

Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators

Jingyue Ju^{*†‡}, Dae Hyun Kim^{*§}, Lanrong Bi^{*†}, Qinglin Meng^{*†¶}, Xiaopeng Bai^{*†¶}, Zengmin Li^{*†}, Xiaoxu Li^{*†}, Mong Sano Marma^{*†}, Shundi Shi^{*}, Jian Wu^{*†¶}, John R. Edwards^{*†}, Airen Romu^{*}, and Nicholas J. Turro^{*†¶}

^{*}Columbia Genome Center, Columbia University College of Physicians and Surgeons, New York, NY 10032; and Departments of [†]Chemical Engineering, [¶]Chemistry, and [§]Biomedical Engineering, Columbia University, New York, NY 10027

Contributed by Nicholas J. Turro, October 26, 2006 (sent for review October 1, 2006)

DNA sequencing by synthesis (SBS) on a solid surface during polymerase reaction offers a paradigm to decipher DNA sequences. We report here the construction of such a DNA sequencing system using molecular engineering approaches. In this approach, four nucleotides (A, C, G, T) are modified as reversible terminators by attaching a cleavable fluorophore to the base and capping the 3'-OH group with a small chemically reversible moiety so that they are still recognized by DNA polymerase as substrates. We found that an allyl moiety can be used successfully as a linker to tether a fluorophore to 3'-O-allyl-modified nucleotides, forming chemically cleavable fluorescent nucleotide reversible terminators, 3'-O-allyl-dNTPs-allyl-fluorophore, for application in SBS. The fluorophore and the 3'-O-allyl group on a DNA extension product, which is generated by incorporating 3'-O-allyl-dNTPs-allyl-fluorophore in a polymerase reaction, are removed simultaneously in 30 s by Pd-catalyzed deallylation in aqueous buffer solution. This one-step dual-deallylation reaction thus allows the reinitiation of the polymerase reaction and increases the SBS efficiency. DNA templates consisting of homopolymer regions were accurately sequenced by using this class of fluorescent nucleotide analogues on a DNA chip and a four-color fluorescent scanner.

DNA chip

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). Accurate high-throughput DNA sequencing methods are needed to explore the complete human genome sequence for applications in clinical medicine and health care. To overcome the limitations of the current electrophoresis-based sequencing technology (2–5), a variety of new DNA-sequencing methods have been investigated. Such approaches include sequencing by hybridization (6), mass spectrometry-based sequencing (7–9), sequence-specific detection of single-stranded DNA using engineered nanopores (10), and sequencing by ligation (11). More recently, DNA sequencing by synthesis (SBS) approaches such as pyrosequencing (12), sequencing of single DNA molecules (13), and polymerase colonies (14) have been widely explored.

The concept of DNA SBS was 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 (15). Pyrosequencing, which was developed based on this concept and an enzymatic cascade, has been explored for genome sequencing (16). However, there are inherent difficulties in this method for determining the number of incorporated nucleotides in homopolymeric regions 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. It is thus 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 the polymerase reaction rather than relying on a complex enzymatic cascade. The SBS scheme based on fluorescence detection coupled with a chip format has the potential to markedly increase the throughput of DNA sequencing projects. Consequently, several groups have investigated such a system with an aim to construct an ultra high-throughput DNA sequencing method (17, 18). Thus far, no complete success of using such a system to unambiguously sequence DNA has been published.

Previous work in the literature exploring the SBS method is mostly focused on designing and synthesizing a cleavable chemical moiety that is linked to a fluorescent dye to cap the 3'-OH group of the nucleotides (19–21). The rationale is that, after the fluorophore is removed, the 3'-OH would be regenerated to allow subsequent nucleotide addition. However, no success has been reported for the incorporation of such a nucleotide with a cleavable fluorescent dye on the 3' position by DNA polymerase into a growing DNA strand. The reason is that the 3' position on the deoxyribose is very close to the amino acid residues in the active site of the polymerase, and the polymerase is therefore sensitive to modification in this area of the ribose ring, especially with a large fluorophore (22).

