.
90. "*Rapid Sizing of Short Tandem Repeat Alleles Using Energy Transfer Fluorescent Primers and Capillary Array Electrophoresis*" Y. Wang, J. Ju, B. A. Carpenter, J. M. Atherton, R. A. Mathies and G. F. Sensabaugh. *Anal. Chem.* 1995, **67**, 1197.
91. "*Troika Acids: Synthesis, Structure and Fragmentation Pathways of Novel a-(Hydroxyimino)-Phosphonoacetic Acids*" B. A. Kashemirov, J. Ju, R. Bau and C. E. McKenna. *J. Am. Chem. Soc.* 1995, **117**, 7285-7286.
92. "*Some Phosphonic Acid Analogs as Inhibitors of Pyrophosphate-dependent Phospho-fructokinase, A Novel Target in Toxoplasma Gondii*" Z. Peng, J. M. Mansour, F. Araujo, J. Ju, C. E. McKenna and T. E. Mansour, *Biochemical Pharmacology* 1995, **49**, 105-113.
93. "*E/Z Stereoisomer Assignment by ^{13}C NMR in Trifunctional Phosphonate a-Oximes and a-Arylhydrazones*" C. E. McKenna, B. A. Kashemirov, J. Ju. *J. Chem. Soc. Chem. Commun.*, 1994, 1211-1212.
94. "*Synthesis and HIV-1 Reverse Transcriptase Inhibition Activity of Functionalized Pyrophosphate Analogues*" C.E. McKenna, A. Khare, J. Ju, Z. Li, G. Duncan, Y. Cheng and R. Kilkuskie. *Phosphorus Sulfur.* 1993, **76**, 139-142.
95. "*Simple and Conjugate Bifunctional Thiophosphonates: Synthesis and Potential as Anti-Viral Agents*" C.E. McKenna, Z. Li, J. Ju, P-T. Pham, R. Kilkuskie, T-L. Loo and J. Straw. *Phosphorus Sulfur.* 1993, **74**, 469-470.
96. "*Stereoselective Aldol Coupling of Cobalt-Complexed Alkynyl Aldehydes.*" J. Ju, B.R. Reddy, M. Khan and K.M. Nicholas. *Journal of Organic Chemistry,* 1989, **54**, 5426-5428.

Structures of Ternary Complexes of Rat DNA Polymerase β, a DNA Template-Primer, and ddCTP

Huguette Pelletier, Michael R. Sawaya, Amalendra Kumar, Samuel H. Wilson, Joseph Kraut

Two ternary complexes of rat DNA polymerase β (pol β), a DNA template-primer, and dideoxycytidine triphosphate (ddCTP) have been determined at 2.9 Å and 3.6 Å resolution, respectively. ddCTP is the triphosphate of dideoxycytidine (ddC), a nucleoside analog that targets the reverse transcriptase of human immunodeficiency virus (HIV) and is at present used to treat AIDS. Although crystals of the two complexes belong to different space groups, the structures are similar, suggesting that the polymerase-DNA-ddCTP interactions are not affected by crystal packing forces. In the pol β active site, the attacking 3′-OH of the elongating primer, the ddCTP phosphates, and two Mg^{2+} ions are all clustered around Asp^{190}, Asp^{192}, and Asp^{256}. Two of these residues, Asp^{190} and Asp^{256}, are present in the amino acid sequences of all polymerases so far studied and are also spatially similar in the four polymerases—the Klenow fragment of *Escherichia coli* DNA polymerase I, HIV-1 reverse transcriptase, T7 RNA polymerase, and rat DNA pol β—whose crystal structures are now known. A two-metal ion mechanism is described for the nucleotidyl transfer reaction and may apply to all polymerases. In the ternary complex structures analyzed, pol β binds to the DNA template-primer in a different manner from that recently proposed for other polymerase-DNA models.

DNA replication (*1*) is a highly complex biological process, even for a relatively simple organism such as *Escherichia coli*. During replication, the double helical DNA molecule is unwound, and the two resultant single strands of DNA act as templates to guide the synthesis, one complementary base at a time, of antiparallel primer strands. Although many auxiliary proteins such as ligases, helicases, and topoisomerases are usually involved, the chemical reaction at the core of DNA replication, the nucleotidyl transfer reaction, is catalyzed by DNA polymerases and may be depicted as follows:

$$\text{Template—primer}_n + \text{dNTP} \rightarrow$$
$$\text{template-primer-dNMP}_{n+1} + \text{PP}_i$$

where dNTP (2′-deoxyribonucleoside 5′-triphosphate) represents any one of four deoxynucleotides (dATP, dGTP, dCTP, and dTTP), and dNMP and PP_i represent 2′-deoxyribonucleoside 5′-monophosphate and pyrophosphate, respectively (Fig. 1).

Inhibition of a polymerase that effects genomic replication can be fatal to an organism. In a common type of polymerase inhibition, 2′,3′-dideoxynucleotides (ddNTPs) act as chain terminators of the primer strand. The ddNTPs differ from their cellular dNTP counterparts by the absence of an attacking 3′-hydroxyl group (3′-OH) (Fig. 2) and therefore, once a dideoxynucleotide is successfully incorporated into a growing primer strand, there can be no further incorporation of subsequent nucleotides. A well-known example of this kind of inhibition involves HIV-1 reverse transcriptase (RT), which is the polymerase responsible for the replication of the HIV genome. 3′-azido-2′,3′-dideoxythymidine (AZT), 2′,3′-dideoxyinosine (ddI), and 2′,3′-dideoxycytidine (ddC) are all anti-HIV drugs (*2*, *3*) that become potent chain termination inhibitors of RT after they are converted by cellular kinases (*4*, *5*), in vivo, to their corresponding nucleoside 5′-triphosphates, AZT-TP, ddATP (*6*), and ddCTP, respectively. In that all polymerases probably share a common catalytic mechanism, it is not surprising that some toxic effects of these drugs have been attributed to inhibition of host-cell polymerases, perhaps including the pol β described here (*7*–*9*). Therefore, a detailed understanding of the nucleotidyl transfer reaction, as well as the mechanism of inhibition of viral and host cell polymerases by nucleoside analogs, may lead to the design of more potent and less toxic HIV-1 RT inhibitors for use in the treatment of AIDS.

Despite limited sequence similarity to the Klenow fragment (KF) of *E. coli* DNA pol I [the only other polymerase for which a crystal structure (*10*) was known at the time], the crystal structure determinations of HIV-1 RT (*11*, *12*) revealed a common polymerase fold consisting of three distinct subdomains (designated fingers, palm, and thumb because of the resemblance to a hand) forming an obvious DNA binding channel. The strongest structural overlap between KF and RT comprised a trio of carboxylic acid residues located in the palm subdomain (*11*, *12*). These observations led to the hypothesis that perhaps all polymerases share a common nucleotidyl transfer mechanism centered around the highly conserved carboxylic acid residues (*11*). Strengthening this argument somewhat was the subsequent crystal structure determination of an RNA polymerase (RNAP) from bacteriophage T7, which showed strong structural similarities with KF (*13*). How-

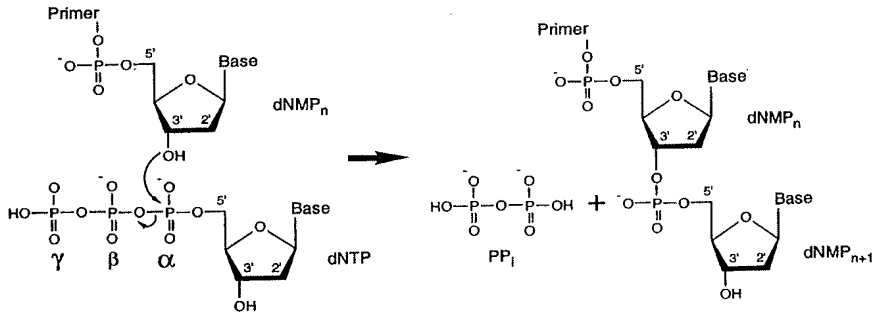


Fig. 1. The nucleotidyl transfer reaction. The 3′-OH group of the terminal dNMP on the primer strand attacks the 5′–α phosphate of an incoming dNTP, and a newly formed phosphodiester linkage results in elongation of the primer strand by one dNMP. After release of pyrophosphate (PP_i), the catalytic cycle is complete and the 3′-OH group of the newly incorporated dNMP is now ready to attack yet another incoming dNTP. Only the 3′ end of a primer is extended so that DNA polymerization is said to proceed in a 5′ to 3′ direction. If the polymerase molecule does not release the template-primer before incorporation of a second dNMP, the mode of DNA synthesis is said to be "processive", but if the polymerase releases the template-primer after each successive incorporation of a dNMP, the mode of DNA synthesis is said to be "distributive".

H. Pelletier, M. R. Sawaya, and J. Kraut are in the Department of Chemistry, University of California, San Diego, CA 92093–0317, USA. A. Kumar and S. H. Wilson are at the Sealy Center for Molecular Science, University of Texas Medical Branch, Galveston, TX 77555–1051, USA.

ever, because structural evidence for a common nucleotidyl transfer mechanism has so far been limited to comparisons among polymerases from a bacterium (KF), a virus (RT), and a phage (RNAP), perhaps the most convincing evidence for this hypothesis is provided by the crystal structure determination of a eukaryotic polymerase, rat DNA pol β (*14*). Sequence alignments show that pol β is so distantly related, even from its eukaryotic relatives, polymerases α, γ, δ, and ε, that it stands in a class of its own along with only one other polymerase, terminal deoxynucleotidyltransferase (TdT) (*15*). The crystal structure of pol β nevertheless revealed a polymerase fold consisting of palm, fingers, and thumb (along with an additional 8-kD domain attached to the fingers), and the most striking structural similarity with its distant relatives, KF, RT, and RNAP, is a portion of the palm that bears the highly conserved carboxylic acid residues (*14*). This suggests that despite the large differences in size (pol β, at 39 kD, is the smallest polymerase known), in function [although pol β may play a role in DNA replication (*16, 17*), its primary function is in DNA repair (*18–20*)], and in fidelity [pol β is the most error prone eukaryotic polymerase studied to date (*21, 22*)], pol β probably shares a common nucleotidyl transfer catalytic mechanism with all other polymerases.

Taking advantage of the chain termination method of polymerase inhibition with ddNTPs, we have succeeded in growing crystals of rat pol β complexed with two pseudo substrates, namely, (i) a DNA template-primer in which the 3′ end of the primer has been "terminated" by ddCMP, and (ii) ddCTP. In preparation for crystallization experiments, we mixed pol β with the DNA template-primer shown below and a large excess of ddCTP on the assumption that, prior to crystallization, the following reaction would occur:

```
3'GCCGCGGGAAA5'
5'CGGCGC      3' + ddCTP →

   3'GCCGCGGGAAA5'
   5'CGGCGCC_dd   3' + PP_i
```

where C_{dd} is the newly incorporated ddCMP. If pol β then were to try to incorporate another nucleotide onto the primer terminus, a second nucleotidyl transfer reaction could not occur because the recently incorporated ddCMP lacks a 3′-OH group. This should result in a pseudo Michaelis-Menten ternary complex in which both "substrates" are present, namely, a nonreactive template-primer and a nucleoside triphosphate. Crystals were obtained, and the subsequent structure determinations revealed that this must have been what happened. Electron density maps showed a primer strand that was seven nucleotides long, although we started with a primer that was only six nucleotides in length, and the 3′ (deoxy) terminus of the primer was positioned next to strong electron density resembling a nucleoside triphosphate, presumably ddCTP.

Such a detailed view of the active site in the ternary complex allows us to propose a two-metal ion mechanism for the nucleotidyl transfer reaction that is similar, in many ways, to the two-metal ion mechanism previously proposed for another type of phosphoryl transfer reaction—the exonuclease reaction of the 3′→5′ exonuclease of *E. coli* DNA pol I (*23, 24*). Our proposed nucleotidyl transfer mechanism probably applies to all polymerases, but when we attempt to extend that mechanism to the other three polymerases—KF, RT, and RNAP—for which the crystal structures are known, a problem arises: in our structures, pol β is bound to the DNA in a manner that differs from the recently proposed polymerase-DNA models for all three of these polymerases.

Crystallizations and preliminary diffraction studies. Recombinant rat DNA pol β (*25*) was expressed in *E. coli* and purified as described (*26*). After purification, the protein was washed three times in a microconcentrator (Centricon-10, Amicon) with a buffer solution (10 mM ammonium sulfate, 0.1 M tris, pH 7.0), then concentrated to 20 mg/ml and stored at −80°C in sealed Eppendorf tubes (120-μl portions). Prior to crystallization, a protein-DNA-ddCTP sample was prepared at room temperature; approximately 1.2 mg of the 11-nucleotide (nt) template and 0.8 mg of the 6-nt primer (*27*) were dissolved in 240 μl of a buffer solution (20 mM $MgCl_2$, 0.1 M MES, pH 6.5) and the mixture was left in a sealed Eppendorf tube for 1 hour to allow annealing of the template-primer (*28*). Two portions of pol β (240-μl of a solution containing 20 mg/ml) were then thawed and added, and the protein-template-primer sample was allowed to stand for an additional hour. A 4-μmol sample of ddCTP (in 40 μl of H_2O) (*29*) was the last component to be added, resulting in a reaction mixture containing pol β at approximately 10 mg/ml, 10 mM $MgCl_2$, and an excess of template:primer:ddCTP in molar ratios of 3:4:30, respectively, relative to the amount of protein. The reaction, nucleotidyl transfer of ddCMP to the primer 3′ terminus, was allowed to proceed for 2 hours before crystallization trays were set up (*30*).

Two different crystal forms were observed, depending on the concentration of lithium sulfate in the reservoir solution. One crystal form, obtained with lithium sulfate concentrations from 40 to 75 mM, was hexagonal and grew to dimensions of 0.8 by 0.8 by 0.6 mm in about 2 weeks. These crystals belong to space group $P6_1$ ($a = b = 94.9$, $c = 117.6$ Å), with one complex molecule per asymmetric unit. Under similar conditions, but at lithium sulfate concentrations from 75 to 150 mM, platelike crystals grew to dimensions of

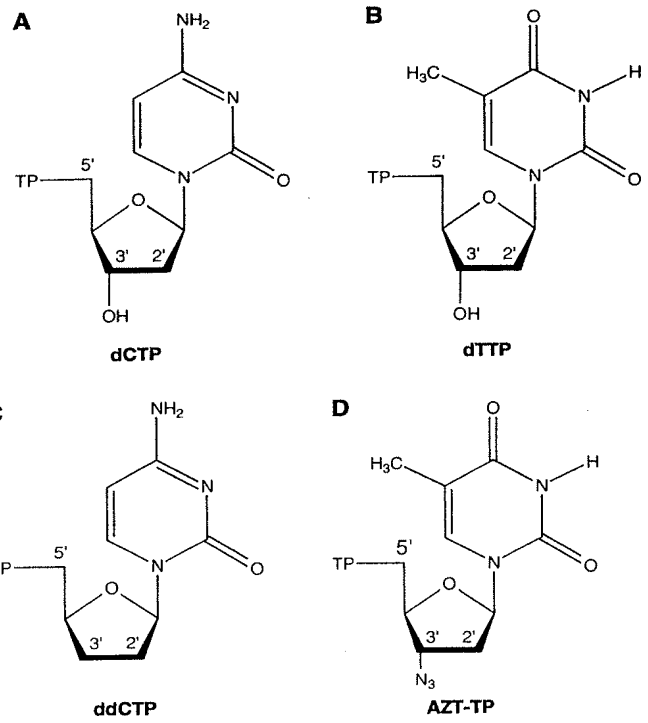


Fig. 2. Two normal cellular nucleotides, (**A**) 2′-deoxycytidine 5′-triphosphate (dCTP) and (**B**) 2′-deoxythymidine 5′-triphosphate (dTTP), and their anti-HIV (drug) counterparts (**C**) 2′,3′-dideoxycytidine 5′-triphosphate (ddCTP) and (**D**) 3′-azido-2′,3′-dideoxythymidine 5′-triphosphate (AZT-TP). The triphosphate moiety, which is linked via a phosphoester bond to the 5′ carbon of the ribose, is designated TP.

1.0 by 0.6 by 0.2 mm in a few days. These crystals belong to space group $P2_1$ (a = 106.3, b = 56.8, c = 86.7 Å, and β = 106.4°) and there are two complex molecules per asymmetric unit. Often both crystal forms grew in the same drop, and the crystals on which data were collected (Table 1) were grown at the same concentration of lithium sulfate, 75 mM. The unusually large excess of ddCTP (1:30 molar ratio) was required in order to obtain the $P6_1$ crystals, but the $P2_1$ crystals could be grown under much lower ddCTP excesses (1:10 molar ratio). Extreme purity of all components in the crystallization medium, particularly the DNA samples (*27*), seemed to be an absolute requirement for growing both types of ternary complex crystals.

Attempts were made to obtain ternary complex crystals of rat pol β, a DNA template-primer, and AZT-TP (Fig. 2D) (*31*) under similar conditions, even though incorporation of AZT-MP would result in a mismatched base pair (of a G-T type) at the primer terminus (*22*). Orthorhombic crystals grew in space group $P2_12_12$ (a = 188.4, b = 67.7, c = 39.1 Å) with one pol β molecule in the asymmetric unit. A 4.0 Å data set was collected and preliminary structural studies (*32*) showed that, because of crystal packing, it was not possible for the template-primer to occupy the pol β binding channel. Failure of pol β to form a tight complex with the DNA template-primer under these conditions might be attributed to steric hindrance by the azido group of a newly incorporated AZT-MP on the primer terminus.

Efforts to obtain a binary complex of pol β and a DNA template-primer alone (neither ddCTP nor AZT-TP) resulted in crystals that grew under much different conditions, but were nonetheless isomorphous with the $P2_12_12$ (AZT-TP) crystals mentioned above. Failure of pol β to bind to the DNA in this case could be due to the higher salt concentration of the crystallization medium (about 250 mM salt compared to 75 mM). Because one crystallization medium contained AZT-TP and the other did not, we calculated $F_{O(AZT\text{-}TP)} - F_{O(apo)}$, α_c, difference Fourier maps to see whether an AZT-TP binding site could be located. Strong electron density was observed in an area of the map adjacent to Arg^{149}, which is near the catalytically important residues, Asp^{190}, Asp^{192}, and Asp^{256}. This pol β–AZT-TP binary complex which, as discussed below, is probably not catalytically relevant, is similar to a KF-dNTP binary complex in which the dNTP bound to Arg residues near the catalytically important carboxylic acid residues of KF (*33*).

Human pol β, which has been cloned and expressed (*34*, *35*) in a manner similar to that of rat pol β, shares more than 95 percent sequence similarity with rat pol β, so it was somewhat surprising when attempts to obtain ternary complex crystals of human pol β, a DNA template-primer, and ddCTP under the same conditions described above for the rat enzyme resulted in crystals that grew in two previously unobserved orthorhombic crystal forms. One form has unit cell parameters a = 158, b = 108, c = 60 Å, with probably two complex molecules in the asymmetric unit, but the crystals diffract only to about 5 Å resolution. In contrast, the second crystal form diffracts quite well (to about 3.3 Å), but its unit cell parameters of a = 465, b = 168, c = 56 Å are so large that special data collection techniques would be required.

Structure determination and refinement. Data collection and refinement statistics for the structure determinations of the two ternary complexes of rat pol β, a DNA template-primer, and ddCTP are listed in Table 1. Structure solutions utilized the refined atomic coordinates of the high resolution (2.3 Å) structure of the 31-kD domain of rat pol β (*14*). The molecular replacement programs of XPLOR (*36*) gave clear rotation solutions for the 31-kD domain of both ternary complexes, but only after results from classical cross-rotation searches had been filtered through the Patterson-correlation (PC) refinement steps (*37*). PC refinement techniques were particularly powerful for our structure determinations because independent rigid body movements of the fingers, palm, and thumb subdomains of the 31-kD domain could be allowed during PC refinement of the rotation searches. Results from subsequent translation searches gave solutions for the $P6_1$ complex structure that were, in general, at higher peak height to background ratios than translation solutions for the $P2_1$ complex structure, but the highest translation peaks in both cases nevertheless were the correct solutions (*38*).

The 31-kD partial structure solutions obtained by molecular replacement techniques were subjected to rigid body refinement by XPLOR (*36*), where the entire 31-kD domain was first allowed to move as a rigid body, then later, the fingers, palm, and thumb subdomains were allowed to move as independent rigid bodies simultaneously. Typical R factors at this stage were about 50 percent. After subsequent positional refinement with the least squares program package TNT (*39*) had lowered the R factors of the partial solutions to about 45 percent, we calculated $F_0 - F_c$, α_c, difference Fourier maps that revealed clear electron density for many of the backbone phosphates of a double-stranded DNA molecule as well as the three phosphates of a ddCTP nucleotide, and even portions of the 8-kD domain were evident at this early stage. Cycles of model building and least squares refinement improved the electron density for the rest of the DNA as well as the 8-kD domain for both complex structures, and once the R factors had dropped below 30 percent, refinement of individual isotropic temperature factors also improved the maps and facilitated refinement.

Although we were unable to discern the DNA base sequences at these resolutions, the directionality (5′ → 3′) of the DNA strands was evident early in our modeling efforts, hence we knew that the 3′ terminus of either the template strand or the primer strand was positioned at the pol β active site. What ultimately distinguished the template from the primer was that we were

Table 1. Data collection and refinement statistics. X-ray diffraction data were collected on a multiwire area detector (*98*) (San Diego Multiwire Systems) with monochromatized CuKα radiation (Rigaku rotating anode x-ray generator), and intensity observations for each data set were processed with a local UCSD Data Collection Facility software package (*99*). Reflections from 20 Å to the maximum resolution were included in all least squares refinement steps. The final structures for both complexes include all residues, with the exception of residues 1 to 8 of the disordered NH_2-terminus and residues 246 to 248 of a disordered surface loop. There are a few missing side chain atoms in both coordinate sets that are mainly in lysine and arginine residues of the 8-kD domain. Omit maps were used to confirm the modeling of the DNA template-primer, the ddCTP nucleotide, and the cis-peptide bond between Gly^{274} and Ser^{275}.

Data collection							Refinement			
Space group	d_{min} (Å)	$\frac{I}{\sigma}$*	Reflections		Completeness (%)	R_{sym}†	Atoms‡	rms deviation§		Final R‖
			Total	Unique				Bond (Å)	Angle (°)	
$P6_1$	2.9	1.8	53,583	13,281	99	0.087	2914	0.020	2.9	0.193
$P2_1$	3.6	1.8	25,046	10,650	96	0.059	5753	0.018	2.9	0.199

*Average ratio of observed intensity to background in the highest resolution shell of reflections. †$R_{sym} = \Sigma|I_{obs} - I_{avg}|/\Sigma I_{avg}$. ‡The number of nonhydrogen atoms includes 31 and 4 water oxygens for the $P6_1$ and the $P2_1$ structures, respectively. §The rms bond and rms angle values are the deviations from ideal values of the bond lengths and bond angles in the final model. ‖Final $R = \Sigma|F_{obs} - F_{calc}|/\Sigma F_{obs}$, including all data between 20 Å and the maximum resolution.

able to model in seven nucleotides for one of the DNA strands and at least 8 nt for the second DNA strand. Provided that no unexpected side reactions had occurred during crystallization, we knew that the primer could be no longer than 7 nt, and therefore the DNA strand containing 8 nt was designated the template. We then concluded that the first three bases of the template (AAA) are disordered in both crystal structures. This interpretation of the data is in agreement with the idea that the 3′ terminus of the primer should be positioned at the polymerase active site. Analysis of the DNA in our refined structures with the program CURVES II (*40*) indicated that the DNA is predominantly B form. Our DNA may have some A-DNA characteristics, however, in that the minor groove width appears to increase as the DNA approaches the pol β active site; the section of the double-stranded DNA that is removed from the active site and protrudes into solution is characteristic of B-DNA with a minor groove width of 11 Å, whereas nearer to the active site, the minor groove width is almost 15 Å (typical A-DNA has a minor groove width of about 17 Å).

Description of the structures. When the pol β ternary complex structures are compared with the structure of the apo enzyme (Fig. 3), the most apparent differences consist of large movements of the 8-kD NH_2-terminal domain relative to the fingers, palm, and thumb of 31-kD COOH-terminal domain. The 8-kD domain is tethered to a proteolytically sensitive hinge region (residues 80–90) and changes from an open conformation in the apo structure to more closed conformations in the complex structures. Because of the precarious position of the 8-kD domain in the pol β apo structure, this type of conformational change seemed inevitable even before the structures of the ternary complexes were determined. The only other significant conformational changes on complex formation were noticeable rigid-body movements of the thumb and, to a lesser degree, the fingers, resulting in a somewhat more tightly closed hand in the ternary com- plexes. A greater degree of flexibility on the part of the thumb subdomain has also been observed in other polymerase-DNA structures (*12, 23, 24, 41*). A least squares superposition of the 31-kD domain of the apo structure (*14*) on the 31-kD domain of one of the ternary complex structures ($P6_1$) resulted in a root-mean-square (rms) deviation in α carbon positions of 2.5 Å, whereas when the fingers, palm, and thumb subdomains were treated separately, the rms deviations in α carbon positions were only 0.71, 0.69, and 0.82 Å, respectively.

