Chapter 6
An Integrated System for DNA Sequencing by Synthesis

John R. Edwards, Dae Hyun Kim and Jingyue Ju

Columbia Genome Center, Columbia University College of Physicians and Surgeons, Russ Berrie Medical Science Pavilion, 1150 St. Nicholas Avenue, New York, NY 10032, USA

Contents

Abstract

The completion of the Human Genome Project 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. Recently, much progress has been made on the fundamental sciences required to make SBS a viable sequencing technology. One of the unique features of this approach is that many of the steps required are compatible in a modular fashion allowing for the best solution at each stage to be effectively integrated. Recent advances include emulsion-PCR based DNA template preparation, the design and synthesis of novel reporter nucleotides and new surface attachment chemistries for DNA template. The integration of these advances will lead to the development of a high-throughput DNA sequencing system in the near future.

1. INTRODUCTION

DNA sequencing is a fundamental tool for biological science. The completion of the Human Genome Project has set the stage for screening genetic mutations to

PERSPECTIVES IN BIOANALYSIS, VOLUME 2
ISSN 1871-0069 DOI: 10.1016/S1871-0069(06)02006-4

identify disease genes on a genome-wide scale (Collins *et al.*, 2003). Recent estimates seem to indicate that the number of genes is relatively constant among vertebrates (IHGS Consortium, 2004). These results point to other factors such as gene variation, regulation and alternative splicing that are involved in the fundamental differences that separate humans from other animals as well as account for many of the primary differences that make-up each individual. DNA sequencing is a primary driving force behind the search for the fundamental regulatory regions that account for these differences. Decreased cost of sequencing is critical to the comparative genomic efforts including the ultimate goals of personalized medicine based on genetic and genomic information. Accuracy, speed and size of the instrument are critical considerations for the development of new DNA analysis methods that can be used directly in the hospitals and clinical settings, for forensics or for pathogen detection in the field. Accuracy is essential for genetic mutation detection and haplotype analysis.

The Sanger dideoxy chain-termination method (Sanger *et al.*, 1977) is currently the technique of choice for large-scale DNA sequencing projects. Widely used automated versions of this method employ either four differently end-labeled fluorescent primers or terminators to generate all the possible DNA fragments complementary to the template to be analyzed. The fragments terminating with the four different bases (A, C, G, T) are then separated at single-base pair resolution on sequencing gels and identified by the four distinct fluorescent emissions (Smith *et al.*, 1986; Prober *et al.*, 1987). Application of laser induced fluorescence for DNA sequencing is a major advancement for the automated DNA sequencing technology that makes large-scale genome sequencing initiatives possible. An "ideal" set of fluorophores for four-color Sanger DNA sequencing must consist of four different fluorophores. These fluorophores should have similar high molar absorbance at a common excitation wavelength, high fluorescence quantum yields, exhibit strong and well-separated fluorescence emissions and introduce the same relative mobility shift of the DNA sequencing fragments. These criteria cannot be met optimally by the spectroscopic properties of single fluorescent dye molecules, and indeed are poorly satisfied by the initially used sets of fluorescent tags. Ju *et al.*, (1995) overcame these constraints imposed by the use of single dyes and developed fluorescence energy transfer dyes for DNA sequencing that meet the performance criteria set out as above (Ju *et al.*, 1995). The higher sensitivity offered by these new sets of fluorescent dyes also allows the direct sequencing of large-template DNA (> 30 kb) with read lengths of over 700 bases per sequencing reaction, leading to significant progress in the large scale genome sequencing and mapping projects (Marra *et al.*, 1996; Lee *et al.*, 1997; Heiner *et al.*, 1998).

DNA sequencing by synthesis (SBS) is based on polymerase reaction, a 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 extension and detection of next nucleotide. The goal is to read out the DNA sequence serially during the polymerase reaction. This stands 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 of easy scale-up in parallel without the need for separations. Currently available fluorescent array scanners can easily detect over 100,000 sample spots arrayed on a glass surface (Schena *et al.*, 1995). Such array scanners allow fast screening of large areas with high resolution, allowing automated detection of hundreds of thousands and even millions of samples simultaneously.

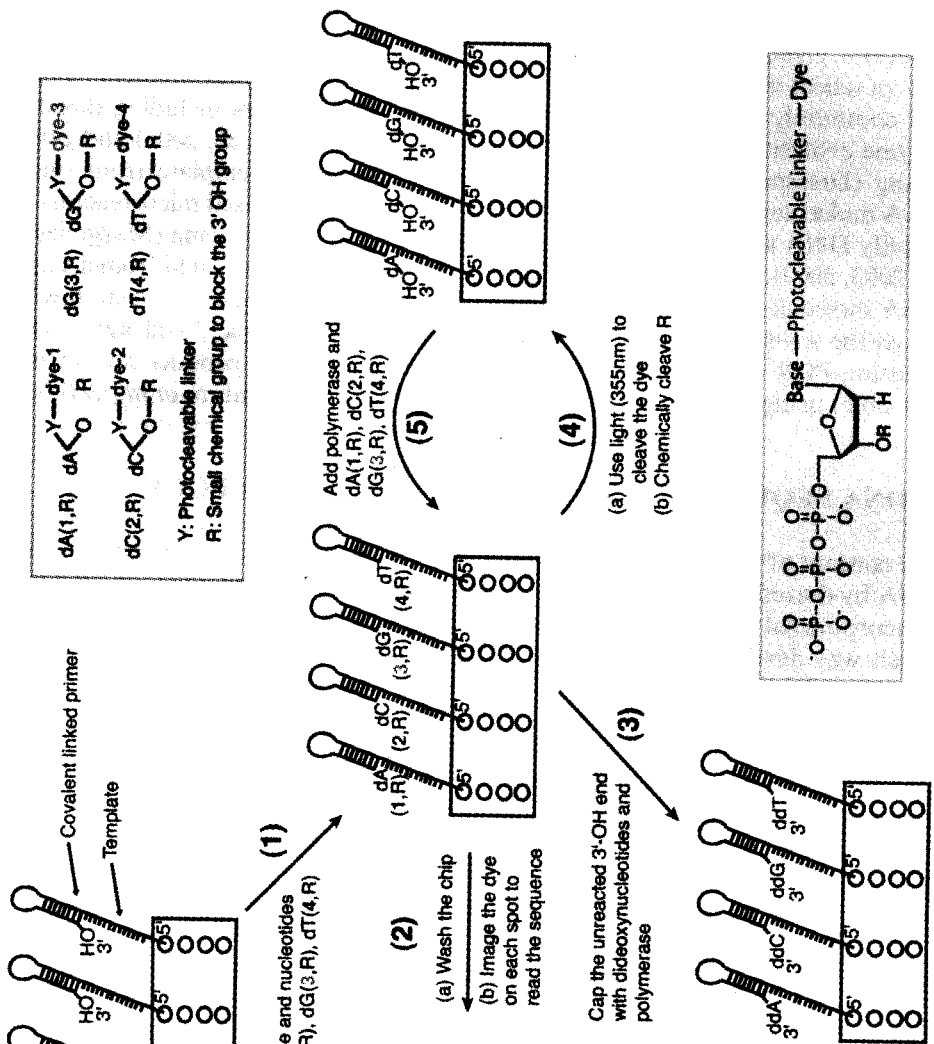
Several groups have recently reported significant advances in implementing new practical strategies for DNA sequencing. In these reports, emulsion-PCR, one commonly used technique for various biological assays including directed enzyme evolution (Tawfik and Griffiths, 1998; Ghadessy *et al.*, 2001) and genotyping (Dressman *et al.*, 2003) was used for generating template from single DNA molecules. Margulies and colleagues used emulsion-based microreactors to amplify DNA templates in a one-tube reaction for pyrosequencing (Margulies *et al.*, 2005, 2007). Beads containing the amplified templates generated from a single DNA molecule were then isolated in individual wells and reagents were flowed across the wells for the pyrosequencing reactions. Shendure and colleagues used emulsion-PCR on 1-μm beads to prepare DNA template to produce seven bases at a time, using a "sequencing by ligation" approach (Shendure *et al.*, 2005).

