Electronic Supplementary Material (ESI) for ChemComm. This journal is © The Royal Society of Chemistry 2015

# Design and synthesis of fluorescence-labeled nucleotide with cleavable azo

# linker for DNA sequencing

Lianjiang Tan<sup>†</sup>, Yazhi Liu<sup>†,I</sup>, Qinglai Yang<sup>†</sup>, Xin-Yan Wu<sup>I</sup>, Xiaowei Li<sup>§</sup>, Bing Gong<sup>¶,‡</sup> Yu-Mei Shen<sup>†,\*</sup>

and Zhifeng Shao§

 <sup>†</sup>Shanghai Center for Systems Biomedicine, Key Laboratory of Systems Biomedicine (Ministry of Education),
Collaborative Innovation Center of Systems Biomedicine, and <sup>§</sup>Bio-ID Center, School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
<sup>†</sup>Key Laboratory for Advanced Materials and Institute of Fine Chemicals, East China University of Science and Technology, Shanghai 200237, China
<sup>‡</sup>College of Chemistry, Beijing Normal University, Beijing 100875, China

<sup>‡</sup>Department of Chemistry, University at Buffalo, State University of New York, Buffalo, NY 14260, USA

# Supporting Information

#### Contents

General Methods	S2
Experimental Section	S2-S5
Supplementary Data	S6-S7
DNA Extension/Cleavage Cycles	S7-S11
Copies of <sup>1</sup> NMR, <sup>13</sup> C NMR and <sup>31</sup> P NMR spectra	S12-S20

#### **General Methods**

Chemicals needed in this work were purchased from commercial sources (Alfa and Shanghai Chemical Reagent Company) and directly used without further purification. All the chemicals were stored in the presence of CaCl<sub>2</sub> in a desiccator under vacuum at room temperature and protected from light. All reactions were followed by thin-layer chromatography (precoated 0.25 mm silica gel plates from Aldrich) and silica gel column chromatography was carried out with silica gel 60 (mesh 200-400). The NMR spectra were recorded on a Varian MERCURY plus-400 (400 MHz for <sup>1</sup>H-NMR; 100 MHz for <sup>13</sup>C-NMR; 162 MHz for <sup>31</sup>P-NMR) spectrometer with chemical shifts reported in ppm relative to the residual deuterated solvents or the internal standard tetramethylsilane. High-resolution mass spectrometry analysis was carried out using a Quadrupole-Time-of-Flight Premier Mass Spectrometer (Waters, US) equipped with an electrospray interface. Data were collected in centroid mode from massto-charge ratio (m/z) 50 to 1000 at scan time of 0.25 s with an interval of 30 s. High-performance liquid chromatography (HPLC) was carried out on a Prominence LC-20A HPLC (Shimadzu, Japan) at ambient temperature. An SB-C18 analytical column ( $4.6 \times 250$  mm, 5 µm, Agilent) was used and the traces were recorded with a UV detector at 254 nm. Conditions for compound 1: 1 mL/min, methanol in 8.6 mM Et<sub>3</sub>N and 100 mM HFIP aqueous solution with 15 % methanol in 5 min, followed by 15 %~40 % methanol in 5 min and 40 %~70 % methanol in 30 min. Conditions for compound 5 and its cleavage products: 1 mL/min, methanol in water with 2 % methanol in 0.01 min, followed by 5 % methanol in 5 min, 20 % methanol in 15 min, 40 % methanol in 50 min, and 80 % methanol in 50.01 min. The measurements were conducted in triplicate. The fluorescence spectra of dUTP-azo linker-5(6)TAMRA (1) and TAMRA in water were recorded on a LS 55 luminescence spectrometer (PerkinElmer, US), applying an excitation wavelength of 560 nm.