It is known that some modified DNA polymerases are highly tolerable for nucleotides with extensive modifications with bulky groups such as energy transfer dyes at the 5-position of the pyrimidines (T and C) and 7-position of purines (G and A) (23, 24). The ternary complexes of a rat DNA polymerase, a DNA template-primer, and dideoxycytidine triphosphate have been determined (22), which supports this fact. We thus reasoned that if a unique fluorescent dye is linked to the 5-position of the pyrimidines (T and C) and the 7-position of purines (G and A) via 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. Based on this rationale, we proposed an SBS approach using cleavable fluorescent nucleotide analogues as reversible terminators to sequence surface-immobilized DNA [supporting information (SI) Fig. 6] (25). In this approach, the nucleotides are modified at two specific locations so that they are still recognized by DNA polymerase as substrates: (i) a different fluorophore with a distinct fluorescent emission is linked to each of the four bases through a cleavable linker and (ii) the

Author contributions: J.J. and N.J.T. designed research; D.H.K., L.B., Q.M., X.B., Z.L., X.L., M.S.M., S.S., J.W., J.R.E., and A.R. performed research; and J.J., D.H.K., and Q.M. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

Abbreviations: SBS, sequencing by synthesis; bodipy, 4,4-difluoro-4-bora-3 α ,4 α -diazaindacene; ROX, 6-carboxy-X-rhodamine; R6G, 6-carboxyrhodamine-6G.

[†]To whom correspondence may be addressed at: Room 405A, Russ Berrie Medical Science Pavilion, Columbia Genome Center, Columbia University College of Physicians and Surgeons, New York, NY 10032. E-mail: ju@c2b2.columbia.edu or njt3@columbia.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0609513103/DC1.

© 2006 by The National Academy of Sciences of the USA

APPLIED BIOLOGICAL
SCIENCES

CHEMISTRY

3'-OH group is capped by a small chemically reversible moiety. DNA polymerase incorporates only a single nucleotide analogue complementary to the base on a DNA template covalently linked to a surface. After incorporation, the unique fluorescence emission is detected to identify the incorporated nucleotide and the fluorophore is subsequently removed. The 3'-OH group is then chemically regenerated, which allows the next cycle of the polymerase reaction to proceed. Because the large surface on a DNA chip can have a high density of different DNA templates spotted, each cycle can identify many bases in parallel, allowing the simultaneous sequencing of a large number of DNA molecules. We have previously established the feasibility of performing SBS on a chip using four photocleavable fluorescent nucleotide analogues (26) and discovered that an allyl group can be used as a cleavable linker to bridge a fluorophore to a nucleotide (27). We have also reported the design and synthesis of two photocleavable fluorescent nucleotides as reversible terminators for polymerase reaction (28, 29).

Our previous research efforts have firmly established the molecular level strategy to rationally modify the nucleotides by attaching a cleavable fluorescent dye to the base and capping the 3'-OH with a small chemically reversible moiety for SBS. This approach was recently adopted by Genomics Industry to potentially provide a platform for DNA sequencing (30). Here we report the design and synthesis of four chemically cleavable fluorescent nucleotide analogues as reversible terminators for SBS. Each of the nucleotide analogues contains a 3'-O-allyl group and a unique fluorophore with a distinct fluorescence emission at the base through a cleavable allyl linker. We first established that these nucleotide analogues are good substrates for DNA polymerase in a solution-phase DNA extension reaction and that the fluorophore and the 3'-O-allyl group can be removed with high efficiency in aqueous solution. We then performed SBS using these four chemically cleavable fluorescent nucleotide analogues as reversible terminators to identify ≈ 20 continuous bases of a DNA template immobilized on a chip. Accurate DNA sequences were obtained for DNA templates containing homopolymer sequences. The DNA template was immobilized on the surface of the chip that contains a PEG linker with 1,3-dipolar azide-alkyne cycloaddition chemistry. These results indicated that successful cleavable fluorescent nucleotide reversible terminators for four-color DNA sequencing by synthesis can be designed by attaching a cleavable fluorophore to the base and capping the 3'-OH with a small chemically reversible moiety so that they are still recognized by DNA polymerase as substrates. Further optimization of the approach will lead to even longer sequencing read lengths.