2. DNA SEQUENCING BY SYNTHESIS METHODOLOGY

The concept of DNA SBS was first revealed in 1988 with an attempt to sequence DNA by detecting the pyrophosphate group that is generated when a nucleotide is incorporated in a DNA polymerase reaction (Hyman, 1988). Pyrosequencing, which was developed based on this concept, has been explored for DNA sequencing (Ronaghi *et al.*, 1996). In this approach, each of the four dNTPs is added sequentially with a cocktail of enzymes and substrates in addition to the usual polymerase reaction components. If the added nucleotide is complementary with the first available base on the template, the nucleotide will be incorporated and a pyrophosphate will be released. The released pyrophosphate is converted to ATP by sulfurylase, and visible light is subsequently produced by firefly luciferase. If the added nucleotide is not incorporated, no light will be produced and the nucleotide will simply be degraded by the enzyme apyrase.

Pyrosequencing has been applied to single nucleotide polymorphism (SNP) detection (Ronaghi *et al.*, 1998) and genome sequencing (Margulies *et al.*, 2005). However, there are inherent difficulties in this method for determining the number of incorporated nucleotides in homopolymeric regions (e.g., a string of several Ts in a row) of the template. Additionally, each of the four nucleotides needs to be added and detected separately, which increases the overall detection time. The accumulation of undegraded nucleotides and other components could also lower the accuracy of the method when sequencing a long DNA template. Ideally, as one examines the fundamental limitations towards miniaturization, it is desirable to have a simple method to directly detect a reporter group attached to the nucleotide that is incorporated into a growing DNA strand in polymerase reaction, rather than relying on a complex enzymatic cascade.

We recently described an integrated SBS approach (Ju *et al.*, 2003) that is illustrated in Figure 1. This method relies on the use of a DNA polymerase


Covalent linked primer
Template
HO 3'
5'
(1)
Add polymerase and nucleotides
dA(1,R), dC(2,R), dG(3,R), dT(4,R)
dA(1,R) dA Y—dye-1 O R
dG(3,R) dG Y—dye-3 O R
dC(2,R) dC Y—dye-2 O R
dT(4,R) dT Y—dye-4 O R
Y: Photocleavable linker
R: Small chemical group to block the 3' OH group
(2)
(a) Wash the chip
(b) Image the dye on each spot to read the sequence
Fluorescence emission (nm)
510 550 602 650
A C G T
(3)
Cap the unreacted 3'-OH end with dideoxynucleotides and polymerase
ddA 3'
ddC 3'
ddG 3'
ddT 3'
(4)
(a) Use light (355nm) to cleave the dye
(b) Chemically cleave R
(5)
Add polymerase and dA(1,R), dC(2,R), dG(3,R), dT(4,R)
Base — Photocleavable Linker — Dye
OR H

reaction to read-out the DNA sequence, using novel reporter nucleotides for signal detection. After completion of the addition of each nucleotide the attached fluorescent reporter group is detected, determining the identity of the added nucleotide. The 3′ -OH moiety of each reporter nucleotide is also blocked by a functional group, which prevents the DNA polymerase from adding additional nucleotides. This blocking group needs to be easily removed, to generate a free 3′ -OH group for subsequent round of extension. This system increases the ability to accurately sequence through homopolymeric regions in the DNA template, as the addition of individual nucleotides are detected independently. In order to design an ideal system for SBS, new nucleotide analogues with the above properties must be developed.

Taking this and other factors into account, the following requirements must be met to make an entire SBS system into an efficient sequencing technology:

1. Standard cloning techniques to amplify DNA must be replaced by a high-throughput method for DNA template preparation.
2. After initial amplification, DNA templates must be physically arrayed in a format that allows each template to be probed multiple times.
3. Nucleotides must be reversible terminators (3′-OH is blocked) so that only a single nucleotide is added each step during SBS.
4. The 3′-OH blocking group used in SBS must be easily removed after detection for subsequent nucleotide addition.
5. The entire system must allow for simple washing and reagent additions between detection cycles.

Emulsion-PCR, which has been shown to have potential to address DNA template preparation for various sequencing platforms (Margulies *et al.*, 2005; Shendure *et al.*, 2005), can be readily adapted to the SBS approach shown in Figure 1. The remainder of this review is divided into two sections that describe advances in DNA attachment chemistries and in the synthesis of novel reporter

Fig. 1. In the SBS approach, a chip is constructed with immobilized DNA templates that are able to self-prime for initiating the polymerase reaction. Four nucleotide analogues are designed such that each is labeled with a unique fluorescent dye on the specific location of the base, and a small chemical group (R) to cap the 3′-OH group. Upon addition of the four nucleotide analogues and DNA polymerase, only the nucleotide analogue complementary to the next nucleotide on the template is incorporated by polymerase on each spot of the chip (step 1). After removing the excess reagents and washing away any unincorporated nucleotide analogues, a four-color fluorescence imager is used to image the surface of the chip, and the unique fluorescence emission from the specific dye on the nucleotide analogues on each spot of the chip will yield the identity of the nucleotide (step 2). After imaging, the small amount of unreacted 3′-OH group on the self-primed template moiety will be capped by excess ddNTPs (ddATP, ddGTP, ddTTP and ddCTP) and DNA polymerase to avoid interference with the next round of synthesis (step 3). The dye moiety will then be cleaved by light (~355 nm) and the R protecting group will be removed chemically to generate a free 3′-OH group with high yield (step 4). The self-primed DNA moiety on the chip at this stage is ready for the next cycle of the reaction to identify the next nucleotide sequence of the template DNA (step 5) (Ju *et al.*, 2003).

nucleotides. Each new advance further enables SBS to be developed into a viable DNA sequencing technology.

2.1. DNA attachment chemistries on surfaces

A variety of attachment chemistries have been used for the immobilization of DNA on surfaces. These chemistries have been driven by a wide range of applications including gene expression analysis using microarrays, chip based genotyping and SBS methods. For SBS, the primary requirement in the development of DNA immobilization chemistry is that the coupling reactions must produce high yields under conditions that are compatible with routine handling of DNA. These coupling conditions must not interrupt the phosphodiester bonds that comprise the DNA phosphate backbone, and should also not modify the primary amines found in guanines, cytosines and adenines.

One commonly used method is to attach streptavidin to a solid surface and then bind biotinylated DNA molecules to the immobilized streptavidin. While the streptavidin–biotin interaction is quite strong (Weber *et al.*, 1989), the hydrogen bonding interaction may prove problematic in multiple rounds of SBS. Streptavidin is also a large tetrameric protein, and its size will limit the number of available binding sites on the solid surface, limiting the number of DNA binding sites on the surface, resulting in decreased density of DNA molecules on the surface and decreased achievable read-length. Ideally, DNA would be attached covalently to the surface to eliminate any loss during multiple washing steps between nucleotide additions. The development of a chemoselective coupling chemistry for the immobilization of DNA on a solid surface is essential for accurate gene-expression measurement (Schena *et al.*, 1995) and polymorphism or mutation detection (Wang *et al.*, 1998; Debouck and Goodfellow, 1999). Because covalent coupling chemistries have been shown to typically lead to more stable DNA arrays than non-covalent chemistries, a variety of covalent coupling methods have been used for DNA immobilization on a solid surface (Beier and Hoheisel, 1999; Adessi *et al.*, 2000; Lindroos *et al.*, 2001). However, an additional improvement of the coupling chemistry for immobilizing DNA on a surface is required to achieve high selectivity and coupling efficiency. One ideal property required for the functional groups to be coupled (one from the DNA and the other from the surface) is the stability of the groups in aqueous conditions, which are typically needed to perform the coupling reaction.

