Molecular Engineering of Novel Nucleotide Analogues for DNA Sequencing and Analysis

Jia Guo

Submitted in partial fulfillment of
the requirements for the degree of Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2009
ABSTRACT

Molecular Engineering of Novel Nucleotide Analogues for DNA Sequencing and Analysis

Jia GUO

DNA sequencing by synthesis (SBS) on a solid surface during the polymerase reaction can decipher multiple DNA sequences in parallel. The first part of this thesis presents the development of a DNA sequencing method that is a hybrid between the Sanger dideoxy chain terminating reaction and SBS. In this approach, four nucleotides, modified as reversible terminators by capping the 3’-OH group with a small reversible moiety so that they are still recognized by DNA polymerase as substrates to extend the DNA chain, are used in combination with a small percentage of four cleavable fluorescent dideoxynucleotides to perform SBS. Sequences are determined by the unique fluorescence emission of each fluorophore on the DNA products terminated by ddNTPs. Upon removing the 3’-OH capping group from the DNA products generated by incorporating the 3’-O-modified dNTPs and the fluorophore from the DNA products terminated with the ddNTPs, the polymerase reaction reinitiates to continue the sequence determination. Various DNA templates, including those with homopolymer regions were accurately sequenced with readlengths of over 30 bases using this hybrid SBS method on
a chip and a four-color fluorescent scanner. To further extend the read-length of this hybrid sequencing method, a consecutive DNA sequencing by primer reset approach is developed. Upon removing the sequenced DNA strand and reattaching the original primer to allow the extension of this primer with a combination of natural and modified nucleotide analogues to the end of the first round sequence, the hybrid SBS can be carried out from that point to decipher the adjacent cluster of bases on the template. The sequencing read-length of a DNA template immobilized on a chip is almost doubled using this primer reset approach.

Single nucleotide polymorphisms (SNPs) are important markers for disease gene identification and for pharmacogenetic studies. The second part of this thesis describes the design, synthesis and evaluation of a chemically cleavable biotinylated nucleotide analogue, ddATP-N₃-biotin, for multiplex SNP analysis by MALDI-TOF MS. This nucleotide analogue has a biotin moiety attached to the 7-position of 2',3'-dideoxyadenosine 5'-triphosphate through a chemically cleavable azide-based linker. We have demonstrated that this ddATP-N₃-biotin is faithfully incorporated by the DNA polymerase Thermo Sequenase. The generated DNA extension products can be efficiently isolated by a streptavidin-coated surface and recovered under a mild chemical cleavage conditions. Single and multiple primer extension reactions were performed using ddATP-N₃-biotin to generate and isolate DNA extension products for MALDI-TOF MS analysis.

DNA microarray technology offers a paradigm for the study of genome-wide patterns of gene expression. The cDNA labeling step plays an important role in the accuracy and reproducibility of a microarray experiment. The third part of this thesis
focuses on the development of a click chemistry based cDNA labeling strategy for microarray analysis. In this approach, azide modified nucleotide analogues along with natural nucleotides are incorporated in reverse transcription reactions with RNA samples as templates. The azide groups on the generated cDNAs are coupled with alkyne functionalized fluorophores by click chemistry. Due to the high stability of the azide and alkyne groups in aqueous solution, the cDNAs are labeled efficiently with sufficient amount of the fluorescent molecules for microarray analysis using this approach.
# TABLE OF CONTENTS

Table of Contents ...........................................................................................................................................i
List of Figures and Tables .................................................................................................................................iv
Abbreviations and Symbols ...............................................................................................................................xiv
Acknowledgements ..........................................................................................................................................xvii
Dedication .........................................................................................................................................................xix

---

**Part I. Four-Color DNA Sequencing with 3’-O-modified Nucleotide Reversible Terminators and Chemically Cleavable Fluorescent Dideoxynucleotides**

Chapter 1: Introduction to DNA Sequencing and Analysis Technologies

1.1 Introduction ...........................................................................................................................................2
1.2 Background and significance .................................................................................................................3
1.3 Sanger dideoxynucleotide sequencing .................................................................................................8
1.4 DNA sequencing by hybridization ....................................................................................................11
1.5 DNA sequencing by MALDI-TOF MS ...............................................................................................12
1.6 Nanopore DNA sequencing ...............................................................................................................16
1.7 DNA sequencing by ligation ..............................................................................................................18
1.8 DNA sequencing by synthesis (SBS) ...............................................................................................23
Chapter 2: Four-color DNA Sequencing with 3’-O-Modified Nucleotide Reversible Terminators and Chemically Cleavable Fluorescent Dideoxynucleotides

2.1 Introduction ..............................................................................36
2.2 Experimental Rationale and Overview .................................38
2.3 Results and Discussion ...........................................................41
2.4 Materials and Methods ............................................................71
2.5 Conclusion .............................................................................96

References ......................................................................................100

Chapter 3: Consecutive Rounds of DNA Sequencing by Primer Reset

3.1 Introduction ..............................................................................102
3.2 Experimental Rationale and Overview .................................105
3.3 Results and Discussion ............................................................109
3.4 Materials and Methods ............................................................119
3.5 Conclusion .............................................................................128

References ......................................................................................130

Part II. Design and Synthesis of Chemically Cleavable Biotinylated Dideoxynucleotides for DNA Analysis with Mass Spectrometry

4.1 Introduction ..............................................................................133
4.2 Experimental Rationale and Overview .................................139
4.3 Results and Discussion ............................................................142
4.4 Materials and Methods ................................................................. 150
4.5 Conclusion .................................................................................... 155
References ......................................................................................... 156

Part III. Fluorescent Labeling of cDNA Probes with Click Chemistry for Microarray Analysis
5.1 Introduction .................................................................................. 159
5.2 Experimental Rationale and Overview ....................................... 163
5.3 Results and Discussion ................................................................. 165
5.4 Materials and Methods ................................................................. 173
5.5 Conclusion .................................................................................... 179
References ......................................................................................... 181
List of Figures and Tables

Part I. Four-Color DNA Sequencing with 3’-O-modified Nucleotide Reversible Terminators and Chemically Cleavable Fluorescent Dideoxynucleotides

Chapter 1: Introduction to DNA Sequencing and Analysis Technologies

Figure 1.1. Chemical structures of 2’-deoxyribonucleotides. Each nucleotide is composed of a base (adenine, guanine, cytosine or thymine), a sugar, and a phosphate group. (page 4)

Figure 1.2. DNA molecular structures. (A) 3-D computer rendered model of the DNA double helix. (B) A cartoon depicting two DNA molecules held together by hydrogen bonds between the paired bases. (C) A zoomed in section of the double helix, which shows the specific chemical structures of the bases. The efficient hydrogen bonding is only allowed between A and T, or between G and C. (page 6)

Figure 1.3. DNA replication. The DNA double helix is unwound and each strand acts as a template to incorporate complementary bases into new growing DNA strands. (page 7)

Figure 1.4. The mechanism of the DNA replication reaction. Incorporation of a nucleotide at the 3’ end of a growing DNA strand is a fundamental biological process, in which the base-pairing between the incoming nucleotide and the DNA template strand guides the generation of a new DNA strand. DNA polymerase catalyzes the extension of the growing DNA strand by incorporating the incoming nucleotide at the 3’-OH end with the formation of a phosphodiester bond and the release of pyrophosphate. (page 9)

Figure 1.5. Sanger dideoxy sequencing. To generate the DNA sequencing ladder, dNTPs and dye-labeled ddNTPs are combined to perform the DNA replication reaction. The incorporation of dNTPs generates a free hydroxyl group at the 3’ end of the DNA growing strand, which allows this DNA strand to be further extended; while the incorporation of dye-labeled ddNTPs eliminates the 3’-OH group of the DNA strand and terminates the strand extension. DNA fragments of different lengths produced by the DNA replication reaction are separated by gel electrophoresis. The unique fluorescence emission of each of the four dye-labeled ddNTPs indicates the sequence of the DNA template. (page 10)

Figure 1.6. Sequencing by hybridization. To resequence a specific base on the reference genome, four features are present on a chip, each identical except for the
base in the middle. The identity of the query base is determined by the differential hybridization of the unknown DNA sample to each of the four features, which contain bases complementary to the query sequence. (page 13)

**Figure 1.7.** Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS). Analyte molecules (such as DNA sequencing fragments) and matrix molecules (typically ultraviolet (UV) or infrared (IR) light-absorbing small organic molecules) are mixed in solution. After crystallization on a sample plate, analyte molecules are gently desorbed and ionized by UV laser irradiation. The resulting charged ions are accelerated under a constant electric voltage, which causes them to fly towards the ion detector. The time of the flight can be used to determine the masses of the charged ions. (page 15)

**Figure 1.8.** The α-hemolysin protein self-assembles in a lipid bilayer to form an ion channel and a nucleic acid stretch passes through it (top), generating corresponding electronic signatures (bottom). (page 17)

**Figure 1.9.** Hypothesized plot of translocation time versus blockade current when DNA molecules (A), (B), (C) are passing through the nanopore. (page 19)

**Figure 1.10.** Exonuclease-assisted nanopore DNA sequencing. An engineered α-hemolysin protein with an internal cyclodextrin adapter is placed in a well that contains two electrodes on either side of the pore. As the exonuclease directs individual DNA bases, in sequence, through the nanopore, each base transiently binds to the cyclodextrin and generates distinct changes in the electronic current, which indicate the sequence of the original DNA strand. (page 19)

**Figure 1.11.** Detector oligonucleotides for sequencing by ligation. (A) 8-mers detector oligonucleotides consist of five normal bases at the 3’ end, three universal bases at the 5’ end, and a unique fluorophore attached to the last base. Although each of the four fluorophores corresponds to four types of dinucleotides (B), with one base already known from the previous round in these dinucleotides, the other base can be unambiguously identified with the fluorescence signal. (page 21)

**Figure 1.12.** DNA sequencing by ligation. Each sequencing cycle consists of ligation, imaging, capping, and cleavage (Step 1-4). After repeating this process for 5-7 cycles (Step 5), a new primer with one base shorter length is introduced (Step 6) to perform a second round of sequencing (Step 7). Once five such sequencing rounds are completed, every base on the DNA strand except the last one has been interrogated twice, and the fluorescence signal indicates the sequence of the DNA template. (page 22)

**Figure 1.13.** DNA pyrosequencing. DNA polymerase catalyzes the incorporation of the complementary nucleotide into a growing DNA strand, which leads to the release of a pyrophosphate (PPi) molecule. This P Pi molecule is converted to ATP by sulfurylase, and visible light is subsequently produced by luciferase. The excess
nucleotides and ATP are then digested by apyrase. (page 24)

**Figure 1.14.** DNA sequencing by synthesis with chemically modified nucleotides. DNA polymerase catalyzes the incorporation of a fluorescently labeled nucleotide reversible terminator, which is complementary to the next base on the unknown DNA template. After removing unincorporated nucleotide analogues and other excess reagents, a fluorescence imager is used to identify the base just incorporated. Then the fluorophore and the 3’ capping moiety are chemically cleaved to reinitiate the DNA polymerase reaction. (page 26)

**Figure 1.15.** The 3D structure of a ternary complex of a rat DNA polymerase, a DNA template-primer, and a dideoxycytidine triphosphate (ddCTP). This figure, which shows the active site of the polymerase in the context of the polymerase DNA complex, illustrates that the 3’ position of the dideoxyribose ring is very crowded, while ample space is available at the 5 position of the cytidine base. (page 28)

**Figure 1.16.** SBS approach conducted in a high-throughput manner. Different DNA templates are immobilized on a chip to initiate the sequencing reaction. DNA polymerase catalyzes the incorporation of the complementary nucleotide analogues (step 1), which generate unique fluorescence emissions on the spot (step 2). This unique fluorescence signal indicates the identity of the incorporated nucleotide. After capping the unextended primers with ddNTPs (step 3), the fluorophore moiety and the 3’ capping group are chemically removed (step 4). With a reconstituted free 3’-OH group, the sequencing primer is ready for the next sequencing cycle (step 5). (page 29)

**Chapter 2: Four-color DNA Sequencing with 3’-O-Modified Nucleotide Reversible Terminators and Chemically Cleavable Fluorescent Dideoxynucleotides**

**Figure 2.1.** The hybrid DNA sequencing approach between the Sanger dideoxy chain terminating reaction and sequencing by synthesis. In this approach, four nucleotides (3’-O-R₁-dNTPs) modified as reversible terminators by capping the 3’-OH with a small reversible moiety R₁ so that they are still recognized by DNA polymerase as substrates, are used in combination with a small percentage of four cleavable fluorescent dideoxynucleotides (ddNTP-R₂-fluorophores) to perform SBS. DNA sequences are determined by the unique fluorescence emission of each fluorophore on the DNA products terminated by ddNTPs. Upon removing the 3’-OH capping group R₁ from the DNA products generated by incorporating the 3’-O-R₁-dNTPs, and the cleavage of the R₂ linker to remove the fluorophore from the DNA products terminated with the ddNTPs, the polymerase reaction reinitiates to continue the sequence determination. (page 40)

**Figure 2.2.** Structures of the nucleotide reversible terminators, 3’-O-N₃-dATP, 3’-O-N₃-dCTP, 3’-O-N₃-dGTP, and 3’-O-N₃-dTTP. (page 42)
Figure 2.3. Synthesis of 3’-O-azidomethyl-dATP. (page 44)

Figure 2.4. Synthesis of 3’-O-azidomethyl-dTTP. (page 44)

Figure 2.5. Synthesis of 3’-O-azidomethyl-dCTP. (page 45)

Figure 2.6. Synthesis of 3’-O-azidomethyl-dGTP. (page 45)

Figure 2.7. Staudinger reaction with TCEP to regenerate the 3’-OH group of the DNA extension product. (page 47)

Figure 2.8. The polymerase extension scheme (left) and MALDI-TOF MS spectra of the four consecutive extension products and their cleavage products (right) using the nucleotide reversible terminators (3’-O-N3-dNTPs). Primer extended with 3’-O-N3-dCTP (1) (A), and its cleavage product 2 (B); Product 2 extended with 3’-O-N3-dGTP (3) (C), and its cleavage product 4 (D); Product 4 extended with 3’-O-N3-dATP (5) (E), and its cleavage product 6 (F); Product 6 extended with 3’-O-N3-dTTP (7) (G), and its cleavage product 8 (H). After brief incubation in a TCEP aqueous solution the azidomethyl moiety capping the 3’-OH group of the DNA extension products is completely removed to continue the polymerase reaction. (page 48)

Figure 2.9. Structures of cleavable fluorescent deoxyxynucleotide terminators (A) ddNTP-N3 (version I)-fluorophores and (B) ddNTP-N3 (version II)-fluorophores, with the 4 fluorophores having distinct fluorescent emissions: Bodipy-FL-510 ($\lambda_{\text{abs (max)}} = 502$ nm, $\lambda_{\text{em (max)}} = 510$ nm), R6G ($\lambda_{\text{abs (max)}} = 525$ nm, $\lambda_{\text{em (max)}} = 550$ nm), ROX ($\lambda_{\text{abs (max)}} = 585$ nm, $\lambda_{\text{em (max)}} = 602$ nm), Cy5 ($\lambda_{\text{abs (max)}} = 649$ nm, $\lambda_{\text{em (max)}} = 670$ nm). (page 50)

Figure 2.10. Synthesis of the cleavable linker (version I). (page 52)

Figure 2.11. Synthesis of ddCTP-N3 (version I)-Bodipy-FL-510. (page 53)

Figure 2.12. Synthesis of ddUTP-N3 (version I)-R6G. (page 55)

Figure 2.13. Synthesis of ddATP-N3 (version I)-ROX. (page 56)

Figure 2.14. Synthesis of ddGTP-N3 (version I)-Cy5. (page 57)

Figure 2.15. Synthesis of the cleavable linker (version II). (page 58)

Figure 2.16. Synthesis of ddCTP-N3 (version II)-Bodipy-FL-510. (page 58)

Figure 2.17. Synthesis of ddUTP-N3 (version II)-R6G. (page 59)
Figure 2.18. Synthesis of ddATP-N3 (version II)-ROX. (page 60)

Figure 2.19. Synthesis of ddGTP-N3 (version II)-Cy5. (page 61)

Figure 2.20. Staudinger reaction with TCEP to cleave the N3-fluorophore from the dideoxynucleotide. (page 63)

Figure 2.21. A polymerase reaction scheme (top) to yield DNA extension products by incorporating each of the four ddNTP-N3 (version I)-fluorophores and the subsequent cleavage reaction to remove the fluorophores from the DNA extension products. MALDI-TOF MS spectra (bottom) showing efficient base specific incorporation of the ddNTP-N3-fluorophores and the subsequent cleavage of the fluorophores from the DNA extension products: (A) primer extended with ddATP-N3 (version I)-ROX (1) (peak at 9,180 m/z); (B) its cleavage product 2 (8,417 m/z); (C) primer extended with ddCTP-N3 (version I)-Bodipy-FL-510 (3) (peak at 8,915 m/z); (D) its cleavage product 4 (8,394 m/z); (E) primer extended with ddGTP-N3 (version I)-Cy5 (5) (peak at 9,261 m/z), (F) its cleavage product 6 (8,432 m/z); (G) primer extended with ddUTP-N3 (version I)-R6G (7) (peak at 9,082 m/z) and (H) its cleavage product 8 (8,395 m/z). (page 64)

Figure 2.22. A polymerase reaction scheme (top) to yield DNA extension products by incorporating each of the four ddNTP-N3 (version II)-fluorophores and the subsequent cleavage reaction to remove the fluorophores from the DNA extension products. MALDI-TOF MS spectra (bottom) showing efficient base specific incorporation of the ddNTP-N3-fluorophores and the subsequent cleavage of the fluorophores from the DNA extension products: (A) primer extended with ddATP-N3 (version II)-ROX (1) (peak at 9,003 m/z); (B) its cleavage product 2 (8,403 m/z); (C) primer extended with ddCTP-N3 (version II)-Bodipy-FL-510 (3) (peak at 8,738 m/z); (D) its cleavage product 4 (8,380 m/z); (E) primer extended with ddGTP-N3 (version II)-Cy5 (5) (peak at 9,140 m/z), (F) its cleavage product 6 (8,419 m/z); (G) primer extended with ddUTP-N3 (version II)-R6G (7) (peak at 8,905 m/z) and (H) its cleavage product 8 (8,381 m/z). (page 66)

Figure 2.23. (A) A hybrid SBS scheme for 4-color sequencing on a chip using the nucleotide reversible terminators (3’-O-N3-dNTPs) and cleavable fluorescent dideoxynucleotide terminals (ddNTP-N3 (version I)-fluorophores). (B) The 4-color fluorescence images for each step of SBS: (1) incorporation of 3’-O-N3-dCTP and ddCTP-N3 (version I)-Bodipy-FL-510; (2) cleavage of N3 (version I)-Bodipy-FL-510 and 3’-CH2N3 group; (3) incorporation of 3’-O-N3-ddATP and ddATP-N3 (version I)-Rox; (4) cleavage of N3 (version I)-ROX and 3’-CH2N3 group; (5) incorporation of 3’-O-N3-ddTTP and ddTTP-N3 (version I)-R6G; (6) cleavage of N3 (version I)-R6G and 3’-CH2N3 group; (7) incorporation of 3’-O-N3-ddGTP and ddGTP-N3 (version I)-Cy5; (8) cleavage of N3 (version I)-Cy5 and 3’-CH2N3 group; images 9–63 are similarly produced. (C) A plot (4-color sequencing data) of raw fluorescence emission intensity obtained using 3’-O-N3-dNTPs and ddNTP-N3 (version I)-fluorophores at the four designated emission wavelengths. (page 69)
Figure 2.24. A plot (4-color sequencing data) of raw fluorescence emission intensity obtained using 3’-O-N3-dNTPs and ddNTP-N3 (version I)-fluorophores at the four designated emission wavelengths of the four cleavable fluorescent dideoxynucleotides. (page 72)

Table 2.1 Volumes of solution A and B in each SBS cycle during de novo DNA sequencing. (page 97)

Chapter 3: Consecutive Rounds of DNA Sequencing by Primer Reset

Figure 3.1. Structures of nucleotide reversible terminators, 3’-O-R1-dATP, 3’-O-R1-dCTP, 3’-O-R1-dGTP and 3’-O-R1-dTTP. With the 3’-OH capped by a small reversible moiety R1, these nucleotide analogues are still recognized as good substrates by DNA polymerase. Upon removing the 3’-OH capping group R1 from the DNA extension product generated by incorporation of 3’-O-R1-dNTPs, the polymerase reaction reinitiates to continue the sequence determination. (page 103)

Figure 3.2. Structures of cleavable fluorescent dideoxynucleotide terminators, ddCTP-R2-Bodipy-FL 510, ddUTP-R2-R6G, ddATP-R2-ROX and ddGTP-R2-Cy5. Each of the four fluorophores is attached to the 5-position of pyrimidines and the 7-position of purines through a chemically cleavable linker R2. After incorporation of these dideoxynucleotide analogues, the fluorophores can be removed from the DNA extension products. (page 104)

Figure 3.3. The hybrid DNA sequencing approach between the Sanger dideoxy chain terminating reaction and sequencing by synthesis. In this approach, four nucleotides (3’-O-R1-dNTPs) modified as reversible terminators by capping the 3’-OH with a small reversible moiety R1 so that they are still recognized by DNA polymerase as substrates, are used in combination with a small percentage of four cleavable fluorescent dideoxynucleotides (ddNTP-R2-fluorophores) to perform SBS. DNA sequences are determined by the unique fluorescence emission of each fluorophore on the DNA products terminated by ddNTPs. Upon removing the 3’-OH capping group R1 from the DNA products generated by incorporating the 3’-O-R1-dNTPs, and the cleavage of the R2 linker to remove the fluorophore from the DNA products terminated with the ddNTPs, the polymerase reaction reinitiates to continue the sequence determination. (page 106)

Figure 3.4. Consecutive rounds of DNA sequencing by template “walking”. Upon ligation of universal primers A and B to both ends of each DNA template, a DNA library is prepared (1). Different DNA templates are immobilized on a PEG (polyethylene glycol) functionalized surface to initiate the sequencing reaction (2). The sequencing primer is then annealed to the DNA template (3). A first round of SBS extends the sequencing primer to produce the maximal read-length (4). After denaturing the sequencing primers, including those terminated with ddNTP analogues,
the original sequencing primer is reattached (5). Then unmodified or 3'-O-modified nucleotides are used to extend the primer approximately to the end of the first round sequence (6). Following that, the second round of SBS is performed to further extend the read-length. (page 108)

**Figure 3.5** Consecutive rounds of DNA sequencing by template “walking”. After the first sequencing round with the hybrid SBS approach, the extended primer is denatured. Upon reattachment of the original primer, three normal nucleotides are combined with another 3'-O-modified nucleotide to extend the primer to the end of the first round sequence. Then the second sequencing round is performed to further extend the read-length. (page 110)

**Figure 3.6** The polymerase extension scheme (left) and MALDI-TOF MS spectra of the three consecutive extension products and their cleavage products (right) using the combination of dATP, dCTP, dTTP and 3'-O-N3-dGTP. Primer extended with 3'-O-N3-dGTP (1) (A), and its cleavage product 2 (B); product 2 extended with 3'-O-N3-dGTP (3) (C), and its cleavage product 4 (D); product 4 extended with dATP, dCTP and 3'-O-N3-dGTP (5) (E), and its cleavage product 6 (F). (page 112)

**Figure 3.7.** (A) A general “template walking” scheme followed by hybrid SBS. The polymerase reaction with dATP, dCTP, dTTP and 3'-O-N3-dGTP extends the primer until it reaches the next C on the DNA template. After regenerating the 3'-OH group by using TCEP, the second and the third extension and cleavage reactions are performed in a similar manner. Then a subsequent round of hybrid SBS is carried out to generate the four-color DNA sequencing raw data (B). (page 115)

**Figure 3.8.** Consecutive rounds of DNA sequencing by primer reset on a chip. (A) Consecutive DNA sequencing scheme. The first round sequencing is carried out with the hybrid SBS approach. Upon the removal of the extended primer and the subsequent annealing of the original primer, the polymerase reaction with dATP, dCTP, dTTP and 3'-O-N3-dGTP extends the primer until the incorporation of 3'-O-N3-dGTP. After regenerating the 3'-OH group by using TCEP, the second extension and cleavage reactions are performed in a similar manner. Then a second round of sequencing is performed with the hybrid SBS approach to interrogate the identity of the consecutive bases. (B) A plot (four-color sequencing data) of raw fluorescence emission intensity vs. the progress of sequencing extension during the first and the second rounds of SBS. (page 117)

**Table 3.1** Volumes of solution A and B in each SBS cycle during de novo DNA sequencing. (page 126)

**Table 3.2** Volumes of solution A and B in each SBS cycle during the first round (A) and the second round (B) of de novo DNA sequencing. (page 126)

---

**Part II. Design and Synthesis of Chemically Cleavable Biotinylated**
Dideoxynucleotides for DNA Analysis with Mass Spectrometry

**Figure 4.1.** The structures of biotinylated dideoxynucleotides, biotin-11-ddATP, biotin-11-ddCTP, biotin-11-ddGTP and biotin-16-ddUTP. (page 135)

**Figure 4.2.** DNA sequencing by MALDI-TOF MS. (A) The DNA sequencing fragments are generated by the polymerase reaction by using the combination of dNTPs and biotion-ddNTPs. Then the pure DNA extension products are isolated with solid phase capture for MALDI-TOF MS analysis. (B) Illustration of the DNA sequencing data generated from the DNA fragments. (page 136)

**Figure 4.3.** Multiplex genotyping using molecular affinity and mass spectrometry. The biotinylated DNA extension products generated by single base extension using four biotin-ddNTPs are isolated on a streptavidin-coated surface. After removing the unextended primers and other reaction components, the pure DNA fragments are released for subsequent MALDI-TOF MS analysis. (page 138)

**Figure 4.4.** Affinity capture and chemical cleavage for isolating biotinylated DNA fragments on a streptavidin-coated surface. The DNA extension products with a cleavable biotin moiety at 3’ end are captured and then cleaved chemically to release the DNA fragments for analysis by mass spectrometry, leaving the biotin moiety still bound to the surface. (page 141)

**Figure 4.5.** Synthesis of ddATP-N₃-biotin. (page 143)

**Figure 4.6.** Staudinger reaction with TCEP to cleave the biotin moiety and release the DNA extension products from the streptavidin-coated surface. (page 145)

**Figure 4.7.** DNA extension reaction using ddATP-N₃-biotin and cleavage of the generated DNA fragment captured on a solid surface. DNA polymerase incorporates ddATP-N₃-biotin into a growing DNA stand, generating the cleavable biotinylated DNA fragment. Chemical cleavage using TCEP of this DNA fragment captured on the streptavidin-coated surface releases the DNA fragment, with the boitin moiety remaining on the surface. (page 146)

**Figure 4.8.** MALDI-TOF mass spectra of the DNA extension product generated from ddATP-N₃-biotin and the subsequent cleavage of the generated DNA fragment captured on a solid surface. (A) The DNA polymerase incorporated ddATP-N₃-biotin, yielding DNA extension product 6. (B) This generated DNA fragment was captured on a streptavidin-coated surface and a chemical cleavage reaction using TCEP was carried out to release the DNA fragment 7. (page 148)

**Figure 4.9.** (A) Multiplex DNA analysis scheme. Four DNA primers with different molecular weights were extended with ddATP-N₃-biotin by DNA polymerase. After solid phase capture and elimination of the unextended primers, a chemical cleavage reaction was carried out under mild conditions to release the DNA extension products.
Table 4.1 Sequences of oligonucleotide primers and the synthetic DNA template for the multiplex DNA extension reaction. (page 154)

Part III. Fluorescent Labeling of cDNA Probes with Click Chemistry for Microarray Analysis

Figure 5.1. Chemical structures of Cy3-dCTP and Cy5-dCTP. (page 161)

Figure 5.2. Direct cDNA fluorescent labeling for microarray experiments. Cy3-labeled dCTP and Cy5-labeled dCTP are incorporated during cDNA synthesis from control and test mRNA samples respectively. After degrading the mRNAs, the Cy3-labeled cDNA and Cy5-labeled cDNA are purified, mixed and hybridized on a DNA microarray chip. (page 162)

Figure 5.3. Chemical structure of aminoallyl-dUTP. (page 164)

Figure 5.4. Indirect cDNA fluorescent labeling for microarray experiments. The control and test mRNA samples are reverse-transcribed into cDNA using reverse transcriptase, dNTPs and aminoallyl-dUTP. Then the amino modified cDNA from the two samples are labeled with Cy3 and Cy5 respectively, followed by mixing and hybridization on the microarray. (page 164)

Figure 5.5. Indirect cDNA fluorescent labeling with click chemistry for microarray analysis. Taking the control and test mRNAs as templates, their corresponding cDNAs are synthesized using reverse transcriptase, dNTPs and azide-modified dUTP (N₃-dUTP). After degrading the mRNAs, the azide labeled cDNA from the two samples are coupled with alkyne modified Cy3 and Cy5 respectively. Subsequently, the two fluorescently labeled cDNA samples are mixed and hybridized on the microarray. (page 166)

Figure 5.6. Synthesis of N₃-dUTP. (page 168)

Figure 5.7. Synthesis of alkyne functionalized Cy3 and Cy5. (page 168)

Figure 5.8. Primer extension using N₃-dUTP along with reverse transcriptase and the subsequent fluorescent labeling with click chemistry. Taking RNA as the template, reverse transcriptase extended the DNA primer with N₃-dUTP. This azido-labeled DNA extension product was then coupled with alkyne functionalized Cy3 and Cy5. (page 169)

Figure 5.9. MALDI-TOF mass spectra of the DNA extension product generated from N₃-dUTP and the subsequent fluorescent labeling with click chemistry. (A) Reverse transcriptase incorporated N₃-dUTP, yielding DNA extension product 9. (B) This
resulting DNA fragment was coupled with Cy3 by click chemistry to generate Cy3 labeled DNA product 10. (B) The coupling reaction was carried out in a similar manner to convert the azido-modified DNA fragment into Cy5 labeled DNA product 11. (page 171)

**Figure 5.10.** (A) The scheme of fluorescent labeling of cDNA probes with click chemistry. The RNA sample was reverse-transcribed into their corresponding cDNA using reverse transcriptase, dNTPs and N3-dUTP. Then the azido labeled cDNA was coupled with alkyne functionalized Cy3 and Cy5 respectively. (B) UV-Vis absorption spectra of the Cy3 and Cy5 labeled cDNA probes. (page 172)

**Table 5.1** Sequences of the DNA primer and the RNA template for single base extension reaction. (page 177)
## Abbreviations and Symbols

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATP</td>
<td>adenosine 5’-triphosphate</td>
</tr>
<tr>
<td>Bodipy</td>
<td>4,4-difluoro-5,7-dimethyl-4-bora-3α,4α-diaza-s-indacene</td>
</tr>
<tr>
<td>bp</td>
<td>base pair</td>
</tr>
<tr>
<td>CHIP</td>
<td>chromatin immunoprecipitation</td>
</tr>
<tr>
<td>Cy3</td>
<td>cyanine-3</td>
</tr>
<tr>
<td>Cy5</td>
<td>cyanine-5</td>
</tr>
<tr>
<td>dA</td>
<td>deoxyadenosine</td>
</tr>
<tr>
<td>dC</td>
<td>deoxycytidine</td>
</tr>
<tr>
<td>ddNTP</td>
<td>dideoxynucleoside triphosphate</td>
</tr>
<tr>
<td>dG</td>
<td>deoxyguanosine</td>
</tr>
<tr>
<td>DMAP</td>
<td>4-dimethylaminopyridine</td>
</tr>
<tr>
<td>DMF</td>
<td>N,N-dimethylformamide</td>
</tr>
<tr>
<td>DMSO</td>
<td>dimethyl sulfoxide</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>dNTP</td>
<td>deoxynucleoside triphosphate</td>
</tr>
<tr>
<td>DSC</td>
<td>N,N’-disuccinimidyl carbonate</td>
</tr>
<tr>
<td>dT</td>
<td>deoxythymidine</td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>dU</td>
<td>deoxyuridine</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>Et&lt;sub&gt;3&lt;/sub&gt;N</td>
<td>triethylamine</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>HPLC</td>
<td>high performance liquid chromatography</td>
</tr>
<tr>
<td>IR</td>
<td>infrared</td>
</tr>
<tr>
<td>i-Pr</td>
<td>isopropyl</td>
</tr>
<tr>
<td>MALDI-TOF</td>
<td>matrix-assisted laser desorption ionization time-of-flight</td>
</tr>
<tr>
<td>MS</td>
<td>mass spectrometry</td>
</tr>
<tr>
<td>NHS</td>
<td>N-hydroxy succinimidyl</td>
</tr>
<tr>
<td>NRT</td>
<td>nucleotide reversible terminator</td>
</tr>
<tr>
<td>OPC</td>
<td>oligonucleotide purification cartridge</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PEG</td>
<td>polyethylene glycol</td>
</tr>
<tr>
<td>PPi</td>
<td>pyrophosphate</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>ROX</td>
<td>6-carboxy-X-rhodamine</td>
</tr>
<tr>
<td>R6G</td>
<td>6-carboxyrhodamine 6G hydrochloride</td>
</tr>
<tr>
<td>SAGE</td>
<td>serial analysis of gene expression</td>
</tr>
<tr>
<td>SBH</td>
<td>sequencing by hybridization</td>
</tr>
<tr>
<td>SBS</td>
<td>sequencing by synthesis</td>
</tr>
<tr>
<td>SNP</td>
<td>single nucleotide polymorphism</td>
</tr>
<tr>
<td>SPSC</td>
<td>sodium phosphate sodium chloride</td>
</tr>
<tr>
<td>TCEP</td>
<td>tris(2-carboxyethyl) phosphine</td>
</tr>
<tr>
<td>TEAB</td>
<td>triethylammonium bicarbonate</td>
</tr>
<tr>
<td>THF</td>
<td>tetrahydrofuran</td>
</tr>
<tr>
<td>UV</td>
<td>ultraviolet</td>
</tr>
</tbody>
</table>
$\lambda_{\text{abs}}$ maximum absorption wavelength (nm)

$\lambda_{\text{em}}$ maximum emission wavelength (nm)
Acknowledgements

First and foremost, I would like to express my most profound appreciation and
grateful appreciation to my advisors, Professor Jingyue Ju and Professor Nicholas Turro, for their
supervision, support and advice throughout the whole course of my graduate study. It is
my privilege to work under the guidance of these truly great scientists and superb
teachers. Their enthusiasm for science and professionalism in research have been and will
always be an inspiration to me.

