RSC Advances

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/advances

DNA sequencing by synthesis using 3′-*O*-azidomethyl nucleotide reversible terminators and surface-enhanced Raman spectroscopic detection

Mirkó Palla,[a,b] Wenjing Guo,[a,c] Shundi Shi,[a,c] Zengmin Li,[a,c] Jian Wu,[a] Steffen Jockusch,[e] Cheng Guo,[a,c] James J. Russo,[a,c] Nicholas J. Turro,[e] Jingyue Ju[a,c,d,*]

[a]Center for Genome Technology and Biomolecular Engineering, Departments of [b]Mechanical Engineering, [c]Chemical Engineering, [d]Pharmacology and [e]Chemistry, Columbia University, 3000 Broadway, New York, NY 10027, USA.

[*]To whom correspondence should be addressed:

      e-mail: dj222@columbia.edu

      Tel.: 212-851-9191

Abstract: As an alternative to fluorescence-based DNA sequencing by synthesis (SBS), we report here an approach using an azido moiety (N_3) that has an intense, narrow and unique Raman shift at 2125 cm^{-1}, where virtually all biological molecules are transparent, as a label for SBS. We first demonstrated that the four 3'-*O*-azidomethyl nucleotide reversible terminators (3'-*O*-azidomethyl-dNTPs) displayed surface enhanced Raman scattering (SERS) at 2125 cm^{-1}. Using these 4 nucleotide analogues as substrates, we then performed a complete 4-step SBS reaction. We used SERS to monitor the appearance of the azide-specific Raman peak at 2125 cm^{-1} as a result of polymerase extension by a single 3'-*O*-azidomethyl-dNTP into the growing DNA strand and disappearance of this Raman peak with cleavage of the azido label to permit the next nucleotide incorporation, thereby continuously determining the DNA sequence. Due to the small size of the azido label, the 3'-*O*-azidomethyl-dNTPs are efficient substrates for the DNA polymerase. In the SBS cycles, the natural nucleotides are restored after each incorporation and cleavage, producing a growing DNA strand that bears no modifications and will not impede further polymerase reactions. Thus, with further improvements in SERS for the azido moiety, this approach has the potential to provide an attractive alternative to fluorescence-based SBS.

Introduction

DNA sequencing is a fundamental tool in biological and medical research, and is especially important for the paradigm of personalized medicine. Various new DNA sequencing methods have been investigated with the aim of eventually realizing the goal to sequence the full genome of an individual human at a cost of only $1,000. Currently, the dominant method is sequencing by synthesis (SBS),[1-14] an approach that determines DNA sequences during the polymerase reaction. The most widely used high-throughput SBS technology[11] uses cleavable fluorescent nucleotide reversible terminator (NRT) sequencing chemistry that we developed previously.[3,8] These cleavable fluorescent NRTs were designed based on the following rationale: each of the four nucleotides (A, C, G, T) is modified by attaching a unique cleavable fluorophore to the base and capping the 3'-OH group with a small reversible chemical moiety so that they are still recognized by DNA polymerase as substrates. Thus the cleavable fluorescent NRTs involve two site modifications:[3,8,11] a fluorescent dye to serve as a reporter group on the base and a small chemical moiety to cap the 3'-OH group to temporarily terminate the polymerase reaction after nucleotide incorporation for sequence determination. After incorporation and signal detection, the fluorophore is cleaved and the 3'-OH capping moiety removed to resume the polymerase reaction in the next cycle. These cleavable fluorescent NRTs have proved to be good substrates for reengineered polymerases and have been used extensively in next generation DNA sequencing systems.[8,11] Moreover, they enable accurate determination of homopolymer sequences, since only one base is identified in each cycle.