We have explored Click Chemistry (Seo *et al.*, 2003, 2004), specifically azide–alkyne cycloaddition, to immobilize DNA on a glass chip for SBS (Figure 2). An amino modified glass surface is reacted with a bifunctional linker containing an NHS-ester and an alkyne on either end to functionalize the surface. Azido-labeled PCR product (created from a PCR reaction using an azido-labeled primer) is then attached to the surface using a copper catalyzed 1,3 dipolar azide–alkyne cycloaddition reaction. Alkaline conditions are then used to remove the unattached DNA strand, leaving only a single stranded DNA on the surface.

We have shown that DNA templates can be spotted in high density using the Click Chemistry with standard microarray spotters (Figure 3). In order to

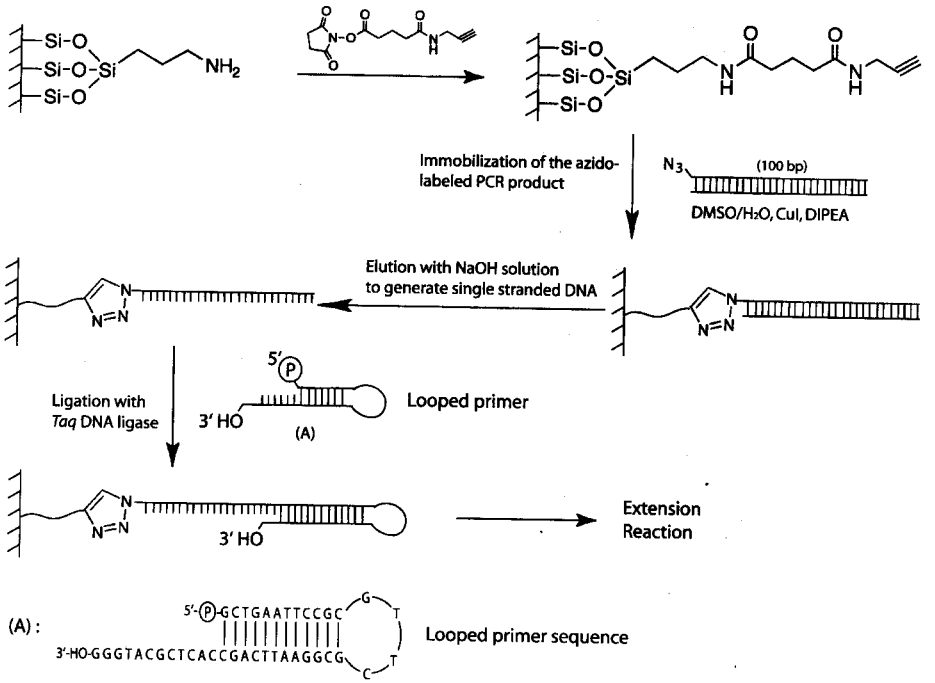


Fig. 2. Immobilization of an azido-labeled PCR DNA product on an alkynyl-functionalized surface and a ligation reaction between the immobilized single-stranded DNA template and a loop primer to form a self-priming DNA moiety on the chip. The sequence of the loop primer is shown in (A).

prevent dissociation of the primers and template during washing cycles of SBS, we ligated a looped primer directly to the template (Figure 2). This looped primer was designed such that the primer sequence self-hybridizes in a very efficient manner, making a universal primer adapter ideal for SBS. The loop sequence was carefully chosen to increase the stability of the hairpin structure (Nakano *et al.*, 2002), and has been shown to efficiently prime templates for SBS reactions (Seo *et al.*, 2005).

2.2. Novel reporter nucleotides

In order to design the reporter nucleotides used in the SBS extension reaction, it is important to examine the structure of the polymerase enzyme complexed with a DNA template, a primer and an incoming nucleotide in polymerase reaction. The 3D structure of the ternary complexes (Pelletier *et al.*, 1994) comprising a rat DNA polymerase, a DNA template-primer and a dideoxycytidine triphosphate (ddCTP) is shown in Figure 4.

What is apparent from this structure is 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 near the active amino acid residues of the polymerase and is therefore very crowded. Any group that is attached at the 3′-position of

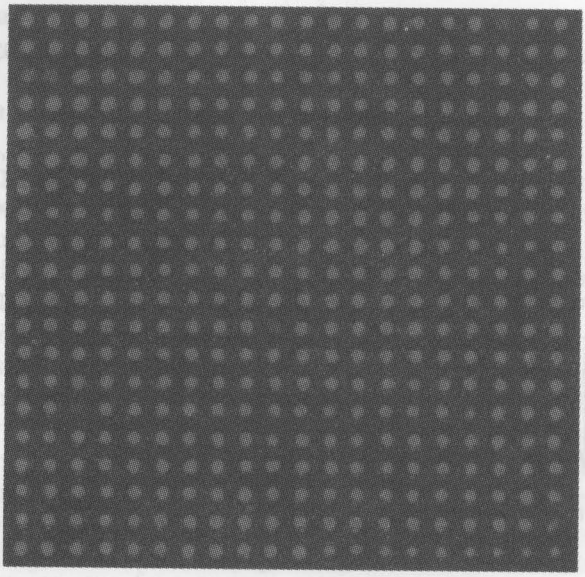


Fig. 3. Azido-labeled fluorescently modified DNA spotted on an alkyne surface using a standard microarray spotter demonstrating that the Click Chemistry used to covalently bind DNA templates to a solid surface can be easily scaled up with conventional high-throughput array spotting techniques.

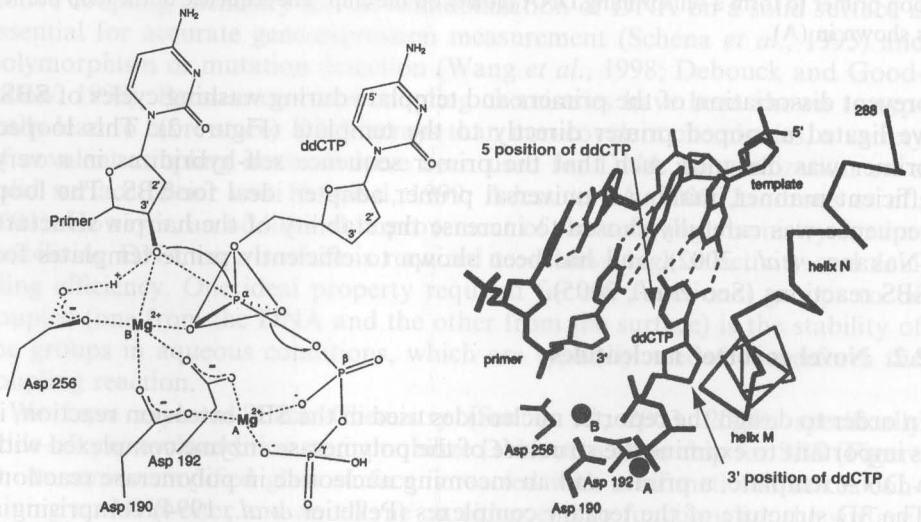


Fig. 4. The 3D structure of the ternary complexes comprising a rat DNA polymerase, a DNA template-primer and a dideoxycytidine triphosphate (ddCTP). The left side of the illustration shows the mechanism for the addition of ddCTP and the right side of the illustration shows the active site of the polymerase in the context of the polymerase–DNA complex. Note that the 3′-position of the dideoxyribose ring is very crowded, while ample space is available at the 5-position of the cytidine base.

the sugar must be small as to not interfere with the polymerase reaction. Large bulky dye molecules have been attached at the 5-position of pyrimidines and the 7-position of purines and used in enzymatic incorporation reactions such as in Sanger dideoxy-sequencing (Zhu *et al.*, 1994; Rosenblum *et al.*, 1997; Duthie *et al.*, 2002). We thus reasoned that if a unique fluorescent dye is attached to 5-position of the pyrimidines (T and C) and 7-position of purines (G and A) through a cleavable linker, and a small chemical moiety is used to cap the 3′-OH group, the resulting nucleotide analogues should be able to incorporate into the growing DNA strand as terminators. Upon removing the fluorophore and the 3′-OH capping group, the polymerase reaction will proceed to incorporate the next nucleotide analogue and detect the next base, as shown in Figure 1.

