Electronic Supplementary Material (ESI) for ChemComm.

Supporting Information

Synthesis and evaluations of an acid-cleavable, fluorescently labeled nucleotide as a reversible terminator for DNA sequencing

Lianjiang Tan[†], Yazhi Liu[‖], Xiaowei Li[§], Xin-Yan Wu[‖], Bing Gong[¶,‡], Yu-Mei Shen[†,*] and Zhifeng Shao[§]

[†]Shanghai Center for Systems Biomedicine, Key Laboratory of Systems Biomedicine (Ministry of Education), Shanghai Jiao Tong University, Shanghai 200240, China.

[‖]Key Laboratory for Advanced Materials and Institute of Fine Chemicals, East China University of Science and Technology, Shanghai 200237, China.

[§]Bio-ID Center, School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China.

[¶]College of Chemistry, Beijing Normal University, Beijing 100875, China.

[‡]Department of Chemistry, University at Buffalo, State University of New York, Buffalo, NY 14260, United States.

E-mail: yumeishen@sjtu.edu.cn

Materials and Methods

Materials and Reagents. Chemical reagents were purchased from Alfa, Aldrich and Aladin unless otherwise indicated. All the reagents were used without further purification. Oligonucleotides used as primers or template were designed by ourselves and synthesized by Sangon Biotech. Therminator DNA polymerase and 1x ThermoPol reaction buffer were provided by New England Biolabs.

General Methods. All synthetic reactions were monitored 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 ^{1}H-NMR; 100 MHz for ^{13}C-NMR; 162 MHz for ^{31}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 mass-to-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 at ambient temperature. An SB-C18 analytical column (4.6 × 250 mm, 5 μm, Agilent) was used and the traces were recorded with a UV detector at 254 nm. Conditions: 1 mL/min, methanol in 0.1 % Et_3N aqueous solution with 0-20 % methanol in 35 min.

Experimental Details and Results

2-((2-(2-hydroxyethoxy)propan-2-yl)oxy) ethyl 4-methylbenzenesulfonate (2). Compound **1** (1.0 g, 6.098 mmol) and triethylamine (0.43 mL, 3.1 mmol) were dissolved in DCM (7.5 mL), followed by tosyl chloride (0.291 g, 1.524 mmol) was added in one portion. The resulting mixture was stirred for 3h at room temperature. And the solvent was removed under vacuum. The residue was purified by column chromatography (eluted with 30 % ethyl acetate in petroleum ether) to give **2** as a clear, colorless oil (380 mg, 78.4 % yield)**:** ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.79 (d, J = 8.0Hz, 2 H), 7.34 (d, J = 8.0Hz, 2 H), 4.14 (t, J = 4.8Hz, 2H), 3.71-3.63 (m, 4H), 3.49 (t, J = 4.8Hz, 2H), 2.45 (s, 3H), 1.32 (s, 6H); ^{13}C NMR (100 MHz, $CDCl_3$): δ ppm 144.9, 133.1, 129.9, 128.0, 100.3, 69.7, 62.2, 62.1, 58.8, 24.8, 21.7.

2-((2-(2-azidoethoxy)propan-2-yl)oxy)ethanol (3). Compound **2** (187 mg, 0.59 mmol) was dissolved in DMF (2.5 mL), followed by sodium azide (84.1 mg, 1.29 mmol) was added and stirred at 80 °C for 10 h. The reaction mixture was extracted with ethyl ether three times. The combined organic extract was washed with brine and dried over

anhydrous Na_2SO_4. After filtration and evaporation, the residue was purified by column chromatography (eluted with 30 % ethyl acetate in petroleum ether) to give compound **3** as a colorless oil (69 mg, 62.3%). ^{1}H NMR (400 MHz, $CDCl_3$): δ 3.76 (t, J = 4.0Hz, 2H), 3.66~3.60 (m, 4H), 3.39 (t, J = 4.0Hz, 2H), 2.07 (s, 1H), 1.42(s, 6H); ^{13}C NMR (100 MHz, $CDCl_3$): δ 100.5, 62.3, 62.2, 60.1, 51.1, 24.9.

2-((2-(2-aminoethoxy)propan-2-yl)oxy)ethanol (dimethyl ketal linker 4). A suspension of compound **3** (46 mg, 0.243 mmol), 10% Pd/C (5 mg) in MeOH (3 mL) was stirred under 1 atm of hydrogen pressure for 8 h, and then filtered. The filtrate was concentrated *in vacuo* to afford the product **4** (25 mg, 64%) as a colorless oil. This compound was used in the next step without further purification (60 mg, 83% yield): ^{1}H NMR ($CDCl_3$, 400 MHz): δ 3.76 (t, J = 4.4Hz, 2H), 3.62~3.60 (m, 4H), 2.91 (t, 2H, J = 4.8Hz), 2.83 (s, 1H), 1.41(s, 6H); ^{13}C NMR ($CDCl_3$, 100 MHz): *δ* 99.8, 62.4, 62.2, 61.6, 41.6, 24.9.

2-(6-(dimethylamino)-3-(dimethyliminio)-3H-xanthen-9-yl)-4-((2-((2-(2-hydroxyethoxy)propan-2-yl)oxy)ethyl)carbamoyl)benzoate (5). To a solution of DMKL **4** (10 mg, 0.061 mmol) in DMF (1.5 mL) was added 5(6)-TAMRA, SE (20 mg, 0.038 mmol) and Et_3N (80 μL, 0.57 mmol). The mixture was protected from light and stirred at room temperature for 3 h. Then the solvent was removed under vacuum and the residue was purified with RP-HPLC to obtain the product (retention time = 24.79 min) as a triethylammonium salt with a red solid (15 mg, 69 %): ^{1}H NMR (400 MHz, CD_3OD): δ ppm 8.52 (d, J = 1.6Hz, 1H), 8.06 (dd, J = 2.0Hz; 8.0Hz, 1H), 7.36 (d, J = 7.6Hz, 1H), 7.25 (d, J = 9.2Hz, 2H), 7.02 (dd, J = 2.8Hz, 9.6Hz, 2H), 6.91 (d, J = 2.4Hz, 2H), 3.72(t, J = 6 Hz, 2H), 3.67～3.60 (m, 4H), 3.59～3.55 (m, 2H), 3.27 (s, 12H)), 1.40 (s, 3H), 1.39 (s, 3H).

2-(6-(dimethylamino)-3-(dimethyliminio)-3H-xanthen-9-yl)-4-((2-((2-(2-((((2,5-dioxopyrrolidin-1-yl)oxy) carbonyl)oxy)ethoxy)propan-2-yl)oxy)ethyl)carbamoyl)benzoate (6). Compound **5** (9.0 mg, 0.0156 mmol) and Et_3N (15 μL, 0.158 mmol) were dissolved in CH_3CN (anhy, 1.5 mL), then N, N′-disuccinimidyl carbonate (26 mg, 0.102 mmol) was added. After the reaction mixture was stirred at room temperature for 4 h, the solution was used directly for the next step without further purification.

dUTP-DMKL-5(6)TAMRA (7). The crude compound **6** was added to the 0.1 M sodium bicarbonate buffer (pH 9.16, 1.5 mL) of **dUTP(AP$_3$)** (16 mg, 0.031 mmol), the resulting reaction mixture was protected from light and periodically

vortexed for 2 h. After lyophilization, the resulting residue was dissolved in triethylammonium acetate (TEAA) buffer (20 mM, pH 7.0, 1.0 mL) and purified with RP-HPLC on a 250 × 4.6 mm C_{18} column to obtain the product (retention time = 27.9 min) as a triethylammonium salt. The product was then precipitated from NaCl/ethanol to remove significant amounts of TEA salts and finally give sodium salt of **7** as a red solid (5.76 mg, 33.1 %). Mobile phase: A, 0.1% TEA in water; B, methanol. Elution was performed by a linear gradient of 0-20% B for 35 min at a flow rate of 1 mL/min: ^{1}H-NMR (400 MHz, D_2O): δ 8.31(s, 1H), 8.27 (d, J = 7.6Hz, 1H), 7.68 (s, 1H), 7.52 (d, J = 7.6Hz, 1H), 7.26 (d, J = 9.2Hz, 2H), 7.01 (d, J = 9.6Hz, 1H), 6.95 (d, J = 9.6Hz, 1H), 6.86 (d, J = 9.6Hz, 1H), 6.75 (s, 1H), 6.66 (s, 1H), 5.70 (s, 1H), 4.40 (s, 1H), 4.04～4.00 (m, 4H), 3.87～4.73 (m,7H), 3.62～3.58 (m, 2H), 3.25 (s, 6H), 3.23 (s, 6H), 2.19～1.91 (m, 2H), 1.48 (s, 3H),1.46 (s, 3H); ^{31}P NMR (D_2O, 162 MHz): -19.33, -10.82, -5.37; HRMS: calc for $C_{45}H_{52}N_6O_{22}P_3$ $[M+3H]^-$ 1121.2348, found 1121.2373; calc for $C_{45}H_{51}N_6O_{22}P_3Na$ $[M+2H+Na]^-$ 1143.2167, found 1143.2161.