My colleagues at the Columbia Genome Center, Department of Chemical
Engineering and Chemistry deserve my gratitude, too. My special thanks goes to Dr.
Zengmin Li, who has given me tremendous help in every aspect of the work since the
first day I joined our group. His dedication to science and broad knowledge in organic
chemistry have set a great example for me. I would like to thank Shundi Shi for helping
me design and set up the experiments for biological evaluations and applications of our
modified nucleotides. Her valuable advice and guidance are greatly appreciated. I also
owe my gratitude to Dr. James Russo for always being there to give the right advice and
spending his precious time to read and comment on this thesis. I am indebted to Drs.
Xiaoxu Li, Sergey Kalachikov, Dae Hyun Kim, Jian Wu, John Edwards, Lanrong Bi and
Chandima Abeywickrama for teaching me all the nuances of different experimental
procedures and working with me on various projects. Many thanks go to Drs. Shenglong
Zhang, Qinglin Meng, Huanyan Cao, and Mong Marma for their helpful discussions and
their sincere cooperation on various projects. I also thank Dr. Shiv Kumar for his
insightful views on chemical biology and many valuable comments and suggestions. It is
an honor for me to work with Yu Lin, Chunmei Qiu, Cheng Guo, Hernan Rengifo,
Wenjing Guo, Xinxin Zhu, Irina Morozova and Minchen Chien.

I would like to express my gratitude to Drs. Angel Marti and Steffen Jockusch for their great help on photochemistry studies. Many thanks to Dr Xuegong Lei, Dr Yi Rao and Dr. Yongjun Li for teaching me how to use various photochemical instruments and sharing their research experiences.

My special thanks go to Professor Louis Brus and Professor Ruben Gonzalez for being on my graduate committee and defense committee. I am very grateful for their precious time and efforts to evaluate and instruct my research.

I would like to thank my wonderful parents for their endless love, constant support and thorough understanding. I owe my deepest gratitude to my wife Ning Xu. I feel extremely fortunate to have such a considerate and unselfish soul mate in my life. The days and nights we spent together in Russ Berrie 403 will always be one of the most valuable memories in my heart. Without their love, support and encouragement, this thesis would not have been possible.

Finally, I would like to thank everybody who has helped me to successfully accomplish this research. I apologize for not mentioning them all by name.
Dedicated to my parents and Ning
Part I. Four-Color DNA Sequencing with \( 3'\)-O-modified Nucleotide Reversible Terminators and Chemically Cleavable Fluorescent Dideoxynucleotides
Chapter 1: Introduction to DNA Sequencing and Analysis Technologies

1.1 INTRODUCTION

The completion of the Human Genome Project at the dawn of the 21st century[1] was a monumental achievement resulting from the combined efforts of scientists worldwide. The engine behind this decade-long genome sequencing project was the Sanger sequencing method,[2] which is still considered the gold standard in high-throughput genome sequencing centers. The prolonged success of Sanger sequencing is due to its basic, efficient, and elegant method for producing dideoxy-terminated sequencing products which can be easily separated electrophoretically and detected using laser induced fluorescence.[3,4] With further improvements in this DNA sequencing technology, including the use of fluorescence energy transfer dyes,[5] engineered DNA polymerases,[6] and capillary array electrophoresis,[7] as well as progress in the areas of sample preparation, informatics, and sequence analysis software,[8,9] the Sanger sequencing platform has been able to maintain its dominance in the sequencing world. However, a challenge of using electrophoresis for DNA separation is its limitation for very high throughput and the complexity involved with automation, although some level of increased parallelization may be achieved using miniaturized lab-on-a-chip based approaches.[10, 11]

To overcome the limitations of the current electrophoresis-based Sanger sequencing technology, a variety of new DNA sequencing methods have been investigated. Such approaches include sequencing by hybridization,[12] mass spectrometry
based sequencing,\textsuperscript{[13-15]} sequence-specific detection of single-stranded DNA using engineered nanopores\textsuperscript{[16]} and sequencing by ligation.\textsuperscript{[17]} More recently, DNA sequencing by synthesis (SBS) approaches such as pyrosequencing\textsuperscript{[18]} and sequencing by synthesis using cleavable fluorescent reversible terminators\textsuperscript{[19, 20]} have been widely explored. This chapter will review the current standard sequencing technology and its limitations as well as compare the progress that has been made on a variety of new DNA sequencing methods.

1.2 BACKGROUND AND SIGNIFICANCE

In 1953, James Watson and Francis Crick deciphered the 3-dimensional structure of deoxyribonucleic acid (DNA), which contains the genetic instructions used in the development and functioning of all known living organisms and some viruses. The main role of DNA molecules is the long-term storage of information, which is needed to construct other components of cells, such as proteins and RNA molecules. Chemically, DNA consists of two long polymers of simple units called nucleotides, with backbones made of sugars and phosphate groups joined by ester bonds. These two strands run in opposite directions to each other and are therefore anti-parallel. Attached to each sugar is one of four molecules, often referred to as bases, adenine (A), guanine (G), thymine (T), and cytosine (C). Adenine and guanine are purine derivatives, while thymine and cytosine are pyrimidine derivatives. The heterocyclic amine bases are bonded to the C1’ of the deoxyribose, and the phosphoric acid is bonded by a phosphate ester linkage to the C5’ sugar position (Fig. 1.1). Held together by hydrogen bonds between purine and pyrimidine bases, the two DNA strands feature the phosphate and the sugar units
Figure 1.1. Chemical structures of 2'-deoxyribonucleotides. Each nucleotide is composed of a base (adenine, guanine, cytosine or thymine), a sugar, and a phosphate group.
positioned outside and the bases positioned inside. Under physiological conditions, adenine is always paired with thymine, and guanine always with cytosine (Fig. 1.2). The specific base pairings are referred to as “Waston and Crick base-pairing” or “complementary base-pairing”. The “backbone” of DNA is composed of sugars linked by phosphate groups, where the 3’-hydroxyl group of one nucleotide is attached to the 5’-phosphate group of the adjacent nucleotide through a phosphodiester bond. The ends of the DNA strand consist of a phosphate group at the 5’ end and a hydroxyl group at the 3’ end. Since the bases in the nucleotides are the only variable component in DNA molecules, it is the sequence of these bases (A, G, T and C) that determines specific genetic information.

To generate two daughter cells during cell division, the primary concern is the maintenance of the original cell’s genome, which requires that the genomic material stored in chromosomes must be replicated at extremely high accuracy. The structure of the DNA double helix and the precise complementary nature of the DNA strands allow for this highly accurate replication mechanism. DNA replication occurs in a “semiconservative” manner, where the two complementary DNA strands are partially separated, with each separated DNA strand serving as a template for the reproduction of its complementary strand (Fig. 1.3). This replication process has extremely high efficiency and fidelity in vivo. Generally, hundreds of nucleotides can be incorporated into a DNA strand in a second, with an error rate less than $10^{-8}$ per base pair.\textsuperscript{[21]}

During DNA replication, an enzyme called DNA polymerase catalyzes this primer elongation reaction. In the presence of divalent metal ions such as magnesium, DNA polymerase incorporates the complementary nucleotides to the 3’ end of a growing
Figure 1.2. DNA molecular structures. (A) 3-D computer rendered model of the DNA double helix. (B) A cartoon depicting two DNA molecules held together by hydrogen bonds between the paired bases. (C) A zoomed in section of the double helix, which shows the specific chemical structures of the bases. The efficient hydrogen bonding is only allowed between A and T, or between G and C.
Figure 1.3. DNA replication. The DNA double helix is unwound and each strand acts as a template to incorporate complementary bases into new growing DNA strands.
DNA strand (Fig. 1.4). This DNA primer extension reaction involves a nucleophilic attack of the hydroxyl group at the 3’ end of the primer on the alpha phosphorus atom of an incoming nucleotide. With the release of pyrophosphate (PPI) as the by-product, the phosphodiester bond is formed. Proceeding always in the 5’ to 3’ direction, this DNA polymerase catalyzed replication is a template-directed event. Specifically, the DNA polymerase only incorporates the incoming nucleotide, which is complementary to the next nucleotide on the DNA template.

This precise complementary feature of the DNA replication reaction is the basis of various emerging DNA sequencing technologies. This chapter will present these DNA sequencing technologies in more detail.

1.3 SANGER DIDEOXYNUCLEOTIDE SEQUENCING

In 1977, Sanger and co-workers developed a DNA sequencing approach by generating DNA sequencing fragments terminated by dideoxynucleotides.[2] Sanger’s dideoxy sequencing method has become the standard approach for large scale sequencing projects, including the Human Genome Project completed in 2001.[1]

The current automated Sanger sequencing methodology utilizes four natural 2’-deoxyribonucleotides (dNTPs) along with four fluorescently labeled 2’, 3’-dideoxyribonucleotides (ddNTPs) to decipher the DNA sequences (Fig. 1.5). With the absence of a hydroxyl group on the 3’ position on the sugar moiety, the latter nucleotide analogues serve as terminators in the polymerase reaction. Thus, the DNA replication reaction using a mixture of natural dNTPs and a much lower amount of chain terminating
**Figure 1.4.** The mechanism of the DNA replication reaction. Incorporation of a nucleotide at the 3’ end of a growing DNA strand is a fundamental biological process, in which the base-pairing between the incoming nucleotide and the DNA template strand guides the generation of a new DNA strand. DNA polymerase catalyzes the extension of the growing DNA strand by incorporating the incoming nucleotide at the 3’-OH end with the formation of a phosphodiester bond and the release of pyrophosphate.
Figure 1.5. Sanger dideoxy sequencing. To generate the DNA sequencing ladder, dNTPs and dye-labeled ddNTPs are combined to perform the DNA replication reaction. The incorporation of dNTPs generates a free hydroxyl group at the 3' end of the DNA growing strand, which allows this DNA strand to be further extended; while the incorporation of dye-labeled ddNTPs eliminates the 3'-OH group of the DNA strand and terminates the strand extension. DNA fragments of different lengths produced by the DNA replication reaction are separated by gel electrophoresis. The unique fluorescence emission of each of the four dye-labeled ddNTPs indicates the sequence of the DNA template.
ddNTPs produces DNA extension products of various length, all of which consist of natural nucleotides and a single fluorescently labeled dideoxy-nucleotide analogue at their 3’ ends. These DNA fragments are then separated by gel electrophoresis based on their different lengths. After scanning the sequence ladder with an excitation laser, the unique fluorescent emission of these fluorophores indicates the identity of the dideoxy-nucleotide analogue at the 3’ end of each DNA strand and thereby the sequence of the original DNA template. Many advances in all aspects of the Saner sequencing platform, including the use of fluorescence energy transfer dyes, engineered DNA polymerases, and capillary array electrophoresis, have led to the routine use of automated machines in large sequencing facilities.

Current DNA sequencers based on the Sanger sequencing approach can generate over 700 bases of clearly readable sequence in a single run, and allow up to 384 DNA templates to be analyzed in parallel. Although Sanger sequencing has been the gold standard in genome research for over three decades, this electrophoresis-based sequencing technology has some intrinsic difficulties in achieving high throughput and miniaturization. Hence, cost is the major limiting factor for applying this sequencing technology to achieve personalized medicine and clinical diagnosis.

1.4 DNA SEQUENCING BY HYBRIDIZATION

Sequencing by hybridization (SBH) is an inexpensive alternative to the commonly used Sanger dideoxy sequencing approach. SBH takes advantage of differential oligonucleotide hybridization to identify the set of constituent subsequences present in a DNA template. One approach is to immobilize unknown DNA samples to a
surface, which are subsequently hybridized with labeled oligonucleotides of known sequence. The sequence of the unknown DNA can be determined by the extent to which the labeled oligonucleotides bind to it. An alternative method is to attach the oligonucleotides of known sequence to a substrate and to hybridize with a labeled unknown DNA. For example, to resequence each base pair of a reference genome, four features, each of which consists of 1,000,000 copies of a defined 25-bp oligonucleotide, are constructed on a chip (Fig. 1.6).\textsuperscript{[23]} The only variable base on the oligonucleotides among these four features is the base in the middle, which is an A, C, T or G. The sequence surrounding this middle base is identical for all the four features and matches the reference genome. Upon hybridizing the labeled unknown DNA sample to the chip, the feature that yields the strongest signal among the four clearly indicates the identity of the specific base in the DNA sample.

SBH technology has its unique advantages and challenges. It can identify a decent amount of sequence (\(>10^9\) bases) from many distinct chromosomes simultaneously. Nonetheless, this hybridization-based approach has inherent difficulties to achieve optimal stability for all probes under the same conditions. Low efficiency probes do not produce strong enough signals even the target is fully matched, while probes with very stable full match hybrids also produce fairly strong signals with mismatch targets. In the short term, SBH has great potential as a tool to analyze the genotype of focused genomic positions, for example, the common SNPs in the human population.\textsuperscript{[24, 25]}

**1.5 DNA SEQUENCING BY MALDI-TOF MS**

Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry
Figure 1.6. Sequencing by hybridization. To resequence a specific base on the reference genome, four features are present on a chip, each identical except for the base in the middle. The identity of the query base is determined by the differential hybridization of the unknown DNA sample to each of the four features, which contain bases complementary to the query sequence.
(MALDI-TOF MS) has emerged as a rapid, precise and efficient analytical tool in DNA sequencing and genotyping. It has the potential to address the problems typically encountered when using fluorescence capillary electrophoresis sequencing techniques, such as poor resolution in GC rich regions owning to compression,\cite{26,27} and ambiguous detection of heterozygotes.

DNA sequencing by mass spectrometry separates and characterizes DNA ladder fragments based on their mass. In this approach, a matrix solution (typically UV or IR absorbing small organic molecules) is mixed with the DNA fragments of different length (Fig. 1.7).\cite{28} With this mixture added on a flat sample plate and allowed to crystallize, a laser is introduced to hit the resulting crystal and ionize the sample. The currently accepted ionization mechanism is that the collision of DNA fragments with the matrix molecules, which are initially ionized by the laser, results in charge transfer from the matrix molecules to the DNA fragments. When an electric field is applied to these ionized DNA fragments, they will fly through a high-vacuum tube toward the detector. Based on their mass to charge ratio, short DNA fragments with lower molecular mass fly faster and arrive at the detector sooner than the longer ones. The time for the fragments to arrive at the detector is then used to compute the mass of the DNA fragments, which indicates the sequence of the original DNA template.

Compared with gel electrophoresis-based sequencing systems, MALDI-TOF MS produces high resolution short DNA sequencing data of less than 100 base pairs, rapid fragment separation on a microsecond time scale and the complete elimination of the issue of compressions. The challenge this technology faces is the difficulty to analyze DNA fragments longer than 100 bases reliably and consistently with sufficient mass
Figure 1.7. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS). Analyte molecules (such as DNA sequencing fragments) and matrix molecules (typically ultraviolet (UV) or infrared (IR) light-absorbing small organic molecules) are mixed in solution. After crystallization on a sample plate, analyte molecules are gently desorbed and ionized by UV laser irradiation. The resulting charged ions are accelerated under a constant electric voltage, which causes them to fly towards the ion detector. The time of the flight can be used to determine the masses of the charged ions.
resolution. With a reliable sequencing read-length of 30 base pairs, MALDI-TOF MS sequencing methods are ideal for directed mutation screening in indel (insertion/deletion) regions.

### 1.6 NANOPORE DNA SEQUENCING

Nanopore DNA sequencing has been envisioned as one of the most promising technologies for achieving real time, ultra-fast, true single molecule DNA sequencing. In this approach, nanopores are immersed in a conducting fluid and an electric current can be detected when voltage is applied across it. Nanopore DNA sequencing and analysis usually involve the measurement of changes in this electric current while a nucleic acid molecule is pulled through it.

The $\alpha$–hemolysin channel has been demonstrated to analyze nucleic acids at the single molecule level.$^{[16]}$ Self-assembled in a lipid bilayer membrane, the $\alpha$–hemolysin channel features a 2.6 nm-diameter vestibule and 1.5 nm-diameter limiting aperture (the narrowest point of the pore)$^{[29-31]}.$ With the $\alpha$–hemolysin channel in a membrane immersed in an aqueous ionic salt solution, a distinguishable ionic current is generated when an appropriate voltage is applied. Due to the size limit of the aperture, double-stranded DNA molecules with the diameter of approximately 2.0 nm are not able to pass through the $\alpha$–hemolysin channel. Driven by the applied electric field, the nucleic acids pass through the nanopore and partially block the ionic current. The generated electronic signature (Fig. 1.8) indicates the passage of the particular DNA molecule.$^{[31]}$ Under steady voltage, the duration of the blocked ionic current is proportional to the length
Figure 1.8. The α-hemolysin protein self-assembles in a lipid bilayer to form an ion channel and a nucleic acid stretch passes through it (top), generating corresponding electronic signatures (bottom).
of the DNA strand. The variation of the ionic current during this process infers the steric and electronic properties of the nucleotides passing through the nanopore. Thus the length and the sequence of the specific DNA molecule are identified with this characteristic ionic current signature (Fig. 1.9).

An alternative approach uses exonuclease to liberate individual nucleoside monophosphates from a strand of DNA and identifies these bases in order of release (Fig. 1.10). In order to differentiate the four nucleoside monophosphates (dAMP, dCMP, dTMP and dGMP), a molecular adapter cyclodextrin is covalently attached to the inside of the $\alpha$-hemolysin pore. As single bases released by an exonuclease pass through the pore, they transiently interact with the cyclodextrin and create a characteristic signal for each type of base.

Although nanopore DNA sequencing is at its early proof of principle stage, this sequencing platform has got intensive improvements in many aspects and has great potential to be further developed into a rapid, accurate and inexpensive sequencing technology.

### 1.7 DNA SEQUENCING BY LIGATION

Sequencing by ligation is a DNA sequencing method that takes advantage of the enzyme DNA ligase to determine the identity of the nucleotide at a given position in a DNA strand. In this approach, fluorescently labeled oligonucleotides (8-mers) are used in the ligation reaction to interrogate the sequence of the DNA molecule (Fig. 1.11A). The first two bases of the oligonucleotides at the 3’ end determine the color of the fluorophore and each fluorophore corresponds to four different dinucleotides (Fig. 1.11B). If the first or second base in the dinucleotide is known, the color of the fluorophore is
Figure 1.9. Hypothesized plot of translocation time versus blockade current when DNA molecules (A), (B), (C) are passing through the nanopore.

Figure 1.10. Exonuclease-assisted nanopore DNA sequencing. An engineered α-hemolysin protein with an internal cyclodextrin adapter is placed in a well that contains two electrodes on either side of the pore. As the exonuclease directs individual DNA bases, in sequence, through the nanopore, each base transiently binds to the cyclodextrin and generates distinct changes in the electronic current, which indicate the sequence of the original DNA strand.
unambiguously related with the other base (Fig. 1.11C). The next three nucleotides in
detector oligonucleotides are degenerate nucleotides. With 64 ($4^3$) versions for each
particular dinucleotide, only one of these variants will be perfectly complementary to the
sequence of the DNA template and be ligated to the sequencing primer. The last three
nucleotides consisting of universal bases, which indiscriminately pair with any bases, are
the same for all the detector oligonucleotides. To sequence an unknown DNA molecule
immobilized on a surface, a universal sequencing primer is annealed (Fig. 1.12). Upon
adding the whole set of fluorescently labeled oligonucleotides and DNA ligase, only the
one perfectly complementary to the DNA template is ligated (Step 1). After removing the
excess reagent and washing away any unligated oligonucleotides, a 4-color fluorescence
imager is used to register the fluorescence signal, which indicates the identity of the
second base in the detector oligonucleotide if the first base is known (Step 2). After
imaging, the small amount of the unextended primers will be capped by removing the
phosphate group by phosphatase, to avoid interference with the next round of sequencing
(Step 3). The dye moiety together with the last three universal bases will then be cleaved
chemically to regenerate a phosphate group at the 5’ end (Step 4). The sequencing primer
at this stage is ready for the next cycle of the ligation reaction (Step 5). After five to
seven sequencing cycles, the initial primer and all extended portions are denatured from
the DNA template and discarded. Then a new sequencing primer, which is one base
shorter than the initial primer, is annealed to the unknown DNA template to repeat the
sequencing cycles (Step 6 & 7). After five rounds of primer reset, each base on the
template is interrogated in two independent ligation reactions by two different primers
Figure 1.11. Detector oligonucleotides for sequencing by ligation. (A) 8-mer detector oligonucleotides consist of five normal bases at the 3’ end, three universal bases at the 5’ end, and a unique fluorophore attached to the last base. Although each of the four fluorophores corresponds to four types of dinucleotides (B), with one base already known from the previous round in these dinucleotides, the other base can be unambiguously identified with the fluorescence signal.
Figure 1.12. DNA sequencing by ligation. Each sequencing cycle consists of ligation, imaging, capping, and cleavage (Step 1-4). After repeating this process for 5-7 cycles (Step 5), a new primer with one base shorter length is introduced (Step 6) to perform a second round of sequencing (Step 7). Once five such sequencing rounds are completed, every base on the DNA strand except the last one has been interrogated twice, and the fluorescence signal indicates the sequence of the DNA template.
(Step 8). Although each fluorophore corresponds to four different dinucleotides, starting from the first known nucleotide the whole sequence of the template can be unambiguously determined by the fluorescence signal obtained in each ligation reaction.

One advantage of this technology is that sequencing by ligation can proceed in both the 5’-to-3’ and 3’-to-5’ directions. Additionally, interrogating every base in two independent ligation reactions leads to a sequencing accuracy in excess of 99.9%. Nevertheless, the major challenge for this sequencing approach still is its relatively short read-length. Thus, ligation strategies are ideal for resequencing specific genomic regions.

1.8 DNA SEQUENCING BY SYNTHESIS (SBS)

To overcome the limitations of the Sanger sequencing technology, the sequencing by synthesis (SBS) approach has emerged as a viable candidate for a massively parallel high throughput sequencing platform. SBS takes advantage of the polymerase reaction to identify each incorporated nucleotide.\textsuperscript{[34]} In each sequencing cycle, the polymerase reaction is temporarily terminated once the nucleotide analogue is incorporated. After sequence determination, the polymerase reaction is reinitiated to identify the next base on the DNA template. By interrogating one base on each DNA template in one sequencing cycle, SBS approaches can easily scale-up over Sanger sequencing to analyze millions of DNA templates simultaneously.

Pyrosequencing, which was developed based on this concept, has been explored for DNA sequencing.\textsuperscript{[35]} In this approach, each of the four natural nucleotides (dATP, dCTP, dGTP and dTTP) is added sequentially during the DNA polymerase reaction (Fig. 1.13).\textsuperscript{[18]} If the added nucleotide is complementary to the first available base on the
Figure 1.13. DNA pyrosequencing. DNA polymerase catalyzes the incorporation of the complementary nucleotide into a growing DNA strand, which leads to the release of a pyrophosphate (PPI) molecule. This PPI molecule is converted to ATP by sulfurylase, and visible light is subsequently produced by luciferase. The excess nucleotides and ATP are then digested by apyrase.
template, the nucleotide will be incorporated, leading to pyrophosphate (PPi) release. Through the quantitative conversion of pyrophosphate to ATP by sulfurylase, the subsequent visible light signal produced by firefly luciferase is detected.

Pyrosequencing has been applied to study single nucleotide polymorphisms (SNPs)\[^{18}\] and sequence DNA.\[^{36}\] However, conventional pyrosequencing has inherent difficulties for deciphering homopolymeric regions of the DNA templates.\[^{18}\] The reason is that the light signal intensity is not exactly proportional to the amount of PPi released, especially when the homopolymeric region has more than five bases. Although Wu et al. have solved this problem of determining the number of nucleotides in the homopolymeric regions by using nucleotide reversible terminators,\[^{37}\] other aspects of pyrosequencing still need improvement. For example, each of the four nucleotides has to be added and detected separately, which increases the overall sequencing time. More importantly, without the competition among the four nucleotides as substrates in the polymerase reaction, the sequencing accuracy is reduced.

To decrease the overall detection time, improve the sequencing accuracy and miniaturize the instrument, Ju and co-workers have developed an integrated SBS approach for a high throughput sequencing platform.\[^{38}\] This method relies on using the polymerase reaction to read out the DNA sequence. After the incorporation of fluorescently labeled nucleotides, the fluorescent signal is detected to determine the identity of the incorporated nucleotides (Fig. 1.14). In order to temporarily pause the polymerase reaction and to accurately sequence through homopolymeric regions, the 3’ hydroxyl group of the nucleotide must be blocked by a chemical moiety. The fluorophore
Figure 1.14. DNA sequencing by synthesis with chemically modified nucleotides. DNA polymerase catalyzes the incorporation of a fluorescently labeled nucleotide reversible terminator, which is complementary to the next base on the unknown DNA template. After removing unincorporated nucleotide analogues and other excess reagents, a fluorescence imager is used to identify the base just incorporated. Then the fluorophore and the 3’ capping moiety are chemically cleaved to reinitiate the DNA polymerase reaction.
and the 3’ blocking group then need to be efficiently removed in a manner compatible with DNA stability and function for subsequent sequence determination.

In order to design the structures of cleavable fluorescent nucleotide reversible terminators (NRTs) used in the SBS reaction, it is necessary to examine the 3-D structure of the polymerase enzyme complex with a DNA template, a primer and an incoming nucleotide during a polymerase reaction (Figure 1.15). It is apparent that the 5-position of the cytosine points away from the catalytic pocket of the enzyme, while the 3’-position of the ribose ring in ddCTP is in a very crowded space near the active amino acid residues of the polymerase. Thus, to ensure these chemically modified nucleotide analogues to be recognized as good substrates by the DNA polymerase, the 3’-OH capping moiety must be small and the bulky fluorescent moiety has to be attached at the 5 position of pyrimidines and the 7 position of purines so as not to interfere with the polymerase reaction.

The general sequencing by synthesis approach using cleavable fluorescent nucleotide analogues is shown in Figure 1.16. A chip is constructed with a large number of different immobilized DNA templates. Four nucleotide analogues are designed in such a way that each is chemically modified with a unique fluorescent dye on the base, and a small chemical moiety (R) to cap the 3’-OH group. Upon adding the four nucleotide analogues and DNA polymerase, only the nucleotide analogue complementary to the next available nucleotide on the template is incorporated (Step 1). Once the excess reagents and any unincorporated nucleotide analogs are washed away, a four color fluorescence imager is used to register the signal. The unique fluorescent emission from each spot on the chip will yield the identity of the nucleotide (Step 2). After imaging, the small amount
Figure 1.15. The 3D structure of a ternary complex of a rat DNA polymerase, a DNA template-primer, and a dideoxycytidine triphosphate (ddCTP). This figure, which shows the active site of the polymerase in the context of the polymerase DNA complex, illustrates that the 3' position of the dideoxyribose ring is very crowded, while ample space is available at the 5 position of the cytidine base.
Figure 1.16. SBS approach conducted in a high-throughput manner. Different DNA templates are immobilized on a chip to initiate the sequencing reaction. DNA polymerase catalyzes the incorporation of the complementary nucleotide analogues (step 1), which generate unique fluorescence emissions on the spot (step 2). This unique fluorescence signal indicates the identity of the incorporated nucleotide. After capping the unextended primers with ddNTPs (step 3), the fluorophore moiety and the 3’ capping group are chemically removed (step 4). With a reconstituted free 3’-OH group, the sequencing primer is ready for the next sequencing cycle (step 5).
of the unextended sequencing primers will be capped by excess ddNTPs and DNA polymerase to avoid interference with the next sequencing cycle (Step 3). The fluorescent dye will then be chemically cleaved and the 3’ capping moiety will be simultaneously removed to regenerate the 3’-OH group (Step 4). The sequencing primer at this stage is ready for the next cycle of reaction to identify the next base of the DNA template (Step 5).