Fluorescence-based methods have many advantages in terms of detection sensitivity. However, because of the large size of the fluorophores and their broad emission spectra, specific polymerase and reaction conditions need to be optimized for sequencing reactions. In addition, the current cleavable fluorescent NRTs used in SBS leave a modified group on the base of the growing DNA strand after cleavage of the fluorophore, limiting sequencing readlength. As an alternative to fluorescence-based DNA SBS, we report here an approach of using an azido moiety (N_3) that has an intense, narrow and unique Raman shift at 2125 cm^{-1}, where virtually all biological molecules are transparent, as a label for SBS. Raman spectroscopy, a technique based on the inelastic scattering of light from molecules, is an inherently less sensitive technique than fluorescence

detection. However, recent advances in surface-enhanced Raman scattering (SERS) have increased the sensitivity to make even single molecule detection possible.[15] Enhancement factors of 10^9 have been reported using nanostructured surfaces, such as arrays of regularly ordered vertical posts coated with Au.[16,17] Recently, the unique Raman peaks (2000-2800 cm^{-1}) of the azido, alkynyl and cyano groups have been explored as Raman tags for cellular imaging.[18,19] Detection of SERS from nucleotide bases has been proposed to determine DNA sequences[20,21] but no successful examples have been published.

We first demonstrated that the four 3'-*O*-azidomethyl NRTs (3'-*O*-azidomethyl-dNTPs)[10] (Fig. 1) displayed SERS at 2125 cm^{-1}. Using these 4 nucleotide analogues as substrates, we then performed a complete 4-step SBS reaction. We used SERS to monitor the appearance of the azide-specific Raman peak at 2125 cm^{-1} as a result of polymerase extension by a single 3'-*O*-azidomethyl-dNTP into the growing DNA strand and disappearance of this Raman peak with cleavage of the azido label to permit the next nucleotide incorporation, thereby continuously determining the DNA sequence. Due to the small size of the azido label, the 3'-*O*-azidomethyl-dNTPs are efficient substrates for the DNA polymerase. In the SBS cycles, the natural nucleotides are restored after each incorporation and cleavage, producing a growing DNA strand that bears no modifications and will not impede further polymerase reactions.

We demonstrate successful SBS using the four 3'-*O*-azidomethyl-dNTPs and SERS on a commercially available Klarite gold-coated surface, with a synthetic template and primer. In our SERS-SBS approach, the 3'-OH group of the nucleotides is capped with an azidomethyl moiety,[10] as shown in Fig. 1, to temporarily terminate the polymerase reaction after incorporation into the growing DNA strand. Instead of attaching a fluorescent dye as a reporter group on the base of the nucleotide, here the azidomethyl group performs a dual function, serving as a reversible 3'-OH capping moiety and as the SERS reporter group to indicate the incorporation of a nucleotide into a growing DNA strand for sequence determination. The azido group is ideally suited as a Raman reporter group, because of its distinct Raman shift at ~2125 cm^{-1}, a spectral region where DNA, proteins and other biological molecules do not elicit signals.

There are several advantages to using an N_3 group as a label for SBS. Because these 3'-*O*-azidomethyl-dNTPs (3'-*O*-N_3-dNTPs) do not require the attachment of fluorescent tags, their synthesis is simpler and more cost effective. They are much smaller than their counterparts with fluorescent tags, which increase their incorporation efficiency by DNA polymerase. In addition, the extended DNA strand is identical to natural DNA, unlike many current SBS approaches, which require the use of modified nucleotides that leave short remnants of the linkers after cleavage of the fluorescent tags;[8,11,12] as these remnants build up in the extended DNA chains, they are increasingly likely to alter DNA structure and impede further nucleotide incorporation by polymerase. After cleavage by tris(2-carboxyethyl)phosphine (TCEP), the N_3 is completely destroyed yielding no Raman signal at ~2125 cm^{-1}. This method to remove the azidomethyl moiety for regenerating the 3'-OH group to allow the next cycle of nucleotide incorporation and detection is well established and has been shown to be highly compatible with DNA stability.[10]

Results and discussion

Before performing SBS with the four 3'-*O*-N_3-dNTPs, we confirmed that they generated distinct Raman signals relative to the natural nucleotides. As shown in Fig. 2, the four 3'-*O*-N_3-dNTPs displayed enhanced Raman scattering at ~2125 cm^{-1} on Klarite SERS substrates, while the natural dNTPs produced only a background signal. A typical SERS enhancement factor of ~8 ×10^5 was observed on Klarite substrates using a small molecule [2-azido-3-benzyloxy)propanoic acid] that contains an azido group (see ESI†).