Table S1 Cleavage of compound **5** at pH 4.1 (25 °C)

Time (min)	Conversion of compound **5** (%)
1	59.75
2	73.11
4	94.36
6	98.18
8	99.48
10	100

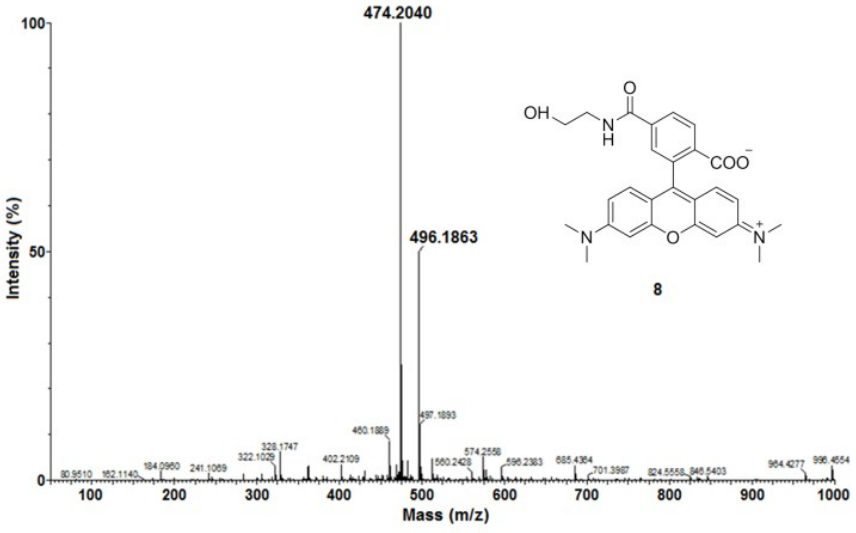


Figure S1. ESI-HRMS spectrum of the cleavage product **8.**

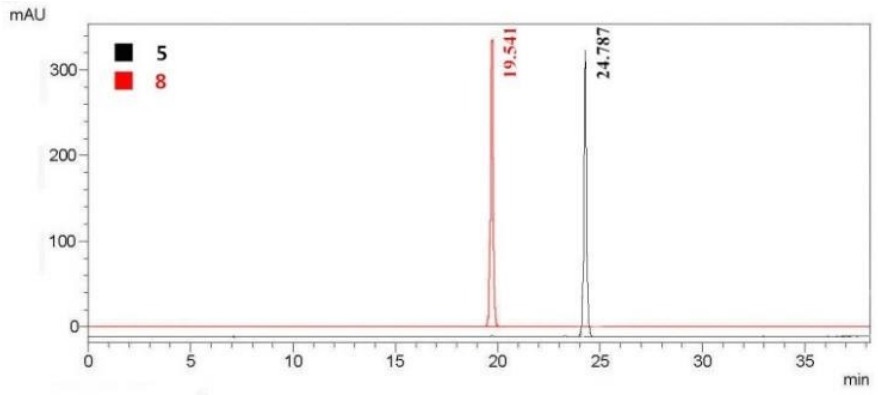


Figure S2. HPLC chromatogram of cleavage experiment. Conditions: Agilent Eclipse Plus C18 column 5 μm, 4.6 mm × 250 mm, flow rate 1.0 mL/min, gradient 0 min 70 % A, 30 % B →20 min 40 % A, 60 % B→20 min 20 % A, 80% B (A: 0.1 % TEA, B: Methanol).

Primer extension and acid-induced cleavage of the extension products. A 24-mer oligonucleotide, 5'-GAGGAAAGGGAAGGGAAAGGAAGG-3' (molecular weight of 7663), and a 5'-Dylight 800 modified 24-mer oligonucleotide with the same base sequence (Dylight 800-5'-GAGGAAAGGGAAGGGAAAGGAAGG-3'), were used as a primer for MALDI-TOF MS and denaturing gel electrophoresis, respectively. Several oligonucleotides with different base sequences were used as templates, as listed in Table S1. The molecular weight of template 3 was 11999. 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 directly to perform extension reaction. The components were mixed in the order in a microfuge tube on ice: 10 μL of 1x ThermoPol reaction buffer, 3 μL of 1 M NaCl, 2 μL of annealed template/primer (1 μg/μL), 10 μL of 1 mM **dUTP-DMKL-5(6)TAMRA (7)**, 1 μL of Therminator DNA polymerase (2U/μL) and Type I water to a final volume of 50 μL. After gentle vortex mix, the mixture was incubated at 72 °C for 5 min and then air cooled to room temperature. The extension products were obtained by phenol chloroform extraction, ethanol precipitation, and vacuum drying. Then the extension products were treated with 10 μL of a disodium phosphate-phosphoric acid buffer (pH 2.3) for 90 s at 37 °C to remove the fluorophore. The cleavage process was stopped by adding an appropriate amount of Tris-hydrochloric acid buffer (pH 8.5), yielding cleavage products. The cleavage products were used as the primer for the next cycle of extension. At each step, a portion of the extension or cleavage products were dissolved in water for denaturing gel electrophoresis and/or MALDI-TOF MS.

Table S2. Sequence of templates.

No.	Sequence *
1	3’-CTCCTTTCCCTTCCCTTTCCTTCC**A**TCGATCGCCATGTGC-5’
2	3’-CTCCTTTCCCTTCCCTTTCCTTCC**AA**CGATCGCCATGTGC-5’
3	3’-CTCCTTTCCCTTCCCTTTCCTTCC**AAAA**GTCGCCATGTGC-5’
4	3’-CTCCTTTCCCTTCCCTTTCCTTCC**AAAAAAAAAAAA**GTGC-5’
5	3’-CTCCTTTCCCTTCCCTTTCCTTCC**A**TCGA-5’
6	3’-CTCCTTTCCCTTCCCTTTCCTTCC**G**TCGA-5’
7	3’-CTCCTTTCCCTTCCCTTTCCTTCC**T**ACGA-5’
8	3’-CTCCTTTCCCTTCCCTTTCCTTCC**C**TCGA-5’

* The position in the template strands denoting the nucleotide to be incorporated is in bold.

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 NaOH and were denatured into single strands by heating to 95 °C for 3 min and then cooling to room temperature rapidly. Then 1 μL of the samples (~ 15 ng/μL DNA) were mixed with 1.5 μL 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’-CAGTCGGTGATAGAGTGGTGCGCGTGCGCGTCAAAAAAGAGAATGAGGAACCCGGGGCAG-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 1X 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 1X B&W Buffer twice. The number of template molecules bound to one bead was about 10^6. 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 1X B&W Buffer twice, and used for cycles of extension and cleavage. The **dUTP-DMKL-5(6)TAMRA (7)** 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 disodium phosphate-phosphoric acid buffer (pH 2.3) for 90 s 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 ± 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.