At present, three platforms based on the concept of SBS are in reasonably widespread use for massively parallel DNA sequencing: the Roche/454 FLX Pyrosequencer, the Illumina/Solexa Genome Analyzer, and the Applied Biosystems SOLiD™ System. The method used with the Roche/454 Pyrosequencer and the Applied Biosystems SOLiD sequencer to amplify single-stranded DNA copies from a fragment library is based on emulsion PCR,[36] and bridge PCR amplification has been adopted for the Illumina Genome Analyzer.[40] Recently, two other platforms for single molecule DNA sequencing by synthesis have been developed: for the Helicos Heliscope™ sequencer, the single-molecule DNA sample is directly immobilized on a solid surface, whereas the polymerase rather than the DNA library is captured on the surface in Pacific Biosciences SMRT system. With sequencing read length from 14 to 30 bases in the next generation DNA sequencing systems, massively parallel digital gene expression analogous to a high-throughput SAGE[41] approach has been reported reaching single copy transcript sensitivity,[42] and CHIP-seq[43-45] based on sequencing tags of around 25 bases has led to many new discoveries in genome function and regulation. With wide applications in genome biology and biomedical research, the sequencing by synthesis approach has great potential to revolutionize the practice of medicine.
The first part of this thesis describes our efforts to explore the novel sequencing chemistry of the SBS approach. Chapter 2 presents the development of a hybrid sequencing approach by integrating the advantages of Sanger sequencing and the SBS method. Additionally, a primer-reset approach allows the further extension of sequencing read-length, which is described in Chapter 3.
REFERENCES


Chapter 2: Four-color DNA Sequencing with 3’-\(O\)-Modified Nucleotide Reversible Terminators and Chemically Cleavable Fluorescent Dideoxynucleotides

2.1 INTRODUCTION

DNA sequencing is driving genomics research and discovery. The completion of the Human Genome Project has set the stage for screening genetic mutations to identify disease genes on a genome-wide scale.\(^1\) The development of accurate, cost-effective and high-throughput DNA sequencing technologies are required to explore the functions of the complex genome for applications in clinical medicine and healthcare. Decreased cost of sequencing is critical to the comparative genomics efforts including the ultimate goals of personalized medicine based on genetic and genomic information. Accuracy, speed, and size of the instrument are among the vital considerations for the development of new DNA analysis methods that can be implemented directly in the hospital and clinical settings, such as forensics and pathogen detection.

The Sanger dideoxy chain-termination method\(^2\) has been the gold standard in genome research for over three decades. Despite Sanger sequencing’s success, the electrophoresis-based sequencing technologies have some shortcomings due to the difficulty in achieving high throughput and the complexity involved in the automation. To overcome the limitations of the Sanger sequencing technology, a variety of new methods have been investigated. Such approaches include sequencing by hybridization,\(^3\) mass spectrometry sequencing,\(^4-6\) nanopore-based sequencing of single-stranded DNA,\(^7\) sequencing by ligation,\(^8\) and single molecule DNA sequencing.\(^9,10\) Among
these novel approaches for DNA sequencing, the sequencing by synthesis (SBS) approach has emerged as a viable candidate for a massively parallel high throughput sequencing platform. SBS takes advantage of the polymerase reaction, which is the key process for DNA replication inside cells. The basic concept of SBS is to use DNA polymerase to extend a primer that is hybridized to a template by a single nucleotide, determine its identity, and then proceed to the next nucleotide, eventually reading out the entire DNA sequence serially. In contrast to Sanger sequencing, in which fluorescently labeled DNA fragments of different sizes are all generated in a single reaction and then separated and detected, SBS approaches have an advantage in that individual bases are detected simultaneously without the need for separation. Thus, SBS can easily scale-up to beyond the potential of Sanger dideoxy-sequencing techniques. Currently array scanners already exist that can easily detect over 100,000 sample spots arrayed on a glass surface.\textsuperscript{[11]} Advanced array scanners enable fast screening of large areas with high resolution, allowing automated detection of hundreds of thousands and even millions of samples simultaneously.

Previously, the Ju laboratory has developed a general strategy to rationally design cleavable fluorescent nucleotide reversible terminators (NRTs) for four-color DNA sequencing by synthesis.\textsuperscript{[12-16]} In this approach, four nucleotides (A, C, G and T) are modified as NRTs by attaching a cleavable fluorophore to a specific location on the base and capping the 3’-OH with a small chemically reversible moiety in such a way that they are still recognized as substrates by DNA polymerase. DNA templates consisting of homopolymer regions were accurately sequenced by this approach.\textsuperscript{[16]} A recently developed fluorescent DNA sequencing-by-synthesis system based on a similar design of
the cleavable fluorescent NRTs has already found wide applications in genome biology.\textsuperscript{[17-19]} In addition, our group has used 3'-O-modified NRTs to solve the homopolymer sequencing problem in conventional pyrosequencing.\textsuperscript{[20]}

### 2.2 EXPERIMENTAL RATIONALE AND OVERVIEW

This chapter describes an alternative sequencing method that is a hybrid between the Sanger dideoxy chain terminating reaction and SBS, and discusses the advantages that come with this hybrid sequencing approach. The fundamental difference between the two methods is that the Sanger approach produces every possible complementary DNA extension fragment for a given DNA template and obtains the sequence after the separation of these fragments by detecting the fluorescent terminated base, while SBS relies on identification of each base as the DNA strand is extended by cleavable fluorescent NRTs that temporarily pause the DNA synthesis for sequence determination. The limiting factor for increasing sequencing throughput in the Sanger method, as mentioned previously, is in the DNA separation using electrophoresis and finite parallelization of the capillaries.

Challenges using SBS with cleavable fluorescent NRTs involve the further improvement of the DNA polymerase that efficiently recognizes the modified nucleotides. In addition, the first generation of the cleavable fluorescent NRTs synthesize the DNA strand with an amino linker modification on the base during SBS, which might interfere with the activity of the polymerase for subsequent chain elongation. The advantage in the Sanger based method is clearly in the dideoxy chain fragment producing reaction. Once the DNA strand is terminated by incorporation of a fluorescent dideoxynucleotide, it is no
longer involved in further DNA extension reactions. Therefore, the DNA polymerase extension reaction occurs with only natural nucleotides with high efficiency leading to read-length of over 700 base pairs. The most attractive features in the SBS sequencing approach are the massively parallel readout capability using a high-density DNA chip and simplified sample preparation method without the need to use electrophoresis to separate the DNA products. We have explored the integration of the advantageous features of the two methods to develop a hybrid DNA sequencing approach. In this approach, four nucleotides, modified as reversible terminators by capping the 3’-OH with a small reversible moiety so that they are still recognized as substrates by DNA polymerase, are used in combination with a small percentage of four cleavable fluorescent dideoxynucleotides to perform SBS. DNA sequences are determined by the unique fluorescent emission of each fluorophore of the DNA products terminated by ddNTPs. Upon removing the 3’-OH capping group from the DNA products generated by incorporating the 3’-O-modified dNTPs and the fluorophore from the DNA products terminated with the ddNTPs, the polymerase reaction reinitiates to continue the sequence determination (Fig. 2.1).

Using an azidomethyl group as a chemically reversible capping moiety in the 3’-O-modified dNTPs, and an azido-based cleavable linker to attach the fluorophores to ddNTPs, we synthesized 3’-O-N3-dNTPs and ddNTP-N3-fluorophores for the hybrid sequencing approach. The azidomethyl capping moiety on the 3’-OH group and the cleavable fluorophore on the DNA extension products are efficiently removed after fluorescence detection for sequence determination using a chemical method that is compatible with DNA structure and stability. Various DNA templates, including those
**Figure 2.1.** The hybrid DNA sequencing approach between the Sanger dideoxy chain terminating reaction and sequencing by synthesis. In this approach, four nucleotides (3'-O-R1-dNTPs) modified as reversible terminators by capping the 3'-OH with a small reversible moiety R1 so that they are still recognized by DNA polymerase as substrates, are used in combination with a small percentage of four cleavable fluorescent dideoxynucleotides (ddNTP-R2-fluorophores) to perform SBS. DNA sequences are determined by the unique fluorescence emission of each fluorophore on the DNA products terminated by ddNTPs. Upon removing the 3'-OH capping group R1 from the DNA products generated by incorporating the 3'-O-R1-dNTPs, and the cleavage of the R2 linker to remove the fluorophore from the DNA products terminated with the ddNTPs, the polymerase reaction reinitiates to continue the sequence determination.
with homopolymer regions, were accurately sequenced with read-length of over 30 bases using this hybrid SBS method.

2.3 RESULTS AND DISCUSSION

2.3.1 Design and synthesis of 3’-O-modified NRTs

A critical requirement for using SBS methods to unambiguously sequence DNA is a suitable chemical moiety to cap the 3’-OH of the nucleotide such that it temporarily terminates the polymerase reaction to allow the identification of the incorporated nucleotide. A stepwise addition of separate nucleotides with a free 3’-OH group has inherent difficulties in detecting sequences in homopolymeric regions. Capping the 3’-OH group of the nucleotides with a reversible moiety allows for the addition of all four nucleotides simultaneously in performing SBS, thereby increasing accuracy and reducing the number of cycles needed. However, it is essential that the capping group be efficiently removed from the DNA extension products to regenerate the 3’-OH group for continuous polymerase reaction.

Our previous research efforts have firmly established the molecular level strategy to rationally modify the nucleotides by capping the 3’-OH with a small chemically reversible moiety for SBS. Building on our successful 3’-O-modification strategy, we synthesized and evaluated four 3’-O-azidomethyl-modified NRTs (3’-O-N3-dNTPs) (Fig. 2.2) for use in the hybrid SBS approach. The 3’-O-modified NRTs containing an azidomethyl group to cap the 3’-OH on the sugar ring were synthesized based on a method similar to that reported by Zavgorodny et al. The synthesis of 3’-O-azidomethyl-dATP started from commercially available nucleoside 1, which has the
Figure 2.2. Structures of the nucleotide reversible terminators, 3’-O-N₃-dATP, 3’-O-N₃-dCTP, 3’-O-N₃-dGTP, 3’-O-N₃-dTTP.
5'-hydroxyl group and the 6-amino group protected (Fig. 2.3). With a modified Pummerer’s rearrangement, nucleoside 1 was converted to 3’-O-methylthiomethyl nucleoside 2 by treating with a mixture of acetic acid, acetic anhydride and DMSO for 48 hours. Activation of nucleoside 2 with sulfurylchloride yielded 3’-O-chloromethyl intermediate, which reacted smoothly in situ with sodium azide to afford 3’-O-azidomethyl nucleotide. After desilylation, nucleoside 3 was obtained in a one-pot reaction. Then phosphorylation of the 5’-hydroxyl group in nucleoside 3 afforded the target 3’-O-azidomethyl-dATP 4. Similarly, 3’-O-azidomethyl-dTTP 8 (Fig. 2.4) and 3’-O-azidomethyl-dATP 12 were synthesized (Fig. 2.5).

The synthesis of 3’-O-azidomethyl-dGTP takes advantage of a different protection strategy before introducing azidomethyl to the 3’-hydroxyl group (Fig. 2.6). With the 5’-hydroxyl group and the 2-amino group protected, nucleoside 13 was converted to 3’-O-methylthiomethyl nucleoside 14 by a modified Pummerer’s rearrangement. Then the 6-hydroxyl group in 14 was protected with diphenyl carbamoyl group by reaction with diphenyl carbamoyl chloride and diisopropylethylamine (DIEA), yielding nucleoside 15. After the conversion of the 3’-O-methylthiomethyl group in 15 to a 3’-O-azidomethyl group and subsequent desilylation, precursor 16 was obtained. Phosphorylation of the 5’-hydroxyl group in 16 afforded the target 3’-O-azidomethyl-dGTP 17.

The 3’-O-azidomethyl group on the DNA extension product generated by incorporating each of the NRTs is efficiently removed by the Staudinger reaction using aqueous Tris(2-carboxy-ethyl) phosphine (TCEP) solution[23] followed by hydrolysis to yield a free 3’-OH group for elongating the DNA chain in subsequent cycles of the
Figure 2.3. Synthesis of 3'-O-azidomethyl-dATP.

Figure 2.4. Synthesis of 3'-O-azidomethyl-dTTP.
**Figure 2.5.** Synthesis of 3'-O-azidomethyl-dCTP.

**Figure 2.6.** Synthesis of 3'-O-azidomethyl-dGTP.
hybrid SBS (Fig. 2.7).

2.3.2 Continuous polymerase extension by using 3’-O-modified NRTs and characterization by MALDI-TOF mass spectrometry

In order to verify that 3’-O-modified NRTs (3’-O-N_3-dNTPs) are incorporated accurately in a base specific manner in the polymerase reaction, four continuous DNA extension and cleavage reactions were carried out in solution using 3’-O-N_3-dNTPs as substrates. This allowed the isolation of the DNA product at each step for detailed molecular structure characterization as shown in Fig. 2.8. The first extension product 5’-primer-C-N_3-3’ (1) was desalted and analyzed using MALDI-TOF MS (Fig. 2.8A). This product was then incubated in aqueous TCEP solution to remove the azidomethyl moiety to yield the cleavage product (2) with a free 3’-OH group, which was also analyzed using MALDI-TOF MS (Fig. 2.8B). As can be seen from Fig. 2.8A, the MALDI-TOF MS spectrum consists of a distinct peak corresponding to the DNA extension product 5’-primer-C-N_3-3’ (1) (m/z 8,310), which confirms that the NRT is incorporated base specifically by DNA polymerase into a growing DNA strand. Fig. 2.8B shows the cleavage result on the DNA extension product. The extended DNA mass peak at m/z 8,310 completely disappeared while the peak corresponding to the cleavage product 5’-primer-C-3’ (2) appears as the sole dominant peak at m/z 8,255, which establishes that TCEP incubation completely cleaves the 3’-O-azidomethyl group with high efficiency. The next extension reaction was carried out using this cleaved product, which now has a free 3’-OH group, as a primer to yield a second extension product, 5’-primer-CG-N_3-3’ (3) (m/z 8,639, Fig. 2.8C). In a similar way as described above, the
Figure 2.7. Staudinger reaction with TCEP to regenerate the 3'-OH group of the DNA extension product.
Figure 2.8. The polymerase extension scheme (left) and MALDI-TOF MS spectra of the four consecutive extension products and their cleavage products (right) using the nucleotide reversible terminators (3'-O-N$_3$-dNTP). Primer extended with 3'-O-N$_3$-dCTP (1) (A), and its cleavage product 2 (B); Product 2 extended with 3'-O-N$_3$-dGTP (3) (C), and its cleavage product 4 (D); Product 4 extended with 3'-O-N$_3$-dATP (5) (E), and its cleavage product 6 (F); Product 6 extended with 3'-O-N$_3$-dTTP (7) (G), and its cleavage product 8 (H). After brief incubation in a TCEP aqueous solution the azidomethyl moiety capping the 3'-OH group of the DNA extension products is completely removed to continue the polymerase reaction.
extension product (3) was cleaved to generate product (4) for further MS analysis yielding a single peak at $m/z$ 8,584 (Fig. 2.8D). The third extension reaction to yield 5’-primer-CGA-N$_3$-3’ (5) ($m/z$ 8,952, Fig. 2.8E), the fourth extension to yield 5’-primer-CGAT-N$_3$-3’ (7) ($m/z$ 9,256, Fig. 2.8G) and their cleavage to yield products (6) ($m/z$ 8,897, Fig. 2.8F) and (8) ($m/z$ 9,201, Fig. 2.8H) were similarly carried out and analyzed by MALDI-TOF MS. These results demonstrate that all four 3’-O-azidomethyl modified NRTs are successfully synthesized and efficiently incorporated base-specifically into the growing DNA strand in a continuous polymerase reaction as reversible terminators and the 3’-OH capping group on the DNA extension products is quantitatively cleaved by TCEP.

2.3.3 Design and synthesis of cleavable fluorescent dideoxynucleotide terminators

To demonstrate the feasibility of carrying out the hybrid SBS on a DNA chip, we designed and synthesized two sets of cleavable fluorescent dideoxynucleotide terminators, ddNTP-N$_3$ (version I)-fluorophores and ddNTP-N$_3$ (version II)-fluorophores (Fig. 2.9). The ddNTP-N$_3$-fluorophores will be used in combination with the four NRTs (Fig. 2.2) to perform the hybrid SBS. Modified DNA polymerases have been shown to be highly tolerant to nucleotide modifications with bulky groups at the 5-position of pyrimidines (C and U) and the 7-position of purines (A and G). Thus, we attached each unique fluorophore to the 5 position of C/U and the 7 position of A/G through a cleavable linker. The cleavable linker also features an azido modified moiety as a trigger for cleavage. Thus treatment with TCEP allows the generation of 3’-OH and the cleavage of fluorophores on DNA extension products simultaneously. Two critical properties of the
Figure 2.9. Structures of cleavable fluorescent dideoxynucleotide terminators (A) ddNTP-N₂ (version I)-fluorophores and (B) ddNTP-N₂ (version II)-fluorophores, with the 4 fluorophores having distinct fluorescent emissions: Bodipy-FL-510 ($\lambda_{\text{abs (max)}} = 502$ nm; $\lambda_{\text{em (max)}} = 510$ nm), R6G ($\lambda_{\text{abs (max)}} = 525$ nm; $\lambda_{\text{em (max)}} = 550$ nm), ROX ($\lambda_{\text{abs (max)}} = 585$ nm; $\lambda_{\text{em (max)}} = 602$ nm), Cy5 ($\lambda_{\text{abs (max)}} = 649$ nm; $\lambda_{\text{em (max)}} = 670$ nm).
cleavable linker determines the sequencing efficiency: (1) the stability of the linker under DNA extension conditions and (2) the cleavage efficiency during treatment with TCEP. The ideal linker should possess both high stability in the polymerase reaction and high cleavage efficiency once treated with TCEP. To fulfill these requirements, we designed and synthesized two sets of cleavable linkers and their corresponding cleavable fluorescent dideoxynucleotide terminators (Fig. 2.9).

The cleavable linker (version I) was synthesized according to reported approaches\[^{24}\] (Fig. 2.10). Treatment with 2-bromomethyl-1,3-dioxolane, potassium carbonate and sodium iodide converted commercially available ethyl 3-hydroxybenzoate 18 to ether 19. The ring opening reaction was then performed in the presence of a Lewis acid (tin tetrachloride) and azidotrimethylsilane to afford compound 20. Hydrolysis of ethyl benzoate 20 by treatment with sodium hydroxide yielded 21, which was converted to compound 22 by reacting with ethyl bromoacetate. Activation of the benzoic acid 22 with N, N’-disuccinimidyl carbonate (DSC) and 4-dimethylaminopyridine (DMAP) yielded a N-hydroxysuccinimidyl (NHS) ester intermediate, which was converted to 23 by reacting with N-(2-amino-ethyl)-2,2,2-trifluoroacetamide. Then hydrolysis of 23 produced the cleavable linker (version I) 24.

The synthesis of cleavable fluorescent dideoxynucleotide terminators (ddNTP-N\(_3\) (version I)-fluorophores) was completed by attaching the unique fluorophores to the bases of dideoxynucleotides through the cleavable linker (version I). As shown in Fig. 2.11, coupling the cleavable linker (version I) 24 with Bodipy-FL-510 NHS ester 25 in a solution of aqueous sodium bicarbonate and DMF yielded N\(_3\) (version I)-Bodipy-FL-510 26. Converted to its corresponding NHS ester 27 by treatment with DSC and DMAP,
Figure 2.10. Synthesis of the cleavable linker (version I).
Figure 2.11. Synthesis of ddCTP-N\textsubscript{3} (version I)-Bodipy-FL-510.
compound 26 was coupled with ddCTP-NH₂ to afford the target ddCTP-N₃ (version I)-Bodipy-FL-510. We synthesized ddUTP-N₃ (version I)-R6G (Fig. 2.12), ddATP-N₃ (version I)-ROX (Fig. 2.13) and ddGTP-N₃ (version I)-Cy5 (Fig. 2.14) using a similar approach.

To further demonstrate the feasibility of carrying out the hybrid SBS on a DNA chip, we designed and synthesized another azido-based cleavable linker and its corresponding cleavable fluorescent dideoxynucleotide terminators. The synthesis of the cleavable linker (version II) starts from the commercially available 2-(aminomethyl)-1,3-dioxolane 41 (Fig. 2.15). Protection of the amino group in 41 with a fluorenylmethyloxycarbonyl (Fmoc) group yielded carbamate 43. Treatment with azidotrimethylsilane and tin tetrachloride converted 43 to compound 44. After deprotection of the amino group in 44 with piperidine, the cleavable linker (version II) 45 was obtained.

The synthetic approach to ddNTP-N₃ (version II)-fluorophores is similar to that of ddNTP-N₃ (version I)-fluorophores. Bodipy-FL-510 NHS ester 46 was coupled to the cleavable linker (version II) 45 in a solution of aqueous sodium bicarbonate and DMF, affording N₃ (version II)-Bodipy-FL-510 47 (Fig. 2.16). Treatment with DSC and triethylamine converted 47 to its corresponding NHS ester 48. Then coupling ddCTP-NH₂ with N₃ (version II)-Bodipy-FL-510 NHS ester 48 produced the target ddCTP-N₃ (version II)-Bodipy-FL-510. We synthesized ddUTP-N₃ (version II)-R6G (Fig. 2.17), ddATP-N₃ (version II)-ROX (Fig. 2.18) and ddGTP-N₃ (version II)-Cy5 (Fig. 2.19) using a similar approach.
Figure 2.12. Synthesis of ddUTP-N$_3$ (version I)-R6G.
Figure 2.13. Synthesis of ddATP-N₃ (version I)-ROX.
Figure 2.14. Synthesis of ddGTP-N₃ (version I)-Cy5.
Figure 2.15. Synthesis of the cleavable linker (version II).

Figure 2.16. Synthesis of ddCTP-N$_3$ (version II)-Bodipy-FL-510.
Figure 2.17. Synthesis of ddUTP-\(N_3\) (version II)-R6G.
Figure 2.18. Synthesis of ddATP-N₃ (version II)-ROX.
Figure 2.19. Synthesis of ddGTP-N$_3$ (version II)-Cy5.
Both sets of ddNTP-N$_3$-fluorophores were found to be efficiently and specifically incorporated into the growing DNA strand to permanently terminate the polymerase reaction for sequence determination. The fluorophore on a DNA extension product, which is generated by incorporation of the cleavable fluorescent dideoxynucleotide analogues, is removed rapidly and quantitatively by TCEP from the DNA extension product in aqueous solution, with a mechanism similar to the removal of the 3’-O-azidomehtyl group (Fig. 2.20).

2.3.4 Polymerase extension by using cleavable fluorescent dideoxynucleotides and characterization by MALDI-TOF mass spectrometry

We have synthesized two sets of cleavable fluorescent dideoxynucleotide terminators, ddNTP-N$_3$ (version I)-fluorophores and ddNTP-N$_3$ (version II)-fluorophores (Fig. 2.9). Here we demonstrate that these fluorescent dideoxynucleotides are recognized by polymerase as excellent substrates in the polymerase reaction and the fluorophore from the DNA extension product is cleaved rapidly and completely.

In order to verify that ddNTP-N$_3$ (version I)-fluorophores are incorporated accurately in a base-specific manner in a polymerase reaction, single base extension reactions with four different self-priming DNA templates whose next complementary base was either A, C, G, or T were carried out in solution. After the reaction, the 4 different primer extension products were analyzed by MALDI-TOF MS as shown in Fig. 2.21. Single clear mass peaks at 9,180, 8,915, 9,261 and 9,082 (m/z) corresponding to each primer extension product with no left over starting material were produced by using ddNTP-N$_3$ (version I)-fluorophores (Fig. 2.21 A, C, E, G). For the cleavage experiment,
Figure 2.20. Staudinger reaction with TCEP to cleave the N$_3$-fluorophore from the dideoxynucleotide.
Figure 2.21. A polymerase reaction scheme (top) to yield DNA extension products by incorporating each of the four ddNTP-N$_3$ (version I)-fluorophores and the subsequent cleavage reaction to remove the fluorophores from the DNA extension products. MALDI-TOF MS spectra (bottom) showing efficient base specific incorporation of the ddNTP-N$_3$-fluorophores and the subsequent cleavage of the fluorophores from the DNA extension products: (A) primer extended with ddATP-N$_3$(version I)-ROX (1) (peak at 9,180 m/z), (B) its cleavage product 2 (8,417 m/z); (C) primer extended with ddCTP-N$_3$(version I)-Bodipy-FL-510 (3) (peak at 8,915 m/z), (D) its cleavage product 4 (8,394 m/z); (E) primer extended with ddGTP-N$_3$(version I)-Cy5 (5) (peak at 9,261 m/z), (F) its cleavage product 6 (8,432 m/z); (G) primer extended with ddUTP-N$_3$(version I)-R6G (7) (peak at 9,082 m/z) and (H) its cleavage product 8 (8,395 m/z).
the linker tethering the fluorophore to the dideoxynucleotide was cleaved by incubating the DNA extension products in an aqueous TCEP solution at 65° C for 15 minutes. Fig. 2.21 (B, D, F, H) shows the cleavage results for the DNA products extended with ddNTP-N₃ (version I)-fluorophores. The mass peaks at 9,180, 8,915, 9,261 and 9,082 (m/z) corresponding to the DNA extension products have completely disappeared while single peaks corresponding to the cleavage products appear at 8,417, 8,394, 8,432 and 8,395 (m/z) respectively. These results demonstrate that ddNTP-N₃ (version I)-fluorophores are successfully synthesized and efficiently terminated the DNA synthesis in a polymerase reaction and that the fluorophores are quantitatively cleaved by TCEP. Thus, these dideoxynucleotide analogues meet the key requirements necessary for performing the hybrid SBS in combination with the NRTs.

To verify that ddNTP-N₃ (version II)-fluorophores are excellent substrates for the DNA polymerase, and the fluorophores tethered to the dideoxynucleotides can be cleaved efficiently, we performed similar polymerase extension and cleavage reactions (Fig. 2.22). Single clear mass peaks at 9,003, 8,738, 9,140 and 8,905 (m/z) corresponding to each primer extension product with no left over starting material were produced by using ddNTP-N₃ (version II)-fluorophores (Fig. 2.22 A, C, E, G). For the cleavage experiment, the linker tethering the fluorophore to the dideoxynucleotide was cleaved by incubating the DNA extension products in an aqueous TCEP solution at 65° C for 10 minutes. Fig. 2.22 (B, D, F, H) shows the cleavage results for the DNA products extended with ddNTP-N₃ (version II)-fluorophores. The mass peaks at 9,003, 8,738, 9,140 and 8,905 (m/z) corresponding to the DNA extension products have completely disappeared while single peaks corresponding to the cleavage products appear at 8,403, 8,380, 8,419 and
Figure 2.22. A polymerase reaction scheme (top) to yield DNA extension products by incorporating each of the four ddNTP-N₃ (version II)-fluorophores and the subsequent cleavage reaction to remove the fluorophores from the DNA extension products. MALDI-TOF MS spectra (bottom) showing efficient base specific incorporation of the ddNTP-N₃-fluorophores and the subsequent cleavage of the fluorophores from the DNA extension products: (A) primer extended with ddATP-N₃ (version II)-ROX (1) (peak at 9,003 m/z), (B) its cleavage product 2 (8,403 m/z); (C) primer extended with ddCTP-N₃ (version II)-Bodipy-FL-510 (3) (peak at 8,738 m/z), (D) its cleavage product 4 (8,380 m/z); (E) primer extended with ddGTP-N₃ (version II)-Cy5 (5) (peak at 9,140 m/z), (F) its cleavage product 6 (8,419 m/z); (G) primer extended with ddUTP-N₃ (version II)-R6G (7) (peak at 8,905 m/z) and (H) its cleavage product 8 (8,381 m/z).
8,381 (m/z) respectively. These results demonstrate that ddNTP-N$_3$ (version II)-fluorophores were successfully synthesized and efficiently terminated the DNA synthesis in a polymerase reaction and that the fluorophores are quantitatively cleaved by TCEP. Nonetheless, the cleavable linker (version II) is not as stable as the cleavable linker (version I) (data not shown). Thus we only explore the feasibility of using ddNTP-N$_3$ (version I)-fluorophores to perform the hybrid SBS in combination with the NRTs.

2.3.5 Four-color DNA sequencing on a chip by using cleavable fluorescent dideoxynucleotides and 3’-O-modified NRTs in the hybrid SBS approach

In our four-color hybrid SBS approach, the identity of the incorporated nucleotide is determined by the unique fluorescent emission from the four fluorescent dideoxynucleotide terminators, while the role of the 3’-O-modified NRTs is to further extend the DNA strand to continue the determination of the DNA sequence. Therefore, the ratio between the amount of ddNTP-N$_3$ (version I)-fluorophores and 3’-O-N$_3$-dNTPs during the polymerase reaction determines how much of the ddNTP-N$_3$ (version I)-fluorophores incorporate and thus the corresponding fluorescent emission strength. With a finite amount of immobilized DNA template on a solid surface, initially the majority of the priming strands should be extended with 3’-O-N$_3$-dNTPs, while a relatively smaller amount should be extended with ddNTP-N$_3$-fluorophores to produce fluorescent signals that are above the fluorescent detection system’s sensitivity threshold for sequence determination. As the sequencing cycles continue, the amount of the ddNTP-N$_3$ (version I)-fluorophores need to be gradually increased to maintain the
fluorescence emission strength for detection. Following these guidelines, we performed the hybrid SBS on a chip-immobilized DNA template using the 3'-O-N_3-dNTP/ddNTP-N_3 (version I)-fluorophore combination and the results are shown in Fig. 2.23. The general four-color sequencing reaction scheme on a chip is shown in Fig. 2.23A. The *de novo* sequencing reaction on the chip was initiated by extending the self-priming DNA using a solution containing the combination of the four 3'-O-N_3-dNTPs and the four ddNTP-N_3 (version I)-fluorophores, and 9°N DNA polymerase.