The 8-kD domain has a net charge of +10 (assuming neutral histidines) and binds to single stranded DNA with an association constant of $2 \times 10^5\ M^{-1}$ (*26*). It has no obvious structural equivalent in any of the other polymerases for which crystal structures have been determined, and crosslinking studies with gapped DNA substrates (*42*) suggest that the 8-kD domain is probably responsible for the highly

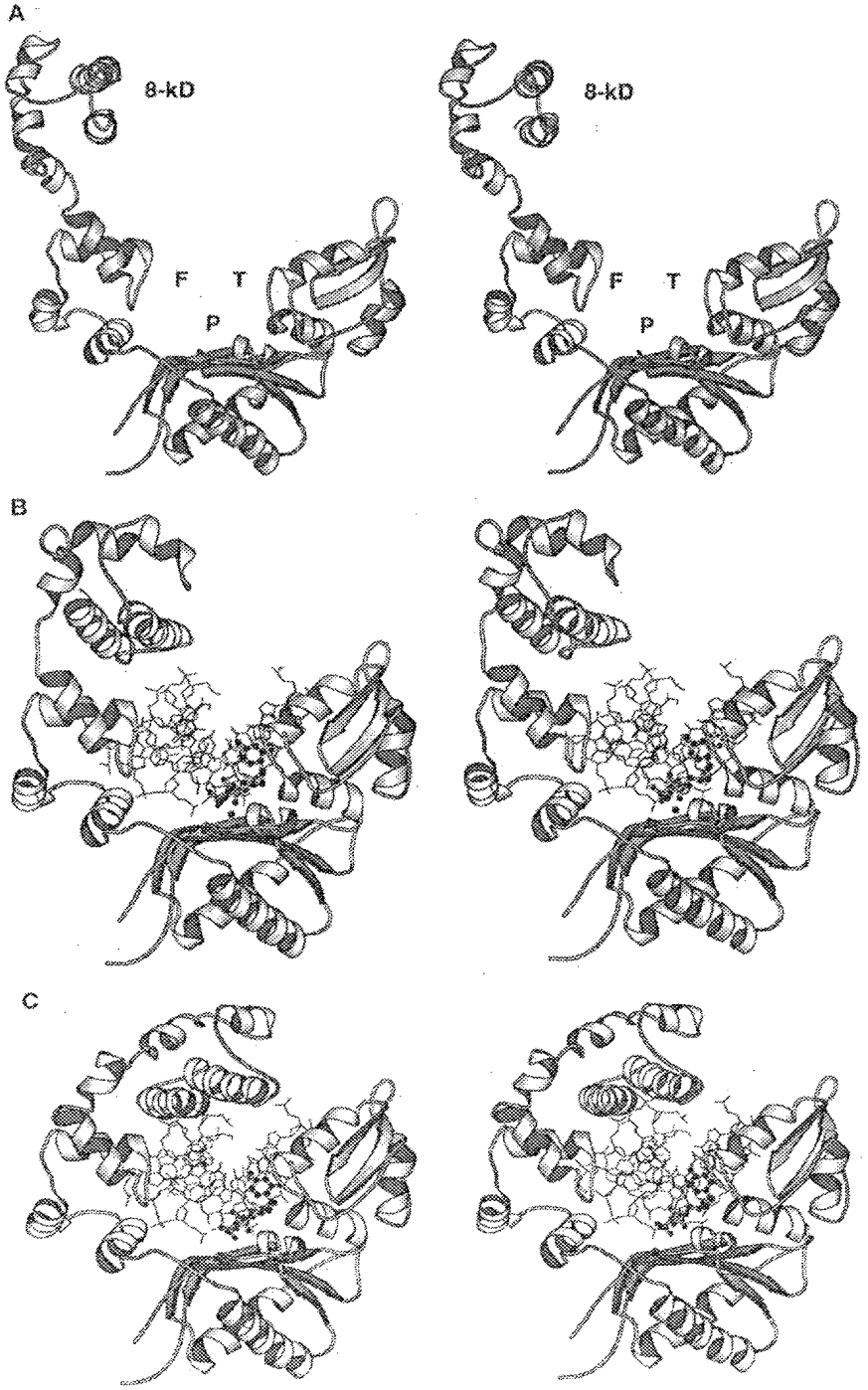


Fig. 3. Stereoview ribbon diagrams (*100*) of (**A**) rat DNA pol β, apo structure (*14*) and (**B** and **C**) ternary complexes of rat DNA pol β with a DNA template-primer and ddCTP in space groups $P6_1$ and $P2_1$, respectively. In (A) the 8-kD domain is designated 8-kD, and the fingers, palm, and thumb subdomains of the 31-kD domain are represented by F, P, and T, respectively. A ball-and-stick representation highlights the ddCTP nucleotide in (B and C). In (B) the positions of the two Mg^{2+} ions are marked with black spheres. These metals ions are not shown in (C) because we were unable to see the Mg^{2+} ions in electron density maps of the lower resolution $P2_1$ ternary complex structure.

processive short-gap filling activities found exclusively in pol β (*43*). It has been proposed that the 31-kD domain binds to the double-stranded segment of the template-primer, and the 8-kD domain binds to the single-stranded template overhang (*44*)—or in the case of binding to a short gap in the DNA, to the 5′-phosphate of the downstream oligonucleotide (*42*). We see some evidence of this in that the 31-kD domain clearly uses its palm, fingers, and thumb to grasp the double-stranded segment of the template-primer while the 8-kD domain, although positioned quite differently in the two complex structures, is nevertheless close to where an extended template would be. Unfortunately, our template overhang was probably a little too short (only four bases—GAAA) to interact strongly with the 8-kD domain, causing the first three bases of the template to be disordered in both crystal structures. It is possible that, because the highly flexible 8-kD domain had no template on which to anchor in our crystallization experiments, its position was determined almost entirely by crystal packing forces, and probably neither of the two conformations of the 8-kD domain seen in Fig. 3, B and C, is correct for template binding in vivo. Nevertheless, kinetic studies of the 31-kD fragment alone showed that pol β can still function as a polymerase without the 8-kD domain, albeit at only about 5 percent of its normal activity (*44*).

Table 2. Hydrogen bond interactions of 3.3 Å or less between pol β and the DNA template-primer.

Residue	Subdomain	Atom	Base*	Atom	Distance (Å)
		Protein to DNA phosphate H bonds			
Gly^{105}	Fingers	N	P-6C	O2P	2.9
Gly^{107}	Fingers	N	P-5G	O2P	2.7
Ser^{109}	Fingers	N	P-5G	O1P	2.9
Ala^{110}	Fingers	N	P-5G	O2P	3.1
Arg^{254}	Fingers	NH2	P-7C	O2P	2.7
Lys^{230}	Palm	N	T-9C	O2P	3.0
Thr^{233}	Palm	N	T-8G	O2P	3.1
Lys^{234}	Palm	N	T-8G	O2P	2.7
Thr^{292}	Thumb	OG1	T-5G	O2P	2.7
Tyr^{296}	Thumb	OH	T-5C	O2P	2.6
		Protein to DNA base H bonds			
Lys^{234}	Palm	NZ	T-7C	O2	2.9
Tyr^{271}	Thumb	OH	P-7C	O2	2.7
Arg^{283}	Thumb	NH1	T-4G	N3	3.2

*DNA bases are designated T or P to distinguish template bases from primer bases, respectively. Starting from the 5′ terminus of each strand, bases are numbered 1 to 11 for the template and are numbered 1 to 7 for the primer. C and G represent cytosine and guanine, respectively, and atom designations follow Protein Data Bank nomenclature.

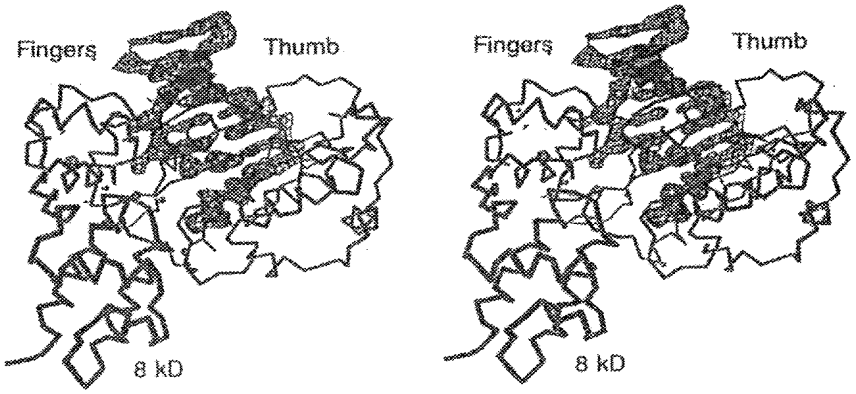


Fig. 4. Omit map of the DNA template-primer and ddCTP overlayed on an α carbon diagram of the refined $P6_1$ pol β structure. The view is that of Fig. 3 rotated by 90° about a horizontal axis in the plane of the page so that the DNA binding channel is now vertical. The template-primer sits on top of the palm subdomain, which is not labeled. Before all omit maps were calculated, the part of the structure in question was deleted from the coordinate file and the remaining partial structure was subjected to 200 cycles of least squares positional refinement in XPLOR (*36*) in order to remove bias from the phases.

In contrast to the 8-kD domain, the rest of the structure (the fingers, palm, and thumb of 31-kD domain, as well as the template-primer and ddCTP substrate) is virtually identical in both crystal forms of the ternary complex. This provides support for the physiological relevance of our complex crystals, at least with respect to the polymerase-DNA-ddCTP interactions. Also strengthening the argument is that, unlike other reported crystals of polymerase-DNA complexes (*12, 41*), our crystals were grown at low, near physiological salt concentrations. Finally, the fact that that a ddCMP was incorporated into our template-primer shows that the nucleotidyl transfer reaction did proceed, at least for one turnover, in the same medium from which crystals were eventually obtained. In that the following discussions do not apply to the 8-kD domain of pol β and will be limited mostly to the 31-kD domain's interactions with DNA and ddCTP, we will henceforth refer only to the complex structure that has been refined to the highest resolution, the $P6_1$ crystal structure.

The DNA binding channel in pol β, just as in KF, RT, and RNAP, is lined with positively charged lysine and arginine side chains, and it has always been a reasonable assumption that their function is to stabilize the negatively charged backbone phosphates of the DNA (*45*). Therefore it was quite surprising that, except for Arg^{254}, which is hydrogen bonded to the phosphate of the newly incorporated ddCMP of the primer strand, there are no direct lysine or arginine interactions with the backbone phosphates of the DNA in our complex (Table 2). Instead, nearly all of the interactions of protein with DNA involve two different clusters of protein backbone nitrogens located at the entrance to the DNA binding channel (Table 2). One cluster, consisting of four of the backbone nitrogens between Gly^{105} and Ala^{110}, is located at the NH_2-terminal end of helix G in the fingers subdomain of pol β and interacts with the phosphates of the primer strand. The second cluster, comprising three of the five backbone nitrogens between Lys^{230} and Lys^{234}, is located in a beta turn, which connects beta strands 3 and 4 of the palm subdomain and interacts with backbone phosphates of the template strand. The only other hydrogen bonded interactions (3.3 Å or less) between pol β and DNA phosphates are between the side chains of Thr^{292} and Tyr^{296}, located on a loop between beta strands 6 and 7 of the thumb subdomain, and the backbone phosphates of the template strand (Table 2).

In addition to our observations that there seemed to be fewer hydrogen bond interactions between pol β and DNA than expected (Table 2), we were also initially

surprised to see that the DNA sits in the binding channel at a slight angle and appears to "run into" alpha helices M and N of the thumb subdomain (Fig. 4). It is possible that the angle between the DNA axis and the apparent axis of the pol β binding channel would change considerably if a longer template were used and, as proposed above, the 8-kD domain participated in the positioning of the template-primer. However, the aesthetically pleasing observation that the base pairs of the DNA are parallel to the beta strands of the palm subdomain (Fig. 5) encourages us to believe that interactions of pol β with the double-stranded segment of any DNA template-primer will not vary much from what is seen in the present ternary complex structures.

Perhaps one of the most unvarying characteristics of B-DNA is that it has a spine of well-ordered water molecules which interacts with the O2 of pyrimidines and the N3's of purines in the minor groove, and it has been proposed that the disruption of this particular water structure is the first step in the B-DNA to A-DNA transition (*46*, *47*). In our complex structure, only three protein side chains come within 3.3 Å of the DNA bases, and they are all located in the shallow minor groove of the template-primer (Table 2). Two of these (Lys^{234} and Tyr^{271}) are hydrogen bonded to the O2 of a template cytidine and the O2 of a primer cytidine, respectively, while another (Arg^{283}) is hydrogen bonded to the N3 of a template guanine. This leads us to propose that perhaps Lys^{234}, Tyr^{271}, and Arg^{283} all function to break up the water structure in the minor groove of the template-primer upon complex formation, resulting in a larger minor groove width, characteristic of A-DNA, at the pol β active site.

Unlike transcription factors and other gene regulatory DNA-binding proteins, polymerases must bind to DNA with little regard for sequence specificity. This is evident from Table 2 where, as discussed above, most of the protein to DNA interactions are nonsequence-specific hydrogen bonds between pol β backbone nitrogens and DNA phosphate oxygens. Even what appear to be sequence-specific interactions between protein side chains and DNA bases, upon closer inspection turn out to be rather nonspecific in that each of the three protein side chains mentioned above—Lys^{234}, Tyr^{271}, and Arg^{283}—that are in contact with DNA bases can act as an unbiased hydrogen bond donor to either the O2 of pyrimidines or the N3 of purines in the minor groove.

Description of the active site. Much of the ddCTP binding pocket in the ternary complex is made up of the 3′ terminus of the primer strand (Fig. 6A, toward the left) and the template overhang (at the top), where the base pairing of ddCTP with a complementary base, guanine, is evident in the crystal structure (Fig. 6A and Table 3). The idea that the template-primer makes up a large part of the binding site in pol β for the incoming ddCTP is consistent with kinetic studies showing that there is a strict kinetic order of binding for the substrates in nucleotidyl transfer reactions catalyzed by pol β (*48*), *E. coli* pol I (*49*), and RT (*50*), the polymerase binds to the template-primer first, and then nucleotide binding takes place. Supporting this view, structural studies of a pol β–dATP binary complex (*14*) and of a KF-dATP binary complex (*33*), both of which were crystallized in the absence of a template-primer, revealed nucleotide binding sites that differed somewhat from that seen in our ternary complex.

Making up most of the right side of the nucleotide binding pocket and interacting primarily with the base and ribose moieties of the ddCTP is a structural motif consisting of a sharp kink, made possible by a cis-peptide bond, between α helices M and N in the thumb subdomain of pol β (Fig. 6A). This was the only cis-peptide found in the pol β apo structure (between Gly^{274} and Ser^{275}) (*14*) and even though it appears to have remained as a cis-peptide upon complex formation with DNA, its proximity to the active site nevertheless suggests that it may play a dynamic role in catalysis. For instance, it is evident that a cis- to trans-peptide bond transition, perhaps occurring during or just after the incorporation of a nucleotide, could result in a large displacement of one or both helices (M and N), which in turn could push the enzyme off the template-primer. Thus, we propose that the cis-peptide in pol β may function to facilitate the product-off step of catalysis, which is typically the steady-state rate-limiting step for polymerases when in a distributive mode of synthesis (*21*, *51*, *52*). Another possibility is that the cis-peptide bond, which links the active site to the template overhang via helix N (Fig. 6A), functions only to facilitate the translation of the enzyme along the template during processive polymerization.

It is quite appropriate that the protein side chain that interacts specifically with the base of ddCTP, Asn^{279} of helix N (Table 3), is unbiased toward all four possible incoming nucleotides in that it can act either as a hydrogen bond acceptor or a hydrogen bond donor. The only other interactions (of 4 Å or less) between pol β and the base moiety of ddCTP are nonspecific van der Waals contacts between the side chain carbon (CB) of Asp^{276} and ring carbon atoms (C4 and C5) of the cytidine base (Table 3).

Pol β interacts intimately with the sugar moiety of ddCTP as is evident from the close van der Waals contacts between the protein backbone atoms of Tyr^{271}, Phe^{272}, and Gly^{274} (*53*), and the ribose ring carbons, C2′ and C3′, of ddCTP (Fig. 6A and Table 3). Because the only difference between a ribonucleotide and a deoxyribonucleotide is a hydroxyl at the C2′ of the ribose ring, the protein backbone segment Tyr^{271}-Gly^{274} may participate in nucleotide selectivity of DNA over RNA for pol β. For example, a DNA polymerase can be converted to a relatively efficient RNA polymerase by changing the metal ion in the reaction medium from Mg^{2+} to Mn^{2+} (*54*). The two metal ions in the pol β active site together bind to all three phosphates (α, β, and γ) of ddCTP (Table 3), and therefore the metal ions probably participate to some extent in the positioning of the incoming nucleotide. As seen in our structure, only a slight change in ddCTP orientation, which could be induced by a change from Mg^{2+} to Mn^{2+} ions in the active site, would be required to reduce steric hindrance at the C2′ ribose of ddCTP, possibly making

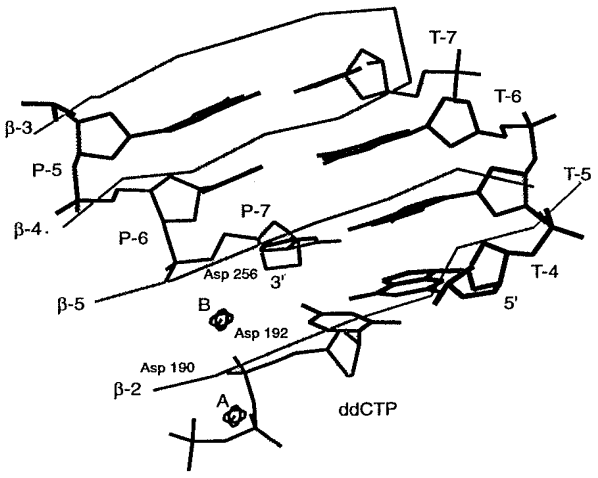


Fig. 5. Alignment of the template-primer base pairs with the beta strands of the palm subdomain. The view is the same as that in Fig. 4. Beta strands are labeled β-2 through β-5. The α carbon positions of the catalytically important residues, Asp^{190}, Asp^{192}, and Asp^{256}, are shown, as well as the nucleotide substrate, ddCTP. The two Mg^{2+} ions in metal sites A and B are labeled A and B, respectively. Base designations defined in Table 2.

RNA just as good a substrate as DNA.

Another example of nucleotide selectivity, this time dTTP over AZT-TP, might also be explained by steric hindrance in the pol β active site, particularly at the ribose C3′. The cellular nucleotide dTTP, which has a hydroxyl group attached to its ribose C3′ (Fig. 2B), binds more tightly to a pol β-DNA complex than does its analog AZT-TP, which has a bulky azido group at C3′ (Fig. 2D) (*55*). Such observations might explain why a drug like AZT specifically targets the HIV-1 RT and not host cell polymerases like pol β; in contrast to pol β, RT shows no selectivity in binding dTTP compared to AZT-TP during reverse transcription (*56*), and thus, perhaps RT lacks the structural equivalent of the pol β "selective" Tyr^{271} to Gly^{274} backbone segment, making RT more susceptible to AZT-TP inhibition.

Three of the six hydrogen bonds between the protein and the negatively charged phosphate moiety of ddCTP involve nitrogen backbone atoms of pol β (Table 3). Although this is reminiscent of the nitrogen clusters mentioned above that help to stabilize the template and primer phosphates in the DNA binding channel, the geometry of the nitrogen backbones differ in that they more closely resemble the mononucleotide binding motifs found in other enzymes. This and other aspects of the interactions between protein and ddCTP phosphates in the pol β active site are discussed in (*14*).

The strongest interactions with ddCTP in the pol β active site, however, are not with protein side chains directly, but rather with two Mg^{2+} ions that in turn coordinate the side chain oxygens of Asp^{190}, Asp^{192}, and Asp^{256} (Fig. 6 and Table 3). These three carboxylic acids are present in all DNA polymerases for which amino acid sequences are known (*15*), and mutagenesis studies show that they are critical for catalysis in pol β (*57*), KF (*58, 59*), and RT (*60*). However, we avoid referring to these three aspartic acids as a "catalytic triad" because only two of the three (Asp^{190} and Asp^{256}) are present in all known RNA polymerase amino acid sequences (*15*). As discussed above, converting a DNA polymerase to a relatively efficient RNA polymerase can be as simple as changing the metal ions in the active site from Mg^{2+} to Mn^{2+} (*54*), and consequently there is not much reason to suppose that the RNA polymerases utilize a nucleotidyl transfer mechanism that differs drastically from the mechanism proposed below for a DNA polymerase such as pol β. In support of this view, mutagenesis experiments that targeted catalytically important residues in KF (*59*) showed that Asp^{882} and Asp^{705} (the equivalents of Asp^{190} and Asp^{256} in pol β) are much more critical to catalysis than the third carboxylic acid of the trio, Glu^{883} (equivalent to Asp^{192} in pol β), which is not conserved in the RNA polymerases. In agreement with these observations and as discussed below, the primary function of Asp^{256} in the pol β active site appears to aid in stabilizing the pentacoordinated α phosphate of the transition state, whereas the primary functions of Asp^{190} and Asp^{192} are to aid in positioning the nucleotide substrate. This suggests that pol β could quite possibly still function in the absence of one of the two latter carboxylic acids, such as Asp^{192}, which would be the case for an RNA polymerase. Perhaps also significant is the observation that the metal-to-oxygen bonds are longer for Asp^{192} than they are for Asp^{190} in the pol β active site (Table 3).

Table 3. Interatomic distances of interest for the pol β active site.

ddCTP moiety	Atom	Base* or residue	Atom	Distance (Å)
		Contacts between ddCTP and DNA or protein		
Base	N4	T-4 G	O6	3.0
	N3	T-4 G	N1	2.7
	O2	T-4 G	N2	2.7
	C4	Asp^{276}	CB	3.5
	C5	Asp^{276}	CB	3.7
	O2	Asn^{279}	ND2	3.0
Ribose	C2′	Tyr^{271}	O	3.5
	C3′	Phe^{272}	O	3.2
	C2′	Gly^{274}	CA	3.2
α Phosphate	PA	P-7 C	O3′	4.3†
β Phosphate	O1B	Ser^{180}	N	3.0
	O1B	Arg^{183}	NH2	2.7
	O3B	Ser^{180}	OG	2.8
γ Phosphate	O1G	Gly^{189}	N	2.8
	O1G	Asp^{190}	N	3.0
	O2G	Arg^{149}	NH2	2.5
		Ligand	Atom	Distance (Å)
Mg^{2+} site A		Asp^{190}	OD1	2.0
		Asp^{192}	OD2	2.7
		ddCTP, β phosphate	OD2	1.7
		ddCTP, γ phosphate	O1G	2.3
		Water	O	2.6
Mg^{2+} site B		Asp^{190}	OD2	2.0
		Asp^{192}	OD1	2.6
		Asp^{256}	OD1	3.0
		ddCTP, α phosphate	O1A	2.7
		P-7 C	O3′	2.9†
		Water	O	—‡

*Base designations are in Table 2. †These distances were obtained with a O3′ atom that had to be modeled into the active site because the newly incorporated ddCMP primer terminus lacks this group. ‡This water is not seen in our crystal structures, but because the Mg^{2+} ion in site B has octahedral geometry, we propose that a water might occupy this empty sixth-ligand position.

Mechanism of nucleotidyl transfer. In 1979, with results from kinetic experiments on *E. coli* pol I that utilized phosphorothioate dATP analogs, Burgers and Eckstein proposed that the pol I catalyzed nucleotidyl transfer reaction had the following properties (*61*): (i) A divalent metal ion (Mg^{2+}) is bound specifically to the β and γ phosphates of the nucleotide (dATP) as a β,γ-bidentate; (ii) the negative charge on the α-phosphate of the nucleotide is neutralized by a positive group on the enzyme; and (iii) attack by the 3′-OH group of the primer on the α phosphate and subsequent release of the PP_i (or Mg-PP_i) proceeds in an in-line fashion.

All three of these properties are in good agreement with the geometry seen in the active site of our ternary complex, and in fact, our active site (Fig. 6) does not differ significantly from that proposed by Burgers and Eckstein [figure 4 of (*61*)]. In agreement with criterion number (i), the pol β active site shows a Mg^{2+} ion (metal site A) bound as a bidentate to the β and γ phosphates of ddCTP (Table 3). As for criterion number (ii), we now know that the originally undefined "positive group on the enzyme" that stabilizes the α phosphate of the nucleotide is simply the second Mg^{2+} ion in metal site B (Fig. 6). In fact, the geometry of the Mg^{2+} ion in site B of our

structure can explain why polymerases are highly selective for only the S, as opposed to the R, absolute configuration of dATPαS phosphorothioate analogs whenever Mg^{2+} is the metal ion in the reaction mixture (*61*). Mg^{2+} coordinates oxygen much more strongly than sulfur, and in our active site the Mg^{2+} ion in metal site B coordinates a specific, nonesterified oxygen of the α-phosphate of the nucleotide. If this particular oxygen were replaced by a sulfur, as would be the case for the *R* configuration of dATPαS, coordination by the Mg^{2+} ion in site B should be weakened considerably, at least according to the geometry our structure. Finally, the geometry of the active site is also in accord with an in-line mechanism, criterion number (iii), for the nucleotidyl transfer reaction. An in-line mechanism, proposed because the polymerase reaction proceeds with inversion of configuration at the α phosphate (*61*), restricts the possible orientation of the attacking group with respect to the leaving group. The attacking and leaving groups must be opposite one another, relative to the α phosphate, in order to occupy the two apical positions of the pentacoordinated α phosphate in the transition state. In the pol β active site, the 3′ carbon of the primer strand (which normally possesses the attacking 3′-OH) lies just opposite, relative to the α phosphate, to the scissile oxygen of the PP_i leaving group.