Results and Discussion

Design and Synthesis of Chemically Cleavable Fluorescent Nucleotide Analogues as Reversible Terminators for SBS. To demonstrate the feasibility of carrying out *de novo* DNA sequencing by synthesis on a chip, four chemically cleavable fluorescent nucleotide analogues (3'-O-allyl-dCTP-allyl-bodipy-FL-510, 3'-O-allyl-dUTP-allyl-R6G, 3'-O-allyl-dATP-allyl-ROX and 3'-O-allyl-dGTP-allyl-bodipy-650/Cy5) (R6G, 6-carboxyrhodamine-6G; ROX, 6-carboxy-X-rhodamine; bodipy, 4,4-difluoro-4-bora-3 α ,4 α -diza-s-indacene; Fig. 1) were designed and synthesized as reversible terminators for DNA polymerase reaction. 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 an allyl carbamate linker. However, due to the close proximity of the 3' position on the sugar ring of a nucleotide to the amino acid residues of the active site of the DNA polymerase, a relatively small allyl moiety was chosen as the 3'-OH reversible capping group. We have found that the fluorophore and the 3'-O-allyl group on a DNA extension product, which is generated by incorporation of the

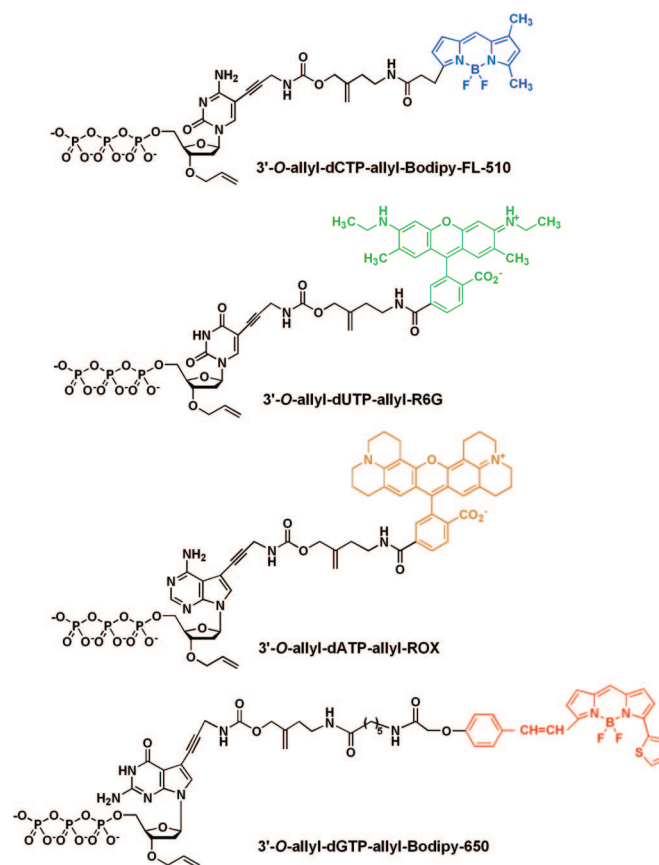


Fig. 1. Structures of 3'-O-allyl-dCTP-allyl-bodipy-FL-510 [$\lambda_{\text{abs(max)}} = 502$ nm; $\lambda_{\text{em(max)}} = 510$ nm], 3'-O-allyl-dUTP-allyl-R6G [$\lambda_{\text{abs(max)}} = 525$ nm; $\lambda_{\text{em(max)}} = 550$ nm], 3'-O-allyl-dATP-allyl-ROX [$\lambda_{\text{abs(max)}} = 585$ nm; $\lambda_{\text{em(max)}} = 602$ nm], and 3'-O-allyl-dGTP-allyl-bodipy-650 [$\lambda_{\text{abs(max)}} = 630$ nm; $\lambda_{\text{em(max)}} = 650$ nm].

chemically cleavable fluorescent nucleotide analogues, are removed simultaneously in 30 s by Pd-catalyzed deallylation in aqueous solution. This one-step dual-deallylation reaction thus allows the re-initiation of the polymerase reaction. The detailed synthesis procedure and characterization of the four nucleotide analogues in Fig. 1 are described in SI Text.