In order to negate any lagging fluorescent signal that is caused by a previously unextended priming strand, a synchronization step was added to reduce the amount of unextended priming strands after the initial extension reaction. A synchronization reaction mixture consisting of just the four 3'-O-N_3-dNTPs in relatively high concentration was used along with the 9°N DNA polymerase to extend any remaining priming strands that retain a free 3'-OH group to synchronize the incorporation. This extension method where the combination of 3'-O-N_3-dNTPs/ddNTP-N_3 (version I)-fluorophores are used, will not have a negative impact on the enzymatic incorporation of the next nucleotide analogue, because after cleavage to remove the 3'-OH capping group, the DNA product extended by 3'-O-N_3-dNTPs carry no modification groups. Previous designs of cleavable fluorescent NRTs left small traces of modification (propargyl amine linker) after the cleavage of the fluorophore on the base of the nucleotide.\cite{16} Successive addition of these NRTs into a growing DNA strand during SBS leads to a newly synthesized DNA chain with, at each base site, a small leftover linker fragment. This may interfere with the ability of the enzyme to efficiently incorporate the
Figure 2.23. (A) A hybrid SBS scheme for 4-color sequencing on a chip using the nucleotide reversible terminators (3'-O-N₃-dNTPs) and cleavable fluorescent dideoxynucleotide terminators (ddNTP-N₃ (version I)-fluorophores). (B) The 4-color fluorescence images for each step of SBS: (1) incorporation of 3'-O-N₃-dCTP and ddCTP-N₃ (version I)-Bodipy-FL-510; (2) cleavage of N₃ (version I)-Bodipy-FL-510 and 3'-CH₂N₃ group; (3) incorporation of 3'-O-N₃-dATP and ddATP-N₃ (version I)-Rox; (4) cleavage of N₃ (version I)-ROX and 3'-CH₂N₃ group; (5) incorporation of 3'-O-N₃-dTTP and ddTTP-N₃ (version I)-R6G; (6) cleavage of N₃ (version I)-R6G and 3'-CH₂N₃ group; (7) incorporation of 3'-O-N₃-dGTP and ddGTP-N₃ (version I)-Cy5; (8) cleavage of N₃ (version I)-Cy5 and 3'-CH₂N₃ group; images 9–63 are similarly produced. (C) A plot (4-color sequencing data) of raw fluorescence emission intensity obtained using 3'-O-N₃-dNTPs and ddNTP-N₃ (version I)-fluorophores at the four designated emission wavelengths.
next incoming nucleotide, which will undoubtedly lead to loss of synchrony and reduction in the maximal read-length. This challenge might potentially be overcome with further research efforts to reengineer new DNA polymerases that efficiently recognize and accept the modified DNA strand. With the hybrid SBS approach, DNA products extended by ddNTP-N\textsubscript{3} (version I)-fluorophores, after fluorescence detection for sequence determination and cleavage, are no longer involved in the subsequent polymerase reaction cycles because they are permanent terminators. Therefore, further polymerase reaction only occurs on a DNA strand that incorporates the 3’-O-N\textsubscript{3}-dNTPs, which subsequently revert to natural nucleotides upon cleavage of the 3’-OH capping group, and should have no deleterious effect on the ability of polymerase to bind and incorporate subsequent nucleotides for growing the DNA chains.

The four-color images from a fluorescence scanner for each step of the hybrid SBS on a chip is shown in Fig. 2.23B. The first extension of the primer by the complementary fluorescent dideoxynucleotide, ddCTP-N\textsubscript{3} (version I)-Bodipy-FL-510 was confirmed by observing a blue signal (the emission from Bodipy-FL-510) [Fig. 2.23B (1)]. After detection of the fluorescent signal, the surface was immersed in a TCEP solution to cleave both the fluorophore from the DNA product extended with ddNTP-N\textsubscript{3} (version I)-fluorophores and the 3’-O-azidomethyl group from the DNA product extended with 3’-O-N\textsubscript{3}-dNTPs. The surface of the chip was then washed, and a negligible residual fluorescent signal was detected, confirming cleavage of the fluorophore [Fig. 2.23B (2)]. This was followed by another extension reaction using the 3’-O-N\textsubscript{3}-dNTP/ddNTP-N\textsubscript{3} (version I)-fluorophore solution to incorporate the next nucleotide complementary to the subsequent base on the template. The entire process of
incorporation, synchronization, detection and cleavage was performed multiple times to identify 32 successive bases in the DNA template. The plot of the fluorescence intensity vs. the progress of sequencing extension (raw 4-color sequencing data) is shown in Fig. 2.23C. The DNA sequences are unambiguously identified with no errors from the 4-color raw fluorescence data without any processing. Similar four-color sequencing data were obtained for a variety of DNA templates (Fig. 2.24).

2.4 MATERIALS AND METHODS

General Information. All solvents and reagents were reagent grade, purchased commercially and used without further purification unless specified. All chemicals were purchased from Sigma-Aldrich unless otherwise indicated. Oligonucleotides used as primers or templates were synthesized on an Expedite nucleic acid synthesizer (Applied BioSystems) or purchased from Midland. \(^1\)H NMR spectra were recorded on a Bruker DPX-400 (400 MHz) spectrometer and reported in parts per million (ppm) from a CDCl\(_3\), CD\(_3\)OD or DMSO-d\(_6\) internal standard (7.26, 3.31 or 2.50 ppm, respectively). Data were reported as follows: (s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, dd = doublet of doublets, coupling constant \((J)\) in Hz, integration). Proton decoupled \(^{13}\)C NMR spectra were recorded on a Bruker DPX-400 (100 MHz) spectrometer and reported in ppm from a CDCl\(_3\), CD\(_3\)OD, or DMSO-d\(_6\) internal standard (77.0, 49.0, or 39.5 ppm, respectively). Proton decoupled \(^{31}\)P NMR spectra were recorded on a Bruker DPX-300 (121.4 MHz) spectrometer, and reported in ppm from 85% H\(_3\)PO\(_4\) external standard. High-resolution mass spectra (HRMS) were obtained on a JEOL JMS HX 110A mass spectrometer. Mass measurement of DNA was performed on a Voyager DE
Figure 2.24. A plot (4-color sequencing data) of raw fluorescence emission intensity obtained using 3'-O-N$_3$-dNTPs and ddNTP-N$_3$ (version I)-fluorophores at the four designated emission wavelengths of the four cleavable fluorescent dideoxynucleotides.
MALDI-TOF mass spectrometer (Applied Biosystems). 9°N polymerase (exo-)A485L/Y409V was obtained from New England Biolabs. Starting materials 5′-O-(tert-butyldimethylsilyl)thymidine, N^4-benzoyl-5′-O-(tert-butyldimethylsilyl)-2′-deoxycytidine, N^6-Benzoyl-5′-O-(tert-butyldimethylsilyl)-2′-deoxy-adenosine, and N^2-isobutyryl-5′-O-(tert-butyldimethylsilyl)-2′-deoxy-guanosine were purchased from CNH Technologies, Inc (Woburn, MA). 2-(Aminomethyl)-1,3-dioxolane was purchased from TCI America (Portland, OR). Aminopropargyl-ddNTPs (NH_2-ddNTPs) were purchased from PerkinElmer Life and Analytical Sciences (Waltham, MA). Phosphoramidites and columns for oligonucleotide synthesis were purchased from Glen Research (Sterling, VA). The 3′-O-modified nucleotides and fluorescently labeled dideoxynucleotides were purified with reverse-phase HPLC on a 150×4.6 mm C18 column (Supelco), mobile phase: A, 8.6 mM Et_3N / 100 mM 1,1,1,3,3,3-hexafluoro-2-propanol in water (pH 8.1); B, methanol. Elution was performed from 100% A isocratic over 10 minutes followed by a linear gradient of 0-50% B for 20 minutes and then 50% B isocratic over another 30 minutes.

2.4.1 Synthesis of 3′-O-azidomethyl-dATP

N^6-Benzoyl-3′-O-(methylthiomethyl)-5′-O-(tert-butyldimethylsilyl)-2′-deoxyadenosine (Compound 2). To a stirred solution of N^6-Benzoyl-5′-O-(tert-butyldimethylsilyl)-2′-deoxyadenosine (3.0 g; 6.38 mmol) in DMSO (12 ml), acetic acid (5.5 ml) and acetic anhydride (17.6 ml) were added. The reaction mixture was stirred at room temperature for 48 hours. A saturated NaHCO_3 solution (100 ml) was added and the aqueous layer was extracted with CH_2Cl_2 (3×100 ml). The combined organic extract was washed with
saturated NaHCO₃ solution (100 ml) and dried over Na₂SO₄. After concentration, the residue was purified by flash column chromatography (hexane/ethyl acetate, 1:1 to 1:4) to afford 2 as a white powder (2.4 g; 71% yield). ¹H NMR (400 MHz, CDCl₃) δ 9.00 (s, 1H), 8.83 (s, 1H), 8.35 (s, 1H), 8.05 (d, J = 7.6 Hz, 2H), 7.62 (m, 1H), 7.55 (m, 2H), 7.51 (t, J = 7.2 Hz, 1H), 4.73 (m, 2H), 4.68 (m, 1H), 4.24 (m, 1H), 3.88 (dd, J = 11.2, 3.2 Hz, 1H), 2.66-2.74 (m, 2H), 2.35 (s, 3H), 0.94 (s, 9H), 0.13 (s, 6H); HRMS (Fab+) calcd for C₂₅H₃₆O₄N₅S: 530.2257, found: 530.2273.

N⁶-Benzoyl-3'-O-(azidomethyl)-2'-deoxyadenosine (Compound 3). To a stirred solution of 2 (400 mg; 0.76 mmol) in dry CH₂Cl₂ (7 ml) under nitrogen, cyclohexene (400 µl) and SO₂Cl₂ (155 µl; 1.91 mmol, redistilled) were added. The reaction mixture was stirred at 0°C for 2 hours. The solvent was first removed under reduced pressure and then under high vacuum for 10 minutes. The residue was dissolved in dry DMF (5 ml) and reacted with NaN₃ (400 mg; 6.6 mmol) at room temperature for 3 hours. The reaction mixture was dispersed in distilled water (50 ml) and extracted with CH₂Cl₂ (3×50 ml). The combined organic layer was dried over Na₂SO₄ and concentrated under reduced pressure. The residue was dissolved in MeOH (5 ml) and stirred with NH₄F (300 mg; 8.1 mmol) at room temperature for 24 hours. The solvent was removed under reduced pressure. The reaction mixture was concentrated under reduced pressure and partitioned between H₂O and CH₂Cl₂. The organic layer was separated and dried over Na₂SO₄. After concentration, the crude product was purified by flash column chromatography (ethyl acetate/methanol, 100:0 to 98:2) to afford 3 as a white powder (150 mg; 48% yield). ¹H NMR (400 MHz, CDCl₃) δ 8.84 (br s, 1H), 8.70 (br s, 1H), 8.08 (m, 1H), 7.54-7.76 (m, 5H), 6.47 (t, J = 5.6
Hz, 1H), 4.83 (m, 2H), 4.78 (m, 1H), 4.39 (m, 1H), 3.88-4.09 (m, 2H), 3.09 (m, 1H), 2.65 (m, 1H); HRMS (Fab+) calcd for C\textsubscript{18}H\textsubscript{19}O\textsubscript{4}N\textsubscript{8}[(M+H)\textsuperscript{+}]: 411.1529, found: 411.1530.

3'-O-Azidomethyl-dATP (Compound 4). Compound 3 (123 mg; 0.3 mmol) and proton sponge (75.8 mg; 0.35 mmol) were dried in a vacuum dessicator over P\textsubscript{2}O\textsubscript{5} overnight before dissolving in trimethyl phosphate (600 µl). Then freshly distilled POCl\textsubscript{3} (40 µl; 0.35 mmol) was added dropwise at 0°C and the mixture was stirred at 0°C for 2 hours. Subsequently, a well-vortexed mixture of tributylammonium pyrophosphate (552 mg) and tributylamine (0.55 ml; 2.31 mmol) in anhydrous DMF (2.33 ml) was added in one portion at room temperature and stirred for 30 minutes. Triethyl ammonium bicarbonate solution (TEAB) (0.1 M; pH 8.0; 15 ml) was then added and the mixture was stirred for 1 hour at room temperature. Then concentrated NH\textsubscript{4}OH (15 ml) was added and stirred overnight at room temperature. The resulting mixture was concentrated under vacuum and the residue was diluted with 5 ml of water. The crude mixture was then purified by anion exchange chromatography on DEAE-Sephadex A-25 at 4°C using a gradient of TEAB (pH 8.0; 0.1-1.0 M). The crude product was further purified by reverse-phase HPLC to afford 4. \textsuperscript{1}H NMR (400 MHz, D\textsubscript{2}O) δ 8.09 (s, 1H), 6.0 (m, 1H), 5.90-6.50 (m, 2H), 4.83 (br s, 2H), 4.15 (m, 1H), 3.66 (d, 2H), 3.19-3.40 (m, 3H), 3.09 (m, 1H), 2.60 (m, 1H); \textsuperscript{31}P NMR (121.4 MHz, D\textsubscript{2}O) δ -10.63 (d, J = 19 Hz, 1P), -11.19 (d, J = 22 Hz, 1P), -23.18 (t, J = 21 Hz, 1P).

2.4.2 Synthesis of 3'-O-azidomethyl-dTTP

3'-O-(Methylthiomethyl)-5'-O-(tert-butyldimethylsilyl)thymidine (Compound 6). To a
stirred solution of 5′-O-(tert-butyldimethylsilyl)thymidine (2.0 g; 5.6 mmol) in DMSO (10.5 ml), acetic acid (4.8 ml) and acetic anhydride (15.4 ml) were added. The reaction mixture was stirred at room temperature for 48 hours. A saturated NaHCO₃ solution (100 ml) was added and the aqueous layer was extracted with ethyl acetate (3×100 ml). The combined organic extract was washed with a saturated solution of NaHCO₃ and dried over Na₂SO₄. After concentration, the crude product was purified by flash column chromatography (hexane/ethyl acetate, 7:3 to 1:1) to afford 6 as a white powder (1.75 g; 75% yield). ¹H NMR (400 MHz, CDCl₃) δ 8.16 (s, 1H), 7.48 (s, 1H), 6.28 (m, 1H), 4.62 (m, 2H), 4.46 (m, 1H), 4.10 (m, 1H), 3.78-3.90 (m, 2H), 2.39 (m, 1H), 2.14 (s, 3H), 1.97 (m, 1H), 1.92 (s, 3H), 0.93 (s, 9H), 0.13 (s, 3H); HRMS (Fab+) cald for C₁₈H₃₃N₂O₅Si [(M+H)⁺]: 417.1879, found: 417.1890.

3′-O-(Azidomethyl)-thymidine (Compound 7). To a stirred solution of 6 (1.095 g; 2.6 mmol) in dry CH₂Cl₂ (10 ml) under nitrogen atmosphere, cyclohexene (1.33 ml) and SO₂Cl₂ (284 µl; 3.5 mmol, redistilled) were added. The reaction mixture was stirred at 0°C for 1.5 hours. The solvent was first removed under reduced pressure and then under high vacuum for 10 minutes. The residue was dissolved in dry DMF (5 ml) and reacted with NaN₃ (926 mg; 15.4 mmol) at room temperature for 3 hours. The reaction mixture was dispersed in distilled water (50 ml) and extracted with CH₂Cl₂ (3×50 ml). The combined organic extract was dried over Na₂SO₄ and concentrated under reduced pressure. The residue was dissolved in MeOH (5 ml) and reacted with NH₄F (600 mg; 16.2 mmol) at room temperature for 24 hours. The reaction mixture was concentrated under reduced pressure and partitioned between H₂O and CH₂Cl₂. The organic layer was
separated and dried over Na$_2$SO$_4$. After concentration, the residue was purified by flash column chromatography (hexane/ethyl acetate, 1:1 to 2:5) to afford 7 as a white powder (550 mg; 71% yield). $^1$H NMR (400 MHz, CDCl$_3$) $\delta$ 8.30 (br s, 1H), 7.40 (s, 1H), 6.14 (t, $J = 6.8$ Hz, 1H), 4.70-4.79 (m, 2H), 4.50 (m, 1H), 4.16 (m, 1H), 3.84-4.01 (m, 2H), 2.45 (m, 2H), 1.95 (s, 3H); HRMS (Fab$^+$) calcd for C$_{11}$H$_{16}$O$_5$N$_5$ [(M+H)$^+$]: 298.1151, found: 298.1146.

3'-O-azidomethyl-dTTP (Compound 8). The preparation procedure was similar to the synthesis of 4. $^1$H NMR (400 MHz, D$_2$O) $\delta$ 8.30 (br s, 1H), 7.40 (s, 1H), 6.16 (t, 1H), 3.81(m, 2H), 3.20-3.40 (m, 3H), 2.45 (m, 2H), 1.93 (s, 3H); $^{31}$P NMR (121.4 MHz, D$_2$O) $\delta$ -11.52 (d, $J = 19$ Hz, 1P), -12.14 (d, $J = 20$ Hz, 1P), -23.19 (t, $J = 21$ Hz, 1P).

2.4.3 Synthesis of 3'-O-azidomethyl-dCTP

$N^4$-benzoyl-3'-O-(methylthiomethyl)-5'-O-(tert-butyldimethylsilyl)-2'-deoxycytidine (Compound 10). To a stirred solution of $N^4$-benzoyl-5'-O-(tert-butyldimethylsilyl)-2'-deoxycytidine (3.5 g; 7.65 mmol) in DMSO (14.7 ml), acetic acid (6.7 ml) and acetic anhydride (21.6 ml) were added. The reaction mixture was stirred at room temperature for 48 hours. A saturated NaHCO$_3$ solution (100 ml) was added and the aqueous layer was extracted with CH$_2$Cl$_2$ (3×100 ml). The combined organic extract was washed with a saturated solution of NaHCO$_3$ and dried over Na$_2$SO$_4$. After concentration, the crude product was purified by flash column chromatography (ethyl acetate/hexane, 2:1 to 9:1) to afford 10 as a white powder (2.9 g; 73% yield). $^1$H NMR (400 MHz, CDCl$_3$) $\delta$ 8.43 (d, $J = 7.1$ Hz, 1H), 7.93 (m, 2H), 7.64 (m, 1H), 7.54 (m, 3H), 6.30 (m, 1H), 4.62-4.70 (m,
2H), 4.50 (m, 1H), 4.19 (m, 1H), 3.84-3.99 (m, 2H), 2.72 (m, 1H), 2.21 (m, 1H), 2.14 (s, 3H), 0.99 (s, 9H), 0.16 (s, 6H); HRMS (Fab+) cald for C_{24}H_{36}O_5N_3SiS [(M+H)^+]: 506.2145, found: 506.2134.

*N*^4^-*Benzoyl-3'-O-(azidomethyl)-2'-deoxycytidine (Compound 11). To a stirred solution of 10 (558 mg; 1.04 mmol) in dry CH_2Cl_2 (8 ml), cyclohexene (560 μl) and SO_2Cl_2 (220 μl; 2.7 mmol, redistilled) were added. The reaction mixture was stirred at 0°C for 1 hour. The volatiles were removed under reduced pressure. The residue was dissolved in dry DMF (5 ml) and reacted with NaN_3 (400 mg; 6.6 mmol) at room temperature for 2 hours. The reaction mixture was dispersed in distilled water (50 ml) and extracted with CH_2Cl_2 (3×50 ml). The combined organic extract was dried over Na_2SO_4 and concentrated under reduced pressure. The residue was dissolved in MeOH (5 ml) and reacted with NH_4F (600 mg; 16.2 mmol) at room temperature for 24 hours. The solvent was removed under reduced pressure. The residue was suspended in water (50 ml) and extracted with CH_2Cl_2 (3×50 ml). The combined organic extract was dried over Na_2SO_4 and concentrated under reduced pressure. The crude product was purified by flash column chromatography (hexane/ethyl acetate, 1:4 to 1:10) to afford 11 as a white powder (200 mg; 50% yield).  

^1^H NMR (400 MHz, CDCl_3) δ 8.30 (d, J = 7.2 Hz, 1H), 7.93 (d, J = 7.5 Hz, 1H), 7.51-7.66 (m, 5H), 6.18 (t, J = 6.4 Hz, 1H), 4.68-4.81 (m, 2H), 4.52 (m, 1H), 4.25 (m, 1H), 3.88-4.08 (m, 2H), 2.69 (m, 1H), 2.50 (m, 2H); HRMS (Fab+) cald for C_{17}H_{19}O_5N_6 [(M+H)^+]: 387.1417, found: 387.1408.

3'-O-Azidomethyl-dCTP (Compound 12). The preparation procedure was similar to the
synthesis of 4. $^1$H NMR (400 MHz, D$_2$O) $\delta$ 7.97 (s, 1H), 6.05 (t, 1H), 4.15 (m, 1H), 3.66 (d, 2H), 3.19-3.42 (m, 3H), 2.42 (m, 2H), 2.02 (br s, 2H); $^{31}$P NMR (121.4 MHz, D$_2$O) $\delta$ -9.91 (d, $J = 19$ Hz, 1P), -10.18 (d, $J = 20$ Hz, 1P), -23.05 (t, $J = 23$ Hz, 1P).

2.4.4 Synthesis of 3’-O-azidomethyl-dGTP

$N^2$-Isobutyryl-3’-O-(methylthiomethyl)-5’-O-(tert-butyldimethylsilyl)-2’-deoxyguanosine (Compound 14). To a stirred solution of $N^2$-isobutyryl-5’-O-(tert-butyldimethylsilyl)-2’-deoxyguanosine (5 g; 11.0 mmol) in dry DMSO (21 ml), acetic acid (10 ml) and acetic anhydride (32 ml) were added. The reaction mixture was stirred at room temperature for 48 hours. A saturated K$_2$CO$_3$ solution (100 ml) was added and the aqueous layer was extracted with ethyl acetate (3×100 ml). The combined organic extract was washed with a saturated NaHCO$_3$ solution and dried over Na$_2$SO$_4$. After concentration, the crude product was purified with flash column chromatography (CH$_2$Cl$_2$/MeOH, 20:1) to afford 14 as a white powder (3.9 g; 69% yield). $^1$H NMR (400 MHz, CDCl$_3$) $\delta$ 12.0 (s, 1H), 8.95 (br s, 1H), 8.09 (s, 1H), 6.24 (t, $J = 6.8$ Hz, 1H), 4.73 (m, 2H), 4.66 (m, 1H), 4.16 (m, 1H), 3.81 (m, 2H), 2.76 (m, 1H), 2.59 (m, 1H), 2.54 (m, 1H), 2.21 (s, 3H), 1.29 (m, 6H), 0.91 (s, 9H), 0.10 (s, 6H); HRMS (Fab+) cald for C$_{22}$H$_{38}$O$_5$N$_3$SiS [(M+H)$^+$]: 512.2363, found: 512.2344.

$N^2$-Isobutyryl-6’-(diphenylcarbamoyl)-3’-O-(methylthiomethyl)-5’-O-(tert-butyldimethylsilyl)-2’-deoxyguanosine (Compound 15). To a stirred solution of 14 (1.0 g; 2.0 mmol) in dry pyridine (22 ml), diphenylcarbamoyl chloride (677 mg; 2.92 mmol) and DIEA (N, N-diisopropylethylamine) (1.02 ml; 5.9 mmol) were added. The reaction
mixture was stirred under nitrogen atmosphere at room temperature for 3 hours. The solvent was removed under high vacuum. The crude product was purified by flash column chromatography (ethyl acetate/hexane, 1:1 to 7:3) to afford 15 as a yellowish powder (1.09 g; 80% yield). $^1$H NMR (400 MHz, CDCl$_3$) δ 8.25 (s, 1H), 7.94 (br s, 1H), 7.37-7.47 (m, 10H), 6.42 (m, 1H), 4.75 (m, 2H), 4.71 (m, 1H), 4.18 (m, 1H), 3.70-3.88 (m, 2H), 2.80 (m, 1H), 2.60 (m, 1H), 2.19 (s, 3H), 1.30 (d, $J = 7.2$ Hz, 6H), 0.93 (s, 9H), 0.14 (s, 6H); HRMS (Fab+) calcd for C$_{35}$H$_{47}$O$_6$N$_6$SiS [(M+H)$^+]: 707.3047$, found: 707.3068.

$N^2$-Isobutyryl-$O^6$-(diphenylcarbamoyl)-3’-O-azidomethyl-2’-deoxyguanosine  (Compound 16). To a stirred solution of 15 (786 mg; 1.1 mmol) in dry CH$_2$Cl$_2$ (8 ml), cyclohexene (560 µl) and SO$_2$Cl$_2$ (180 µl; 2.2 mmol, redistilled) were added. The reaction mixture was stirred under nitrogen atmosphere at 0°C for 1.5 hours. The solvent was first removed under reduced pressure and then under high vacuum for 10 minutes. The residue was dissolved in dry DMF (5 ml) and reacted with NaN$_3$ (600 mg; 10 mmol) at room temperature for 3 hours. The reaction mixture was dispersed in distilled water (50 ml) and extracted with CH$_2$Cl$_2$ (3×50 ml). The combined organic extract was dried over Na$_2$SO$_4$ and concentrated under reduced pressure. The residue was dissolved in MeOH (5 ml) and reacted with NH$_4$F (500 mg; 13.5 mmol) at room temperature for 24 hours. The solvent was removed under reduced pressure. The residue was suspended in water (50 ml) and extracted with CH$_2$Cl$_2$ (3×50 ml). The combined organic extract was dried over Na$_2$SO$_4$ and concentrated under reduced pressure. The crude product was purified by flash column chromatography (hexane/ethyl acetate, 1:1 to 1:5) to afford 16.
as a white powder (230 mg; 36% yield). 1H NMR (400 MHz, DMSO-d6) δ 8.64 (br s, 1H), 7.34-7.48 (m, 10H), 6.36 (t, J = 7.0 Hz), 4.93 (m, 2H), 4.76 (m, 1H), 4.04 (m, 1H), 3.57 (m, 1H), 3.34 (m, 2H), 2.79 (m, 1H), 2.81 (m, 1H), 1.10 (m, 6H); HRMS (Fab+) calcd for C28H30O6N9 [(M+H)+]: 588.2319, found: 588.2343.

3’-O-Azidomethyl-dGTP (Compound 17). The preparation procedure was similar to the synthesis of 4. 1H NMR (400 MHz, D2O) δ 7.93 (d, 1H), 5.95 (t, 1H), 4.52 (m, 1H), 4.25 (m, 1H), 3.92 (m, 2H), 3.19-3.42 (m, 3H), 2.80 (m, 1H), 2.50 (m, 1H), 2.20 (br s, 2H); 31P NMR (121.4 MHz, D2O) δ -10.63 (d, J = 19 Hz, 1P), -11.35 (d, J = 22 Hz, 1P), -23.85 (t, J = 21Hz, 1P).

2.4.5 Synthesis of ddCTP-N3 (version I)-Bodipy-FL-510

Azido (version I)-Bodipy-FL-510 (Compound 26). (2-{2-[3-(2-Amino-ethylcarbamoyl)-phenoxy]-1-azido-ethoxy}-ethoxy)-acetic acid 24 (7.0 mg, 0.019 mmol) prepared according to the literature[24] was dissolved in DMF (300 µl) and 1 M NaHCO3 aqueous solution (100 µl). A solution of Bodipy-FL-510 NHS (N-hydroxysuccinimide) ester (Invitrogen) (5.0 mg, 0.013 mmol) in DMF (400 µl) was added slowly to the above reaction mixture and then stirred at room temperature for 5 hours with exclusion of light. The crude product was purified on a preparative silica gel TLC plate (CHCl3/CH3OH, 1:4) to afford 26 (7.6 mg; 91%). 1H NMR (400 MHz, CD3OD) δ 7.34-7.42 (m, 4H), 710-7.14 (m, 1H), 6.90 (d, J = 4.0 Hz, 1H), 6.29 (d, J = 4.0 Hz, 1H), 6.20 (s, 1H), 5.00 (t, J = 5.2 Hz, 1H), 4.22-4.25 (m, 1H), 4.10-4.14 (m, 1H), 3.96-4.01 (m, 2H), 3.91 (s, 2H), 3.83-3.88 (m, 1H), 3.70-3.71 (m, 2H), 3.43-3.48 (m, 3H), 3.20-3.24 (m, 2H), 2.61-2.65
(m, 2H), 2.57 (s, 3H), 2.49 (s, 3H); MS (Fab+) calcd for C_{29}H_{34}BF_{2}N_{7}O_{7} [(M+H)^+]: 642.4, found: 642.5.

**ddCTP-N\textsubscript{3} (version I)-Bodipy-FL-510 (Compound 28).** To a stirred solution of 26 in dry DMF (2 ml), DSC (N, N’-disuccinimidyl carbonate) (3.4 mg, 13.2 µmol) and DMAP (4-dimethylaminopyridine) (1.6 mg, 13.2 µmol) were added. The reaction mixture was stirred at room temperature for 2 hours. TLC indicated that 26 was completely converted to compound 27, which was directly used to couple with amino-ddCTP (13 µmol) in NaHCO\textsubscript{3}/Na\textsubscript{2}CO\textsubscript{3} buffer (pH = 8.7, 0.1 M) (300 µl). The reaction mixture was stirred at room temperature for 3 hours with exclusion of light. The reaction mixture was purified by a preparative silica gel TLC plate (CH\textsubscript{3}OH/CH\textsubscript{2}Cl\textsubscript{2}, 1:1). The crude product was further purified on reverse-phase HPLC to afford 28 (retention time = 34.0 min). Compound 28 was further evaluated by performing a single base extension reaction to yield a DNA extension product which was characterized by MALDI-TOF MS (m/z 8915) (Fig. 2.21C).

### 2.4.6 Synthesis of ddUTP-N\textsubscript{3} (version I)-R6G

**Azido (version I)-R6G (Compound 30).** The preparation procedure was similar to the synthesis of 26. The crude product was purified by a preparative silica gel TLC plate (CH\textsubscript{3}OH/CH\textsubscript{2}Cl\textsubscript{2}, 2:5) to afford 30 (8.2 mg; 89%). \textsuperscript{1}H NMR (400 MHz, CD\textsubscript{3}OD) δ 8.12-8.08 (m, 2H), 7.68 (d, J = 1.6 Hz, 1H), 7.49-7.45 (m, 1H), 7.38-7.36 (m, 2H), 7.32-7.30 (m, 1H), 7.26-7.22 (m, 1H), 7.14-7.12 (m, 1H), 7.06-7.05 (m, 1H), 6.96 (s, 2H), 6.87 (s, 3H), 5.05 (t, J = 5.0 Hz, 1H), 4.15-4.14 (m, 1H), 4.04-4.03 (m, 1H), 3.94-3.92 (m,
2H), 3.86-3.80 (m, 3H), 3.67-3.62 (m, 6H), 3.51 (q, J = 7.2 Hz, 4H), 2.08 (s, 6H), 1.36 (t, J = 7.2 Hz, 6H); HRMS (Fab+) calcd for C\textsubscript{42}H\textsubscript{46}N\textsubscript{7}O\textsubscript{10} [(M+H)\textsuperscript{+}]: 808.3306, found 808.3267.

\textit{ddUTP-}N\textsubscript{3} (version I)-R6G (Compound 32). The preparation procedure was similar to the synthesis of 28. The crude product was purified on reverse-phase HPLC to afford 32 (retention time = 32.9 min). Compound 32 was further evaluated by performing a single base extension reaction to yield a DNA extension product which was characterized by MALDI-TOF MS (m/z 9082) (Fig. 2.21G).