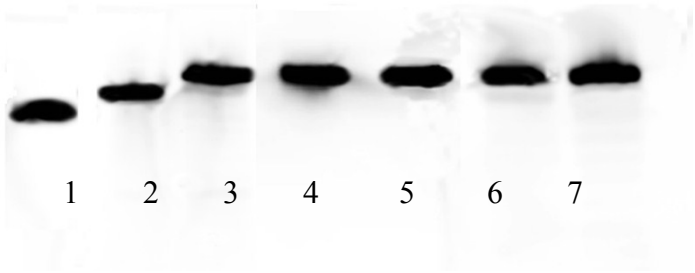


Figure S3. Fluorescent image of a denaturing polyacrylamide gel after electrophoresis. The fluorescence-labeled primer was extended using **dUTP-DMKL-5(6)TAMRA (7)** and templates with different number of successive "A". Lane 1: the primer (Dylight 800-5'-GAGGAAAGGGAAGGGAAAGGAAGG-3'). Lane 2: the fluorescence-labeled 25-mer oligonucleotide (Dylight 800-5'-GAGGAAAGGGAAGGGAAAGGAAGGT-3'). Lane 3: the extended primer using template 1. Lane 4: the extended primer using template 2. Lane 5: the extended primer using template 3. Lane 6: the extended primer using template 4. Lane 7: as lane 4 followed by acid cleavage (90 s at 37 °C). The difference between lane 1 and lane 2 indicates sufficient distinguishbility of the gel for analysis. The similar signals of lanes 3-6 show that the primer was extended by only one **dUTP-DMKL-5(6)TAMRA (7)** at a time even if several successive A's exist in the template. The signal of lane 7 was similar to those of lanes 3-6, since the fluorophore in **dUTP-**

DMKL-5(6)TAMRA (7) had been cleaved in the denaturing process before gel electrophoresis. That is to say, the signals of lanes 3-6 represent the cleaved extension products, and actually the extension products cannot be observed on the gel.

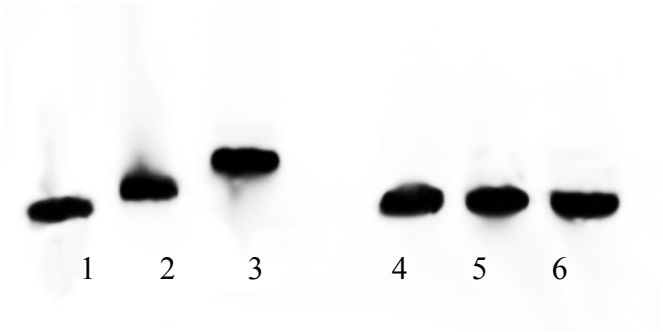


Figure S4. Fluorescent image of a denaturing polyacrylamide gel after electrophoresis. The primer was extended using **dUTP-DMKL-5(6)TAMRA (7)** and different templates. Lane 1: the primer (Dylight 800-5'-GAGGAAAGGGAAGGGAAAGGAAGG-3'). Lane 2: the fluorescence-labeled 25-mer oligonucleotide (Dylight 800-5'-GAGGAAAGGGAAGGGAAAGGAAGGT-3'). Lane 3: extended primer using template 5. Lane 4: primer extension product using template 6. Lane 5: primer extension product using template 7. Lane 6: primer extension product using template 8. In lanes 4-6, the primer was not extended as the signals were similar to that of the primer.

Figure S5. Evaluation of DNA strand stability in a disodium phosphate-phosphoric acid buffer (pH 2.3) by running a denaturing polyacrylamide gel loaded by the DNA strands. 2 μg of oligonucleotide was dissolved in 10 μL of the disodium phosphate-phosphoric acid buffer and incubated at 37 °C for different times before electrophoresis. The upper bands: an untreated 25-mer oligonucleotide. The lower bands: a 24-mer oligonucleotide treated with the disodium phosphate-phosphoric acid buffer (pH 2.3) for different times: 0 h (lane 1); 0.5 h (lane 2); 1 h (lane 3); 3 h (lane 4); 6 h (lane 5); 12 h (lane 6). The signal of the untreated 25-mer oligonucleotide is the control. The stability of the 24-mer oligonucleotide can be evaluated by comparing the signal intensity of the two bands in each lane. The

signal intensity hardly changed over a period of 3 h, and showed a visible decrease at 6 h. When the time was extended to 12 h, the band nearly disappeared.

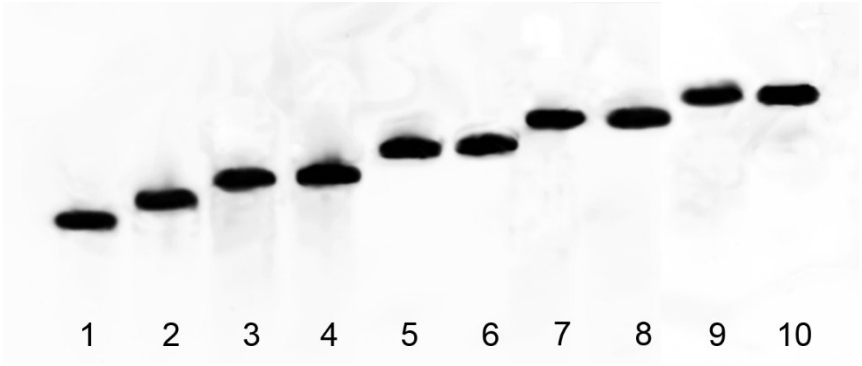
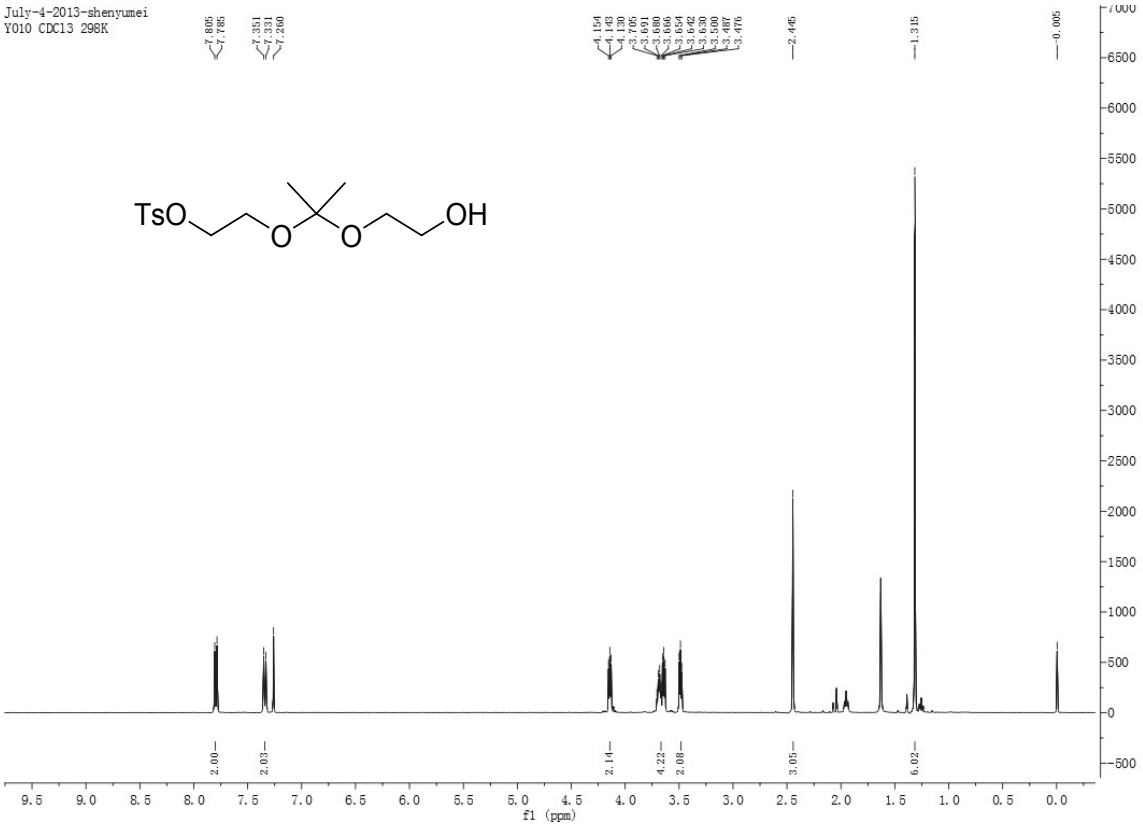