It is not too surprising that the active site of pol β is similar to the active site of the 3′ → 5′ exonuclease domain of *E. coli* pol I, which is known, through extensive structural (*23, 24*) and mutagenesis (*62, 63*) studies, to employ a two-metal ion mechanism for phosphoryl transfer, and like pol β, proceeds with inversion of configuration at the scissile phosphate (*64*). Like the 3′ → 5′ exonuclease, pol β has three carboxylic acids that position two divalent metal ions about 4 Å apart in the active site and, although the orientations of the carboxylic acids are quite different in the two structures, the geometries of the metal sites are strikingly similar. The metal ion in site A in both cases tightly coordinates four oxygen ligands with highly distorted tetragonal geometry and has a fifth, weakly bound water ligand (Table 3). In fact, a better description of the unusual geometry around metal site A in our structure would be that of a square pyramid in which the Mg^{2+} is at the apex, rising about 1.8 Å out of the square plane described by four oxygen ligands (one from Asp^{190}, one from Asp^{192}, one from the β phosphate, and one from the γ phosphate) (Fig. 6) (Table 3). In contrast, the metal ion in site B (for both pol β and the 3′ → 5′ exonuclease) is not as tightly bound as the metal ion in site A and has slightly distorted octahedral geometry (*65*). Just as with the 3′ → 5′ exonuclease, we propose that the primary function of the metal ion in site A is to aid in binding and positioning of the substrate, and the primary function of the metal ion in site B is to help stabilize the pentacoordinated phosphate of the transition state, though both metal sites probably participate in both of these functions to some degree. Finally, as is the case with the 3′ → 5′ exonuclease, we propose that a metal ion activates the attacking oxygen by acting as a Lewis acid, while a protein side chain acts as a proton acceptor (Fig. 6B). For the 3′ → 5′ exonuclease, the metal ion in site A activates an attacking water molecule, while Glu^{357} acts as the proton acceptor and, in the case of pol β, the metal ion in site B activates the attacking 3′-OH of the primer, while Asp^{256} acts as the proton acceptor. All of these features of the transition state of the nucleotidyl transfer reaction of pol β are evident from the crystal structure, and only the missing 3′-OH of the primer strand was added in order to draw the schematic in Fig. 6B.

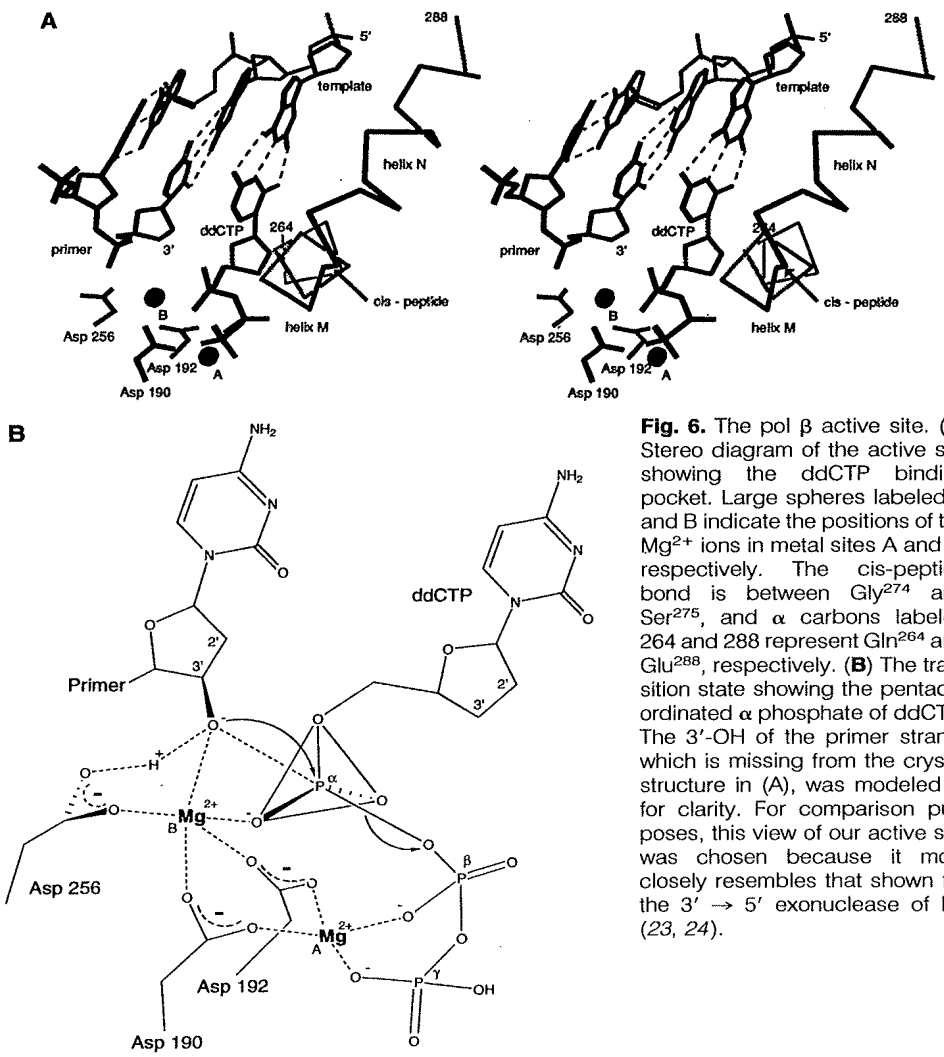


Fig. 6. The pol β active site. (**A**) Stereo diagram of the active site showing the ddCTP binding pocket. Large spheres labeled A and B indicate the positions of the Mg^{2+} ions in metal sites A and B, respectively. The cis-peptide bond is between Gly^{274} and Ser^{275}, and α carbons labeled 264 and 288 represent Gln^{264} and Glu^{288}, respectively. (**B**) The transition state showing the pentacoordinated α phosphate of ddCTP. The 3′-OH of the primer strand, which is missing from the crystal structure in (A), was modeled in for clarity. For comparison purposes, this view of our active site was chosen because it most closely resembles that shown for the 3′ → 5′ exonuclease of KF (*23, 24*).

Despite all the similarities, we caution against thinking of the 3′ → 5′ exonuclease reaction as "the polymerase reaction in reverse." In fact, the polymerase reaction in reverse is pyrophosphorolysis and differs from the exonuclease reaction as follows:

3′ → 5′ Exonuclease reaction:
primer-dNMP + H_2O→primer + dNMP

Pyrophosphorolysis:
primer-dNMP + PP_i→primer + dNTP

so that the metal ion in site A for the 3′ → 5′ exonuclease interacts primarily with a water molecule (*24*), whereas the metal ion in site A for the polymerase interacts primarily with PP_i (or the PP_i moiety of ddCTP) (*65*).

In summary, the overall nucleotidyl transfer reaction, as catalyzed by pol β, probably proceeds as follows (Fig. 6):

i) As pol β binds to the template-primer, Lys^{234}, Tyr^{271}, and Arg^{283} break up the

water structure of the minor groove, causing the minor groove width to increase near the active site.

ii) An incoming nucleotide is positioned in the active site by base-pairing with the template, by a hydrogen bond to Asn^{279}, by a van der Waals contacts with Asp^{276}, by steric hindrance with the protein backbone between Tyr^{271} and Gly^{274}, by six hydrogen bonds between protein and ddCTP phosphates, and by two Mg^{2+} ions that are, in turn, positioned by Asp^{190}, Asp^{192}, and Asp^{256}.

iii) The Mg^{2+} ion in metal site B, acting as a Lewis acid, activates the 3′-OH of the primer terminus, while one of its ligands, Asp^{256}, probably acts as the proton acceptor for the 3′-OH.

iv) After attack on the α-phosphate by the activated 3′-OH, the reaction proceeds through the transition state in which the α phosphate is pentacoordinated, with the attacking 3′ oxygen of the primer terminus and the leaving oxygen of the PP_i group occupying the two apical positions. The pentacoordinated transition state is stabilized by the Mg^{2+} ion in site B, which coordinates both an apical and an equatorial oxygen of the α phosphate.

v) After PP_i (or Mg-PP_i) is released, pol β is ready for another cycle, but only after the enzyme either releases from the template-primer (distributive mode of synthesis) or pulls itself along the template (processive mode of synthesis); one or both of these activities may be facilitated by a conformational change of cis- to trans-peptide at Gly^{274}–Ser^{275}.

Metal ions play crucial roles in the mechanism of pol β and of the 3′ → 5′ exonuclease of pol I. This kind of independence from direct involvement of protein side chains during catalysis has led to proposals that perhaps hydrolysis reactions involving nonprotein catalysts such as self-splicing ribozymes (*66, 67*), where positioning of the two metal ions can be achieved just as easily by RNA backbone phosphates, proceed through a similar two-metal ion mechanism as that proposed for the 3′ → 5′ exonuclease of pol I (*23, 68*). In much the same way, perhaps the nucleotidyl transfer mechanism that we present also applies, to some extent, to those ribozymes that are capable of catalyzing the nucleotidyl transfer reaction (*69, 70*).

Comparisons with other polymerases. As discussed in the accompanying report (*14*), the most obvious structural overlap among all four polymerases, KF, RNAP, RT, and pol β, consists of a conserved pair of carboxylic acid side chains located in the palm subdomain. As revealed by the pol β active site, the primary function of these carboxylic acids is to position two catalytically critical divalent metal ions. The observation that even the most divergent polymerase structures retain this catalytic core is compelling evidence that all polymerases share a common catalytic mechanism (*11, 14*). We believe that the pol β ternary complex structures presented here are physiologically relevant and that the nucleotidyl transfer mechanism that we propose based on these structures represents a common catalytic mechanism found in all polymerases. In the absence of evidence that suggests otherwise, the next logical argument would be that all polymerases, because they share a common catalytic mechanism, should bind to the DNA (or RNA) template-primer in a fashion very similar—at least with respect to the highly conserved catalytic residues—to that in which pol β has attached itself to the DNA template-primer in our ternary complex (Fig. 7, A to C). All of the polymerase structures discussed have a DNA binding channel that can grasp a stationary rodlike DNA template-primer in one of two general directions, so the only other possibility for binding to the template-primer, provided that there are no gross conformational changes on the part of DNA or protein, is shown for the RT structure (Fig. 7D). In order to generate Fig. 7D, the template-primer in Fig. 7C was held stationary while the RT molecule was rotated by 180° about an axis normal to the plane of the page and passing through the polymerase active site. The mode of DNA binding depicted in Fig. 7D, which is opposite to that found in pol β (anti–pol β), is the one proposed for KF (*41*), for RNAP (*13*), and for RT (*11, 12*). We are thus faced with a dilemma.

Three main possibilities can be pursued. First, it could be argued that the other three polymerase-DNA models are basically correct, and that the pol β ternary complex structures presented here are not physiologically relevant. However, as stated above, there is strong evidence in favor of the physiological relevance of our structures, and we therefore bypass this possibility for the present. A second possibility could be that all four of the polymerase-DNA models are correct, including the pol β–DNA structures presented here, and that pol β uses an entirely different mode of template-primer binding from the other polymerases. However, that pol β has a palm subdomain that is structurally homologous with the other polymerases (*14*), and that this palm subdomain also contains the highly conserved catalytic residues, suggests otherwise. The two-metal ion mechanism that we propose for the nucleotidyl transfer reaction seems to depend heavily on the positioning of the metal ions by the conserved carboxylic acid residues. Furthermore, the catalytic site appears to be asymmetric in that the two metal sites possess different geometries and different binding affinities (*65*). This leads us to the third possibility, which is the one we choose to pursue in the following discussions. The other polymerases must bind to a template-primer in a manner similar to that of pol β so as to conserve the critical, asymmetric geometry of the active site. An inescapable inference, then, is that the other polymerase-DNA models are not correct with respect to the directionality of the template-primer in the binding channel. Because of the complexity of the existing data and interpretations, we address each of the

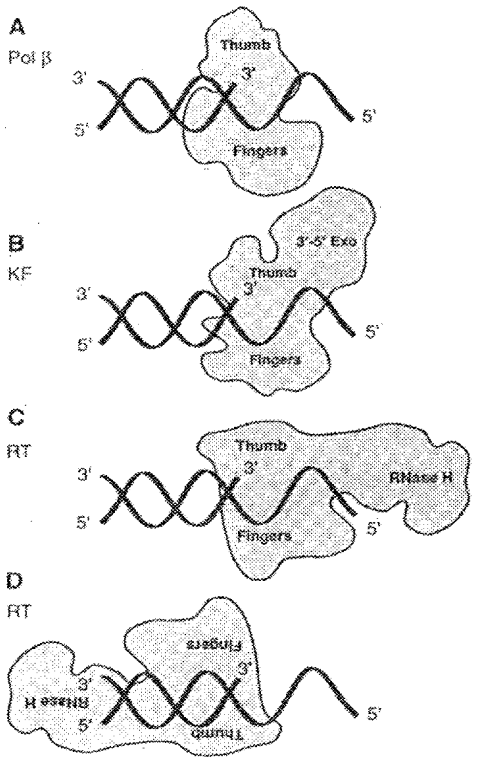


Fig. 7. (**A**) Schematic drawing showing the position of the template-primer as seen in the pol β ternary complex structure and (**B** and **C**) our proposed models for the KF-DNA and RT-DNA complexes, respectively. To generate (B and C) the palms of KF and RT were aligned with the palm subdomain of pol β as viewed in (A). In the text, (A to C) represent the pol β–like mode of binding to the template-primer. To generate (**D**), which represents the anti–pol β mode of binding to the template-primer, the template-primer in (C) was held stationary, and the RT molecule was rotated by 180° about an axis normal to the plane of the page and running through the RT active site. The template-primer was kept stationary in all four diagrams because that is a reasonable representation of in vivo polymerization, where the DNA molecule is usually very long and immovable in comparison to the polymerase molecule. The active site in all four diagrams is at the 3′ terminus of the primer strand, and the view looks down at the palm, which forms the bottom of the template-primer binding channel.

three other polymerases, KF, RNAP, and RT, separately.

DNA pol I of *E. coli* is a 103-kD monomer and can be cleaved by limited proteolysis into a large (68-kD) COOH-terminal fragment [commonly referred to as the Klenow fragment (KF)], and a smaller (35-kD) NH_2-terminal fragment (*1*). KF has both polymerase and 3′ → 5′ exonuclease (editing) activities, whereas the small NH_2-terminal fragment functions solely as a 5′ → 3′ exonuclease. The crystal structure of KF alone revealed two distinct domains (*10*), confirming previous proposals that the polymerase and 3′ → 5′ exonuclease functions of KF lie on separate, independent folding units. From its position in the pol I amino acid sequence, it was then inferred that the missing NH_2-terminal 5′ → 3′ exonuclease domain should be positioned somewhere to the right of the 3′ → 5′ exonuclease as viewed in Fig. 7B (*10*). During nick translation, a function intrinsic to pol I, it is proposed that the polymerase, starting at a nick in the DNA and polymerizing in a 5′ → 3′ direction, works in conjunction with the activities of the 5′ → 3′ exonuclease so that the net result is simply a translation of the nick along the DNA in a 5′ → 3′ direction (*1*). It was this property of pol I that led to proposals that the polymerase domain of pol I bound to the DNA template-primer in a manner not too different from that shown in Fig. 7B, so that the polymerase and the 5′ → 3′ exonuclease were positioned properly, relative to the DNA nick, to perform their separate activities in conjunction with one another (*10*). Subsequent KF-DNA models were proposed to show how the polymerase and the 3′ → 5′ exonuclease of KF could work in conjunction with one another during the DNA editing process (*23, 24, 71, 72*). Once again, these earlier models require that KF binds to the DNA template-primer in a manner similar to that shown in Fig. 7B. Therefore, all of the earlier pol I–DNA models are in agreement with a pol β–like mode of DNA binding.

The recent crystal structure determination of KF crystallized in the presence of a putative template-primer revealed an unexpected complex in which the KF had bound to the template-primer in neither of the two binding modes discussed above, but instead, in such a way that the DNA lay in a separate, less obvious channel running between the thumb subdomain of the polymerase and the 3′ → 5′ exonuclease (*41*). From this structure it was proposed that during polymerization, the KF bent the DNA by 80° so that the DNA entered the larger, more obvious DNA binding channel in a direction opposite to that of all the previously proposed KF-DNA models. We now find ourselves in the awkward position of presenting arguments in favor of the idea that the initial KF-DNA models were correct, at least with respect to the direction of DNA binding, and that the recent KF-DNA crystal structure (*41*) might not be physiologically relevant.

Possibly the most compelling reason to believe that the previously reported KF-DNA crystal structure (*41*) might not be physiologically relevant is that it is not in accord with the nick translation activities of pol I. Other reasons, which concern the nature of the KF-DNA crystals themselves, are (i) the complex crystals were grown under relatively high, nonphysiological salt concentrations; (ii) the template-primer in the crystallization experiments included an unnatural epoxy-ATP; (iii) the template-primer had been unexpectedly modified during crystallization in such a way that the primer strand was longer than the template strand in the crystal structure; and finally, (iv) the complex crystals were isomorphous with the apo KF crystals, which in itself is not a strong argument, but is made stronger by the observation that none of the pol β ternary complex crystals presented here is isomorphous with the apo crystals. Because the 3′ → 5′ exonuclease domain of KF was shown to have a high affinity for the primer terminus during other attempts at obtaining KF-DNA complex crystals (*23*), perhaps the recently reported KF-DNA structure (*41*) is solely an editing complex in which the primer terminus bound to the 3′ → 5′ exonuclease site, and the leftover "tethered" DNA, guided by crystal packing forces, bound wherever it could under the circumstances. This is similar to the argument presented above that the position of the tethered 8-kD domain in our complex crystals, though ordered in both structures, is probably not physiologically relevant.

Further support for the idea that the KF binds to the template-primer in a manner very similar to that of pol β comes from closer inspection of the DNA binding channels of both enzymes. KF possesses what appears to be the structural equivalent to helices M and N of the pol β thumb subdomain. As described above, helices M and N protrude into the pol β DNA binding channel and interact with the nucleotide substrate (Fig. 3). Though not structurally homologous with helices M and N, helices J and K of KF also protrude into the KF-DNA binding channel in an area near the catalytic carboxylic acids that overlaps quite well with helices M and N of pol β, suggesting a similar role for helices J and K in nucleotide specificity for KF.

RNA polymerase (RNAP) from bacteriophage T7 is a monomer of molecular weight 99-kD, and the crystal structure determination of RNAP revealed that it folds into two distinct domains (*13*). One domain possesses a polymerase fold that is highly homologous with the polymerase domain of KF, while a smaller, NH_2-terminal domain, though not an exonuclease, is located in the same relative position as the 3′ → 5′ exonuclease domain of KF. That the polymerase domains of RNAP and KF are so structurally homologous was one of the strongest arguments presented in favor of a model (*13*) wherein RNAP binds to a DNA-RNA duplex in a manner similar to that of the recent KF-DNA model (*41*). However, as pointed out above, we believe that the recent KF-DNA complex structure might not be physiologically relevant and that the KF, instead, probably utilizes a pol β–like mode of binding to the template-primer. Therefore, we propose that like KF, RNAP also utilizes a pol β–like mode of binding to an RNA-DNA duplex.

Other evidence besides homology with KF was also presented in favor of an anti–pol β mode of template-primer binding for RNAP (*13*)—in disagreement with our proposals for an RNAP-RNA-DNA complex structure. Results from mutagenesis studies (*73, 74*) as well as binding studies with proteolytically modified RNAP fragments (*75*) suggest that the NH_2-terminal domain of RNAP binds to the nascent RNA of the emerging RNA-DNA duplex (*73, 75*) and that Gln^{748} of RNAP recognizes the -10 and -11 bases of the DNA promoter upon complex formation (*74*). These observations suggested positioning the RNA-DNA duplex in the template-primer binding channel of RNAP in an anti–pol β fashion (*13*). Although more work will be required to settle this issue, the structural evidence presented by the pol β ternary complex seems very compelling and interpretations of results from the mutagenesis studies as well as binding studies of proteolytically modified RNAP fragments may require a more careful analysis. As an example of the difficulties involved in interpreting mutagenesis and binding studies in the absence of structural data on a polymerase-DNA complex, mutagenesis studies on pol β implicated Arg^{183} as taking part in primer strand recognition upon complex formation (*76*), and binding studies with pyridoxal 5′-phosphate suggested that the 8-kD domain of pol β formed a part of the nucleotide binding pocket (*77*). Neither of these conclusions is in agreement with our ternary complex structures, though the results from these studies, when re-evaluated, are not necessarily in disagreement with our structural work [reference 41 in (*14*)].

Retroviral RTs, responsible for making double-stranded DNA copies of single-stranded RNA viral genomes, had been well studied (*78*) prior to the discovery that the retrovirus HIV-1 was the cause of AIDS (*79–81*). This facilitated the study of the

HIV-1 RT because its mechanism of operation is similar to that of many previously studied RTs (*82*). HIV-1 RT functions as a heterodimer, and the crystal structure determination of RT showed that one monomer of the heterodimer, called p66 because it has a molecular mass of 66 kD, folds into two distinct domains: a typical polymerase domain (palm, fingers, and thumb) and a connected ribonuclease (RNase) H domain (*11, 12*). The second monomer of the RT heterodimer, p51, is simply a copy of p66 in which the 15-kD RNase H COOH-terminal segment has been proteolytically cleaved. In contrast to the fingers, palm, and thumb of the polymerase domain of p66, the fingers, palm, and thumb of p51 occupy relatively different positions in the crystal structure, resulting in something other than a typical polymerase fold for p51 in the p66-p51 heterodimer (*11, 12*). Only p66 is shown in Fig. 7C for clarity and to highlight how the RNase H domain lies in approximately the same direction with respect to the polymerase as the 3′ → 5′ exonuclease domain of KF. The function of the RNase H is to remove the viral RNA template from the RNA-DNA hybrid that results after RNA-directed DNA polymerization (reverse transcription) of the viral genome.

In keeping with our proposals that all polymerases share a common catalytic mechanism, and hence a common template-primer binding mode, we suggest that the polymerase domain of the p66 monomer of RT also binds to a template-primer in a manner similar to that of pol β (Fig. 7C). Some evidence of this exists in that, as with KF and RNAP, RT also seems to possess the structural equivalent to helices M and N of pol β. Though not quite as obvious and pronounced as in the other polymerases, beta strands 12 and 13 of RT do protrude into the template-primer binding channel near the catalytically important carboxylic acid residues in an area that overlaps quite well with helices M and N of pol β and helices J and K of KF, implicating beta strands 12 and 13 of RT as possibly playing a similar role in substrate specificity.

In disagreement with this proposal is the crystal structure of RT complexed with a template-primer, which shows that HIV-1 RT has bound to the template-primer in a manner opposite to that of pol β (*12*) (Fig. 7D may serve as an approximate representation of that structure). We can present some of the same reservations as to the physiological relevance of the RT-DNA complex as were presented above for the KF-DNA complex: (i) The RT-DNA complex crystals were grown at relatively high, nonphysiological salt concentrations; (ii) the template-primer in the crystallization experiments was odd in that it had only a single base template overhang; and (iii) the complex crystals were isomorphous with the apo structure crystals. However, unlike KF, there exists strong evidence in favor of the idea that the RT nevertheless binds to a template-primer in the manner depicted in Fig. 7D, regardless of the physiological relevance of the RT-DNA crystal structure. This evidence comes from kinetic studies that show a tight temporal coupling between the polymerase and RNase H activities of RT during the reverse transcription process (*83*). These results suggest that RT binds to the template-primer in such a way that the RNase domain comes in contact with the RNA template of the emerging RNA-DNA duplex during reverse transcription, which would be possible only if the anti–pol β mode of binding were employed, as shown in Fig. 7D.

Although we have presented strong structural evidence in favor of the idea that RT binds to the template-primer in a manner similar to that of pol β, KF, and RNAP, there seems to be equally strong kinetic evidence (*83*) in support of an anti–pol β mode of template-primer binding for RT during RNase H coupled polymerization activity. Thus it may be that RT can bind to the template-primer in two different catalytically competent ways; an RNase H independent, pol β–like mode of binding and an RNase H–coupled, anti–pol β mode of binding.

The method by which RT makes double-stranded DNA copies of the single-stranded RNA genome is quite complicated (*82*), so it is not too difficult to imagine that one enzyme performing so many different functions might employ two modes of binding to a template-primer (Fig. 7, C and D). Although this appears to be in disagreement with our arguments that the asymmetric geometry of the polymerase active site must be conserved, closer inspection of the RT active site reveals that our arguments may nevertheless hold true. As it turns out, the 180° rotation performed on the RT in Fig. 7C to produce Fig. 7D resulted in an active site where Asp^{185} and Asp^{186} of RT (the equivalent to Asp^{190} and Asp^{192} in pol β) have simply switched positions, and Asp^{110} (the equivalent to pol β's Asp^{256}) has been replaced by Tyr^{183}. Instead of Asp^{110}, Tyr^{183} is now near the primer 3′ terminus for RT in the anti–pol β mode of template-primer binding, suggesting that perhaps Tyr^{183} performs some of the same functions that Asp^{110} performs when RT is in a pol β–like binding mode. Thus the geometry of the active site seems to be conserved. In support of this idea, the YMDD sequence of RT (Tyr^{183}, Met^{184}, Asp^{185}, Asp^{186}) is the most highly conserved amino acid sequence in all known RTs, and mutations to the tyrosine residue of this segment have shown it to be highly critical for catalysis (*84*). None of the other polymerases discussed, neither pol β, KF, nor RNAP, has a structural equivalent to Tyr^{183} in RT.

