Supplementary Information

Photo-caged 2-butene-1,4-dial as an efficient, target-specific photo-crosslinker for covalent trapping of DNA-binding proteins

Jiahui Li, Zenghui Cui, Chaochao Fan, Yifei Zhou, Mengtian Ren and Chuanzheng Zhou*

State Key Laboratory of Elemento-Organic Chemistry, Department of Chemical Biology, College of Chemistry, Nankai University, Tianjin 300071, China.

* E-mail: chuanzheng.zhou@nankai.edu.cn

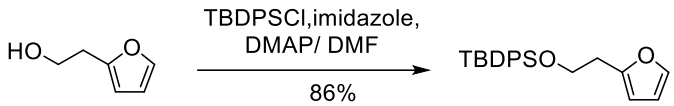
Table of Contents

Materials and methods

General Methods

All chemicals were obtained from commercial supplier and used without further purification. All the solvents used for reactions were distilled under argon after drying over an appropriate drying agent. Unmodified oligonucleotides were purchased from Sangon Biotech. Analytical thin layer chromatography (TLC) was carried out on SiLiDa silica gel GF254 plates, using $UV_{254\ nm}$ or staining with phosphomolybdic acid for visualization. Column chromatography was performed with normal phase silica gel (300 - 400 mesh). The NMR spectra were recorded on a Bruker AV 400 spectrometer. The preparative reverse phase HPLC was conducted on FLEXA HP Series fitted with Venusil prepG C18 column (120 Å, 10 μm, 21.2 mm x 250 mm). HPLC was conducted on Waters HP Series fitted with Ultimate® XB-C18 (10 × 250 mm, 5 μm). UPLC was conducted on Waters ACQUITY UPLC fitted with HSS C18 column (300 Å, 1.7 μm, 2.1 mm X 100 mm). High resolution mass spectra (HRMS) were recorded on Waters Xevo G2-XS Q-TOF mass spectrometry. The mass of intact protein and DNA were obtained by deconvolution of the raw data using MaxEnt1 tool. 365 nm light source was provided by one hand-held LED ultraviolet lamp (7.0 mW cm^{-2}).

Synthesis of 3.

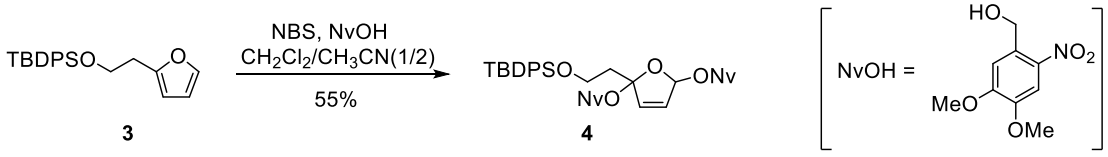


2 **3**

To a solution of **2 (**3.0 g, 26.77 mmol, 1.0 eq.) in dry DMF (60 mL) under an Ar atmosphere was added imidazole (5.38 g, 80.25 mmol, 3.0 eq.), TBDPSCl (9.10 mL, 34.80 mmol, 1.3 eq.) and DMAP (164 mg, 1.34 mmol, 0.05 eq.). After stirring at room temperature for 3 h, saturated sodium bicarbonate solution was added to quench the reaction. The solvent was removed on vacuum and the obtained residue was dissolved in EtOAc and washed with water twice. After drying over $MgSO_4$, the organic phase was concentrated and subjected to purification by flash chromatography (PE/EtOAc, v/v, 30/1-15/1) to afford oily liquid **3** (8.1 g, 23.13 mmol, yield 86%). R_f = 0.32 (PE/EtOAc, v/v, 15/1). ^{1}H NMR (400 MHz, $CDCl_3$) δ: 8.23 (s, 4H), 7.84 (s, 6H), 7.71 (s, 1H), 6.71 (s, 1H), 6.55 (s, 1H), 4.47 (t, *J* = 6.4 Hz, 1H),

3.42 (t, J = 6.4 Hz, 1H), 1.65 (s, 9H). ^{13}C NMR (100.4 MHz, $CDCl_3$) δ (ppm) : 153.0, 140.8, 135.5, 133.6, 129.6, 127.7, 110.1, 106.3, 62.3, 31.6, 26.8, 19.1. ESI-MS：[M+H] $^+$, cal. 351.1780, found 351.1807.

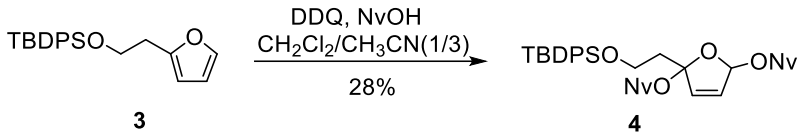
Synthesis of 4.



3 4

To a solution of **3** (1.0 g, 2.86 mmol, 1.0 eq.) and (4,5-dimethoxy-2-nitrophenyl) methanol (NvOH, 2.75 g, 12.87 mmol, 4.5 eq.) in a mixture of CH_3CN/CH_2Cl_2 (v/v 2:1, 70 mL) was added NBS (1.02 g, 5.72 mmol, 2.0 eq.) at -45 ℃. The reaction mixture was stirred for 1.5 h, then warmed to room temperature and quenched with saturated $Na_2S_2O_3$ and $NaHCO_3$ solution. The organic phase was separated and the aqueous phase was extracted with CH_2Cl_2. The combined organic layer was washed with saturated $NaHCO_3$, brine and dried over $MgSO_4$. After solvent removal using a rotary evaporator, the obtained residue was purified by silica gel flash chromatography (PE/EtOAc = 8/1-6/1-5/1-4/1, v/v). The obtained product (a mixture of two isomers) was further purified by preparative reverse phase HPLC to give two isomers in pure forms (1.2 g, 1.56 mmol, total yield 55%). RP HPLC conditions: C18, 5 μm, 21.2 mm X 250 mm. Flow 10 mL/min at rt. Buffer A: water. Buffer B: MeCN. 0-4-14-24-44-54 min, B: 0-5%-60%-80%-100%. **Isomer I** (HPLC RT = 49 min). ^{1}H NMR (400 MHz, $CDCl_3$) δ (ppm): 7.73 (s, 1H), 7.68 (s, 1 H), 7.63 (s, 2 H), 7.62 (s, 2 H), 7.38 - 7.28 (m, 6 H), 7.21 (s, 1 H), 7.18 (s, 1 H), 6.20 (dd, J = 13.3 Hz , J = 6.0 Hz, 2 H), 6.18 (d, J = 5.9 Hz, 1 H), 6.06 (s, 1 H), 5.13-4.97 (m, 2 H), 4.78-4.69 (m, 2 H), 3.99-3.98 (m, 1H), 3.96 (s, 3 H), 3.94 (s, 3 H), 3.91-3.90 (m, 1H), 3.88 (s, 3 H), 3.75 (s, 3 H), 2.36 (t, J = 6.8 Hz, 2 H), 1.00 (s, 9 H). ESI-MS: $[M + Na]^+$, cal. 797.2718, found 797. 2874. **Isomer II** (HPLC RT= 50.5, main): ^{1}H NMR (400 MHz, $CDCl_3$) δ (ppm): 7.68-7.66 (m, 5 H), 7.45-7.35 (m, 7 H), 7.19 (s, 1 H), 7.16 (s, 1H), 6.15 (s, 2 H), 5.66 (s, 1 H), 5.01-4.79 (m,4 H), 3.98 (s, 3H), 3.95-3.94 (m,1H), 3.93 (s, 3 H), 3.88-3.86 (m,1H), 3.85 (s, 3 H), 3.77 (s, 3 H), 2.35-2.23 (m, 2 H), 1.05 (s, 9 H). ^{13}C NMR (400 MHz, $CDCl_3$) δ (ppm):153.9, 153.5, 147.5, 147.0, 138.6, 135.5, 134.0, 133.6, 133.5, 132.1, 130.3, 130.0, 129.7,

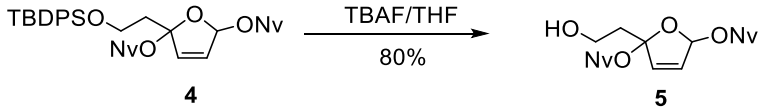
127.7, 113.2, 109.6, 109.3, 107.6, 107.2, 105.8, 67.7, 62.0, 59.8, 56.4, 56.3, 56.1, 42.4, 29.7, 26.8, 19.1. ESI-MS: [M + Na] $^{+}$, cal. 797.2718, found 797. 2758.



3 **4**

To a solution of **3** (2.0 g, 5.71 mmol, 1.0 eq.) and (4,5 - Dimethoxy - 2 -nitrophenyl) methanol (NvOH, 12.2 g, 57.10 mmol, 10 eq.) in a mixture of CH_3CN/CH_2Cl_2 (v/v, 3/1, 60 mL) was added 2,3-Dichloro-5,6-dicyano-1,4-benzoquinone (DDQ) (2.4 g, 10.28 mmol, 1.8 eq.). The reaction mixture was stirred at 35 ℃ for 1.8 h and then quenched with saturated $Na_2S_2O_3$ and $NaHCO_3$ solution. The organic layer was separated and the aqueous phase was extracted with CH_2Cl_2. The combined organic layer was washed with saturated $NaHCO_3$ and $Na_2S_2O_3$ solution and dried over anhydrous sodium sulfate. The crude product was purified as described above to give two isomers in pure forms (1.25 g, 1.62 mmol, total yield 28%).

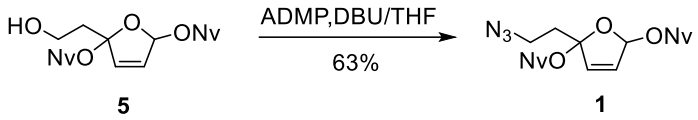
Synthesis of 5.



4 **5**

To a solution of **4** (2.0 g, 2.58 mmol, 1.0 eq.) in THF (20 mL) was added (2.84 mL, 2.84 mmol, 1.1 eq.) solution of TBAF (1 M in THF) at room temperature. After stirring for 2 h, the mixture was condensed on a rotary evaporator. The residue was purified by silica gel flash chromatography (PE/EtOAc = 2/1, v/v) to give **5** (1.11 g, 2.07 mmol, 80%) as a mixture of two isomers (orange solid). A pure form of the major isomer was obtained by further purification by HPLC. ^{1}H NMR (400 MHz, $CDCl_3$) (ppm): 7.66 (s, 1H), 7.45 (s, 1H), 7.29 (s, 1H), 7.14 (s, 1H), 6.25 (d, *J* = 6.8 Hz, 1H), 6.13 (d, *J* = 5.8 Hz, 1H), 5.80 (s, 1H), 5.06-4.86 (m, 4 H), 4.00-3.99 (m, 6 H), 3.96-3.95 (m,1H), 3.94 (s, 3H), 3.93-3.91(m,1H), 3.87 (s, 3H), 2.29-2.23 (m, 1H), 2.21-2.14 (m, 1H).^{13}C NMR (100.4 MHz, $CDCl_3$) δ (ppm): 153.9, 153.6, 147.6, 147.2, 138.8, 138.7, 133.4, 131.6, 131.3, 129.5,114.3, 109.7, 109.4, 107.7, 107.3, 106.1, 67.9, 62.3, 58.7, 56.4, 56.3, 56.2, 41.5, 29.6. ESI-MS: [M + Na] $^{+}$, cal. 559.1540, found. 559.1552.

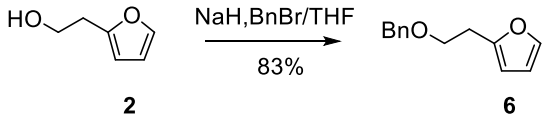
Synthesis of 1.



5 **1**

1,8-Diazabicyclo [5.4.0] undec-7-ene (DBU, 0.93 mL, 6.21 mmol, 3.0 eq.) was added to a stirred solution of **5** (1.11 g, 2.07 mmol, 1.0 eq.) and 2-azido-1,3-dimethylimidazolinium Hexafluorophosphate (ADMP) (1.80 g, 6.21 mmol, 3.0 eq.) in THF (20 mL) at room temperature. After stirring for 2 h, the mixture was condensed on a rotary evaporator. The residue was purified by silica gel flash chromatography (PE/EtOAc = 3/1, v/v) to give **1** (733 mg, 1.31 mmol, 63%) as a mixture of two isomers (bright yellow solid). ^{1}H NMR (400 MHz, $CDCl_3$) δ (ppm): 7.67 (s, 1 H), 7.48 (s, 1 H), 7.27 (s, 1 H), 7.18 (s, 1 H), 6.26 (d, *J* = 5.7 Hz, 1 H), 6.10 (d, *J* = 5.8 Hz, 1 H), 5.80 (s, 1 H), 5.03-4.84 (m, 4 H), 4.0 (s, 6 H), 3.94 (s, 3 H), 3.88 (s, 3 H), 3.58-3.44 (m, 2 H), 2.33-2.18 (m, 2 H). ^{13}C NMR (100.4 MHz, $CDCl_3$) δ (ppm): 153.9, 153.5, 147.6, 147.3, 138.9, 138.7, 134.3, 133.0, 131.5, 129.6, 127.5, 112.7, 109.8, 109.5, 107.7, 107.4, 106.1, 67.9, 62.2, 56.4, 56.3, 56.2, 46.7, 38.4, 29.6. ESI-MS: $[M+H]^+$, cal. 562.1785, found. 562.1813.

Synthesis of 6.

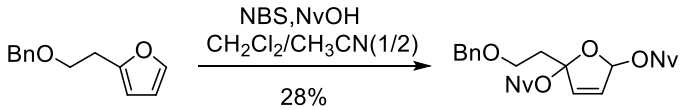


2 **6**

NaH (60%, dispersion in Paraffin Liquid) (326.8 mg, 81.66 mmol, 3.0 eq.) was suspended in 40 ml anhydrous THF under Ar atmosphere, and the mixture was cooled on an ice bath. The cold mixture was added dropwise to a solution of **2** (3.05 g, 27.22 mmol, 1.0 eq.) in 40 ml THF in 30 min. Then the mixture was stirred at room temperature for 3 h, to which benzyl bromide (6.47 ml, 54.44 mmol, 2.0 eq.) was added. After reflux for 2 h, the reaction was cooled on an ice bath, quenched with water and condensed on a rotary evaporator. The residue was diluted with EtOAc, washed by saturated NaCl solution, dried on anhydrous $MgSO_4$. After solvent removal, the residue was purified by silica gel flash chromatography (PE/EtOAc = 20/1-10/1, v/v) to give a yellow oily liquid **6** (4.56 g, 22.56 mmol, 83%). R_f = 0.56 (PE/EtOAc= 20/1, v/v). ^{1}H NMR (400 MHz, $CDCl_3$) δ (ppm): 7.42 (s, 6 H), 6.39 (s, 1 H), 6.18 (s, 1 H),

4.62 (s, 2 H), 3.82 (t, J = 6.8 Hz, 2 H), 3.06 (t, J = 6.7 Hz, 2 H). ^{13}C NMR (100.4 MHz, $CDCl_3$) δ (ppm): 153.0, 141.0, 138.2, 128.3, 127.5, 110.1, 105.9, 72.8, 68.2, 28.8. ESI-MS: $[M+Na]^+$, cal. 225.0891, found 225.0918.

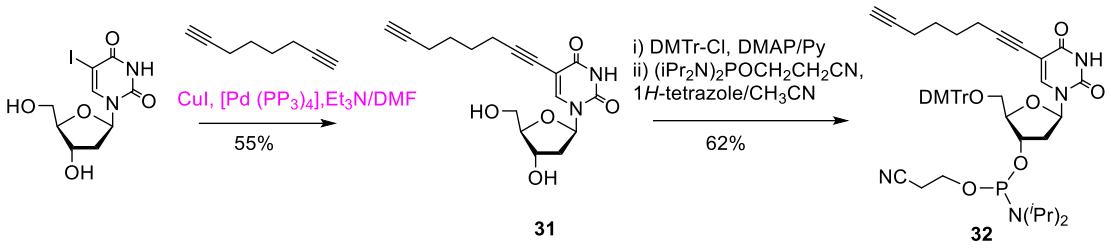
Synthesis of 7.



6 **7**

To a solution of **6** (2.5 g, 12.37 mmol, 1.0 eq.) and (4,5-dimethoxy -2-nitrophenyl) methanol (NvOH, 11.9 g, 55.67 mmol, 2.0 eq.) in a mixture of CH_3CN/CH_2Cl_2 (2/1, v/v, 150 mL) was added NBS (4.4 g, 24.74 mmol, 2.0 eq.) at -45°C. The reaction mixture was stirred for 1.5 h, then warmed to room temperature and quenched with saturated $Na_2S_2O_3$ and $NaHCO_3$ solution. The organic layer was separated and the aqueous phase was extracted with CH_2Cl_2. The combined organic layer was washed with saturated $NaHCO_3$, brine and dried over $MgSO_4$. After solvent removal, the crude product was purified by silica gel flash chromatography (PE/EtOAc = 8/1-6/1-5/1-3/1, v/v) to give orange solid 2.21 g (a mixture of two isomers, yield 28%). Rf = 0.38 (PE/EtOAc = 4/1, v/v). Two pure isomers could be separated to pure forms by preparative reverse phase HPLC: C18, 5 μm, 21.2 mm X 250 mm. Flow 10 mL/min at rt. Buffer A: water. Buffer B: MeCN. 0-6-13-48 min, B: 5%-5%-50%-70%. Form the main isomer, RT = 41 min. ^{1}H NMR (400 MHz, $CDCl_3$) δ (ppm): 7.65 (s, 1 H), 7.44 (s, 1 H), 7.33 (s, 4H), 7.26 (s, 2 H), 7.17 (s, 1 H), 6.15 (dd, J = 16.8 Hz, J = 6.1 Hz, 2 H), 5.73 (s, 1 H), 5.03-4.83 (m, 4 H), 4. 51 (s, 2 H), 3.99 (s, 3 H), 3.92 (s, 3 H), 3.89 (s, 3 H), 3.85 (s, 3 H), 3.78-3.72 (m, 1 H), 3.67-3.59 (m, 1 H), 2.33 (t, J = 6.5 Hz, 2 H). ^{13}C NMR (100.4 MHz, $CDCl_3$) δ (ppm): 153.7, 153.4, 147.4, 147.0, 138.6, 138.5, 138.1, 133.6, 131.8, 130.4, 129.7, 128.2, 127.5, 113.1, 109.6, 109.2, 107.5, 107.1, 105.8, 72.9, 67.6, 65.8, 61.9, 56.3, 56.1, 56.0, 39.3. ESI-MS: $[M+Na]^+$, cal. 649.2001, found 649.2038.

Synthesis of 32[1]



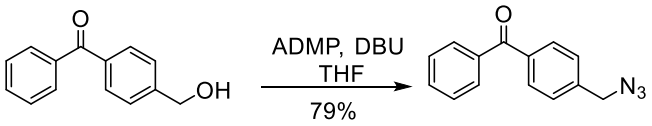
To a thoroughly degassed solution of 5-Iodo-2′-deoxyuridine (2.0 g, 5.65 mmol, 1.0 eq.), $PdCl_2(PPh_3)4$ (647.0 mg, 0.57 mmol, 0.1 eq.) and CuI (220.0 mg, 1.14 mmol, 0.2 eq.) in DMF (30 mL) was added triethylamine (1.6 mL, 11.30 mmol, 2 eq.) under argon atmosphere. The reaction mixture was stirred at room temperature for 10 minutes. 1,7-octadiyne (6.0 g, 56.50 mmol, 10.0 eq.) was added to the reaction mixture. After stirring at room temperature overnight, the mixture was added to 30 ml of mixed CH_3OH and CH_2Cl_2 (v/v, 1/1), and 1.0 g of ion exchange resin Dowex1 × 8 (100-200 mesh) bicarbonate type, followed by stirring for 1 h. Filter the mixture and wash the resin with 50 ml of mixed CH_3OH and CH_2Cl_2 solvent (v/v, 1/1). The filtrate was combined and concentrated on vacuum, then diluted with NaCl solution and extracted with CH_2Cl_2. The organic layer was separated and washed with brine (3 x 50 mL) followed by water (2 x 50 mL), dried over $MgSO_4$. Flash column chromatography eluting with a gradient of CH_2Cl_2/CH_3OH (v/v, 19/1-15/1) provided **31** as colorless oily liquid (1.03 g, 3.10 mmol, 55 %). Rf = 0.42 (CH_2Cl_2 /CH_3OH= 9/1). ^{1}H NMR (400 MHz, $CDCl_3$&DMSO-d6) δ ppm : 7.86 (s, 1H), 7.29 (s, 1H), 5.53 (s, 1H), 5.36 (t, *J* = 6.0 Hz, 1H), 5.30 (t, *J* = 6.7 Hz, 1H), 3.49-3.44 (m, 12H), 3.28-3.20 (m, 1H), 3.11(m,1H), 2.99-2.98 (m, 1H), 2.92- 2.88 (m, 1H), 2.84- 2.80 (m, 2H), 2.79- 2.75 (m, 1H), 2.27 (dd, *J* = 14.6 Hz, *J* = 7.3 Hz, 1H), 1.81 (t, *J* = 7.3 Hz, 1H), 1.69 (m, 1H), 1.57 (m, 2H), 1.51 (t, *J* = 6.2 Hz, 2 H). ESI-MS: $[M+Na]^+$, cal. 355.1270, found 355.1276.[1]

To a solution of **31** (0.84 g, 2.52 mmol, 1.0 eq.) in dry pyridine (20 mL) was added 4,4′-dimethoxytriphenylmethyl chloride (1.11 g, 3.02 mmol,1.2 eq.) under a nitrogen atmosphere and the reaction mixture was stirred 5 h. Then 1 ml CH_3OH was added for quenching the reaction. After concentration on vacuum, the crude mixture was diluted with EtOAc and saturated sodium bicarbonate solution. The mixture was extracted with EtOAc (2 times). The combined organic layer was dried ($MgSO_4$), and then subjected to flash column chromatography eluting with a gradient of EtOAc/PE (+ 1 %Et_3N, v/v, 5/1-1/1) to provide a colorless solid (1.28 g, 2.02 mmol, 80 %). Rf = 0.52 (PE /EtOAc=

1/1). ^{1}H NMR (400 MHz, $CDCl_3$) δ ppm : 8.03 (s, 1H), 7.44-7.42 (m, 2H), 7.35-7.23 (m, 6H), 7.23-7.19 (m, 1H), 6.85 (s, 2H), 6.83 (s, 2H), 6.33 (t, *J* = 6.9 Hz, 1H), 3.78 (s, 6H), 3.42 (m, 2H), 3.41 (m, 1H), 3.30 (m, 1H), 2.52-2.48 (m, 1H), 2.30-2.29 (m, 1H), 2.10-2.07 (m, 2H), 2.00-1.98 (m, 2H), 1.89 (s, 1H), 1.43-1.30 (m, 4H). ESI-MS: $[M+Na]^+$, cal. 657.2577, found 657.2574.

To a solution of the product obtained above (487.0 mg, 0.77 mmol, 1.0 eq.) in dry dichloromethane (10 mL) was added *N*, *N*-diisopropylethylamine (DIPEA) (0.54 ml, 3.08 mmol, 4.0 eq.) followed by 2-cyanoethyl *N*, *N*-diisopropylchlorophosphoramidite (0.26 ml, 1.16 mmol, 1.5 eq.) under an argon atmosphere. The reaction mixture was stirred at room temperature for 2 hours. After concentration on vacuum, the crude mixture was purified by flash column chromatography eluting with ethyl EtOAc/PE (+ 1 % Et_3N, v/v, 3/1-1/1;) to afford **32** (0.40 g, 0.48 mmol, 62 %) as a yellowish solid. Rf = 0.24 (+ 1 % Et_3N, PE /EtOAc= 1/1). ESI-MS: $[M+Na]^+$, cal. 857.3655, found 857.3657.[1]

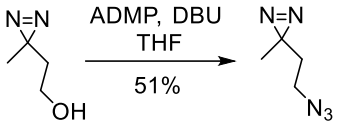
Synthesis of 34.