2.2.1. Nucleotide reporter groups for SBS

The reporter groups on the nucleotides for SBS must be easily detectable and then removed efficiently after detection to ensure maximum efficiency during each SBS cycle. Fluorescent dyes have been widely used as labels of nucleotides including the Sanger sequencing reactions. Braslavsky *et al.*, (2003) explored the use of photobleaching to eliminate the fluorescent signal in between nucleotide additions, without actually removing the fluorophores directly for SBS. However, studies have shown that the photobleached fluorophores remain with the DNA template and interfere with the DNA polymerase activity during the incorporation of each subsequent nucleotide. Thus, the fluorophores on the nucleotides need to be removed efficiently between each cycle of SBS.

Several different labile attachment chemistries for reporters have been used, including disulfide linkages (Mitra *et al.*, 2003) and photocleavable linkers (Ju *et al.*, 2003; Li *et al.*, 2003). The disulfide group can be chemically cleaved using 2-mercaptoethanol after the nucleotide incorporation and detection. However, the disulfide bond can be reversed and becomes destabilized under certain conditions (Pleasants *et al.*, 1989; Huyghues-Despointes and Nelson, 1992). Photocleavable linkers provide an effective and rapid method for removing the fluorophores from the nucleotide by using high intensity photons as reagents. We have developed a set of such photocleavable fluorescent nucleotide analogues (Seo *et al.*, 2005), using 2-nitrobenzyl group as the photocleavable linker to attach the fluorophore to each base at the 5-position of the pyrimidines and the 7-position of the purines (Figure 5).

These nucleotide analogues have been shown to be good substrates for incorporation into DNA by the commonly used DNA polymerase *Thermo Sequenase*, and have also been shown to incorporate efficiently in a primer extension assay. The products from the polymerase extension reaction using the four photocleavable fluorescent nucleotide analogues described in Figure 5 and the photocleavage products are shown in Figure 6. The fluorescent dyes attached to these nucleotides are removed quickly and efficiently under near-UV irradiation in nearly quantitative yield, and these photocleavage conditions are compatible for use with DNA samples (Seo *et al.*, 2005), showing no DNA damage during repeated exposure to photolysis (Figure 7).

As a demonstration of the feasibility of using these nucleotides for SBS, the above-described Click Chemistry for DNA immobilization has been combined

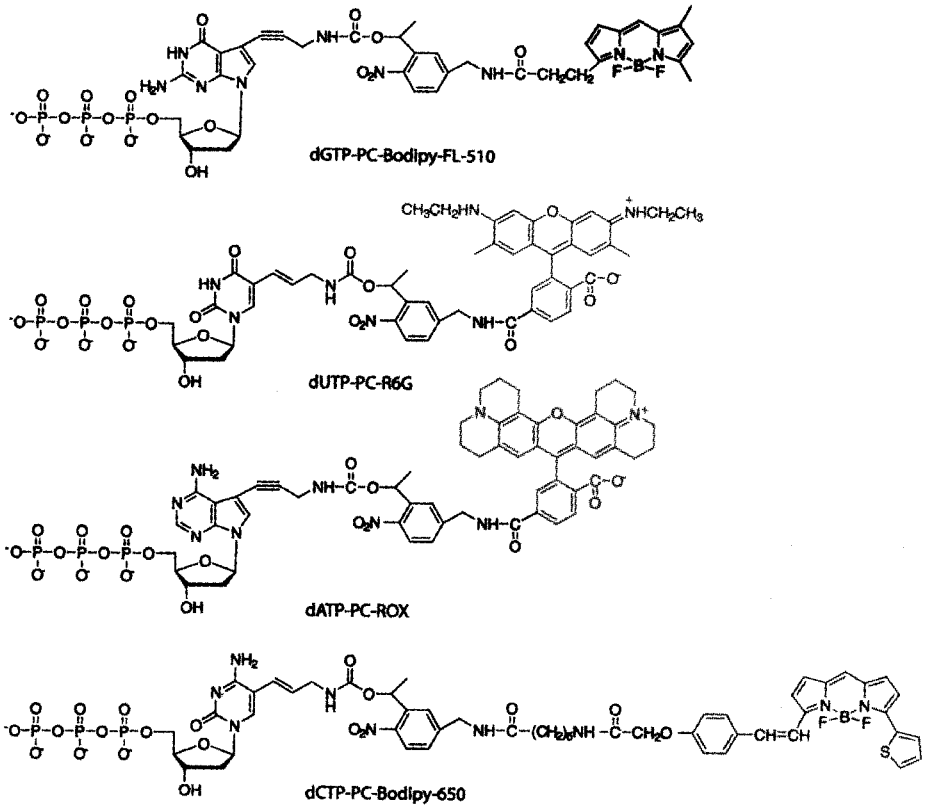
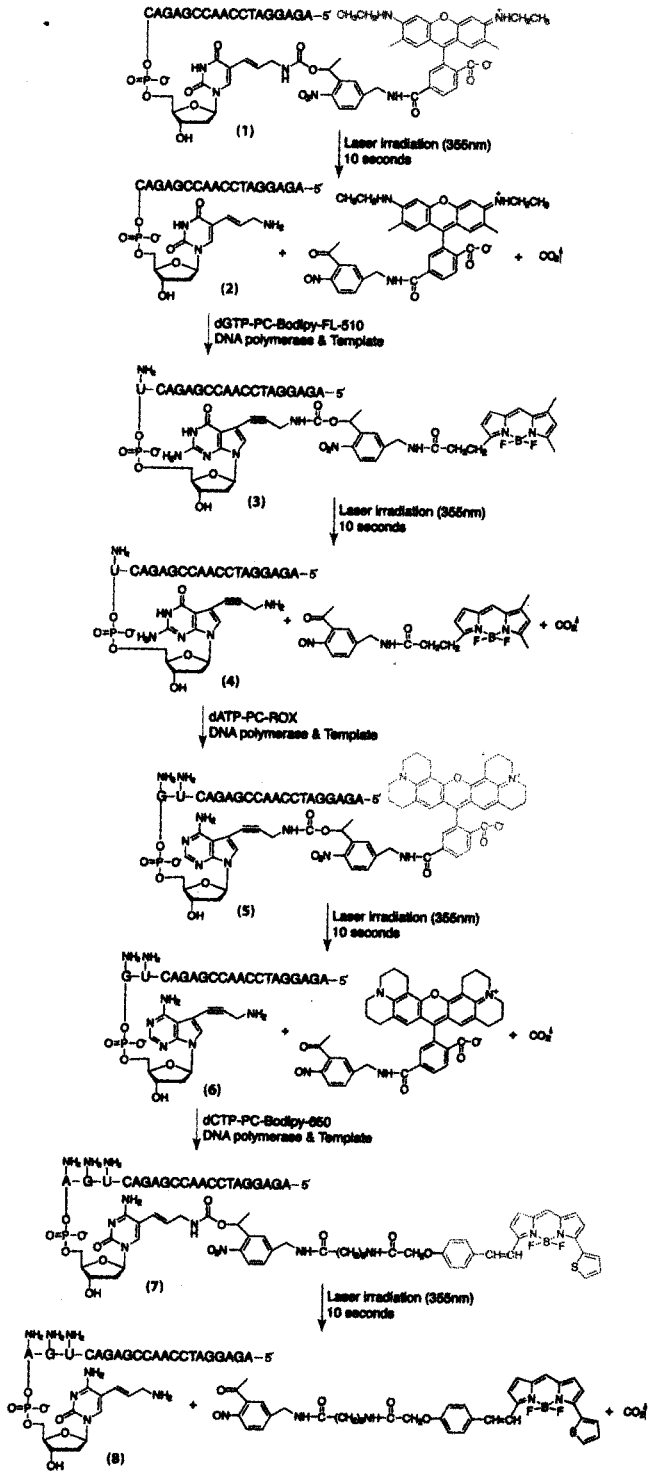