#### **Experimental Section**

4-ethyl-2-((4-(2-hydroxyethyl)phenyl)diazenyl)phenol (3). 274 mg (2 mmol) of 2-(4aminophenyl)ethanol dissolved in mixture of ethanol (5 mL) and water (5 mL), followed by sodium nitrite (145mg, 2.1mmol) dissolved in water (10 mL) and then added dropwise, the solution was cooled to 0 °C. The mixture was stirred for 40 min, sodium hydroxide (80 mg, 2 mmol) was added, followed by 244 mg (2 mmol) of 4-ethylphenol dissolved in mixture of ethanol (45 mL) and water (15 mL) in an ice bath. Then the new synthesized diazonium salt was added dropwise to 4-ethylphenol solution. The mixture was stirred for 5 hrs. The precipitate was filtered and washed with water (3 x 10 mL). Then the precipitate was dried under vacuum overnight to give crude product 225 mg (yield 42%). <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>):  $\delta$  ppm 12.72 (s, 1H), 7.80 (d, 2H, J = 8.4 Hz), 7.75 (d, 1H, J = 2.4 Hz), 7.37 (d, 2H, J = 8.4 Hz), 7.18 (d, 1H, J = 2.4 Hz, 8.4 Hz), 6.95 (d, 2H, J = 8.4 Hz), 3.90 (t, 2H, J = 6.4 Hz), 2.94 (t, 2H, J = 6.4 Hz), 2.69 (q, 2H, J = 7.6 Hz), 1.75 (s, 1H), 1.29 (t, 3H, J = 7.6 Hz). <sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>):  $\delta$ ppm 150.8, 149.5, 142.3, 137.2, 135.8, 133.2, 131.8, 130.1, 122.4, 118.0, 63.4, 39.1, 27.9, 15.7.

### Tert-butyl(2-(2-(4-ethyl-2-((4-(2-

**hydroxyethyl)phenyl)diazenyl)phenoxy)ethoxy)ethoxy)ethyl)carbamate (4).** 50 mg (0.124 mmol) of 2, 2-dimethyl -4-oxo -3, 8, 11-trioxa -5-azatridecan-13-yl 4-methylbenzenesulfonate **2** and 19 mg (0.136 mmol) of K<sub>2</sub>CO<sub>3</sub> was added to a 10 mL flask, followed by 101 mg (0.372 mmol) of 4-ethyl-2-((4-(2-hydroxyethyl)phenyl)diazenyl)phenol **3**. The mixture was heated to 120°C for 2.5 hrs under a nitrogen atmosphere and then cooled to room temperature. Ethyl acetate (20 mL) was added and the organic layer was washed with water (3 x 20 mL) and concentrated. The crude mixture was purified by column chromatography (2:1 PE/EA) to give 56 mg of product **4** (yield 90 %). <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>):  $\delta$  ppm 7.85 (d, 2H, J = 8.4 Hz), 7.47 (d, 1H, J = 2.0 Hz), 7.36 (d, 2H, J = 8.4 Hz), 7.25 (dd, 1H, J = 2.0 Hz, 8.4 Hz), 7.03 (d, 1H, J = 8.4 Hz), 5.05 (s, 1H), 4.34 (t, 2H, J = 4.8 Hz), 3.95~3.91 (m, 4H), 3.78 (t, 2H, J = 4.8 Hz), 3.60 (t, 2H, J = 4.8 Hz), 3.49 (t, 2H, J = 5.2 Hz), 3.26~3.23 (m, 2H), 2.94 (t, 2H, J = 6.4 Hz), 2.65 (q, 2H, J = 7.6 Hz), 1.40 (s, 9H), 1.24 (t, 3H, J = 7.6 Hz). <sup>13</sup>C NMR (100 MHz, CDCl<sub>3</sub>):  $\delta$  ppm

154.6, 152.0, 142.7, 142.1, 137.6, 131.7, 129.8, 123.2, 116.2, 115.5, 79.3, 71.0, 70.3, 70.3, 69.9, 63.4, 40.4, 39.2, 28.5, 28.1, 15.7.