Another aspect of the models in Fig. 7, C and D, is that a much smaller part of RT is in contact with duplex DNA in Fig. 7C, possibly suggesting a more distributive mode of synthesis when RT invokes a pol β–like mode of template-primer binding. A more distributive mode of polymerization has been observed for RT when a DNA template is utilized instead of an RNA template (*85, 86*). All this leads us to propose that RT might use a pol β–like mode of binding for a DNA template, and the reverse, anti–pol β mode of binding for an RNA template, thus making it a reverse transcriptase in another sense of the word. The anti–pol β mode of template-primer binding (Fig. 7D) could be operative during the highly processive, RNase H–coupled, RNA-directed DNA polymerization of the viral minus strand (*82*), whereas a pol β–like mode of template-primer binding (Fig. 7C) could be employed during the more distributive DNA-directed DNA polymerization of the viral plus strand (*82*), when RNase H coupled polymerization is no longer required. The most salient feature of this proposal is that RT may distinguish between the two types of DNA substrates involved: the A form of an RNA-DNA hybrid versus the B form of a DNA duplex. For instance, RT might possess the structural equivalents of pol β's Lys^{234}, Tyr^{271}, and Arg^{283} (which, as discussed above, break up the water structure of the minor groove near the active site) only when RT employs a pol β–like mode of binding. If all DNA-directed DNA polymerization steps require that the water structure in the minor groove of the B-DNA template-primer be disrupted, this would make it unfavorable for RT to bind to a B form substrate in any other way but in a pol β–like fashion.

Although the aforementioned proposals restrict RT from utilizing an anti-pol β mode of binding on B form substrates, they do not necessarily restrict RT from utilizing a pol β–like mode of binding on A-form substrates, which already possess a broad minor groove. It is therefore feasible that other, possibly non–RNase H coupled steps during the replication of the viral genome, such as tRNA primed synthesis of the minus strand strong stop DNA or synthesis of the RNase H insensitive polypurine tract (ppt) (*82*), might employ a pol β–like mode of template-primer binding for RT as well, despite the fact that an RNA template is utilized during these steps. That it might be possible for RT, under certain circumstances, to employ the same mode of bind-

ing for an RNA template as for a DNA template is supported by kinetic studies showing that pol I, under the appropriate conditions, can utilize an RNA template almost as efficiently as its natural DNA template (*87*). In the case of the synthesis of minus stand strong stop DNA, the length of the tRNA primer (18 nt) is the same as the distance between the polymerase active site and the RNase H active site (about 18 bp) (*83*). Accordingly we suggest that perhaps the bulky tRNA molecule attached to the primer strand functions as a steric hindrance to binding at the RNase H active site, forcing RT to use the pol β–like mode of template-primer binding during this step of the cycle. In favor of this argument is the observation that RT polymerization during minus strand strong stop synthesis appears to be more distributive than during reverse transcription of the viral genome (*88, 89*). Also supporting this idea are results from primer utilization studies showing that tRNA primed synthesis of minus strand strong stop DNA, at least in vitro, does not require a specific tRNA such as human $tRNA^3_{Lys}$, which is utilized by RT in vivo (*90*).

Further indication of a possible dual mode of template-primer binding by RT comes from many independent studies. Active site titration studies showed that the RT heterodimer possesses a possible second template-primer binding site, but what is most intriguing about these results is that this second binding site only reveals itself after the template strand of the template-primer has been shortened to 16 nt (*91*). Along the same lines, although a model proposing an RNA-DNA-RNA intermediate for a RT strand transfer mechanism does not suggest a second mode of template-primer binding (*92*), perhaps some of the kinetic and crosslinking data leading up to that model can also be interpreted as evidence in favor of our arguments. Furthermore, kinetic studies on the individual p66 and p51 monomers of RT showed that both of these monomers are fully capable of catalyzing the nucleotidyl transfer reaction, but only under optimal conditions for each monomer (*93*). That the optimal template for p66 was RNA and the optimal template for p51 was DNA strengthened the proposal that the RT heterodimer "may be functionally asymmetric with distinct plus and minus strand polymerases" (*93*). It was further suggested that the p66 monomer of the p66-p51 RT heterodimer was responsible for RNA-directed DNA polymerase functions (reverse transcription), whereas the p51 monomer of the heterodimer was responsible for the DNA-directed DNA polymerization activities of RT (*93*). However, as discussed above, the crystal structure of the p66-p51 RT heterodimer clearly shows that the p51 monomer does not possess a polymerase fold, at least while p51 RT is a part of the p66-p51 heterodimer (*11, 12*). Therefore, our hypothesis that the p66 monomer of the p66-p51 RT heterodimer may act as two different polymerases that happen to share a common active site is not only in agreement with the idea that the RT heterodimer is functionally asymmetric with distinct plus and minus strand polymerases (*93*), but is in accord with the structural work (*11, 12*) as well.

Finally, a dual mode of template binding by RT might go a long way toward explaining why drugs such as AZT can specifically target HIV-1 RT in preference to host cell polymerases. Pol β, KF, RNAP, and even RT (when in a pol β–like mode of template-primer binding) all seem to possess structural features near the active site, such as helices M and N of pol β, that may be responsible for nucleotide selectivity. However, there appears to be no structural equivalent to helices M and N of pol β when RT is in an anti–pol β mode of template-primer binding, which might explain why RT is so error prone and more sensitive to AZT-TP inhibition. In support of this idea are kinetic studies that show decreasing K_i values for AZT-TP in its binding to pol β (K_i = 73 μM) and to RT (K_i = 0.3 μM) when calf thymus DNA is utilized as a template in both cases, as compared to RT when the native RNA template is utilized (K_i = 0.01 μM) (*55*). The decreasing order of K_i values (tighter binding) might reflect the decreasing number of structural elements in the active site that can, through steric hindrance, induce nucleotide specificity, the order of structural hindrance being: the pol β's helices M and N ≫ RT's β strand 12 and 13 (when employing a pol β–like mode of template-primer binding on a DNA template) > RT when employing an anti–pol β mode of binding on an RNA template. With this in mind, one could imagine how AZT, ddC, and ddI resistant RT mutants might arise in AIDS patients after prolonged treatment with these drugs (*94–97*). As discussed above, much of the nucleotide binding pocket in pol β is determined by the exact position of the template-primer, so a single mutation that affects how RT binds to the template-primer could in turn affect the nucleotide binding site. Indeed, it has been proposed that many of the drug resistant mutations in RT, because they are so distant from the active site, interact primarily with the template strand (*11*). Perhaps these RT mutants affect template-primer binding in such a way as to introduce structural elements, similar to helices M and N of pol β, into the nucleotide binding pocket of RT, possibly causing greater selectivity and more resistance to these anti-HIV drugs.

REFERENCES AND NOTES

1. A. Kornberg and T. A. Baker, *DNA Replication* (Freeman, New York, ed. 3, 1991).
2. R. Yarchoan, H. Mitsuya, C. E. Myers, S. Broder, *N. Engl. J. Med.* **321**, 726 (1989).
3. R. F. Schinazi, *Perspect. Drug Discovery Design* **1**, 151 (1993).
4. H. Mitsuya *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 2033 (1987).
5. P. A. Furman *et al.*, *ibid.* **83**, 8333 (1986).
6. ddI is the deamination product of ddA; it is thought that ddIMP, in vivo, is converted to ddAMP by adenylosuccinate synthetase-lyase, which is then converted to ddATP by further phosphorylation [M. A. Johnson *et al.*, *J. Biol. Chem.* **263**, 15354 (1988)].
7. M. A. Waqer, M. J. Evans, K. F. Manly, R. G. Hughes, J. A. Huberman, *J. Cell. Physiol.* **121**, 402 (1984).
8. J.-P. Sommadossi, R. Carlisle, Z. Zhou, *Mol. Pharmacol.* **36**, 9 (1989).
9. Y. C. Cheng, W. Y. Gao, C. H. Chen, M. Vazquez-Padua, M. C. Starnes, *Annals N. Y. Acad. Sci.* **616**, 217 (1990).
10. D. L. Ollis, P. Brick, R. Hamlin, N. G. Xoung, T. A. Steitz, *Nature* **313**, 762 (1985).
11. L. A. Kohlstaedt, J. Wang, J. M. Friedman, P. A. Rice, T. A. Steitz, *Science* **256**, 1783 (1992).
12. A. Jacobo-Molina *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 6320 (1993).
13. R. Sousa, Y. J. Chung, J. P. Rose, B. C. Wang, *Nature* **364**, 593 (1993).
14. M. R. Sawaya, H. Pelletier, A. Kumar, S. H. Wilson, J. Kraut, *Science* **260**, 1930 (1994).
15. M. Delarue, O. Pooh, N. Tordo, D. Moras, P. Argos, *Protein Eng.* **3**, 461 (1990).
16. S. Linn, *Cell* **66**, 185 (1991).
17. T. M. Jenkins, J. K. Saxena, A. Kumar, S. H. Wilson, E. J. Ackerman, *Science* **258**, 475 (1992).
18. D. C. Rein, A. J. Recupero, M. P. Reed, R. R. Meyer, *The Eukaryotic Nucleus: Molecular Biochem. and Macromolecular Assembly*, P. R. Strauss and S. H. Wilson, Eds. (Telford, Caldwell, NJ, 1990), vol. 1, chap. 5.
19. S. H. Wilson, *ibid.*, chap. 8.
20. Certain aspects of the aging process, as well as some types of cancer, may be linked to a breakdown in the DNA repair mechanisms of the cell and hence to pol β in particular. Supporting this view, deleterious mutations in the gene encoding for pol β have been observed in blood samples from patients with Werner syndrome (WS), a rare autosomal recessive disorder characterized by a high DNA mutation rate and the appearance of premature aging [Y. Sadakane, K. Maeda, Y. Kuroda, K. Hori, *Biochem. Biophys. Res. Commun.* **200**, 219 (1994)] and also in tumor samples from patients with colorectal cancer [L. Wang, U. Patel, L. Ghosh, S. Banerjee, *Cancer Res.* **52**, 4824 (1992)].
21. S. S. Carroll and S. J. Benkovic, *Chem. Rev.* **90**, 1291 (1990).
22. R. Green and D. Korn, *J. Biol. Chem.* **245**, 254 (1970). Pol β is so efficient at incorporating mismatched base pairs that it was originally questioned whether pol β utilizes a template at all during DNA polymerization. Fueling the debate was the observation that the only other polymerase sharing sequence similarity with pol β—terminal deoxynucleotidyl-transferase (TdT)—is not a template-directed polymerase.
23. P. S. Freemont, J. M. Friedman, L. S. Beese, M. R. Sanderson, T. A. Steitz, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8924 (1988).
24. L. S. Beese and T. A. Steitz, *EMBO J.* **10**, 25 (1991).
25. B. Z. Zmudzka *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 5106 (1986).
26. A. Kumar *et al.*, *J. Biol. Chem.* **265**, 2124 (1990).
27. The oligonucleotides that were to serve as template and primer, 5′-AAAGGGCGCCG-3′ and 5′-CGGCGC-3′, respectively, were purchased (Integrated DNA Technologies, Inc., Coralville, IA) and stored at −20°C. Occasional DNA samples

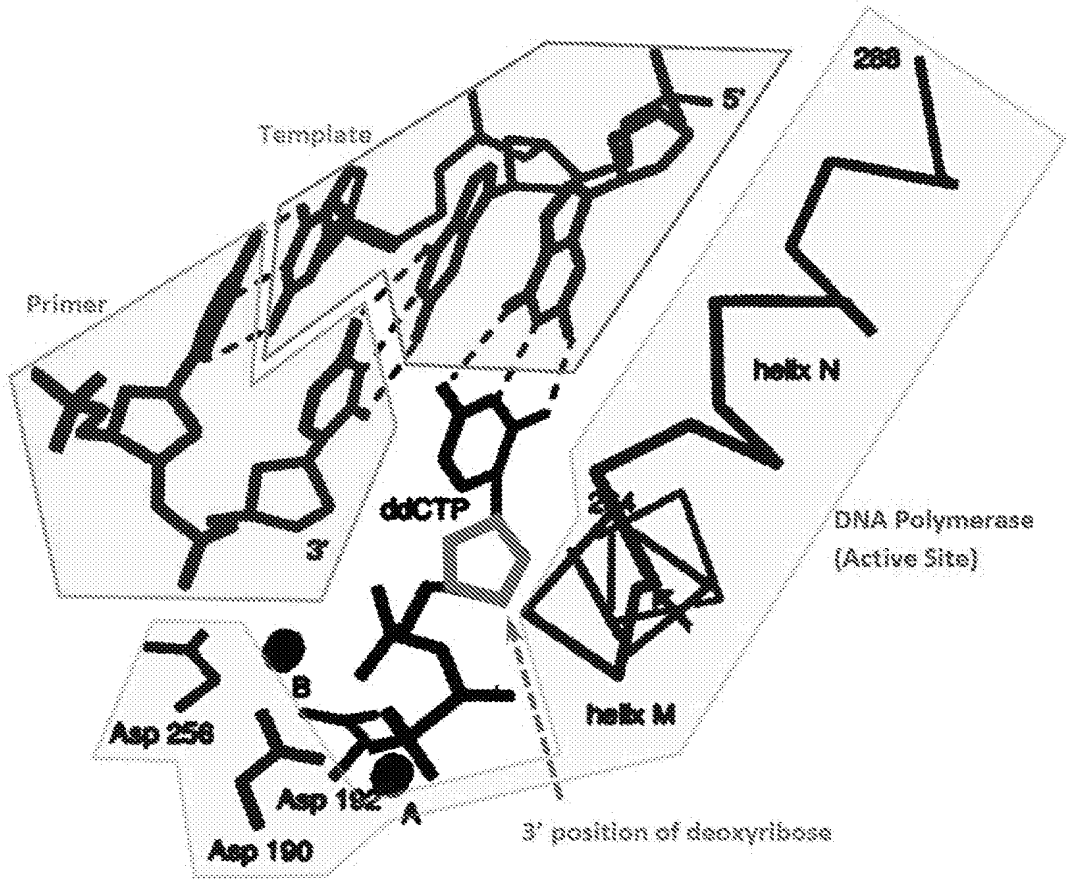
that had a slight yellow hue and were difficult to dissolve in the buffer solution inevitably failed to produce complex crystals. We therefore recommend that a reversed-phase cartridge, or other DNA purification techniques, be employed before the crystallization experiments.
28. A more common method for annealing the DNA template-primer, in which samples are heated to 90° for 3 minutes then allowed to cool slowly to room temperature, was attempted in order to obtain better crystals, but this method resulted in crystals with no noticeable improvements in diffraction power.
29. The ddCTP was purchased (Sigma) as 4 μmol samples and stored at −20°C. The H_2O used throughout all crystallization procedures had been deionized, then distilled, prior to use.
30. Crystals were grown at room temperature in MVD-24 sitting drop trays (Charles Supper Co.), which were subsequently sealed with clear packaging tape (Manco, purchased from Sears) to allow vapor diffusion between the reservoir solution [7 to 9 percent (w/v) PEG 3350, 0.1 M MES, 75 mM lithium sulfate] and the crystallization medium, which was made by mixing 20 μl of the reservoir solution with 20 μl of the protein-DNA-ddCTP sample.
31. AZT-TP was provided by R. F. Schinazi, Emory University, Georgia.
32. A partial (31-kD) structure was determined by molecular replacement techniques, then refined, in a manner very similar to that described in the text for the ternary complex structures. It was difficult to interpret $F_o - F_c$, α_c, difference maps at this resolution (4 Å), but what little positive electron density that was observed in the maps was located adjacent to the fingers subdomain of pol β and was attributed to the missing 8-kD domain. No further work has been done on this structure.
33. L. S. Beese, J. M. Friedman, T. A. Steitz, *Biochemistry* **32**, 14095 (1993).
34. D. N. SenGupta *et al.*, *Biochem. Biophys. Res. Commun.* **136**, 341 (1986).
35. J. Abbotts *et al.*, *Biochemistry* **27**, 901 (1988).
36. A. T. Brunger, *XPLOR Manual* (Yale Univ. Press, New Haven, CT, version 3.1, 1992).
37. A. T. Brunger, *Acta Crystallogr.* **A46**, 46 (1990).
38. The greater difficulty in obtaining clear translation solutions for the $P2_1$ structure could be attributed to the fact that the $P2_1$ structure contains two molecules per asymmetric unit instead of one, as is the case for the $P6_1$ structure. Also, diffraction data on the $P2_1$ crystals were limited to a lower resolution than diffraction data on the $P6_1$ crystals. We were originally unable to obtain clear translation solutions for the $P2_1$ crystals from a low resolution (4.0 Å) data set, and the molecular replacement structure determination of the $P2_1$ crystals had to await the growth of slightly better crystals and the somewhat higher resolution data (3.6 Å) that we present (Table 1).
39. D. E. Tronrud, L. F. TenEyck, B. W. Matthews, *Acta Crystallogr.* **A43**, 489 (1987).
40. R. Lavery and H. Sklenar, *J. Biomol. Struct. Dynam.* 6, 655 (1989).
41. L. S. Beese, V. Derbyshire, T. A. Steitz, *Science* **260**, 352 (1993).
42. R. Prasad, W. A. Beard, S. H. Wilson, *J. Biol. Chem.*, in press.
43. R. K. Singhal and S. H. Wilson, *ibid.* **268**, 15906 (1993).
44. A. Kumar, J. Abbotts, E. M. Karawya, S. H. Wilson, *Biochemistry* **29**, 7156 (1990).
45. J. Warwicker, D. Ollis, F. M. Richards, T. A. Steitz, *J. Mol. Biol.* **186**, 645 (1985).
46. B. N. Conner, T. Takano, S. Tanaka, K. Itakura, R. E. Dickerson, *Nature* **295**, 294 (1982).
47. H. R. Drew and R. E. Dickerson, *J. Mol. Biol.* **151**, 535 (1981).
48. K. Tanabe, E. W. Bohn, S. H. Wilson, *Biochemistry* **18**, 3401 (1979).
49. W. R. McClure, T. M. Jovin, *J. Biol. Chem.* **250**, 4073 (1975).
50. C. Majumdar, J. Abbotts, S. Broder, S. H. Wilson, *ibid.* **263**, 15657 (1988).
51. R. D. Kuchta, V. Mizrahi, P. A. Benkovic, K. A. Johnson, S. J. Benkovic, *Biochemistry* **26**, 8410 (1987).
52. J. E. Reardon and W. H. Miller, *J. Biol. Chem.* **265**, 20302 (1990).
53. Interaction of Gly^{274} with ddCTP is noteworthy because Gly^{274} is at the COOH-terminus of helix M and, as is mentioned earlier, forms a cis-peptide with Ser^{275} of helix N.
54. J. H. Van de Sande, P. C. Loewen, H. G. Khorana, *J. Biol. Chem.* **247**, 6140 (1972).
55. M. H. St. Clair *et al.*, *Antimicrob. Agents Chemother.* **31**, 1972 (1987).
56. J. E. Reardon, *Biochemistry* **31**, 4473 (1992).
57. T. Date, S. Yamamoto, K. Tanihara, Y. Nishimoto, A. Matsukage, *ibid.* **30**, 5286 (1991).
58. A. H. Polesky, T. A. Steitz, N. D. F. Grindley, C. M. Joyce, *J. Biol. Chem.* **265**, 14579 (1990).
59. A. H. Polesky, M. E. Dahlberg, S. J. Benkovic, N. D. F. Grindley, C. M. Joyce, *ibid.* **267**, 8417 (1992).
60. B. A. Larder, D. J. M. Purifoy, K. L. Powell, G. Darby, *Nature* **327**, 716 (1987).
61. P. M. J. Burgers and F. Eckstein, *J. Biol. Chem.* **254**, 6889 (1979).
62. V. Derbyshire *et al.*, *Science* **240**, 199 (1988).
63. V. Derbyshire, N. D. F. Grindley, C. M. Joyce, *EMBO J.* **10**, 17 (1991).
64. A. P. Gupta and S. J. Benkovic, *Biochemistry* **23**, 5874 (1984).
65. In the refined $P6_1$ ternary complex structure, the temperature factor for the Mg^{2+} ion in site B (79 $Å^2$) was much greater than that of the Mg^{2+} ion in site A (6 $Å^2$). Because the electron density in site B could also have been interpreted as a water molecule, we had to rely on local geometry to distinguish the Mg^{2+} ion in site B from a water; in particular, a short coordinate bond of 2.0 Å to the oxygen of Asp^{190} (Table 3) is characteristic of a metal ion. The low occupancy of the metal ion in site B may be attributed to the fact that one of its ligands, the 3′ oxygen of the primer terminus, is missing in the crystal structure. Because we have consistently observed [(*14*), H. Pelletier and M. R. Sawaya, unpublished results] that only the metal ion in site B is present in 31-kD pol β crystals that have been soaked in the presence of $MnCl_2$, we believe that the second metal ion (site A) most likely enters the active site with the nucleotide substrate as a β,γ-bidentate and exists as M^{2+}-PP_i.
66. T. R. Cech, *Science* **236**, 1532 (1987).
67. A. M. Pyle, *ibid.* **261**, 709 (1993).
68. T. A. Steitz and J. A. Steitz, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 6498 (1993).
69. M. D. Been and T. R. Cech, *Science* **239**, 1412 (1988).
70. D. P. Bartel and J. W. Szostak, *ibid.* **261**, 1411 (1993).
71. C. M. Joyce and T. A. Steitz, *Trends Biochem. Sci.* **12**, 288 (1987).
72. T. A. Steitz, L. Beese, P. S. Freemont, J. M. Friedman, M. R. Sanderson, *Cold Spring Harbor Symp. Quant. Biol.* **52**, 465 (1987).
73. D. Patra, E. M. Lafer, R. Sousa, *J. Mol. Biol.* **224**, 307 (1992).
74. C. A. Raskin, G. Diaz, K. Joho, W. T. McAllister, *ibid.* **228**, 506 (1992).
75. D. K. Muller, C. T. Martin, J. E. Coleman, *Biochemistry* **27**, 5763 (1988).
76. T. Date *et al.*, *ibid.* **29**, 5027 (1990).
77. A. Basu, P. Kedar, S. H. Wilson, M. J. Modak, *ibid.* **28**, 6305 (1989).
78. E. Gilboa, S. W. Mitra, S. Goff, D. Baltimore, *Cell* **18**, 93 (1979).
79. F. Barré-Sinoussi *et al.*, *Science* **220**, 868 (1983).
80. M. Popovic, M. G. Sarngadharan, E. Read, R. C. Gallo, *ibid.* **224**, 497 (1984).
81. R. C. Gallo *et al.*, *ibid.*, p. 500.
82. S. P. Goff, *J. Acquired Immune Deficiency Syndrome* **3**, 817 (1990).
83. V. Gopalakrishnan, J. A. Peliska, S. J. Benkovic, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 10763 (1992).
84. S. A. Jablonski and C. D. Morrow, *J. Virol.* **67**, 373 (1993).
85. H. E. Huber, J. M. McCoy, J. S. Seehra, C. C. Richardson, *J. Biol. Chem.* **264**, 4669 (1989).
86. J. E. Reardon, *ibid.* **268**, 8743 (1993).
87. M. Ricchetti and H. Buc, *EMBO J.* **12**, 387 (1993).
88. W. A. Haseltine, D. G. Kleid, A. Panet, E. Rothenberg, D. Baltimore, *J. Mol. Biol.* **106**, 109 (1976).
89. L. I. Lobel and S. P. Goff, *J. Virol.* **53**, 447 (1985).
90. L. A. Kohlstaedt and T. A. Steitz, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 9652 (1992).
91. W. A. Beard and S. H. Wilson, *Biochemistry* **32**, 9745 (1993).
92. J. A. Peliska and S. J. Benkovic, *Science* **258**, 1112 (1992).
93. R. L. Thimmig and C. S. McHenry, *J. Biol. Chem.* **268**, 16528 (1993).
94. B. A. Larder, G. Darby, D. D. Richman, *Science* **243**, 1731 (1989).
95. B. A. Lardar and S. D. Kemp, *ibid.* **246**, 1155 (1989).
96. M. H. St. Clair *et al.*, *ibid.* **253**, 1557 (1991).
97. J. E. Fitzgibbon *et al.*, *Antimicrob. Agents Chemother.* **36**, 153 (1992).
98. R. Hamlin, *Methods Enzymol.* **114**, 416 (1985).
99. A. J. Howard, C. Nielsen, N. H. Xuong, *ibid.*, p. 452 (1985).
100. P. J. Kraulis, *J. Appl. Cryst.* **24**, 946 (1991).
101. Supported in part by NIH grants GM10928 and CA17374 (J.K.), NIH grant ES06839, R. A. Welch Foundation grant H-1265 (S.H.W.), and a grant of computing time from the San Diego Supercomputer Center. We thank the personnel of the N. H. Xuong Laboratory for aid in data collection; C. Nielsen, N. Nguyen, D. Sullivan, V. Ashford, and W. Wolfle for technical assistance in protein preparation. Full coordinates for both ternary complex structures are available from the Brookhaven Protein Data Bank and are designated 1bpf and 1bpg for the $P6_1$ and $P2_1$ structures, respectively.