To verify that these fluorescent nucleotide analogues are incorporated accurately in a base-specific manner in a polymerase reaction, four continuous steps of DNA extension and deallylation were carried out in solution. This allows the isolation of the DNA product at each step for detailed molecular structure characterization by MALDI-TOF mass spectrometry (MS) as shown in Fig. 2. The first extension product 5'-U(allyl-R6G)-3'-O-allyl (1) was purified by HPLC and analyzed using MALDI-TOF MS (Fig. 2A). This product was then deallylated using a Pd-catalyzed deallylation mixture [$1 \times$ Thermopol I reaction buffer/ $\text{Na}_2\text{PdCl}_4/\text{P}(\text{PhSO}_3\text{Na})_3$]. The active Pd catalyst is generated from Na_2PdCl_4 and a ligand $\text{P}(\text{PhSO}_3\text{Na})_3$ (TPPTS) to mediate the deallylation reaction in DNA compatible aqueous condition to simultaneously cleave both the fluorophore and the 3'-O-allyl group (28). The deallylated product (2) was also analyzed by using MALDI-TOF MS (Fig. 2B). As can be seen in Fig. 2A, the MALDI-TOF MS spectrum consists of a distinct peak at m/z 6469 corresponding to the DNA extension product 5'-U(allyl-R6G)-3'-O-allyl (1), which confirms that the nucleotide analogue can be incorporated base specifically among pool of all four (A, C, G, T) by DNA polymerase into a growing DNA strand. Fig. 2B shows the deallylation result of the above DNA product. The peak at m/z 6469 has completely disappeared, whereas the peak corresponding to the dual deallylated product

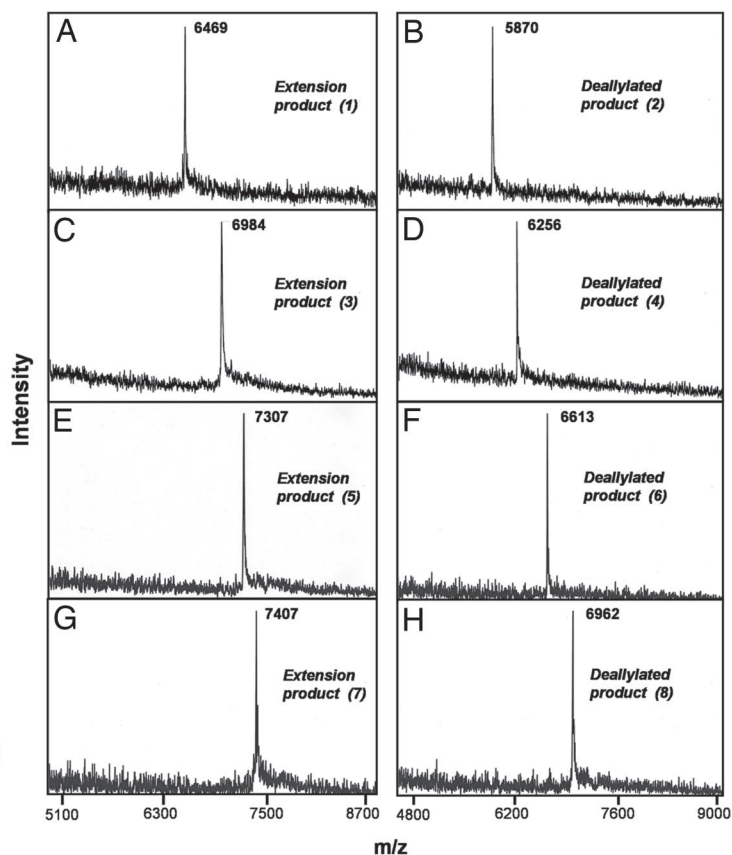
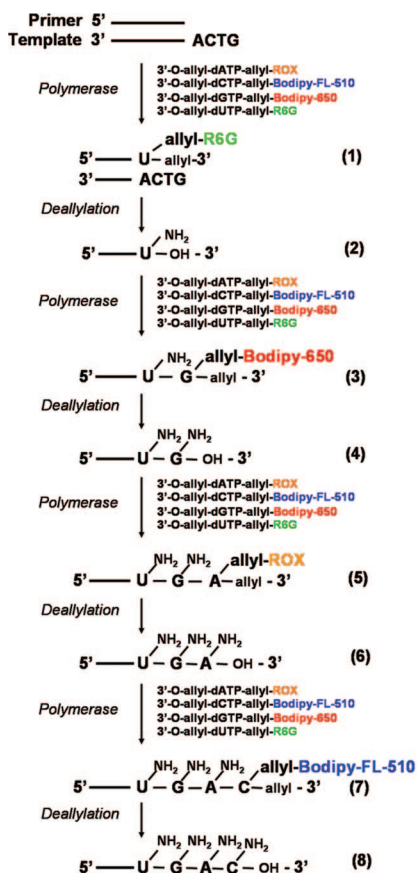