We next performed a complete consecutive 4-step SBS reaction that involved incorporation of each complementary 3'-*O*-N_3-dNTP, detection of the unique N_3 Raman peak at 2125 cm^{-1} for sequence determination, and cleavage of the 3'-*O*-azidomethyl blocking group from the DNA extension product to yield a free 3'-OH group for incorporating the next nucleotide analogue. A template-primer combination was designed in which the next four nucleotides to be added were A, C, G and T. As shown in Fig. 3 (middle), the SBS reaction was initiated with the 13-mer primer annealed to a DNA template. When the first complementary nucleotide, 3'-*O*-N_3-dATP, was used in the polymerase reaction (Fig. 3, middle), it was incorporated into the primer to form a DNA extension product that yielded a unique N_3 Raman peak at 2125 cm^{-1} (Fig. 3a, left). The

expected DNA extension product generated by incorporating a 3'-O-N$_3$-dATP with a molecular weight of 4329 Da was confirmed by MALDI-TOF MS as a single peak in the mass spectrum (Fig. 3a, right). These results indicated that the 3'-O-N$_3$-dATP was quantitatively incorporated into the 13-mer DNA primer. After treatment with TCEP to remove the azidomethyl group from the DNA product and HPLC purification, the Raman peak at 2125 cm^{-1} completely disappeared (Fig. 3b, left), and cleavage was confirmed by a single MS peak at 4274 Da, corresponding to the DNA product with the 3'-O-azidomethyl group removed (Fig. 3b, right). The newly formed DNA extension product with a free 3'-OH group was then used in a second polymerase reaction to incorporate a 3'-O-N$_3$-dCTP. The Raman spectrum again revealed a peak at 2125 cm^{-1} (Fig. 3c, left) and MS gave a single peak at 4621 Da (Fig. 3c, right), indicating incorporation of a 3'-O-N$_3$-dCTP into the growing DNA strand in this cycle. After TCEP treatment, the Raman peak at 2125 cm^{-1} disappeared (Fig. 3d, left), and a single MS peak of the cleavage DNA product appeared at 4566 Da (Fig. 3d, right), which demonstrated the complete removal of the azidomethyl group from the DNA extension product.

Fig. 3e (left) shows the third incorporation of 3'-O-N$_3$-dGTP into this growing DNA strand with reappearance of the N$_3$-dependent Raman signal, and the disappearance of this signal after TCEP cleavage is revealed in Fig. 3f (left). Accurate masses of the corresponding DNA products were obtained by MALDI-TOF MS for the third nucleotide incorporation (4953 Da, Fig. 3e, right) and cleavage reaction (4898 Da, Fig. 3f, right).

Finally, 3'-O-N$_3$-dTTP incorporation in the fourth cycle was demonstrated by the re-emergence of the N$_3$ Raman signal at 2125 cm^{-1} (Fig. 3g, left) and a final removal of the azidomethyl group by TCEP was verified by loss of this peak (Fig. 3h, left). Again, appropriate masses for the corresponding DNA products were obtained by MALDI-TOF MS for the fourth nucleotide incorporation (5260 Da, Fig. 3g, right) and cleavage reaction (5205 Da, Fig. 3h, right).