28 February 1994; accepted 10 May 1994

Analysis of Space Available Within the Active Site of a DNA Polymerase Ternary Complex (*polymerase, DNA template/primer, nucleotide*) to Accommodate a 3'-Capped dNTP

Based on the 3-dimensional structure of the ternary complex (*polymerase, DNA template/primer, nucleotide*) determined by Pelletier et al. (Pelletier et al. "*Structures of ternary complexes of rat DNA polymerase beta, a DNA template-primer, and ddCTP.*" *Science* 1994, 264, 1891-1903), which is cited in U.S. Serial No. 15/380,284 (Ju et al. *Massive parallel method for decoding DNA and RNA*), an analysis was performed to determine the space available for a 3'-O capping group on the 3' carbon of the deoxyribose of the nucleotide. The results indicate that there is only a small space available between amino acids in the active site of the polymerase and the 3' carbon of the deoxyribose of the nucleotide, as shown in the Figure below (corresponding to Fig. 1 of U.S. Serial No. 15/380,284 and to Fig. 6 of Pelletier et al.; color and labels added for clarity). This space can only accommodate a capping group of limited diameter on the 3' position of the deoxyribose of the nucleotide. Pelletier et al. (1994) determined that three amino acids of the polymerase, Tyr 271, Phe272, and Gly274, are in close proximity to the 3' carbon of the deoxyribose of the nucleotide. (Pelletier et al. 1994, Table 3). In Table 3 Pelletier et al. highlight the distances from the nucleotide to these amino acids in the polymerase ternary complex as follows: 3.2 Å between the 3' carbon of the deoxyribose ring and Phe272; 3.2 Å between the 2' carbon of the deoxyribose ring and Gly274; and 3.5 Å between the 2' carbon and Tyr271.

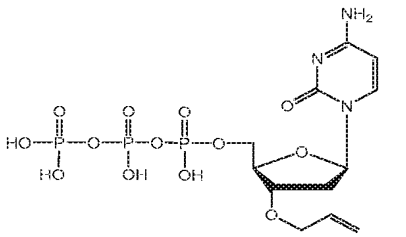
The distances given in Pelletier et al. were used to calculate the available space around the 3' carbon of the deoxyribose ring of the nucleotide. It was determined that the diameter of the available space in the active site of the polymerase ternary complex is approximately 3.7 Å.



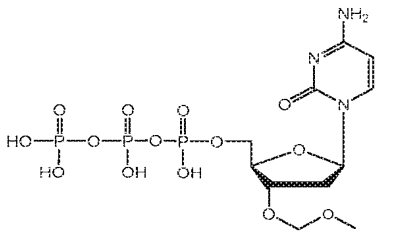
The Figure above from Pelletier et al. (1994) shows the 3-dimensional structure of the ternary complexes of rat DNA polymerase, a DNA template-primer, and dideoxycytidine triphosphate. The active site of the polymerase

is highlighted. Note that the 3' position of the dideoxyribose ring (blue) is very crowded. (See also Brief Description of the Figures of U.S. Serial No. 15/380,284 for FIG. 1).

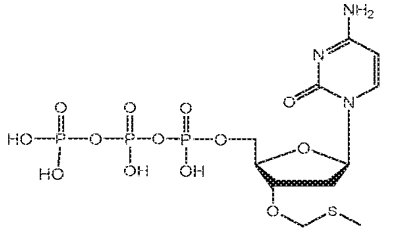
The diameters of five groups possibly useful as 3'-oxygen capping groups for a nucleotide analogue (allyl, methoxymethyl (MOM), methylthiomethyl, azidomethyl, and 2-nitrobenzyl, whose structures are shown below) were calculated. The calculations were based on the lengths and angles of the bonds beyond the 3' carbon of the deoxyribose ring of the nucleotide, and were determined using *Chem3D Pro* software (MC Squared, "*Software Review: CS Chem3D Pro 3.5 and CS MOPAC Pro*" *Electronic Journal of Theoretical Chemistry*, 1997, 2, 215--217).



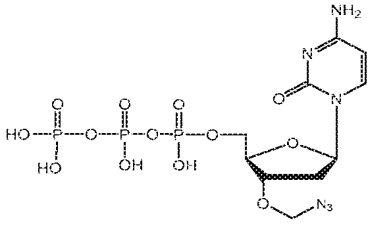
3'-*O*-Allyl-dCTP



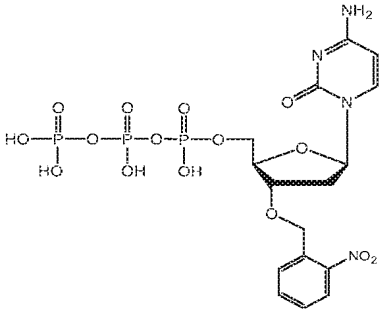
3'-*O*-Methoxymethyl-dCTP



3'-*O*-Methylthiomethyl-dCTP



3'-*O*-Azidomethyl-dCTP



3'-*O*-2-Nitrobenzyl-dCTP

The calculated diameter (D) for each group is as follows:

1. Allyl (--CH_2-CH=CH_2): D = 3.0 Å

2. Methoxymethyl (MOM; --CH_2-OCH_3): D= 2.1 Å

3. Methylthiomethyl (--CH_2-SCH_3): D= 2.4 Å

4. Azidomethyl (--CH_2-N_3): D= 2.1 Å

5. 2-Nitrobenzyl (--$C_7H_6O_2N$): D = 5.0 Å

Conclusion:

The available space in the active site of the polymerase around the 3' position of the deoxyribose ring of the nucleotide in the polymerase ternary complex has a diameter of approximately 3.7 Å. The allyl, MOM, methylthiomethyl and azidomethyl groups (diameters ranging from 2.1 Å to 3.0 Å) will fit into the available space, but the rigid aromatic 2-nitrobenzyl group (diameter 5.0 Å) will not fit into the active site of the polymerase due to its larger diameter.

Nucleic Acids Research, 1994, Vol. 22, No. 20 4259–4267

Termination of DNA synthesis by novel 3′-modified-deoxyribonucleoside 5′-triphosphates

Michael L.Metzker*, Ramesh Raghavachari[1,+], Stephen Richards, Swanee E.Jacutin[1], Andrew Civitello, Kevin Burgess[1] and Richard A.Gibbs
Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030 and [1]Department of Chemistry, Texas A & M University, College Station, TX 77843, USA

Received June 3, 1994; Revised and Accepted July 25, 1994

ABSTRACT

Eight 3′-modified-dNTPs were synthesized and tested in two different DNA template assays for incorporation activity. From this enzymatic screen, two 3′-*O*-methyl-dNTPs were shown to terminate DNA syntheses mediated by a number of polymerases and may be used as alternative terminators in Sanger sequencing. 3′-*O*-(2-Nitrobenzyl)-dATP is a UV sensitive nucleotide and was shown to be incorporated by several thermostable DNA polymerases. Base specific termination and efficient photolytic removal of the 3′-protecting group was demonstrated. Following deprotection, DNA synthesis was reinitiated by the incorporation of natural nucleotides into DNA. The identification of this labile terminator and the demonstration of a one cycle stop-start DNA synthesis are initial steps in the development of a novel sequencing strategy.

INTRODUCTION

2′-Deoxyribonucleoside-5′-triphosphates (dNTPs) modified at their 3′-hydroxyl position can act as terminators of enzyme-directed DNA synthesis (1–12). These nucleotide analogs are useful as DNA sequencing tools, mechanistic probes, antimetabolites, and as antiviral agents. Consequently, such compounds have been used for analytical and therapeutic studies (2). Overall, however, the number of compounds that are well characterized is small, and there is considerable scope for new combinations of terminators and polymerases to be identified.

Among the most familiar terminators of DNA synthesis are the 2′, 3′-dideoxyribonucleoside-5′-triphosphates (ddNTPs) that are the basis for Sanger DNA sequencing (13). In that method oligonucleotide-primed DNA or RNA templates are enzymatically extended in a 5′ → 3′ direction in the presence of a mixture of dNTPs and ddNTPs to generate a population of molecules that are terminated at specific base positions. DNA fragments of different lengths are resolved by denaturing polyacrylamide gel electrophoresis and detected either by radioactive or fluorescent labels to reveal the underlying base sequence. Despite the obvious limitations of gel electrophoresis for sequencing long DNA strands, this method has been the favored approach for more than ten years (13,14).

Improvements to the Sanger protocols are being sought to meet the increasing demands of large scale sequencing of whole genomes (14). We and others (15–18) have independently conceived a radically different, gel-free alternative to the Sanger scheme for DNA sequencing. This method, called the Base Addition Sequencing Scheme (BASS), is based on novel nucleotide analogs that terminate DNA synthesis. BASS involves repetitive cycles of incorporation of each successive nucleotide, *in situ* monitoring to identify the incorporated base, and deprotection to allow the next cycle of DNA synthesis, (Figure 1). Compared to Sanger sequencing, BASS has two major advantages: base resolution would not require gel electrophoresis and there is a tremendous capacity for simultaneous analyses of multiple samples. The complete scheme demands nucleotide analogs that are tolerated by polymerases, spectroscopically distinct for each base, stable during the polymerization phase, and deprotected efficiently under mild conditions in aqueous solution. These stringent requirements are formidable obstacles for the design and synthesis of the requisite analogs.

The investigation of the interactive patterns between various terminating analogs and different enzymes is an important preliminary phase in the development of the BASS method. Consequently, eight 3′-modified-dNTPs were synthesized and examined for their ability to terminate DNA synthesis mediated by a variety of polymerases. The majority of 3′-modified analogs have labile protecting groups that have the potential to be incorporated into BASS. Active combinations of terminators and enzymes were identified using two different primer-template gel assays. One of these compounds, 3′-*O*-(2-nitrobenzyl)-dATP [7], was used to demonstrate one complete cycle of termination, deprotection, and reinitiation of DNA synthesis.

*To whom correspondence should be addressed

+Present address: LI-COR Inc., Biotechnology Division, Lincoln, NE 68504-5000, USA

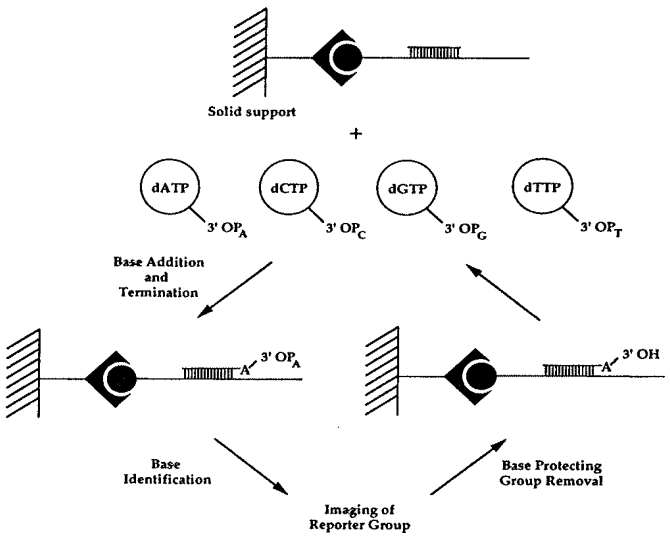


Figure 1. Cartoon of the Base Addition Sequencing Scheme (BASS). A primer is annealed to a biotinylated-labeled template bound to a solid support. Four deoxynucleotides triphosphates that have spectroscopically unique blocking groups attached to the 3'-position are added. Polymerase extension is terminated after the addition of one base. Upon imaging of the reporter group, the protecting group is removed resulting in a 'free' 3'-OH group, allowing the addition of the next complement base.

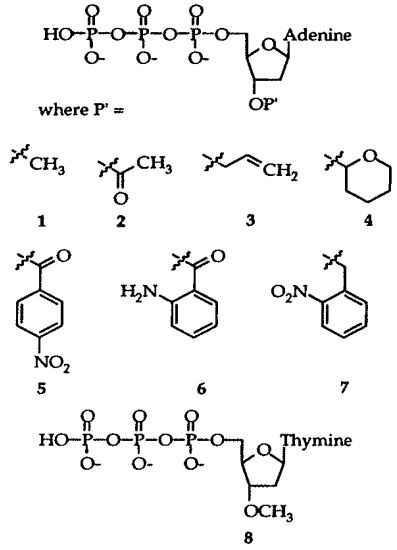


Figure 2. Chemical structures of the 3'-modified-nucleotides. Details of the chemical syntheses are described in the Materials and Methods section.

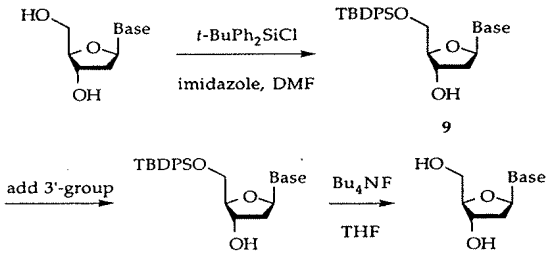
MATERIALS AND METHODS

General

High field Nuclear Magnetic Resonance (NMR) spectra were recorded on a Bruker AC250 (^{1}H at 250 MHz, ^{13}C at 62.9 MHz, ^{31}P at 101.26 MHz) or a Varian XL 200 (^{1}H at 200 MHz, ^{13}C at 50 MHz, ^{31}P at 81 MHz). Ultraviolet (UV) spectra were recorded on a Hewlett-Parkard model 8452A diode array spectrophotometer. Thin layer chromatography was performed on Whatman silica gel 60 A F_{254} plates. Flash chromatography was performed on SP silica gel 60 (230–600 mesh ASTM). Ion-exchange chromatography was performed on Fluka DEAE cellulose C451 (HCO_3^- form). Photodecomposition of 3'-*O*-(2-nitrobenzyl)-dATP [7] was performed using a FisherBiotech transilluminator.

Organic syntheses

The chemical structures of compounds [1]–[8] are shown in Figure 2. Compounds [1]–[4], [6], [8] were prepared according to the general scheme:



The 5'-hydroxyl was protected with a *tert*-butyldiphenylsilyl (TBDPS) group, and the specific addition of the 3'-protecting groups (P') are described below. Desilylations were performed by the addition of 1.0 equiv. of tetrabutylammonium fluoride (Bu_4NF) to the 3'-protected-5'-silyl-adenosine or thymidine derivatives. The reactions were monitored by TLC; after completion (ca. 15 min.), the reactions were quenched with 1.0 equiv. of glacial acetic acid. The solvent was removed, and the residues were purified by silica column chromatography (10% methanol/ethyl acetate).

2'-Deoxy-3'-O-methyladenosine [1]. To 2'-deoxy-5'-*tert*-butyldiphenylsilyladenosine [9] (200 mg, 0.4 mmol) in benzene (5 mL), methyl iodide (568 mg, 4.0 mmol, 10 equiv.), tetrabutylammonium hydroxide (TBAH) (40% solution, 325 μL), and 1 M NaOH (5 mL) were vigorously stirred at 25°C for 16 h. The organic layer was extracted with ethyl acetate and washed with deionized (D.I.) water, saturated NaCl, dried over Na_2SO_4 and purified by flash chromatography using a stepwise gradient (0% methanol/ethyl acetate to 5% methanol/ethyl acetate in 2% intervals) (180 mg, 89%) (19).

The *O*-methyl derivative from the above procedure (80 mg, 0.16 mmol), after desilylation and flash chromatography gave compound [1] as colorless crystals (30 mg, 70%). High resolution mass spectrometry (HRMS) *m/e* calculated for $C_{11}H_{15}N_5O_3$: 265.1172, observed 265.1154.

2'-Deoxy-3'-O-acyladenosine [2]. 2'-Deoxy-5'-*tert*-butyldiphenylsilyladenosine [9] (100 mg, 0.2 mmol), acetic anhydride (28 mg, 0.27 mmol), and 4-dimethylaminopyridine (DMAP) (5 mg, 0.05 mmol) in dry pyridine were stirred at 25°C for 6 h. After removing pyridine under vacuum, the residue was dissolved in D.I. water, extracted in chloroform, washed with D.I. water, 10% HCl, saturated $NaHCO_3$, saturated NaCl, dried over Na_2SO_4 and flash chromatographed (96 mg, 90%).

The 3'-*O*-acyl derivative (100 mg, 0.19 mmol) following desilylation and flash chromatography afforded compoud [2] (44

mg, 80%). HRMS *m/e* calculated for $C_{12}H_{15}N_5O_4$: 293.1121, observed 293.1107.

2'-Deoxy-3'-O-allyladenosine [3]. 2'-Deoxy-5'-*tert*-butyldiphenylsilyladenosine [9] (200 mg, 0.4 mmol) in benzene (5 mL), allyl bromide (484 mg, 4.0 mmol, 10 equiv.), TBAH (40% solution, 390 μL), and 1 M NaOH (5 mL) were stirred at 25°C for 15 h. Following ethyl acetate extraction, the organic phase was washed with D.I. water, saturated NaCl, dried over Na_2SO_4, and flash chromatographed (157 mg, 74%).

The *O*-allyl derivative (198 mg, 0.37 mmol) following desilylation and flash chromatography gave compound [3] (106 mg, 98.5%). HRMS *m/e* calculated for $C_{13}H_{17}N_5O_3$: 291.1328, observed 291.1318.

2'-Deoxy-3'-O-tetrahydropyranyladenosine [4]. 2'-Deoxy-5'-*tert*-butyldiphenylsilyladenosine [9] (2.90 g, 5.92 mmol), dihydropyran (4.89 g, 59.2 mmol, 10 equiv.) and pyridinium nitrobenzenesulfonate (1.67 g, 5.92 mmol) were dissolved in methylene chloride (20 mL) and stirred at 40°C for 20 h. The reaction mixture was washed with D.I. water, saturated NaCl, dried over Na_2SO_4 and flash chromatographed to give a diastereomeric mixture of 2'-deoxy-3'-O-tetrahydropyranyl-5'-*tert*-butyldiphenylsilyladenosine (0.4 g, 12%).

The tetrahydropyran derivative formed above, after desilylation and flash chromatography yielded compound [4] (147 mg, 84%) as a mixture of diastereomers. HRMS *m/e* calculated for $C_{15}H_{21}N_5O_4$: 335.1589, observed 335.1581.

2'-Deoxy-3'-O-(2-aminobenzoyl)adenosine-5'-triphosphate [5]. This compound was prepared according to the procedure of Hiratsuka *et al.* (20) directly from the 2'-deoxyadenosine-5'-triphosphate sodium salt.

2'-Deoxy-3'-O-(4-nitrobenzoyl)adenosine [6]. 2'-Deoxy-5'-*tert*-butyldiphenylsilyladenosine [9] (100 mg, 0.2 mmol), 4-nitrobenzoyl chloride (89 mg, 0.48 mmol), DMAP (5 mg, 0.04 mmol) were dissolved in pyridine and stirred for 8 h. Following solvent removal, the residue was dissolved in chloroform and was washed D.I. water, saturated $NaHCO_3$, D.I. water, saturated NaCl, dried over Na_2SO_4, and flash chromatographed giving a colorless solid (73 mg, 57%).

The 4-nitrobenzoyl derivative (66 mg, 0.1 mmol) following desilylation and flash chromatography gave compound [6] (22 mg, 57%). HRMS *m/e* calculated for $C_{17}H_{16}N_6O_6$: 400.1128, observed 400.1140.

2'-Deoxy-3'-O-(2-nitrobenzyl)adenosine [7]. 2'-Deoxyadenosine (100 mg, 0.4 mmol) [dried by repeated coevaporation with pyridine] was dissolved in hot DMF and cooled to 0°C in an ice bath. To the above solution, NaH (26 mg, 0.52 mmol [50% in mineral oil] in DMF after washing with dry benzene was added and stirred for 45 min. 2-Nitrobenzyl bromide (95 mg, 0.44 mmol) in DMF was added, and the reaction stirred for 3 h. The reaction was quenched with cold D.I. water and stirred overnight. The solid obtained was filtered, dried, and recrystalized in ethanol (122 mg, 79%). HRMS *m/e* calculated for $C_{17}H_{18}N_6O_5$: 386.1335, No *m/e* was observed. Fast atom bombardment MS, nitrobenzyl alcohol (NBA) *m/e* 387.1 (M+1).

2'-Deoxy-3'-O-methylthymidine [8]. 2'-Deoxy-5'-*tert*-butyldiphenylsilylthymidine [9] (100 mg, 0.21 mmol) in benzene (5 mL), methyl iodide (43 mg, 0.3 mmol), TBAH (40% solution, 325 μL), and 1 M NaOH (5 mL) were vigorously stirred at 25°C for 6 h. The organic layer was extracted with ethyl acetate and washed with D.I. water, saturated NaCl, and dried over Na_2SO_4 (100 mg, 98%).

The above sample, following desilylation and purification by flash chromatography, gave compound [8] (36 mg, 88%). HRMS *m/e* calculated for $C_{11}H_{16}N_2O_5$: 256.1059, observed 256.1082.

Syntheses of nucleoside 5'-triphosphates

In general, the 3'-modified nucleoside (1.0 equiv.) was dissolved in trimethylphosphate under nitrogen atmosphere. Phosphorus oxychloride ($POCl_3$) (3.0 equiv.) was added, and the reaction stirred at −10°C for 4 h. The reaction was quenched with a solution of tributylammonium pyrophosphate (5.0 equiv.) in DMF and tributylamine (0.2 mL) (21). After stirring vigorously for 10 min., the reaction was quenched with 2 mL of 2 M TEAB, pH 7.5. The solution was concentrated, and the triphosphate derivative was isolated by linear gradient (0.01 M to 0.5 M TEAB) using a DEAE cellulose (HCO_3^- form) column.

Reverse-phase high performance liquid chromatography (RP-HPLC)

The RP-HPLC hardware system consisted of a Beckman controller and model 100A pumps, a Rheodyne model 7125 injector, an Applied Biosystems (ABI) model 759A absorbance detector, and a Spectra-Physics model SP4600 DataJet integrator. Gradient RP-HPLC was performed using an ABI aquapore OD-300 column (4.6 mm×250 mm) where 'Buffer A' is 100 mM triethylammonium acetate (TEAA), pH 7.0 and __Buffer B' is 100 mM TEAA, 70 % (v/v) acetonitrile. Compounds [1]−[6] and [8] were purified using the following gradient conditions: 0% B, 5 min.; 0% B − 40% B, 60 min.; 40% B − 100% B, 18 min.; 100% B, 5 min. at a flow rate of 0.5 mL per min. Compound [7] was purified using the following gradient conditions: 30% B, 5 min.; 30% B − 70% B, 60 min.; 70% B − 100% B, 15 min.; 100% B, 5 min. The gradient conditions used to analyze individual nucleotides were: 0% B, 5 min.; 0% B − 40% B, 30 min.; 40% B − 100% B, 18 min.; 100% B, 5 min.

Polymerases

Avian Myeloblastosis Virus (AMV) and Moloney Murine Leukemia Virus (M-MuLV) reverse transcriptases, Klenow fragment of DNA polymerase I, and T4 polynucleotide kinase were purchased from Pharmacia. *Bst* DNA polymerase was purchased from Bio-Rad Laboratories. AmpliTaq® DNA polymerase and r*Tth* DNA polymerase were purchased from Perkin Elmer. Sequenase® was purchased from United States Biochemical. $Vent_R$® (exo$^-$) DNA polymerase was kindly provided by New England Biolabs. *Pfu* (exo$^-$) DNA polymerase was purchased from Stratagene.

DNA templates

M13mp19 DNA was obtained from a 250 mL culture by polyethylene glycol precipitation and purified by a QIAGEN-tip 100 column according to the manufacture's protocol. Universal primer (5'-TGTAAAACGACGGCCAGT), biotinylated and unbiotinylated oligonucleotide template (5'-TACGGAGGTGG-ACTGGCCGTCGTTTTACA) and biotinylated oligonucleotide template (5'-TACGGAGGTTTTTGGACTGGCCGTCGTTT-

ACA) were synthesized using an ABI model 380B DNA synthesizer and purified by trityl-on RP-HPLC. All nonradioactive nucleotides were purchased from Pharmacia, and [γ-^{32}P]ATP was purchased from Amersham.

Polymerase incorporation assays

Two different template assays were used to test for 3'-modified nucleotide incorporation. In the first, designated the 'M13mp19-template assay', [^{32}P]-labeled universal primer was annealed to single-stranded M13mp19 DNA (0.1 pmol to 0.45 μg respectively, per 5 μL) in the specific enzyme buffer by heating to 80°C for 5 min. and cooling slowly to 25°C. The subsequent enzymatic extension of the primer–template complex was performed under conditions that are analogous to Sanger sequencing, where the natural nucleotides were mixed with either a dideoxynucleotide or 3'-modified nucleotide terminator to generate a sequencing ladder. For the second assay, designated the 'Oligo-template assay', [^{32}P]-labeled universal primer was annealed to an oligonucleotide template (0.05 pmol to 0.1 pmol respectively, per 5 μL) in the same fashion. Subsequent extensions were performed in the absence of the natural nucleotide when either a dideoxynucleotide or 3'-modified nucleotide was tested.