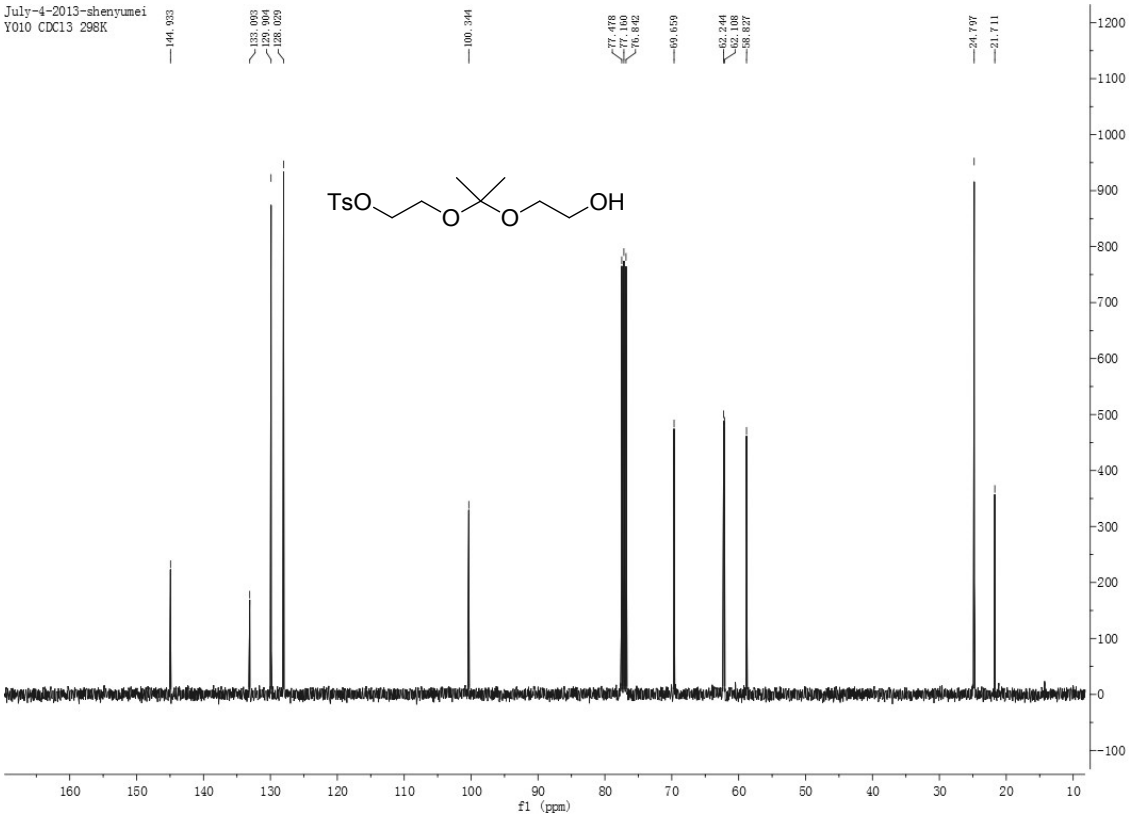
Figure S6. Fluorescent image of a sample-loading denaturing polyacrylamide gel after electrophoresis. The primer underwent four cycles of extension/cleavage reaction using template 3 (3'-CTCCTTTCCCTTCCCTTTCCTTCC**AAAA**GTCGCCATGTGC-5'). Lane 1: the primer (Dylight 800-5'-GAGGAAAGGGAAGGGAAAGGAAGG-3'). Lane 2: the 25-mer oligonucleotide (Dylight 800-5'-GAGGAAAGGGAAGGGAAAGGAAGGT-3'). Lane 3: the first extended primer. Lane 4: as lane 3 followed by the first cleavage. Lane 5: as lane 4 followed by the second extension. Lane 6: as lane 5 followed by the second cleavage. Lane 7: as lane 6 followed by the third extension. Lane 8: as lane 7 followed by the third cleavage. Lane 9: as lane 8 followed by the fourth extension. Lane 10: as lane 9 followed by the fourth cleavage.

Copies of Spectrum of Compounds:

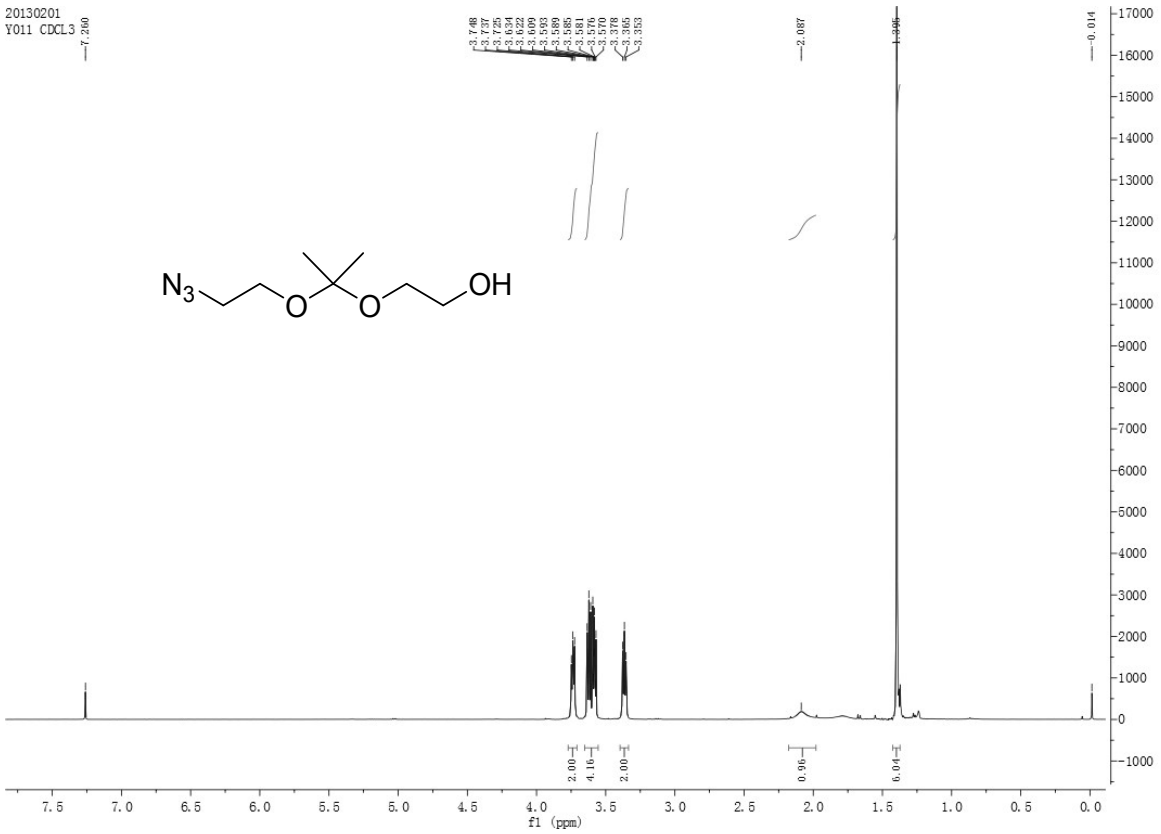
2-((2-(2-hydroxyethoxy)propan-2-yl)oxy)ethyl 4-methylbenzenesulfonate (2)