Fig. 2. The polymerase extension scheme (Left) and MALDI-TOF MS spectra of the four consecutive extension products and their deallylated products (Right). Primer extended with 3'-O-allyl-dUTP-allyl-R6G (1), and its deallylated product 2; Product 2 extended with 3'-O-allyl-dGTP-allyl-bodipy-650 (3), and its deallylated product 4; Product 4 extended with 3'-O-allyl-dATP-allyl-ROX (5), and its deallylated product 6; Product 6 extended with 3'-O-allyl-dCTP-allyl-bodipy-FL-510 (7), and its deallylated product (8). After 30 s of incubation with the palladium/TPPTS mixture at 70°C, deallylation is complete with both the fluorophores and the 3'-O-allyl groups cleaved from the extended DNA products.

5'-U (2) appears as the sole dominant peak at m/z 5870, which establishes that the Pd-catalyzed deallylation reaction completely cleaves both the fluorophore and the 3'-O-allyl group with high efficiency without damaging the DNA. The next extension reaction was carried out by using this deallylated DNA product with a free 3'-OH group regenerated as a primer along with four allyl modified fluorescent nucleotide mixture to yield an extension product 5'-UG(allyl-bodipy-650)-3'-O-allyl (3). As described above, the extension product 3 was analyzed by MALDI-TOF MS producing a dominant peak at m/z 6984 (Fig. 2C), and then deallylated for further MS analysis yielding a single peak at m/z 6256 (product 4) (Fig. 2D). The third extension reaction yielding 5'-UGA(allyl-ROX)-3'-O-allyl (5), the fourth extension reaction yielding 5'-UGAC(allyl-bodipy-FL-510)-3'-O-allyl (7), and their deallylation reactions to yield products 6 and 8 were similarly carried out and analyzed by MALDI-TOF MS as shown in Fig. 2 E-H. The chemical structures of the extension and cleavage products for each step are shown in SI Fig. 7. These results demonstrate that the above-synthesized four chemically cleavable fluorescent nucleotide analogues are successfully incorporated with high fidelity into the growing DNA strand in a polymerase reaction, and furthermore, both the fluorophore and the 3'-O-allyl group are efficiently removed by using a Pd-catalyzed deallylation reaction, which makes it feasible to use them for SBS on a chip.

Four-Color DNA Sequencing with Chemically Cleavable Fluorescent Nucleotide Analogues as Reversible Terminators on a DNA Chip. The chemically cleavable fluorescent nucleotide analogues were then

used in an SBS reaction to identify the sequence of the DNA template immobilized on a solid surface. A site-specific 1,3-dipolar cycloaddition coupling chemistry was used to covalently immobilize the alkyne-labeled self-priming DNA template on the azido-functionalized surface in the presence of a Cu(I) catalyst. The principal advantage offered by the use of a self-priming moiety as compared with using separate primers and templates is that the covalent linkage of the primer to the template in the self-priming moiety prevents any possible dissociation of the primer from the template during the process of SBS. To prevent nonspecific adsorption of the unincorporated fluorescent nucleotides on the surface of the chip, a PEG linker is introduced between the DNA templates and the chip surface (SI Fig. 8). This approach was shown to produce very low background fluorescence after cleavage to remove the fluorophore as demonstrated by the DNA sequencing data described below.

We first performed SBS on a chip-immobilized DNA template that has no homopolymer sequences using the four chemically cleavable fluorescent nucleotide reversible terminators (3'-O-allyl-dCTP-allyl-bodipy-FL-510, 3'-O-allyl-dUTP-allyl-R6G, 3'-O-allyl-dATP-allyl-ROX and 3'-O-allyl-dGTP-allyl-Cy5), and the results are shown in Fig. 3. The structure of the self-priming DNA moiety is shown schematically in Fig. 3A, with the first 13 nucleotide sequences immediately after the priming site. The *de novo* sequencing reaction on the chip was initiated by extending the self-priming DNA using a solution containing all four 3'-O-allyl-dNTPs-allyl-fluorophore, and a 9^N mutant DNA polymerase. To negate any lagging fluorescence signal that is caused by previously unextended