\subsection*{2.4.7 Synthesis of ddATP-}N\textsubscript{3} (version I)-ROX

\textit{Azido (version I)}-ROX (Compound 34). The preparation procedure was similar to the synthesis of 26. The crude product was purified with a preparative silica gel TLC plate (CH\textsubscript{3}OH/CH\textsubscript{2}Cl\textsubscript{2}, 2:5) to afford 34 (6.3 mg; 90%). \textsuperscript{1}H NMR (400 MHz, CD\textsubscript{3}OD) \(\delta\) 8.24 (d, J = 3.2 Hz, 2H), 7.65 (s, 1H), 7.49-7.46 (m, 1H), 7.38-7.35 (m, 1H), 7.32-7.30 (m, 1H), 7.26-7.23 (m, 1H), 7.14-7.12 (m, 1H), 7.05-7.04 (m, 1H), 6.70 (s, 2H), 6.87 (s, 3H), 5.02 (t, J = 4.0 Hz, 1H), 4.26-4.23 (m, 1H), 4.16-4.12 (m, 2H), 4.00-3.97 (m, 2H), 3.90-3.71 (m, 3H), 3.67-3.45 (m, 8H), 3.04-3.01 (m, 4H), 2.66-2.56 (m, 4H), 2.09-2.08 (m, 4H), 1.90-1.89 (m, 4H); HRMS (Fab+) calcd for C\textsubscript{48}H\textsubscript{50}N\textsubscript{7}O\textsubscript{10} [(M+H)\textsuperscript{+}]: 884.3619, found 884.3622.

ddATP- N\textsubscript{3} (version I)-ROX (Compound 36). The preparation procedure was similar to the synthesis of 28. The crude product was purified on reverse-phase HPLC to afford 36
(retention time = 36.1 min). Compound 36 was further evaluated by performing a single base extension reaction to yield a DNA extension product which was characterized by MALDI-TOF MS (m/z 9180) (Fig. 2.21A).

2.4.8 Synthesis of ddGTP-N₃ (version I)-Cy5

Azido (version I)-Cy5 (Compound 38). The preparation procedure was similar to the synthesis of 26. The crude product was purified with a preparative silica gel TLC plate (CH₃OH/CH₂Cl₂, 2:5) to afford 38 (5.6 mg; 84%). ¹H NMR (400 MHz, CD₂OD) δ 8.35-8.28 (m, 2H), 7.90-7.86 (m, 3H), 7.46-7.44 (m, 2H), 7.37-7.35 (m, 2H), 7.30-7.28 (d, J = 8.0 Hz, 1H), 7.12-7.10 (m, 1H), 6.72 (t, J = 12.4 Hz, 1H), 6.37-6.29 (m, 1H), 5.03 (t, J = 4.8 Hz, 1H), 4.25-4.24 (m, 1H), 4.22-4.10 (m, 3H), 4.04-3.98 (m, 3H), 3.92 (s, 2H), 3.89-3.86 (m, 1H), 3.72-3.71 (m, 2H), 3.50-3.47 (m, 2H), 3.41-3.38 (m, 2H), 2.57 (m, 1H), 2.24-2.20 (m, 2H), 1.76 (s, 12H), 1.69-1.65 (m, 2H), 1.43-1.36 (m, 6H); MS (Fab+) calcd for C₄₈H₅₈N₇O₁₃S₂ [(M+H)⁺]: 1006.4, found 1006.6.

ddGTP-N₃ (version I)-Cy5 (Compound 40). The preparation procedure was similar to the synthesis of 28. The crude product was purified by reverse-phase HPLC to afford 40 (retention time = 31.6 min). Compound 40 was further evaluated by performing a single base extension reaction to yield a DNA extension product which was characterized by MALDI-TOF MS (m/z 9261) (Fig. 2.21E).

2.4.9 Synthesis of 2-(2-amino-1-azidoethoxy)ethanol (cleavable linker (version II)) (9H-fluoren-9-yl)methyl (1,3-dioxolan-2-yl)methylcarbamate (Compound 43). A
solution of 9-fluorenymethyl chloroformate (1.3 g; 5 mmol) in ether (30 ml) was cooled in an ice bath. Commercially available 2-(aminomethyl)-1,3-dioxolane (1.03 g; 10 mmol) was added slowly. The reaction mixture was stirred at 0°C for 30 minutes and at room temperature for 30 minutes. After evaporation of the solvent, the residue was purified by flash column chromatography (ethyl acetate/hexane, 1:1) to afford 43 as a white solid (1.6 g; 98% yield). \(^1\)H NMR (400 MHz, CD\(_3\)OD) \(\delta\) 7.80 (d, \(J = 7.2\) Hz, 2H, Ar-H), 7.62 (d, \(J = 7.2\) Hz, 2H, Ar-H), 7.42 - 7.45 (m, 2H, Ar-H), 7.33 - 7.36 (m, 2H, Ar-H), 5.00 (t, \(J = 3.6\) Hz, 1H, O-CH-O), 4.44 (d, \(J = 6.8\) Hz, 2H, CH\(_2\)OC(O)), 4.26 (t, \(J = 6.8\) Hz, 1H, CH-Ar), 4.02 (t, \(J = 4.4\) Hz, 2H, OCH\(_2\), H\(_a\)), 3.94 (t, \(J = 4.4\) Hz, 2H, OCH\(_2\), H\(_b\)), 3.48 (d, \(J = 3.6\) Hz, 2H, NHCH\(_2\)); APCI-MS calcd for C\(_{19}\)H\(_{19}\)NO\(_4\) [(M+H)\(^+\): 326.1, found: 326.1.

**(9H-fluoren-9-yl)methyl 2-azido-2-(2-hydroxyethoxy)ethylcarbamate (Compound 44).**

A stirred solution of 43 (650 mg; 2 mmol) in CH\(_2\)Cl\(_2\) (15 ml) was cooled to -78°C in a dry ice/acetone bath under argon atmosphere. To this solution azidotrimethylsilane (520 \(\mu\)l; 4 mmol) and SnCl\(_4\) (1 M solution in CH\(_2\)Cl\(_2\); 50 \(\mu\)l) were added. The reaction mixture was warmed to room temperature and stirred for 15 hours. After adding CH\(_2\)Cl\(_2\) (150 ml), the organic layer was washed with H\(_2\)O and dried over anhydrous Na\(_2\)SO\(_4\). After evaporation of the solvent, the residue was purified by flash column chromatography (ethyl acetate/hexane, 1:1) to afford 44 as a colorless oil (530 mg; 72% yield). \(^1\)H NMR (400 MHz, CD\(_3\)OD) \(\delta\) 7.80 (d, \(J = 7.2\) Hz, 2H, Ar-H), 7.62 (d, \(J = 7.2\) Hz, 2H, Ar-H), 7.42 - 7.45 (m, 2H, Ar-H), 7.33 - 7.36 (m, 2H, Ar-H), 4.60 (t, \(J = 5.6\) Hz, 1H, N\(_3\)-CH), 4.38 (d, \(J = 6.8\) Hz, 2H, CH\(_2\)OC(O)), 4.20 (t, \(J = 6.8\) Hz, 1H, CH-Ar), 3.62 - 3.88 (m, 4H, OCH\(_2\)CH\(_2\)O), 3.38 (d, \(J = 4.4\) Hz, 2H, NHCH\(_2\)); \(^{13}\)C NMR (100 MHz, CD\(_3\)OD) \(\delta\) 157.3,
144.2, 141.7, 128.2, 127.5, 125.4, 120.4, 90.9, 71.5, 67.3, 62.0, 47.6, 45.1; APCI-MS calcd for C_{19}H_{20}N_{4}O_{4} [(M-N_{2}+H)^+] : 341.2, found: 341.9.

2-(2-Amino-1-azidoethoxy)ethanol (Compound 45). To a stirred solution of 44 (500 mg; 1.36 mmol) in CH$_2$Cl$_2$ (3 ml), piperidine (2 ml; 20.3 mmol) was added. The reaction mixture was stirred at room temperature for 30 minutes, and then most of the solvent and piperidine were removed under vacuum. The residue was purified by flash column chromatography (ethyl acetate/hexane, 1:1 and then pure methanol) to afford 45 as a brown oil (190 mg; 95% yield). $^1$H NMR (400 MHz, CD$_3$OD) $\delta$ 4.78 (t, $J = 5.6$ Hz, 1H, N$_3$-CH), 3.70-3.95 (m, 4H, OCH$_2$CH$_2$O), 2.98 (d, $J = 4.4$ Hz, 2H, NHCH$_2$); $^{13}$C NMR (100 MHz, CD$_3$OD) $\delta$ 90.9, 72.3, 61.9, 45.7; HRMS (FAB+) calcd for C$_4$H$_{10}$N$_4$O$_2$ [(M+H)$^+$]: 147.0882, found: 147.0880.

2.4.10 Synthesis of ddCTP-N$_3$ (version II)-Bodipy-FL-510

Azido (version II)-Bodipy-FL-510 (Compound 47). To a stirred solution of 45 (5 mg; 34.2 µmol) in DMF (450 µl), aqueous NaHCO$_3$ (1 M; 100 µl) was added. The solution was stirred at room temperature for 5 minutes. Bodipy-FL-510 NHS (N-hydroxysuccinimide) ester (5 mg; 12.8 µmol) in DMF (450 µl) was added. The reaction mixture was stirred at room temperature for 6 hours. The crude product was purified with a preparative silica gel TLC plate (CH$_3$OH/CH$_2$Cl$_2$, 1:15) to afford 47. $^1$H NMR (400 MHz, CD$_3$OD) $\delta$ 7.42 (s, 1H), 7.00 (d, $J = 4.0$ Hz, 1H), 6.32 (d, $J = 4.0$ Hz, 1H), 6.21(s, 1H), 4.65 (t, $J = 5.6$ Hz, 1H, N$_3$-CH), 3.62-3.92 (m, 4H, OCH$_2$CH$_2$O), 3.33-3.38 (m, 4H), 3.21 (t, $J = 7.1$ Hz, 2H), 2.51 (s, 3H), 2.28 (s, 3H); APCI-MS calcd for
C_{18}H_{23}BF_{2}N_{6}O_{3} [(M+H)]: 421.2, found: 421.2.

**Azido (version II)-Bodipy-FL NHS ester (Compound 48).** To a stirred solution of 47 in anhydrous DMF (350 µl), DSC (N, N’-disuccinimidyl carbonate) (8.0 mg; 31.2 µmol) and triethylamine (5.0 µl; 35.4 µmol) were added. The reaction was stirred at room temperature for 10 hours and then most of the DMF was removed under vacuum. The crude product was purified quickly by flash column chromatography (CH₃OH/CH₂Cl₂, 1:50) to afford 48, which was used directly for the next step.

**ddCTP- N₃ (version II)-Bodipy-FL (Compound 49).** To a stirred solution of 48 in DMF (300 µl), amino-ddCTP in NaHCO₃-Na₂CO₃ buffer (0.1 M; 300 µl; pH 8.7) was added. The reaction mixture was stirred at room temperature for 7 hours and purified on a preparative silica gel TLC plate (CH₃OH/CH₂Cl₂, 1:1). The crude product was further purified by reverse-phase HPLC to afford 49 (retention time = 31.5 min). Compound 49 was further evaluated by performing a single base extension reaction to yield a DNA extension product which was characterized by MALDI-TOF MS (m/z 8738) (Fig. 2.22C).

### 2.4.11 Synthesis of ddUTP-N₃ (version II)-R6G

**Azido (version II)-R6G (Compound 51).** The preparation procedure is similar to the synthesis of 47. The crude product was purified on a preparative silica gel TLC plate (CH₃OH/CH₂Cl₂, 1:2) to afford 51. $^1$H NMR (400 MHz, CD₃OD) $\delta$ 8.12 (d, $J = 8.1$ Hz, 1H), 8.05 (dd, $J = 1.8$, 8.1 Hz, 1H), 7.66 (d, $J = 1.6$ Hz, 1H), 7.02 (s, 2H), 6.88 (s, 2H), 4.83 (t, $J = 5.6$ Hz, 1H, N₃-CH), 3.65-3.98 (m, 4H, OCH₂CH₂O), 3.48-3.56 (m, 6H), 2.13
(s, 6H), 1.36 (t, J = 5.6 Hz, 6H); APCI-MS calcd for C_{31}H_{34}N_{6}O_{6} [(M+H)^+]: 587.3, found 587.3.

**Azido- (version II)R6G NHS ester (Compound 52).** The preparation procedure is similar to the synthesis of 48. The crude product was purified by flash column chromatography (CH$_3$OH/CH$_2$Cl$_2$, 1:4) to afford 52, which was used directly for the next step.

**ddUTP- N$_3$ (version II)-R6G (Compound 53).** The preparation procedure is similar to the synthesis of 49. The crude product was purified by reverse-phase HPLC to afford 53 (retention time = 32.3 min). Compound 53 was further evaluated by performing a single base extension reaction to yield a DNA extension product which was characterized by MALDI-TOF MS (m/z 8905) (Fig. 2.22G).

### 2.4.12 Synthesis of ddATP-N$_3$ (version II)-ROX

**Azido (version II)-ROX (Compound 55).** The preparation procedure is similar to the synthesis of 47. The crude product was purified on a preparative silica gel TLC plate (CH$_3$OH/CH$_2$Cl$_2$, 1:2) to afford 23. $^1$H NMR (400 MHz, CD$_3$OD) δ 8.03 (d, J = 8.1 Hz, 1H), 7.98 (dd, J = 1.6, 8.1 Hz, 1H), 7.60 (d, J = 1.4 Hz, 1H), 6.75 (s, 2H), 4.79 (t, J = 5.4 Hz, 1H, N$_3$-CH), 3.60-3.92 (m, 4H, OCH$_2$CH$_2$O), 3.40 (d, J = 4.4 Hz, 2H, NHCH$_2$), 3.03-3.10 (m, 12H), 2.64-2.73 (m, 4H), 2.04-2.15 (m, 4H), 1.89-1.99 (m, 4H); APCI-MS calcd for C$_{37}$H$_{38}$N$_6$O$_6$ [(M+H)$^+$]: 663.3, found: 663.6.
Azido (version II)-ROX NHS ester (Compound 56). The preparation procedure is similar to the synthesis of 48. The crude product was purified by flash column chromatography (CH$_3$OH/CH$_2$Cl$_2$, 1:5) to afford 56, which was used directly for the next step.

ddATP-N$_3$ (version II)-ROX (Compound 57). The preparation procedure is similar to the synthesis of 49. The crude product was purified by reverse-phase HPLC to afford 57 (retention time = 36.2 min). Compound 57 was further evaluated by performing a single base extension reaction to yield a DNA extension product which was characterized by MALDI-TOF MS (m/z 9003) (Fig. 2.22A).

2.4.13 Synthesis of ddGTP-N$_3$ (version II)-Cy5

Azido (version II)-Cy5 (Compound 59). The preparation procedure is similar to the synthesis of 47. The crude product was purified on a preparative silica gel TLC plate (CH$_3$OH/CH$_2$Cl$_2$, 1:2) to afford 59. $^1$H NMR (400 MHz, CD$_3$OD) $\delta$ 8.32 (d, $J = 8.0$ Hz, 1H) 7.93 (m, 2H), 7.87 (m, 2H), 7.36 (d, $J = 8.0$ Hz, 1H), 6.25-6.40 (m, 3H), 5.62-5.65 (m, 2H), 4.67 (t, $J = 5.6$ Hz, 1H, N$_3$-CH), 4.11-4.25 (m, 2H), 3.60-3.92 (m, 4H, OCH$_2$CH$_2$O), 3.22-3.56 (m, 10H), 1.81-1.90 (m, 2H), 1.78 (s, 6H), 1.31 (s, 6H), 0.9 (t, $J = 4.2$ Hz, 3H); APCI-MS calcd for C$_{37}$H$_{47}$N$_6$O$_9$S$_2$ [(M+H)$^+$]: 784.3, found: 784.1.

Azido (version II)-Cy5 NHS ester (Compound 60). The preparation procedure is similar to the synthesis of 48. The crude product was purified by flash column chromatography (CH$_3$OH/CH$_2$Cl$_2$, 1:3) to afford 60, which was used directly for the next step.
ddGTP- \( N_3 \) (version II)-Cy5 (Compound 61). The preparation procedure is similar to the synthesis of 49. The crude product was purified by reverse-phase HPLC to afford 61 (retention time = 30.2 min). Compound 61 was further evaluated by performing a single base extension reaction to yield a DNA extension product which was characterized by MALDI-TOF MS (m/z 9140) (Fig. 2.22E).

2.4.14 Continuous polymerase extension using 3’-O-modified NRTs in solution and characterization by MALDI-TOF MS.

We characterized the four NRTs (3’-O-N3-dCTP, 3’-O-N3-dTTP, 3’-O-N3-dATP and 3’-O-N3-dGTP) by performing four continuous DNA-extension reactions sequentially using a self-priming DNA template (5’-ATC-GGC-GCC-GCG-CCT-TGG-CGC-GGC-GC-3’). The four nucleotides in the template immediately adjacent to the annealing site of the primer are 3’-GCTA-5’, which allows the evaluation of the incorporation and cleavage efficiency of the 4 NRTs. First, a polymerase extension reaction using a pool of all four NRTs along with the self-priming DNA template was performed producing a single base extension product. The reaction mixture for this, and all subsequent extension reactions, consisted of 80 pmol of self-priming DNA template, 160 pmol of 3’-O-N3-dNTPs, 1X Thermopol II reaction buffer (New England Biolabs), 40 nmol of MnCl2 and 1 unit of 9°N DNA polymerase (exo-) A485L/Y409V (New England Biolabs) in a total reaction volume of 20 µl. The reaction consisted of incubation at 94°C for 5 minutes, 4°C for 5 minutes and 65°C for 20 minutes. Subsequently, the extension product was desalted by using a ZipTip and analyzed by Voyager DE™ MALDI-TOF mass spectrometry (Applied Biosystems). For cleavage, the desalted DNA
extension product bearing the 3’-O-azidomethyl group was first resuspended with 5 µl of 50 mM EDTA solution to quench the polymerase activity. This DNA solution was then mixed with 10 µl of 225 mM TCEP solution (pH 9.0) and incubated at 65°C for 15 minutes to yield a cleaved DNA product which was characterized by MALDI-TOF MS. The DNA product with the 3’-O-azidomethyl group removed to generate a free 3’-OH group was purified using an Oligonucleotide Purification Cartridge (Applied Biosystems) and used as a primer for a second extension reaction using 3’-O-N₃-dNTPs. The second extended DNA product was then purified by ZipTip and cleaved as described above. The third and the fourth extensions were carried out in a similar manner using the previously extended and cleaved product as the primer.

2.4.15 Polymerase extension using cleavable fluorescent dideoxynucleotide terminators

*ddNTP-N₃ (version I)-fluorophores in solution and characterization by MALDI-TOF MS.*

We characterized the four cleavable fluorescent dideoxynucleotide terminators, ddNTP-N₃ (version I)-fluorophores (ddCTP-N₃ (version I)-Bodipy-FL-510, ddUTP-N₃ (version I)-R6G, ddATP-N₃ (version I)-ROX and ddGTP-N₃ (version I)-Cy5) by performing four separate DNA-extension reactions, each with a different self-priming DNA template allowing the four ddNTP analogues to be incorporated. The resulting DNA extension products were analyzed by MALDI-TOF MS. The following four self-priming DNA templates (26-mer hairpin DNA with a 4-base 5’-overhang) were used for the extension: 5’-GACTGCGCCGCGCCTTTGGCCGCGCGCGC-3’ for ddATP-N₃ (version I)-ROX; 5’-ATCGGCAGCGCGCCTTTGGCCGCGCGC-3’ for ddCTP-N₃
(version I)-Bodipy-FL-510; 5’-GATCGCGCCGCCTTGGCGCGGCGC-3’ for ddGTP-N₃ (version I)-Cy5; and 5’-GTCAGCGCGCCGCTTGGCGCGGCGC-3’ for ddUTP-N₃ (version I)-R6G. Each of the extension reactions consisted of all four ddNTP-N₃ (version I)-fluorophores (120 pmol each of ddCTP-N₃ (version I)-Bodipy-FL-510, ddUTP-N₃ (version I)-R6G, ddATP-N₃ (version I)-ROX and ddGTP-N₃ (version I)-Cy5) along with 60 pmol of the self-priming DNA template, 1X Thermopol II reaction buffer, 40 nmol of MnCl₂ and 1 unit of 9⁰N DNA polymerase (exo-) A485L/Y409V in a total reaction volume of 20 µl. The reaction consisted of incubations at 94°C for 5 minutes, 4°C for 5 minutes, and 65°C for 20 minutes. Subsequently, the extension product was purified by reverse-phase HPLC using established procedures.¹⁶ The fraction containing the desired product was collected and freeze-dried for analysis by MALDI-TOF MS and cleavage. For cleavage of the DNA extension product bearing the ddNTP-N₃ (version I)-fluorophores, the purified DNA product was resuspended in 50 µl of 100 mM TCEP solution (pH 9.0) at 65°C for 15 minutes and then analyzed by MALDI-TOF MS.

2.4.16 Polymerase extension using cleavable fluorescent dideoxynucleotide terminators ddNTP-N₃ (version II)-fluorophores in solution and characterization by MALDI-TOF MS.

We characterized the four cleavable fluorescent dideoxynucleotide terminators, ddNTP-N₃ (version II)-fluorophores (ddCTP-N₃ (version II)-Bodipy-FL-510, ddUTP-N₃ (version II)-R6G, ddATP-N₃ (version II)-ROX and ddGTP-N₃ (version II)-Cy5) by performing four separate DNA-extension reactions, each with a different self-priming
DNA template allowing the four ddNTP analogues to be incorporated. The resulting DNA extension products were analyzed by MALDI-TOF MS. The following four self-priming DNA templates (26-mer hairpin DNA with a 4-base 5’-overhang) were used for the extension: 5’-GACTGCGCCGCGCTTTGGCGCGGC-3’ for ddATP-N3 (version II)-ROX; 5’-ATCGGCGCCCGCCTTTGGCGCGGC-3’ for ddCTP-N3 (version II)-Bodipy-FL-510; 5’-GATCGC GCCGCCGCTTTGGCGCGGC-3’ for ddGTP-N3 (version II)-Cy5; and 5’-GTCAGCGCCGCCCTTTGGCGCGGC-3’ for ddUTP-N3 (version II)-R6G. Each of the extension reactions consisted of all four ddNTP-N3 (version II)-fluorophores (120 pmol each of ddCTP-N3 (version II)-Bodipy-FL-510, ddUTP-N3 (version II)-R6G, ddATP-N3 (version II)-ROX and ddGTP-N3 (version II)-Cy5) along with 60 pmol of the self-priming DNA template, 1X Thermopol II reaction buffer, 40 nmol of MnCl2 and 1 unit of 9oN DNA polymerase (exo-) A485L/Y409V in a total reaction volume of 20 µL. The reaction consisted of incubations at 94°C for 5 minutes, 4°C for 5 minutes, and 65°C for 20 minutes. Subsequently, the extension product was purified by reverse-phase HPLC using established procedures.[16] The fraction containing the desired product was collected and freeze-dried for analysis by MALDI-TOF MS and cleavage. For cleavage of the DNA extension product bearing the ddNTP-N3 (version II)-fluorophores, the purified DNA product was resuspended in 50 µL of 100 mM TCEP solution (pH 9.0) at 65°C for 10 minutes and then analyzed by MALDI-TOF MS.

2.4.17 Four-color DNA sequencing on a chip using cleavable fluorescent dideoxynucleotides and 3’-O-modified NRTs in the hybrid SBS approach.
A DNA chip was constructed by immobilizing a 5’-amino-modified looped oligonucleotide on a CodeLink microarray slide (GE Healthcare). The 5’-amino-labeled self-priming DNA template 5’-NH₂-CAC-TCA-CAT-ATG-TTT-TTT-AGC-TTT-TTT-AAT-TTC-TTA-ATG-ATG-TTG-TTG-CAT-GCG-ACT-TAA-GGC-GCT-TGC-GCC-TTA-AGT-CG-3’ was purchased from IDT (Coralville, IA). The DNA template was dissolved at 40 µM in 50 mM sodium phosphate buffer, pH 8.5 and spotted using a SpotArray 72 arraying robot (Perkin Elmer) onto high density CodeLink microarray slides (GE Healthcare). After spotting, the slides were incubated at ambient temperature (~ 24°C) for 20 hours in a humid chamber containing saturated sodium chloride solution (~ 75% humidity) to allow for 5’-tethering of the spotted amino-modified DNA templates to the slide surface functionalized with succinimide ester groups. After incubation the slides were removed from the humid chamber and stored in a vacuum dessicator at room temperature until further use. The principal advantage of the hairpin structure introduced into the 3’-portion of the self-priming DNA template is its higher stability and the increased priming efficiency for DNA polymerases as compared to a separate primer/template complex, which is prone to spontaneous dissociation.

Ten microliters of a solution consisting of ddCTP-N₃ (version I)-Bodipy-FL-510 (10 fmol), ddUTP-N₃ (version I)-R6G (20 fmol), ddATP-N₃ (version I)-ROX (40 fmol), ddGTP-N₃ (version I)-Cy5 (20 fmol), 3’-O-N₃-dCTP (22 pmol), 3’-O-N₃-dTTP (22 pmol), 3’-O-N₃-dATP (22 pmol), 3’-O-N₃-dGTP (4 pmol), 1 unit of 9°N DNA polymerase(exo-)A485L/Y409V, 20 nmol of MnCl₂ and 1X Thermopol II reaction buffer was spotted on the surface of the chip, where the self-primed DNA moiety was immobilized. The nucleotide complementary to the DNA template was allowed to incorporate into the
primer at 65°C for 15 minutes. To synchronize any unincorporated templates, an extension solution consisting of 38 pmol each of 3’-O-N3-dTTP, 3’-O-N3-dATP, 3’-O-N3-dGTP, and 75 pmol of 3’-O-N3-dCTP, 1 unit of 9°N DNA polymerase(exo-) A485L/Y409V, 20 nmol of MnCl2 and 1X Thermopol II reaction buffer was added to the same spot and incubated at 65°C for 15 minutes. After washing the chip with SPSC buffer containing 0.1% Tween-20 for 1 minute, the surface was rinsed with dH2O, dried briefly and then scanned with a 4-color ScanArray Express scanner (Perkin–Elmer Life Sciences) to detect the fluorescent signal. The 4-color scanner is equipped with four lasers with excitation wavelengths of 488, 543, 594, and 633 nm and emission filters centered at 522, 570, 614, and 670 nm. To perform the cleavage, the glass chip and the spot where the self-primed DNA moiety is immobilized was placed inside a chamber filled with 100 mM TCEP (pH 9.0) and incubated at 65°C for 10 minutes. After washing the surface with dH2O, the chip was scanned again to compare the intensity of fluorescence after cleavage with the original fluorescence intensity. This process was followed by the next polymerase extension reaction using the 3’-O-N3-dNTP/ddNTP-N3 (version I)-fluorophore solution with the subsequent synchronization, washing, fluorescence detection, and cleavage processes performed in the same way as described above. To obtain de novo DNA sequencing data on a DNA template immobilized on a chip, the SBS cycle was repeated multiple times using the combination mixture of solution A consisting of 3’-O-N3-dCTP (3 µM), 3’-O-N3-dTTP (3 µM), 3’-O-N3-dATP (3 µM) and 3’-O-N3-dGTP (0.5 µM) and solution B consisting of ddCTP-N3 (version I)-Bodipy-FL-510 (50 nM), ddUTP-N3 (version I)-R6G (100 nM), ddATP-N3 (version I)-ROX (200 nM) and ddGTP-N3 (version I)-Cy5 (100 nM) in each polymerase extension
reaction. The volumes of solution A and B in each SBS cycle were adjusted to achieve relatively even fluorescence signals (Table 2.1).

2.5 CONCLUSION

We have synthesized and characterized four 3’-O-modified NRTs (3’-O-N₃-dNTPs) along with four cleavable fluorescent dideoxynucleotide terminators (ddNTP-N₃ (version I)-fluorophores) and used them to produce 4-color de novo DNA sequencing data on a chip by the hybrid SBS approach. In doing so, we have combined the advantageous aspects of the Sanger dideoxy termination method and SBS sequencing approaches to sequence DNA unambiguously. A strategy to use a chemically reversible azidomethyl moiety to cap the 3’-OH group of the nucleotide, and to construct the cleavable linker to attach fluorophores to the dideoxynucleotide has been successfully implemented to allow the identification of the incorporated cleavable fluorescent dideoxynucleotide for sequence determination. After cleavage, both the fluorophore from the dideoxynucleotide and the 3’-OH capping group from the NRTs are removed. Removal of the fluorophore after the identification of the base is crucial so that it does not interfere with the fluorescent detection of the next incorporated nucleotide. Regeneration of the 3’-OH group is needed for the subsequent incorporation of the next complementary nucleotide.

With the 3’-O-modified NRTs, after cleavage of the 3’-OH capping group of the DNA extension product, there are no traces of modification, thereby regenerating a natural nucleotide at the terminal 3’-end of the growing DNA strand. Therefore, there
**Table 2.1** Volumes of solution A and B in each SBS cycle during *de novo* DNA sequencing.

<table>
<thead>
<tr>
<th>SBS Cycle</th>
<th>Solution A (µl)</th>
<th>Solution B (µl)</th>
<th>SBS Cycle</th>
<th>Solution A (µl)</th>
<th>Solution B (µl)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>7.3</td>
<td>0.2</td>
<td>17th</td>
<td>6.0</td>
<td>1.5</td>
</tr>
<tr>
<td>2nd</td>
<td>7.3</td>
<td>0.2</td>
<td>18th</td>
<td>6.0</td>
<td>1.5</td>
</tr>
<tr>
<td>3rd</td>
<td>7.3</td>
<td>0.2</td>
<td>19th</td>
<td>6.0</td>
<td>1.5</td>
</tr>
<tr>
<td>4th</td>
<td>7.3</td>
<td>0.2</td>
<td>20th</td>
<td>5.5</td>
<td>2.0</td>
</tr>
<tr>
<td>5th</td>
<td>7.2</td>
<td>0.3</td>
<td>21st</td>
<td>5.5</td>
<td>2.0</td>
</tr>
<tr>
<td>6th</td>
<td>7.2</td>
<td>0.3</td>
<td>22nd</td>
<td>5.5</td>
<td>2.0</td>
</tr>
<tr>
<td>7th</td>
<td>7.2</td>
<td>0.3</td>
<td>23rd</td>
<td>5.0</td>
<td>2.5</td>
</tr>
<tr>
<td>8th</td>
<td>7.2</td>
<td>0.3</td>
<td>24th</td>
<td>5.0</td>
<td>2.5</td>
</tr>
<tr>
<td>9th</td>
<td>7.0</td>
<td>0.5</td>
<td>25th</td>
<td>5.0</td>
<td>2.5</td>
</tr>
<tr>
<td>10th</td>
<td>7.0</td>
<td>0.5</td>
<td>26th</td>
<td>4.5</td>
<td>3.0</td>
</tr>
<tr>
<td>11th</td>
<td>7.0</td>
<td>0.5</td>
<td>27th</td>
<td>4.0</td>
<td>3.5</td>
</tr>
<tr>
<td>12th</td>
<td>7.0</td>
<td>0.5</td>
<td>28th</td>
<td>3.5</td>
<td>4.0</td>
</tr>
<tr>
<td>13th</td>
<td>6.5</td>
<td>1.0</td>
<td>29th</td>
<td>3.0</td>
<td>4.5</td>
</tr>
<tr>
<td>14th</td>
<td>6.5</td>
<td>1.0</td>
<td>30th</td>
<td>2.5</td>
<td>5.0</td>
</tr>
<tr>
<td>15th</td>
<td>6.5</td>
<td>1.0</td>
<td>31st</td>
<td>2.0</td>
<td>5.5</td>
</tr>
<tr>
<td>16th</td>
<td>6.5</td>
<td>1.0</td>
<td>32nd</td>
<td>0</td>
<td>7.5</td>
</tr>
</tbody>
</table>
will be no adverse effect on the DNA polymerase for the incorporation of the next complementary nucleotide. Second, the cleavable fluorescent dideoxynucleotides and 3′-O-modified NRTs are terminators (permanent and reversible, respectively) which allow the interrogation of each base in a serial manner, a key strategy enabling accurate determination of homopolymeric regions of DNA templates. In addition, since all the steps of the nucleotide incorporation, fluorescence detection for sequence determination, cleavage of the fluorophore and the 3′-O-azidomethyl group are performed on a DNA chip, there is no longer a need for electrophoretic DNA fragment separation as in the classical Sanger sequencing method. The hybrid SBS also allows for the addition of all 8 nucleotide substrates simultaneously to unambiguously determine DNA sequences. This ultimately reduces the number of steps needed to complete the sequencing cycle, while increasing the sequencing accuracy due to competition among the substrates in the polymerase reaction.