In summary, with the above experimental results we have demonstrated DNA sequencing by synthesis using 3'-O-azidomethyl-dNTPs that contain an azido moiety (N$_3$) as a label that has an intense, narrow and unique SERS peak at 2125 cm^{-1}, where virtually all biological molecules are transparent (Fig. 1, Fig. 2, and Fig. S2 in ESI†). The azidomethyl group capping the 3'-OH of the nucleotides temporarily terminates the

polymerase reaction after they are incorporated into the growing DNA strand for sequence determination. Treatment with TCEP efficiently removes the azidomethyl group from the DNA extension product, yielding a free 3'-OH group for continuous sequencing. Meanwhile, the strong and distinct azido peak in Raman spectra makes it an excellent reporter group for sequence determination during SBS. The characteristic Raman shift of the DNA extension product produced by incorporation of the 3'-*O*-azidomethyl-dNTPs is clearly distinguishable from those of the cleavage products. Once the complementary nucleotide was incorporated into the growing DNA strand in the polymerase reaction, a Raman peak appeared at 2125 cm^{-1}, otherwise, this portion of the spectrum was at background level. This method of DNA sequencing is realized with only minor nucleotide modifications. The nucleobases are intact and the DNA strand remains in its natural form as it is elongated during SBS, therefore a long readlength would be expected.

Integration of nanoplasmonic systems together with site specific molecular interactions to yield reproducible and optimal SERS signal enhancement can be achieved in the future by attaching either the DNA or the polymerase to the SERS surface, permitting washing away of unreacted nucleotides and other reactants for SERS-SBS. Furthermore, with 4 different SERS labels bearing unique spectral signatures on each of the four nucleotides, a 4-color SERS-SBS platform can be developed in the future. The concept elaborated in this initial exploration of the sensitivity of SERS in detecting the azido group on the DNA nucleotide for SBS can be developed further with the use of improved Raman instrumentation tailored for DNA sequencing, coupled with optimized SERS surfaces.

Experimental

Gold-coated Klarite SERS-active substrates were purchased from Renishaw Diagnostics. The 6 mm × 10 mm chip (consisting of a 4 mm x 4 mm patterned region and an unpatterned Au reference area) was adhered to a standard microscope slide (25 mm × 75 mm) at the foundry. The active area contains an array of micro-scaled inverted pyramids with 1.5 μm well diameter, 2 μm pitch and 1 μm depth coated with a 20 nm chrome adhesion layer below a 400 μm gold layer.

Instrumentation. All Raman/SERS spectra were recorded using a Jobin-Yvon LabRam ARAMIS Raman microscope (Horiba, Japan) in a standard backscattering configuration with a 785 nm excitation laser. The laser beam was focused onto the sample using a 50× long working-distance (NA = 0.5) dry objective (Nikon, Japan). All spectra in this work were obtained with an exposure time of 10 sec, 5 accumulations per spot and at 34 mW laser power before the objective.

Analysis of Raman spectral data. Due to potential non-uniformity of analyte deposition, a data set of 10 spectra acquired at randomly selected regions on the same substrate was obtained. Data were presented as background removed averages of such a data set. Spectra were processed using the Savitzky-Golay fourth derivative method (window size of 25 data points), which can effectively reduce or eliminate possible false correlations resulting from a constant offset or broadband background.[22]

SERS detection of nucleotides. The four 3'-*O*-azidomethyl nucleotide reversible terminators (Fig. 1) were synthesized according to the procedure described previously.[10] Each nucleotide analogue was dissolved in water to form a 100 µM solution, then a 2 µL aliquot (200 pmol) was deposited onto individual Klarite SERS substrates (Renishaw Diagnostics, UK) and dried in ambient air to obtain a uniform molecular deposition for Raman measurements. Equivalent amounts of the natural deoxynucleotides were deposited on 4 separate SERS substrates in the same way.