33 **34**

1,8-Diazabicyclo [5.4.0] undec-7-ene (DBU, 1.06 mL, 7.07 mmol, 3.0 eq.) was added to a stirred solution of **33** (0.50 g, 2.36 mmol, 1.0 eq.) and 2-azido-1,3-dimethylimidazolinium hexafluorophosphate (ADMP) (2.02 g, 7.07 mmol, 3.0 eq.) in THF (60 mL) at room temperature. After 2 h, the mixture was condensed on a rotary evaporator. The residue was purified by silica gel flash chromatography (PE/EtOAc = 4/1, v/v) to give **34** (441 mg, 1.86 mmol, 79%) (colorless liquid). Rf = 0.72. ^{1}H NMR (400 MHz, $CDCl_3$) δ (ppm): 7.68 (t, *J* = 8.4 Hz, 4H), 7.46 (t, *J* = 7.4 Hz, 1H), 7.37 - 7.29 (m, 4H), 4.31 (s, 2H). ^{13}C NMR (100.4 MHz, $CDCl_3$) δ (ppm): 196.0, 140.0, 137.4, 132.6, 130.6, 130.0, 128.4, 127.9, 54.3. ESI-MS: $[M+H]^+$, cal. 238.0980, found. 238.1006.

Synthesis of 36.



35 **36**

1,8-Diazabicyclo [5.4.0] undec-7-ene (DBU, 1.03 mL, 6.89 mmol, 3.0 eq.) was added to a stirred solution of **35** (0.23 g, 2.30 mmol, 1.0 eq.) and 2-azido-1,3-dimethylimidazolinium hexafluorophosphate (ADMP) (1.96 g, 6.89 mmol, 3.0 eq.) in THF (60 mL) at -20 ºC. After 2 h, the mixture was condensed on a rotary evaporator. The residue was purified by silica gel flash chromatography (PE/EtOAc = 2/1, v/v) to give **36** (130 mg, 1.17 mmol, 51%) (colorless liquid). Rf = 0.79. ^{1}H NMR (400 MHz, $CDCl_3$) δ (ppm): 3.20 (t, *J* = 6.8 Hz, 2H), 1.61 (t, *J* = 6.9 Hz, 2H), 1.07 (s, 3H). ^{13}C NMR (100.4 MHz, $CDCl_3$) δ (ppm): 46.2, 33.8, 23.9, 19.8.

Photo-reaction of 7 with glycinamide.

7 (final 0.8 mM) and glycinamide (20 eq.) were dissolved in a 3:2 (v/v) mixture of MeCN and phosphate buffer (40 mM, 100 mM NaCl, pH 7.4). The mixture was irradiated (365 nm, 7.0 mW cm^{-2}) for 7 min, followed by incubation at 37 ºC for 10 min. The reaction mixture was subjected to HPLC purification to afford **10** as the predominant product. ^{1}H NMR of **10** (400 MHz, $CDCl_3$) δ (ppm): 7.29 -7.19 (d, *J* = 6.5 Hz, 4H), 6.17 (s, 1H), 6.0 (s, 1H), 4.87 (d, *J* = 2.9 Hz, 1H), 4.43 (s, 2H), 4.04 (s, 2H), 3.95 (t, *J* = 5.8 Hz, 2H), 2.67 - 2.63 (m, 2H), 2.49 – 2.45 (m, 2H). ^{13}C NMR of **10** (100.4 MHz, $CDCl_3$) δ (ppm): 176.2, 169.3, 144.1, 138.1, 128.5, 127.9, 127.8, 97.7, 72.2, 65.3, 43.6, 28.5, 21.4. ESI-MS: $[M+Na]^+$: Cal. 297.1215, found 297.1216.

Photo-reaction of 7 with $BnNH_2$ and β-mercaptoehtanol (BME)

7 (final 0.8 mM), $BnNH_2$ (20 eq.) and BME (20 eq.) were dissolved in a 1:1 mixture of MeCN and phosphate buffer (40 mM, 100 mM NaCl, pH 7.4). The mixture was irradiated (365 nm, 7.0 mW cm^{-2}) for 7 min, followed by incubation at 37 ºC for 12 h. The reaction was analyzed by UPLC-MS (see Figure S1 for details). The two isomers of product **12** were isolated by HPLC and characterized by NMR. ^{1}H NMR of **12-I** (400 MHz, $CDCl_3$) δ (ppm): ^{1}H NMR (400 MHz, $CDCl_3$) δ: 7.26 -7.14 (d, *J* = 6.5 Hz, 4H),

6.88 (s, 1H), 6.86 (s, 1H), 6.51 (d, *J* = 2.9 Hz, 1H), 6.15 (d, *J* = 2.9 Hz , 1H), 5.00 (s, 2H), 4.33 (s, 2H), 3.56 (t, *J* = 5.8 Hz, 2H), 3.41 (t, *J* = 6.9 Hz, 2H), 2.91 (t, *J* = 6.8 Hz, 2H), 2.67 (t, *J* = 5.8 Hz, 2H). ^{13}C NMR of **12-I** (100.4 MHz, $CDCl_3$) δ (ppm): 138.2, 137.8, 133.9, 128.7, 128.3, 127.6, 127.5, 126.4, 121.5, 113.2, 72.9, 69.7, 60.3, 51.3, 40.2, 25.4. ESI-MS: $[M+H]^+$: Cal. 368.1684, found 368.1686.

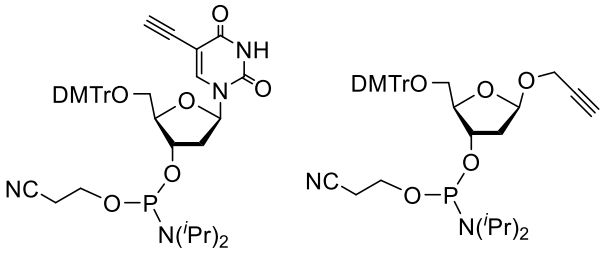
^{1}H NMR of **12-II** (400 MHz, $CDCl_3$) δ (ppm): 7.23 -7.13 (m, 8H), 6.78 (s, 1H), 6.77 (s, 1H), 6.38 (d, *J* = 3.6 Hz, 1H), 5.96 (d, *J* = 3.5 Hz, 1H), 5.26 (s, 2H), 4.36 (s, 2H), 3.54-3.48 (m, 4H), 2.69 (t, *J* = 7.0 Hz, 2H), 2.36 (t, *J* = 5.8 Hz, 2H). ^{13}C NMR of **12-II** (100.4 MHz, $CDCl_3$) δ (ppm): 138.5, 138.1, 134.1, 128.6, 128.3, 127.6, 127.1, 125.6, 118.4, 118.3, 73.0, 69.1, 60.1, 47.0, 39.9, 27.9. ESI-MS: $[M+H]^+$: Cal. 368.1684, found 368.1686.

Photo-reaction of 7 with peptides

To a solution of peptide (**13**, **16** or **19**, final 10 μM) in phosphate buffer (40 mM, pH 7.4) was added a certain amount of **7** (dissolved in 1:1 (v/v) mixture of MeCN and H_2O). The mixture was irradiated (365 nm, 7.0 mW cm^{-2}) for 7 min, followed by incubation at 37 ºC. Aliquots were taken at 20 min and analyzed by UPLC-MS. UPLC conditions: HSS C18 column (300 Å, 1.7 μm, 2.1 mm X 100 mm), flow 0.4 mL/min at 35 ºC; Gradient: 0-6 min, B: 5-85%. A: H_2O; B: MeCN.

Oligonucleotides synthesis

5fC modified oligonucleotide was prepared according to reported protocol.[2] Octadiynyl-dU phosphoramidite **32**, ethynyl-dU phosphoramidite **37** and propargyl-ribo phosphoramidite **38** [1, 3, 4] were employed for synthesis of the alkyne modified oligonucleotides.



37 **38**

Oligonucleotides containing alkyne modifications were synthesized in 1 μM scale on a ABI 394 DNA synthesis instrument. For incorporation of the modified phosphoramidite **32**, **37** and **38**, the

coupling time was extended to 10 min. After synthesis, solid supports were treated with 1 mL 28% aqueous ammonia at room temperature overnight. The crude products were purified by denaturing PAGE (20% polyacrylamide with 7 M urea) to give pure oligos whose integrity was verified by mass spectroscopy (Table S1).

Click chemistry

Alkyne-modified oligonucleotides in water (5 nmol, 10 μL) was mixed with 4-50 eq. of **1**, **34** or **36** (0.25 M in H_2O/MeCN, v/v =1/1). To the mixture was added 25 eq. of THPTA/$CuSO_4$ (0.05 M in water), and then 40 eq. of sodium ascorbate (0.1 M in H_2O, freshly prepared before use). Degas was applied by bubbling nitrogen gas for one minute. After incubation at room temperature for 15 -60 minutes, the reaction was subjected to purification by reverse phase HPLC (Venusil prepG C18 column, 120 Å, 10 μm 21.2 mm x 250 mm; flow 3 mL/min at room temperature; Gradient: 0-11-20 min, B: 5-40-40%. A: 100 mM HFIP and 7 mM TEA in H_2O; B: CH_3OH) to give PBDA-, BP- and DA-modified oligonucleotides in > 99% purity (Table S1).

Reconstitution of nucleosome core particles

The dsDNA sequence used for nucleosome reconstitution was 145 bp 601-DNA, which forms stable, well-positioned nucleosome core particles.[5] Following the procedure we reported previously,[6-8] we prepared PBDA-, BP- and DA-modified 601-DNA via ligation of synthesized oligonucleotides (see Figure S8). Then modified 601-DNA and Xenopus Laevis histone octamer were employed to reconstitute nucleosome core particles based on salt gradient dialysis.[8, 9] The reconstitution efficiency was analyzed by nucleoprotein gel electrophoresis (10 × 8 × 0.15 cm, 5 %, acrylamide/bisacrylamide, 59:1, 0.6 × TBE buffer, run at 4 ℃ using 0.2 × TBE buffer).

Photo-crosslink of PBDA-, BP- and DA-modified DNA with histones within nucleosome core particles

The reconstituted nucleosomes (approximately 45.2 nM, total 220 μL) were irradiated (365 nm) for varying times. Aliquot (4 μL) was removed and mixed with 5×SDS loading buffer, followed by 10% SDS-PAGE analysis. For samples with protease K treatment, protease K (0.1 μg) was added before

mixing with 5×SDS loading buffer.

Expression and purification of hTDG

The catalytic domain of human TDG (a.a. 82- a.a. 308) was constructed in a pET-28a vector and expressed in *Escherichia coli* BL21(DE3) cells (stratagene) as previously described.[10] Briefly, cells were grown in Luria broth (typically 2 L×2) at 15℃ until $A_{600\ nm}$ = 0.75. Then expression was induced by adding 0.25 mM IPTG (isopropyl β-D-thiogalactoside) and incubation was continued at 15 ℃ for about 15 hours. Cells were harvested by centrifugation and re-suspended in ~40 mL×2 lysis buffer (0.05 M sodium phosphate pH 8.0, 0.3 M NaCl, 0.02 M imidazole, 0.01 M β-mercaptoethanol), which was sonicated at 4 ℃ (Pulse 0.4 s, stop 0.8 s, 5 min ×4). The supernatant was obtained by centrifugation of the cell lysate. The Ni - column (HisTrapTM HP column 1 X 5 mL, GE Healthcare) was equilibrated with buffer A (containing 300 mM NaCl, 20 mM imidazole, 50 mM phosphate pH 8.0 and 10 mM β-mercaptoethanol), and the supernatant was loaded into the Ni - column. The column was washed with 85 mL elution buffer B (2 mL/min, containing 1 M NaCl, 20 mM imidazole, 50 mM phosphate pH 8.0 and 10 mM β-mercaptoethanol), followed by 73 mL elution buffer A for column balance. Then hTDG was eluted with elution buffer A and C (Gradient: 0 min– 35 min- 29 min, C: 0%-6%-24%, flow rate 2 mL/min). Elution buffer C (containing 300 mM NaCl, 1 M imidazole, 50 mM phosphate pH 8.0 and 10 mM β-mercaptoethanol). The combined fractions of pure hTDG was dialyzed overnight versus storage buffer (20 mM HEPES 7.5, 0.1 M NaCl, 1 mM DTT, 0.5 mM EDTA, 5% glycerol), concentrated to about 0.1 mM, and stored at -80 ℃. The concentration of $TDG^{82\text{-}308}$ was 27.3 μM calculated by the extinction coefficient. The purity and integrity of hTDG was verified by SDS-PAGE and mass spectroscopy.

Photo-crosslink of PBDA-modified DNA with TDG

dsDNA **26**-**29** (final 62.5 nM) and hTDG (final 625 nM) in HEPES buffer (20 mM, pH 7.5, 100 mM NaCl, 0.2 mM EDTA, 0.05 mg/ml BSA) was incubated at 4 ℃ for 2.5 h, and then at 37 ℃ for 10 min. After photo-irradiation for 7 min, the reaction was further incubated 37 ℃ for 10 min. Aliquot was removed and analyzed by Tricine-SDS-PAGE (15.5%). The gel was visualized by means of fluorescence (fluorescein, FAM) imaging, followed by Coomassie blue staining.

Pull down DNA binding proteins from cell lysate using PBDA-modified DNA

Human TDG (a.a. 82- a.a. 308) was expressed in *Escherichia coli* BL21(DE3) cells (stratagene) and cell lysis was proceeded as described above. After centrifugation, the supernatant was collected and protein

concentration was determined to be 2.82 mg/mL by BCA Protein Quantification Kit. The obtained cell lysate was used for following pull-down study.

An aqueous solution of dsDNA **30** (10 μM, 120 μL) and a proper amount of cell lysate (1.41 mg/ml) was mixed in a phosphate buffer (10 mM, pH 7.5, 50 mM NaCl, 1 mM EDTA) and incubated at 4 ºC for 2.5 h, then at 37 ºC for 10 min. After photo-irradiation for 7 min, the mixture was further incubated 37 ºC for 30 min. The obtained sample was mixed with 0.5 mL binding buffer (10 mM Tris-HCl, pH7.5, 1 mM EDTA, 1M NaCl, 0.05%Tween-20) and 200 μL streptavidin beads (＞500 pmol/mg, HY-k0208, MCE), followed by shaking for 30 minutes. The sample was placed on a magnet for 5 min, and supernatant was discarded. The beads were washed with binding buffer (1 mL) twice, high salt buffer (1 M NaCl, 20 mM Tris-HCl pH7.5 and 1 mM EDTA, 1 mL) twice, SDS solution (0.1% in PBS, 1 mL) twice, urea solution (6 M in water, 1 mL) twice, PBS with 1 M NaCl (1 mL) four times, and finally PBS (1 mL) twice. Next, 25 μL of low salt buffer (0.15 M NaCl, 20 mM Tris-HCl pH 7.5 and 1 mM EDTA) and 5 × SDS loading buffer (6.25 μL) were added to the beads, followed by incubation at 70 ºC for 3 min. The supernatant was separated and analyzed by 15% SDS-PAGE. The gel was visualized by means of fluorescence (fluorescein, FAM) imaging, followed by either Coomassie blue staining or anti-TDG western blot analysis.

Supplementary figures

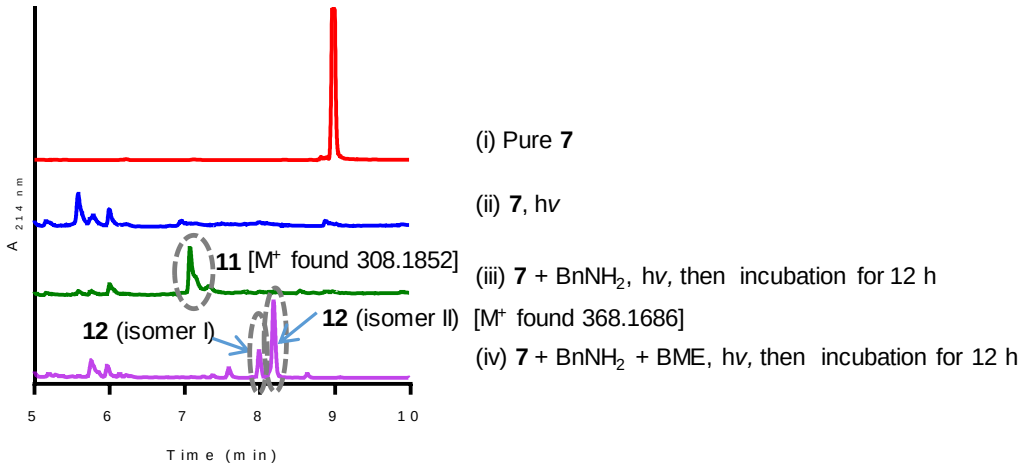


Figure S1. UPLC-MS analysis of photo-reactions between **7** with $BnNH_2$ and β-mercaptoehtanol (BME). **Reaction conditions:** **7** (final 0.8 mM), $BnNH_2$ (20 eq.) and BME (20 eq.) were dissolved in a 1:1 mixture of MeCN and phosphate buffer (40 mM, 100 mM NaCl, pH 7.4). The mixture was irradiated (365 nm, 7.0 mW cm^{-2}) for 7 min, followed by incubation at 37 ℃ for 12 h.

UPLC conditions: HSS C18 column (300 Å, 1.7 μm, 2.1 mm X 100 mm), flow 0.4 mL/min at 35 ºC; Gradient: 0-10 min, B: 5-100%. A: H_2O; B: MeCN.

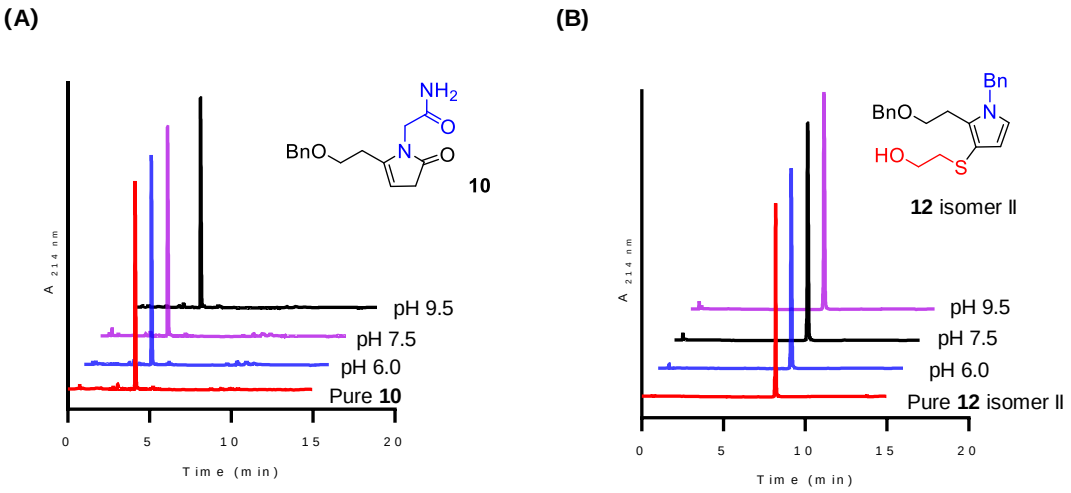


Figure S2. UPLC analysis of stability of **10** (A) and **12** (B) in different pH conditions.

Reaction conditions: A solution of substrate (final concentration 4 mM) in an appropriate buffer (MES, phospahte and sodium borate buffer were used for pH 6.0, 7.5, 9.5 respectively, final concentration 20 mM, 100 mM NaCl) was incubated at 37 ℃. Aliquots were taken at 12 h and analyzed by UPLC-MS.

UPLC conditions: HSS C18 column (300 Å, 1.7 μm, 2.1 mm X 100 mm), flow 0.4 mL/min at 35 ºC; Gradient: 0-10 min, B: 5-100%. A: H_2O; B: MeCN.

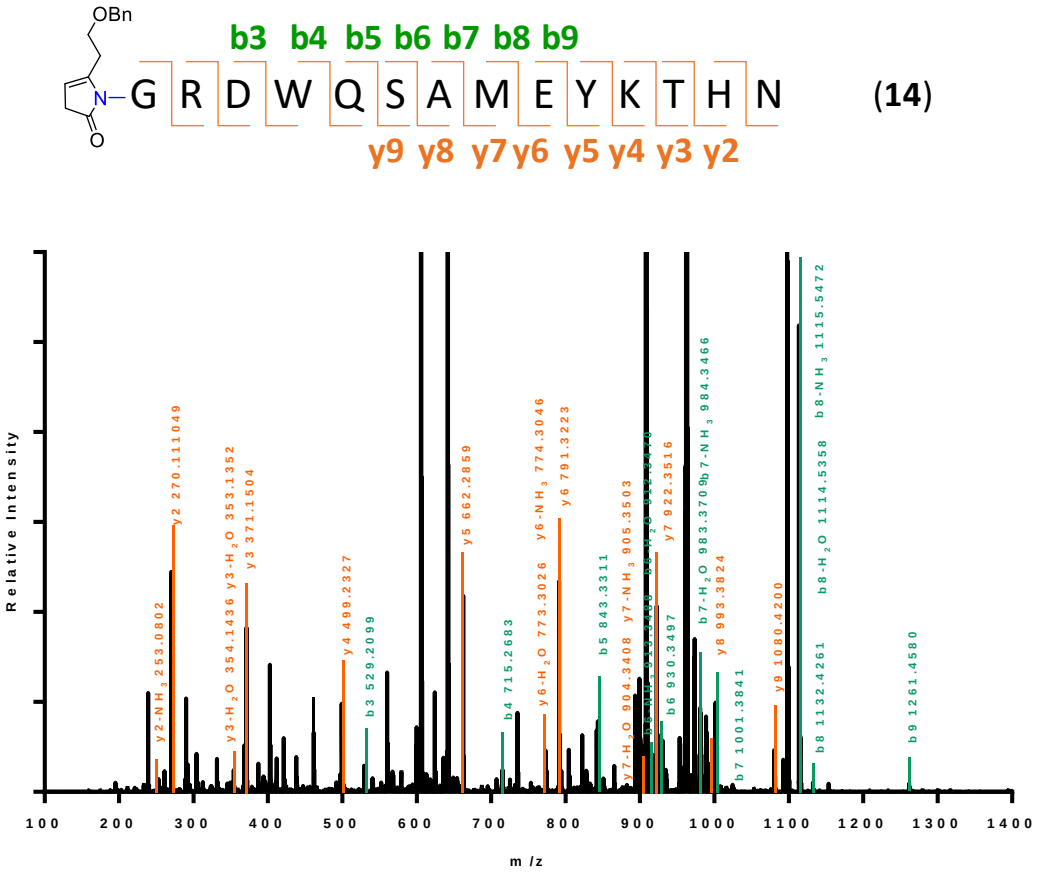


Figure S3. MS/MS spectra of peptide 14 which has one pyrrolinone modification on the N-terminal α-amino group.