Fig. 5. Structures of four nucleotide analogues labeled through a photocleavable linker (PC) using four fluorophores with distinct fluorescent emissions, dGTP-PC-Bodipy-FL-510 ($\lambda_{abs(max)} = 502$ nm; $\lambda_{em(max)} = 510$ nm), dUTP-PC-R6G ($\lambda_{abs(max)} = 525$ nm; $\lambda_{em(max)} = 550$ nm), dATP-PC-ROX ($\lambda_{abs(max)} = 585$ nm; $\lambda_{em(max)} = 602$ nm) and dCTP-PC-Bodipy-650 ($\lambda_{abs(max)} = 630$ nm; $\lambda_{em(max)} = 650$ nm).

with the universal looped primer to perform a series of extension reactions on a chip surface using all four photocleavable fluorescent nucleotide analogues. The principal advantage offered by the use of a self-priming moiety, as compared to using separate primers and templates, is that the covalent linkage of the primer to the template in the self-priming moiety completely prevents any possible dissociation of the primer from the template, even under vigorous washing conditions. Furthermore, the possibility of mispriming is considerably reduced,

Fig. 6. Products of DNA extension reaction and photolysis generated in solution phase to characterize the four different photocleavable fluorescent nucleotide analogues (dUTP-PC-R6G, dGTP-PC-Bodipy-FL-510, dATP-PC-ROX and dCTP-PC-Bodipy-650). After each extension reaction, the DNA extension product is analyzed by MALDI-TOF MS measurement to verify that it is the correct extension product (see Figure 7). Photolysis is then performed to produce a DNA product that is used as a primer for the next DNA extension reaction.


CAGAGCCAACCTAGGAGA–5′
(1)
Laser irradiation (355nm)
10 seconds
(2)
dGTP-PC-Bodipy-FL-510
DNA polymerase & Template
(3)
Laser irradiation (355nm)
10 seconds
(4)
dATP-PC-ROX
DNA polymerase & Template
(5)
Laser irradiation (355nm)
10 seconds
(6)
dCTP-PC-Bodipy-650
DNA polymerase & Template
(7)
Laser irradiation (355nm)
10 seconds
(8)

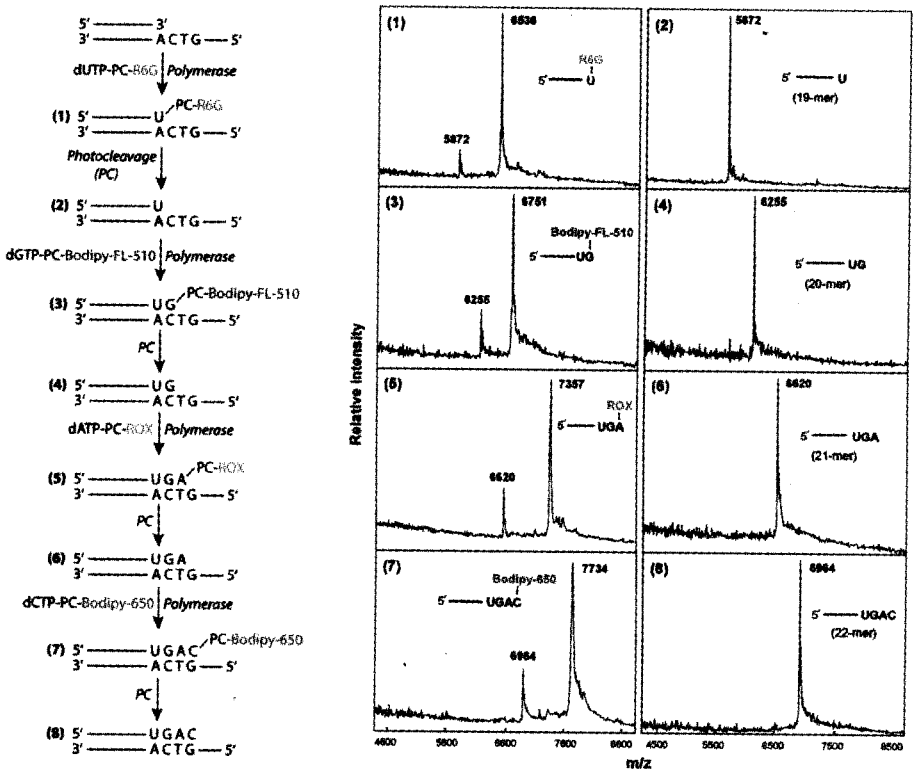


Fig. 7. The polymerase extension scheme (left) and MALDI-TOF MS spectra of the four consecutive extension products and their photocleavage products (right). Primer extended with dUTP-PC-R6G (1), and its photocleavage product 2; product 2 extended with dGTP-PC-Bodipy-FL-510 (3), and its photocleavage product 4; product 4 extended with dATP-PC-ROX (5), and its photocleavage product 6; product 6 extended with dCTP-PC-Bodipy-650 (7), and its photocleavage product 8. After 10 s of irradiation with a laser at 355 nm, photocleavage is complete with all the fluorophores cleaved from the extended DNA products. After each nucleotide is incorporated, the MS spectrum of each extension product is obtained to verify whether the incorporation was successful. In the spectra (1, 3, 5, 7) a small photocleavage product is visible due to the laser irradiation used in MALDI-TOF MS analysis. After photolysis, the spectra (2, 4, 6, 8) are obtained demonstrating complete photocleavage of each fluorophore from the extended DNA products.

and a universal loop primer can be used for all the templates allowing enhanced accuracy and ease of operation.

The four-color SBS sequencing results (Seo *et al.*, 2005) are shown in Figure 8. The structure of the self-priming DNA moiety is shown schematically in the upper panel, with the first 12-nucleotide sequence immediately after the priming site. The sequencing reaction on the chip was initiated by extending the self-priming DNA using dATP-PC-ROX (complementary to the T on the template) and Thermo Sequenase DNA polymerase. After washing, the extension of the primer by a single fluorescent nucleotide was confirmed by observing an orange signal (the emission signal from ROX) in a four-color fluorescent microarray scanner