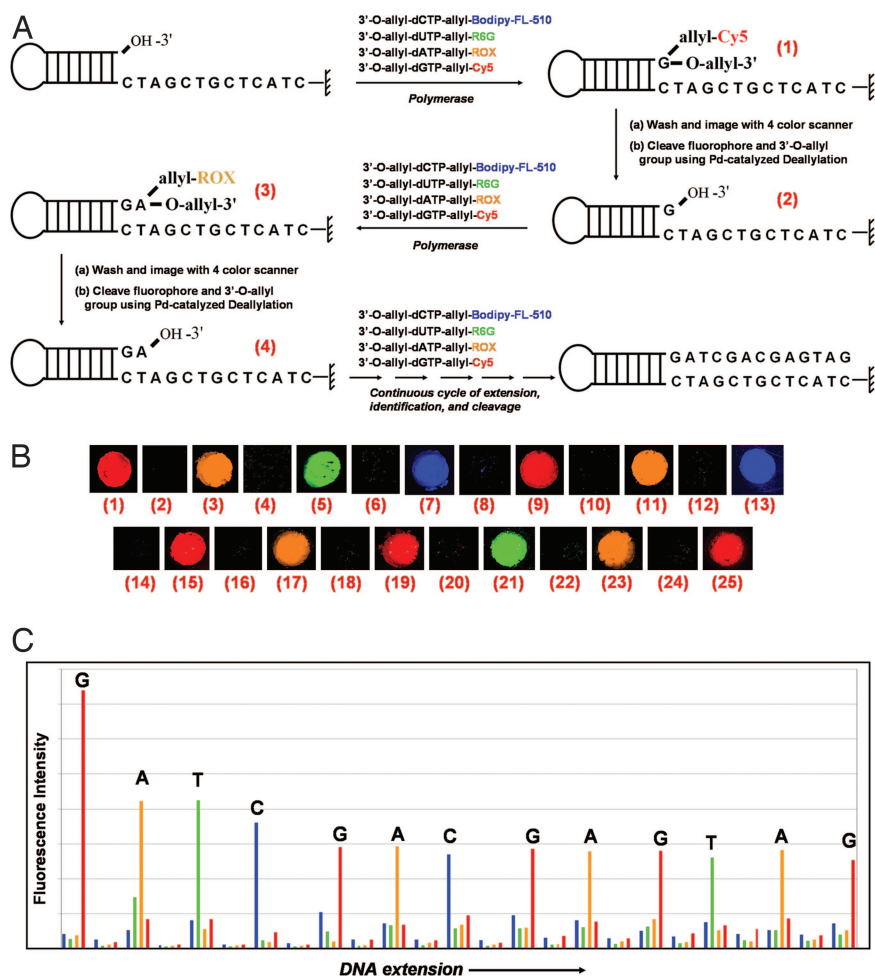


Fig. 3. Four-color sequencing by synthesis data on a DNA chip. (A) Reaction scheme of SBS on a chip using four chemically cleavable fluorescent nucleotides. (B) The scanned four-color fluorescence images for each step of SBS on a chip: (1) incorporation of 3'-O-allyl-dGTP-allyl-Cy5; (2) cleavage of allyl-Cy5 and 3'-allyl group; (3) incorporation of 3'-O-allyl-dATP-allyl-ROX; (4) cleavage of allyl-ROX and 3'-allyl group; (5) incorporation of 3'-O-allyl-dUTP-allyl-R6G; (6) cleavage of allyl-R6G and 3'-allyl group; (7) incorporation of 3'-O-allyl-dCTP-allyl-bodipy-FL-510; (8) cleavage of allyl-bodipy-FL-510 and 3'-allyl group; images 9–25 are similarly produced. (C) A plot (four-color sequencing data) of raw fluorescence emission intensity at the four designated emission wavelength of the four chemically cleavable fluorescent nucleotides vs. the progress of sequencing extension.