### 2-(6-(dimethylamino)-3-(dimethyliminio)-3H-xanthen-9-yl)-4-((2-(2-(2-(4-ethyl-2-((4-(2-

dUTP-AZO-5(6)TAMRA (1). To a 10 mL flask containing 13 mg (0.016 mmol) of 2-(6-(dimethylamino)-3-(dimethyliminio)-3H-xanthen-9-yl)-4-((2-(2-(2-(4-ethyl-2-((4-(2-hydroxyethyl) phenyl)diazenyl)phenoxy)ethoxy)ethoxy)ethyl)carbamoyl)benzoate 7, 2 mL of dry acetonitrile was added, followed by 24.6 mg (0.096 mmol) of DSC, 3.9 mg (0.032 mmol) of DMAP and 5.2 mg (0.04 mmol) DIPEA. The mixture was stirred for 6 hrs. Then the crude product was dissolved in 0.8 mL of acetonitrile. The prepared **dUTP(AP\_3)** (20 mg, 38.4 mmol) dissolved in 1.5 mL of Na<sub>2</sub>CO<sub>3</sub>/NaHCO<sub>3</sub> buffer was added, followed by 6  $\mu$ L triethylamine. The mixture was stirred for 3 hrs and then concentrated. The crude product was purified by preparative HPLC with C-18 reverse phase column (250 mm x 4.6 mm) to give the terminal product 1 (3.5 mg, yield 16%). <sup>1</sup>H NMR (400 MHz, D<sub>2</sub>O):  $\delta$  ppm 8.25 (s, 1H), 7.83 (s, 1H), 7.41 (d, 2H, J = 7.2 Hz), 7.16 (s, 2H), 6.92~6.66 (m, 7H), 6.55~6.49 (m, 2H), 6.40 (s, 2H), 5.93 (s, 1H), 4.49 (s, 1H), 4.24~3.99 (m, 8H), 3.87~3.68 (m, 9H), 3.10 (s, 3H), 3.08 (s, 3H), 3.06 (s, 3H), 3.03 (s, 3H), 2.86-2.88 (m, 2H), 2.19~2.07 (m, 3H), 1.97~1.92 (m, 1H), 0.64-0.66 (m, 3H); <sup>31</sup>P NMR (D<sub>2</sub>O, 162 MHz):  $\delta$  -6.76, -10.82, -20.07; HRMS: calcd for C<sub>60</sub>H<sub>66</sub>N<sub>8</sub>O<sub>23</sub>P<sub>3</sub> [M+3H]+ 1359.3454, found 1359.3459.

A disulfide linker-based reversible terminator dUTP-S-S-TAMRA. The detailed synthetic procedure and characterization for the reversible terminator dUTP-S-S-TAMRA were reported in Ref. (1). <sup>1</sup>H-NMR (400 MHz, D<sub>2</sub>O):  $\delta$  8.29 (s, 1H), 8.18 (d, J = 7.6 Hz, 1H), 7.80 (s, 1H), 7.60 (d, J = 8.0 Hz, 1H), 7.30 (d, J = 9.6 Hz, 1H), 7.18 (d, J = 9.6 Hz, 1H), 7.00~7.05 (m, 1H), 6.91 (d, J = 9.6 Hz, 1H), 6.70~6.76 (m, 2H), 5.84~5.88 (m, 1H), 5.47 (s, 1H), 4.49 (s, 1H), 4.10~4.16 (m, 2H), 3.96 (s, 2H), 3.85~3.90 (m, 3H), 3.72~3.77 (m, 2H), 3.25 (d, J = 9.6 Hz, 8H), 3.17 (t, J = 4.8 Hz, 2H), 2.94~2.99 (m, 2H), 2.65~ 2.75 (m, 2H), 2.26~2.32 (m, 1H), 2.04~2.15 (m, 1H), 1.24~1.31 (m, 1H). HRMS: calc for C<sub>42</sub>H<sub>46</sub>N<sub>6</sub>O<sub>19</sub>S<sub>2</sub>P<sub>3</sub> [M-H]<sup>-</sup> 1095.1459, found 1095.1483.



Scheme S1. Schematic cleavage of the azo linker 5.



Fig. S1. HPLC spectra of the azo linker 5 (a) and the cleavage products 5a and 5b (b).



Fig. S2. HRMS spectra of the cleavage products 5a and 5b from the azo linker 5.



Fig. S3. Fluorescence spectra of TAMRA and the nucleotide analogue dUTP-azo linker-5(6)TAMRA (1) in water at the concentration of 10  $\mu$ M.