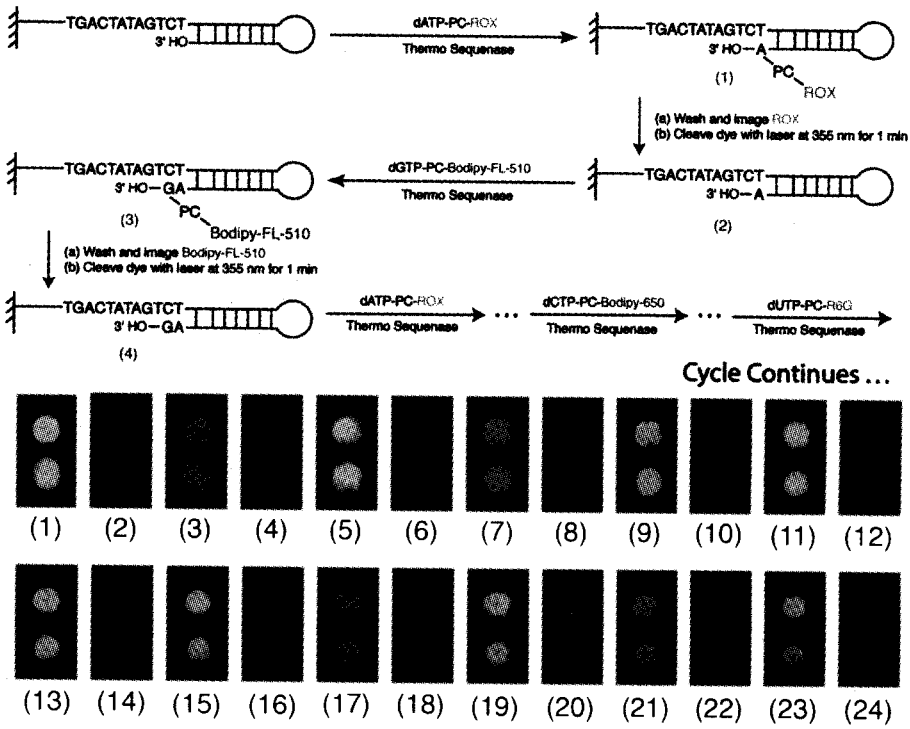


Fig. 8. Schematic representation of SBS on a chip using four PC fluorescent nucleotides (Upper panel) and the scanned fluorescence images for each step of SBS on a chip (Lower panel). (1) Incorporation of dATP-PC-ROX; (2) photocleavage of PC-ROX; (3) incorporation of dGTP-PC-Bodipy-FL-510; (4) photocleavage of PC-Bodipy-FL-510; (5) incorporation of dATP-PC-ROX; (6) photocleavage of PC-ROX; (7) incorporation of dCTP-PC-Bodipy-650; (8) photocleavage of PC-Bodipy-650; (9) incorporation of dUTP-PC-R6G; (10) photocleavage of PC-R6G; (11) incorporation of dATP-PC-ROX; (12) photocleavage of PC-ROX; (13) incorporation of dUTP-PC-R6G; (14) photocleavage of PC-R6G; (15) incorporation of dATP-PC-ROX; (16) photocleavage of PC-ROX; (17) incorporation of dGTP-PC-Bodipy-FL-510; (18) photocleavage of PC-Bodipy-FL-510; (19) incorporation of dUTP-PC-R6G; (20) photocleavage of PC-R6G; (21) incorporation of dCTP-PC-Bodipy-650; (22) photocleavage of PC-Bodipy-650; (23) incorporation of dATP-PC-ROX and (24) photocleavage of PC-ROX.

(Figure 8 [1]). After detection of the fluorescent signal, the surface was irradiated at 355 nm using an Nd-YAG laser to cleave the fluorophore. The surface was then washed, and a negligible residual fluorescent signal was detected, confirming the complete photocleavage of the fluorophore (Figure 8 [2]). This was followed by incorporation of the next fluorescent nucleotide complementary to the subsequent base on the template. The entire process of incorporation, detection and photocleavage was performed multiple times using the four photocleavable fluorescent nucleotide analogues to identify 12 successive bases in the DNA template. Thus, two conditions for a future SBS system have been satisfied – these newly developed nucleotide analogues have been shown to be excellent substrates for the

DNA polymerase, and the fluorophore group could be cleaved efficiently from the nucleotide using near UV irradiation. These results are important with respect to enhancing the speed of each cycle in SBS for high-throughput DNA analysis.

2.3. Blocking of 3′-hydroxyl groups

Another critical requirement to sequence DNA unambiguously using SBS methods is a suitable chemical moiety to cap the 3′-OH of the nucleotide so that it terminates the polymerase reaction after addition of a single nucleotide. This stepwise addition then allows the identification of the incorporated nucleotide at each step. A 3′-OH capping group on each of the nucleotides also permits all four nucleotides to be present together during the SBS extension reaction, resulting in a significant decrease in the number of cycles needed for sequencing. The 3′-OH capping group then also needs to be labile, and be able to be efficiently removed to regenerate the 3′-OH to permit the polymerase reaction to proceed to the next round. Conversely, the stepwise addition of nucleotides with a free 3′-OH group would present inherent difficulties in the detection of the sequence of homopolymeric regions.

The principal challenge posed by this requirement is the incorporation ability of the 3′-modified nucleotide by DNA polymerase into the growing DNA strand. Several groups have focused on the design and synthesis of nucleotides that have a photocleavable fluorophore on the 3′-position, as a simple way to cap the 3′-OH directly with the reporter group (Metzker *et al.*, 1994; Welch and Burgess, 1999). The rationale of this scheme is that after the fluorophore is removed, the 3′-OH group would be regenerated, and allow subsequent nucleotide addition. However, the incorporation by DNA polymerase of such a nucleotide with a photocleavable fluorescent dye on the 3′-position into a growing DNA strand has not been successfully reported so far. This is primarily due to the difficulty of DNA polymerases recognizing nucleotides with the 3′-position modified with large fluorophores. As noted earlier (Figure 4), the 3′-position on the sugar ring of a nucleotide is very close to the amino acid residues in the active site of the DNA polymerase. Thus, any bulky modification at this position will sterically hinder the DNA polymerase and prevent the nucleotide from being incorporated.

A second challenge critical to the overall efficiency and final achievable read length of SBS is the efficient removal of the 3′-OH capping group once the fluorescence signal is detected. Any DNA strand that has a remaining 3′-OH blocking group will inhibit the polymerase reaction and therefore lose its contribution to detect the next base in the template. Furthermore, the subsequent removal of the 3′-OH blocking group in subsequent SBS rounds would liberate that molecule for further addition steps, contributing to asynchronous "noise". Since each cycle of SBS sequencing essentially requires the complete removal of the 3′-OH capping group, a rapid and highly efficient process is required. It is important to use a small functional group that provides no hindrance to the DNA polymerase, while also stable enough to withstand DNA extension reaction conditions, and able to be removed easily and rapidly to regenerate a free 3′-OH under specific conditions.

Recently, we have developed a photocleavable fluorescent nucleotide with an allyl group capping the 3′-OH. These nucleotide analogues have been shown to act as substrates for a mutant DNA polymerase (Ruparel *et al.*, 2005). Figure 9 shows the synthetic scheme for the preparation of this novel nucleotide analogue, 3′-*O*-allyl-dUTP-PC-Bodipy-FL-510. Our selection of an allyl group is based on the fact that an allyl moiety, being relatively small and inert, would not provide significant hindrance for the polymerase reaction, and therefore would allow the incoming 3′-*O*-allyl modified nucleotide analogue to be accepted as a substrate by DNA polymerase. The entire cycle of a polymerase reaction using 3′-*O*-allyl-dUTP-PC-Bodipy-FL-510 as a reversible terminator is depicted in Figure 10. The extension product 11 obtained using 3′-*O*-allyl-dUTP-PC-Bodipy-FL-510 and a DNA polymerase was purified using HPLC and analyzed using MALDI-TOF MS. The base in the template immediately adjacent to the priming site was 'A'. Thus, if 3′-*O*-allyl-dUTP-PC-Bodipy-FL-510 was accepted by the polymerase as a terminator, the primer would extend by one base and then the reaction would terminate. Our results indicate that this was indeed the case.