In our four-color hybrid SBS approach, the identity of the incorporated nucleotide is determined by the unique fluorescent emission from each of the four fluorescent dideoxynucleotides. Therefore, the ratio between the amount of 3′-O-N3-dNTPs and ddNTP-N3 (version I)-fluorophores during the polymerase reaction determines how much of the ddNTP-N3 (version I)-fluorophores are incorporated and thus the corresponding fluorescent emission strength. This ratio also determines the final read length in the hybrid SBS approach. We have experimentally determined the ratio of the two sets of the nucleotides to yield a sequencing read length of 32 bases. The ultimate read length for this hybrid SBS system will then depend on three factors: the number of starting DNA molecules on each spot of a DNA chip, the reaction efficiency, and the detection
sensitivity of the system. The reaction efficiency is a function of the sum of the incorporation yield of the modified nucleotides and the cleavage yield of the fluorophore and the 3’-OH capping group. The read-length with the Sanger sequencing methods commonly reach over 700 bps. The hybrid SBS approach described here may have the potential to reach this read-length, especially with improvements in the sensitivity of the fluorescent detection system, where single or relatively few molecules can be reliably detected.

With sequencing read length from 14 to 30 bases in the next generation of DNA sequencing systems, massively parallel digital gene expression analogous to a high-throughput SAGE$^{[25]}$ approach has been reported reaching single copy transcript sensitivity$^{[26]}$ and CHIP-Seq$^{[17-19]}$ based on sequencing tags of approximately 25 bases has led to many new discoveries in genome function and regulation. It is well established that millions of different PCR templates can be generated on a solid surface through emulsion PCR or clonal amplification.$^{[19,27]}$ Thus, future implementation of the hybrid SBS approach on a high-density bead array platform will provide a high-throughput and accurate DNA sequencing system with wide applications in genome biology and biomedical research.
REFERENCES


Chapter 3: Consecutive DNA Sequencing by Primer Reset

3.1 INTRODUCTION

The completion of the Human Genome Project\textsuperscript{[1]} has increased the need for high-throughput DNA sequencing technologies aimed at uncovering the genomic contributions to diseases. The DNA sequencing by synthesis (SBS) approach has shown great promise as a new platform for deciphering the genome.\textsuperscript{[2,3]} Fluorescent-based SBS methods have almost unlimited ability for parallelization, restricted only by the resolution of the imaging system. Sequencing read lengths of about 35 bases are sufficient for tag-based sequencing projects.\textsuperscript{[4-7]} However, re-sequencing of exons for disease mutation analysis, metagenomic sequencing, and \textit{de novo} sequencing require longer reads. Even though somewhat mitigated by the development of methods for obtaining paired reads, lengths of at least 50 bases would be desirable to assure proper assemblies, especially in highly repetitive genomes.

Recently, as described in detail in Chapter 2, we have developed a hybrid SBS approach.\textsuperscript{[8]} which integrates the advantages of the Sanger dideoxy chain terminating reaction and SBS. In this approach, four nucleotides (Fig. 3.1), modified as reversible terminators by capping the 3’-OH with a small reversible moiety so that they are still recognized by DNA polymerase as substrates to extend the DNA chain, are used in combination with a small percentage of four cleavable fluorescent dideoxynucleotides (Fig. 3.2) to perform SBS. Sequences are determined by the unique fluorescence emission of each fluorophore on the DNA products terminated by ddNTPs, while the role of the 3’-O-modified dNTPs is to further extend the DNA strand to continue the
Figure 3.1. Structures of nucleotide reversible terminators, 3'-O-R₁-dATP, 3'-O-R₁-dCTP, 3'-O-R₁-dGTP and 3'-O-R₁-dTTP. With the 3’-OH capped by a small reversible moiety R₁, these nucleotide analogues are still recognized as good substrates by DNA polymerase. Upon removing the 3’-OH capping group R₁ from the DNA extension product generated by incorporation of 3’-O-R₁-dNTPs, the polymerase reaction reinitiates to continue the sequence determination.
Figure 3.2. Structures of cleavable fluorescent dideoxynucleotide terminators, ddCTP-R$_2$-Bodipy-FL 510, ddUTP-R$_2$-R6G, ddATP-R$_2$-ROX and ddGTP-R$_2$-Cy5. Each of the four fluorophores is attached to the 5-position of pyrimidines and the 7-position of purines through a chemically cleavable linker R$_2$. After incorporation of these dideoxynucleotide analogues, the fluorophores can be removed from the DNA extension products.
determination of the DNA sequence. Upon removing the 3’-OH capping group from the DNA products generated by incorporating the 3’-O-modified dNTPs and the fluorophore from the DNA products terminated with the ddNTPs, the polymerase reaction reinitiates to continue the sequence determination (Fig. 3.3). Such incorporation, fluorescence measurement and dye removal is repeatedly conducted until the detectable fluorescence intensity is not distinguishable from background, indicating a situation in which all the elongated primers are terminated with ddNTPs. The successful implementation of this sequencing approach is essentially dependent on the read length of the target DNA template. One of the major factors that determine the read length when performing SBS is the number of available templates. Since the incorporation of fluorescently labeled dideoxynucleotide analogues into a strand of DNA permanently terminates further extensions of that template, the total number of sequenceable templates decreases after each cycle of SBS reaction. Various strategies can be used to minimize this rate of template reduction, including the utilization of a more sensitive imaging system, fluorophores with higher quantum yield, more efficient sequencing chemistry, etc. Among those, a powerful method termed “template walking” can potentially diminish the negative effect of template termination or reduction and extend the read length of SBS at least two to three-fold.

3.2 EXPERIMENTAL RATIONALE AND OVERVIEW

The fundamental rationale behind this “template walking” strategy is the removal of the sequenced strand and reattachment of the original primer to allow the extension of the primer with a combination of natural and modified nucleotides to the end of the first
Figure 3.3. The hybrid DNA sequencing approach between the Sanger dideoxy chain terminating reaction and sequencing by synthesis. In this approach, four nucleotides (3'-O-R1-dNTPs) modified as reversible terminators by capping the 3'-OH with a small reversible moiety R1 so that they are still recognized by DNA polymerase as substrates, are used in combination with a small percentage of four cleavable fluorescent dideoxynucleotides (ddNTP-R2-fluorophores) to perform SBS. DNA sequences are determined by the unique fluorescence emission of each fluorophore on the DNA products terminated by ddNTPs. Upon removing the 3'-OH capping group R1 from the DNA products generated by incorporating the 3'-O-R1-dNTPs, and the cleavage of the R2 linker to remove the fluorophore from the DNA products terminated with the ddNTPs, the polymerase reaction reinitiates to continue the sequence determination.
round sequence so that SBS can be carried out from that point. Since the sequencing primers are stripped away, including those terminated with ddNTPs, all the templates become available for the next sequencing round. Given that “template walking” is carried out with either natural or 3’-O-modifed nucleotides, the subsequent round of SBS is performed on nascent DNA strands for maximum read-length. In general, three steps are involved in this approach: 1) annealing of the original primer and extending the chain to the end of the previous round sequence, 2) performing SBS to further extend the sequencing read-length, 3) denaturing the elongated primer to recover a single-stranded DNA for another round of primer annealing. These steps are carried out repeatedly until the target DNA is sequenced in its entirety (Fig. 3.4). The advantage of primer resetting lies in its ability to restore all the templates after the denaturation step, including those that are terminated with ddNTPs, so the next cycle of SBS can restart with potentially the same amount of sequenceable DNA as the previous round. This strategy is not sensitive to the type of library (genomic, cDNA or other), to the method of amplification prior to sequencing (spotting of clones, ePCR, or polony PCR), or to the mode of sequencing (hybrid SBS, pyrosequencing, or SBS). Hence it has wide applications in sequencing technologies.

Upon annealing a sequencing primer to the linear DNA template immobilized on a microarray chip, the first round SBS is performed with the hybrid sequencing approach described above. Immediately after the first round of SBS, all of the elongated primers terminated with ddNTPs are removed from the template by denaturing. The templates are freed again and available for further sequencing reactions. To fill the gap between the first and second rounds of SBS, the original starting primer is annealed to the template
Figure 3.4. Consecutive rounds of DNA sequencing by template “walking”. Upon ligation of universal primers A and B to both ends of each DNA template, a DNA library is prepared (1). Different DNA templates are immobilized on a PEG (polyethylene glycol) functionalized surface to initiate the sequencing reaction (2). The sequencing primer is then annealed to the DNA template (3). A first round of SBS extends the sequencing primer to produce the maximal read-length (4). After denaturing the sequencing primers, including those terminated with ddNTP analogues, the original sequencing primer is reattached (5). Then unmodified or 3'-O-modified nucleotides are used to extend the primer approximately to the end of the first round sequence (6). Following that, the second round of SBS is performed to further extend the read-length.
and enzymatic incorporation is conducted using three dNTPs and the fourth nucleotide reversible terminator as substrates (Fig. 3.5). Primer elongation will only stop once a nucleotide reversible terminator is incorporated. After incorporation, a specific chemical reaction is applied to regenerate the 3’-OH which ensures consecutive incorporation in the next cycle. Repeated cycles of such incorporation and cleavage will extend the sequencing primer to the end of the first round sequence. Then the second round of SBS is conducted with the combination of nucleotide reversible terminators and cleavable fluorescent dideoxynucleotide terminators as incorporation substrates. Another continuous stretch of bases on the template can be continuously revealed, leading to the doubling of the original read length. The SBS-walking-SBS process is repeated to generate maximum read length.

3.3 RESULTS AND DISCUSSION

3.3.1 Polymerase extension using dATP, dCTP, dTTP and 3′-O-N3-dGTP and characterization by MALDI-TOF MS.

A critical requirement for using the “template walking” strategy to consecutively sequence DNA is a suitable chemical moiety to cap the 3’-OH of one of the nucleotides such that it temporarily terminates the polymerase reaction. A stepwise addition of separate natural nucleotides has inherent difficulties in achieving high incorporation accuracy and minimizing the overall extension time. Using all four natural nucleotides simultaneously during the polymerase reaction will extend the primer to the end. Thus the primer extension with three natural nucleotides and one 3′-O-modified nucleotide reversible terminator increases the nucleotide incorporation accuracy and reduces the
**Figure 3.5** Consecutive rounds of DNA sequencing by template “walking”. After the first sequencing round with the hybrid SBS approach, the extended primer is denatured. Upon reattachment of the original primer, three normal nucleotides are combined with another 3'-O-modified nucleotide to extend the primer to the end of the first round sequence. Then the second sequencing round is performed to further extend the read-length.
number of extension reactions needed. Upon removing the 3’-OH capping moiety from the DNA extension product, the primer consisting of only natural bases does not interfere with the ability of the polymerase to efficiently incorporate the subsequent nucleotides. Our previous research efforts has firmly established the molecular level strategy to modify the nucleotides by capping their 3’-OH with a small reversible moiety for SBS. Building on our successful 3’-O-modification approaches, we explored the “template walking” strategy by using dATP, dCTP, dTTP and 3’-O-N₃-dGTP (Fig. 2.2) to elongate the primer to the suitable position. In human genomic DNA, the homopolymeric regions consisting of AT pairs are more common than the ones with GC pairs. Additionally, 3’-O-modified dGTP incorporates with the best inefficiency among the four chemically modified nucleotides. Thus the combination of 3’-O-N₃-dGTP with the other three natural nucleotides is likely to further decrease the number of extensions.

To verify that dATP, dCTP, dTTP and 3’-O-N₃-dGTP incorporate accurately in a base-specific manner, and extend the primer efficiently until the polymerase reaction is temporarily terminated by the incorporation of 3’-O-N₃-dGTP, three continuous DNA extension and cleavage reactions were carried out in solution. This allowed the isolation of the DNA product at each step for detailed molecular structure characterization as shown in Fig. 3.6. The first extension product 5’-primer-G-N₃-3’ (1) was desalted and analyzed using MALDI-TOF MS (Fig. 3.6A). This product was then incubated in aqueous TCEP solution to remove the azidomethyl moiety to yield the cleavage product (2) with a free 3’-OH group, which was also analyzed using MALDI-TOF MS (Fig. 3.6B). As can be seen from Fig. 3.6A, the MALDI-TOF MS spectrum consists of a distinct peak corresponding to the DNA extension product 5’-primer-G-N₃-3’ (1) (m/z
Figure 3.6 The polymerase extension scheme (left) and MALDI-TOF MS spectra of the three consecutive extension products and their cleavage products (right) using the combination of dATP, dCTP, dTTP and 3’-O-N3-dGTP. Primer extended with 3’-O-N3-dGTP (1) (A), and its cleavage product 2 (B); product 2 extended with 3’-O-N3-dGTP (3) (C), and its cleavage product 4 (D); product 4 extended with dATP, dCTP and 3’-O-N3-dGTP (5) (E), and its cleavage product 6 (F).
6,515), which confirms that the 3’-O-N₃-dGTP is incorporated base specifically by DNA polymerase into a growing DNA strand and terminates the polymerase reaction temporarily. Fig. 3.6B shows the cleavage result on the DNA extension product. The extended DNA mass peak at \( m/z \) 6,515 completely disappeared while the peak corresponding to the cleavage product 5’-primer-G-3’ (2) appears as the sole dominant peak at \( m/z \) 6,460, which establishes that TCEP incubation completely cleaves the 3’-O-azidomethyl group with high efficiency. The next extension reaction was carried out using this cleaved product, which now has a free 3’-OH group, as a primer to yield a second extension product, 5’-primer-GG-N₃-3’ (3) (\( m/z \) 6,844, Fig. 3.6C). As described above, the extension product (3) was cleaved to generate product (4) for further MS analysis yielding a single peak at \( m/z \) 6,460 (Fig. 3.6D). The third extension reaction to yield 5’-primer-GGACG-N₃-3’ (5) (\( m/z \) 7,775 Fig. 3.6E) with three nucleotides incorporated instead of one and its cleavage to yield product (6) (\( m/z \) 7,720 Fig. 3.6F) were similarly carried out. These results demonstrate that the four nucleotides, dATP, dCTP, dTTP and 3’-O-N₃-dGTP, were incorporated base-specifically into the growing DNA strand in a polymerase reaction and the continuous primer extension was efficient until terminated by the incorporation of 3’-O-N₃-dGTP.

3.3.2 Four-color DNA sequencing with a “walked” DNA template on a chip.

To verify that the combination of dATP, dCTP, dTTP and 3’-O-N₃-dGTP efficiently and accurately extends the DNA primer immobilized on a microarray chip, three continuous DNA extension and cleavage reactions were performed on the surface. The subsequent DNA sequencing with the hybrid SBS approach⁸ characterized the
primer extension efficiency and accuracy. In our “template walking” approach, the role of
the three natural nucleotides is to further extend the DNA strand, whereas 3’-O-N3-dGTP
is used to temporarily terminate the primer extension reaction and increase the nucleotide
incorporation accuracy due to the competition among the substrates. Therefore, the ratio
between the natural nucleotides and 3’-O-N3-dGTP determines whether the primer will
be extended efficiently in a base-specific manner. With a 3’-OH capped by an
azidomethyl group, 3’-O-N3-dGTP incorporates slower than the other three nucleotides.
Thus a relatively high concentration of 3’-O-N3-dGTP, compared with the natural
nucleotides, is required to maximize the efficiency and accuracy of the primer extension
reaction.

The scheme of the “template walking” strategy on a DNA chip is shown in Fig.
3.7A. Initiated by extending the self-priming DNA by using a solution consisting of
dATP, dCTP, dTTP, 3’-O-N3-dGTP and 9°N DNA polymerase, the polymerase reaction
elongated the DNA strand until the 3’-O-N3-dGTP was incorporated. To negate any
unterminated DNA strands, 3’-O-N3-dGTP and 9°N DNA polymerase were added to
synchronize the DNA strands that retain a free 3’-OH group. Treatment of the DNA
extension products with TCEP cleaved the azidomethyl group[8] and regenerated the
3’-OH, which allowed the subsequent primer extension. The second and the third
incorporation, synchronization, and cleavage reactions were carried out in a similar
manner to extend the primer by 32 bases (Fig. 3.7A). To achieve the high signal to
background ratio and the sufficient read-length in the subsequent SBS, almost all the
primers are required to be extended to the proposed position.

The de novo DNA sequencing reaction following the “template walking” was
Figure 3.7. (A) A general “template walking” scheme followed by hybrid SBS. The polymerase reaction with dATP, dCTP, dTTP and 3'-O-N$_3$-dGTP extends the primer until it reaches the next C on the DNA template. After regenerating the 3'-OH group by using TCEP, the second and the third extension and cleavage reactions are performed in a similar manner. Then a subsequent round of hybrid SBS is carried out to generate the four-color DNA sequencing raw data (B).
performed with our previously developed hybrid SBS approach. The entire process of incorporation of 3’-O-N\textsubscript{3}-dNTP/ddNTP-N\textsubscript{3} (version I)-fluorophore, synchronization by 3’-O-N\textsubscript{3}-dNTP and cleavage with TCEP was performed multiple times to unambiguously identify 13 bases from the four-color raw fluorescence data without any processing (Fig. 3.7B). The homopolymeric region of the DNA template was accurately determined. These results demonstrate that the DNA primer immobilized on a chip was accurately and efficiently elongated to the proposed position by multiple cycles of extension with dATP, dCTP, dTTP and 3’-O-N\textsubscript{3}-dGTP and subsequent cleavage. Thus the combination of the four nucleotides meets the key requirements necessary for performing the consecutive DNA sequencing by primer reset on a chip.

### 3.3.3 Consecutive DNA sequencing by primer reset on a chip.

In our consecutive DNA sequencing approach, the sequencing reactions were carried out on a linear DNA template immobilized on a chip. Upon annealing the sequencing primer to the linear template, the first round of sequence determination was conducted with the hybrid SBS strategy. The principle advantage of this hybridization-based primer/template complex is the simple removal of the extended primer, compared to a covalently bounded self-primed DNA template. This permits the next cycle of SBS to have potentially the same amount of sequenceable templates as the previous round.

The scheme of the consecutive DNA sequencing on a chip approach is shown in Fig. 3.8A. The first round \textit{de novo} sequencing reaction on the chip was initiated by extending the annealed primer using a solution consisting of the four 3’-O-N\textsubscript{3}-dNTPs, the
Figure 3.8. Consecutive rounds of DNA sequencing by primer reset on a chip. (A) Consecutive DNA sequencing scheme. The first round sequencing is carried out with the hybrid SBS approach. Upon the removal of the extended primer and the subsequent annealing of the original primer, the polymerase reaction with dATP, dCTP, dTTP and 3'-O-N3-dGTP extends the primer until the incorporation of 3'-O-N3-dGTP. After regenerating the 3'-OH group by using TCEP, the second extension and cleavage reactions are performed in a similar manner. Then a second round of sequencing is performed with the hybrid SBS approach to interrogate the identity of the consecutive bases. (B) A plot (four-color sequencing data) of raw fluorescence emission intensity vs. the progress of sequencing extension during the first and the second rounds of SBS.
four ddNTP-N₃ (version I)-fluorophores and 9°N DNA polymerase. To minimize the spontaneous dissociation of the primer from the template, washing buffer containing a relatively high concentration of salt is required. After the identity of the incorporated nucleotide is determined by the unique fluorescent emission, a synchronization reaction mixture consisting of just the four 3’-O-N₃-dNTPs was used along with the 9°N DNA polymerase to extend any remaining priming strands that retain a free 3’-OH group. Then the surface was immersed in a TCEP solution to cleave the fluorophore and the 3’ azidomethyl capping group to reinitiate the sequencing reaction. This entire process of incorporation, identification, synchronization and cleavage was performed multiple times to unambiguously identify 15 bases on a linear DNA template (Fig. 3.8B).

To reset the primer for the next round of SBS, the extended primers at the end of the first sequencing round were denatured and removed from the linear template immobilized on the surface. Upon annealing the original sequencing primer, “template walking” was initiated by extending the annealed primer using a solution consisting of dATP, dCTP, dTTP, 3’-O-N₃-dGTP and 9°N DNA polymerase (Fig. 3.8A). The incorporation of three natural nucleotides generates a free hydroxyl group at the 3’ end of the primer, which allows the next nucleotide incorporation; whereas the primer extension with 3’-O-N₃-dGTP yields a 3’-OH capped DNA strand, which temporarily terminates the DNA extension reaction. Thus the combination of dATP, dCTP, dTTP and 3’-O-N₃-dGTP during the polymerase reaction extended the DNA primer by C, A, T three bases and the incorporation of the subsequent 3’-O-N₃-dGTP terminated the primer elongation. To negate any unterminated DNA strands, 3’-O-N₃-dGTP and 9°N DNA polymerase were added to synchronize the DNA strands that retain a free 3’-OH group.
Treatment of the DNA extension products with TCEP cleaved the azidomethyl group\(^8\) and regenerated the 3'-OH, which allowed the subsequent primer extension. The second incorporation, synchronization, and cleavage reactions were carried out in a similar manner to extend the primer to the end of the first sequencing round (Fig. 3.8A). During this “template walking” process, the primer was extended by either the three natural nucleotides or 3'-O-N\(_3\)-dGTP, which reverted to a natural nucleotide after cleavage. Therefore, this extended primer only consisting of natural nucleotides should have no deleterious effect on the ability of the polymerase to bind and to incorporate the subsequent nucleotides in the second round of SBS. With the hybrid SBS strategy as described above, the second round \textit{de novo} sequencing unambiguously identified another 10 bases with no errors from the four-color raw fluorescence data (Fig. 3.8B). Combined with the first round sequencing data, in total 25 bases were accurately identified with our consecutive DNA sequencing by primer reset approach.

\section*{3.4 MATERIALS AND METHODS}

\subsection*{3.4.1 Polymerase extension using dATP, dCTP, dTTP and 3'-O-N\(_3\)-dGTP and characterization by MALDI-TOF MS.}

We characterized the “template walking” strategy in solution phase by performing three continuous DNA-extension reactions sequentially using a DNA template (5'-TAC-CCG-GAG-GCC-AAG-TAC-GGC-GGG-TAC-CTT-GAC-AAT-GTG-TAC-ATC-AAC-ATC-ACC-TAC-CAC-CAT-GTC-AGT-CAC-GGT-TGG-ATC-CTC-TAT-TGT-GTC-CGG-G-3') and a DNA primer (5'-GTT-GAT-GTA-CAC-ATT-GTC-AA-3'). The five nucleotides in the template immediately adjacent to the annealing site of
the primer are 3'-CCTGC-5', which allows the evaluation of the extension and cleavage efficiency of the “template walking” strategy. First, a polymerase extension reaction using a pool of dATP, dCTP, dTTP and 3’-O-N$_3$-dGTP along with the DNA template and primer was performed producing a DNA extension product. The reaction mixture for this, and all subsequent extension reactions, consisted of 100 pmol of the DNA template, 80 pmol of the DNA primer, 150 pmol of dATP, dCTP and dTTP, 250 pmol of 3’-O-N$_3$-dGTP, 1X Thermopol II reaction buffer (New England Biolabs), 40 nmol of MnCl$_2$ and 1 unit of 9°N DNA polymerase (exo-) A485L/Y409V (New England Biolabs) in a total reaction volume of 20 µl. The reaction consisted of incubation at 94°C for 5 minutes, 4°C for 5 minutes and 65°C for 20 minutes. Subsequently, the extension product was desalted by using a ZipTip and analyzed by Voyager DE™ MALDI-TOF mass spectrometry (Applied Biosystems). For cleavage, the desalted DNA extension product bearing the 3’-O-azidomethyl group was first resuspended with 5 µl of 50 mM EDTA solution to quench the polymerase activity. This DNA solution was then mixed with 10 µl of 225 mM TCEP solution (pH 9.0) and incubated at 65°C for 15 minutes to yield a cleaved DNA product which was characterized by MALDI-TOF MS. The DNA product with the 3’-OH group regenerated was purified using an Oligonucleotide Purification Cartridge (Applied Biosystems) and used as a primer for a second extension reaction using dATP, dCTP, dTTP and 3’-O-N$_3$-dGTP. The second extended DNA product was then purified with a ZipTip and cleaved as described above. The third extension was carried out in a similar manner using the previously extended and cleaved product as the primer.
3.4.2 Four-color DNA sequencing with a “walked” DNA template on a chip.

A DNA chip was constructed by immobilizing a 5’-amino-modified looped oligonucleotide on a CodeLink microarray slide (GE Healthcare). The 5’-amino-labeled self-priming DNA template 5’-NH2-CAC-TCA-CAT-ATG-TTT-TTT-AGC-TTT-TTT-AAT-TTC-ATG-ATG-TTG-TTG-CAT-GCG-ACT-TAA-GGC-GCT-TGC-GCC-TTA-AGT-CG-3’ was purchased from IDT (Coralville, IA). The DNA template was dissolved at 40 µM in 50 mM sodium phosphate buffer, pH 8.5 and spotted using a SpotArray 72 arraying robot (Perkin Elmer) onto high density CodeLink microarray slides (GE Healthcare). After spotting, the slides were incubated at ambient temperature (~ 24°C) for 20 hours in a humid chamber containing saturated sodium chloride solution (~ 75% humidity) to allow for 5’-tethering of the spotted amino-modified DNA templates to the slide surface functionalized with succinimide ester groups. After incubation the slides were removed from the humid chamber and stored in a vacuum dessicator at room temperature until further use.

Ten microliters of a solution consisting of dATP (30 pmol), dTTP (30 pmol), dCTP (30 pmol), 3’-O-N3-dGTP (50 pmol), 1 unit of 9°N DNA polymerase(exo-) A485L/Y409V, 20 nmol of MnCl2 and 1X Thermopol II reaction buffer was spotted on the surface of the chip, where the self-primed DNA moiety was immobilized. The nucleotides complementary to the DNA template were allowed to incorporate into the primer at 65°C for 15 minutes. To synchronize any unterminated templates, an extension solution consisting of 50 pmol of 3’-O-N3-dGTP, 1 unit of 9°N DNA polymerase(exo-) A485L/Y409V, 20 nmol of MnCl2 and 1X Thermopol II reaction buffer was spotted in the same position and incubated at 65°C for 15 minutes. After washing the chip with
SPSC buffer containing 0.1% Tween-20 for 1 minute, the surface was rinsed with dH$_2$O. To perform the cleavage, the glass chip and the spot where the self-primed DNA moiety is immobilized was placed inside a chamber filled with 100 mM TCEP (pH 9.0) and incubated at 65°C for 10 minutes. After washing the chip with SPSC buffer containing 0.1% Tween-20 for 1 minute, the surface was rinsed with dH$_2$O and dried briefly. The second and third DNA extension, synchronization and cleavage reactions were carried out in a similar manner.

Then ten microliters of a solution consisting of ddCTP-N$_3$ (version I)-Bodipy-FL-510 (10 fmol), ddUTP-N$_3$ (version I)-R6G (20 fmol), ddATP-N$_3$ (version I)-ROX (40 fmol), ddGTP-N$_3$ (version I)-Cy5 (20 fmol), 3’-O-N$_3$-dCTP (22 pmol), 3’-O-N$_3$-dTTP (22 pmol), 3’-O-N$_3$-dATP (22 pmol), 3’-O-N$_3$-dGTP (4 pmol), 1 unit of 9°N DNA polymerase(exo-) A485L/Y409V, 20 nmol of MnCl$_2$ and 1X Thermopol II reaction buffer was spotted on the surface of the chip, where the extended DNA moiety was immobilized. The nucleotide complementary to the DNA template was allowed to incorporate into the primer at 65°C for 15 minutes. To synchronize any unincorporated templates, an extension solution consisting of 38 pmol each of 3’-O-N$_3$-dTTP, 3’-O-N$_3$-dATP, 3’-O-N$_3$-dGTP, and 75 pmol of 3’-O-N$_3$-dCTP, 1 unit of 9°N DNA polymerase(exo-) A485L/Y409V, 20 nmol of MnCl$_2$ and 1X Thermopol II reaction buffer was spotted at the same position and incubated at 65°C for 15 minutes. After washing the chip with SPSC buffer containing 0.1% Tween-20 for 1 minute, the surface was rinsed with dH$_2$O, dried briefly and then scanned with a 4-color ScanArray Express scanner (Perkin–Elmer Life Sciences) to detect the fluorescence signal. The 4-color scanner is equipped with four lasers with excitation wavelengths of 488, 543, 594, and
633 nm and emission filters centered at 522, 570, 614, and 670 nm. To perform the cleavage, the glass chip and the spot where the self-primed DNA moiety is immobilized was placed inside a chamber filled with 100 mM TCEP (pH 9.0) and incubated at 65°C for 10 minutes. After washing the surface with dH₂O, the chip was scanned again to compare the intensity of fluorescence after cleavage with the original fluorescence intensity. This process was followed by the next polymerase extension reaction using the 3’-O-N₃-dNTP/ddNTP-N₃ (version I)-fluorophore solution with the subsequent synchronization, washing, fluorescence detection, and cleavage processes performed as described above. To obtain de novo DNA sequencing data on a DNA template immobilized on a chip, the SBS cycle was repeated multiple times using the combination mixture of solution A consisting of 3’-O-N₃-dCTP (3 μM), 3’-O-N₃-dTTP (3 μM), 3’-O-N₃-dATP (3 μM) and 3’-O-N₃-dGTP (0.5 μM) and solution B consisting of ddCTP-N₃ (version I)-Bodipy-FL-510 (50 nM), ddUTP-N₃ (version I)-R6G (100 nM), ddATP-N₃ (version I)-ROX (200 nM) and ddGTP-N₃ (version I)-Cy5 (100 nM) in each polymerase extension reaction. The volumes of solution A and B in each SBS cycle were adjusted to achieve relatively even fluorescence signals (Table 3.1).