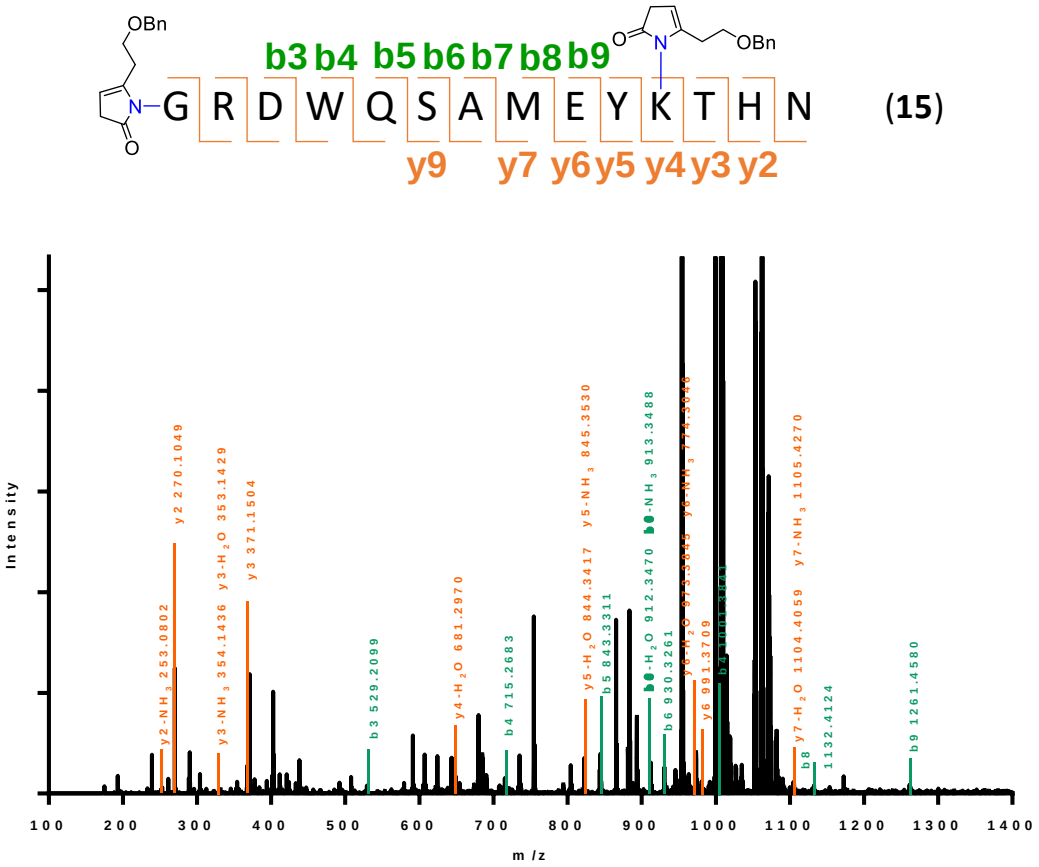


Figure S4. MS/MS spectra of peptide **15** which contains pyrrolinone modifications on both the N-terminal α-amino group and ε-amino group of Lys.

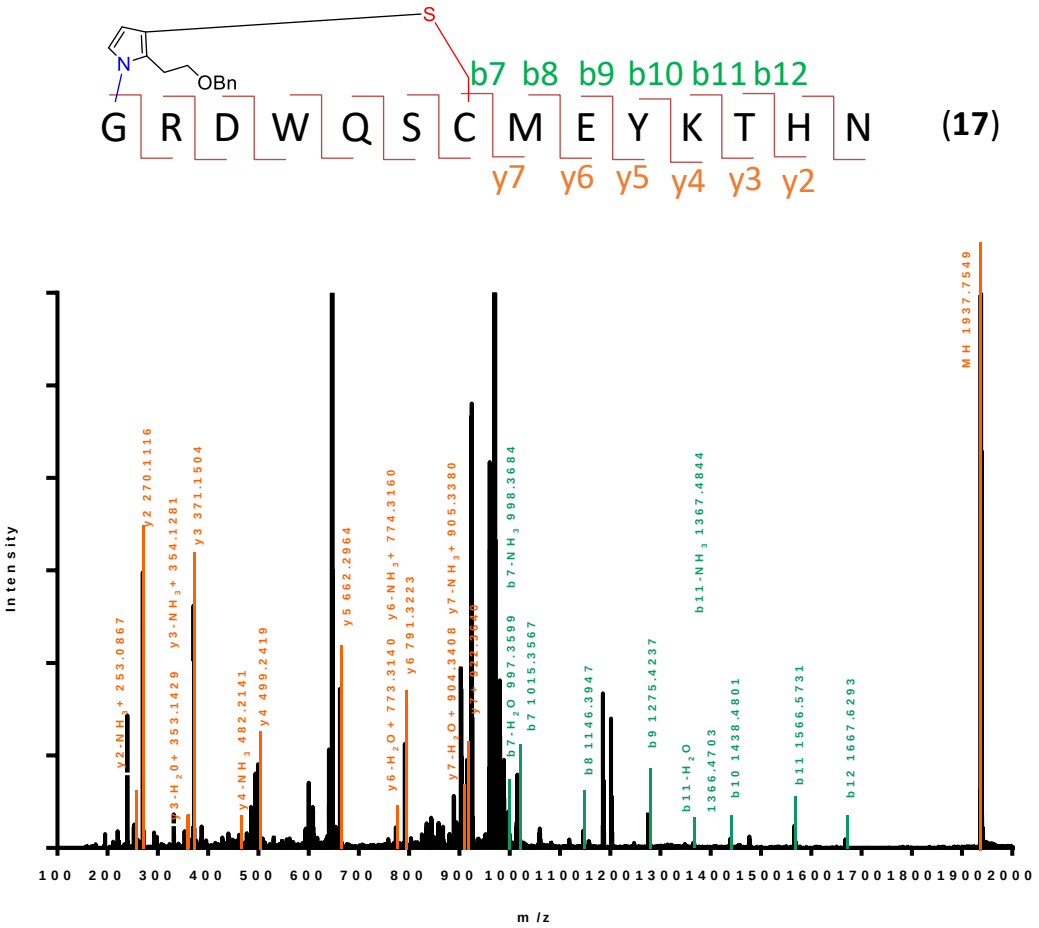


Figure S5. MS/MS spectra of peptide **17** which contains a cyclic pyrrole adduct between the N-terminal α-amino group and Cys residue.

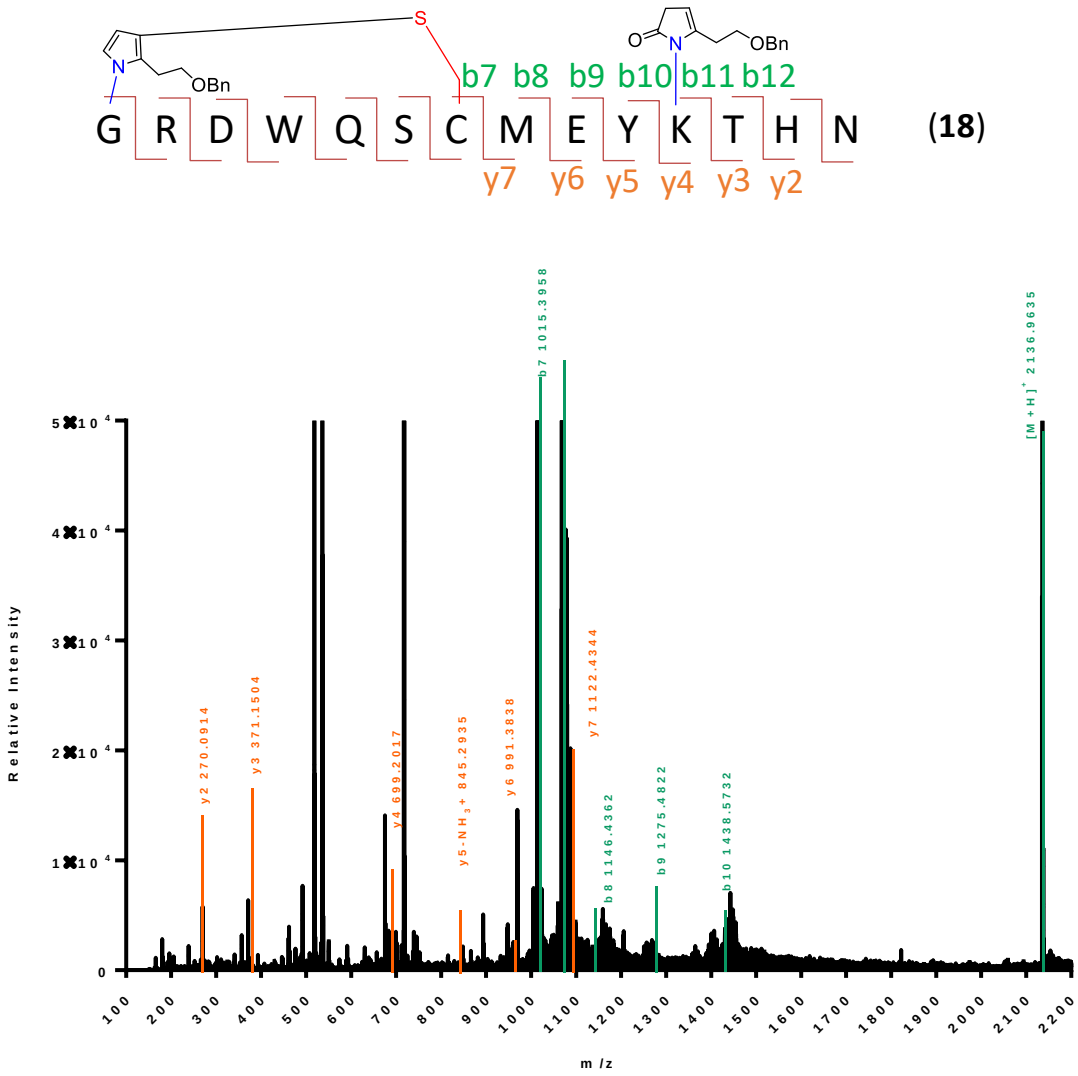


Figure S6. MS/MS spectra of peptide **18**.

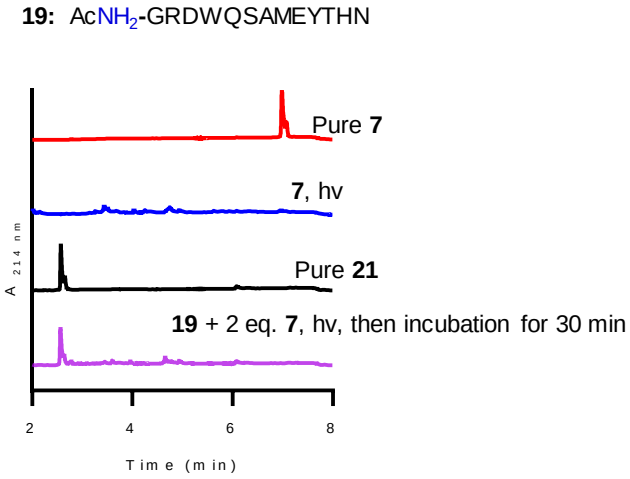


Figure S7. UPLC analysis of photo-reactions between **7** with peptide **19**.

Reaction conditions: To a solution of peptide **19** (final 10 μM) in phosphate buffer (40 mM, pH 7.4) was added 2 eq. of **7** (dissolved in 1:1 mixture of MeCN and H_2O). The mixture was irradiated (365 nm, 7.0 mW cm^{-2}) for 7 min, followed by incubation at 37 ºC. An aliquot was taken at 30 min and analyzed by UPLC-MS. UPLC conditions: HSS C18 column (300 Å, 1.7 μm, 2.1 mm X 100 mm), flow 0.4 mL/min at 35 ºC; Gradient: 0-6 min, B: 5-85%. A: H_2O; B: MeCN.

(A)

601-F1: (FAM)-ATCGATGTATATATCTGACACGTGCCTGGA

601-F2: pGACTAGGGAGTAATCCCCTTGGCGGTTAAAACGCG

601-F4: pAGCTGTCTACGACCAATTGAGCGGCCTCGGCACCGGGATTCTGAT

601-T1: CTCCCTAGTCTCCAGGCACG

601-T2: CGCTGTCCCCCGCGTTTTAA

601-T3: GTAGACAGCTCTAGCACCGC

601-R1: ATCAGAATCCCGGTGCCGAGGCCGCTCAATTGGTC

601-R2: pGTAGACAGCTCTAGCACCGCTAAAACGCAC

601-R3: pGTACGCGCTGTCCCCCGCGTTTTAACCGCCAAGGG

601-R4: pGATTACTCCCTAGTCTCCAGGCACGTGTCAGATATATACATCGAT

601-T4: AGCTGTCTACGACCAATTGA

601-T5: CAGCGCGTACGTGCGTTTAA

601-T6: GGGAGTAATCCCCTTGGCGG

601-F3-a: pGGGGACAGCGCGTACGTG

601-F3-b1: pCGTTTXAGCGGTGCTAG

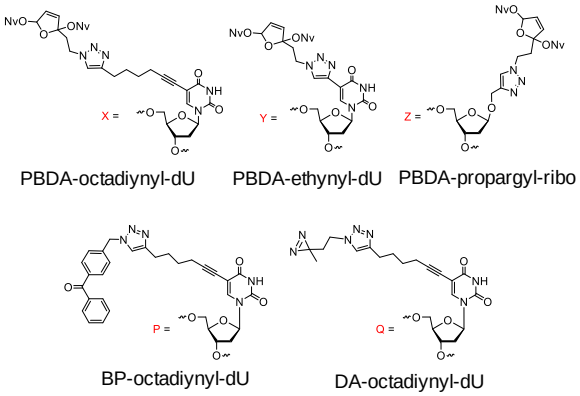
601-F3-b2: pCGTTTYAGCGGTGCTAG

601-F3-b3: pCGTTTZAGCGGTGCTAG

601-F3-b4: pCGTTTPAGCGGTGCTAG

601-F3-b5: pCGTTTQAGCGGTGCTAG

601-T7: CGCTAAAACGCACGTACGCG



(B)

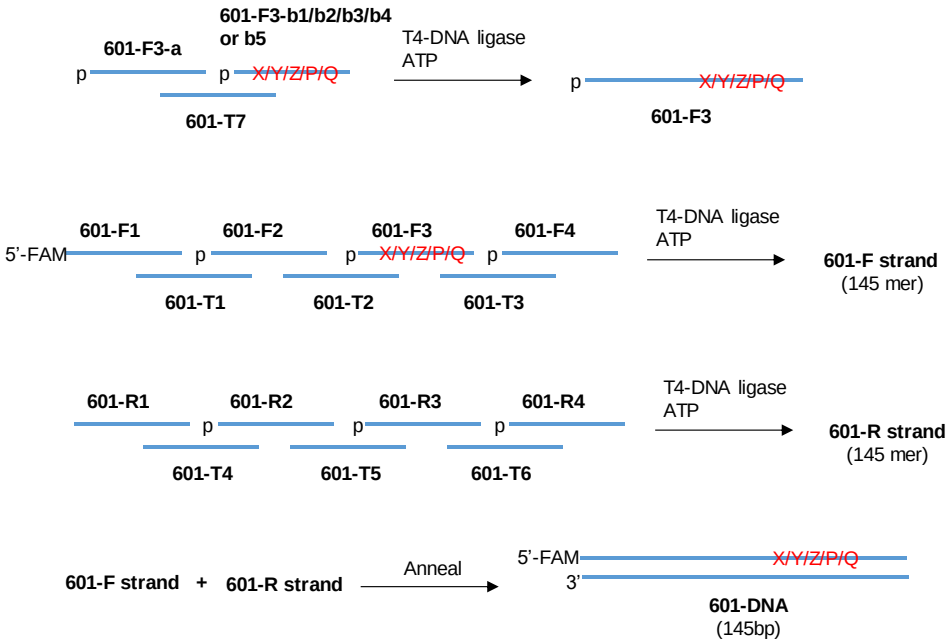


Figure S8. Preparation of 601-DNA. (A) DNA sequence. (B) Strategy of preparation of 601-DNA. "p" denotes 5'-phosphate group.

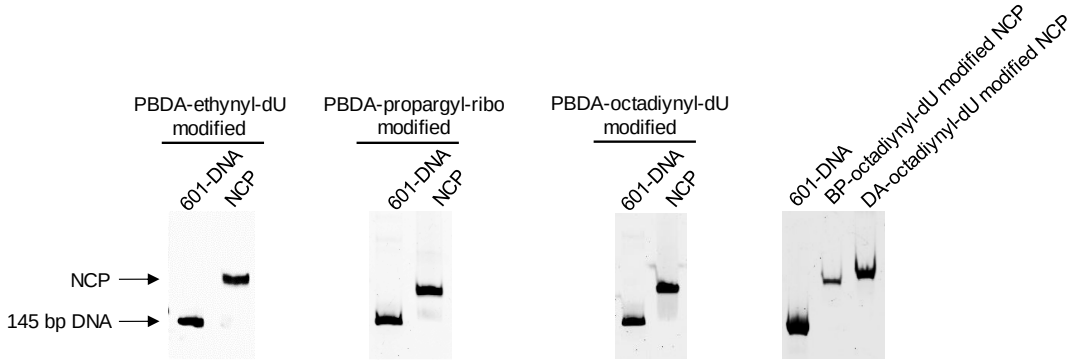


Figure S9. Non-denaturing PAGE (5%) analysis of the reconstitution yields of nucleosome core particles. One of the 5′ end of the 601 DNA is labeled with FAM. The picture was obtained by fluorescence (FAM) imaging of the gel.

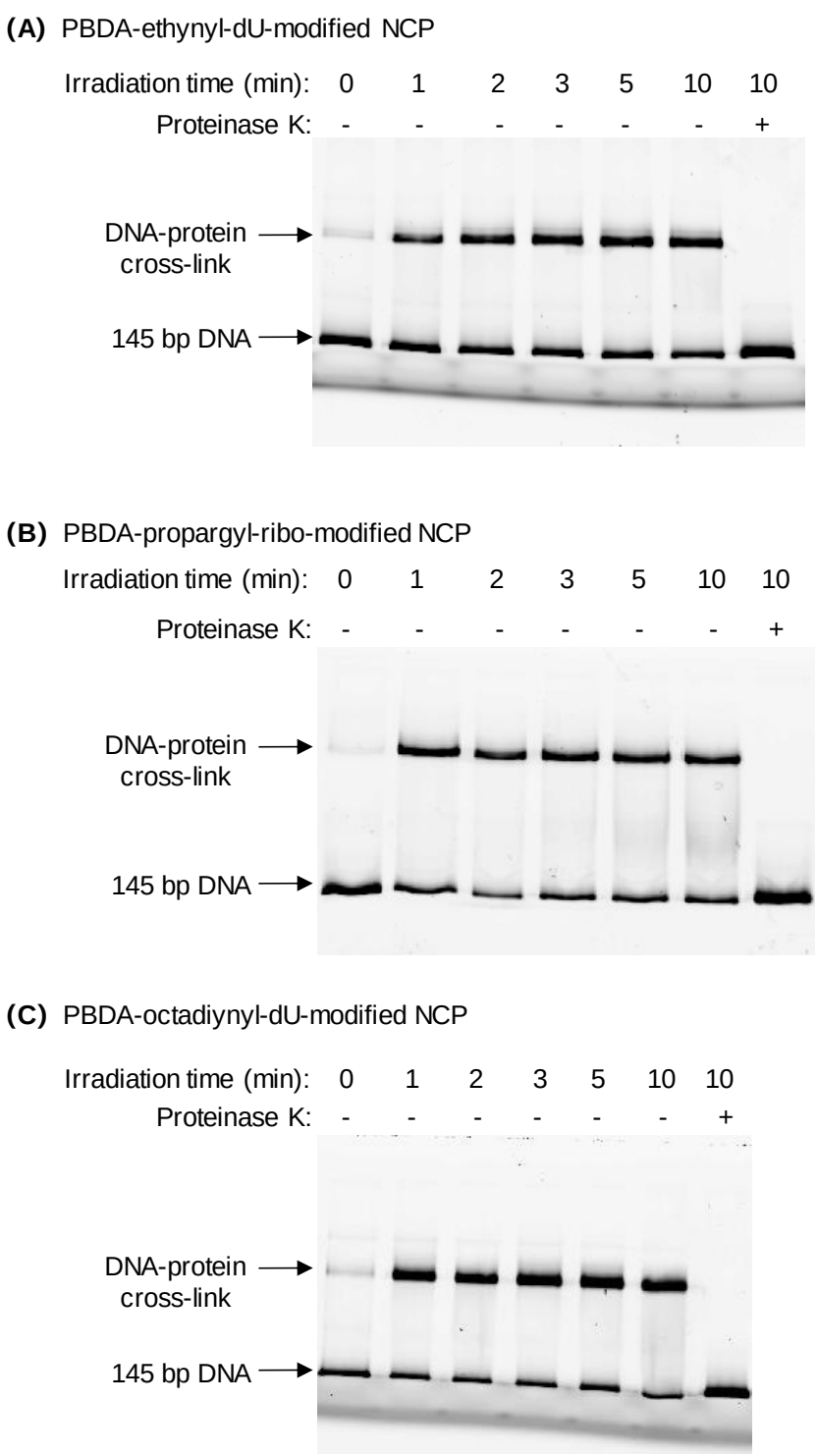


Figure S10. SDS-PAGE (10%) analysis of photo-crosslink of PBDA-modified DNA with histones within nucleosome core particles (NCPs). One of the 5′ end of the dsDNA is labeled with FAM. The picture was obtained by fluorescence (FAM) imaging of the gel.

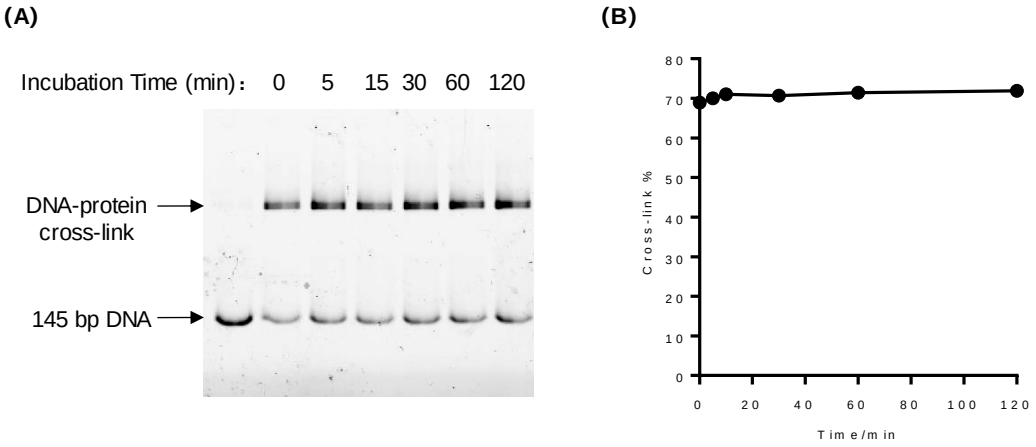


Figure S11. Stability of DNA-histone cross-link formed by photo-irradiation of NCP containing a PBDA-octadiynyl-dU modification. PBDA-octadiynyl-dU modified NCP was irradiated for 7 min, followed by incubation at 37 ºC. Aliquots were removed at different time points and analyzed by SDS-PAGE. (A) SDS-PAGE analysis of the stability of the DNA-histone cross-link. The gel was visualized by fluorescence (FAM) imaging. The yields of DNA-protein cross-link were plotted against incubation time and the obtained graph is shown in (B).