priming strand, a synchronization step was added to reduce the amount of unextended priming strands after the extension with the fluorescent nucleotides. A synchronization reaction mixture consisting of all four 3'-O-allyl-dNTPs (Fig. 4), which have a higher polymerase incorporation efficiency due to the lack of a fluorophore compared with the bulkier 3'-O-allyl-dNTPs-allyl-fluorophore, was used along with the 9[°]N mutant DNA polymerase to extend any remaining priming strand that has a free 3'-OH group to synchronize the incorporation. The extension by 3'-O-allyl-dNTPs also enhances the enzymatic incorporation of the next nucleotide analogue, because after cleavage to remove the 3'-O-allyl group, the DNA product extended by 3'-O-allyl-dNTPs carry no modification groups. After washing, the extension of the primer by only the complementary fluorescent nucleotide was confirmed by observing a red signal (the emission from Cy5) in a four-color fluorescent scanner (Fig. 3B1). After detection of the fluorescent signal, the chip surface was immersed in a deallylation mixture [1× Thermopol I reaction buffer/Na₂PdCl₄/P(PhSO₃Na)₃] and incubated for 5 min at 60°C to cleave both the fluorophore and 3'-O-allyl group simultaneously. The chip was then immediately immersed in a 3 M Tris-HCl buffer (pH 8.5) and incubated for 5 min at 60°C to remove the Pd complex. The surface was then washed, and a negligible residual fluorescent signal was detected to confirm

cleavage of the fluorophore. This was followed by another extension reaction using 3'-O-allyl-dNTPs-allyl-fluorophore mixture to incorporate the next fluorescent nucleotide complementary to the

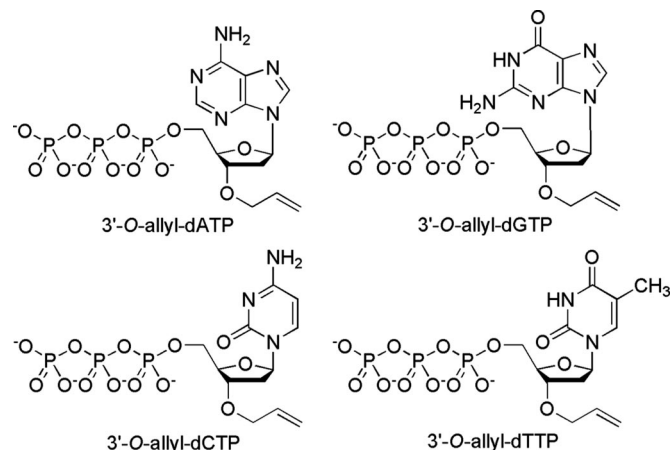


Fig. 4. Structures of 3'-O-allyl-dATP, 3'-O-allyl-dCTP, 3'-O-allyl-dGTP, and 3'-O-allyl-dTTP.