After confirming that the extension reaction was successful, we irradiated it with near UV light at 355 nm for 10 s to cleave the fluorophore from the DNA, generating product 12. In the SBS system, this step would ensure that there would be no carryover of the fluorescence signal into the next incorporation cycle, so as to prevent the generation of ambiguous data at each step. The photocleavage product 12 was then incubated with a palladium catalyst system in aqueous solution to perform deallylation. The deallylated DNA product 13 carrying a free 3′-OH group was purified by reverse phase HPLC and then used as a primer in a second DNA extension reaction to prove that the regenerated 3′-OH was capable of allowing the polymerase reaction to continue. For the extension reaction, we used a photocleavable fluorescent nucleotide dGTP-PC-Bodipy-FL-510 and Thermo Sequenase DNA polymerase. The extension product 14 was irradiated as above, for 10 s to generate photocleavage product 15

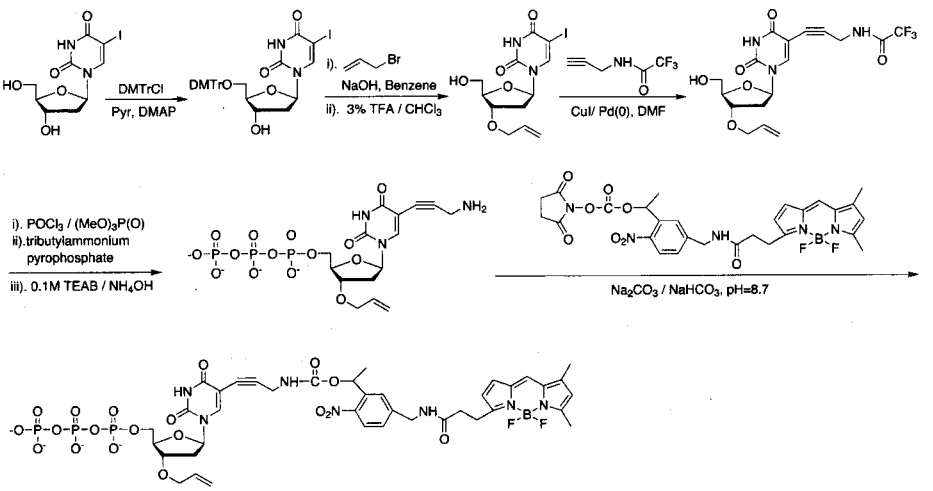


Fig. 9. Synthesis of 3′-*O*-allyl-dUTP-PC-Bodipy-FL-510.

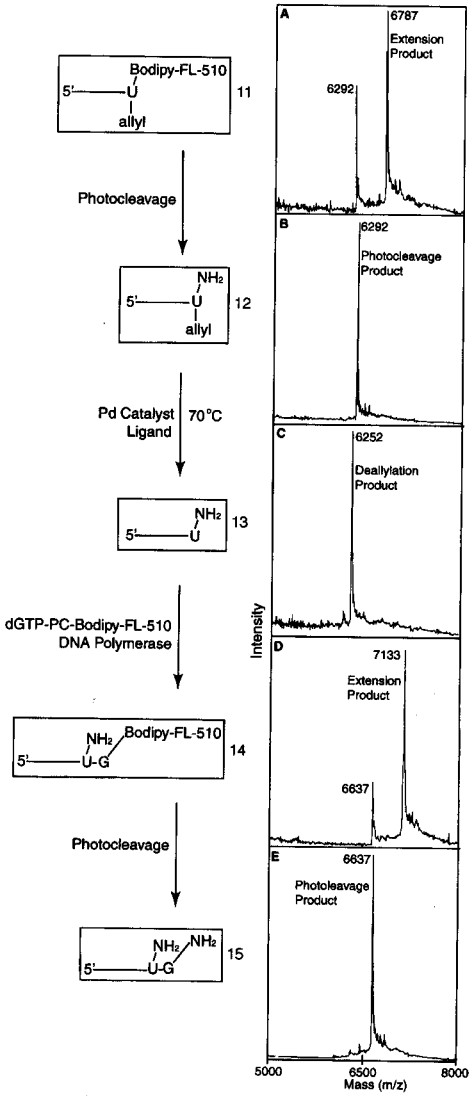


Fig. 10. MALDI-TOF MS results for each step of a polymerase reaction cycle using 3′-*O*-allyl-dUTP-PC-Bodipy-FL-510 as a reversible terminator. (A) Peak at m/z 6787 corresponding to the primer extension product 11 obtained using 3′-*O*-allyl-dUTP-PC-Bodipy-FL-510 and the 9°N Polymerase. (B) Peak at m/z 6292 corresponding to the photocleavage product 12. (C) Peak at m/z 6252 corresponding to the photocleavage product without the allyl group 13 obtained after incubation with the catalyst and ligand at 70°C. (D) Peak at m/z 7133 corresponding to the extension product 14 from the purified deallylated product using dGTP-PC-Bodipy-FL-510 and Thermo Sequenase DNA polymerase. (E) Peak at m/z 6637 corresponding to the photocleavage product 15.

and hence complete an entire reversible termination and extension cycle. The deallylation reaction was shown to achieve near quantitative yield under mild reaction conditions in an aqueous environment, providing an effective method to modulate the 3′-OH group of the nucleotides for SBS.

3. CONCLUSION

A substantial number of advances have been made toward the goal of making DNA SBS a viable technology for genomic research. This includes the rapid large-scale amplification of genomic libraries through emulsion-PCR, new developments in DNA attachment chemistries that allow increased array densities and novel reporter nucleotides as reversible terminators for polymerase reaction. These nucleotide analogues allow the enzymatic addition of a single nucleotide, direct detection to determine its identity, efficient removal of the reporter fluorophore and the 3′-OH blocking group to allow subsequent nucleotide additions. The integration of these developments will lead SBS to be developed into a high-throughput DNA sequencing platform for the era of personalized medicine.

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health Grants P50 HG002806 and R01 HG003582, and the Packard Fellowship for Science and Engineering.

REFERENCES

Adessi, C., Matton, G., Ayala, G., Turcatti, G., Mermod, J. J., Mayer, P. and Kawashima, E. (2000). Solid phase DNA amplification: characterization of primer attachment and amplification mechanisms. Nucleic Acids Res. 28, e87.

Beier, M. and Hoheisel, J. D. (1999). Versatile derivatisation of solid support media for covalent bonding on DNA-microchips. Nucleic Acids Res. 27, 1970–1977.

Braslavsky, I., Hebert, B., Kartalov, E. and Quake, S. R. (2003). Sequence information can be obtained from single DNA molecules. Proc. Natl. Acad. Sci. USA 100, 3960–3964.

Collins, F. S., Green, E. D., Guttmacher, A. E. and Guyer, M. S. (2003). A vision for the future of genomics research. Nature 422, 835–847.

Debouck, C. and Goodfellow, P. N. (1999). DNA microarrays in drug discovery and development. Nat. Genet. 1, 48–50.

Dressman, D., Yan, H., Traverso, G., Kinzler, K. W. and Vogelstein, B. (2003). Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. Proc. Natl. Acad. Sci. USA 100, 8817–8822.

Duthie, R. S., Kalve, I. M., Samols, S. B., Hamilton, S., Livshin, I., Khot, M., Nampalli, S., Kumar, S. and Fuller, C. W. (2002). Novel cyanine dye-labeled dideoxynucleoside triphosphates for DNA sequencing. Bioconjug. Chem. 13, 699–706.

Ghadessy, F. J., Ong, J. L. and Holliger, P. (2001). Directed evolution of polymerase function by compartmentalized self-replication. Proc. Natl. Acad. Sci. USA 98, 4552–4557.

Heiner, C. R., Hunkapiller, K. L., Chen, S. M., Glass, J. I. and Chen, E. Y. (1998). Sequencing multimegabase-template DNA with BigDye terminator chemistry. Genome Res. 8, 557–561.

Huyghues-Despointes, B. M. and Nelson, J. W. (1992). Stabilities of disulfide bond intermediates in the folding of apamin. Biochemistry 31, 1476–1483.

Hyman, E. D. (1988). A new method of sequencing DNA. Anal. Biochem. 174, 423–436.

International Human Genome Sequencing Consortium (2004). Finishing the euchromatic sequence of the human genome. Nature 431, 931–945.

Ju, J., Li, Z., Edwards, J. R. and Itagota, Y. (2003). Massive parallel method for decoding DNA and RNA. US Patent No. 6 664 079.

Ju, J., Ruan, C., Fuller, C. W., Glazer, A. N. and Mathies, R. A. (1995). Fluorescence energy transfer dye-labeled primers for DNA sequencing and analysis. Proc. Natl. Acad. Sci. USA 92, 4347–4351.