Sequencing by synthesis reactions using 3'-*O*-azidomethyl nucleotide reversible terminators. Polymerase extension reactions consisted of 20 pmol of a synthetic 51-mer DNA template (5'-GAGGCCAAGTACGGCGGGTACGTCCTTGACAATGTGT-ACATCAACATCACC-3'), 60 pmol of primer (5'-CACATTGTCAAGG-3') or a previously extended and TCEP-cleaved DNA product, 100 pmol of a single 3'-*O*-azidomethyl nucleotide reversible terminator (3'-*O*-N$_3$-dATP, 3'-*O*-N$_3$-dCTP, 3'-*O*-N$_3$-dGTP, or 3'-*O*-N$_3$-dTTP),[10] 1× ThermoPol reaction buffer (New England Biolabs, MA), 2 unit Therminator^{TM} III DNA polymerase and deionized H$_2$O in a total volume of 20 μL. Reactions were conducted in a thermal cycler (MJ Research, MA). After initial incubation at 94 °C for 20 sec, the reaction was performed for 36 cycles at 80 °C for 20 sec, 45 °C for 40 sec and 65 °C for 90 sec.

After the extension reaction, a small aliquot of the DNA product was desalted using a C18 ZipTip column (Millipore, MA) and analyzed by MALDI-TOF MS (ABI Voyager, DE). The remaining DNA product was concentrated under vacuum and purified by reverse phase HPLC on an XTerra MS C18, 2.5 μm 4.6 mm x 50 mm column (Waters, MA) to obtain the pure DNA extension product. Mobile phase: A, 8.6 mM triethylamine/100 mM 1,1,1,3,3,3-hexafluoro-2-propanol in water (pH 8.1); B, methanol. Elution was performed at 40 °C with a 0.5 mL/min flow rate, and from 88% A / 12% B to 66% A / 34% B linear gradient for 90 min. The purified DNA product was used for Raman measurements and in the subsequent cleavage reaction.

Cleavage reactions to remove the 3'-O-azidomethyl group from the DNA extension products with TCEP to regenerate the 3'-OH group were carried out by dissolving 100 pmol extension products in 10 μL of 100 mM TCEP solution (pH 9.0), and incubating at 65 °C for 25 min. Following dilution in 1 mL deionized H_2O and purification in an Amicon Ultra-0.5 centrifugal filter unit with Ultracel-3 membrane (Millipore) to remove the TCEP, small aliquots of the resulting solution were used to obtain the MALDI-TOF mass spectrum after ZipTip cleanup and the Raman spectrum. The remaining DNA cleavage product was used as primer in the subsequent polymerase extension reaction. The mechanism of the TCEP cleavage reaction to remove the azidomethyl group from the DNA product is shown in Scheme S1 (see ESI†). The second through fourth extension/cleave cycles were carried out in a similar manner by using the previously extended and cleaved product as the primer. Four consecutive nucleotide additions are shown in Scheme S2 (see ESI†).

SERS detection of SBS products. Raman spectra of the newly purified DNA extension and cleavage products were acquired with a drop coating method on the SERS surface, in which an aliquot was deposited and dried in ambient air to obtain a uniform layer. Specifically, a 3 μL aliquot of the 100 μM DNA extension product was deposited onto a Klarite SERS-active substrate (Renishaw Diagnostics); similarly, 30 pmol of cleavage products were spotted for SERS measurement.

RSC Advances Accepted Manuscript

†Electronic supplementary information (ESI) available: schemes for mechanisms to cleave the 3'-azidomethyl group from the DNA product and polymerase extension reaction with the 3'-N_3-dNTPs; details of the SERS enhancement factor measurement and calculation; and extended Raman spectra of a typical SBS cycle. See DOI: 10.1039/b000000x/

References:

(1) E. D. Hyman, *Anal. Biochem.,* 1988, **174**, 423-436.

(2) M. Ronaghi, M. Uhlén and P. Nyrén, *Science,* 1998, **281**, 363-365.

(3) J. Ju, Z. Li, J. R. Edwards and Y. Itagaki, *US Pat.*, 6 664 079, 2003.

(4) Z. Li, X. Bai, H. Ruparel, S. Kim, N. J. Turro and J. Ju, J. *Proc. Natl. Acad. Sci. USA*, 2003, **100**, 414-419.

(5) I. Braslavsky, B. Hebert, E. Kartalov, and S. R. Quake, *Proc. Natl. Acad. Sci. USA,* 2003, **100**, 3960-3964.