For each reaction, 5 μL aliquots of the annealed primer-template samples were dispensed into separate tubes containing 5 μL mixtures of each enzyme and nucleotides in their specific buffers. The final buffer conditions, concentrations of nucleotides, enzymatic units, and incubation temperatures are given in Table 1. The reactions were incubated for 10 min. and then stopped by the addition of 5 μL of stop solution containing 98% D.I. formamide, 10 mM EDTA, pH 8.0, 0.025% bromophenol blue, and 0.025% xylene cyanol. The samples were heated to 85°C for 3 min., chilled on ice, and either 4 μL (M13mp19-template Assay) or 3 μL (Oligo-template assay) were loaded on a 10% or 20% polyacrylamide gel, respectively. Following electrophoresis, the gel was fixed in an aqueous 10% acetic acid, 10 % methanolic solution (v/v), dried, and autoradiographed on HyperfilmTM-MP (Amersham).

Biotinylated Oligo-template assay

The conditions of this assay are similar to the Oligo-template assay except prior to primer annealing, 2.0 pmol of biotinylated template was captured on 10 mL streptavidin coated magnetic beads (Dynal Dynabeads® M-280) in 1 M NaCl for 15 min. After washing the bound template in the specific enzyme buffer, 0.1 pmol of [^{32}P]-labeled universal primer was annealed to an oligonucleotide template and extension reactions were performed as described in the Oligo-template assay.

RESULTS

Syntheses and purification

Compounds [**1**] through [**8**] were synthesized, desilylated, phosphorylated and purified as described above. In general, the yields of 3'-protection reactions ranged from 57% to 98% with the exception of 3'-*O*-THP-dATP [**4**]. This low yield (12%) is believed to be due to the acidic properties of the silica gel hydrolyzing the linkage to the THP group. For the desilylation reactions, the yields ranged from 70% to 98%, except for 3'-*O*-

Table 1. Specific enzymatic conditions for both the M13mp19-template and Oligo-template assays

Enzymatic Conditions	AMV-RT	M-MuLV-RT	Klenow fragment	Sequenase®	*Bst* DNA polymerase	AmpliTaq® DNA polymerase	*Pfu*(exo⁻) DNA polymerase	r*Tth* DNA polymerase	Vent(exo⁻)® DNA polymerase
Buffers									
[Tris-HCl] [$MgCl_2$]	50 mM, pH 8.3 8 mM 30 mM KCl	50 mM, pH 8.3 8 mM 10 mM DTT	10 mM, pH 8.5 5 mM	40 mM, pH 7.5 5 mM 50 mM NaCl	10 mM, pH 8.5 10 mM	10 mM, pH 8.5 10 mM 50 mM NaCl	20 mM, pH 8.75 2 mM $MgSO_4$ 10 mM KCl 10 mM $(NH_4)_2SO_4$ 0.1% Triton X-100 0.1 mg/mL BSA	50 mM, pH 8.3 8 mM 10 mM DTT	20 mM, pH 8.8 2 mM $MgSO_4$ 10 mM KCl 10 mM $(NH_4)_2SO_4$ 0.1% Triton X-100
Incubation Temperature (°C)	37	37	37	37	65	68	75	74	72
Units	1.3	1.9	1.0	1.3	0.1	0.3	0.1	0.3	0.1
M13mp19-Templ.									
[dATP] (μM)	40	N/D	N/D	10	1	2.5	N/D	N/D	N/D
[dCTP] (μM)	40	N/D	N/D	10	1	2.5	N/D	N/D	N/D
[dGTP] (μM)	40	N/D	N/D	10	1	2.5	N/D	N/D	N/D
[dTTP] (μM)	40	N/D	N/D	20	1	2.5	N/D	N/D	N/D
[ddATP] (μM)	2	N/D	N/D	0.25	10	150	N/D	N/D	N/D
[ddCTP] (μM)	2	N/D	N/D	0.25	10	75	N/D	N/D	N/D
[ddGTP] (μM)	2	N/D	N/D	0.25	10	15	N/D	N/D	N/D
[ddTTP] (μM)	2	N/D	N/D	0.5	10	150	N/D	N/D	N/D
Oligo-Templ.									
[dATP] (μM)	0.5	2.5	0.5	2.5	0.025	0.25	2.5	0.5	0.5
[dCTP] (μM)	0.5	2.5	0.5	2.5	0.025	0.25	2.5	2.5	0.5
[dTTP] (μM)	2.5	2.5	0.5	2.5	0.025	0.25	2.5	0.5	0.5
[ddATP] (μM)	2.5	250	0.05	2.5	0.25	2.5	I/T	125	125
[ddGTP] (μM)	0.5	250	0.05	2.5	0.25	0.25	I/T	25	25
[ddTTP] (μM)	5	5	0.5	2.5	2.5	25	I/T	100	100

N/D means the assay conditions were not determined and I/T means ddNTP termination was incomplete.

(4-nitrobenzoyl)-dATP [6]. The yield was reduced in that case to 57% due probably to a rearrangement of the 4-nitrobenzoyl group from the 3'-position to the 5'-position (data not shown). The phosphorylation yields of compounds [1]–[4] and [6]–[8] ranged from 25% to 40%. Thymidine analogs including 3'-*O*-methyl-dTTP [8], however, had to be handled more cautiously since they were more rapidly degraded by tributylammonium pyrophosphate than were the adenosine analogs.

The 3'-modified-dNTPs were further purified by RP-HPLC to ≥ 99% prior to the polymerase assay. Each nucleotide synthesis initially contained several major peaks that were individually tested in the Oligo-template assay to determine the active species. In general, the adenosine analogs contained both the natural dATP and 3'-modified-dATP.

Termination assays

A series of polymerases were chosen to test the candidate 3'-modified terminators, based on their broad template specificities and their commercial availability. The conditions for screening compounds [1]–[8] using each enzyme were first defined by a series of control polymerization experiments. For the M13mp19-template assay, a range of dNTP and ddNTP concentrations was identified that gave a clear sequencing ladder. Each test gel subsequently contained constant dNTP/ddNTP ratios in control lanes for three bases, while the concentrations of the test compound and its corresponding ddNTP were varied.

The Oligo-template assay was also standardized before testing each 3'-modified-dNTP for termination. The synthetic template contained all four bases to allow for the incorporation of the remaining natural nucleotides, so that other aspects of the enzyme performance could be identified. We found that all the polymerases misincorporated other dNTPs in the absence of the complement dNTP, and this nucleotide readthrough was concentration dependent. Thus, minimum dNTP concentrations that gave efficient incorporation, but no apparent misincorporation were first defined in this assay. These dNTP concentrations were then used to determine the minimum ddNTP concentration that yielded complete termination. *Pfu* (exo$^-$) DNA polymerase was excluded from the Oligo-template assay since a ddNTP concentration that yielded complete termination for this enzyme could not be identified. In each of the Oligo-template gels, the reactions contained all the required nucleotides except the natural nucleotide corresponding to the analog tested. The samples routinely used in this assay were a blank control (absence of corresponding dNTP or ddNTP), a titration of the corresponding ddNTP, a readthrough control (presence of corresponding dNTP), and a titration of the corresponding 3'-modified-dNTP.

Terminator screen

Table 2 summarizes the data from the enzymatic screen of compounds [1]–[8]. Three main classes of activity were defined: termination, inhibition, and inactive. Termination was apparent when the reaction containing the test compound mimicked the migration pattern of the ddNTP control. Inhibition was revealed when the presence of the test compound prevented the polymerase from incorporating the natural nucleotides. No activity was recorded when the 3'-modified-dNTPs mimicked either the blank or readthrough controls. In addition, a fourth effect related to an alteration in enzymatic fidelity is discussed below. Compounds [1], [8], and [7] showed specific termination and were further evaluated with respect to their concentration dependent effects.

3'-*O*-methyl-dATP [1] incorporation

The M13mp19-template assay in Figure 3A shows the incorporation of 3'-*O*-methyl-dATP [1] by AMV-RT. The termination of DNA synthesis by 3'-*O*-methyl-dATP [1] mimics the ddATP controls in a concentration dependent manner, although each band appears to migrate slightly slower. From the comparison of termination band intensities, it can be estimated that 3'-*O*-methyl-dATP [1] is approximately 200 to 250-fold less efficiently incorporated by AMV-RT than ddATP (compare lanes 6 and 8: 5 mM and 1 mM, respectively). In Figure 3B, the Oligo-template gel also shows the incorporation of 3'-*O*-methyl-dATP [1] by AMV-RT. In addition to termination, some readthrough was also observed due to the presence of contaminating dATP. All RP-HPLC purified 3'-modified-dATPs (compounds [1]–[7]) showed approximately 1% dATP contamination, and these trace levels could not be removed by subsequent RP-HPLC.

3'-*O*-Methyl-dATP [1] was also incorporated by M-MuLV-RT and inhibited DNA syntheses by r*Tth* and Vent$_{R}$® (exo$^-$)

Table 2. Activity matrix of RP-HPLC purified 3'-protecting dNTPs challenged against commercially available polymerases

3'-modified-dATP (except compound [8])	AMV-RT	M-MuLV-RT	Klenow fragment	Sequenase®	*Bst* DNA polymerase	AmpliTaq® DNA polymerase	Vent$_{R}$(exo$^-$)® DNA polymerase	r*Tth* DNA polymerase
[1] *O*-methyl	Termination	Termination*	-	-	-	-	Inhibition	Inhibition*
[2] *O*-acyl	-	-	-	-	-	-	Inhibition	-
[3] *O*-allyl	-	-	-	-	-	-	Termination*	-
[4] *O*-tetrahydropyran	-	-	-	-	-	-	-	-
[5] *O*-(4-nitrobenzoyl)	-	-	-	-	-	-	-	-
[6] *O*-(2-aminobenzoyl)	-	-	-	-	-	-	-	-
[7] *O*-(2-nitrobenzyl)	-	-	-	Inhibition	Termination	Termination*	Termination*	-
[8] 3'-*O*-methyl-dTTP	-	Inhibition	-	Inhibition	Termination	Termination	Termination	Termination

All compounds were assayed at a final concentration of 250 μM according to the conditions specified in Table 1. '–' means no activity was detected, 'Termination' means that the termination bands mimic ddNTP termination bands, and 'Inhibition' means the rate of DNA synthesis is reduced in a nonspecific manner. '*' means the activity was incomplete at a final concentration of 250 μM .

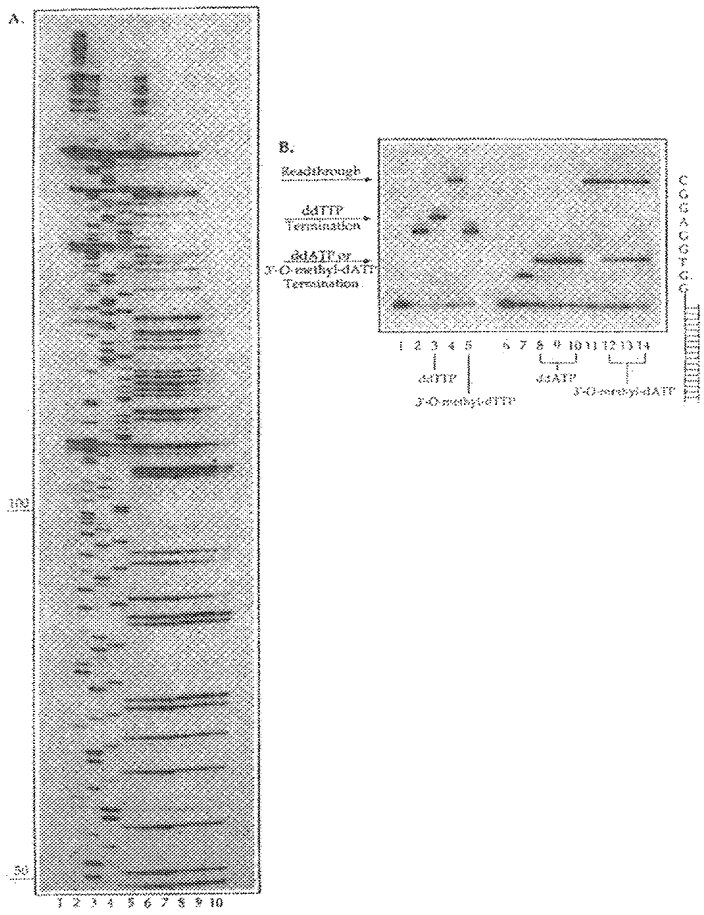


Figure 3. Incorporation of 3′-*O*-methyl-dATP by AMV-RT. (A) M13mp19-template assay: In addition to the conditions specified in Table 1, lane 1 contained no ddNTPs, and lanes 2–5 contained ddTTP, ddGTP, ddCTP, and ddATP, respectively. Lanes 6 and 7 contained 5 μM and 10 μM of ddATP, respectively. Lanes 8–10 contained 1 mM, 5 mM, or 10 mM of 3′-*O*-methyl-dATP, respectively. (B) Conditions for the AMV-RT Oligo-template assay were used where lanes 1 and 6 contained no dNTPs or ddNTPs. Lanes 2–5 contained dATP, dCTP and ddGTP. In addition, lane 3 contained ddTTP, lane 4 contained dTTP, and lane 5 contained 500 μM of 3′-*O*-methyl-dTTP. Lanes 7–14 contained dCTP, dTTP and ddGTP. In addition, lane 8–10 contained 0.1 μM, 0.5 μM, and 2.5 μM of ddATP; lane 11 contained dATP; and lanes 12–14 contained 10 μM, 50 μM, and 250 μM of 3′-*O*-methyl-dATP.

DNA polymerases. In the M-MuLV-RT Oligo-template assay, we observed a batch-to-batch difference in assaying both 3′-*O*-methyl-dATP [**1**] and 3′-*O*-acyl-dATP [**2**], (data not shown). While, the ddNTP controls gave similar results in both M-MuLV-RT batch assays, 3′-*O*-methyl-dATP [**1**] showed minor termination, and both 3′-*O*-dATP analogs ([**1**] and [**2**]) showed partial inhibition of DNA synthesis in M-MuLV-RT batch 1. In contrast, 3′-*O*-methyl-dATP [**1**] showed significant termination of DNA synthesis, and both 3′-*O*-dATP analogs ([**1**] and [**2**]) showed significant readthrough in M-MuLV-RT batch 2. This observation illustrates the importance of assaying candidate compounds with multiple enzyme batches.

3′-*O*-methyl-dTTP [8] incorporation

Unlike 3′-*O*-methyl-dATP [**1**], 3′-*O*-methyl-dTTP [**8**] was not incorporated by AMV-RT, (Figure 3B). A similar result was obtained by using the AMV-RT M13mp19-template assay (data not shown). However, 3′-*O*-methyl-dTTP [**8**] was efficiently incorporated by AmpliTaq® DNA polymerase in the M13mp19-template assay, and the termination pattern mimicked the ddTTP DNA ladders in a concentration dependent manner, (Figure 4A). Additional bands, however, were observed in the 3′-*O*-methyl-dTTP [**8**] ladders that were generated by both AmpliTaq® DNA polymerase (see arrows) and *Bst* DNA polymerase (data not shown) assays. The position of these additional bands corresponded to ddATP termination bands suggesting misincorporation of the thymidine analog in place of the deoxyadenosine analog. The efficiency of incorporation relative to ddTTP was, therefore, not determined in the M13mp19-template assays because of the additional bands.

Figure 4B shows the results of challenging 3′-*O*-methyl-dTTP [**8**] against four thermostable DNA polymerases in the Oligo-template assay. 3′-*O*-Methyl-dTTP [**8**] terminates all four polymerases in a concentration dependent manner. From the comparison of band intensities of 3′-*O*-methyl-dTTP [**8**] to ddTTP termination products, it can be estimated that 3′-*O*-methyl-dTTP [**8**] is approximately 50-fold less efficiently incorporated by *Bst* DNA polymerase than ddTTP, 20-fold less efficiently incorporated by AmpliTaq® DNA polymerase than ddTTP, 2-fold less efficiently incorporated by r*Tth* DNA polymerase than ddTTP, and 10-fold less efficiently incorporated by $Vent_R$® (exo⁻) DNA polymerase than ddTTP.

Further investigation of a different nucleotide composition in the Oligo-template assay revealed altered base specificity of 3′-*O*-methyl-dTTP [**8**]. In the absence of dATP, 3′-*O*-methyl-dTTP [**8**] was incorporated by both *Bst* and AmpliTaq® DNA polymerases, and the termination bands mimicked the ddATP controls, (Figure 4C). It is noteworthy that in addition to 3′-*O*-methyl-dTTP [**8**] incorporation in Figure 4C, significant levels of readthrough were observed. Since the Oligo-template assay was performed in the absence of a deoxyadenosine analog, the readthrough must reflect the misincorporation of dNTPs in the noncomplement base position. This result suggests that 3′-*O*-methyl-dTTP [**8**] alters the base specific properties of DNA polymerases, not only in its incorporation, but in the incorporation of other natural nucleotides present in the reaction. This result also highlights the importance of the precise conditions of incorporation assays.

3′-*O*-(2-nitrobenzyl)-dATP [7] incorporation by *Bst* DNA polymerase

3′-*O*-(2-Nitrobenzyl)-dATP [**7**] terminated *Bst* DNA synthesis in a base specific manner. In the *Bst* M13mp19-template assay, however, the termination did not correspond to any of the ddNTP controls, (Figure 5A). This result made it difficult to assign a specific mode of action to 3′-*O*-(2-nitrobenzyl)-dATP [**7**]. The assignment was resolved using the biotinylated Oligo-template assay that gave strong evidence for termination. As shown in Figure 5B, 3′-*O*-(2-nitrobenzyl)-dATP [**7**] terminates *Bst* DNA synthesis where the DNA fragment migrates slower than the ddATP control, (compare lanes 3 and 7). 3′-*O*-(2-Nitrobenzyl)-dATP [**7**] is a UV sensitive compound that undergoes an intramolecular rearrangement by irradiation to give dATP and nitrosobenzaldehyde (22). Following UV exposure, the DNA termination fragment migrated at the same rate as the ddATP control, (compare lanes 3 and 8). This observation suggests that the 2-nitrobenzyl group can significantly alter the mobility of DNA in the gel and may provide a possible reason for the

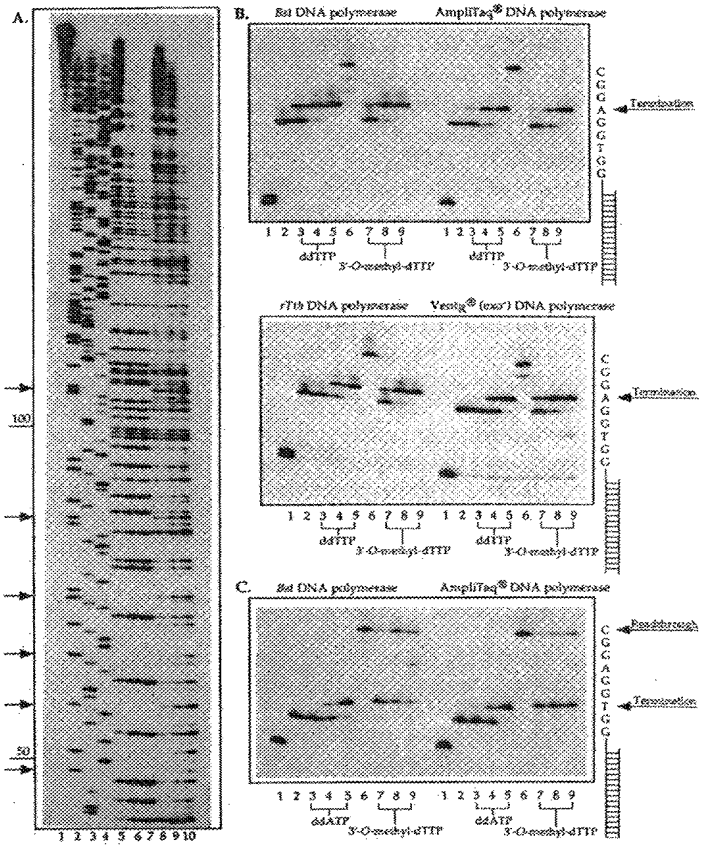


Figure 4. Incorporation of 3′-*O*-methyl-dTTP by *Bst*, AmpliTaq® , r*Tth* , and Vent$_R$® (exo⁻) DNA polymerases. (A) M13mp19-template assay (AmpliTaq®): Enzymatic conditions in Table 1 were used where lane 1 contained no ddNTPs, and lanes 2−5 contained ddATP, ddCTP, ddGTP, and ddTTP, respectively. Lanes 6 and 7 contained 450 μM and 750 of μM ddTTP, respectively. Lanes 8−10 contained 1 mM, 2.5 mM, or 5 mM of 3′-*O*-methyl-dTTP, respectively. Arrows correspond to termination bands that are observed in the ddATP control (compare lanes 2 with 8−10). (B) Conditions for the Oligo-template assay were used for *Bst*, AmpliTaq® , r*Tth* , and Vent$_R$® (exo⁻) DNA polymerases. Lane 1 contained no dNTPs or ddNTPs. Lanes 2−7 contained dATP, dCTP and ddGTP. In addition, lanes 3−5 contained (*Bst*) 0.1 μM, 0.5 μM and 2.5 μM ddTTP; (AmpliTaq®) 1.0 μM, 5.0 μM and 25 μM ddTTP; (r*Tth*) and (Vent$_R$® (exo⁻)) 4 μM, 20 μM, and 100 μM ddTTP, respectively; lane 6 contained dTTP; and lanes 7−9 contained (*Bst*) 4 μM, 20 μM, and 100 μM of 3′-*O*-methyl-dTTP; (AmpliTaq®), (r*Tth*), and (Vent$_R$® (exo⁻)) 20 μM, 100 μM, and 500 μM of 3′-*O*-methyl-dTTP, respectively. (C) Oligo-template assay (*Bst*, and AmpliTaq® DNA polymerases): Lane 1 contained no dNTPs or ddNTPs. Lanes 2−9 contained dCTP, dTTP and ddGTP. In addition, lanes 3−5 contained (*Bst*) 0.01 μM, 0.05 μM, and 0.25 μM of ddATP and (AmpliTaq®) 0.1 μM, 0.5 μM, and 2.5 μM of ddATP, respectively; lane 6 contained dATP; and lanes 7−9 contained (*Bst*) and (AmpliTaq®) 20 μM, 100 μM, and 500 μM of 3′-*O*-methyl-dTTP, respectively.

complicated pattern observed in the M13mp19-template assay. However, after exposing the DNA fragments to UV, a complex pattern was still observed in the M13mp19-template assay (data not shown).

The biotinylated Oligo-template assay was used to examine the reinitiation of DNA synthesis after 3′-*O*-(2-nitrobenzyl)-dATP [7] incorporation and UV deprotection. Following 3′-*O*-(2-nitrobenzyl)-dATP [7] termination, the solid-phase bound template was thoroughly washed and exposed to UV for 10 min. All the nucleotides required for the reinitiation of *Bst* DNA synthesis were then added in the absence of a deoxyadenosine analog. DNA synthesis clearly continues subsequent to 3′-*O*-(2-nitrobenzyl)-dATP [7] incorporation and UV deprotection, (Figure 5C).

DISCUSSION

In this study, the behavior of a series of novel nucleotide analogs was evaluated. Two different template tests were developed using various nucleotide compositions to assay the 3′-modified-dNTPs for incorporation by a variety of polymerases. Three compounds, 3′-*O*-methyl-dATP [**1**], 3′-*O*-methyl-dTTP [**8**], and 3′-*O*-(2-nitrobenzyl)-dATP [7], proved to be interesting DNA synthesis terminators. In Sanger sequencing, the 3′-*O*-methyl analogs generated clean terminating ladders, thus demonstrating their possible role as alternative terminators to ddNTPs. One of the labile nucleotide analogs, 3′-*O*-(2-nitrobenzyl)-dATP [7] was incorporated by *Bst* DNA polymerase. Subsequent removal of the 2-nitrobenzyl group and the reinitiation of DNA synthesis were used to demonstrate one complete cycle in BASS DNA sequencing.

Some notable correlations were observed between the structure of the 3′-modified-dNTPs and the specificity of enzymatic incorporation. First, the protecting groups containing ether linkages at the 3′-position, compounds [**1**], [**3**], [**4**], [**7**], and [**8**] were incorporated by some of the polymerases. With the exception of inhibition of Vent$_R$® (exo⁻) DNA synthesis by 3′-*O*-acyl-dATP [**2**], those substrates containing ester linkages at the 3′-position, i.e., compounds [**2**], [**5**], and [**6**], were not accepted for any of the polymerases examined here. A comparison of compounds [**5**] and [**6**] with [**7**] illustrates this point. Both 3′-*O*-benzoyl-dATPs [**5**] and [**6**] were determined to be inactive in all the polymerase assays, whereas the 3′-*O*-(2-nitrobenzyl)-dATP [7] was incorporated by *Bst*, AmpliTaq® , and Vent$_R$® exo⁻) DNA polymerases and inhibited Sequenase® DNA synthesis. Second, no consistent activity pattern between the polymerases and the analogs could be discerned. However, the incorporation of 3′-*O*-methyl-dATP [**1**] and of 3′-*O*-methyl-dTTP [**8**] was favored by reverse transcriptases and by thermostable DNA polymerases, respectively.