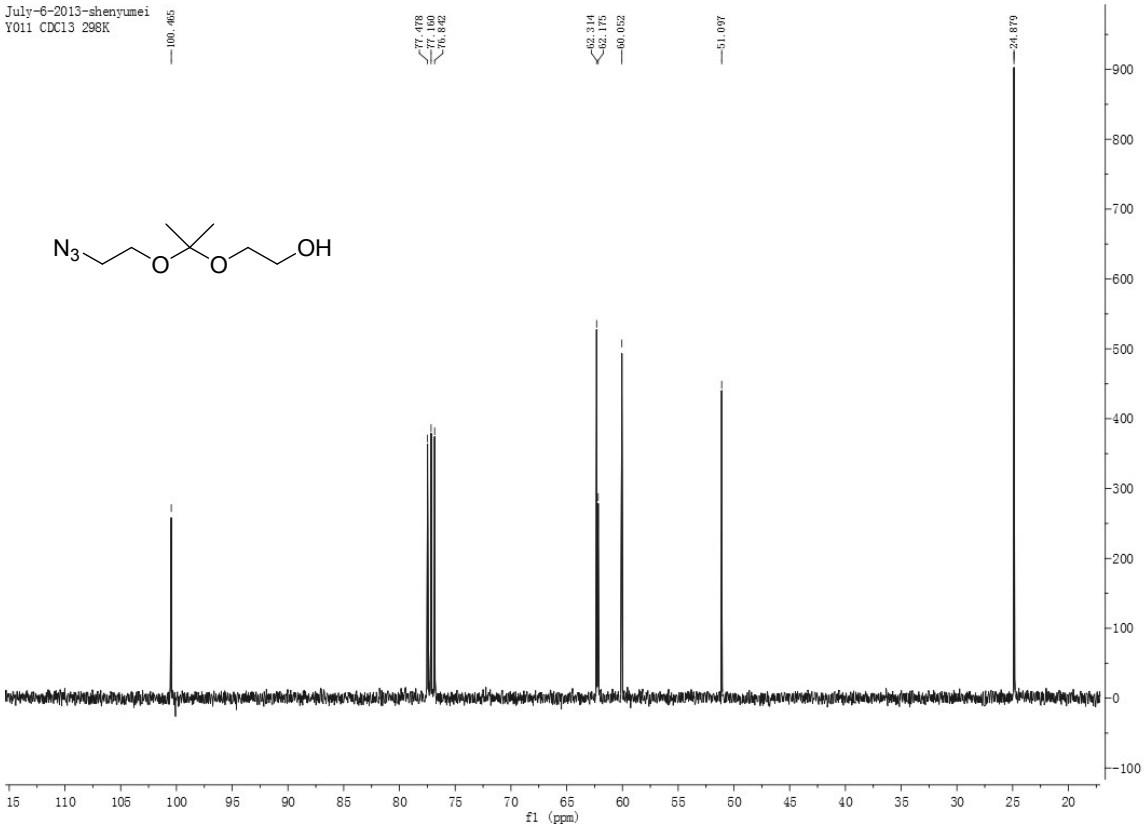
2-((2-(2-hydroxyethoxy)propan-2-yl)oxy)ethyl 4-methylbenzenesulfonate (2)



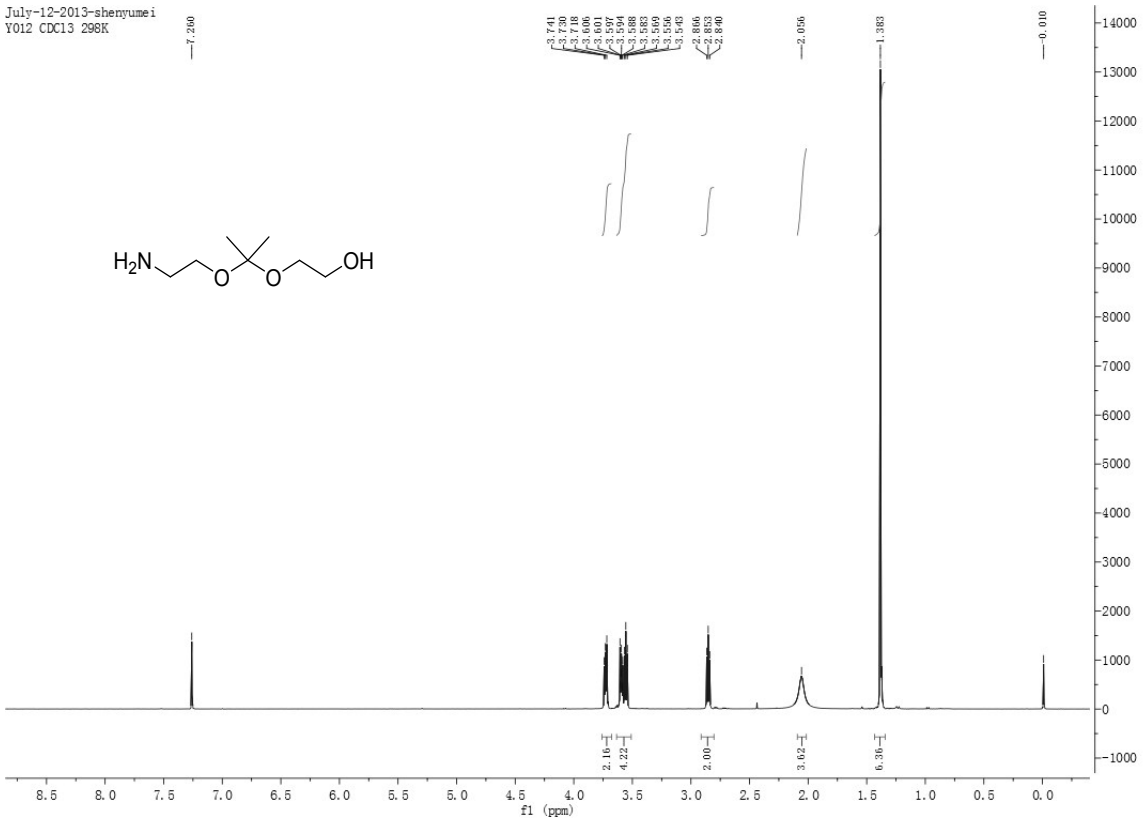
2-((2-(2-azidoethoxy)propan-2-yl)oxy)ethanol (3)



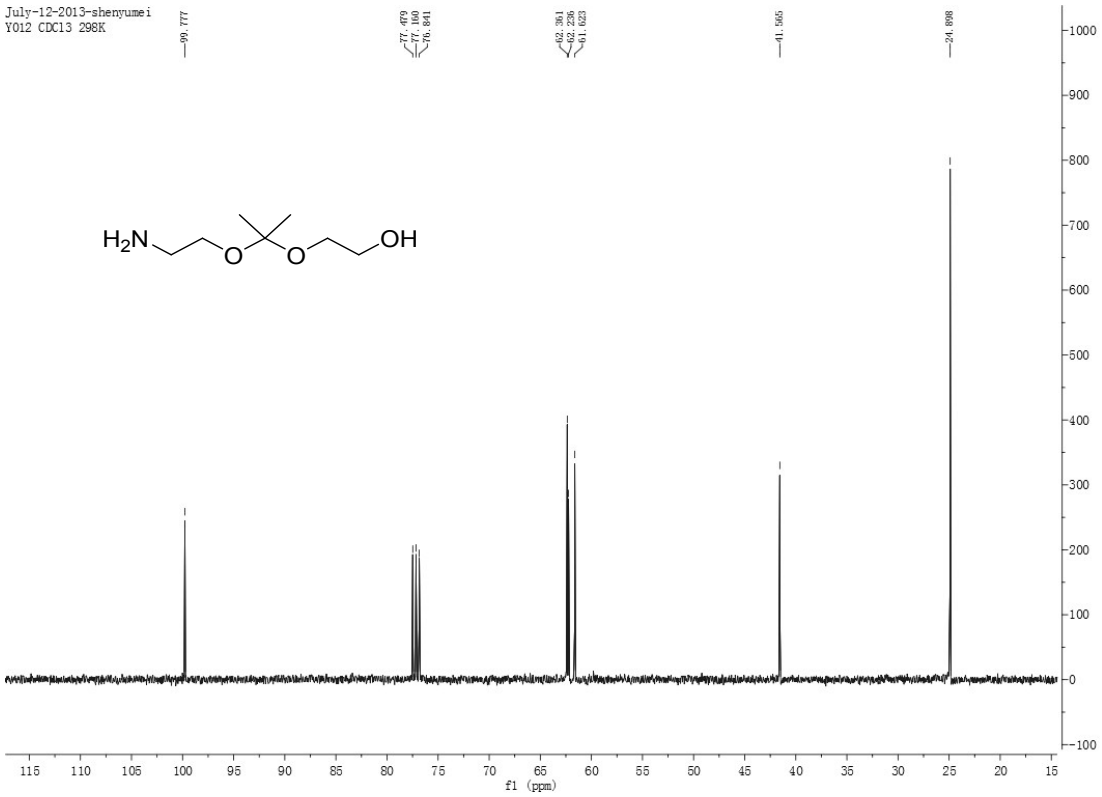
2-((2-(2-azidoethoxy)propan-2-yl)oxy)ethanol (3)