(6) H. Ruparel, L. Bi, Z. Li, X. Bai, D. H. Kim, N. J. Turro and J. Ju, *Proc. Natl. Acad. Sci. USA,* 2005, **102**, 5932-5937.

(7) M, Margulies, M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y.-J. Chen, *et al.*, *Nature,* 2005, **437**, 376-380.

(8) J. Ju, D. H. Kim, L. Bi, Q. Meng, X. Bai, Z. Li, X. Li, M. S. Marma, S. Shi, J. Wu, J. R. Edwards, A. Romu and N. J. Turro, *Proc. Natl. Acad. Sci. USA,* 2006, **103**, 19635-19640.

(9) J. Wu, S. Zhang, Q. Meng, H. Cao, Z. Li, X. Li, S. Shi, D. H. Kim, L. Bi, N. J. Turro and J. Ju, *Proc. Natl. Acad. Sci., USA* 2007, **104**, 16462-16467.

(10) J. Guo, N. Xu, Z. Li, S. Zhang, J. Wu, D. H. Kim, S. M. Marma, Q. Meng, H. Cao, X. Li, S. Shi, L. Yu, S. Kalachikov, J. J. Russo, N. J. Turro and J. Ju, *Proc. Natl. Acad. Sci. USA,* 2008, **105**, 9145-9150.

(11) D. R. Bentley, S. Balasubramanian, H. P. Swerdlow, G. P. Smith, J. Milton, C. G. Brown, K. P. Hall, D. J. Evers, C. L. Barnes, H. R. Bignell, *et al.*, *Nature,* 2008, 456, 53-

59.

(12) T. D. Harris, P. R. Buzby, H. Babcock, E. Beer, J. Bowers, I. Braslavsky, M. Causey, J. Colonell, J. Dimeo, J. W. Efcavitch, *et al.*, *Science,* 2008, **320**, 106-109.

(13) J. Eid, A. Fehr, J. Gray, K. Luong, J. Lyle, G. Otto, P. Peluso, D. Rank, P. Baybayan, B. Bettman, *et al.*, *Science,* 2009, **323**, 133-138.

(14) J. M. Rothberg, W. Hinz, T. M. Rearick, J. Schultz, W. Mileski, M. Davey, J. H. Leamon, K. Johnson, M. J. Milgrew, M. Edwards, *et al.*, *Nature,* 2011, **475**, 348-352.

(15) S. L. Kleinman, E. Ringe, N. Valley, K. L. Wustholz, E. Phillips, K. A. Scheidt, G. C. Schatz and R. P. Van Duyne, *J. Am. Chem. Soc.,* 2011, **133**, 4115-4122.

(16) W. D. Li, F. Ding, J. Hu and S. Y. Chou, *Opt. Express,* 2011, **19**, 3925-3936.

(17) M. S. Schmidt, J. Hübner and A. Boisen, *Adv Mater* 2012, **24**, OP11-OP18.

(18) H. Yamakoshi, K. Dodo, A. Palonpon, J. Ando, K. Fujita, S. Kawata and M. Sodeoka, *J. Am. Chem. Soc.,* 2012, **134**, 20681–20689.

(19) L. Lin, X. Tian, S. Hong, P. Dai, Q. You, R. Wang, L. Feng, C. Xie, Z. Q. Tian and X. Chen, *Angew. Chem. Int. Ed.,* 2013, **52**, 7266–7271.

(20) K. Kneipp, H. Kneipp, V. B. Kartha, R. Manoharan, G. Deinum, I. Itzkan, R. R. Dasari and M. S. Feld, *Phys. Rev.,* 1998, **57**, R6281–R6284.

(21) T. Vo-Dinh, D. L. Stokes, G. D. Griffin, M. Volkan, U. J. Kim and M. I. Simon, *J. Raman Spectrosc.,* 1999, **30**, 785–793.