The data presented here are consistent with previous reports (1,3−9,11). A number of other 3′-modified-dNTPs have been synthesized and tested by various incorporation assays. Some examples from this nucleotide set include the 3′-*O*-methyl (5−7), 3′-amino (3), 3′-azido (1,6), 3′-chloro (8), 3′-fluoro (4), and 3′-mercapto (11) nucleoside-5′-triphosphates. The compounds in this study, with the exception of 3′-*O*-methyl-dATP [**1**], represent novel 3′-modified-dNTPs. The activity of 3′-*O*-methyl-dATP [**1**] has previously been shown to be incorporated by AMV-RT (4,6) and was shown to be inactive using Klenow fragment (5). We have expanded the list of polymerases to include those enzymes developed for DNA sequencing purposes (23−26).

The assay conditions are important determinants for evaluating the combined behavior of enzymes and 3′-modified-dNTPs. Here, two different template assays were used to enhance the chances of detecting enzymatic activity. Typically in the past, either a 'short' or a 'long' template assay has been used to monitor nucleotide incorporation (23−27) and misincorporation (28−30), and it is not clear that either template assay mimics the natural environment of DNA replication more than the other (31). In addition to patterns of termination, the sensitivity of these assays

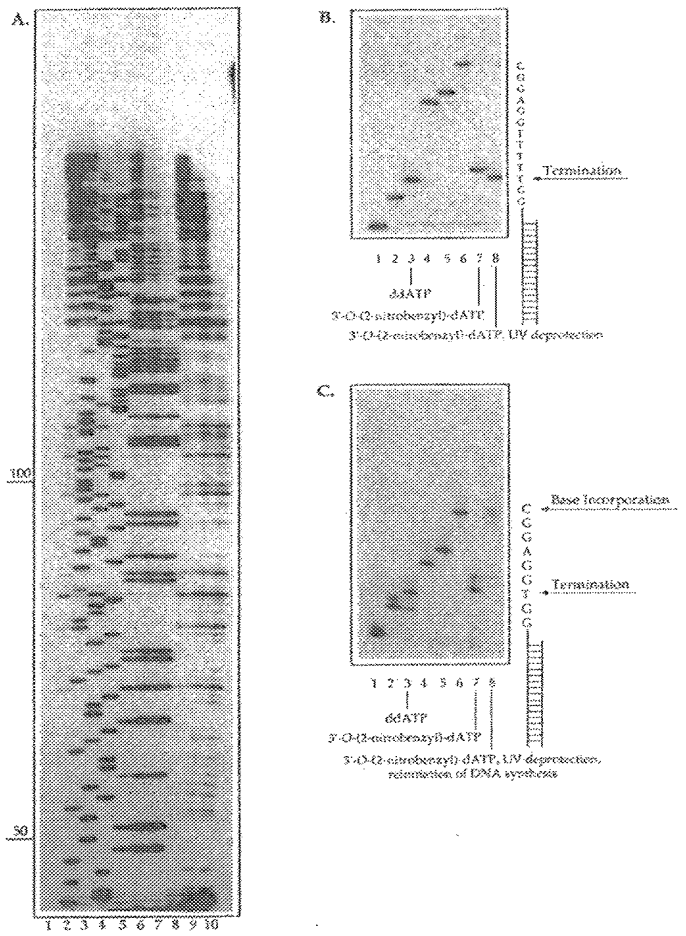


Figure 5. Incorporation of 3′-*O*-(2-nitrobenzyl)-dATP by *Bst* DNA polymerase. **(A)** M13mp19-template: In addition to the conditions specified in Table 1, lane 1 contained no ddNTPs, and lanes 2−5 contained ddTTP, ddGTP, ddCTP, and, ddATP, respectively. Lanes 6 and 7 contained 25 μM and 50 μM of ddATP, respectively. Lanes 8−10 contained 0.25 mM, 0.5 mM, or 1 mM of 3′-*O*-(2-nitrobenzyl)-dATP, respectively. **(B)** *Bst* Biotinylated Oligo-template assay conditions specified in Table 1 were used. Lane 1 contained no dNTPs or ddNTPs. Lanes 2−8 contained 0.05 μM of dCTP and 2.5 μM ddGTP. In addition, lane 3 contained 0.05 μM of dTTP and 5 μM of ddATP, lane 4 contained 0.05 μM of dATP, lanes 5 contained 0.05 μM of dATP and 50 μM of ddTTP, lane 6 contained 0.05 μM of dATP and dTTP. Lanes 7 and 8 contained 0.05 μM of dTTP and 250 μM of 3′-*O*-(2-nitrobenzyl)-dATP. Lane 8 was exposed to UV for 10 minutes. **(C)** The reactions conditions for lanes 1−8 are equivalent to (B) except for lane 8. Following UV deprotection, the reinitiation reaction were performed under the conditions specified in Table 1 and 0.05 μM of dCTP and dTTP and 2.5 μM ddGTP.

revealed isolated cases of DNA synthesis inhibition. The mode of inhibition is potentially of biological significance and is worthy of further investigation.

The short and long template assays employed in this study were tested either in the absence or in the presence of its competing dNTP, respectively. Each template assay, therefore, generated additional information not provided by the other. The Oligo-template assay was particularly useful for examining the incorporation of 3′-modified-dNTPs at reduced concentrations and revealed the presence of natural nucleotide contamination in the RP-HPLC purified 3′-modified-dNTP solutions. The M13mp19-template assay revealed the reduction of enzymatic fidelity in the incorporation of 3′-*O*-methyl-dTTP [**8**] by the thermostable DNA polymerases which was subsequently confirmed in the Oligo-template assay. These observations were possible because the assays were performed by varying the nucleotide composition and were not dependent of the length of the template.

The 3′-modified-dNTPs synthesized here were tested in a Sanger-type DNA sequencing scheme as alternative chain terminators. The 3′-*O*-methyl analogs best fit this class of dideoxy sequencing terminators because they generated clean sequencing ladders. In comparison to the ddNTPs that terminate the syntheses of DNA polymerases, both 3′-*O*-methyl-dATP [**1**] and 3′-*O*-methyl-dTTP [**8**] were less efficiently incorporated by these enzymes. The incorporation efficiency of the 3′-*O*-methyl analogs relative to their ddNTP analogs was variable and dependent upon the enzyme examined. For instance, the incorporation efficiency 3′-*O*-methyl-dTTP [**8**] ranged from 50-fold less for *Bst* DNA polymerase to 2-fold less for r*Tth* DNA polymerase relative to ddTTP.

Several unexpected results were observed in the use of the 3′-*O*-methyl-dNTPs as terminators in DNA sequencing. First, the 3′-*O*-methyl-dTTP [**8**] reduced enzyme fidelity. Also, we have not identified a polymerase that will incorporate both 3′-*O*-methyl-dATP [**1**] and 3′-*O*-methyl-dTTP [**8**]. These problems could be resolved by increasing the concentration of dATP in the former case to titrate away artifact termination bands and by testing more polymerases in the latter case. This issue, however, could also be resolved by substituting 2′-deoxy-3′-*O*-methyl-uracil-5′-triphosphate (3′-*O*-methyl-dUTP) for 3′-*O*-methyl-dTTP [**8**]. Kutateladze *et al.* have shown the incorporation of 3′-*O*-methyl-dATP, 3′-*O*-methyl-dCTP, 3′-*O*-methyl-dGTP, and 3′-*O*-methyl-dUTP by AMV-RT in a base specific manner (7). The use of these 3′-*O*-methyl terminators, however, has not been exploited in DNA sequencing strategies. We have synthesized the 3′-*O*-methyl-dUTP, and we are currently evaluating its activity with AMV-RT in the M13mp19-template assay. The eventual utility of the 3′-*O*-methyl terminators in Sanger sequencing and whether these novel terminators can eliminate artifacts such as gel compression remains to be determined.

The base specific incorporation of a labile terminator, the efficient removal of its 3′-protecting group, and the reinitiation of DNA synthesis demonstrate a one-cycle example of BASS. 3′-*O*-(2-Nitrobenzyl)-dATP [**7**] is a UV sensitive compound that decomposes to dATP, and both the incorporation and deprotection characteristics of this nucleotide analog made it an ideal compound to test the feasibility of this system. Here, it has been shown that 3′-*O*-(2-nitrobenzyl)-dATP [**7**] terminates *Bst* DNA synthesis in a base specific manner, and that following incorporation into DNA, the 2-nitrobenzyl group can be efficiently removed by UV. This UV deprotection was demonstrated by a band shifting pattern of the terminated reaction that mimicked the ddATP control and by the reinitiation of DNA synthesis in the Oligo-template assay, (Figures 5B and 5C, respectively). The band shifting experiment was crucial in demonstrating base specificity in that it discriminates between true incorporation of the nucleotide analog and nonspecific termination followed by incorporation of the natural, contaminating dNTP that we and others (7) have observed in the synthesis of 3′-modified-dNTPs.

Identification of a labile terminator sets the stage in the development of BASS. Investigations into the efficiencies of each step are required to further develop this strategy into a working model. We are currently evaluating both different base analogs containing the 2-nitrobenzyl protecting group and 3'-*O*-fluorescently labeled dNTPs for enzymatic activity. Efforts into high-density solid support technologies and quantitative imaging systems are also being pursued to integrate BASS into a feasible alternative to Sanger sequencing.

ACKNOWLEDGEMENTS

The authors thank M.Ali Ansari-Lari and Björn Andersson for critical review of the manuscript. This work was supported in part as a pilot project in USPH 5 P30 HG00210 and by a grant to KB from the Texas Advanced Technology Program, 003604-021 and 010366-174. KB also acknowledges support from NIH and The Robert A.Welch Foundation and thanks the NIH for a Research Career Development Award and the Alfred P.Sloan Foundation for a fellowship.

REFERENCES

1. Mitsuya, H., Weinhold, K. J., Furman, P. A., St. Clair, M. H., Nusinoff-Lehrman, S., Gallo, R. C., Bolognesi, D., Barry, D. W., and Broder, S. (1985) *Proc. Natl. Acad. Sci.* 82, 7096.
2. Mitsuya, H., Yarchoan, R., and Broder, S. (1990) *Science*, 249, 1533–1544.
3. Chidgeavadze, Z. G., and Beabealashvilli, R. Sh. (1984) *Nucleic Acids Res.*, 12, 1671–1686.
4. Chidgeavadze, Z. G., Scamrov, A. V., Beabealashvilli, R. Sh., Kvasyuk, E. I., Zaitseva, G. V., Mikhailopulo, I. A., Kowollik, G., and Langen, P. (1985) *FEBS*, 183, 275–278.
5. Beabealashvilli, R. Sh., Scamrov, A. V., Kutateladze, T. V., Mazo, A. M., Krayevsky, A. A., and Kukhanova, M. K. (1986) *Biochim. Biophys. Acta*, 868, 136–144.
6. Chidgeavadze, Z. G., Beabealashvilli, R. Sh., Krayevsky, A. A., and Kukhanova, M. K. (1986) *Biochim. Biophys. Acta*, 868, 145–152.
7. Kutateladze, T. V., Kritzyn, A. M., Florentjev, V. L., Kavsan, V. M., Chidgeavadze, Z. G., and Beabealashvilli, R. Sh. (1986) *FEBS*, 207, 205–212.
8. Krayevsky, A. A., Kukhanova, M. K., Atrazhev, A. M., Dyatkina, N. B., Papchikhin, A. V., Chidgeavadze, Z. G., and Beabealashvilli, R. Sh. (1988) *Nucleosides and Nucleotides*, 7, 613–617.
9. Chinchaladze, D. Z., Prangishvili, D. A., Scamrov, A. V., Beabealashvilli, R. Sh., Dyatkina, N. B. and Krayevsky, A. A. (1989) *Biochim. Biophys. Acta*, 1008, 113–115.
10. Krayevsky, A. A., Tarussova, N. B., Kukhanova, M. K. Balzarini, J., DeClercq, E., Karamov, E. V., and Lukashov, V. V. (1991) *Nucleic Acids Res.*, Symposium Series No. 24, 13–16.
11. Yuzhakov, A. A., Chidgeavadze, Z. G., and Beabealashvilli, R. Sh. (1992) *FEBS*, 306, 185–188.
12. Semizaroz, D. G., Victorova, L. S., Krayevsky, A. A., and Kukhanova, M. K. (1993) *FEBS*, 327, 45–48.
13. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci.*, 74, 5463–5467.
14. Hunkapiller, T., Kaiser, R. J., Koop, B. F., and Hood, L. (1991) *Science*, 254, 59–67.
15. Tsien, R. Y., Ross, P., Fahnestock, M., and Johnston, A. J., PCT number WO 91/06678, *'DNA sequencing.'*, filed: October, 26, 1990, published: May 16, 1991.
16. Cheeseman, P. C., US patent number 5,302,509, *'Method for sequencing polynucleotides.'*, filed: February, 27, 1991, published: April 12, 1994.
17. Wilhelm, A., DE, patent number 41 41 178 A1, *' Methods for sequencing nucleic acids.'*, filed: December 13, 1991, published: June 17, 1993.
18. Rosenthal, A., Close, K., and Brenner, S., PCT number WO 93/21340, *'DNA sequencing method'*, filed: April 22, 1992, published: October 28, 1993.
19. Both ^{1}H NMR and ^{13}C NMR spectra were performed for all the nucleosides and are available upon request.
20. Hiratsuka, T. (1983) *Biochim. Biophys. Acta*, 742, 496–508.
21. Ludwig, J. (1981) *Acta Biochim. Biophys. Acad. Sci. Hung.*, 16, 131–133.
22. Pillai, V. N. R. (1980) *Synthesis*, 1–26.
23. Tabor, S. and Richardson, C. C. (1987) *Proc. Natl. Acad. Sci.*, 84, 4767–4771.
24. Innis, M. A., Myambo, K. B., Gelfand, D. H., and Brow, M. A. D. (1988) *Proc. Natl. Acad. Sci.*, 85, 9436–9440.
25. McClary, J. Ye, S. Y., Hong, G. F., and Witney, F. (1991) *DNA Sequence*, 1, 173–180.
26. Kong, H., Kucera, R. B., and Jack, W. E. (1993) *J. Biol. Chem.*, 268, 1965–1975.
27. Huang, P., Farquhar, D., and Plunkett, W. (1990) *J. Biol. Chem.* 265, 11914–11918.
28. Hillebrand, G. G., McCluskey, A. H., Abbott, K. A., Revich, G. G., and Beattie, K. L. (1984) *Nucleic Acids Res.*, 12, 3155–3171.
29. Skinner, J. A. and Eperon, I. C. (1986) *Nucleic Acids Res.*, 14, 6945–6964.
30. Yu, H. and Goodman, M. F. (1992) *J. Biol. Chem.* 267, 10888–10896.
31. Goody, R. S., Müller, B., and Restle, T. (1991) *FEBS*, 291, 1–5.

Published online 18 September 2007

Nucleic Acids Research, 2007, Vol. 35, No. 19 6339–6349
doi:10.1093/nar/gkm689

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,*]

[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

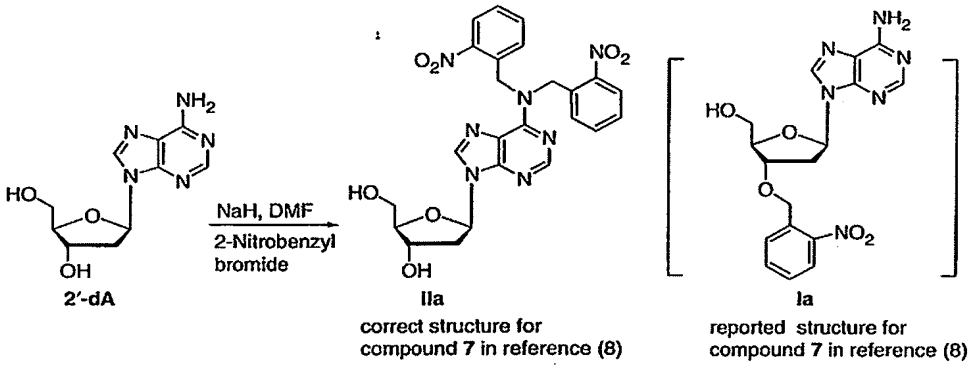
The Human Genome Project has facilitated the sequencing of many species, yet the current Sanger method is too expensive, labor intensive and time consuming to accomplish medical resequencing of human genomes *en masse*. Of the 'next-generation' technologies, cyclic reversible termination (CRT) is a promising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

INTRODUCTION

Next-generation technologies are being developed to advance sequencing to the $100 000, and eventually the $1000 genome. A number of strategies, albeit at different stages of development, have been proposed including pyrosequencing, sequencing-by-ligation (SBL), cyclic reversible termination (CRT), real-time sequencing and nanopore sequencing (1–6). CRT is a promising approach, which is comprised of a three-step process of incorporating modified nucleotides, fluorescence imaging and deprotecting after which the cycle begins again (5,6). CRT reactions can be performed in a high-density format using single-molecule arrays (3) or oligonucleotide arrays (5,7), eliminating the requirement for gel electrophoresis while significantly increasing sequence throughput. At the center of this technology is the reversible terminator (RT), whereby DNA polymerases exhibit specific and efficient incorporation of the modified nucleotide into the growing primer strand, with deprotection chemistries resulting in the efficient removal of the terminating group.

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

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



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

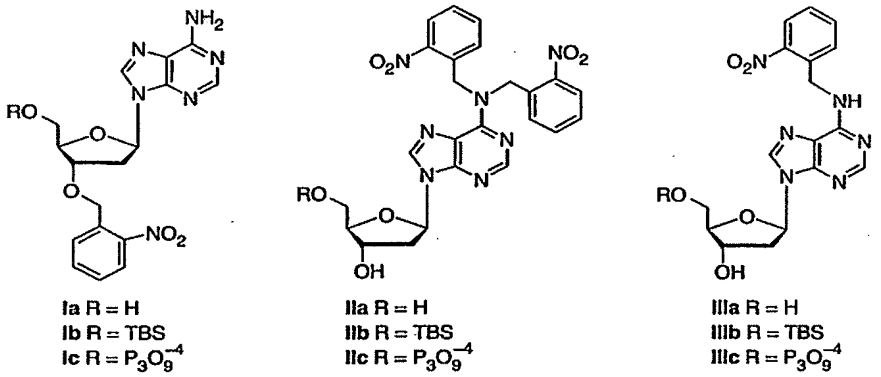


Figure 1. Structures of 3′-*O*-alkylated and N^6-alkylated 2′-deoxyadenosine analogs.

by DNA polymerase incorporation assays. This finding led to the synthesis and characterization of a third set of N^6-(2-nitrobenzyl)-2′-deoxyadenosine analogs **IIIa–IIIc** (Figure 1), of which compound **IIIb** was originally identified by us as an intermediate during the ultraviolet (UV) light-induced deprotection of compound **IIb** to its corresponding parent analog. Here, we report that the active triphosphate described by Metzker *et al.* (8) is actually the N^6-(2-nitrobenzyl)-dATP **IIIc**, representing a novel, 3′-unblocked terminator of DNA synthesis and an ideal candidate as a RT for the CRT approach.

MATERIALS AND METHODS

Reagents and materials

Chemical reagents and solvents were purchased from Alfa Aesar, Sigma-Aldrich, or EM Sciences. Oligonucleotides were purchased from Integrated DNA Technologies. 2′-Deoxyadenosine triphosphate (dATP), 2′,3′-dideoxyadenosine triphosphate (ddATP) and Q Sepharose Fast Flow anion-exchange resin were purchased from GE Healthcare Life Sciences. All DNA polymerases and apyrase were purchased from New England Biolabs, with the exception of Ampli*Taq*FS being purchased from Applied Biosystems (AB). Analytical silica gel 60 F_{254} 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*-butyldimethylsilyl (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 ($DT_{1/2}$) were determined from kinetic plots at which 50% of the compound was deprotected (i.e. loss of the 2-nitrobenzyl group). UV deprotection experiments were performed in triplicate for each compound.

Oligonucleotide template assay

As described for the polymerase end-point (PEP) assays, compounds **Ic**, **IIc** and **IIIc** 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′-TACGGAGCAGTTTTTACTGGC CGTCGTTTTACA, 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 $(NH_4)_2SO_4$; 10 mM KCl; 2 mM $MgSO_4$; 0.1% Triton X-100, New England BioLabs). We found that the addition of Triton X-100 stimulated the activity of many of the enzymes tested (data not shown), which is consistent with other reports (14,15). For all polymerases evaluated in this study, 5 nM BODIPY-FL labeled primer-1 was annealed with 40 nM of oligoTemplate-2 (5′-TACGGAGCAGTACTGGCC GTCGTTTTACA, interrogation base is underlined and bolded) in 1× 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 IC_{50} value for nucleotide titrations to 1.25 nM (i.e. [primer] = [primer plus incorporated nucleotide]). Polymerase reactions were incubated at their appropriate temperature (Supplementary Table 1) for 10 min, then cooled to 4°C and quenched with 10 μl of stop solution (98% deionized formamide; 10 mM Na_2EDTA, pH 8.0; 25 mg/ml Blue Dextran, MW 2 000 000). Stopped reactions were heated to 90°C for 30 s and then placed on ice. The extension products were analyzed on a 10% Long Ranger (Cambrex) polyacrylamide gel using an AB model 377 DNA sequencer, the quantitative data of which are displayed as a linear–log plot of product formation versus compound concentration. PEP assays were performed in triplicate, for each DNA polymerase/nucleotide analog combination, to calculate the average $IC_{50} \pm 1$ SD.

Compound **IIIc** and ddATP were then titrated using the PEP assay with the eight DNA polymerases (unit activities defined in Supplementary Table 1) in the concentration range of either 0.1 nM to 100 nM, 1 nM to 1 μM, 10 nM to 10 μM or 100 nM to 100 μM. Average $IC_{50} \pm 1$ SD values were calculated for compound **IIIc** and ddATP using oligoTemplate-2 as described above.

PEP termination assays for compound IIIc

OligoTemplate-2 was substituted with 40 nM of oligoTemplate-3 (5′-CCGTTTTTTTTTTACTGGCCGT CGTTTTACAGCCGCCGCCGCCGAACCGAGAC-Biotin, interrogation bases are underlined and bolded), annealed to 5 nM BODIPY-FL labeled primer-1 and assayed as described above. The PEP titrations were then performed at 1×, 5× and 25× IC_{50} values for compound **IIIc**, using the eight DNA polymerases, and reported as % primer product, % first-base product and % second-base product.

PEP mismatch assays for compound IIIc

OligoTemplate-2 was substituted with either 40 nM of oligoTemplate-4 (5′-TACGGAGCTGAACTGGCCGTC GTTTTACA), 40 nM of oligoTemplate-5 (5′-TACG GAGCAGCACTGGCCGTCGTTTTACA) or 40 nM of oligoTemplate-6 (5′-TACGGAGCAGGACTGGCCGT CGTTTTACA, interrogation bases are underlined and bolded), annealed to 5 nM BODIPY-FL labeled primer-1, and assayed as described above. Compound **IIIc** and dATP were assayed in the concentration range of 100 nM to 100 μM, and ddATP in the range of 500 nM to 500 μM. Average $IC_{50} \pm 1$ SD values were calculated for dATP and compound **IIIc** using oligoTemplate-4, -5 and -6, as described above.

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

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

Five-base sequencing experiment using compound IIIc as reversible terminator

An 80 nM solution of oligoTemplate-3 in 1 M NaCl and 1× 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-2 (5′-GGCGGCGGCG GCTGTAAAACGACGGCCAGT) 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 **IIIc** in 1× ThermoPol buffer (reaction volume: 20 μl) at 65°C for 6 min, then placed on ice. Compound **IIIc**-incorporated beads were washed four times with 50 μl W10 washing solution (10 mM Tris–HCl, pH 8.0; 10 mM Na_2EDTA; 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 Na_2EDTA, pH 8.0; 16.6 mg/ml Blue Dextran, MW 2 000 000), 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 1× apyrase buffer (100 U/l apyrase in 1× ThermoPol), three times with 50 µl W10 washing solution, twice with 50 µl 1× ThermoPol buffer and then once with 20 µl 1× 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 Ranger polyacrylamide gel using an AB model 377 DNA sequencer.

Ab initio calculations

Adenine–thymine and N^6-(2-nitrobenzyl)-adenine–thymine base pairs were created with the nucleobases planar to each other, using Watson–Crick hydrogen-bond distances of 2.82 Å ($N^1 \ldots H–N^3$) and 2.91 Å ($N^6–H \ldots O^4$) (16), utilizing ChemDraw and Chem3D Ultra 9.0 software packages (CambridgeSoft). A series of 3D, N^6-(2-nitrobenzyl)-adenine–thymine base pairs were then created by rotating the 2-nitrobenzyl group, pivoted on the N^6-position of adenine, 360° at 5° intervals, with the nitro group at 0°, 30°, 45°, 60° and 90° intervals relative to the plane of the phenyl group. Chem3D structures were then further optimized using the GAMESS program (17). Restricted Hartree–Fock (RHF) energy calculations were initially determined using the STO-3G atomic orbital/shell data set, and each nitro group conformation (i.e. 0°, 30°, 45°, 60° or 90° intervals) was plotted as RHF energy versus degrees of rotation of the 2-nitrobenzyl group. Our calculations revealed that a 45° rotation of the nitro group, relative to the plane of the phenyl group, gave the lowest RHF energy calculations. The N^6-(2-nitrobenzyl)-adenine–thymine base-pair conformations were further characterized using the more stringent 6-31G* atomic orbital/shell calculations and plotted as RHF energy versus degrees of rotation of the 2-nitrobenzyl group (data not shown). To evaluate the efficacy of our existing software tools, we modeled the natural adenine–thymine base pair and compared our results to those reported by Šponer *et al.* (18,19). Here, we used the 6-31G** atomic orbital/shell set, with the *X*, *Y*, *Z* coordinates described in Ref. (19) to perform the calculations. Showing good agreement with these reports (Supplementary Table 3), this served as an independent validation of our method.