3.4.3 Consecutive DNA sequencing by primer reset on a chip.

A DNA chip was constructed by immobilizing a 5’-amino-modified linear oligonucleotide on a CodeLink microarray slide (GE Healthcare). The 5’-amino-labeled linear DNA template 5’-NH₂-CCT-TTA-ATT-TTG-GCT-TTT-AAT-TGG-CTT-GCT-TTG-GTT-AAC-TTG-GTT-GTT-GCA-TGC-CCA-TGC-GAG-TGC-GAG-TGC-ACG-TGG-CGC-AGC-AGG-TCA-3’ purchased from IDT (Coralville, IA) was
immobilized on a microarray slide using a similar procedure as that for the looped DNA
template. Then 10 µl of 1X Thermo Pol buffer was spotted on the surface of the chip,
where the linear DNA template was immobilized. After incubation at 65°C for 10
minutes, the buffer was replaced with 10 µl of annealing mixture consisting of 35 pmol
of the primer (5’-TGA-CCT-GCT-GCG-CCA-CGT-GCA-CTC-GCA-CTC-GCA-
TGG-G-3’) (IDT, IA) and 1X Thermo Pol buffer. The primer was allowed to anneal to
the linear template at 65°C for 30 minutes. To eliminate any unannealed DNA template, a
fresh annealing mixture was spotted in the same position and incubated at 65°C for 30
minutes. Then the chip was washed with the 1X Thermo Pol buffer in preparation for
SBS.

Ten microliters of a solution consisting of ddCTP-N₃ (version I)-Bodipy-FL-510
(10 fmol), ddUTP-N₃ (version I)-R6G (20 fmol), ddATP-N₃ (version I)-ROX (40 fmol),
ddGTP-N₃ (version I)-Cy5 (20 fmol), 3’-O-N₃-dCTP (22 pmol), 3’-O-N₃-dTTP (22 pmol),
3’-O-N₃-dATP (22 pmol), 3’-O-N₃-dGTP (4 pmol), 1 unit of 9⁰N DNA polymerase(exo-)
A485L/Y409V, 20 nmol of MnCl₂ and 1X Thermopol II reaction buffer was spotted on
the surface of the chip, where the primed DNA template was immobilized. The
nucleotide complementary to the DNA template was allowed to incorporate into the
primer at 65°C for 15 minutes. To synchronize any unincorporated templates, an
extension solution consisting of 38 pmol each of 3’-O-N₃-dTTP, 3’-O-N₃-dATP,
3’-O-N₃-dGTP, and 75 pmol of 3’-O-N₃-dCTP, 1 unit of 9⁰N DNA polymerase(exo-)
A485L/Y409V, 20 nmol of MnCl₂ and 1X Thermopol II reaction buffer was spotted in
the same position and incubated at 65°C for 15 minutes. After washing the chip with
SPSC buffer containing 0.1% Tween-20 for 1 minute, the surface was rinsed with 1X
Thermo Pol buffer, dried briefly and then scanned with a 4-color ScanArray Express scanner (Perkin–Elmer Life Sciences) to detect the fluorescence signal. The 4-color scanner is equipped with four lasers with excitation wavelengths of 488, 543, 594, and 633 nm and emission filters centered at 522, 570, 614, and 670 nm. To perform the cleavage, the glass chip and the spot where the DNA moiety is immobilized was placed inside a chamber filled with 100 mM TCEP (pH 9.0) and incubated at 65°C for 10 minutes. After washing the surface with 1X Thermo Pol buffer, the chip was scanned again to compare the intensity of fluorescence after cleavage with the original fluorescence intensity. This process was followed by the next polymerase extension reaction using the 3’-O-N3-dNTP/ddNTP-N3 (version I)-fluorophore solution with the subsequent synchronization, washing, fluorescence detection, and cleavage processes performed as described above. To obtain the first round of de novo DNA sequencing data on a DNA template immobilized on a chip, the SBS cycle was repeated multiple times using the combination mixture of solution A consisting of 3’-O-N3-dCTP (3 µM), 3’-O-N3-dTTP (3 µM), 3’-O-N3-dATP (3 µM) and 3’-O-N3-dGTP (0.5 µM) and solution B consisting of ddCTP-N3 (version I)-Bodipy-FL-510 (50 nM), ddUTP-N3 (version I)-R6G (100 nM), ddATP-N3 (version I)-ROX (200 nM) and ddGTP-N3 (version I)-Cy5 (100 nM) in each polymerase extension reaction. The volumes of solution A and B in each SBS cycle were adjusted to achieve relatively even fluorescence signals (Table 3.2 A).

To denature all the extended primers, the glass chip and the spot where the DNA template is immobilized was placed inside a chamber filled with dH2O and incubated 95°C for 15 minutes. The original primer was annealed using a similar procedure as that
Table 3.1 Volumes of solution A and B in each SBS cycle during de novo DNA sequencing.

<table>
<thead>
<tr>
<th>SBS Cycle</th>
<th>Solution A (µl)</th>
<th>Solution B (µl)</th>
<th>SBS Cycle</th>
<th>Solution A (µl)</th>
<th>Solution B (µl)</th>
<th>SBS Cycle</th>
<th>Solution A (µl)</th>
<th>Solution B (µl)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>7.3</td>
<td>0.2</td>
<td>6th</td>
<td>6.7</td>
<td>0.8</td>
<td>11th</td>
<td>5.4</td>
<td>2.1</td>
</tr>
<tr>
<td>2nd</td>
<td>7.2</td>
<td>0.3</td>
<td>7th</td>
<td>6.5</td>
<td>1.0</td>
<td>12th</td>
<td>5.1</td>
<td>2.4</td>
</tr>
<tr>
<td>3rd</td>
<td>7.1</td>
<td>0.4</td>
<td>8th</td>
<td>6.3</td>
<td>1.2</td>
<td>13th</td>
<td>4.5</td>
<td>3.0</td>
</tr>
<tr>
<td>4th</td>
<td>7.0</td>
<td>0.5</td>
<td>9th</td>
<td>6.0</td>
<td>1.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5th</td>
<td>6.8</td>
<td>0.7</td>
<td>10th</td>
<td>5.7</td>
<td>1.8</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3.2 Volumes of solution A and B in each SBS cycle during the first round (A) and the second round (B) of de novo DNA sequencing.

A

<table>
<thead>
<tr>
<th>SBS Cycle</th>
<th>Solution A (µl)</th>
<th>Solution B (µl)</th>
<th>SBS Cycle</th>
<th>Solution A (µl)</th>
<th>Solution B (µl)</th>
<th>SBS Cycle</th>
<th>Solution A (µl)</th>
<th>Solution B (µl)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>7.3</td>
<td>0.2</td>
<td>6th</td>
<td>7.1</td>
<td>0.4</td>
<td>11th</td>
<td>6.0</td>
<td>1.5</td>
</tr>
<tr>
<td>2nd</td>
<td>7.3</td>
<td>0.2</td>
<td>7th</td>
<td>6.8</td>
<td>0.7</td>
<td>12th</td>
<td>5.8</td>
<td>1.7</td>
</tr>
<tr>
<td>3rd</td>
<td>7.2</td>
<td>0.3</td>
<td>8th</td>
<td>6.8</td>
<td>0.7</td>
<td>13th</td>
<td>5.4</td>
<td>2.1</td>
</tr>
<tr>
<td>4th</td>
<td>7.2</td>
<td>0.3</td>
<td>9th</td>
<td>6.7</td>
<td>0.8</td>
<td>14th</td>
<td>5.0</td>
<td>2.5</td>
</tr>
<tr>
<td>5th</td>
<td>7.1</td>
<td>0.4</td>
<td>10th</td>
<td>6.5</td>
<td>1.0</td>
<td>15th</td>
<td>4.5</td>
<td>3.0</td>
</tr>
</tbody>
</table>

B

<table>
<thead>
<tr>
<th>SBS Cycle</th>
<th>Solution A (µl)</th>
<th>Solution B (µl)</th>
<th>SBS Cycle</th>
<th>Solution A (µl)</th>
<th>Solution B (µl)</th>
<th>SBS Cycle</th>
<th>Solution A (µl)</th>
<th>Solution B (µl)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>7.3</td>
<td>0.2</td>
<td>5th</td>
<td>6.5</td>
<td>1.0</td>
<td>9rd</td>
<td>5.0</td>
<td>2.5</td>
</tr>
<tr>
<td>2nd</td>
<td>7.2</td>
<td>0.3</td>
<td>6th</td>
<td>6.2</td>
<td>1.3</td>
<td>10th</td>
<td>4.3</td>
<td>3.2</td>
</tr>
<tr>
<td>3rd</td>
<td>7.0</td>
<td>0.5</td>
<td>7th</td>
<td>5.9</td>
<td>1.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4th</td>
<td>6.8</td>
<td>0.7</td>
<td>8th</td>
<td>5.5</td>
<td>2.0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
for the first sequencing round. Then ten microliters of a solution consisting of dATP (30 pmol), dTTP (30 pmol), dCTP (30 pmol), 3’-O-N3-dGTP (50 pmol), 1 unit of 9°N DNA polymerase(exo-) A485L/Y409V, 20 nmol of MnCl2 and 1X Thermopol II reaction buffer was spotted on the surface of the chip, where the primed DNA moiety was immobilized. The nucleotides complementary to the DNA template were allowed to incorporate into the primer at 65°C for 15 minutes. To synchronize any unterminated templates, an extension solution consisting of 50 pmol 3’-O-N3-dGTP, 1 unit of 9°N DNA polymerase(exo-) A485L/Y409V, 20 nmol of MnCl2 and 1X Thermopol II reaction buffer was placed at the same spot and incubated at 65°C for 15 minutes. After washing the chip with SPSC buffer containing 0.1% Tween-20 for 1 minute, the surface was rinsed with 1X Thermo Pol buffer. To perform the cleavage, the glass chip and the spot where the primed DNA moiety is immobilized was placed inside a chamber filled with 100 mM TCEP (pH 9.0) and incubated at 65°C for 10 minutes. After washing the chip with SPSC buffer containing 0.1% Tween-20 for 1 minute, the surface was rinsed with 1X Thermo Pol buffer and dried briefly. The second DNA extension, synchronization and cleavage were carried out in a similar manner.

To obtain the second round of de novo DNA sequencing data on a DNA template immobilized on a chip, the SBS cycle carried out in a similar way as the first round SBS was repeated multiple times using the combination mixture of solution A consisting of 3’-O-N3-dCTP (3 µM), 3’-O-N3-dTTP (3 µM), 3’-O-N3-dATP (3 µM) and 3’-O-N3-dGTP (0.5 µM) and solution B consisting of ddCTP-N3 (version I)-Bodipy-FL-510 (50 nM), ddUTP-N3 (version I)-R6G (100 nM), ddATP-N3 (version I)-ROX (200 nM) and ddGTP-N3 (version I)-Cy5 (100 nM) in each polymerase extension
reaction. The volumes of solution A and B in each SBS cycle were adjusted to achieve relatively even fluorescence signals (Table 3.2 B).

3.5 CONCLUSION

We have demonstrated the feasibility of using the combination of dATP, dCTP, dTTP and 3′-O-N₃-dGTP to perform “template walking” during consecutive de novo DNA sequencing on a chip with the hybrid SBS approach. In our consecutive DNA sequencing strategy, the number of extension and cleavage cycles during “template walking” determines the DNA sequence interrogated in different sequencing rounds. Therefore, the sequencing data obtained in continuous sequencing rounds can be controlled to be adjacent to each other or have minimum overlaps. This consecutive DNA sequencing approach extends the sequencing read-length continuously and does not require much bioinformatic processing, unlike other paired-end approaches, which identify the sequence of both ends of the template and leave the sequence unknown in the middle. Additionally, extended with either natural nucleotides or 3′-O-N₃-dGTP, which reverts to a natural nucleotide upon cleavage, the starting primers prior to SBS in every sequencing round have no traces of modification. Thus they will not interfere with the fidelity and efficiency of the DNA polymerase for the subsequent round of hybrid SBS.

We have experimentally determined the ratio of 3′-O-N₃-dGTP and the other three natural nucleotides to extend the primer and fill the gap between the first and the second sequencing rounds. The ultimate read-length of this consecutive sequencing strategy obviously depends on two factors: the read-length of each sequencing round and the number of sequencing rounds. The read-length of the pyrosequencing method[13,14]
commonly reaches more than 200 base pairs. The consecutive DNA sequencing approach described here may have the potential to extend this read-length several folds.

Re-sequencing of exons for disease gene discovery, metagenomic sequencing, and de novo sequencing, especially in highly repetitive genomes, requires optimal read-lengths of more than 100 base pairs, which has not been achieved consistently in most of the massively parallel sequencing systems.\textsuperscript{[15-17]} Our consecutive sequencing approach is not sensitive to the type of library, to the method of amplification prior to sequencing, or the mode of sequencing. Thus, further implementation of this consecutive sequencing approach on a high-density array platform will provide a high-throughput and long read-length DNA sequencing system with wide applications in genome biology and biomedical research.
REFERENCES


Part II. Design and Synthesis of Chemically Cleavable Biotinylated Dideoxynucleotides for DNA Analysis with Mass Spectrometry
4.1 INTRODUCTION

With the completion of the human genome sequence map,\textsuperscript{[1]} the ability to resequence candidate genes rapidly and accurately will contribute greatly to the understanding of the molecular basis of disease and the development of new therapeutics. The current state-of-the-art DNA sequencing technology, such as that used in the Human Genome Project, uses capillary electrophoresis to separate the DNA fragments terminated by the incorporation of fluorescent dideoxynucleotides.\textsuperscript{[2-5]} Although this technology addresses the read-length requirements of large scale DNA sequencing projects, the accuracy required for detecting genetic variations needs to be improved for genome research and clinical diagnosis. For example, electrophoresis-based DNA sequencing approaches have inherent difficulties to unambiguously identify many heterozygous bases. Additionally, the problem of compression\textsuperscript{[6,7]} in GC rich regions can lead to poor resolution in DNA sequencing data. Furthermore, low signal to background ratio is usually obtained for the first few bases after the primer site owing to the excess dye-labeled primers or dye-labeled terminators, which makes it difficult to decipher these bases.

A variety of novel DNA sequencing technologies have been developed to address these limitations. In particular, DNA analysis with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) has emerged as a powerful and efficient tool in DNA sequencing and genotyping. With the elimination of the electrophoretic separation, MALDI-TOF MS based methods have great potential to overcome the difficulties typically encountered by electrophoresis sequencing techniques,
including inaccurate data in GC rich regions and ambiguous heterozygote detection.\textsuperscript{[8,9]} In addition, DNA analysis by MALDI-TOF MS has also been widely explored for multiplex genotyping of single nucleotide polymorphism (SNPs).\textsuperscript{[10,11]} To obtain high mass resolution in MS spectra, one critical requirement for MS-based DNA analysis is stringent sample purity. The target DNA molecules must be isolated from alkali metal ions and falsely terminated DNA fragments, which reduce peak resolution and lead to inaccurate base identification.\textsuperscript{[12]} To address this issue, the use of biotinylated dideoxynucleotides (biotin-ddNTPs) has been developed in our laboratory to isolate DNA extension products.\textsuperscript{[13-17]}

DNA sequencing by MS methods apply the similar procedure as the Sanger dideoxy sequencing\textsuperscript{[2]} to generate the DNA sequencing fragments. Theoretically, the resolution of this MS based sequencing approach can be as good as one dalton. Nonetheless, the accurate measurement of the mass of the DNA sequencing fragments requires the isolation of the pure DNA extension products. The most widely used approach for purifying DNA samples for MS analysis relies on the high affinity between biotin and streptavidin.\textsuperscript{[18,19]} Four natural nucleotides (dNTPs) along with biotinylated dideoxynucleotides (biotin-ddNTPs), each of which has a biotin moiety attached to the base of the dideoxynucleotides via a linker arm (Fig. 4.1), are used to generate DNA sequencing fragments by DNA polymerase reaction (Fig. 4.2A). Upon binding to a streptavidin-coated surface, these DNA fragments carrying a biotin moiety are isolated from all the excess primers, salts, and falsely terminated fragments. After their release from the surface, the pure DNA fragments are analyzed by MALDI-TOF MS to generate the mass spectrum as shown in Fig. 4.2B. The first peak in the spectrum corresponds to
Figure 4.1. The structures of biotinylated dideoxynucleotides, biotin-11-ddATP, biotin-11-ddCTP, biotin-11-ddGTP and biotin-16-ddUTP.
Figure 4.2. DNA sequencing by MALDI-TOF MS. (A) The DNA sequencing fragments are generated by the polymerase reaction by using the combination of dNTPs and biotin-ddNTPs. Then the pure DNA extension products are isolated with solid phase capture for MALDI-TOF MS analysis. (B) Illustration of the DNA sequencing data generated from the DNA fragments.
the DNA extension product generated by the incorporation of a biotin-ddNTP that is complementary to the next available base on the template. The mass difference between this DNA extension product and the primer is used to identify the biotin-ddNTP incorporated. By the analysis of the mass differences between each peak and its preceding peak, the sequence of the template is deciphered. This DNA sequencing approach is effective in removing the peaks corresponding to the unextended primer and the falsely terminated DNA fragments in the MS spectra, thereby increasing the sequencing accuracy and capacity. By designing the structures of biotin-ddNTPs with different linker arms, the mass differences in the spectra can be adjusted to achieve higher resolution. For example, the smallest mass difference among the four biotin-ddNTPs (Fig. 4.1) is 16 daltons, compared with 9 daltons among the four standard ddNTPs. DNA fragments with larger molecular weights typically feature wider peaks in their mass spectra, resulting in difficulties in heterozygote detection. This issue is also addressed by using biotin-ddNTPs with the enlarged mass differences, which improve the mass resolution and facilitate the heterozygote analysis.

SNPs, the most common genetic variations in the human genome, are important markers for disease gene identification and for pharmacogenetic studies. A rapid, precise, and cost-effective method is required for large-scale SNP genotyping. MALDI-TOF MS coupled with single base extension has been successfully developed for SNP detection. To increase the mass resolution and the scope of multiplexing, our laboratory has developed a multiplex genotyping by MS approach using solid phase capturable biotin-ddNTPs. In this method, primers with different molecular masses are designed to anneal next to the polymorphic sites (Fig. 4.3). The single base extension of
Figure 4.3. Multiplex genotyping using molecular affinity and mass spectrometry. The biotinylated DNA extension products generated by single base extension using four biotin-ddNTPs are isolated on a streptavidin-coated surface. After removing the unextended primers and other reaction components, the pure DNA fragments are released for subsequent MALDI-TOF MS analysis.
these primers using biotin-ddNTPs produces 3’-biotinylated DNA products, which are then captured on a streptavidin-coated surface and isolated from the unextended primers and other components of the extension reaction. The pure DNA extension products are subsequently released from the surface by denaturing the biotin-streptavidin interaction and analyzed by MALDI-TOF MS. The mass difference between the DNA extension product and its corresponding primer reveals the identity of the nucleotide at the polymorphic site. With the DNA extension products isolated by solid phase capture prior to MS analysis, the resulting mass spectrum is free of unextended primer peaks and their associated dimers, leading to an increased accuracy and scope of multiplexing. Additionally, the DNA extension products terminated by four biotin-ddNTPs with significantly distinct molecular masses improve the resolution and accuracy of peak detection in the mass spectrum.

The biotin-streptavidin interaction is one of the strongest known non-covalent bonds and has been widely used for biological purification. Nonetheless, the release of the biotinylated DNA extension products from the streptavidin-coated surface requires harsh conditions such as ammonium hydroxide at 60°C, or formamide at 94°C. Furthermore, the biotin moiety carried on the recovered DNA fragments contains sulfur, which has four major stable isotopes. It reduces the resolution of genotyping analysis using MS. These issues in DNA analysis with MALDI-TOF MS can be addressed by using chemically cleavable biotinylated dideoxynucleotides.

4.2 EXPERIMENTAL RATIONALE AND OVERVIEW

The chemically cleavable biotinylated dideoxynucleotides are designed in such a
way that the biotin moiety is covalently attached to the 5-position of pyrimidines and the 7-position of purines through a chemically cleavable linker. These chemically modified dideoxynucleotides may have great potential to be recognized as good substrates for the modified polymerase, owing to the high tolerance of the polymerase to nucleotide modifications with bulky groups at the 5-position of C/U and the 7 position of A/G. The DNA extension reaction with these dideoxynucleotide analogues generates DNA fragments with a cleavable biotin moiety at their 3’ end. A general scheme of capture and cleavage of these DNA fragments from the streptavidin-coated surface is shown in Fig. 4.4. At the end of the primer extension reaction, the reaction mixture consists of excess primers, enzymes, salts, falsely terminated DNA fragments, and desired DNA fragments. This reaction mixture is passed over a streptavidin-coated surface (e.g., magnetic beads) and allowed to incubate. The biotinylated DNA fragments are captured at the 3’ end by the biotin-streptavidin interaction, while everything else in the mixture is washed away. Then the fragments are released into solution by cleaving the linker with a specific and efficient chemical method, while the biotin remains attached to the streptavidin that is covalently bound to the surface. The pure DNA fragments can then be analyzed by mass spectrometry. It is advantageous to avoid the dissociation of the stable biotin-streptavidin interaction and to cleave the biotin moiety since sulfur has several relatively abundant isotopes.

This part of the thesis describes the design, synthesis and evaluation of a chemically cleavable biotinylated dideoxynucleotide, ddATP-N₃-biotin, for the DNA polymerase reaction to isolate extension products for MS analysis. This dideoxynucleotide analogue bears a biotin moiety attached to the 7-position of ddATP via
Figure 4.4. Affinity capture and chemical cleavage for isolating biotinylated DNA fragments on a streptavidin-coated surface. The DNA extension products with a cleavable biotin moiety at the 3′ end are captured and then cleaved chemically to release the DNA fragments for analysis by mass spectrometry, leaving the biotin moiety still bound to the surface.
an azido-based linker. We demonstrate that ddATP-N$_3$-biotin can be faithfully incorporated into a growing DNA strand in a base-specific manner during the polymerase reaction, producing DNA extension products with a cleavable biotin at the 3’ end. These DNA fragments can be isolated on a streptavidin-coated surface and the captured DNA fragments can be efficiently released for MS analysis by incubating in a TCEP solution. Such dideoxynucleotide analogues that carry a chemically cleavable biotin moiety as shown in this study will allow the isolation and purification of DNA fragments under mild conditions for MS-based genetic analysis by DNA sequencing or multiplex SNP detection. Furthermore, these dideoxynucleotide analogs may also have wide applications in isolating DNA-protein complexes under non-denaturing conditions.

### 4.3 RESULTS AND DISCUSSION

#### 4.3.1 Design and synthesis of ddATP-N$_3$-biotin

To demonstrate of the feasibility of carrying out DNA analysis with chemically cleavable biotinylated dideoxynucleotides, we designed and synthesized ddATP-N$_3$-biotin, which bears a biotin moiety attached to the 7-position of the adenine base via an azido-based cleavable linker. The successful implementation of this approach requires the biotin moiety to be removed from the DNA fragment specifically and efficiently in a DNA compatible manner. Previously, we established that the chemically cleavable fluorescent dideoxynucleotides with the azido-based linker could be used successfully for the hybrid SBS approach, and the fluorophores were shown to be completely cleaved by the TCEP treatment.\[^{23}\] The azido-based linker was thus chosen as the chemically cleavable moiety to link a biotin to ddATP in this study.
Figure 4.5. Synthesis of ddATP-N$_3$-biotin.
The synthesis of ddATP-N\textsubscript{3}-biotin was achieved by attaching the biotin moiety and the modified ddATP to the two ends of the cleavable linker. As shown in Fig. 4.5, coupling the cleavable linker 1 with biotin NHS ester 2 in a solution of aqueous sodium bicarbonate and DMF yielded N\textsubscript{3}-biotin 3. Treatment with DSC and DMAP converted compound 3 to its corresponding NHS ester 4, which was subsequently coupled with ddATP-NH\textsubscript{2} to afford the target ddATP-N\textsubscript{3}-biotin 5.

The ddATP-N\textsubscript{3}-biotin was demonstrated to be efficiently incorporated into the growing DNA strand to terminate the polymerase reaction for DNA analysis. The biotin moiety generated by incorporation of this dideoxynucleotide analogue is removed rapidly and quantitatively by TCEP from the DNA extension product, with the mechanism shown in Fig 4.6. The Staudinger reaction is used to reduce the azido group in the cleavable linker into an amino group with TCEP. The generated intermediate hydrolyzes efficiently in aqueous solution to cleave the biotin moiety.

\section*{4.3.2 Polymerase extension using ddATP-N\textsubscript{3}-biotin and characterization by MALDI-TOF mass spectrometry}

To verify that ddATP-N\textsubscript{3}-biotin can be faithfully and efficiently incorporated by Thermo Sequenase in DNA polymerase reactions, we conducted a single base extension reaction with a self-priming DNA template whose next complementary base was A. Should the ddATP-N\textsubscript{3}-biotin be an effective substrate for the DNA polymerase, the extension reaction would be terminated after the incorporation of this dideoxynucleotide analogue, generating DNA fragment 6 in Fig. 4.7. After the reaction, the extension products were desalted and analyzed by MALDI-TOF MS as shown in Fig. 4.8A. A
Figure 4.6. Staudinger reaction with TCEP to cleave the biotin moiety and release the DNA extension products from the streptavidin-coated surface.
DNA extension reaction using ddATP-N₃-biotin and cleavage of the generated DNA fragment captured on a solid surface. DNA polymerase incorporates ddATP-N₃-biotin into a growing DNA stand, generating the cleavable biotinylated DNA fragment. Chemical cleavage using TCEP of this DNA fragment captured on the streptavidin-coated surface releases the DNA fragment, with the boitin moiety remaining on the surface.

Figure 4.7. DNA extension reaction using ddATP-N₃-biotin and cleavage of the generated DNA fragment captured on a solid surface. DNA polymerase incorporates ddATP-N₃-biotin into a growing DNA stand, generating the cleavable biotinylated DNA fragment. Chemical cleavage using TCEP of this DNA fragment captured on the streptavidin-coated surface releases the DNA fragment, with the boitin moiety remaining on the surface.
single clear mass peak at 8,889 corresponding to the single base extension product with no left over starting material was produced using ddATP-N₃-biotin. This result demonstrates that the ddATP-N₃-biotin is a good substrate for the DNA polymerase, Thermo Sequenase.

Rapid and efficient chemical cleavage for the recovery of the DNA extension products captured on a streptavidin-coated surface is another critical requirement for successful application of cleavable biotinylated dideoxynucleotides in DNA analysis. To demonstrate the feasibility of carrying out this chemical cleavage by using TCEP, the DNA extension fragment with a biotin moiety at its 3’ end was immobilized on streptavidin-coated magnetic beads. After washing away unextended primers and other reaction components, treatment with an aqueous TCEP solution released the DNA extension fragment 7 (Fig. 4.7) and left the biotin moiety remaining on the streptavidin-coated surface. The analysis of the cleaved DNA fragment by MALDI-TOF MS produced a single peak (Fig. 4.8B), whose mass corresponded to the DNA cleavage product 7 shown in Fig. 4.7. This experiment indicates that chemical cleavage with TCEP efficiently breaks the linker under mild conditions and releases DNA molecules free of biotin.

### 4.3.3 Multiplex DNA extension by using ddATP-Ν₃-biotin

We further investigated the application of this dideoxynucleotide analogue bearing the cleavable biotin moiety in the multiplex analysis of DNA extension products. Four primers with different molecular masses (Table 4.1) were simultaneously extended with ddATP-N₃-biotin, yielding extension products of 20-26 base pairs in length (Fig.
Figure 4.8. MALDI-TOF mass spectra of the DNA extension product generated from ddATP-N₃-biotin and the subsequent cleavage of the generated DNA fragment captured on a solid surface. (A) The DNA polymerase incorporated ddATP-N₃-biotin, yielding DNA extension product 6. (B) This generated DNA fragment was captured on a streptavidin-coated surface and a chemical cleavage reaction using TCEP was carried out to release the DNA fragment 7.
Figure 4.9. (A) Multiplex DNA analysis scheme. Four DNA primers with different molecular weights were extended with ddATP-N$_3$-biotin by DNA polymerase. After solid phase capture and elimination of the unextended primers, a chemical cleavage reaction was carried out under mild conditions to release the DNA extension products. (B) MALDI-TOF mass spectra of the generated DNA fragments.
Then these DNA fragments were captured on a streptavidin-coated surface to test the chemical cleavage and the recovery of DNA extension products. After washing the beads, an aqueous solution of TCEP was utilized to break the linker tethering the DNA fragments to the biotin moiety. The released DNA fragments were analyzed by MALDI-TOF MS (Fig. 4.9B). Four clear peaks corresponded to the cleaved fragments of all four extension products, indicating that the mild chemical cleavage successfully recovered multiple DNA samples captured on a surface.

4.4 MATERIALS AND METHODS

General Information. All solvents and reagents were reagent grade, purchased commercially and used without further purification unless specified. All chemicals were purchased from Sigma-Aldrich unless otherwise indicated. Oligonucleotides used as primers or templates were synthesized on an Expedite nucleic acid synthesizer (Applied BioSystems) or purchased from Midland. $^1$H NMR spectra were recorded on a Bruker DPX-400 (400 MHz) spectrometer and reported in parts per million (ppm) from a CD$_3$OD internal standard (3.31). Data were reported as follows: (s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, dd = doublet of doublets, coupling constant (J) in Hz, integration). Mass measurement of DNA was performed on a Voyager DE MALDI-TOF mass spectrometer (Applied Biosystems). Thermo Sequenase polymerase was obtained from GE Healthcare. 2-(Aminomethyl)-1,3-dioxolane was purchased from TCI America (Portland, OR). Aminopropargyl-ddATP (NH$_2$-ddATP) was purchased from PerkinElmer. The ddATP-N$_3$-biotin was purified with reverse-phase HPLC on a 150×4.6 mm C18 column (Supelco), mobile phase: A, 8.6 mM Et$_3$N / 100 mM
1,1,3,3,3-hexafluoro-2-propanol in water (pH 8.1); B, methanol. Elution was performed from 100% A isocratic over 10 minutes followed by a linear gradient of 0-50% B for 20 minutes and then 50% B isocratic over another 30 minutes.

4.4.1 Synthesis of ddATP-$N_3$-biotin

Azido-biotin (Compound 3). (2-{2-[3-{2-Amino-ethylcarbamoyl}-phenoxy]-1-azido-ethoxy}-ethoxy)-acetic acid 1 (7.0 mg, 0.019 mmol), prepared according to the literature\textsuperscript{[24]}, was dissolved in DMF (300 µl) and 1 M NaHCO$_3$ aqueous solution (100 µl). A solution of Biotin NHS (N-hydroxysuccinimide) ester 2 (7.0 mg, 0.021 mmol) in DMF (400 µl) was added slowly to the above reaction mixture and then stirred at room temperature for 5 hours with exclusion of light. The crude product was purified on a preparative silica gel TLC plate (CH$_2$Cl$_2$/CH$_3$OH, 1:1) to afford 3 (8.8 mg; 78%). $^1$H NMR (400 MHz, CD$_3$OD) δ7.39-7.48 (m, 3H), 7.17 (d, $J = 8.6$ Hz, 1H), 5.07 (t, $J = 5.2$ Hz, 1H), 4.46-4.49 (m, 1H), 4.17-4.46 (m, 3H), 3.89-4.05 (m, 4H), 3.76 (s, 2H), 3.44-3.55 (m, 4H), 3.07-3.12 (m, 1H), 2.88-2.94 (m, 1H), 2.60-2.73 (m, 2H), 2.20-2.26 (m, 3H), 1.39-1.72 (m, 9H); MS (Fab+) calcd for C$_{25}$H$_{36}$N$_7$O$_8$S [(M+H)$^+$]: 594.2, found: 594.3.

ddATP-$N_3$-biotin (Compound 5). To a stirred solution of 3 in dry DMF (2 ml), DSC (N, N’-disuccinimidyl carbonate) (3.4 mg, 13.2 µmol) and DMAP (4-dimethylaminopyridine) (1.6 mg, 13.2 µmol) were added. The reaction mixture was stirred at room temperature for 2 hours. TLC indicated that 3 was completely converted to compound 4, which was directly used to couple with amino-ddATP (1 µmol) in NaHCO$_3$/Na$_2$CO$_3$ buffer (pH = 8.7, 0.1 M) (300 µl). The reaction mixture was stirred at room temperature for 3 hours
with exclusion of light. The reaction mixture was purified on a preparative silica gel TLC plate (CH$_3$OH/CH$_2$Cl$_2$, 1:1). The crude product was further purified by reverse-phase HPLC to afford 5 (retention time = 30.3 min). Compound 5 was further evaluated by performing a single base extension reaction to yield a DNA extension product that was characterized by MALDI-TOF MS (m/z 8889) (Fig. 4.8).