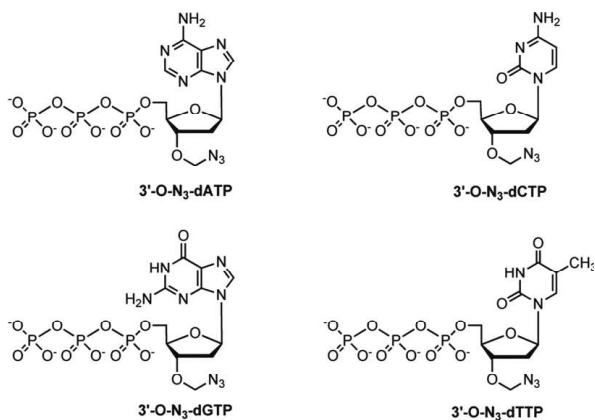(22) D. Zhang and D. Ben-Amotz, *Appl. Spectrosc.,* 2000, **54**, 1379-1383.

Fig. 1 Structures of the four 3'-*O*-azidomethyl nucleotide reversible terminators: 3'-*O*-N$_3$-dATP, 3'-*O*-N$_3$-dCTP, 3'-*O*-N$_3$-dGTP, and 3'-*O*-N$_3$-dTTP.
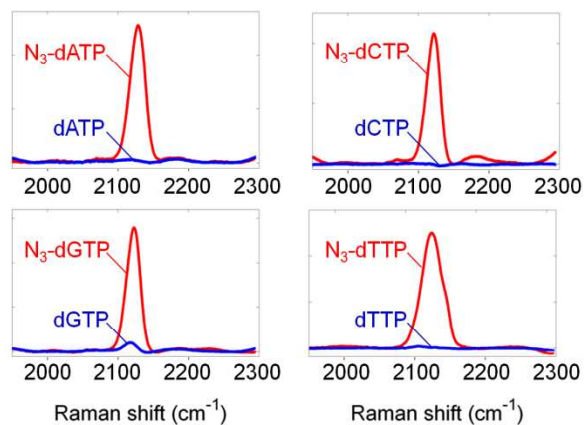