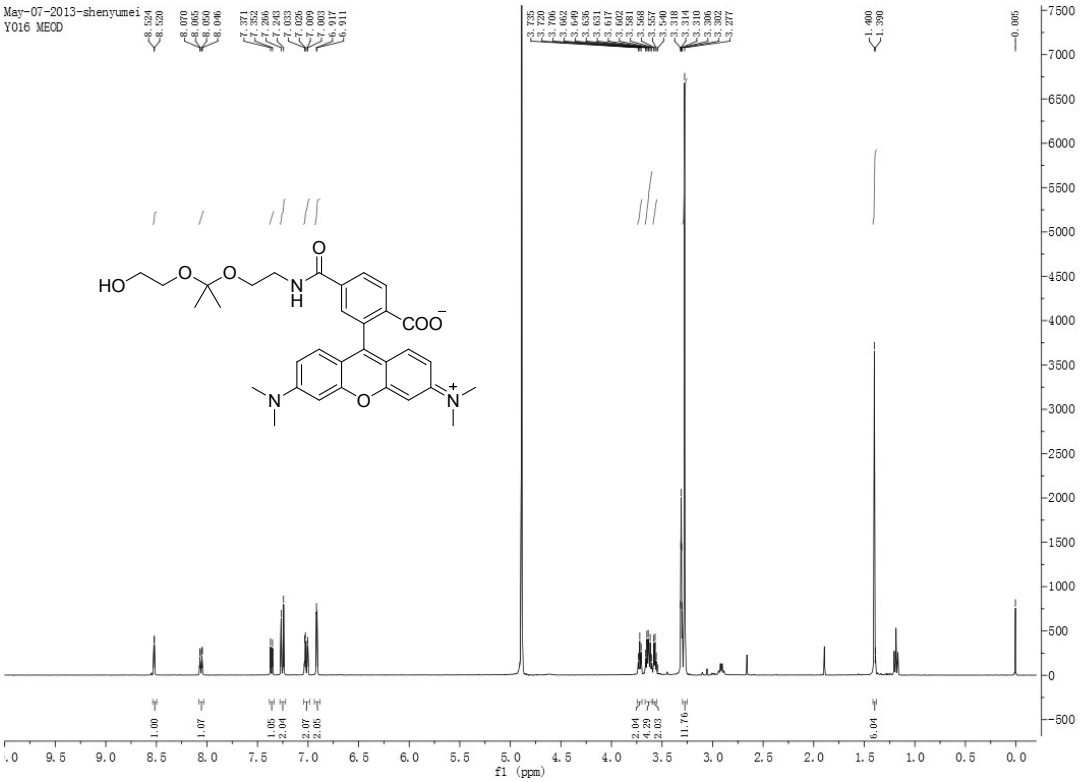
2-((2-(2-aminoethoxy)propan-2-yl)oxy)ethanol (4)



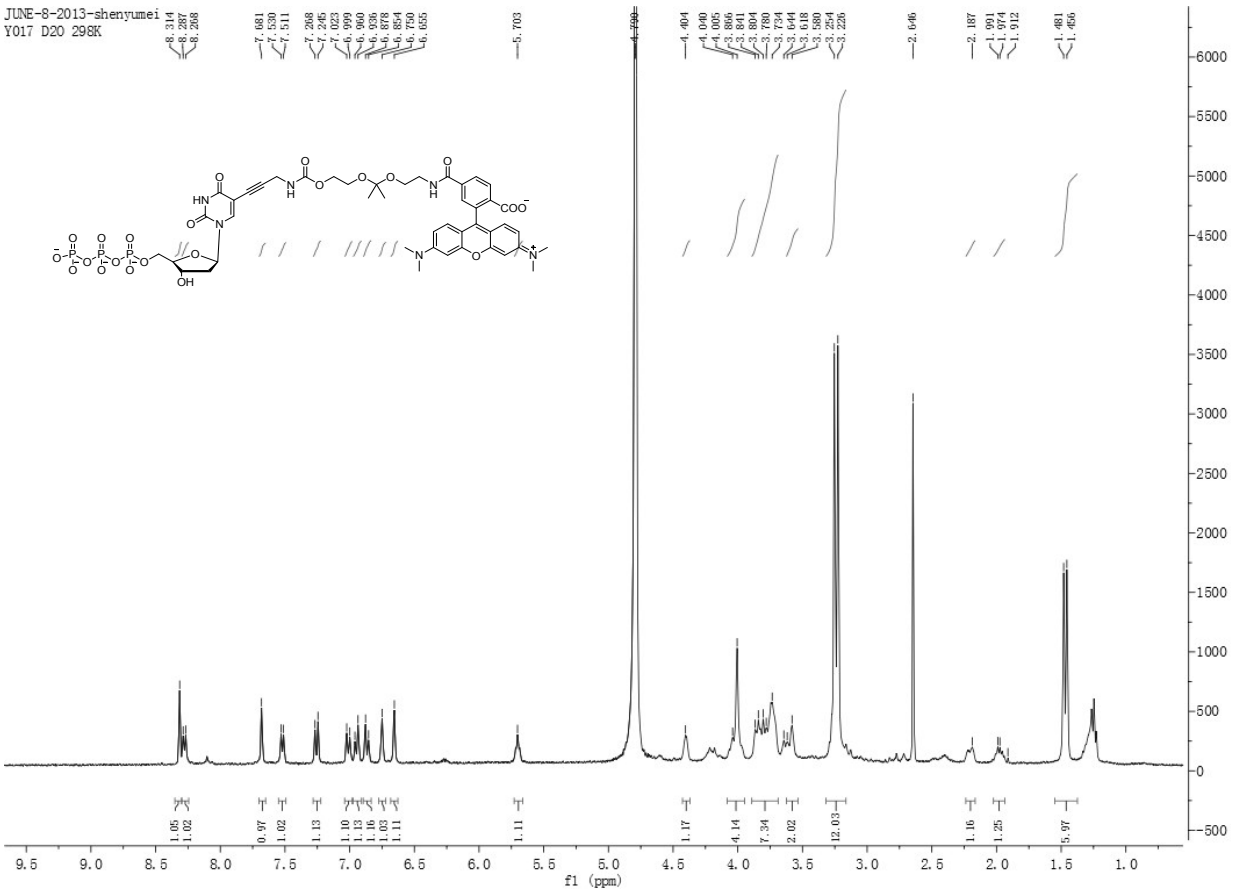
2-((2-(2-aminoethoxy)propan-2-yl)oxy)ethanol (4)