### 4.4.2 Polymerase extension using ddATP-N$_3$-biotin and characterization by MALDI-TOF mass spectrometry

We characterized ddATP-N$_3$-biotin by performing a single base DNA extension reaction with a self-priming DNA template (5’-GAC-TGC-GCC-GCG-CCT-TGG-CGC-GGC-GC-3’). The extension reaction mixture was composed of 160 pmol of ddATP-N$_3$-biotin along with 80 pmol of the self-priming DNA template, 4 µl of Thermo Sequenase reaction buffer, and 4 units of the DNA polymerase Thermo Sequenase in a total reaction volume of 40 µl. The reaction consisted of 25 cycles of 94°C for 20 seconds, 48°C for 40 seconds, and 72°C for 20 seconds. A portion of the extension products (20 µl) were desalted by using a ZipTip, dried, and resuspended in 1 µl of deionized water. After mixing with 1 µl of matrix solution, which contained 3-hydroxypicolinic acid and ammonium citrate in 50% acetonitrile, the DNA extension products were analyzed by Voyager DE™ MALDI-TOF mass spectrometry.

The rest of the DNA extension products (20 µl) were purified by solid phase capture on streptavidin-coated magnetic beads (Seradyn, Ramsey, MN) to remove unextended primers and other reaction components. The captured DNA fragments on the streptavidin-coated magnetic beads were washed and then released with TCEP for
MALDI-TOF MS analysis. 20 µl of the streptavidin-coated magnetic beads were washed with modified binding and washing (B/W) buffer (0.5 mM Tris–HCl buffer, 2 M NH₄Cl, 1 mM EDTA, pH 7.0) and resuspended in 20 µl of modified B/W buffer. The extension products were added to the suspended beads and incubated for 30 minutes. After capture, the beads were washed twice with modified B/W buffer, twice with 0.1 M triethyl ammonium acetate (TEAA) buffer and twice with deionized water. Then the extension products were cleaved and released from the beads by treatment with 20 µl of TCEP solution (100 mM, pH 9.0) at 65°C for 10 minutes, precipitated with 100% ethanol at 4°C for 30 minutes, and centrifuged (Eppendorf 5417 C/R) at 4°C and 14000 r.p.m. for 45 minutes. The purified extension products were analyzed by MALDI-TOF MS.

4.4.3 Multiplex Genotyping using ddATP-N₃-biotin

We conducted the multiplex extension reaction using ddATP-N₃-biotin, four primers with different molecular masses, and a synthetic template containing the sequence related to a portion of exon 8 of the p53 gene. The sequences of the primers and the template are listed in Table 4.1. The reaction mixture consisted of 250 pmol of ddATP-N₃-biotin, 60 pmol of the template, 50 pmol each of the four primers, 4 µl of Thermo Sequenase reaction buffer, and 4 units of the DNA polymerase Thermo Sequenase in a total reaction volume of 40 µl. The concentrations of the template and all four primers were determined experimentally to generate relatively even peaks in the mass spectrum. The extension reaction, solid-phase capture, and the subsequent chemical cleavage were carried out with a procedure similar to that for self-priming DNA templates. The released DNA fragments in TCEP solution were precipitated and analyzed.
**Table 4.1** Sequences of oligonucleotide primers and the synthetic DNA template for the multiplex DNA extension reaction.

<table>
<thead>
<tr>
<th>Primers</th>
<th>Primer sequences</th>
<th>Mass (Da)</th>
<th>Mass of cleavage products (Da)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5’-TGTCGAACCTCCACGCAA-3’</td>
<td>5742</td>
<td>6193</td>
</tr>
<tr>
<td>2</td>
<td>5’-TAGATGACCCTGCTTGT-3’</td>
<td>6084</td>
<td>6535</td>
</tr>
<tr>
<td>3</td>
<td>5’-ACGGAGAAACGAAGAAGAAGG-3’</td>
<td>6595</td>
<td>7046</td>
</tr>
<tr>
<td>4</td>
<td>5’-AAAGGATAGGACTCATCACCATTAG-3’</td>
<td>7683</td>
<td>8134</td>
</tr>
</tbody>
</table>

Template sequence

5’-GAAGGAGACACGGCCAGAGGAGGTCTTGTCCGTGTTTGTTGCGTGAGTTCGACAAAGGCAGGTCTCATAATGGTGATGAGTCTCTATCCTTTTTCATCTCGTCCGT-3’
by MALDI-TOF MS.

4.5 CONCLUSION

We have synthesized a chemically cleavable biotinylated dideoxynucleotide analogue, ddATP-N₃-biotin, and evaluated its application for rapid and efficient recovery of DNA fragments captured on a solid surface under mild conditions. This dideoxynucleotide analogue has been demonstrated to be a good substrate for the DNA polymerase Thermo Sequenase. The relatively bulky cleavable biotin moiety introduced at the 7-position of ddATP does not interfere with its recognition by Thermo Sequenase. Generated by incorporation of ddATP-N₃-biotin, the biotin moiety at the 3’ end of the DNA extension products is used to immobilize the DNA fragments to a streptavidin-coated surface. Subsequently, the DNA extension products are released by using TCEP without any strong base or other denaturing reagents. These DNA fragments recovered by chemical cleavage are free of the biotin moiety, which facilitates the MS analysis with higher resolution.

All the four chemical cleavable biotinylated dideoxynucleotides, ddATP-N₃-biotin, ddUTP-N₃-biotin, ddGTP-N₃-biotin, and ddCTP-N₃-biotin, can be synthesized using a similar procedure to that described here. This set of dideoxynucleotide analogues will be a valuable tool for developing DNA sequencing by MS, multiplex genotyping, and other novel genetic analysis techniques.
REFERENCES


Part III. Fluorescent Labeling of cDNA Probes with Click Chemistry for Microarray Analysis
5.1 INTRODUCTION

The completion of the Human Genome Project and the achievements of similar projects in other species have opened the door to a postgenome era. One of the most ambitious objectives of this new period is to understand how the genomes of various species function, and how the gene products interact to produce complex living organisms. The development of new technologies has resulted in powerful tools to provide answers to some of these questions on a genome-wide scale. Among these genomic technologies, DNA microarrays have enormous importance and many exciting prospects, despite the advent of next generation sequencing approaches. Although the most common use of DNA microarrays is to study gene expression,\textsuperscript{[1,2]} they also have wide applications in disease diagnosis,\textsuperscript{[3-5]} candidate gene identification,\textsuperscript{[6,7]} and biochemical pathway construction.\textsuperscript{[8,9]}

Nonetheless, DNA microarrays technology suffers from inadequate precision due to many variations during the experimental process.\textsuperscript{[10,11]} A reliable and accurate microarray experiment requires several important factors to be considered: 1) the quality and quantity of the DNA probes immobilized on the surface, 2) the quality and quantity of the RNA samples, 3) the method of labeling cDNA samples, 4) the hybridization condition, and 5) the scanning procedure. Many efforts have been made to optimize and standardize each of these steps, especially regarding the labeling of cDNA samples.\textsuperscript{[12-15]} Depending on different detection systems, radioactive, chromogenic, or fluorescent labels have been applied for DNA microarrays technology. Historically, the cDNAs have been labeled by the incorporation of nucleotides featuring beta-emitting radioisotopes such as
$^{32}$P or $^{33}$P at the $\alpha$ or $\gamma$ phosphate position and detected by exposure to X-ray film. Despite its relatively high efficiency, radioactive labeling approach is not suitable for high-density DNA microarray platforms.

Among the various cDNA labeling approaches, fluorescent labeling has many advantages, including the versatility in its excitation and emission wavelengths and its minimal hazardousness. The two major cDNA labeling strategies with fluorescent dyes are: direct fluorescent labeling and indirect fluorescent labeling. With the direct cDNA labeling method, fluorescently modified deoxynucleotides such as Cy3-labeled dCTP and Cy5-labeled dCTP (Fig. 5.1) are incorporated during first-strand cDNA synthesis from two mRNA samples, control and test respectively (Fig. 5.2). After reverse transcription, the mRNAs are degraded by sodium hydroxide. The remaining single-stranded cDNA mixtures are purified to remove unincorporated nucleotides, salts, enzymes, and other components. Then the Cy5-labeled cDNA and Cy3-labeled cDNA are mixed and hybridized to a microarray. The main advantage of the direct labeling approach is its simplicity, since it avoids the need for a post-labeling reaction. However, Cy5-labeled dCTP is less efficiently incorporated than Cy3-labeled dCTP, which may introduce the most important source of systematic error in two-color microarray experiments. An additional disadvantage of the direct labeling method is its cost, which is two times more expensive than the indirect labeling approach.

In the indirect labeling approach, the first-strand of cDNAs are synthesized with the combination of dNTPs and an aminoallyl-modified nucleotide such as aminoallyl-dUTP (Fig. 5.3). Then the aminoallyl-functionalized cDNAs are coupled with
Fig. 5.1. Chemical structures of Cy3-dCTP and Cy5-dCTP.
Figure 5.2. Direct cDNA fluorescent labeling for microarray experiments. Cy3-labeled dCTP and Cy5-labeled dCTP are incorporated during cDNA synthesis from control and test mRNA samples respectively. After degrading the mRNAs, the Cy3-labeled cDNA and Cy5-labeled cDNA are purified, mixed and hybridized on a DNA microarray chip.
Cy3 or Cy5. After purification, the Cy3-labeled control sample and the Cy5-labeled test sample are mixed and applied to the microarray (Fig. 5.4). With the less bulky aminoallyl group attached on the base, aminoallyl-dUTP is incorporated more efficiently than Cy3 or Cy5-labeled dCTP, which leads to higher signal intensity in the indirect labeling method. In addition, with the post-labeling approach, the possibility of equally labeling the control and test samples with Cy3 and Cy5 respectively is significantly increased.

Nonetheless, in the post-labeling step of this conventional indirect labeling method, N-hydroxysuccinimide (NHS) ester modified fluorescent dyes are easily hydrolyzed in aqueous solution, leading to low efficiency of the coupling chemistry. In addition, the amino groups on the bases of dATP, dGTP and dCTP could be coupled with the NHS ester dyes. Thus, these generated side products will interfere with the subsequent hybridization reaction. The third part of this thesis describes the development of a click-chemistry based approach to fluorescently label cDNAs in an efficient, inexpensive, and unbiased manner.

5.2 EXPERIMENTAL RATIONALE AND OVERVIEW

Here, we describe a molecular approach to efficiently and specifically label cDNA by the Cu(I)-mediated Huisgen cycloaddition reaction (click chemistry). In this post-synthetic functionalization approach, the quantitative conversion of reporter groups into fluorescently labeled products must fulfill the following requirements: (1) The chemically modified nucleotide analogues must be good substrates for reverse transcriptase. (2) The reporter groups generated by the incorporation of these nucleotide analogues must be coupled with fluorophores with high efficiency and specificity. (3)
Cy5 respectively, followed by mixing and hybridization on the aminoallyl-dUTP samples are reverse-transcribed into cDNA using reverse transcriptase, dNTPs and aminoallyl-dUTP. Then the amino modified cDNA from the two samples are labeled with Cy3 and Cy5 respectively, followed by mixing and hybridization on the microarray.

**Figure 5.3.** Chemical structure of aminoallyl-dUTP.

**Figure 5.4.** Indirect cDNA fluorescent labeling for microarray experiments. The control and test mRNA samples are reverse-transcribed into cDNA using reverse transcriptase, dNTPs and aminoallyl-dUTP. Then the amino modified cDNA from the two samples are labeled with Cy3 and Cy5 respectively, followed by mixing and hybridization on the microarray.
This post-labeling reaction must be DNA compatible. Among a handful of chemical moieties that feature the required attributes, azide and alkyne functions have been shown to be elegantly suited for biomolecular coupling. Our preliminary investigations have demonstrated that DNA polymerase is highly tolerant to nucleotide modifications with bulky groups at the 5-position of pyrimidines and the 7-position of purines. Thus, the chemically modified dUTP with an azide group attached to the 5-position could probably be recognized as a good substrate by reverse transcriptase. Additionally, both azide and alkyne groups are fairly stable in aqueous solution, resulting in the high specificity of the labeling reaction. Furthermore, due to its high efficiency and biocompatibility, click chemistry has wide applications in post-synthetic DNA labeling and detection.

With all three requirements for indirect cDNA labeling met by click chemistry, we designed and synthesized N₃-dUTP and alkyne functionalized fluorophores for this post-synthetic labeling approach (Fig. 5.5). Taking mRNAs from control and test samples as templates, the first strand of cDNA is synthesized by using reverse transcriptase along with N₃-dUTP and dNTPs. After degrading the RNA strands, the azide modified cDNAs are purified and coupled with alkyne functionalized fluorophores by Cu(I) catalyzed [3+2] cycloaddition (click chemistry). The resulting fluorescently labeled cDNAs from the two samples are then mixed for microarray analysis.

5.3 RESULTS AND DISCUSSION

5.3.1 Design and synthesis of N₃-dUTP and alkyne functionalized fluorophores

To demonstrate the feasibility of the indirect cDNA labeling approach by click
Figure 5.5. Indirect cDNA fluorescent labeling with click chemistry for microarray analysis. Taking the control and test mRNAs as templates, their corresponding cDNAs are synthesized using reverse transcriptase, dNTPs and azide-modified dUTP (N₃-dUTP). After degrading the mRNAs, the azide labeled cDNA from the two samples are coupled with alkyne modified Cy3 and Cy5 respectively. Subsequently, the two fluorescently labeled cDNA samples are mixed and hybridized on the microarray.
chemistry, we designed and synthesized azide functionalized deoxyuridine triphosphate (N₃-dUTP) and alkyne functionalized fluorophores. A relatively small azidomethyl group is attached to the 5-position of uridine to ensure this chemically modified nucleotide analogue is incorporated efficiently in a base-specific manner by reverse transcriptase. As shown in Fig. 5.6, N₃-dUTP 3 was synthesized from 5-(hydroxymethyl)-2’-deoxyuridine 1. Among the three hydroxyl groups on hydroxyl-dU 1, only the one at the 5-position reacted with chlorotrimethylsilane in 1,4-dioxane and then sodium azide in DMF, leading to 5-(azidomethyl)-2’-deoxyuridine 2. N₃-dU 2 was converted to the target N₃-dUTP 3 with an established phosphorylation procedure.[21] Two alkyne functionalized fluorophores with unique fluorescence emissions were prepared from dye NHS esters by treatment with propargyl amine at room temperature for 6 hours (Fig. 5.7). It was demonstrated that N₃-dUTP is efficiently incorporated by reverse transcriptase during the cDNA synthesis. The azide groups on the generated cDNAs could couple with alkyne functionalized fluorophores quantitatively in aqueous solution to fluorescently label the cDNAs for microarray analysis.

5.3.2 N₃-dUTP incorporation by reverse transcriptase and the subsequent fluorescent labeling with click chemistry

To verify that N₃-dUTP incorporates efficiently during the reverse transcription reaction, a single base DNA extension reaction was carried out in solution by using RNA as the template and N₃-dUTP as the substrate (Fig. 5.8). This allowed the isolation of the DNA extension product for detailed molecular structure characterization. The sequences of the RNA template and the DNA primer were designed in such a way that the next two
**Figure 5.6.** Synthesis of $\text{N}_3$-dUTP.

**Figure 5.7.** Synthesis of alkyne functionalized Cy3 and Cy5.
Figure 5.8. Primer extension using $N_3$-dUTP along with reverse transcriptase and the subsequent fluorescent labeling with click chemistry. Taking RNA as the template, reverse transcriptase extended the DNA primer with $N_3$-dUTP. This azido-labeled DNA extension product was then coupled with alkyne functionalized Cy3 and Cy5.
available bases on the template were A and C, which allowed only a single N$_3$-dUTP to be incorporated into the growing DNA primer. The extension product 9 was desalted and analyzed using MALDI-TOF MS (Fig. 5.9A). The mass spectrum consists of a distinct peak corresponding to the DNA extension product 9 (m/z 7,744), which confirms that N$_3$-dUTP has successfully synthesized and efficiently incorporated by reverse transcriptase into the growing DNA strand.

To verify that the azide modified DNA, generated by the incorporation of N$_3$-dUTP, can be efficiently labeled by click chemistry, the coupling reaction between DNA extension product 9 and alkyne functionalized fluorophores were carried out using Cu(I) as the catalyst (Fig. 5.9B and C). The reaction products 10 and 11 were desalted and characterized by MALDI-TOF MS, yielding a single peak at 8,410 and 8,436 respectively. These results demonstrate that the azide modified DNA can be quantitatively labeled with alkyne functionalized fluorophores by click chemistry.

5.3.3 Fluorescent labeling of cDNA probes with click chemistry

In our cDNA labeling by click chemistry approach, an azide modified nucleotide analogue is used to perform the cDNA synthesis and alkyne functionalized fluorophores are utilized to label the resulting cDNA by click chemistry. The general cDNA labeling scheme for microarray analysis is shown in Fig. 5.10A. The stability of the azide and the alkyne groups in aqueous solution lead to high specificity and efficiency of this coupling reaction. In contrast, the conventional cDNA labeling approach uses NH$_2$-dUTP to synthesize the cDNA, which is subsequently coupled with NHS ester modified fluorophores (Fig. 5.4). The hydrolysis of the NHS ester group on the fluorophores in
Figure 5.9. MALDI-TOF mass spectra of the DNA extension product generated from N₃-dUTP and the subsequent fluorescent labeling with click chemistry. (A) Reverse transcriptase incorporated N₃-dUTP, yielding DNA extension product 9. (B) This resulting DNA fragment was coupled with Cy3 by click chemistry to generate Cy3 labeled DNA product 10. (C) The coupling reaction was carried out in a similar manner to convert the azido-modified DNA fragment into Cy5 labeled DNA product 11.
Figure 5.10. (A) The scheme of fluorescent labeling of cDNA probes with click chemistry. The RNA sample was reverse-transcribed into their corresponding cDNA using reverse transcriptase, dNTPs and N₃-dUTP. Then the azido labeled cDNA was coupled with alkyne functionalized Cy3 and Cy5 respectively. (B) UV-Vis absorption spectra of the Cy3 and Cy5 labeled cDNA probes.
aqueous solution lowers the labeling efficiency. Furthermore, the coupling reaction between the NHS ester modified fluorophores and the amino groups on the purines and the cytosines of the cDNA could decrease the labeling specificity. In the subsequent hybridization reaction, these side products will interfere with the formation of the hydrogen bond between the cDNAs and the DNA probes immobilized on the microarray.

The preparation of our fluorescently labeled cDNA samples was initiated by the synthesis of azide modified cDNAs. The random primers were annealed to the RNA templates and extended by the incorporation of $\text{N}_3$-dUTP along with dNTPs. Using Mn$^{2+}$ as the cofactor, we determined the optimal ratio of $\text{N}_3$-dUTP and dTTP for reverse transcriptase experimentally. Once the azide functionalized cDNAs were successfully synthesized, we degraded the RNA strands with sodium hydroxide. Then the purified cDNA strands were coupled with alkyne functionalized Cy3 and Cy5 respectively by click chemistry. These two fluorescently labeled cDNA samples were purified and analyzed by UV-Vis absorption spectroscopy. The UV-Vis absorption spectra of cDNA samples labeled by Cy3 and Cy5 contains a major peak around 260 nm resulting from the DNA bases and peaks centered at 550 nm and 650 nm corresponding to Cy3 and Cy5 respectively (Fig. 5.10B). These results illustrate that $\text{N}_3$-dUTP is successfully incorporated in the reverse transcription reaction and the resulting cDNAs are successfully labeled by alkyne functionalized Cy3 and Cy5.

5.4 MATERIALS AND METHODS

General Information. All solvents and reagents were reagent grades, purchased commercially and used without further purification unless specified. All chemicals were
purchased from Sigma-Aldrich unless otherwise indicated. Oligonucleotides used as primers or templates were synthesized on an Expedite nucleic acid synthesizer (Applied BioSystems) or purchased from Midland. $^1$H NMR spectra were recorded on a Bruker DPX-400 (400 MHz) spectrometer and reported in parts per million (ppm) from a CD$_3$OD internal standard (3.31). Data were reported as follows: (s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, dd = doublet of doublets, coupling constant ($J$) in Hz, integration). Mass measurement of DNA was performed on a Voyager DE MALDI-TOF mass spectrometer (Applied Biosystems). Cy3 NHS ester and Cy5 NHS ester were obtained from GE Healthcare. The N$_3$-dUTP was purified by reverse-phase HPLC on a 150×4.6 mm C18 column (Supelco), mobile phase: A, 8.6 mM Et$_3$N / 100 mM 1,1,1,3,3,3-hexafluoro-2-propanol in water (pH 8.1); B, methanol. Elution was performed from 100% A isocratic over 10 minutes followed by a linear gradient of 0-50% B for 20 minutes and then 50% B isocratic over another 30 minutes.

5.4.1 Synthesis of N$_3$-dUTP

N$_3$-dUTP (Compound 3). 5-(Azidomethyl)-2'-deoxyuridine 2 (42.5 mg; 0.15 mmol), prepared according to the literature$^{[25]}$, and proton sponge (37.9 mg; 0.18 mmol) were dried in a vacuum dessicator over P$_2$O$_5$ overnight before dissolving in trimethyl phosphate (300 µl). Then freshly distilled POCl$_3$ (20 µl; 0.17 mmol) was added dropwise at 0°C and the mixture was stirred at 0°C for 2 hours. Subsequently, a well-vortexed mixture of tributylammonium pyrophosphate (276 mg) and tributylamine (0.27 ml; 1.15 mmol) in anhydrous DMF (1.16 ml) was added in one portion at room temperature and stirred for 30 minutes. Triethyl ammonium bicarbonate solution (TEAB) (0.1 M; pH 8.0;
7.5 ml) was then added and the mixture was stirred for 1 hour at room temperature. Then concentrated NH₄OH (7.5 ml) was added and stirred overnight at room temperature. The resulting mixture was concentrated under vacuum and the residue was diluted with 2.5 ml of water. The crude mixture was then purified by anion exchange chromatography on DEAE-Sephadex A-25 at 4°C using a gradient of TEAB (pH 8.0; 0.1-1.0 M). The crude product was further purified by reverse-phase HPLC to afford 3. Compound 3 was further evaluated by performing a single base extension reaction to yield a DNA extension product which was characterized by MALDI-TOF MS (m/z 7744) (Fig. 5.9A).

5.4.2 Synthesis of alkyne functionalized Cy3 and Cy5

**Alkyne-Cy3 (Compound 6).** To a stirred solution of propargylamine 4 (10 µl; 156 µmol) in DMF (450 µl), aqueous NaHCO₃ (1 M; 100 µl) was added. The solution was stirred at room temperature for 5 minutes. Cy3 NHS (N-hydroxysuccinimide) ester 5 (5 mg; 6.5 µmol) in DMF (450 µl) was added. The reaction mixture was stirred at room temperature for 6 hours. The crude product was purified on a preparative silica gel TLC plate (CH₃OH/CH₂Cl₂, 1:2) to afford 6 (3.2 mg 74%). H NMR (400 MHz, CD₃OD) δ 8.60 (t, J = 13.6 Hz, 1H), 7.93-8.00 (m, 5H), 7.41-7.45 (m, 2H), 6.52-6.59 (m, 2H), 4.18-4.28 (m, 4H), 3.95 (d, J = 4.8 Hz, 2H), 2.59 (t, J = 2.8 Hz, 1H), 2.25 (t, J = 7.2 Hz, 2H), 1.72-1.91 (m, 14H), 1.45-1.53 (m, 6H); APCI-MS calcd for C₃₄H₄₁N₃O₇S₂ [(M+H)⁺]: 667.2, found: 667.1.

**Alkyne-Cy5 (Compound 8).** The preparation procedure was similar to the synthesis of 6. The crude product was purified on a preparative silica gel TLC plate (CH₃OH/CH₂Cl₂,
1:2) to afford 8 (3.3 mg 75%). $^1$H NMR (400 MHz, CD$_3$OD) $\delta$ 8.33 (t, $J = 13.6$ Hz, 2H), 7.89-7.93 (m, 4H), 7.34-7.38 (m, 2H), 6.70 (d, $J = 12.4$ Hz, 2H), 6.34-6.39 (m, 2H), 4.13-4.23 (m, 4H), 3.95 (d, $J = 4.8$ Hz, 2H), 2.59 (t, $J = 2.8$ Hz, 1H), 2.24 (t, $J = 7.2$ Hz, 2H), 1.69-1.92 (m, 14H), 1.40-1.52 (m, 6H); APCI-MS calcd for C$_{36}$H$_{43}$N$_3$O$_7$S$_2$ [(M+H)$^+$]: 693.2, found: 692.9.

5.4.3 $N_3$-dUTP incorporation by reverse transcriptase

We characterized the $N_3$-dUTP by performing a single base DNA extension reaction using reverse transcriptase, a synthetic DNA primer and an RNA template. The sequences of the DNA primer and the RNA template are listed in Table 5.1. The reaction mixture consisted of 60 pmol of the DNA primer, 80 pmol of the RNA template, 200 pmol of $N_3$-dUTP, 1X reverse transcription buffer (Invitrogen), 10 unit of Superase-In (Ambion), 200 nmol of dithiothreitol (DTT), 40 nmol of MnCl$_2$ and 200 units of SuperScript II reverse transcriptase (Invitrogen) in a total reaction volume of 20 $\mu$l. After incubation at 45°C for 50 minutes, the extension product was desalted using a ZipTip and analyzed by Voyager DE™ MALDI-TOF mass spectrometry (Applied Biosystems).

5.4.4 Fluorescent labeling of azide modified DNA with click chemistry

The fluorescent labeling of azide modified DNA, which was generated by DNA primer extension with $N_3$-dUTP, was characterized by MALDI-TOF MS. A coupling mixture was prepared by mixing tetrakis-(acetonitrile)copper (I) hexafluorophosphate (5 $\mu$l; 16 mM/DMSO), tris-(benzyltriazolymethyl) amine (TBTA) (5 $\mu$l; 16 mM/DMSO),
Table 1.1 Sequences of the DNA primer and the RNA template for single base extension reaction.

<table>
<thead>
<tr>
<th>DNA primer sequence</th>
<th>Mass (Da)</th>
<th>Mass of the extension product (Da)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5’-CGA-AAA-AAG-CTT-TAT-GAA-ATT-TTG-3’</td>
<td>7399</td>
<td>7744</td>
</tr>
</tbody>
</table>

RNA template sequence

sodium ascorbate (5 μl; 10 mM/H₂O), azide modified DNA (5 μl; 20 μM/H₂O) and alkyne functionalized Cy3 or Cy5 (5 μl; 10 mM/H₂O). After incubation at room temperature overnight, the coupling product was purified by reverse-phase HPLC. The fraction containing the desired product was collected and dried for analysis by MALDI-TOF MS.

5.4.5 Fluorescent labeling of cDNA probes with click chemistry

To synthesize the first-strand cDNA, a reaction mixture was prepared by mixing 10 nmol each of dATP, dCTP, and dGTP, 5 nmol of dTTP, 10 nmol of N₃-dUTP, 9 μg of total RNA,²⁶ 240 ng of random primers, 1X reverse transcription buffer, 10 unit of Superase-In, 200 nmol of dithiothreitol (DTT), 40 nmol of MnCl₂ and 200 unit of SuperScript II reverse transcriptase in a total reaction volume of 20 μl. After the synthesis of first-strand cDNA at 45°C for 50 minutes, the RNA strands were degraded by adding 30 μl of H₂O along with 21.5 μl of 1 M NaOH and incubating at 65°C for 15 minutes. Upon adding 21.5 μl of 1 M HCl, the cDNA samples were cleaned up using a Microcon YM-50 filter (Millipore) and resuspended in 20 μl of H₂O.

To fluorescently label the resulting cDNA samples with click chemistry, a coupling mixture, prepared by mixing the generated cDNA (10 μl), tetrakis-(acetonitrile)copper (I) hexafluorophosphate (10 μl; 16 mM/DMSO), tris-(benzyltriazolylmethyl) amine (TBTA) (10 μl; 16 mM/DMSO), sodium ascorbate (10 μl; 10 mM/H₂O), and alkyne functionalized Cy3 or Cy5 (10 μl; 10 mM/H₂O), was incubated at room temperature overnight. Upon removing the excess fluorophores with a Microcon YM-50 filter, the purified cDNA samples were resuspended in 6 μl of TE
buffer and then analyzed by UV-Vis absorption spectroscopy.

5.5 CONCLUSION

We have synthesized N$_3$-dUTP along with alkyne functionalized fluorophores and used them to prepare fluorescently labeled cDNA samples by an indirect click chemistry labeling approach, which has the following advantages. By using N$_3$-dUTP with a less bulky group attached to the base in the reverse transcription reaction, this strategy should outperform the direct labeling method by increasing the signal intensity through the incorporation of more reporter groups and also by maximizing the possibility of equal labeling with Cy3 and Cy5. In addition, improved labeling efficiency and specificity should be achievable owing to the high stability of the azide and alkyne groups in aqueous solution and the inherent properties of click chemistry.

We have experimentally determined the ratio of N$_3$-dUTP and dTTP to incorporate an optimal number of azide groups for the subsequent labeling reaction. These azide groups on the cDNAs are efficiently coupled with the alkyne functionalized fluorophores by click chemistry, yielding the fluorescently labeled cDNA samples for microarray analysis. The efficiency and specificity of this indirect cDNA labeling approach depend on two factors: the incorporation of the modified nucleotide analogue carrying the reporter group by reverse transcriptase and the coupling reaction between this reporter group and the functionalized fluorophores. We have investigated this indirect labeling approach by using N$_3$-dUTP and alkyne functionalized fluorophores. Alternatively, alkyne modified dUTP and azide functionalized fluorophores could be synthesized with a similar procedure to demonstrate the robustness of this strategy. This
post-synthetic labeling by click chemistry approach has wide applications in labeling DNA with a variety of tag molecules, such as fluorophores and biotin moieties. These labeled DNA samples will be valuable reagents for many biological investigations as well as developing new genetic analysis techniques.
REFERENCES