Fig. 2 Raman signal of 3'-*O*-azidomethyl nucleotide reversible terminators (N$_3$-dNTPs) (red) and natural dNTPs (blue). In all 4 cases, there is a 10^2-fold signal increase at the expected Raman shift of ~2125 cm^{-1} due to the N$_3$ group.
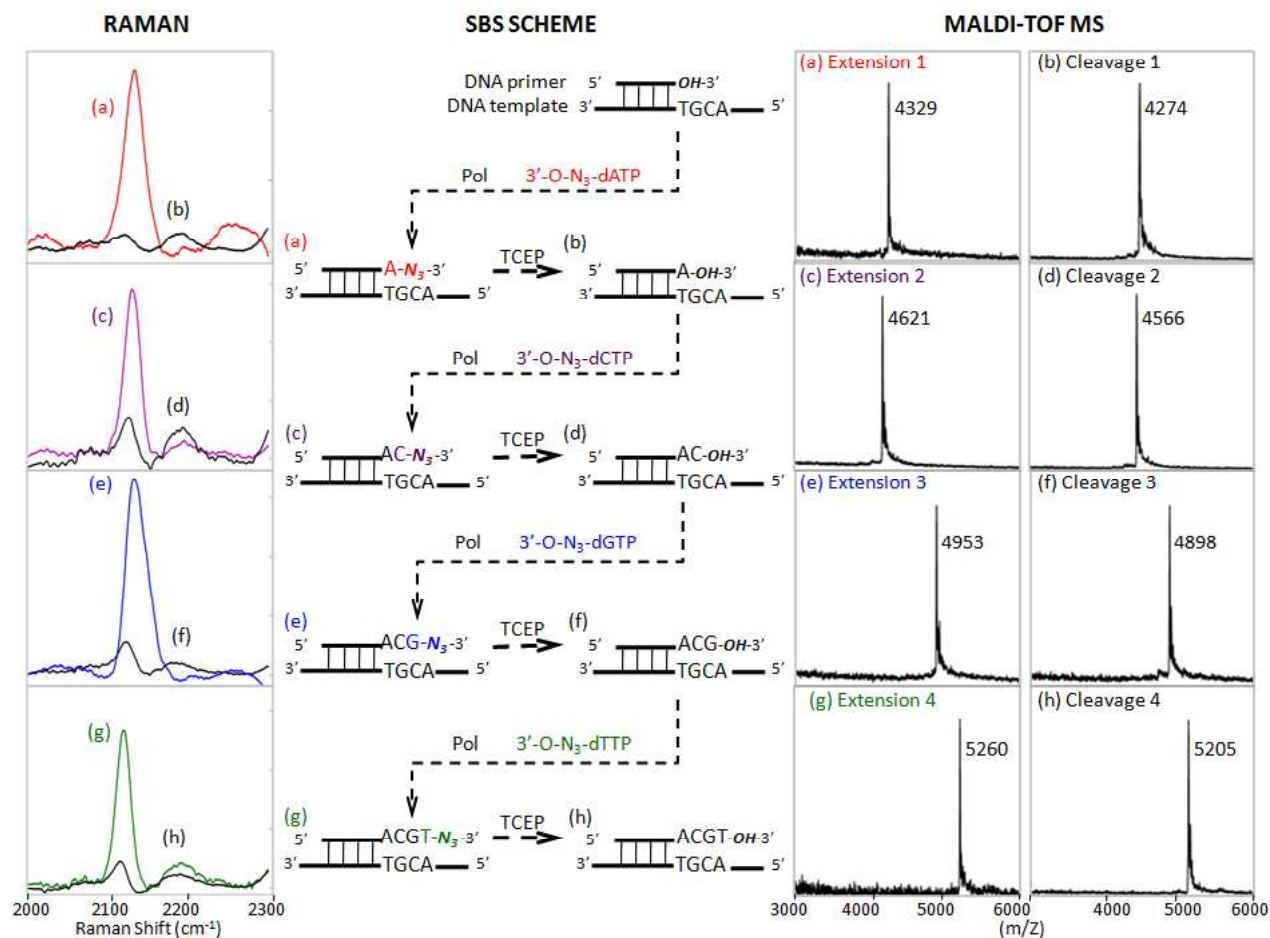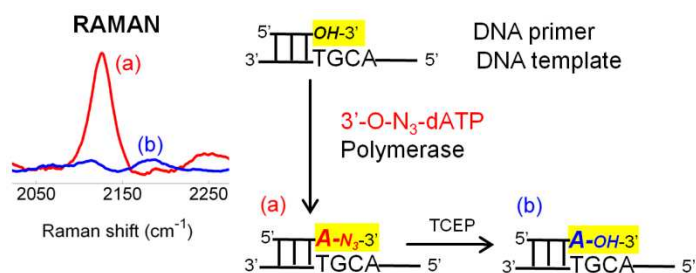
Fig. 3 Experimental scheme of continuous DNA sequencing by synthesis (middle) using four 3'-*O*-azidomethyl nucleotide reversible terminators (3'-*O*-N$_3$-dNTPs) with Raman (left) and MALDI-TOF MS spectra (right) obtained for each DNA product at each step. Only the 2000-2300 cm^{-1} Raman spectral window is shown with the azide peak appearing at ~2125 cm^{-1}. Pol = DNA polymerase.

Table of contents entry

RAMAN

(a)

(b)

2050   2150   2250

Raman shift (cm^{-1})

5'   *OH-3'*   DNA primer
3'   TGCA — 5'   DNA template

$3'-O-N_3-dATP$
Polymerase

(a) ↓

5'   *A-N_3-3'*
3'   TGCA — 5'

→ TCEP →

(b)

5'   *A-OH-3'*
3'   TGCA — 5'

DNA sequencing using surface enhanced Raman spectroscopy for the detection of the unique Raman shift of the azido group was demonstrated using 3′-O-azidomethyl nucleotide reversible terminators in polymerase reaction.