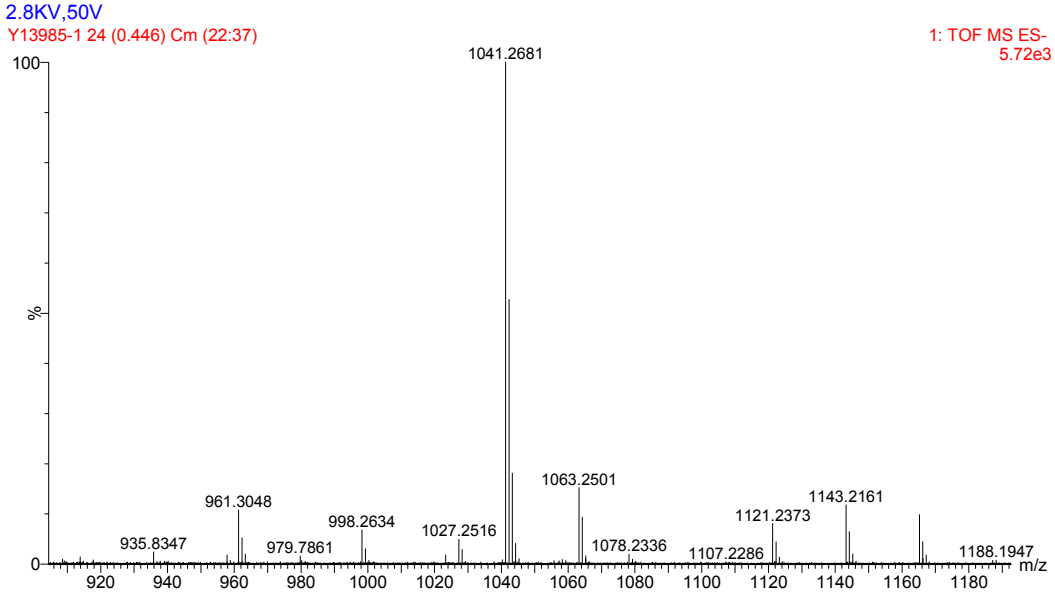
2-(6-(dimethylamino)-3-(dimethyliminio)-3H-xanthen-9-yl)-4-((2-((2-(2-((((2,5-dioxopyrrolidin-1-yl)oxy)carbonyl)oxy)ethoxy)propan-2-yl)oxy)ethyl)carbamoyl)benzoate (5)



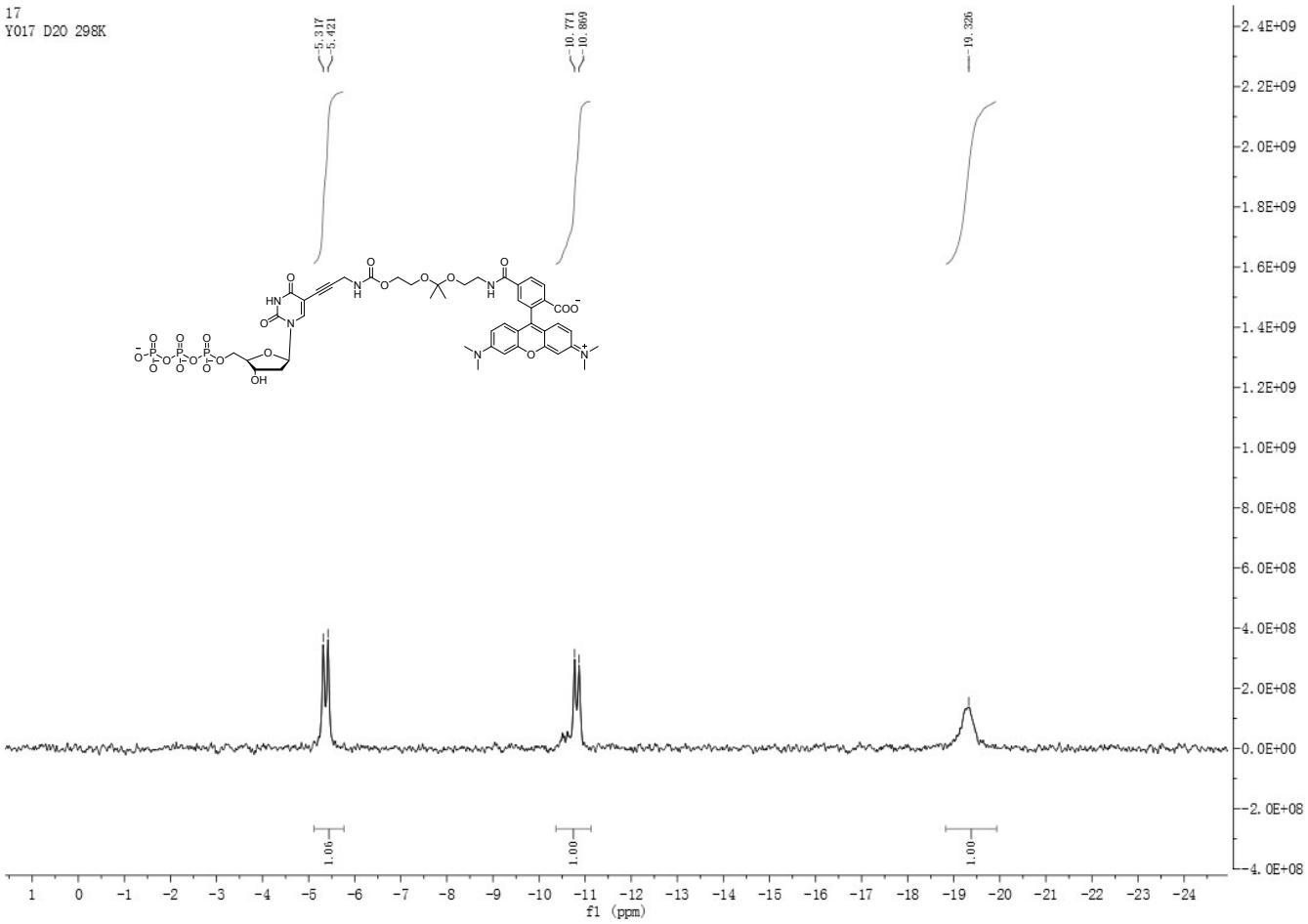
^{1}H-NMR for dUTP-DMKL-5(6)TAMRA (7)



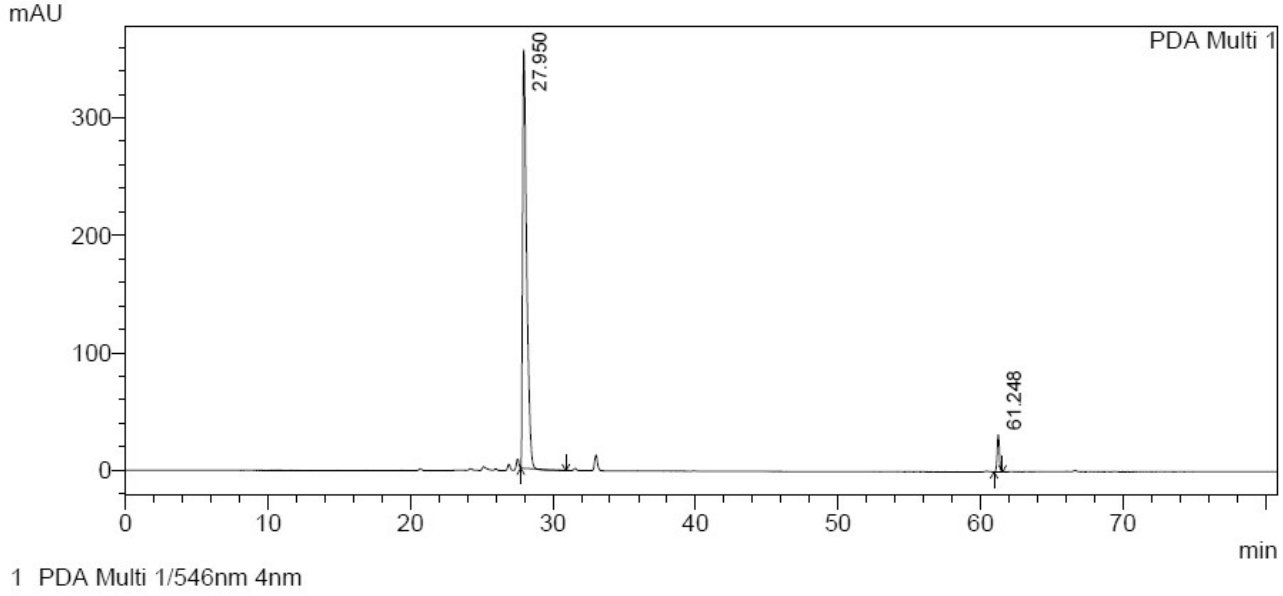
HRMS for dUTP-DMKL-5(6)TAMRA (7)