#### **DNA Extension/Cleavage Cycles**

**Primer extension and cleavage of the extension product.** A 24-mer oligonucleotide, 5'-GAGGAAAGGGAAAGGGAAAGGAAGG-3' (molecular weight of 7663), and a fluorescence-labeled 24-mer oligonucleotide with the same base sequence (Dylight 800-5'-GAGGAAAGGGAAAGGGAAAGGAAGG-3'), were used as a primer for MALDI-TOF MS and denaturing gel electrophoresis, respectively. Two oligonucleotides with slightly different base sequences

were used as templates, and another three oligonucleotides were for comparison purpose, as listed in Table S1. The molecular weight of the template 2 was 11998. The primer and template were annealed in a Tris-EDTA buffer (TE, pH 7.5) according to the following protocol: 95 °C for 3 min, followed by a decrease to 4 °C at a ramp rate of 0.1 °C /sec and holding at 4 °C. The annealed primer and template mixture was used to perform extension reaction. The components were mixed in the order in a microfuge tube on ice: 2 µL of 10× Klenow reaction buffer, 1 µL of 1 M NaCl, 2 µL of annealed template/primer (1 µg/µL), 5 µL of 1 mM dUTP-azo linker-5(6)TAMRA (1), 0.4 µL (2 U) of Klenow Fragment exo<sup>-</sup> DNA polymerase and Type I water to a final volume of 20 µL. After gentle vortex mix, the mixture was incubated at 37 °C for 12 min and then heated at 72 °C for 10 min, followed by cooling to 16 °C. The extension products were obtained by phenol chloroform extraction, ethanol precipitation, and vacuum drying. The products were treated with 30 mM sodium dithionite in NaAc/HAc buffer (pH 5.2) for 4.5 min at 37 °C to cleave the fluorophore. The cleavage process was stopped by adding an appropriate amount of Tris-HCl buffer (pH 8.5), yielding cleavage products that could be used for the next cycle of extension and cleavage. At each step, a portion of the extension or cleavage products was dissolved in water for denaturing gel electrophoresis and MALDI-TOF MS.

Denaturing Polyacrylamide Gel Electrophoresis and MALDI-TOF MS. DNA sequencing gel electrophoresis was performed on 12 % denaturing polyacrylamide gels containing 7 M of urea, using a Sequi-Gen GT Nucleic Acid Electrophoresis Cell (Bio-Rad, US). Before electrophoresis, the prepared gels were run at a voltage of 2000 V and a constant power of 40 W for 30 min at 55 °C. In the meantime, the sequencing reaction samples were mixed with a small amount of 0.1 M sodium hydroxide and were denatured into single strands by heating to 95 °C for 3 min and then cooling to room temperature rapidly. Then 1  $\mu$ L of the samples (~ 15 ng/ $\mu$ L DNA) were mixed with 1.5  $\mu$ L of 6×

loading buffer containing marker dyes, and the mixtures were then loaded on each well of the gels. Maintained at 55 °C, the gels were run under the above-mentioned conditions for 70 min. The resultant gels were observed using an Odyssey Infrared Imaging System (LI-COR Biosciences, US) under excitation of 785 nm laser. The samples at each step were resolved in water, purified by HPLC, desalted by using the ZipTip protocol and then analyzed on an Autoflex MALDI-TOF mass spectrometer (Bruker Daltonics, US). 3-Hydroxypropionaldehyde (3-HPA) was used as matrix. The samples were spotted on the layer of the matrix and measured in negative ion mode.

Superparamagnetic Beads Experiments. Dynabeads® magnetic beads (10 mg/mL, Thermo Fisher, US) were used as substrates for cycles of extension and cleavage reactions. A 60-mer 5'-dual biotin modified oligonucleotide (dual biotin-5'-CAGTCGGTGATAGAGTGGTGCGCGTGCGCGTCAA AAAAGAGAATGAGGAACCCGGGGGCAG-3') was used as the template. A 23-mer 5'-AMCA modified oligonucleotide (AMCA-5'-CTGCCCCGGGTTCCTCATTCTCT-3') was used as the primer. The beads were washed with 1× B&W Buffer for three times before use. The biotinylated template was added to the washed beads, which were incubated for 15 min at room temperature under gentle rotation to immobilize the template. The template-coated beads were separated with a magnet for 2 min and washed with 1× B&W Buffer twice. The number of template molecules bound to one bead was about 10<sup>6</sup>. Then the primer was added to the beads and an annealing process as mentioned above was performed. The annealed primer and template-coated beads were separated with the magnet, washed with 1× B&W Buffer twice, and used for cycles of extension and cleavage. The dUTP-azo linker-5(6)TAMRA (1) was incorporated into the primer on the beads using the same extension protocol as described above. The resultant beads were separated, washed, and immersed in the 30 mM sodium dithionite in NaAc/HAc buffer (pH 5.2) for 4.5 min at 37 °C to cleave the fluorophore. The beads were then separated and washed for the next cycle of extension and cleavage. The beads at each step were observed using a Ti-U inverted phase contrast microscope (Nikon, Japan). The fluorescence emission in the TAMRA channel under excitation of 560 nm laser was detected for both the DNA-coated beads and blank beads. For each image, 20 beads were analyzed and the emission intensity was given as average  $\pm$  standard deviation. The emission in the AMCA channel under excitation of 350 nm was also detected to ensure that the amount of primer on the beads kept approximately constant during the cycles of extension and cleavage.