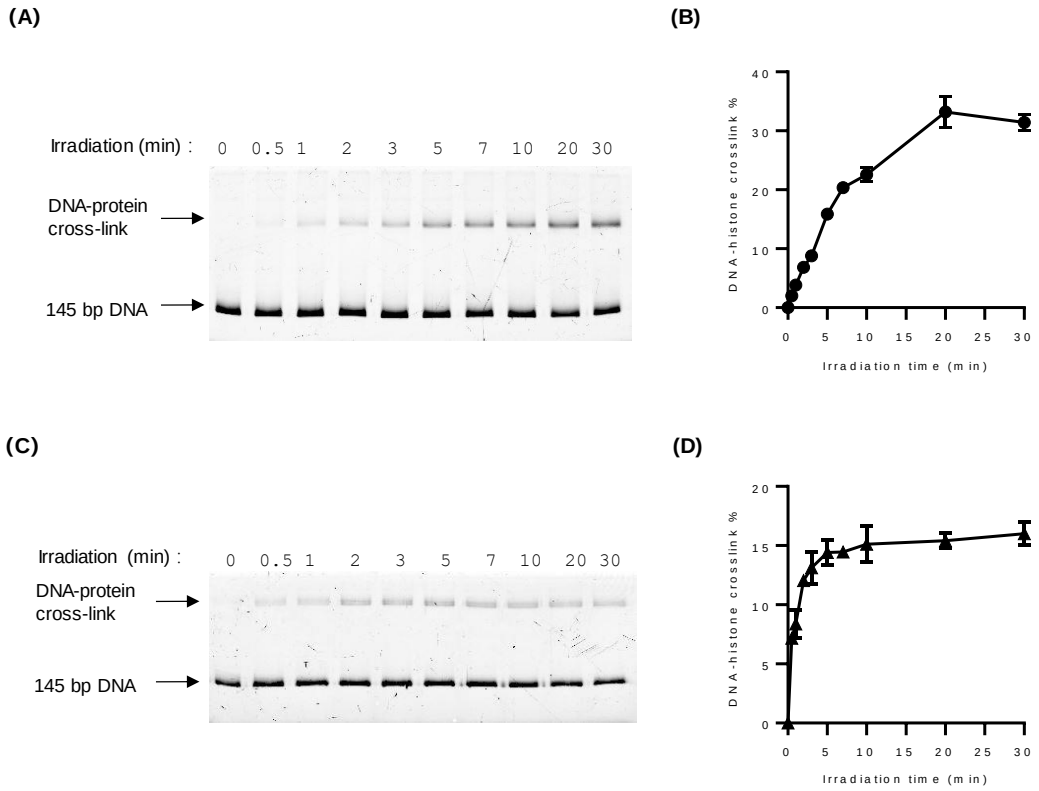


Figure S12. Analysis of photo-crosslink of BP-modified DNA and DA-modified DNA with histones within nucleosome core particles (NCPs). (A) SDS-PAGE (10%) analysis of photo-crosslink of BP-modified DNA with histones within NCPs. The quantification of the results is shown in (B). (C) SDS-PAGE (10%) analysis of photo-crosslink of DA-modified DNA with histones within NCPs. The quantification of the results is shown in (D). One of the 5′ end of the dsDNA is labeled with FAM. The picture was obtained by fluorescence (FAM) imaging of the gel.

dsDNA **26**: 5'-FAM-CAGCTCTGTACGYGAGCAGTGGA-3'
3'-TGTCGAGACATGFACTCGTCACC-5'

F = 5fC; Y = PBDA-octadiynyl-dU

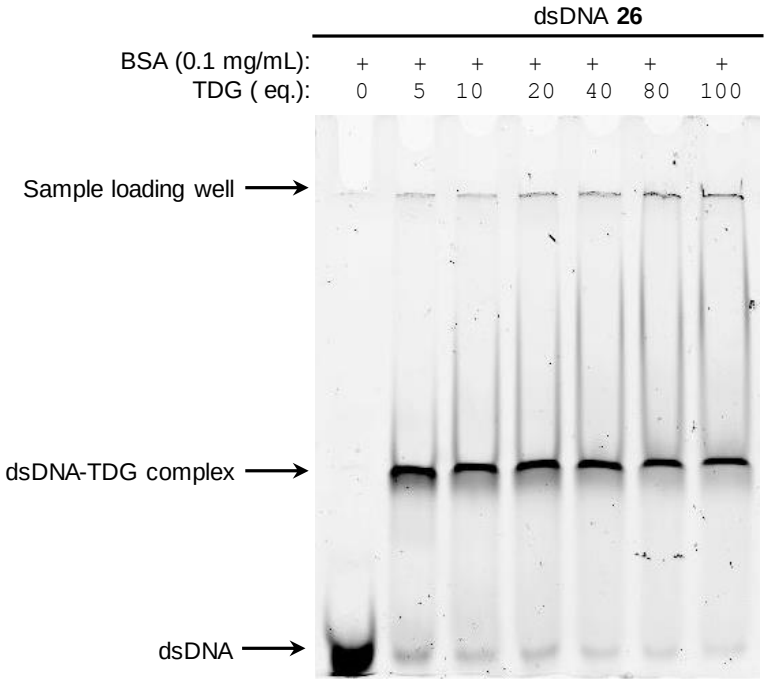


Figure S13. Gel shift assay of the binding of TDG to dsDNA **26**.
To a solution of dsDNA **26** (final 62.5 nM) in HEPES buffer (20 mM, pH 7.5, 100 mM NaCl, 0.2 mM EDTA, 0.1 mg/ml BSA) was added different equivalent of TDG. The mixture was incubated at 4ºC for 2.5 h. Aliquot was removed and analyzed by 5% non-denatured PAGE. The picture was obtained by fluorescence (FAM) imaging of the gel.

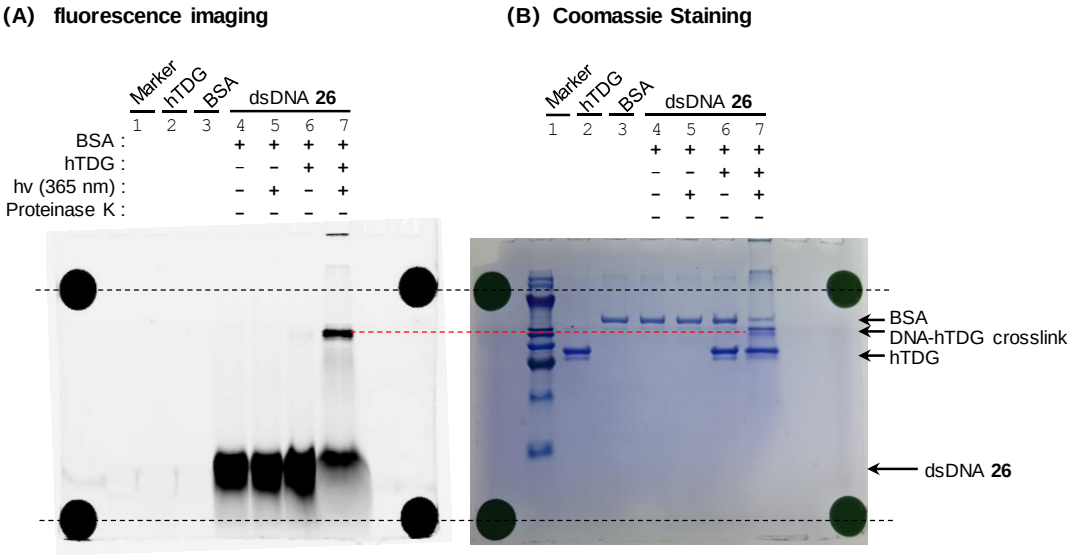


Figure S14. SDS-PAGE analysis of photo-crosslinking of dsDNA **26** with hTDG. The gel was firstly visualized by means of fluorescence (fluorescein, FAM) imaging (A), followed by Commassie Staining (B). Reaction conditions: dsDNA **26** (final 62.5 nM) and hTDG (final 625 nM) in HEPES buffer (20 mM, pH 7.5, 100 mM NaCl, 0.2 mM EDTA, 0.05 mg/ml BSA) was incubated at 4 ℃ for 2.5 h, and then at 37 ℃ for 10 min. After photo-irradiation for 7 min, the reaction was further incubated 37 ℃ for 10 min. Aliquot was removed and analyzed by Tricine-SDS-PAGE (15.5%).

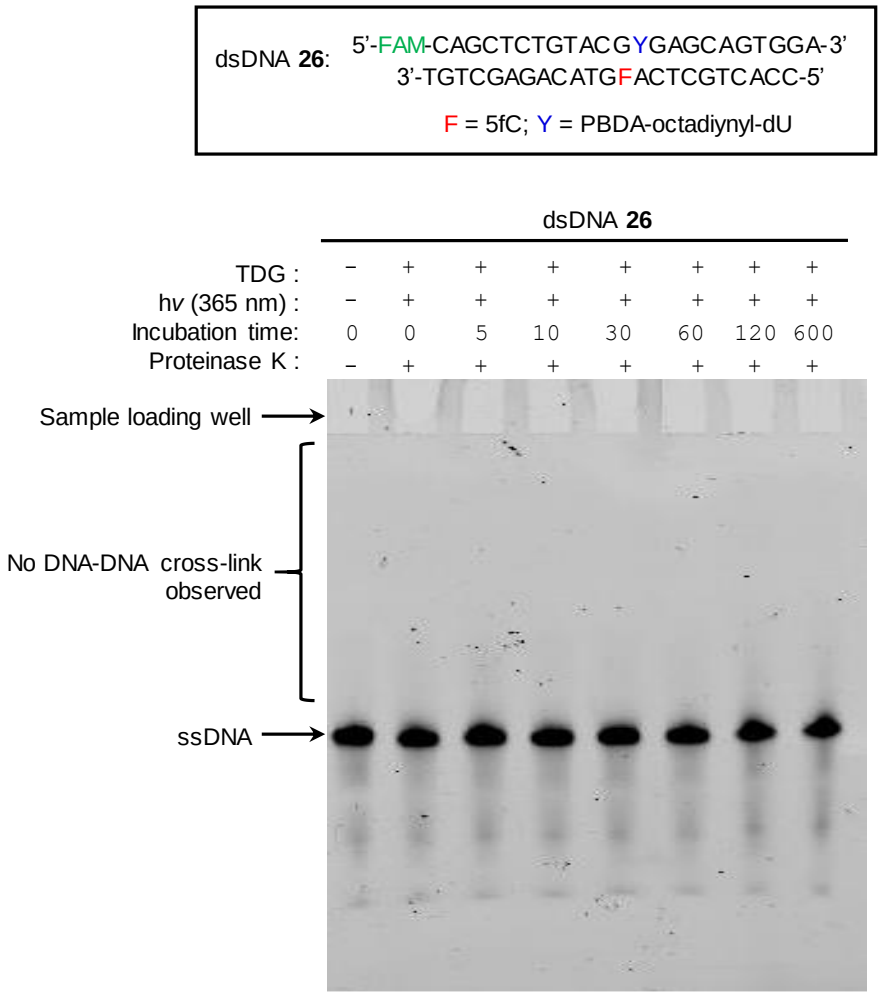


Figure S15. 8% denaturing PAGE analysis of the reaction between dsDNA **26** and TDG upon photo-irradiation. dsDNA **28** was pre-incubated with 10 eq. of TDG followed by photo-irradiation (365 nm) for 7 min. The reaction mixture was incubated at 37 ºC. Aliquots were removed at different time points and treated with proteinase K, followed by 8% denaturing PAGE analysis. The picture was obtained by fluorescence (FAM) imaging of the gel.

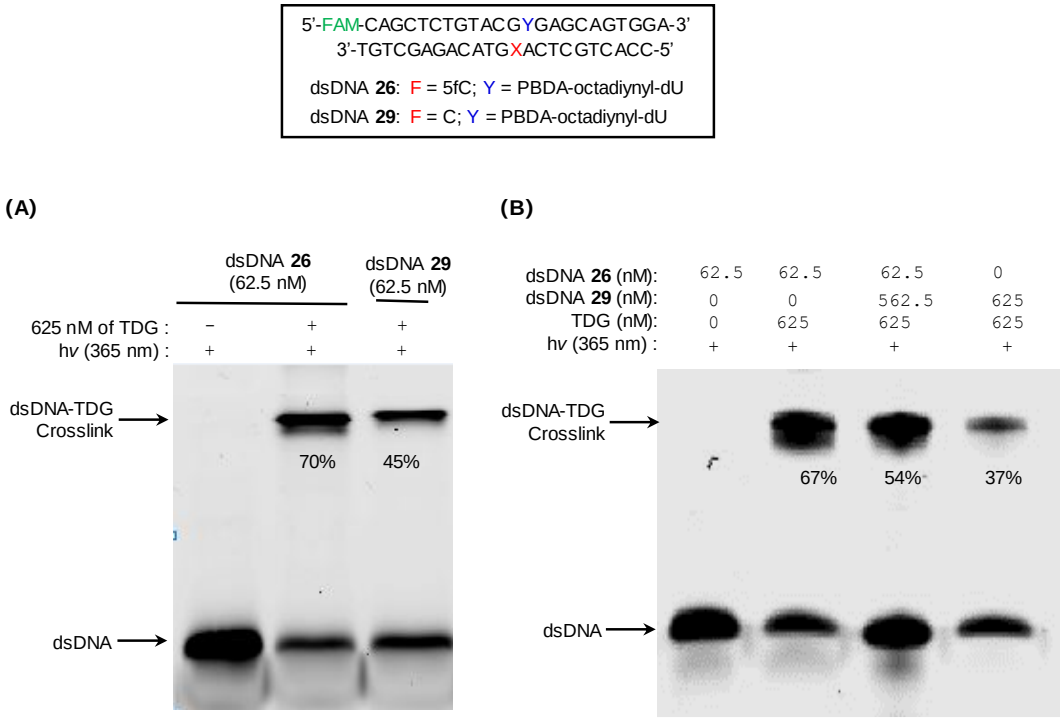


Figure S16. Tricine-SDS-PAGE analysis of the photo-crosslink of dsDNA **26** and dsDNA **29** with TDG. A mixture of dsDNA (concentration is indicated on the graphs) and TDG (625 nM) in HEPES buffer (20 mM, pH 7.5, 100 mM NaCl, 0.2 mM EDTA, 0.1 mg/ml BSA) was incubated at 4 ℃ for 2.5 h, and then at 37 ℃ for 10 min. After photo-irradiation for 7 min, the reaction was further incubated 37 ℃ for 10 min. Aliquot was removed and analyzed by Tricine-SDS-PAGE (15.5%). The picture was obtained by fluorescence (FAM) imaging of the gel.

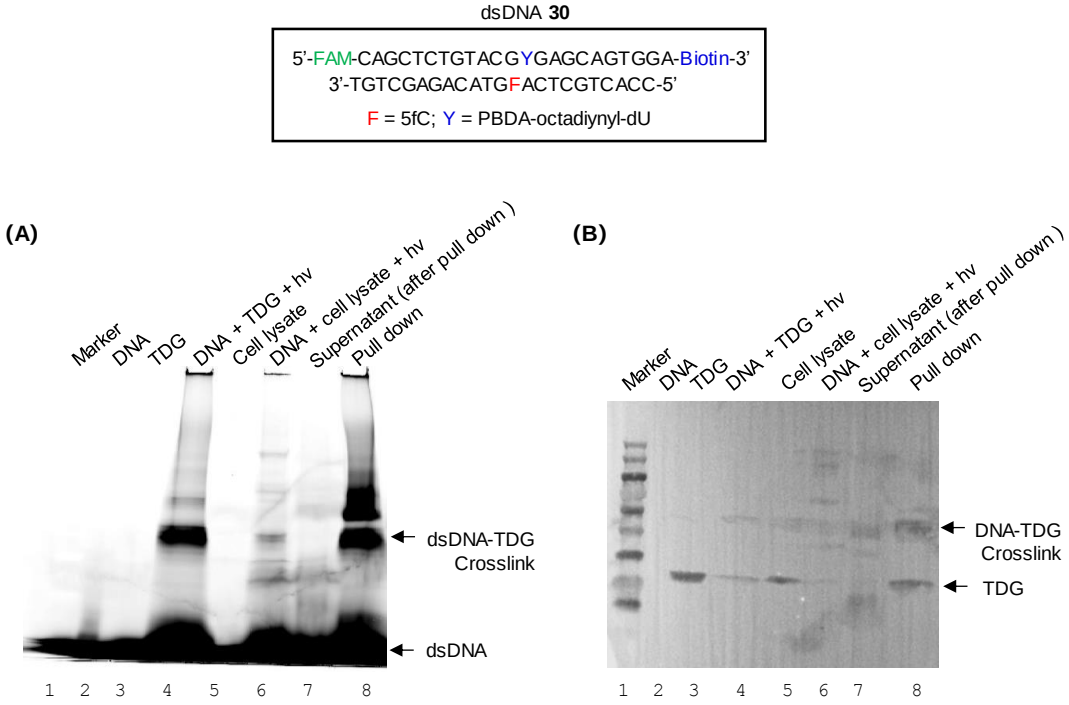
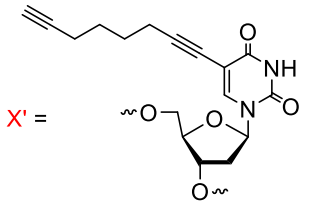


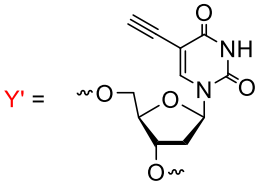
Figure S17. Pull down of TDG from cell lysate using PBDA-modified dsDNA **30**. The gel was visualized by fluorescence (FAM) imaging (A), followed by anti-TDG western blot (B).

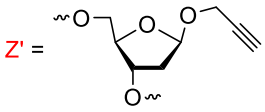
Supplementary tables

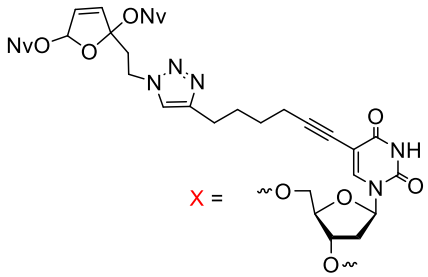
Table S1. Sequence and mass of synthesized oligonucleotides.

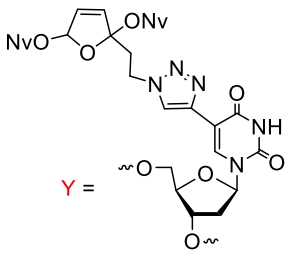
Oligo name	Sequence(5′-3′)	$[MH]^+$ cal.	$[MH]^+$ Found
601-F3-b1′	pCGTTTX′AGCGGTGCTAG	5402.5	5402.8
601-F3-b2′	pCGTTTY′AGCGGTGCTAG	5321.4	5321.6
601-F3-b3′	pCGTTTZ′AGCGGTGCTAG	5242.4	5244.0
601-F3-b1	pCGTTTXAGCGGTGCTAG	5963.3	5964.6
601-F3-b2	pCGTTTYAGCGGTGCTAG	5882.6	5883.5
601-F3-b3	pCGTTTZAGCGGTGCTAG	5857.2	5856.7
601-F3-b4	pCGTTTPAGCGGTGCTAG	5639.3	5639.6
601-F3-b5	pCGTTTQAGCGGTGCTAG	5527.4	5526.9
601-F3-X	pGGGGACAGCGCGTACGTGCGTTTXAGCGGTGCTAG	11552.2	11552.2
601-F3-Y	pGGGGACAGCGCGTACGTGCGTTTYAGCGGTGCTAG	11471.2	11471.5
601-F3-Z	pGGGGACAGCGCGTACGTGCGTTTZAGCGGTGCTAG	11632.2	11633.6
dsDNA **26**-a′	5′-FAM-CAGCTCTGTACGX′GAGCAGTGGA	7732.2	7731.8
dsDNA **28**-a′	5′-FAM-CAGCX′CTGTACGTGAGCAGTGGA	7732.2	7731.8
dsDNA **27**-a′	5′-FAM-CAGCTCTGTACGTGAGCAGX′GGA	7732.2	7731.9
dsDNA **26**-a	5′-FAM-CAGCTCTGTACXTGAGCAGTGGA	8293.4	8294.2
dsDNA **28**-a	5′-FAM-CAGCXCTGTACTTGAGCAGTGGA	8293.4	8293.3
dsDNA **27**-a	5′-FAM-CAGCTCTGTACTTGAGCAGXGGA	8293.4	8293.5
dsDNA **26**-b	CCACTGCTCA5fCGTACAGAGCTGT	7012.6	7012.8
dsDNA **30**-a′	5′-FAM-CAGCTCTGTACGX′GAGCAGTGGA-Biotin	8301.8	8303.0
dsDNA **30**-a	5′-FAM-CAGCTCTGTACGXGAGCAGTGGA-Biotin	8863.9	8863.1

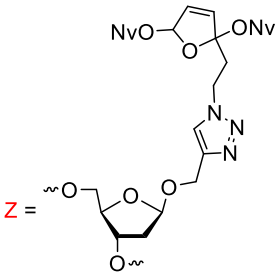


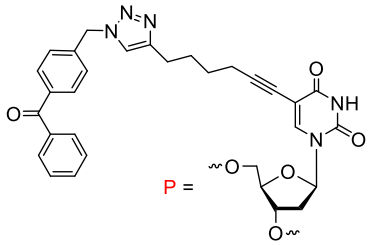


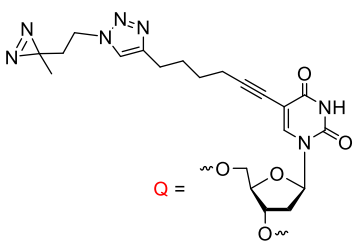


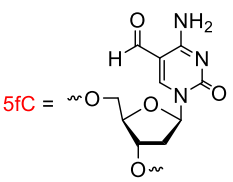




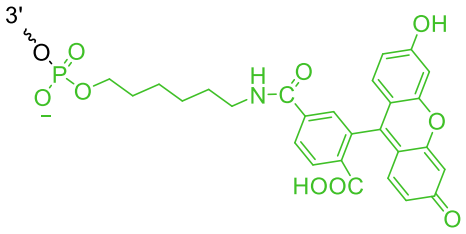








5'-FAM =



3'-Biotin =

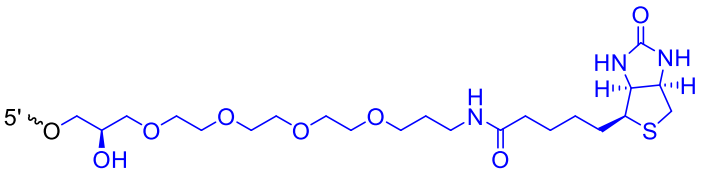


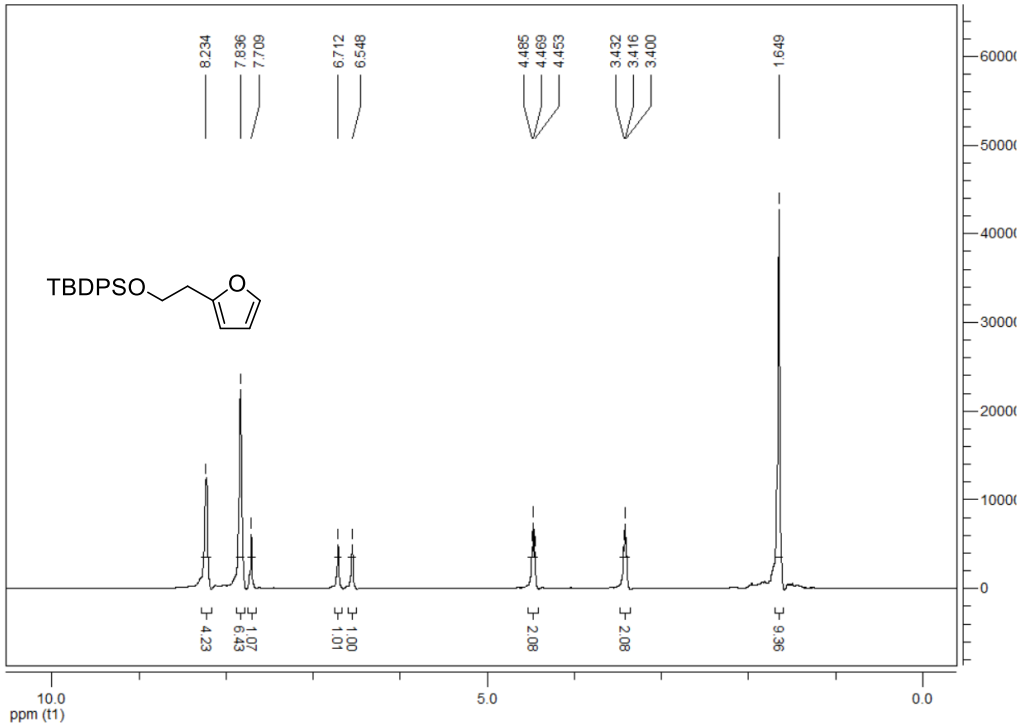
Table S2. Pull-down proteins by dsDNA **30** from cell lysate[a]

Organism	Rank	Protein name	Peptide count	Unit peptide count	Cover Percent
Human	1	Desmoplakin	29	28	8.29%
	2	Thymine DNA glycosylase	40	12	22.41%
	3	Hornerin	13	12	4.56%
	4	Serum albumin	10	7	10.76%
	5	Junction plakoglobin	8	7	11.37%
	6	Desmoglein-1	6	6	5.24%
	7	alpha skeletal muscle	6	6	21.57%
	8	Annexin (Fragment)	4	4	18.75%
	9	Caspase-14	3	3	11.98%
	10	Dermcidin	3	3	27.27%
E. Coli.	1	50S ribosomal protein L1	10	9	39.74%
	2	30S ribosomal protein S3	11	8	21.68%
	3	Elongation factor Tu (Fragment)	8	8	26.89%
	4	30S ribosomal protein S4	8	7	30.58%
	5	Succinate dehydrogenase iron-sulfur subunit	7	7	26.47%
	6	Succinate--CoA ligase [ADP-forming] subunit alpha	7	6	19.72%
	7	Phosphoribosylaminoimidazole-succinocarboxamide synthase	6	6	20.68%
	8	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase	6	6	21.60%
	9	Glycerol kinase	6	6	11.75%
	10	30S ribosomal protein S2	7	5	13.40%

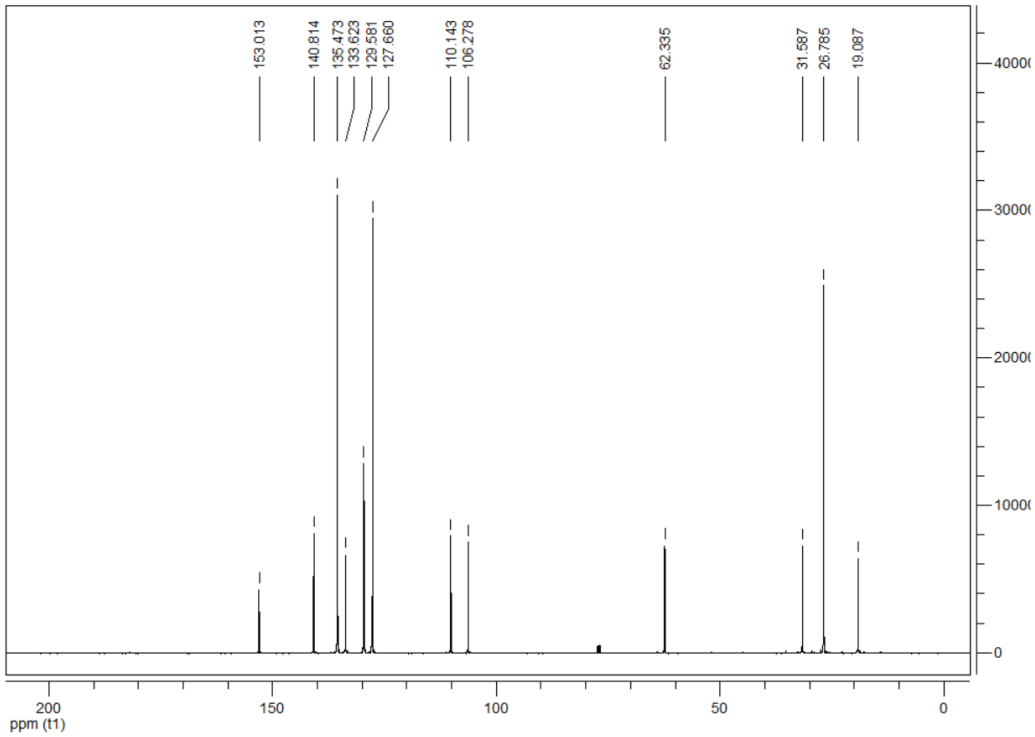
[a] *Escherichia coli* cells that overexpress human TDG were harvested, lysed, and centrifuged; and the supernatant was collected for a pull-down assay. The pull down sample was analysed by 15% SDS PAGE. The intense band (1 × 1.5 cm) was excised and subjected to in-gel trypsin digestion followed by UPLC-MS/MS analysis. Tandem mass spectra were searched against uniprot_*Escherichia_coli* database and uniprot_*Homo_sapiens* database. The proteins identified are ranked by the unit peptide count and only Top 10 are shown in the table.