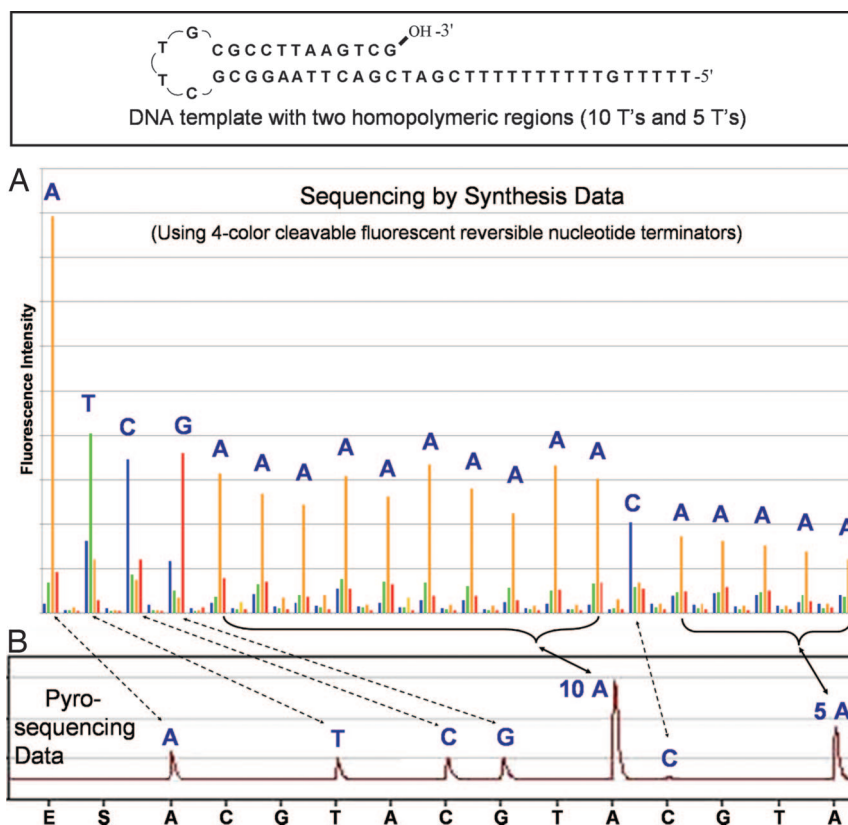


Fig. 5. Comparison of four-color sequencing by synthesis and pyrosequencing data. (A) Four-color DNA sequencing raw data with our sequencing by synthesis chemistry using a template containing two homopolymeric regions. The individual base (A, T, C, G), the 10 repeated A's, and the five repeated A's are clearly identified. The small groups of peaks between the identified bases are fluorescent background from the DNA chip, which does not build up as the cycle continues. (B) The pyrosequencing data of the same DNA template containing the homopolymeric regions (10 T's and five T's). The first four individual bases are clearly identified. The two homopolymeric regions (10 A's) and (five A's) produce two large peaks, which are very difficult to identify the exact sequence from the data.

subsequent base on the template. The entire process of incorporation, synchronization, detection, and cleavage was performed multiple times using the four chemically cleavable fluorescent nucleotide reversible terminators to identify 13 successive bases in the DNA template. The fluorescence image of the chip for each nucleotide addition is shown in Fig. 3B, and a plot of the fluorescence intensity vs. the progress of sequencing extension (raw four-color sequencing data) is shown in Fig. 3C. The DNA sequences are unambiguously identified from the four-color raw fluorescence data without any processing.

Comparison of Four-Color SBS with Pyrosequencing. To further verify the advantage of SBS method using the four chemically cleavable fluorescent nucleotide reversible terminators, we carried out similar sequencing reaction as described above on a DNA template which contained two separate homopolymeric regions (stretch of 10 T's and five T's) as shown in Fig. 5A. These sequencing raw data were produced by adding all four fluorescent nucleotide reversible terminators together to the DNA template immobilized on the chip followed by synchronization reaction with four 3'-O-allyl-dNTPs, detecting the unique fluorescence emission for sequence determination, then cleaving the fluorophore and the 3'-O-allyl group in one step to continue the sequencing cycles. All 20 bases including the individual base (A, T, C, G), the 10 repeated A's and the five repeated A's are clearly identified. The small groups of peaks between the identified bases are fluorescent background from the DNA chip, which does not build up as the cycle continues. Fig. 5B shows the pyrosequencing data of the same DNA template containing the homopolymeric sequences. The first four individual bases are clearly identified. The two homopolymeric regions (10

A's) and (five A's) produce two large peaks, but it is very difficult to identify the exact sequence from the data.

Conclusion

We have synthesized and characterized four chemically cleavable fluorescent nucleotide analogues and used them to produce four-color *de novo* DNA sequencing data on a chip. In doing so, we have achieved two critical requirements for using SBS method to sequence DNA unambiguously. First, a strategy to use a chemically reversible moiety to cap the 3'-OH group of the nucleotide has been successfully implemented so that the nucleotide terminates the polymerase reaction to allow the identification of the incorporated nucleotide. In addition, these reversible terminators allow for the addition of all four nucleotides simultaneously in performing SBS. This ultimately reduces the number of cycles needed to complete the sequencing cycle, increases sequencing accuracy due to competition of the four nucleotides in the polymerase reaction, and enables accurate determination of homopolymeric regions of DNA template. Second, efficient removal of both the fluorophore and the 3'-OH capping group after the fluorescence signal detection have successfully been carried out, which increases the overall efficiency of SBS.

The key factor governing the sequencing read length of our four-color SBS approach is the step-wise yield that is determined by the nucleotide incorporation efficiency and the yield of the cleavage of the fluorophore and the 3'-OH capping group from the DNA extension products. This stepwise yield is $\approx 99\%$ based on measurement of the DNA products in solution phase. The yield on the surface is difficult to measure precisely due to fluctuations in the fluorescence imaging using the current manual fluorescent scanner

Explore Litigation Insights

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

Real-Time Litigation Alerts



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

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

Advanced Docket Research



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

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

Analytics At Your Fingertips



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

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

API

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

LAW FIRMS

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

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

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.