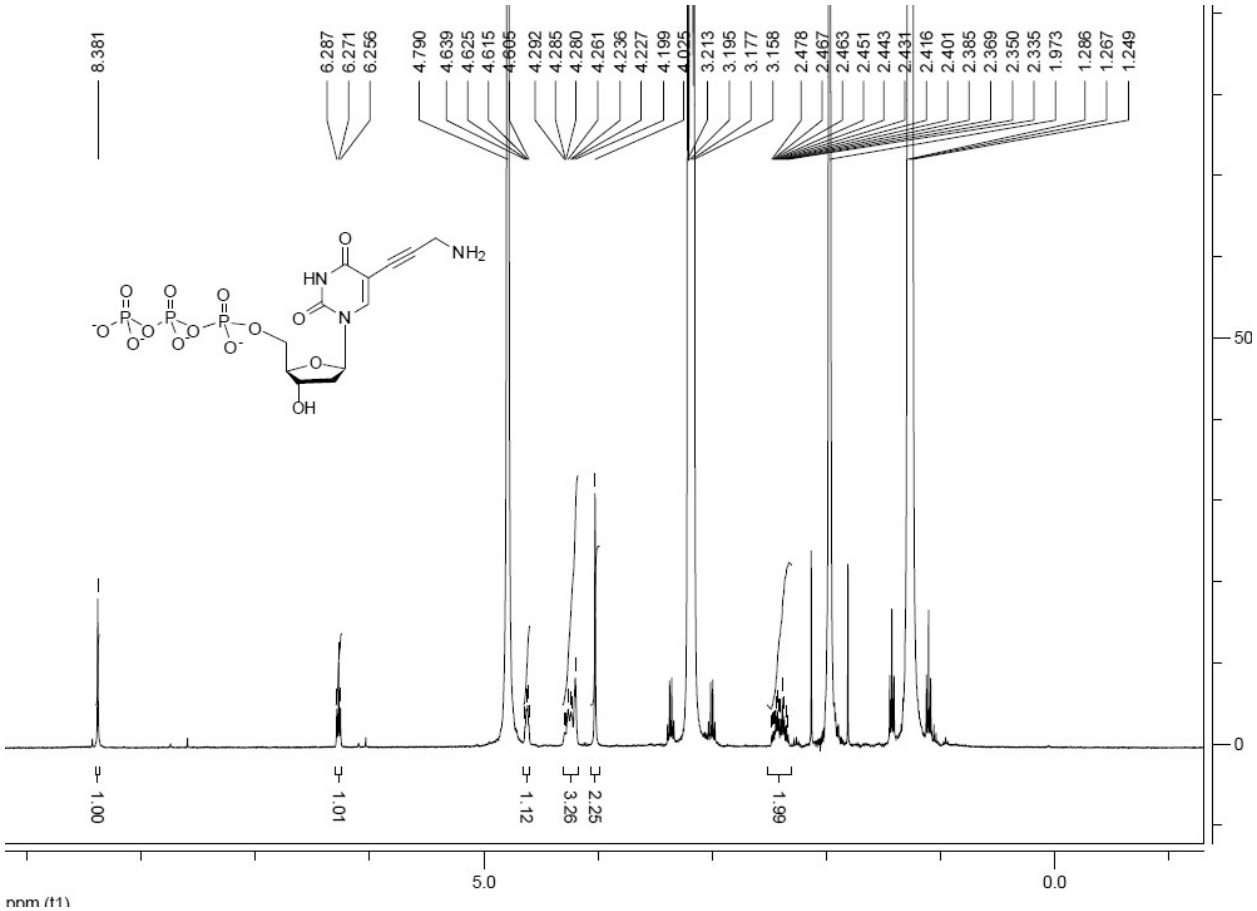
^{31}P-NMR for dUTP-DMKL-5(6)TAMRA (7)



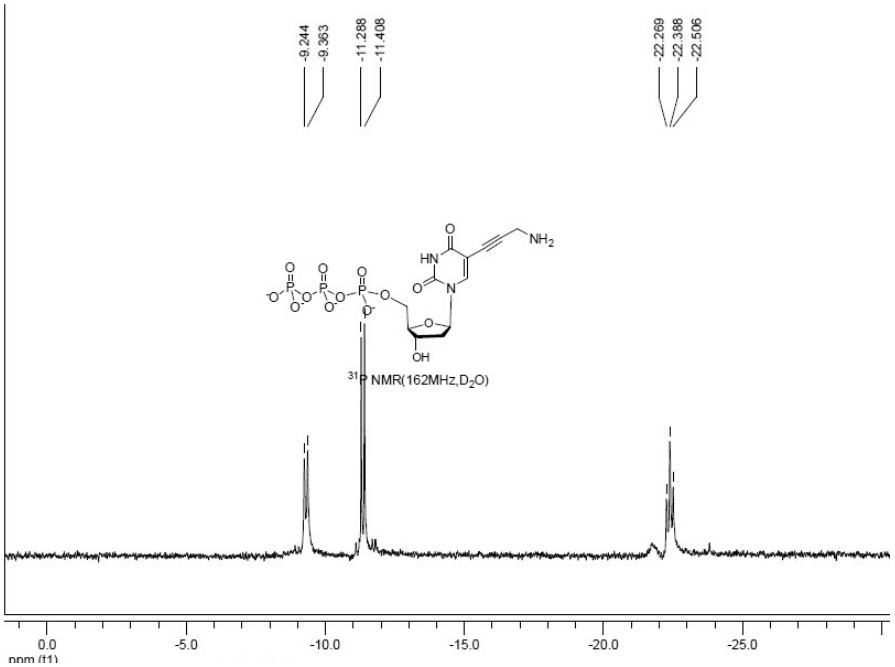
HPLC for dUTP-DMKL-5(6)TAMRA (7)



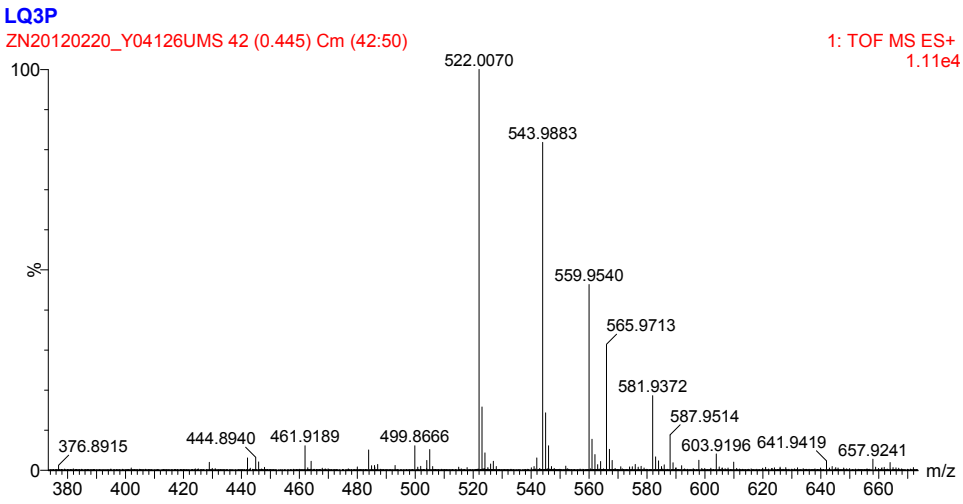
^{1}H-NMR for dUTP(AP$_3$)



^{31}P-NMR for dUTP(AP$_3$)



HRMS for dUTP(AP$_3$)



HPLC purity for dUTP(AP_3)