No.	Sequence
1	3'-CTCCTTTCCCTTCCCTTCCATCGATCGCCATGTCG-5'
2	3'-CTCCTTTCCCTTCCCTTCCCAAAAGTCGCCATGTGC-5'
3	3'-CTCCTTTCCCTTCCCTTTCCTTCCGTCGATCGCCATGTCG-5'
4	3'-CTCCTTTCCCTTCCCTTCCTACGATCGCCATGTCG-5'
5	3'-CTCCTTTCCCTTCCCTTCCCTCGATCGCCATGTCG-5'

Table S1. Sequence of Templates.





dUTP-azo linker-5(6)TAMRA (1) and different templates. Lane 1: the primer (Dylight 800-5'-

GAGGAAAGGGAAGGGAAAGGAAGG-3'). Lane 2: the 25-mer oligonucleotide(Dylight 800-5'-

GAGGAAAGGGAAAGGGAAAGGAAGGT-3'). Lane 3: extended primer using template 1 (cleaved). Lane 4: primer extension product using template 3. Lane 5: primer extension product using template 4. Lane 6: primer extension product using template 5. In lanes 4-6, the primer was not extended as the signals were similar to that of the primer.



## Reference

(1) Tang, D.; Jiang, M.; Li, X.; Kang, Y.; Wu, X.; Gong, B.; Shao, Z.; Zhao, X.; Shen, Y. Chem. J. Chinese U. 2014, 35, 2346-2352.

# Copies of <sup>1</sup> H-NMR, <sup>13</sup>C NMR and <sup>31</sup>P NMR Spectra for the Compounds





4-ethyl-2-((4-(2-hydroxyethyl)phenyl)diazenyl)phenol (3)



## 4-ethyl-2-((4-(2-hydroxyethyl)phenyl)diazenyl)phenol (3)



tert-butyl(2-(2-(2-(4-ethyl-2-((4-(2-hydroxyethyl)phenyl) diazenyl) phenoxy) ethoxy) ethoxy) ethyl) carbamate (4)





# tert-butyl(2-(2-(4-ethyl-2-((4-(2-hydroxyethyl)phenyl)diazenyl) phenoxy)ethoxy) ethyl) carbamate (4)





## 

2-(6-(dimethylamino)-3-(dimethyliminio)-3H-xanthen-9-yl)-4-((2-(2-(2-(4-ethyl-2-((4-(2-hydroxyethyl)phenyl)phenoxy)ethoxy)ethoxy)ethoxy)ethyl)carbamoyl)benzoate (6)



# 2-(6-(dimethylamino)-3-(dimethyliminio)-3H-xanthen-9-yl)-4-((2-(2-(2-(4-ethyl-2-((4-(2-hydroxyethyl)phenyl)phenoxy)ethoxy)ethoxy)ethoxy)ethoxy)benzoate (6)

LYZ013



## <sup>1</sup>H-NMR for dUTP-azo-5(6)TAMRA (1)



## <sup>31</sup>P-NMR for dUTP-azo-5(6)TAMRA (1)



## HRMS for dUTP-azo-5(6)TAMRA (1)



HPLC purity for dUTP-azo-5(6)TAMRA (1)











## HRMS for dUTP(AP<sub>3</sub>)