RESULTS

Nucleosides and nucleotides synthesis

To synthesize the 3′-*O*-(2-nitrobenzyl)-2′-deoxyadenosine analog **Ia**, the 5′-hydroxyl and 6-amino groups of 2′-deoxyadenosine were protected with *tert*-butyldimethylsilyl (TBS) and *tert*-butyloxycarbonyl (Boc) groups, respectively, to yield intermediate **1**, according to Furrer and Giese (20). Transformation of this precursor into intermediate **2** occurred via deprotection and selective reprotection procedures, as outlined in Scheme 2. Alkylation of intermediate **2** with 2-nitrobenzyl bromide, using phase transfer catalysis under basic conditions, gave the desired 3′-*O* alkylated intermediate **3** in 91% yield. The *bis*-Boc groups were removed by heating on silica gel under vacuum (20) to give compound **Ib** in 91% yield, followed by the removal of the 5′-*O*-TBS group with tetra(*n*-butyl)ammonium fluoride to give compound **Ia** in 23% yield. Synthesis of the triphosphate was performed using the 'one-pot' procedure described by Ludwig (21), followed by purification using Q Sepharose FF anion-exchange chromatography to yield compound **Ic** as an ammonium salt.

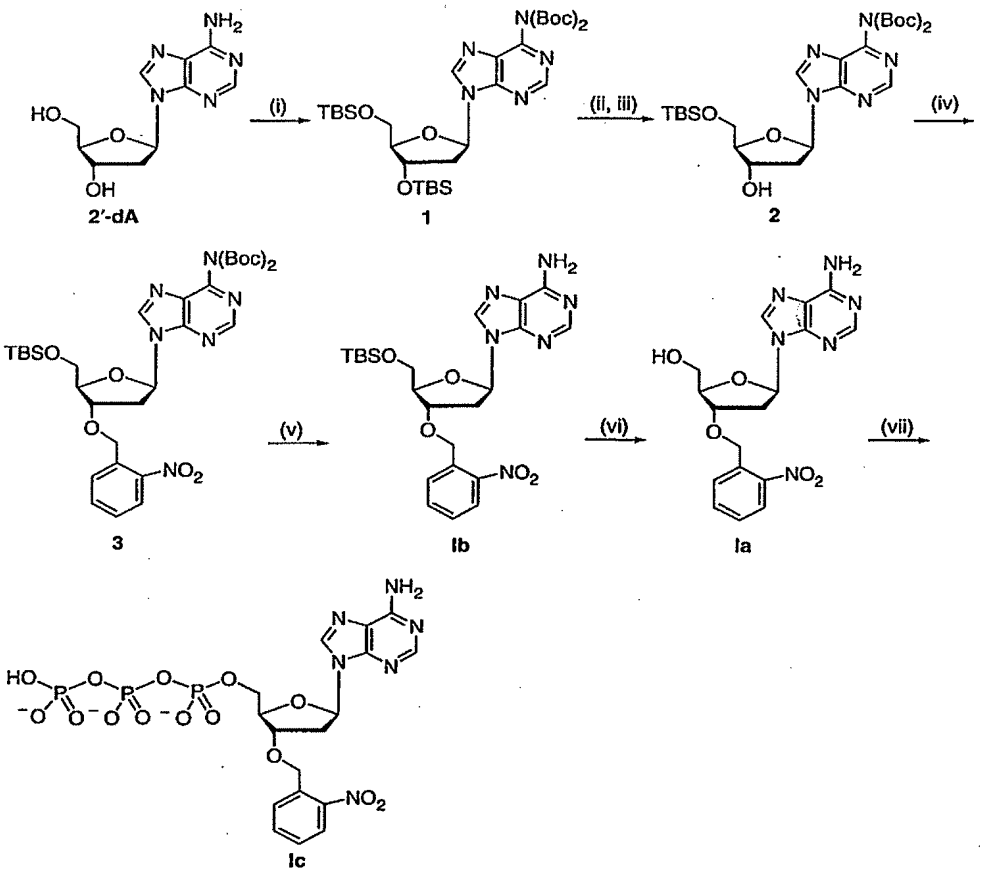
Treatment of 2′-deoxyadenosine with NaH in DMF at 0°C followed by 2-nitrobenzyl bromide gave *bis*-N^6, N^6-(2-nitrobenzyl)-2′-deoxyadenosine **IIa** in 22% yield (Scheme 3). The assignment of the **IIa** structure was based on the ^1H NMR spectra (in DMSO-d_6), which showed a total of eight aromatic hydrogens derived from the two 2-nitrobenzyl moieties and two D_2O-exchangeable signals from 3′- and 5′-hydroxyl groups of the 2′-deoxyribose. Selective 5′-*O*-TBS protection gave compound **IIb** in 57% yield. Phosphorylation of compound **IIa** was performed using the same procedure described for compound **Ia**, with the exception that purification of triphosphate **IIc** was achieved by preparative HPLC.

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

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

Identification of active nucleoside triphosphate IIIc

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

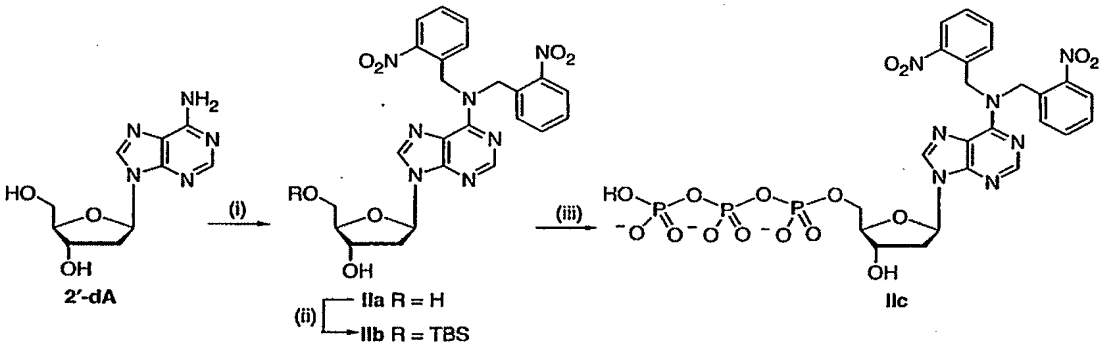


Scheme 2. Synthesis of 3′-*O*-(2-nitrobenzyl)-2′-deoxyadenosine analogs **Ia–Ic**. (i) TBSCl, imidazole, DMF, room temperature, overnight; Boc_2O, DMAP, DMF, room temperature, overnight, 83%; (ii) *n*-Bu_4NF, THF, 0°C, then gradually warmed to room temperature, 96%; (iii) TBSCl, imidazole, DMF, 83%; (iv) *n*-Bu_4NOH, NaI, NaOH, CH_2Cl_2/H_2O, 2-nitrobenzyl bromide, room temperature, 91%; (v) SiO_2, high vacuum, 70–80°C, 24 h, 91%; (vi) *n*-Bu_4NF, THF, 0°C, then gradually warmed to room temperature, 23%; (vii) $POCl_3$, $(MeO)_3PO$, minus 20°C; (*n*-$Bu_3NH)_2H_2P_2O_7$, *n*-Bu_3N, DMF; 1 M $HNEt_3HCO_3$, 31%.

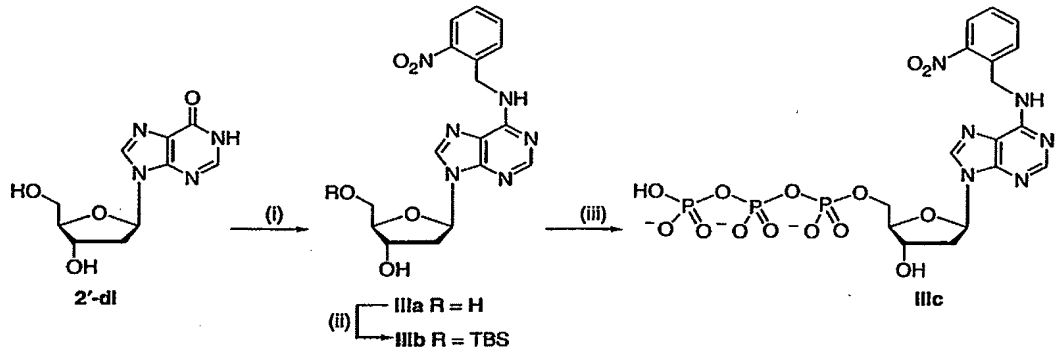
a second round of RP-HPLC purification was performed. The dATP contamination could, however, be substantially reduced using the Mop-Up assay (23). Next, we examined the rates of UV deprotection for 5′-*O*-TBS analogs **Ib** and **IIb**, conducted at wavelengths of 302 and 365 nm. Compound **Ib** exhibited 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 N^6-(2-nitrobenzyl)-2′-deoxyadenosine analog.

Following synthesis of N^6-(2-nitrobenzyl)-dATP **IIIc**, we examined its incorporation using *Bst* DNA polymerase, which showed efficient, base-specific termination of DNA synthesis at a final concentration of 200 nM (Figure 2A). The 2-nitrobenzyl group was efficiently removed from the DNA duplex using a custom built UV deprotector (Supplementary Figure 1), evidenced by the band shift of the extended dye-primer to the termination position of the first thymidine of the oligonucleotide template. Examination of the UV deprotection data for compound **IIIb** revealed a first-order reaction with $DT_{1/2}$ values of 46 and 144 s at 302 and 365 nm, respectively (Figure 2B). The UV deprotection data suggest that the attachment of a single 2-nitrobenzyl group to either the 3′-*O* or the N^6-aromatic amine position does not significantly alter the rate of reaction for UV light-induced cleavage.

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



Scheme 3. Synthesis of N^6,N^6-*bis*-(2-nitrobenzyl)-2′-deoxyadenosine analogs IIa–IIc. (i) NaH, 2-nitrobenzyl bromide, DMF, 0°C, 22%; (ii) TBSCl, imidazole, DMF, 57%; (iii) $POCl_3$, $(MeO)_3PO$, minus 20°C; $(n\text{-}Bu_3NH)_2H_2P_2O_7$, $n\text{-}Bu_3N$, DMF; 1 M $HNEt_3HCO_3$, yield not determined.



Scheme 4. Synthesis of N^6-(2-nitrobenzyl)-2′-deoxyadenosine analogs IIIa–IIIc. (i) BOP, DIPEA, 2-nitrobenzyl amine, DMF, room temperature, 20 h, then 50°C 3 h, 98%; (ii) TBSCl, imidazole, DMF, 63%; (iii) $POCl_3$, $(MeO)_3PO$, minus 20°C; $(n\text{-}Bu_3NH)_2H_2P_2O_7$, $n\text{-}Bu_3N$, DMF; 1 M $HNEt_3HCO_3$, 60%.

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

Compound IIIc is active with a variety of DNA polymerases

Numerous groups have employed qualitative, Sanger-based assays to estimate incorporation efficiencies, although these methods are not feasible for assaying modified analogs in the absence of natural nucleotides (8,25,26). This led us to develop a quantitative, PEP assay, which could be utilized for high-throughput screening of modified nucleotides against commercially available DNA polymerases. The PEP assay is designed with a polymerase concentration in excess of the primer/template complex, thereby limiting the reaction to nucleotide binding and coupling steps. The desired nucleotide is then titrated across the appropriate concentration range to observe extension of a dye-primer by gel electrophoresis. The end-point concentration, or 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).

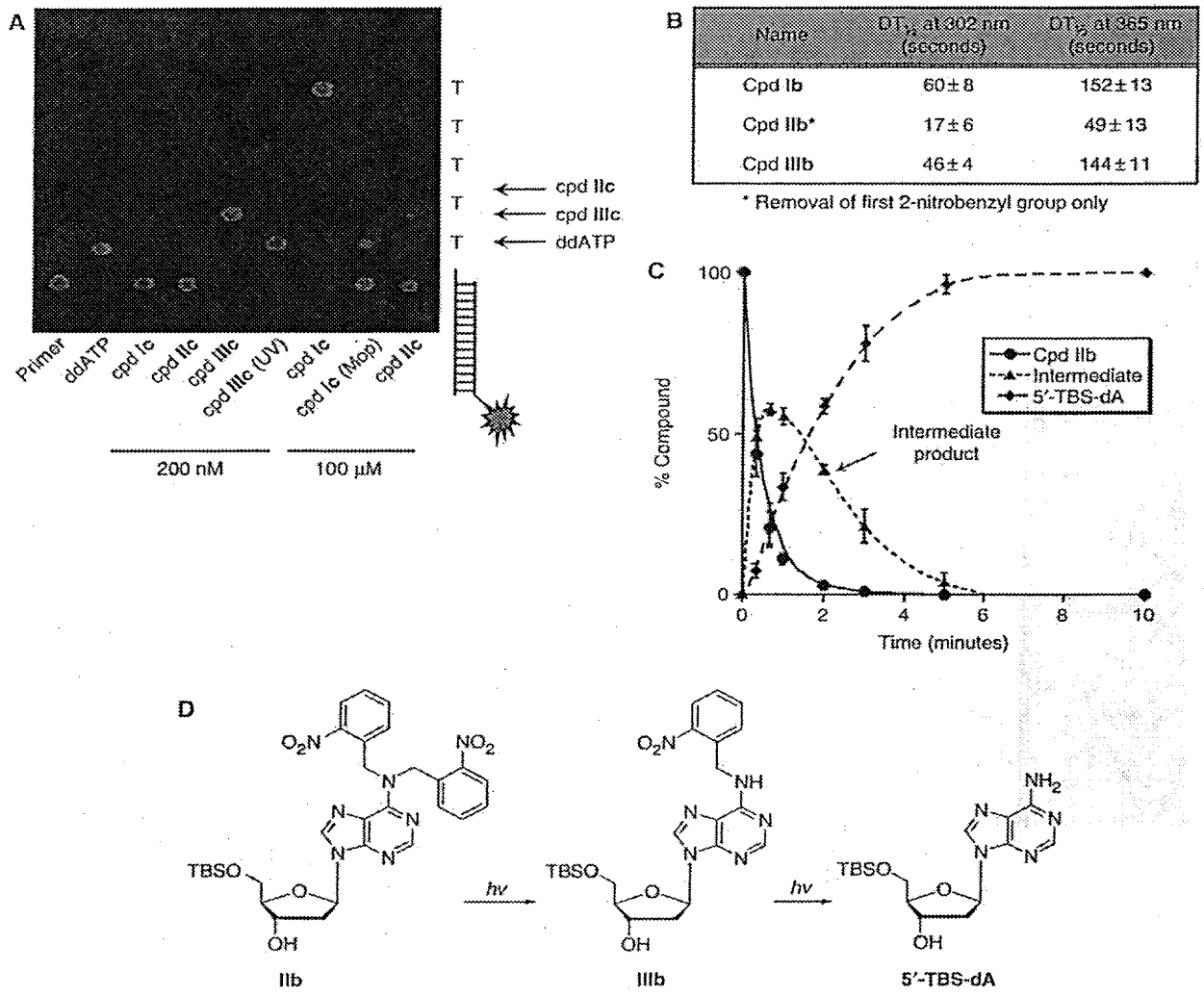


Figure 2. Identification of N^6-(2-nitrobenzyl)-2′-deoxyadenosine triphosphate **IIIc** as the active analogue. (A) OligoTemplate assay for compounds **Ic**, **IIc** and **IIIc**, analyzed on a 377 DNA sequencer. The final concentration for ddATP was 50 μM. (B) Summary of the UV deprotection experiments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

For these enzymes, increasing the number of units yielded an increase in IC_{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_{50} value for dATP (Supplementary Table 1, highlighted blue boxes).

N^6-(2-Nitrobenzyl)-dATP **IIIc** was tested for base-specific incorporation using eight polymerases with the PEP assay and compared with assay data for dATP and ddATP. Compound **IIIc** was incorporated by all polymerases examined, with IC_{50} values ranging from 2.1 nM to 2.1 μM (Table 1). Five of the eight polymerases revealed a less than 4-fold preference for dATP over compound **IIIc**, with Therminator and Vent(exo−) showing the least bias of 1:1.3, providing evidence that compound **IIIc** is incorporated almost as efficiently as dATP itself. In all cases except Ampli*Taq*FS, compound **IIIc** was preferred over ddATP with an incorporation bias range of 3.8×10^{-3}:1 to 0.32:1, respectively. The F667Y mutation in Ampli*Taq*FS has been shown to prefer ddATP over dATP, with a ratio of 0.59:1 (25), which is in good agreement with the ratio of 0.62:1 reported in Table 1.

Next, we examined the ability of N^6-(2-nitrobenzyl)-dATP **IIIc** to terminate DNA synthesis using an oligonucleotide template containing a stretch of 10 thymidine bases. As expected, *Bst* DNA polymerase extended the growing primer utilizing dATP as a substrate in a concentration-dependent manner. At a concentration of 25× its IC_{50} value, the enzyme completely extended the 10 thymidine template and partially misincorporated

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

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

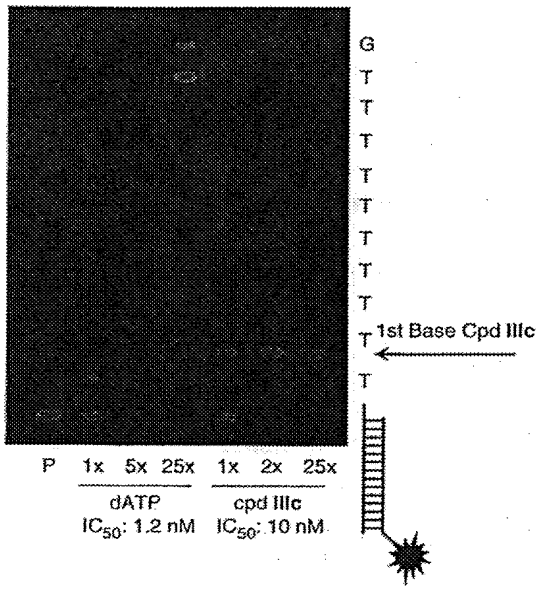


Figure 3. PEP termination assay comparing dATP with compound IIIc using *Bst* DNA polymerase.

dATP against a 'G' template base at the 11th base position (Figure 3). In contrast, the N^6-(2-nitrobenzyl)-dATP **IIIc** efficiently incorporated and terminated *Bst* DNA synthesis at the first-base position up to 25× its IC_{50} value. The difference in electrophoretic mobility of the first-base product for dATP and that of compound **IIIc** is due to the N^6-attached 2-nitrobenzyl group.

IC_{50} values for compound **IIIc** were retitrated using the poly(dT) template. In some cases, IC_{50} values differed from those in Table 1, reflecting DNA sequence context effects. PEP termination assays were performed for the eight polymerases at 1×, 5× and 25× the IC_{50} values (Table 2). Three of the four Family A DNA polymerases resulted in efficient termination of DNA synthesis at the first-base position using compound **IIIc**, while all Family B DNA polymerases showed significant, but variable, levels of second-base product. Therminator provided the most extreme example with ~98.5% of the growing primer extended as a second-base product. The majority

Table 2. PEP termination results for compound IIIc

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

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

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

PEP discrimination assays were then performed to evaluate the specificity of N^6-(2-nitrobenzyl)-dATP **IIIc** against mismatched template bases (i.e. A, C or G). Comparing IC_{50} values of mismatched versus matched bases for compound **IIIc** revealed nucleotide discrimination of greater than two orders of magnitude (Figure 4A). Surprisingly, the cytosine–adenine mismatch revealed the highest discrimination ratio of 1100-fold over the complement thymidine base, and only slightly less than that for dATP (Supplementary Table 2). The ddATP analog did not show mismatch incorporation at a final concentration of 500 μM (data not shown). These data suggest that compound **IIIc** incorporates as a base-specific terminator and reveals nucleotide characteristics more similar to dATP than those of ddATP, which we attribute to the presence of the 3′-OH group.

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

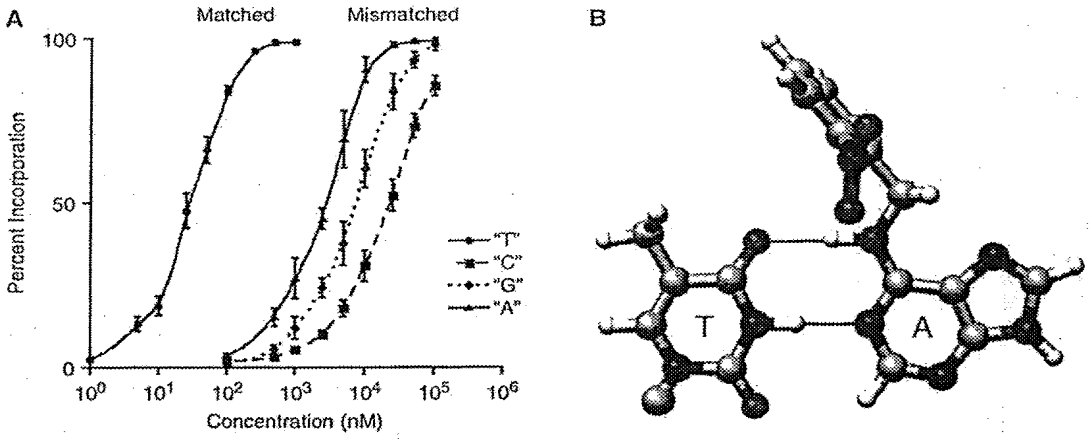


Figure 4. (A) PEP discrimination curves for compound IIIc. (B) Proposed thymine–N^6-(2-nitrobenzyl)-adenine base pair.

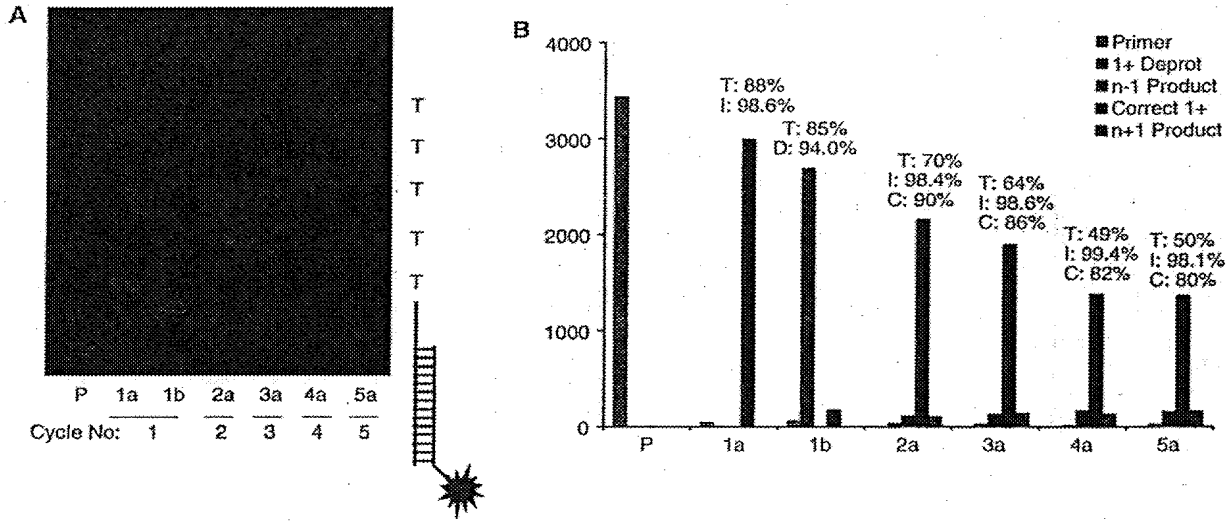


Figure 5. (A) Five-base CRT sequencing with N^6-(2-nitrobenzyl)-dATP IIIc. Lanes: 'P' represents the dye-primer, '1a–5a' represents final incorporations of compound IIIc at different cycles, and '1b' represents the first base addition followed by UV deprotection. (B) Histogram plot of the gel image in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

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

Five-base CRT sequencing with compound IIIc

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

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

DISCUSSION

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

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

Although the cycle efficiency is primarily the product of incorporation and deprotection efficiencies of the RT, other factors including dephasing (natural nucleotide carryover) and accumulating 'molecular scars' (residual linker structures left over after deprotection) can also influence the efficiency in an adverse manner. The N^6-attachment to the adenine nucleobase is also unique, differing from that of the traditional 7-deaza position of BigDye terminators (34) and 3′-*O*-allyl terminators (7,10). Upon chemical deprotection of the 3′-*O*-allyl terminators, a residual propargyl amino group remains on the nucleobase, resulting in an accumulating molecular scar with subsequent CRT cycles. The primary advantage of N^6-alkylation is that, upon directed photocleavage with 365 nm UV light, the modified nucleotide is transformed back into its natural state without molecular scarring (Figure 2D). The enhanced enzymatic properties and the N^6-cleavage site, transforming the efficiently incorporated RT back into natural DNA, are anticipated to improve the cycle efficiency and read-length of the CRT method.

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

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

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

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

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

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

REFERENCES

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