Lee, L. G., Spurgeon, S. L., Heiner, C. R., Benson, S. C., Rosenblum, B. B., Menchen, S. M., Graham, R. J., Constantinescu, A., Upadhya, K. G. and Cassel, J. M. (1997). New energy transfer dyes for DNA sequencing. Nucleic Acids Res. 25, 2816–2822.

Li, Z., Bai, X., Ruparel, H., Kim, S., Turro, N. J. and Ju, J. (2003). A photocleavable fluorescent nucleotide for DNA sequencing and analysis. Proc. Natl. Acad. Sci. USA 100, 414–419.

Lindroos, K., Liljedahl, U., Raitio, M. and Syvänen, A. (2001). Minisequencing on oligonucleotide microarrays: comparison of immobilisation chemistries. Nucleic Acids Res. 29, e69.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y. J., Chen, Z., Dewell, S. B., Du, L., Fierro, J. M., Gomes, X. V., Godwin, B. C., He, W., Helgesen, S., Ho, C. H., Irzyk, G. P., Jando, S. C., Alenquer, M. L., Jarvie, T. P., Jirage, K. B., Kim, J. B., Knight, J. R., Lanza, J. R., Leamon, J. H., Lefkowitz, S. M., Lei, J., Li, K. L., Lohman, H., Lu, V. B., Makhijani, K. E., McDade, M. P., McKenna, E. W., Myers, M., Nickerson, E., Nobile, J. R., Plant, R., Puc, B. P., Ronan, M. T., Roth, G. T., Sarkis, G. J., Simons, J. F., Simpson, J. W., Srinivasan, M., Tartaro, K. R., Tomasz, A., Vogt, K. A., Volkmer, G. A., Wang, S. H., Wang, Y., Weiner, M. P., Yu, P., Begley, R. F. and Rothberg, J. M. (2005). Genome sequencing in microfabricated high-density picolitre reactors. Nature 437, 376–380.

Margulies,, M., Jarvie,, T. P., Knight,, J. R. and Simons,, J. F. (2007). The 454 Life Sciences picoliter sequencing system. In: K. R. Mitchelson (Ed), New High Throughput Technologies for DNA Sequencing and Genomics (pp. 151–186). Elsevier, Amsterdam.

Marra, M., Weinstock, L. A. and Mardis, E. R. (1996). End sequence determination from large insert clones using energy transfer fluorescent primers. Genome Res. 6, 1118–1122.

Metzker, M. L., Raghavachari, R., Richards, S., Jacutin, S. E., Civitello, A., Burgess, K. and Gibbs, R. A. (1994). Termination of DNA synthesis by novel 3′-modified-deoxyribonucleoside 5′-triphosphates. Nucleic Acids Res. 22, 4259–4267.

Mitra, R. D., Shendure, J., Olejnik, J., Krzymanska-Olejnik, E. and Church, G. M. (2003). Fluorescent in situ sequencing on polymerase colonies. Anal. Biochem. 320, 55–65.

Nakano, M., Moody, E. M., Liang, J. and Bevilacqua, P. C. (2002). Selection for thermodynamically stable DNA tetraloops using temperature gradient gel electrophoresis reveals four motifs: d(cGNNAg), d(cGNABg), d(cCNNGg), and d(gCNNGc). Biochemistry 41, 14,281–14,292.

Pelletier, H., Sawaya, M. R., Kumar, A., Wilson, S. H. and Kraut, J. (1994). Structures of ternary complexes of rat DNA polymerase beta, a DNA template-primer, and ddCTP. Science 264, 1891–1903.

Pleasants, J. C., Guo, W. and Rabenstein, D. L. (1989). A comparative study of the kinetics of selenol/diselenide and thiol/disulfide exchange reactions. J. Am. Chem. Soc. 111, 6553–6558.

Prober, J. M., Trainor, G. L., Dam, R. J., Hobbs, F. W., Robertson, C. W., Zagursky, R. J., Cocuzza, A. J., Jensen, M. A. and Baumeister, K. (1987). A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. Science 238, 336–341.

Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlén, M. and Nyren, P. (1996). Real-time DNA sequencing using detection of pyrophosphate release. Anal. Biochem. 242, 84–89.

Ronaghi, M., Uhlén, M. and Nyren, P. (1998). A sequencing method based on real-time pyrophosphate. Science 281, 363–365.

Rosenblum, B. B., Lee, L. G., Spurgeon, S. L., Khan, S. H., Menchen, S. M., Heiner, C. R. and Chen, S. M. (1997). New dye-labeled terminators for improved DNA sequencing patterns. Nucleic Acids Res. 25, 4500–4504.

Ruparel, H., Bi, L., Li, Z., Bai, X., Kim, D. H., Turro, N. J. and Ju, J. (2005). Design and synthesis of a 3′-*O*-allyl photocleavable fluorescent nucleotide as a reversible terminator for DNA sequencing by synthesis. Proc. Natl. Acad. Sci. USA 102, 5932–5937.

Sanger, F., Nicklen, S. and Coulsen, A. R. (1977). Sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463–5467.

Schena, M., Shalon, D., Davis, R. W. and Brown, P. O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270, 467–470.

Seo, T. S., Bai, X., Kim, D. H., Meng, Q., Shi, S., Ruparel, H., Li, Z., Turro, N. J. and Ju, J. (2005). Four-color DNA sequencing by synthesis on a chip using photocleavable fluorescent nucleotides. Proc. Natl. Acad. Sci. USA 102, 5926–5931.

Seo, T. S., Bai, X., Ruparel, H., Li, Z., Turro, N. J. and Ju, J. (2004). Photocleavable fluorescent nucleotides for DNA sequencing on a chip constructed by site-specific coupling chemistry. Proc. Natl. Acad. Sci. USA 101, 5488–5493.

Seo, T. S., Li, Z., Ruparel, H. and Ju, J. (2003). Click chemistry to construct fluorescent oligonucleotides for DNA sequencing. J. Org. Chem. 68, 609–612.

Shendure, J., Porreca, G. J., Reppas, N. B., Lin, X., McCutcheon, J. P., Rosenbaum, A. M., Wang, M. D., Zhang, K., Mitra, R. D. and Church, G. M. (2005). Accurate multiplex polony sequencing of an evolved bacterial genome. Science 309, 1728–1733.

Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R., Heiner, C., Kent, S. B. and Hood, L. E. (1986). Fluorescence detection in automated DNA sequence analysis. Nature 321, 674–679.

Tawfik, D. S. and Griffiths, A. D. (1998). Man-made cell-like compartments for molecular evolution. Nat. BioTech. 16, 652–656.

Wang, D. G., Fan, J. B., Siao, C. J., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., Kruglyak, L., Stein, L., Hsie, L., Topaloglou, T., Hubbell, E., Robinson, E., Mittmann, M., Morris, M. S., Shen, N., Kilburn, D., Rioux, J., Nusbaum, C., Rozen, S., Hudson, T. J., Lipshutz, R., Chee, M. and Lander, E. S. (1998). Large-scale identification, mapping and genotyping of single-nucleotide polymorphisms in the human genome. Science 280, 1077–1082.

Weber, P. C., Ohlendorf, D. H., Wendoloski, J. J. and Salemme, F. R. (1989). Structural origins of high-affinity biotin binding to streptavidin. Science 243, 85–88.

Welch, M. B. and Burgess, K. (1999). Synthesis of fluorescent, photolabile 3′-*O*-protected nucleoside triphosphates for the base addition sequencing scheme. Nucleosides Nucleotides 18, 197–201.

Zhu, Z., Chao, J., Yu, H. and Waggoner, A. S. (1994). Directly labeled DNA probes using fluorescent nucleotides with different length linkers. Nucleic Acids Res. 22, 3418–3422.