NMR spectra

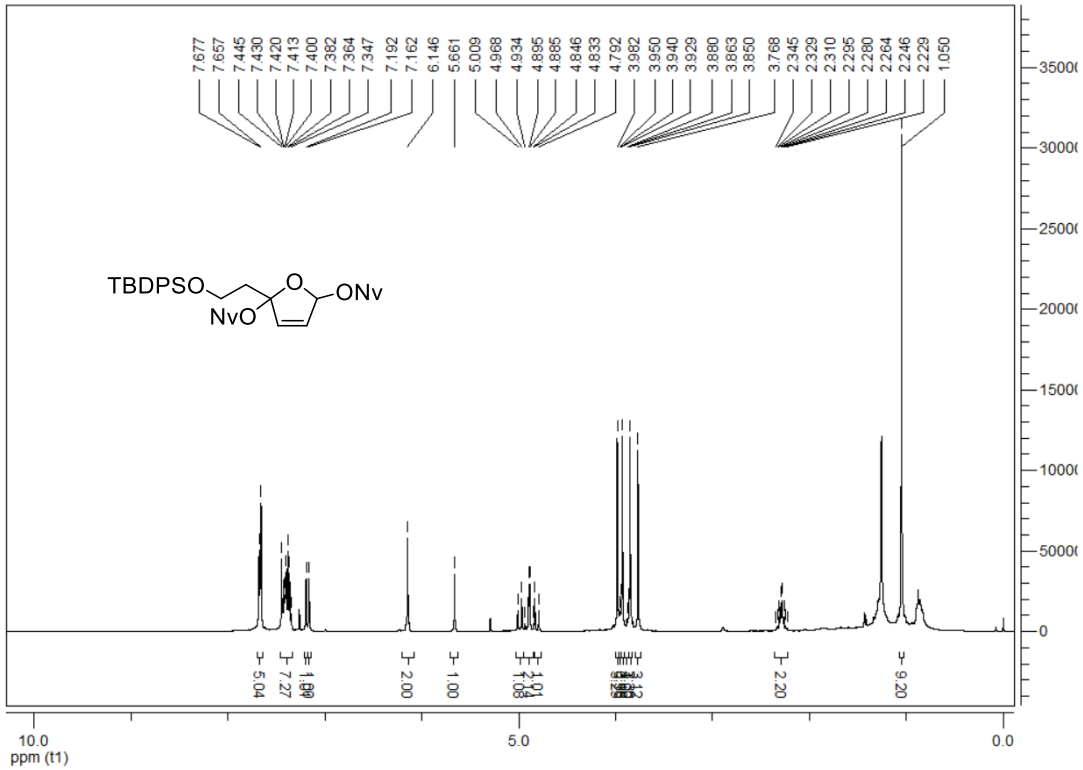
^{1}H NMR of 3:



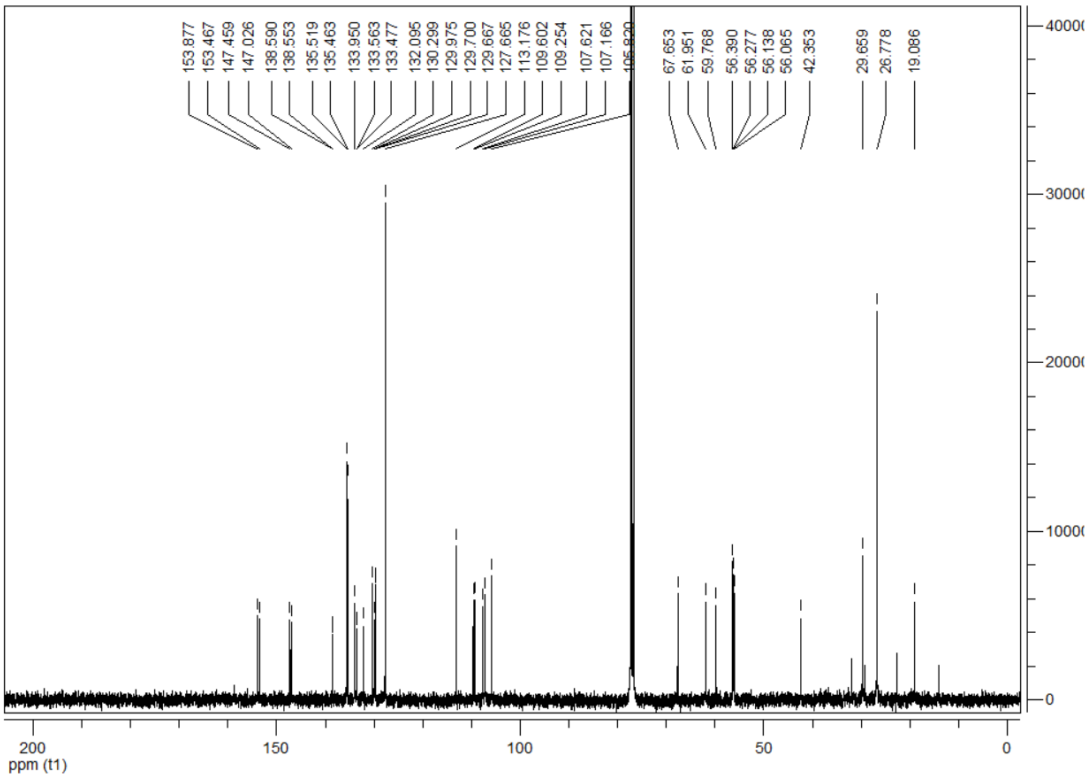
^{13}C NMR of 3:



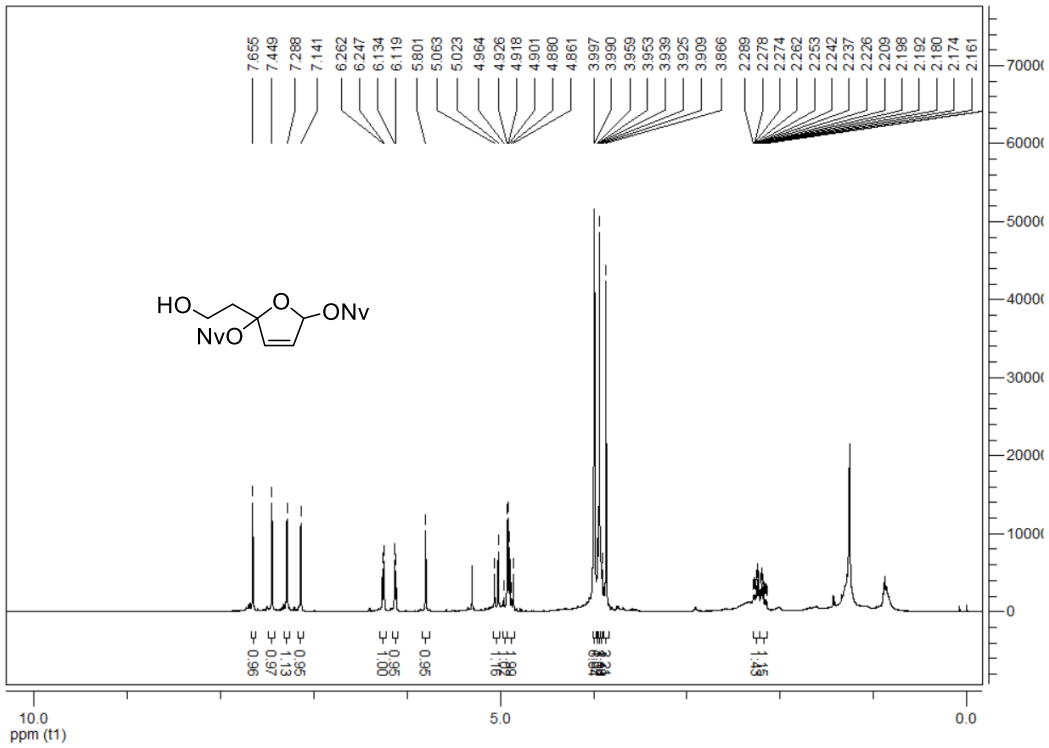
^{1}H NMR of 4 (Isomer II):



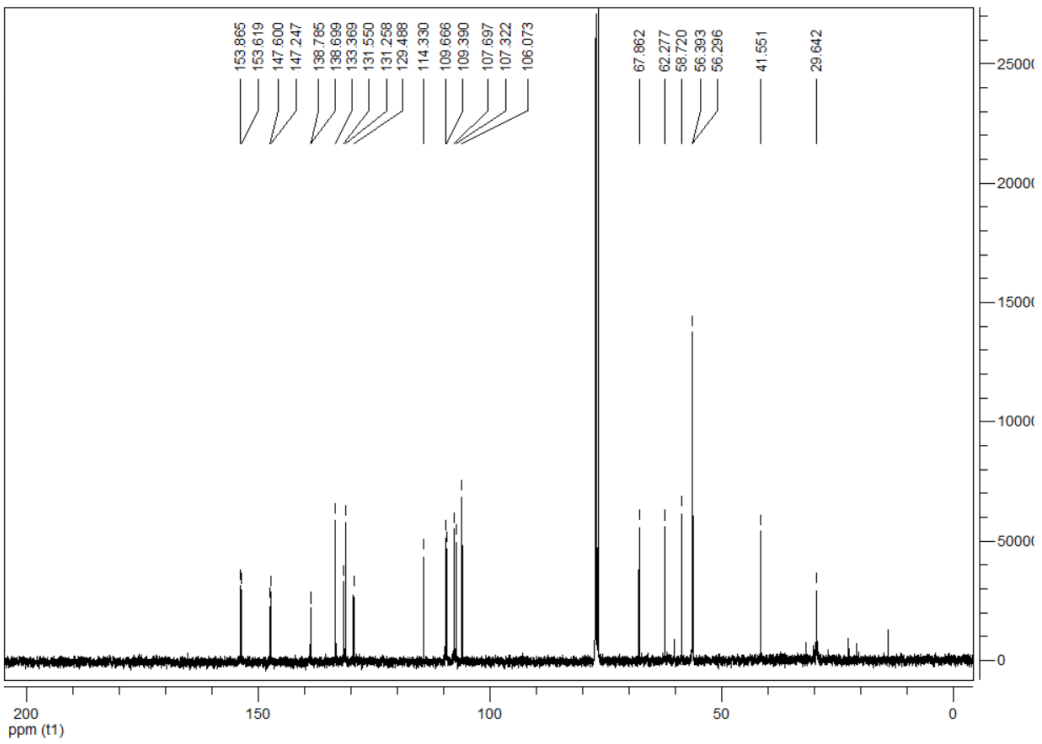
^{13}C NMR of 4 (Isomer II):



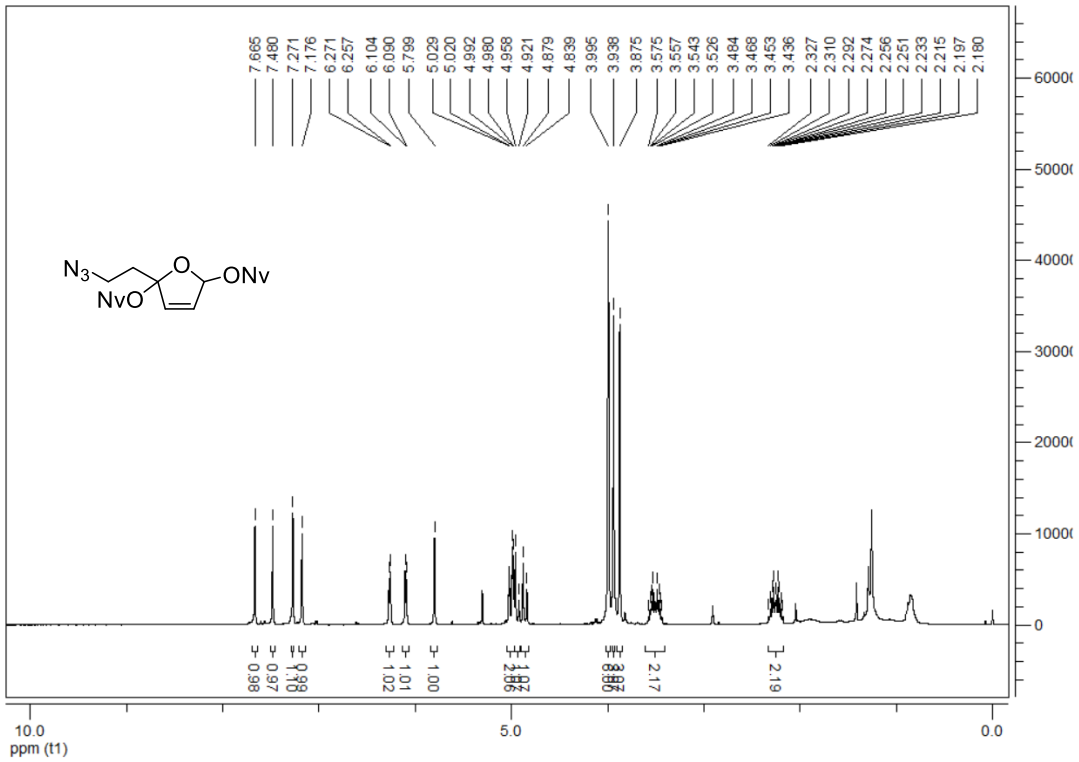
^{1}H NMR of 5:



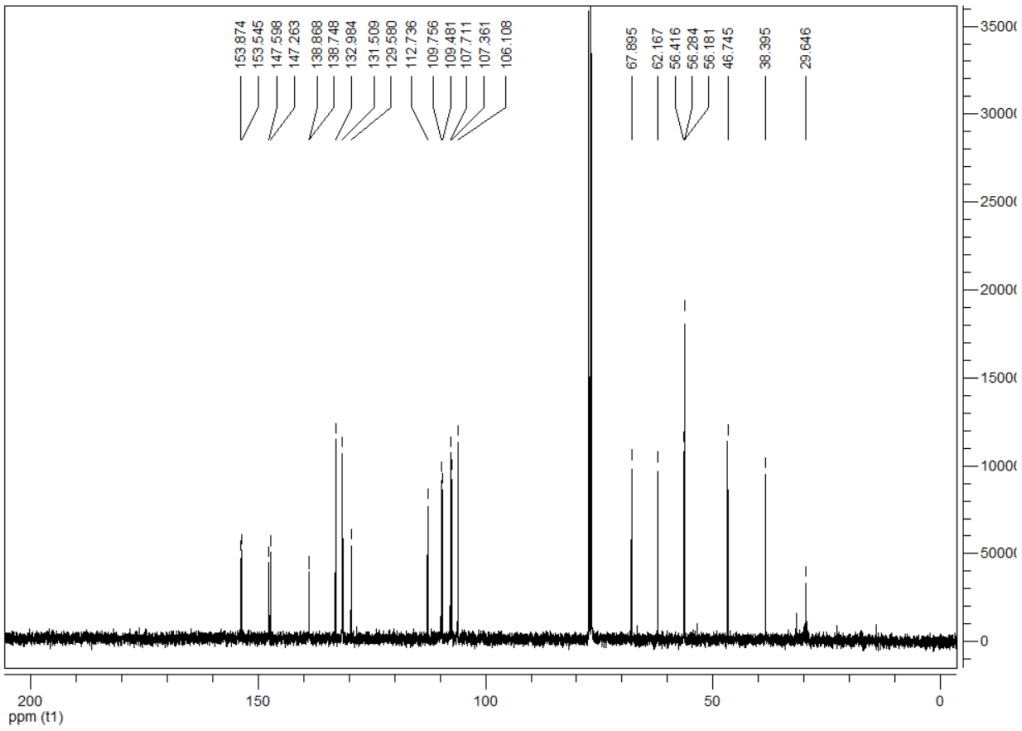
^{13}C NMR of 5:



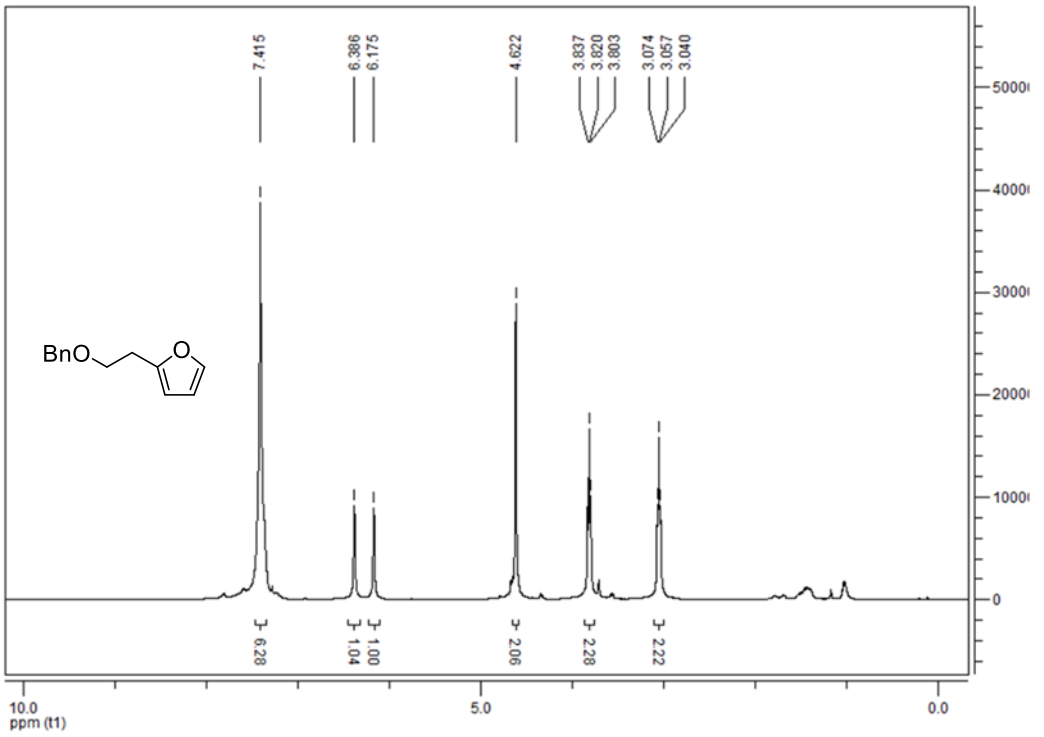
^{1}H NMR of 1:



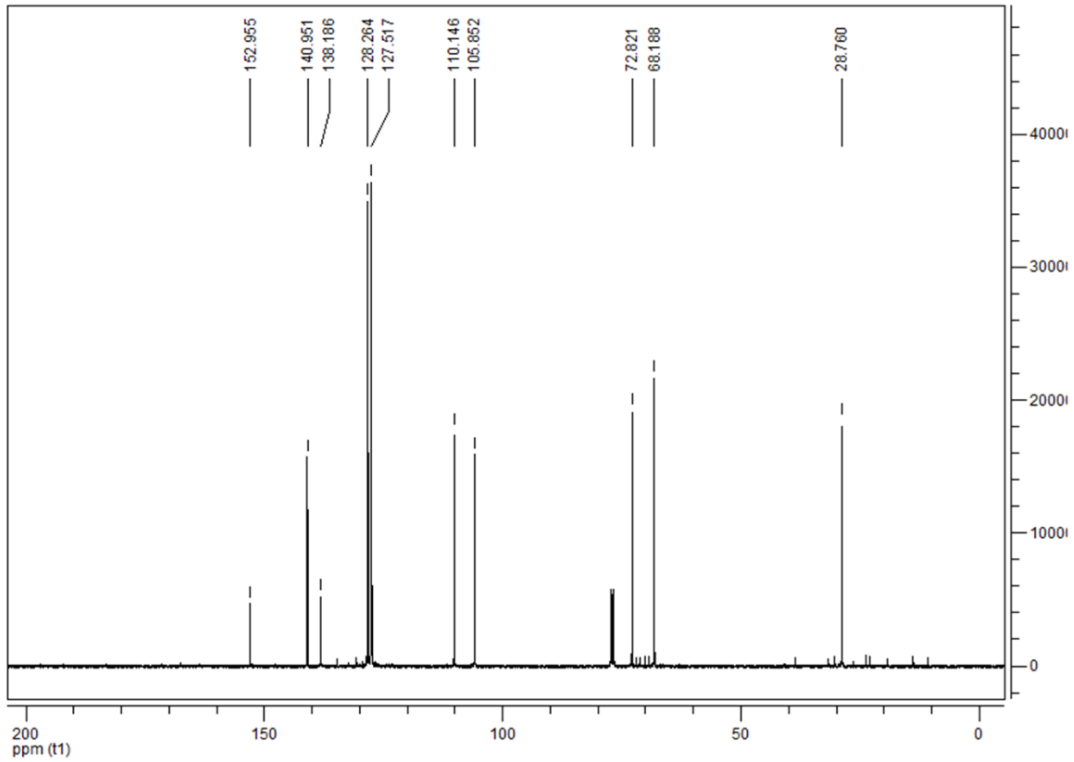
^{13}C NMR of 1:



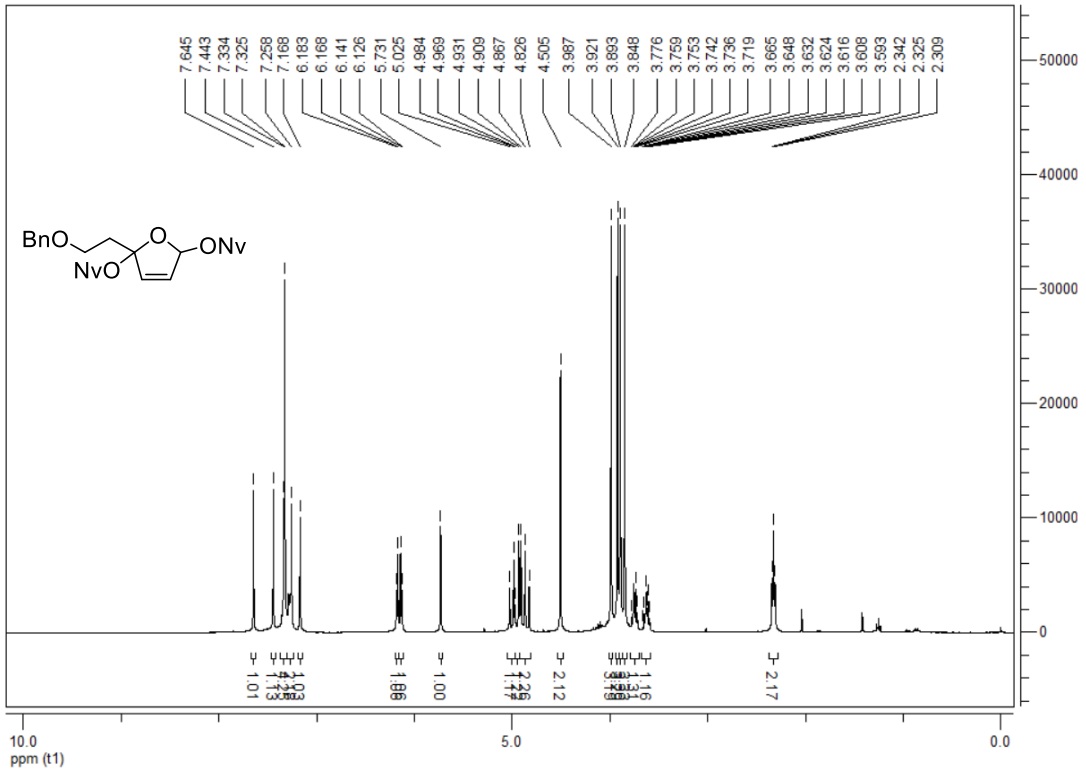
^{1}H NMR of 6:



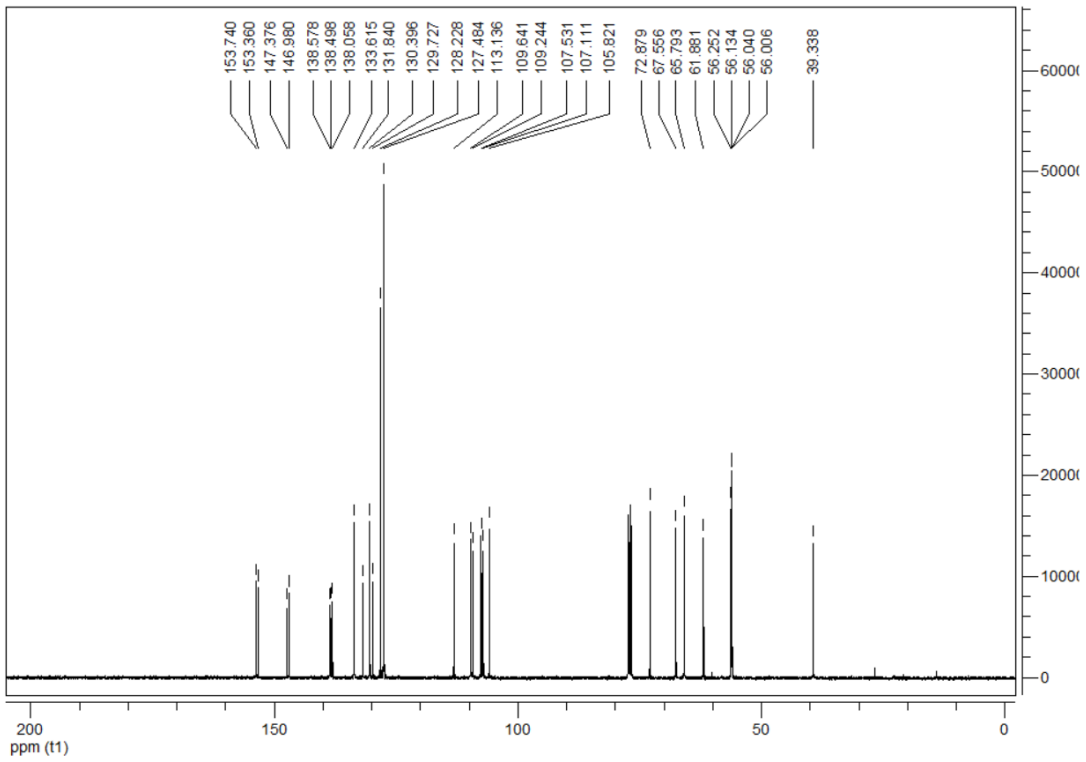
^{13}C NMR of 6:



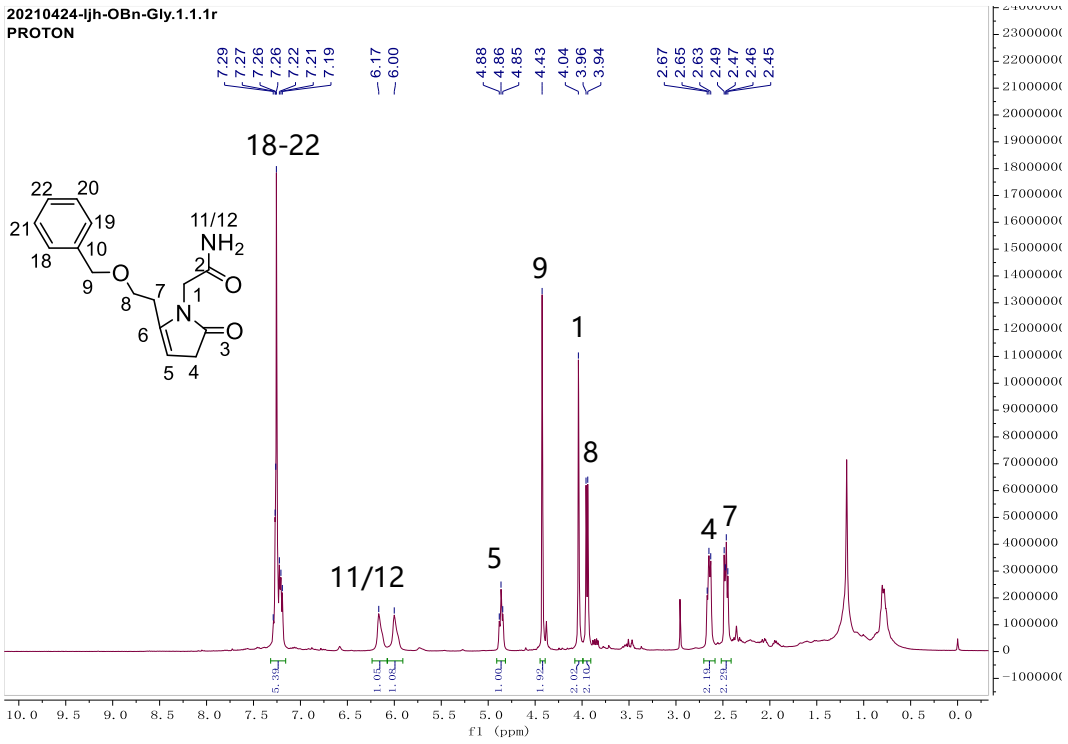
^{1}H NMR of 7:



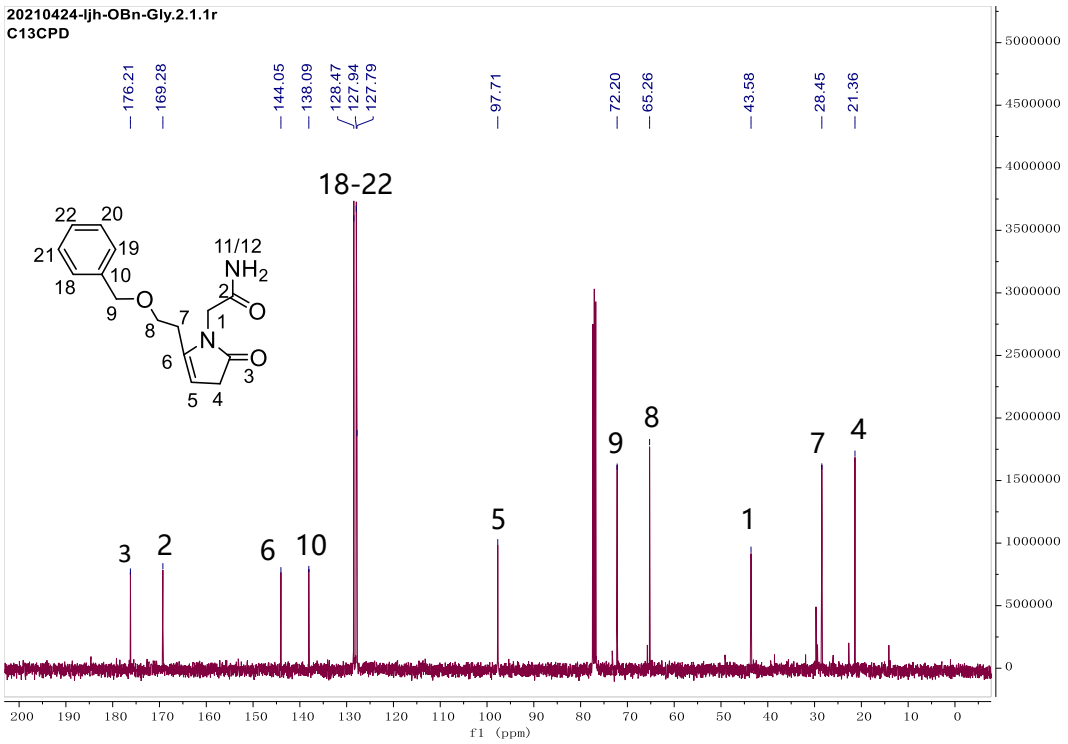
^{13}C NMR of 7:



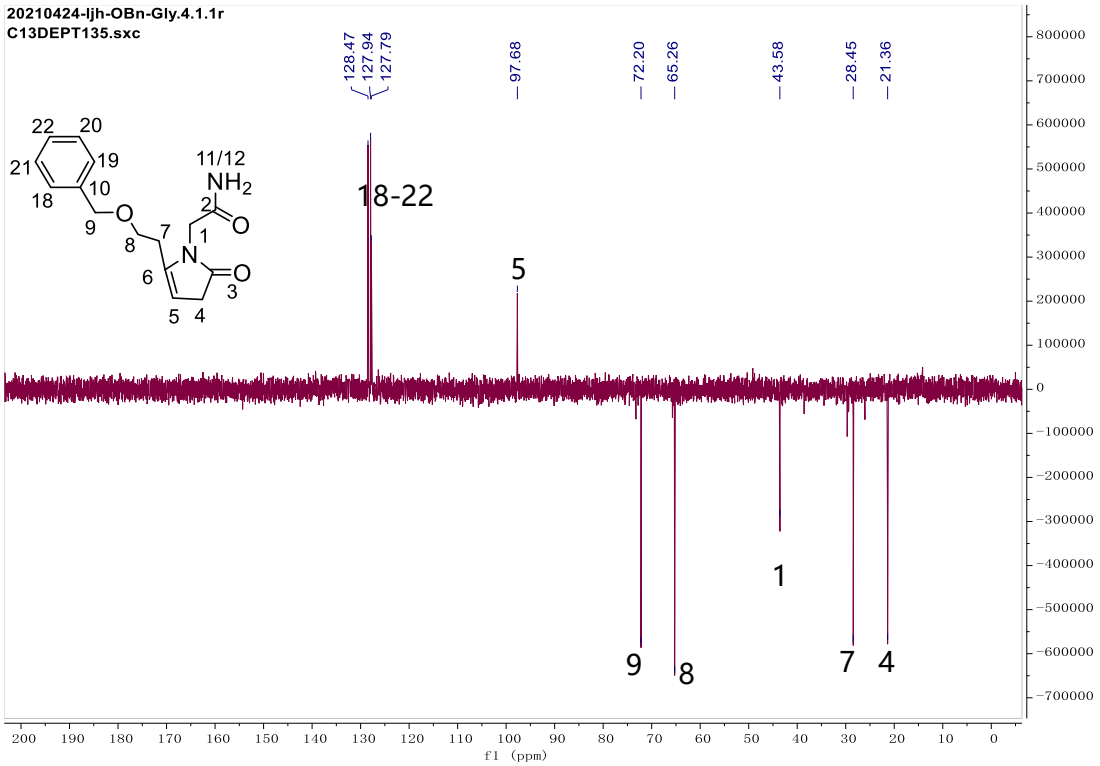
^{1}H NMR of 10:



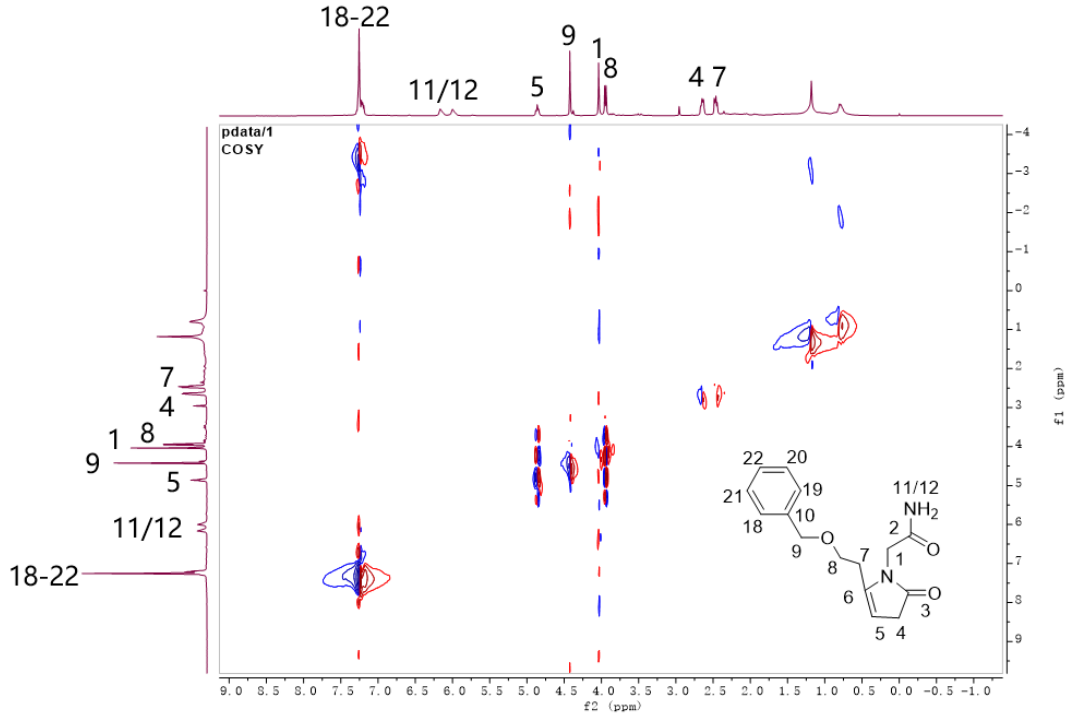
^{13}C NMR of 10:



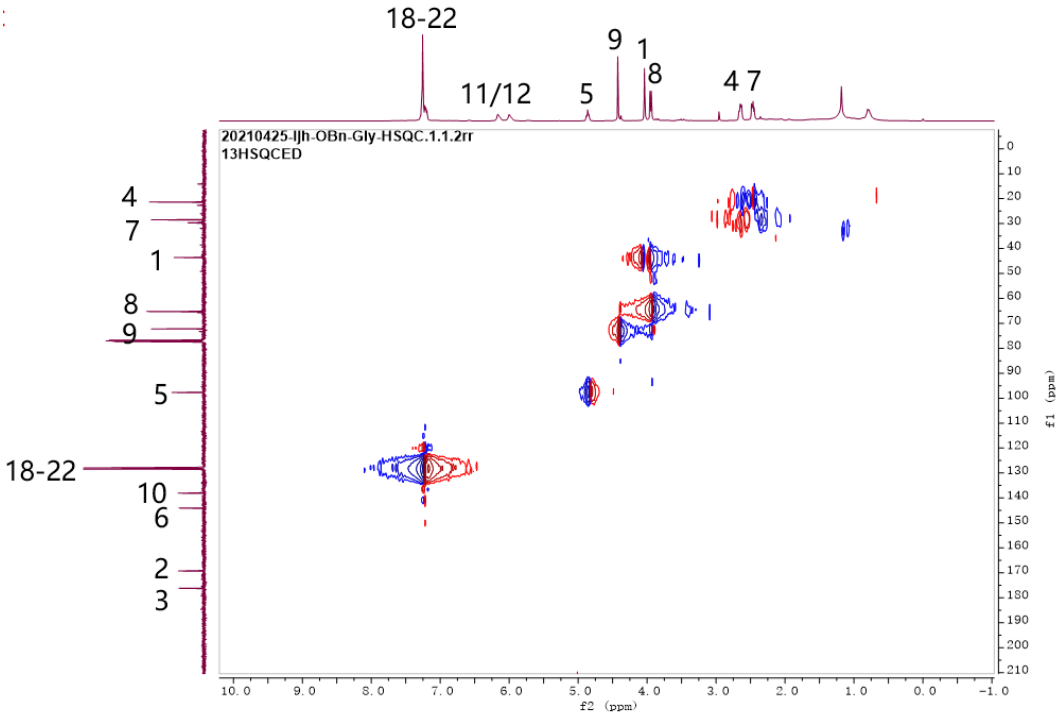
DEPT 135 of 10:



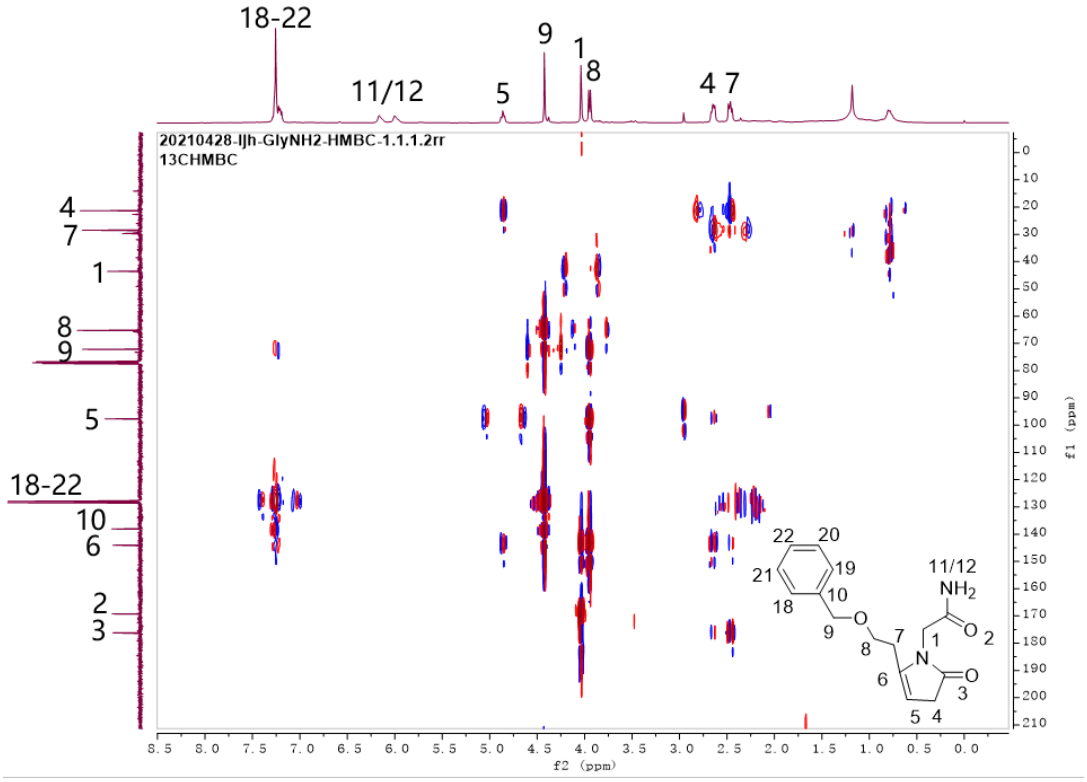
COSY of 10:



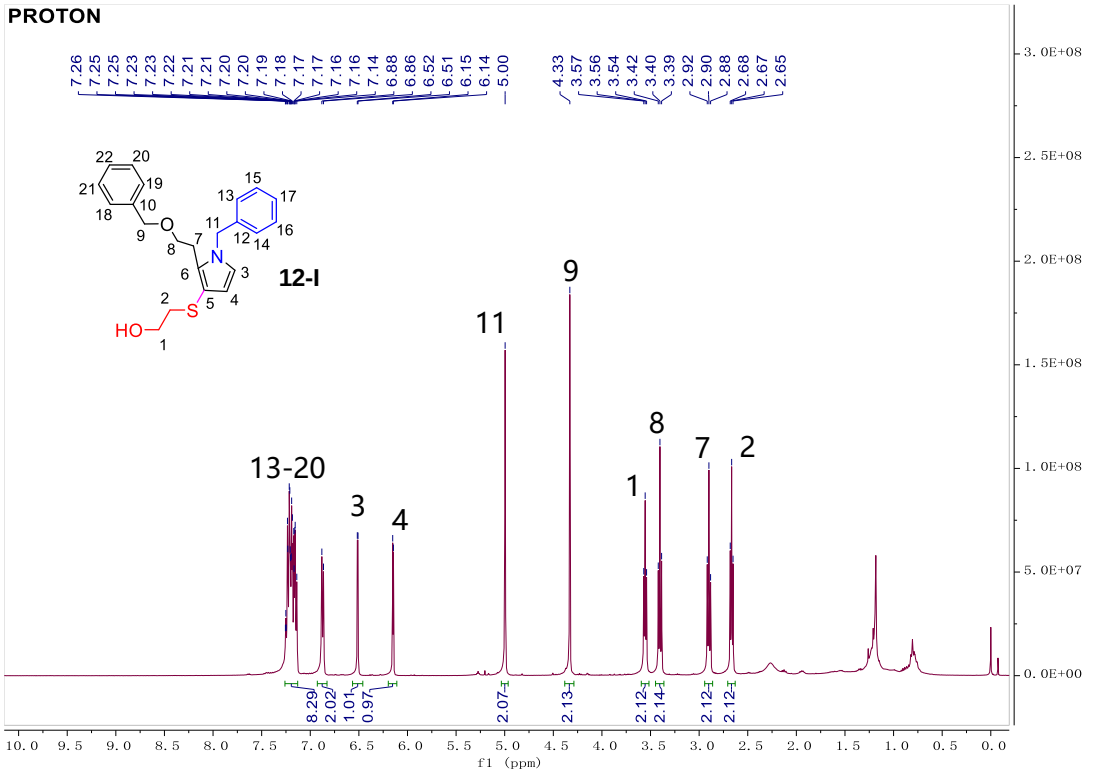
HSQC of 10:



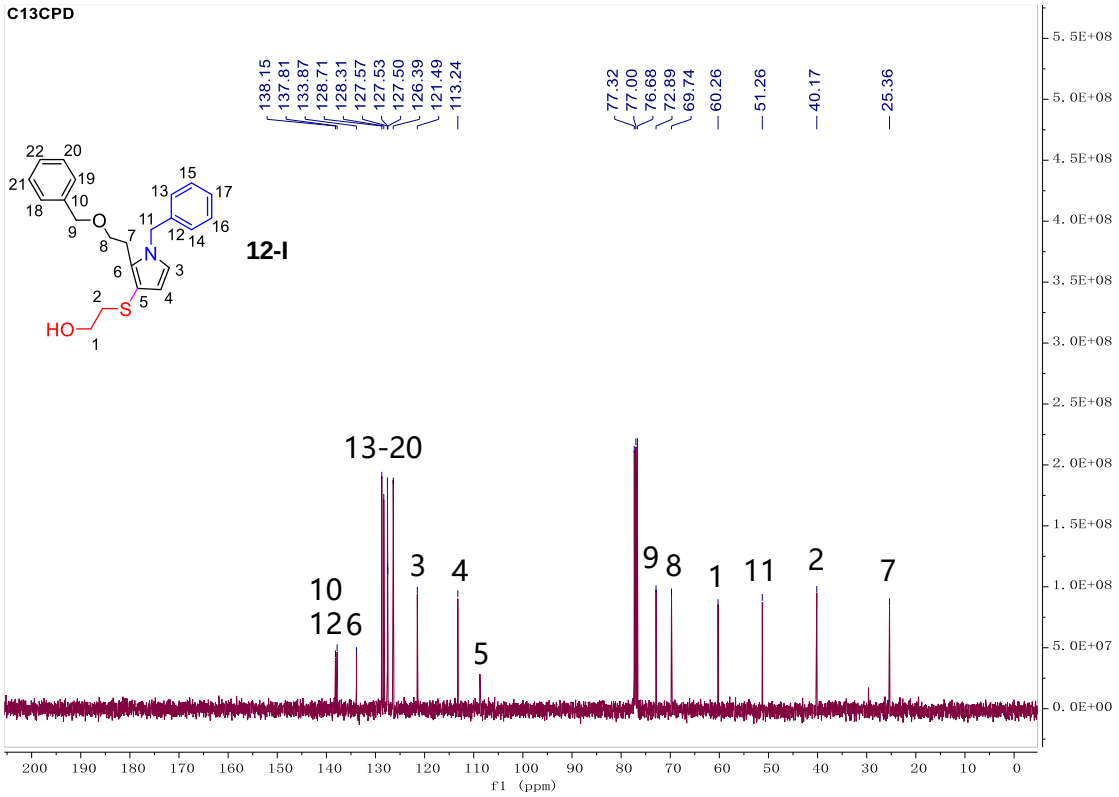
HMBC of 10:



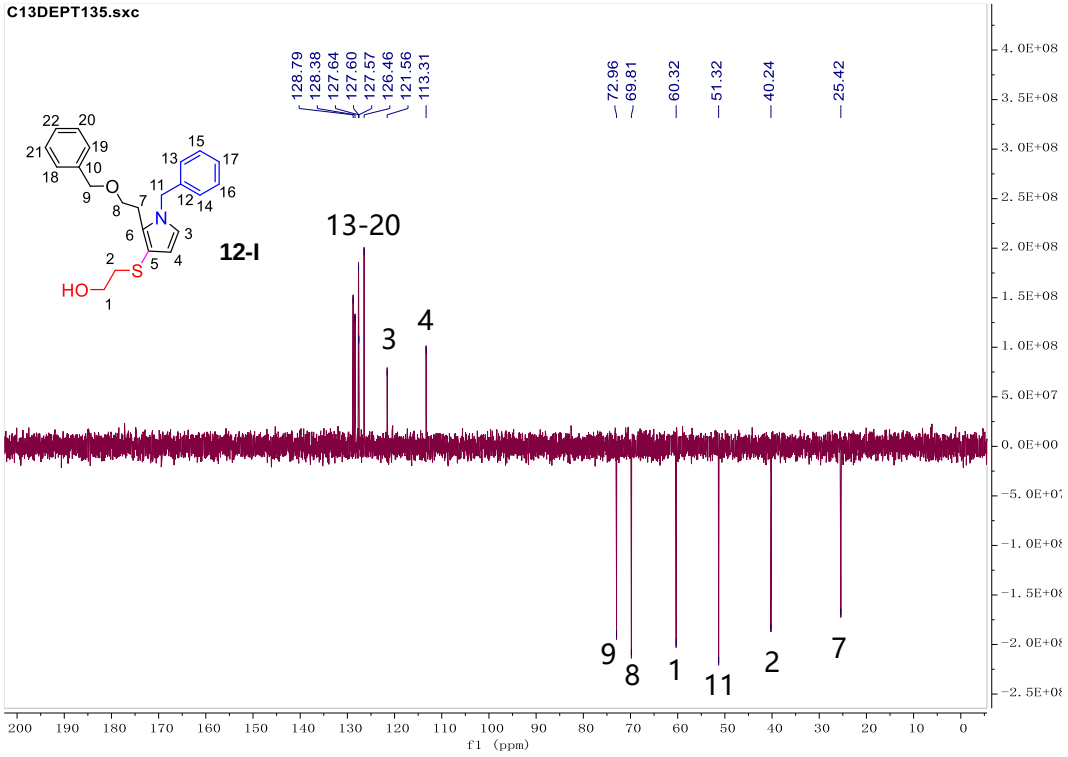
^{1}H NMR of 12-I:



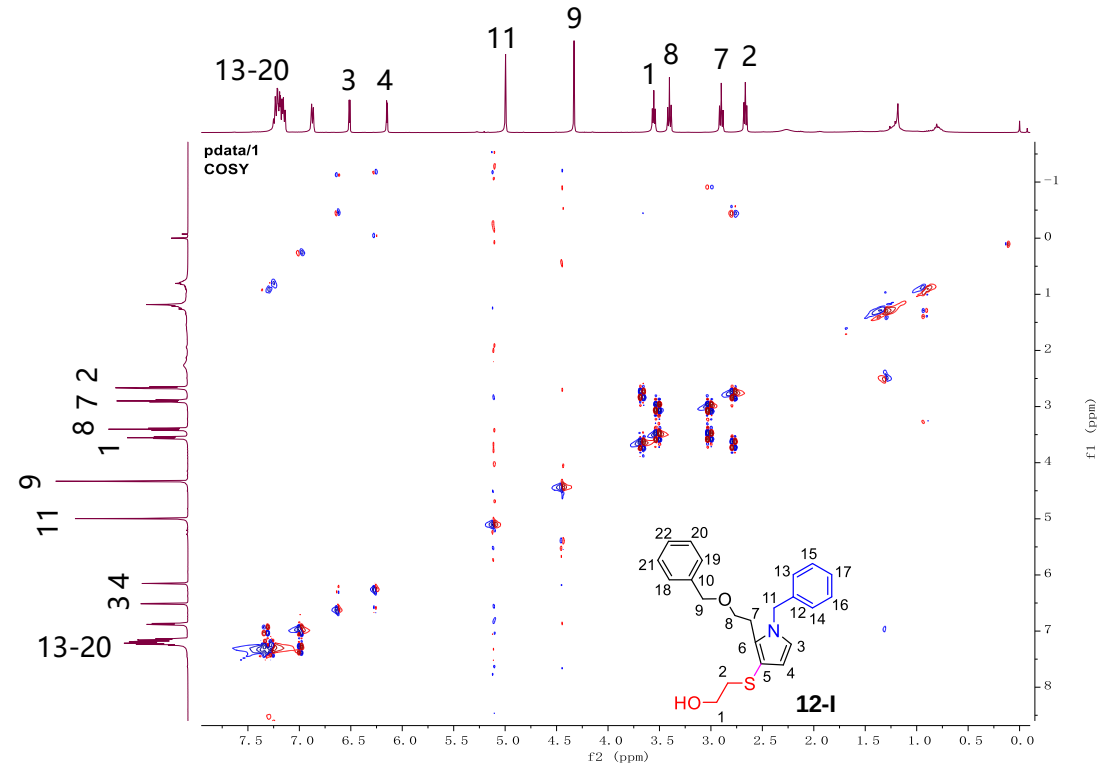
^{13}C NMR of 12-I:



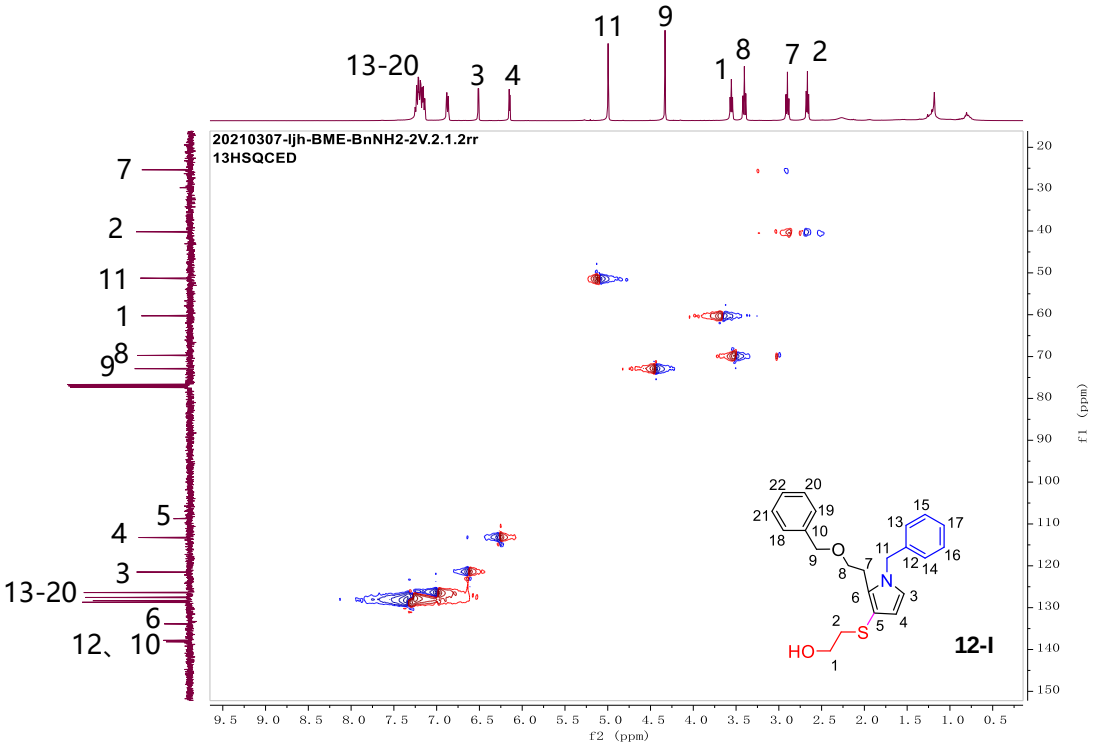
DEPT 135 of 12-I:



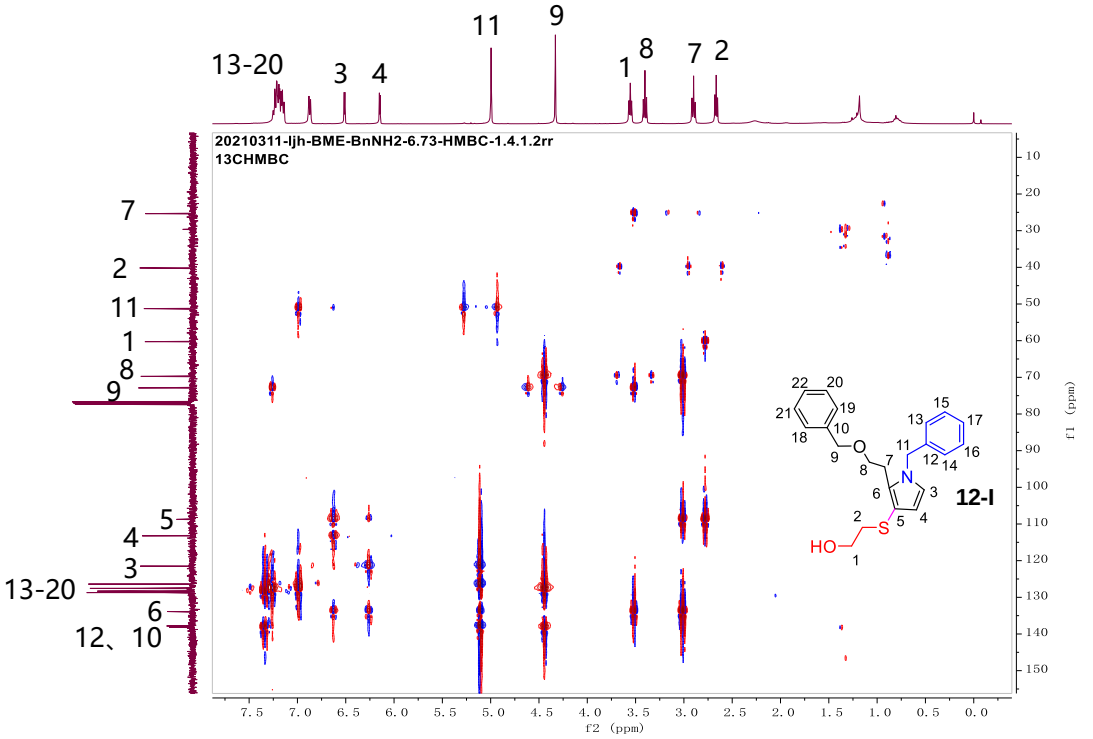
COSY of 12-I:



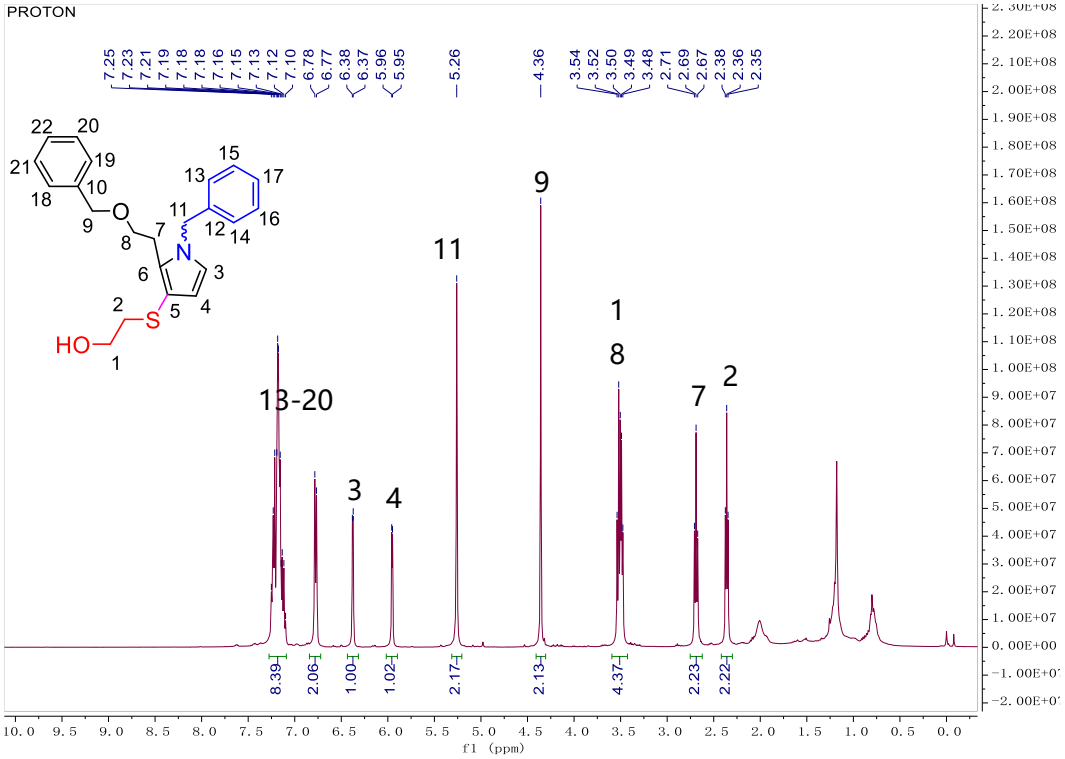
HSQC of 12-I:



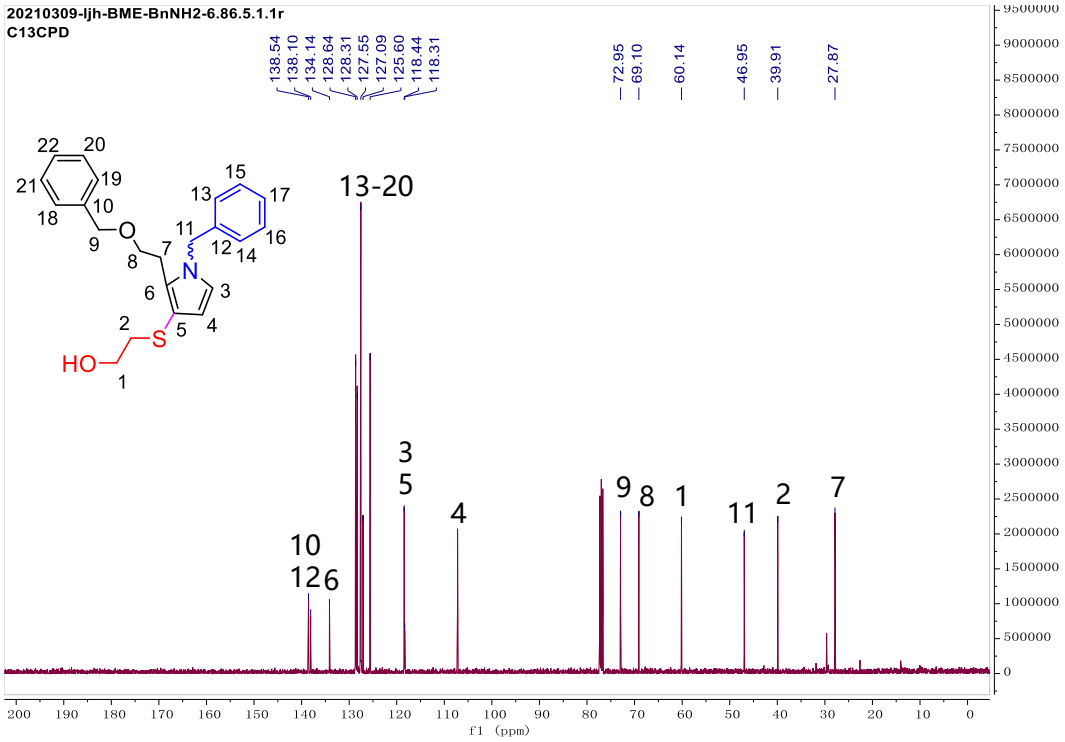
HMBC of 12-I:



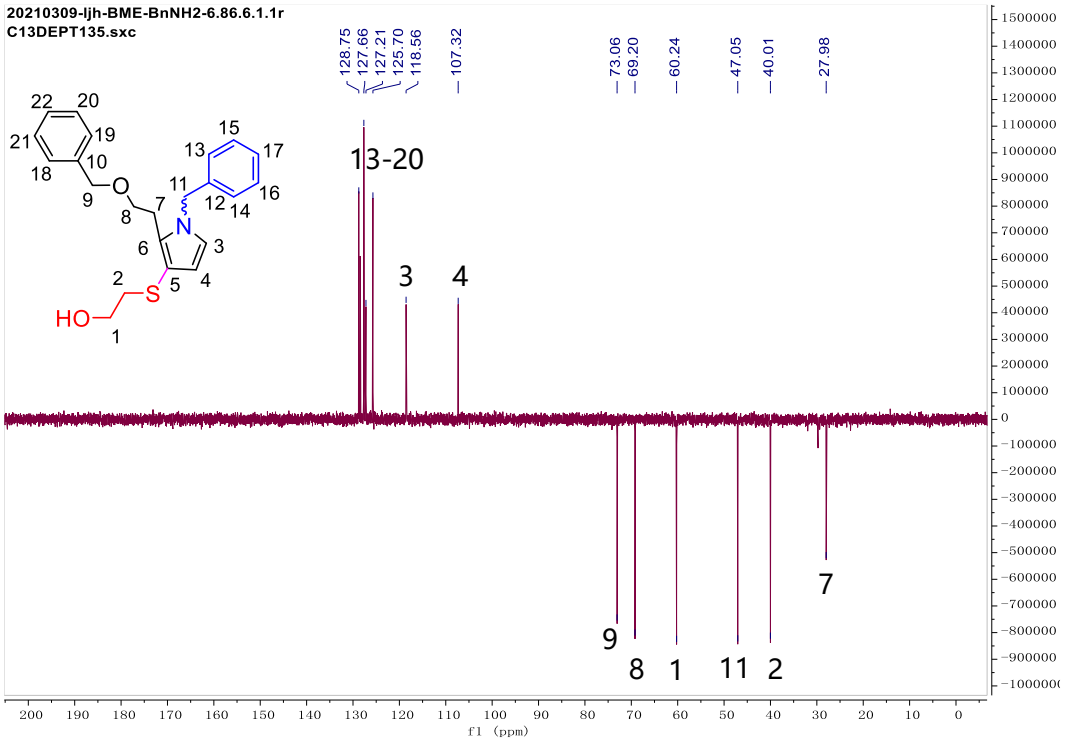
^{1}H NMR of 12-II:



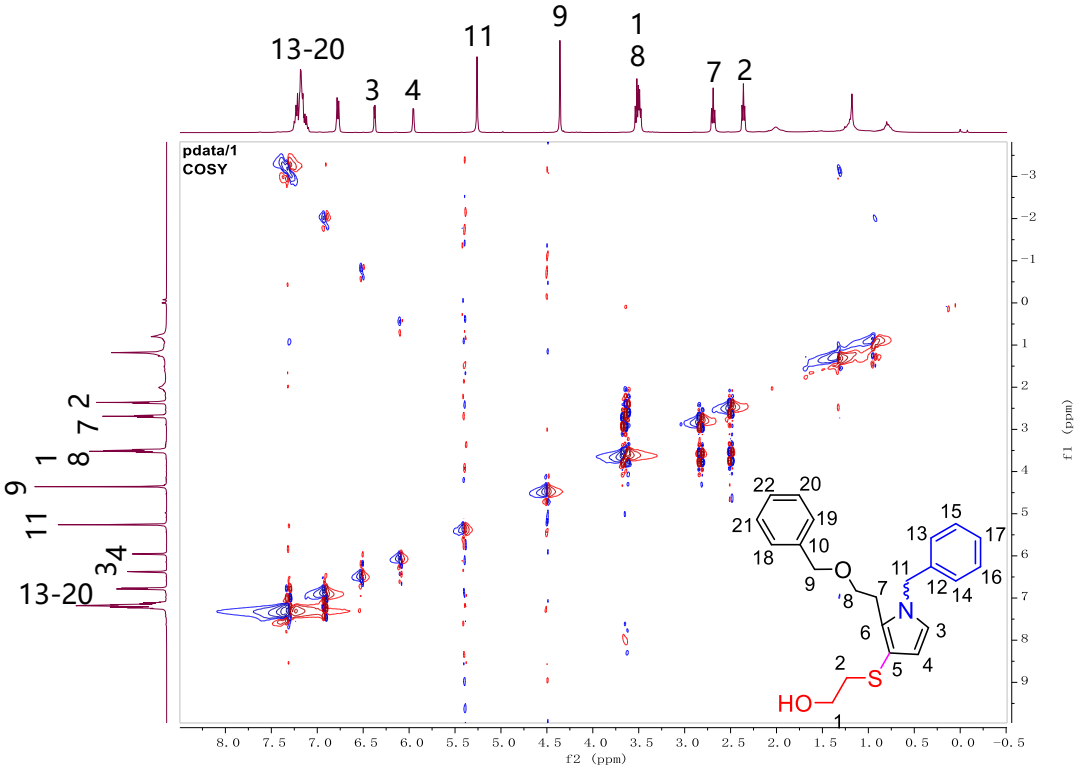
^{13}C NMR of 12-II:



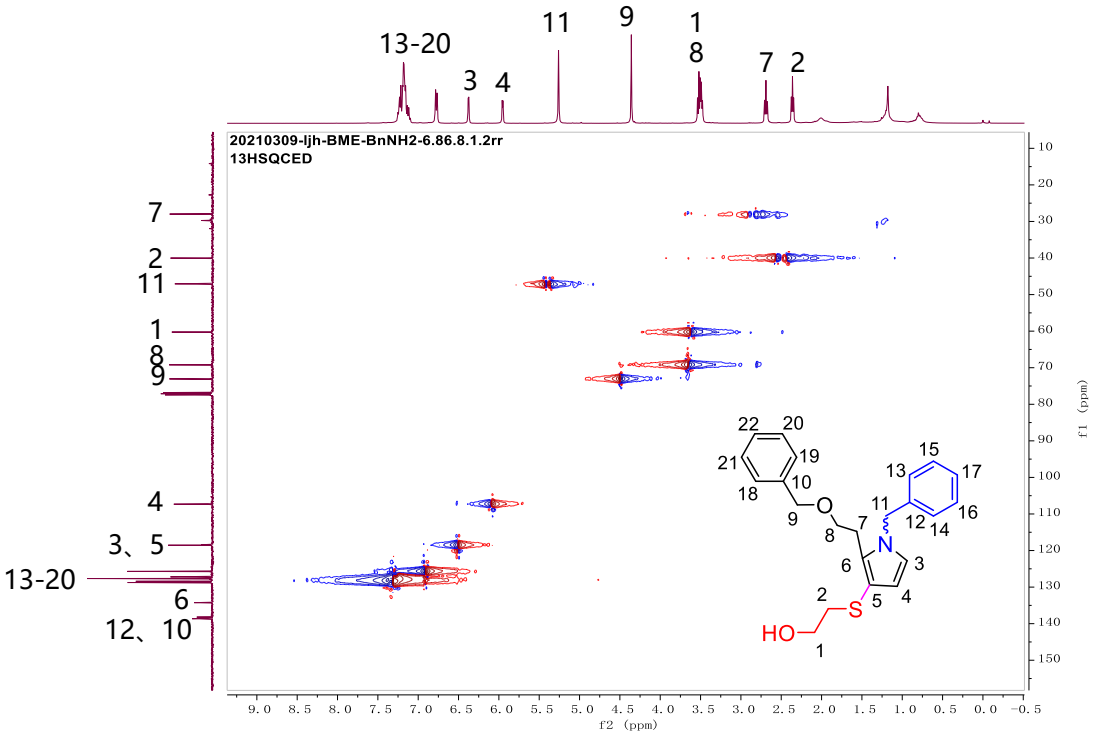
DEPT 135 of 12-II:



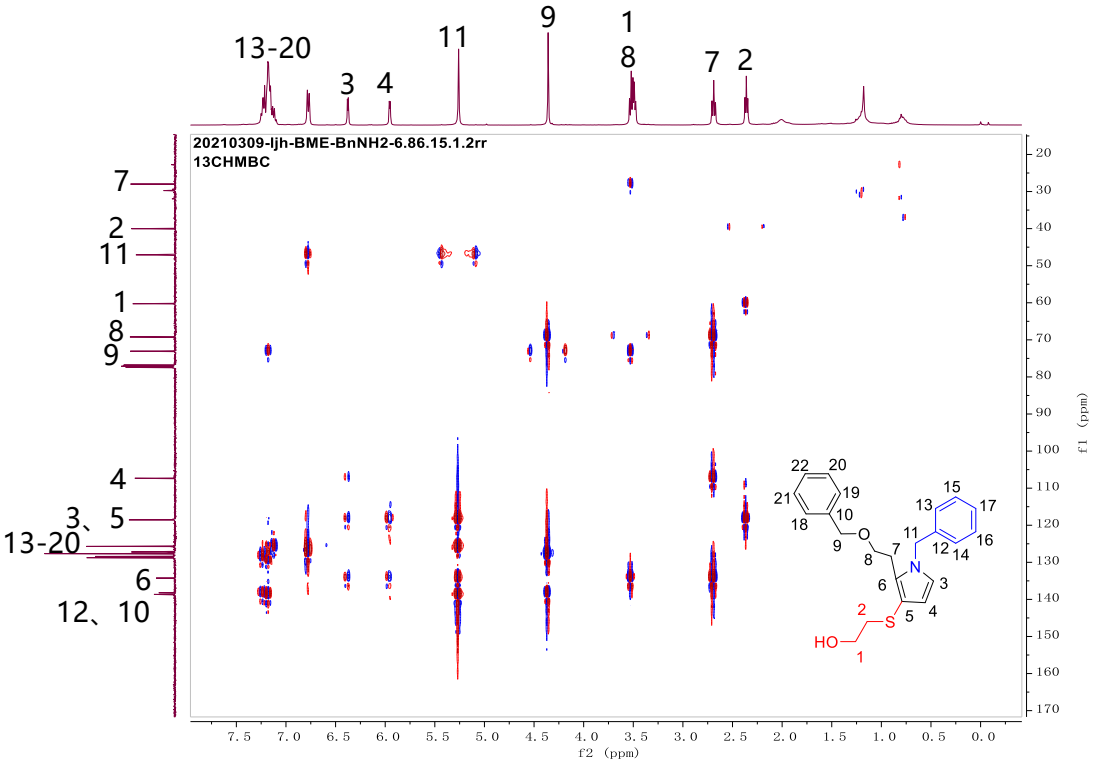
COSY of 12-II:



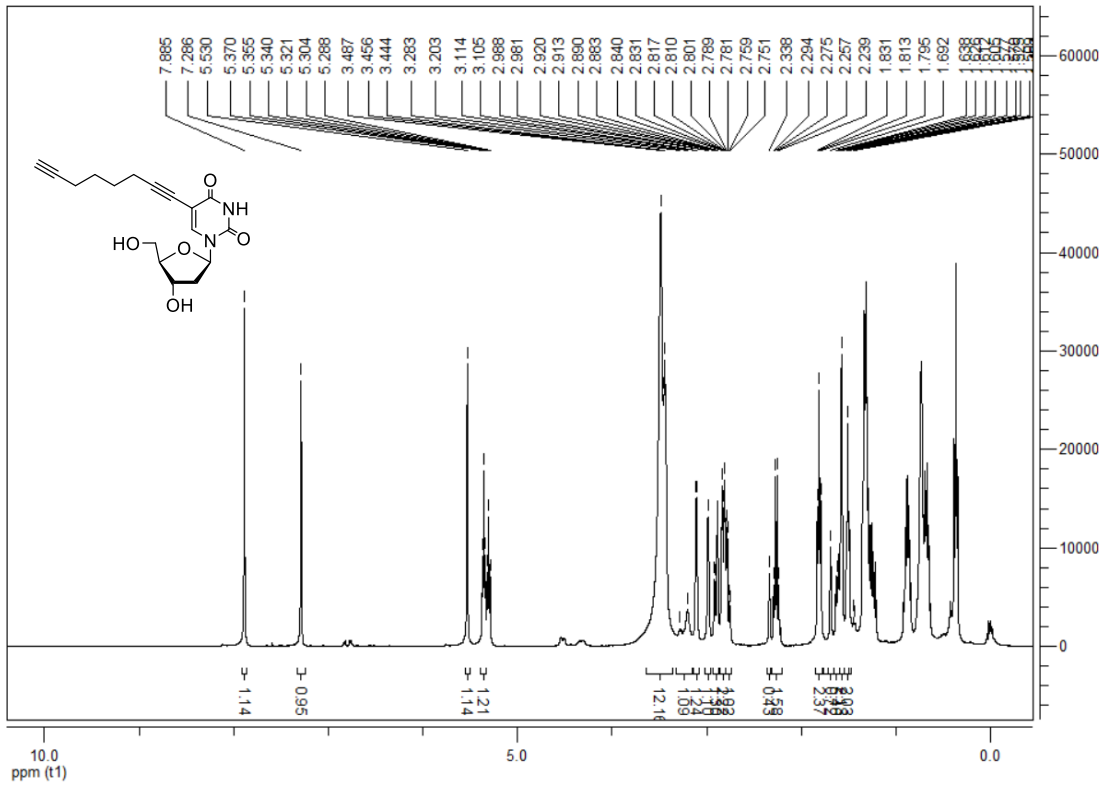
HSQC of 12-II:



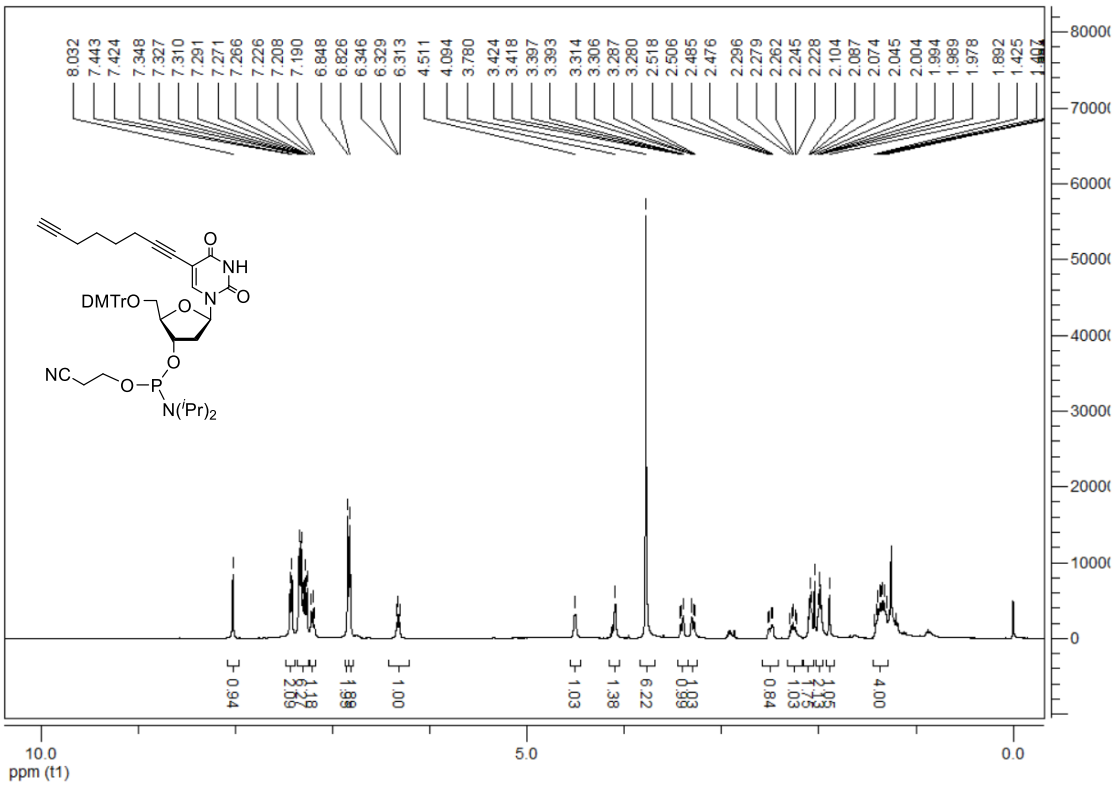
HMBC of 12-II:



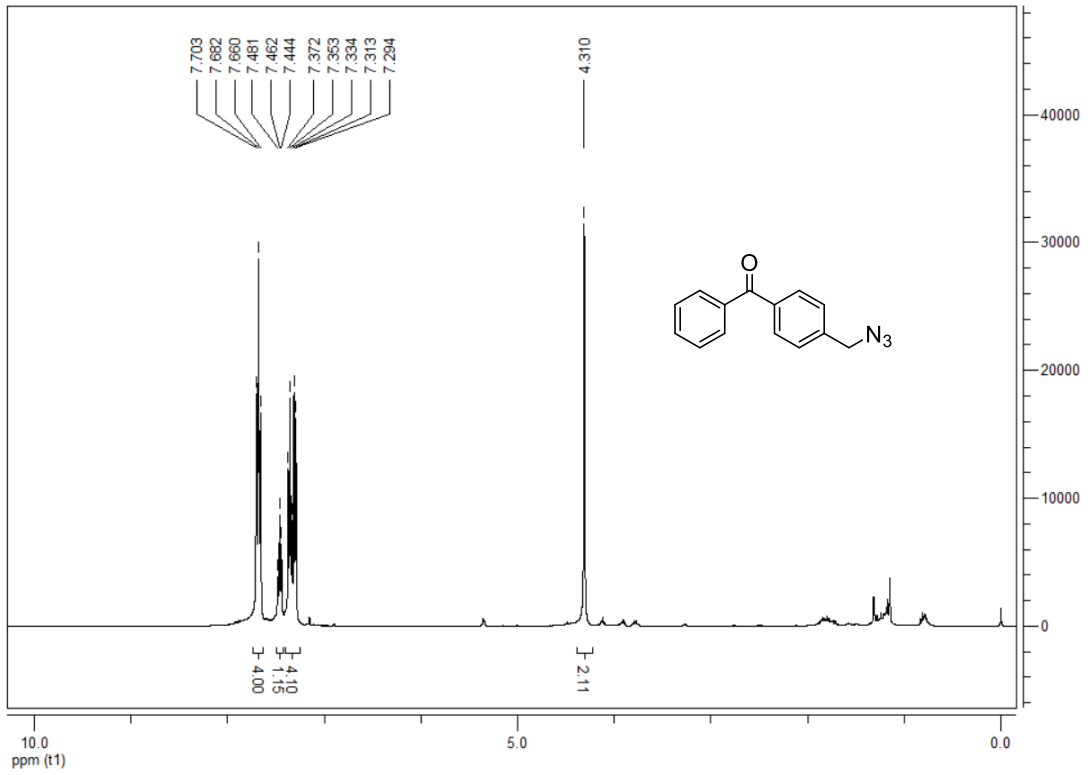
^{1}H NMR of 31:



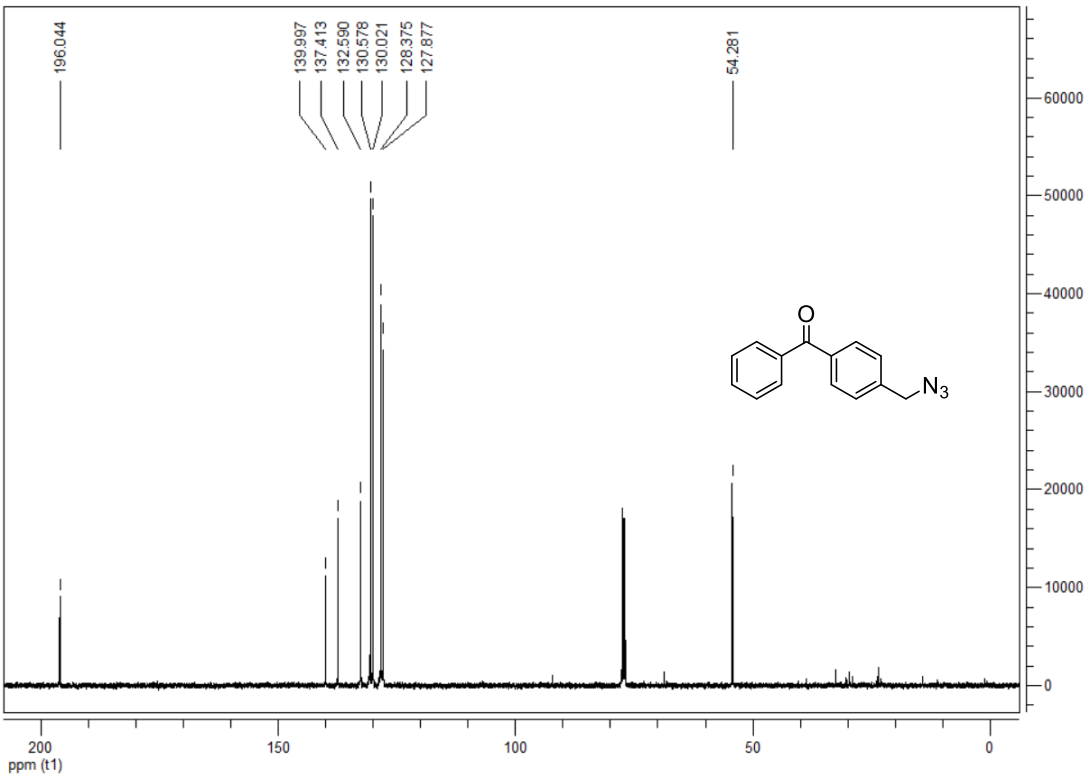
^{1}H NMR of 32:



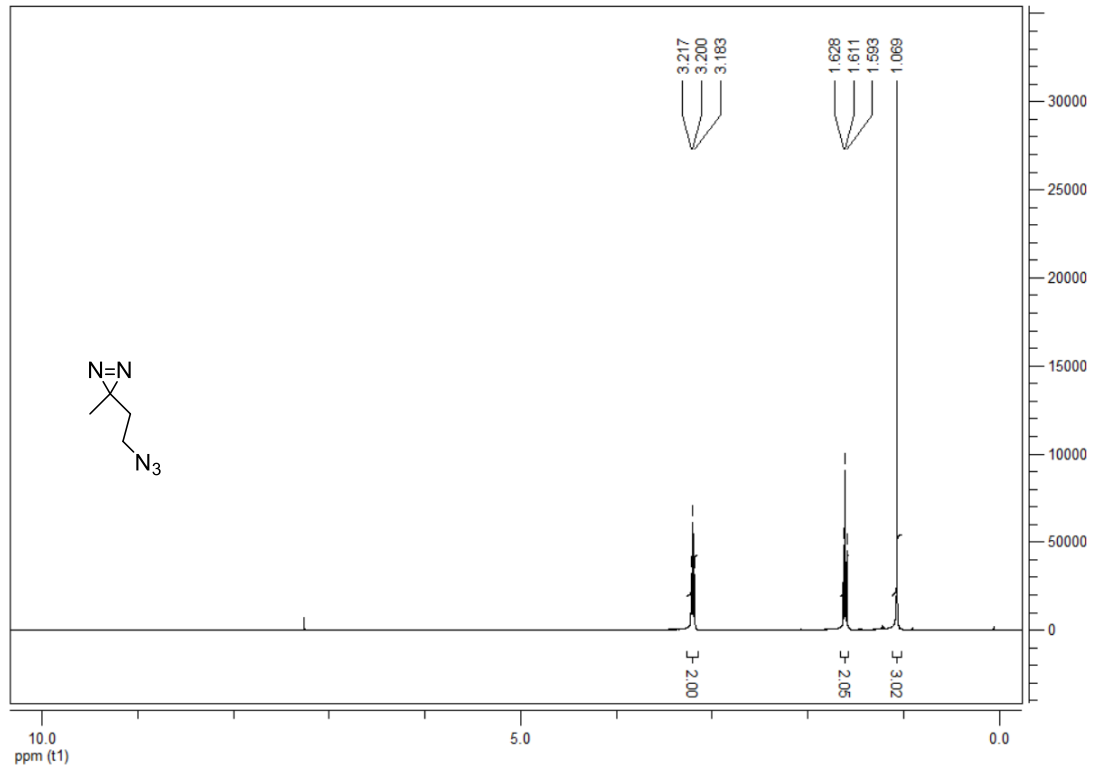
[1]H NMR of 34:



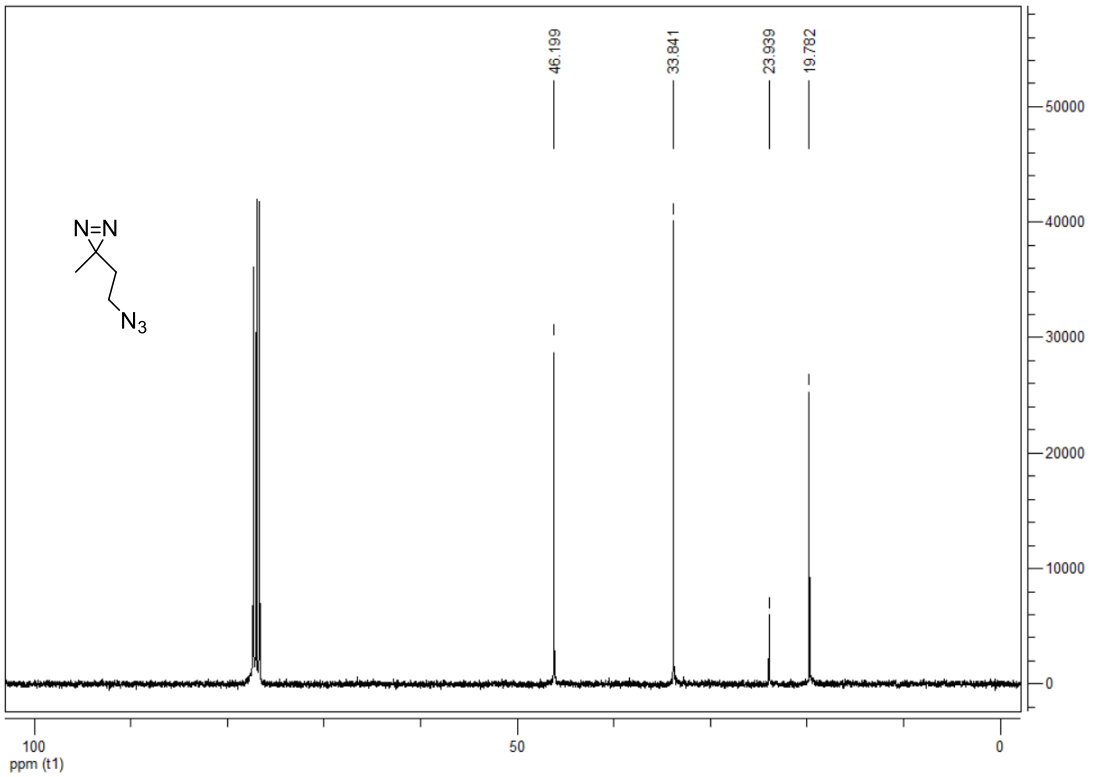
[13]H NMR of 34:



^{13}H NMR of 36:



^{13}H NMR of 36:



Supplementary references

1. J. Gierlich, G. A. Burley, P. M. E. Gramlich, D. M. Hammond and T. Carell, Click chemistry as a reliable method for the high-density postsynthetic functionalization of alkyne-modified DNA, *Org. Lett.*, 2006, **8**, 3639-3642.
2. A. S. Schroeder, J. Steinbacher, B. Steigenberger, F. A. Gnerlich, S. Schiesser, T. Pfaffeneder and T. Carell, Synthesis of a DNA Promoter Segment Containing All Four Epigenetic Nucleosides: 5-Methyl-, 5-Hydroxymethyl-, 5-Formyl-, and 5-Carboxy-2'-Deoxycytidine, *Angew. Chem. Int. Ed.*, 2014, **53**, 315-318.
3. F. Seela and V. R. Sirivolu, Nucleosides and Oligonucleotides with Diynyl Side Chains: Base Pairing and Functionalization of 2'-Deoxyuridine Derivatives by the Copper(I)-Catalyzed Alkyne□Azide 'Click(Cycloaddition, *Helv. Chim. Acta.*, 2007, **90**, 535-552.
4. Y. Hari, M. Nakahara and S. Obika, Triplex-forming ability of oligonucleotides containing 1-aryl-1,2,3-triazole nucleobases linked via a two atom-length spacer, *Biorg. Med. Chem.*, 2013, **21**, 5583-5588.
5. D. Vasudevan, E. Y. D. Chua and C. A. Davey, Crystal structures of nucleosome core particles containing the '601' strong positioning sequence, *J. Mol. Biol.*, 2010, **403**, 1-10.
6. M. Ren, M. M. Greenberg and C. Zhou, Participation of Histones in DNA Damage and Repair within Nucleosome Core Particles: Mechanism and Applications, *Acc. Chem. Res.*, 2022, **55**, 1059-1073.
7. M. Ren, M. Shang, H. Wang, Z. Xi and C. Zhou, Histones participate in base excision repair of 8-oxodGuo by transiently cross-linking with active repair intermediates in nucleosome core particles, *Nucleic Acids Res.*, 2021, **49**, 257-268.
8. F. Li, Y. Zhang, J. Bai, M. M. Greenberg, Z. Xi and C. Zhou, 5-Formylcytosine yields DNA–protein cross-links in nucleosome core particles, *J. Am. Chem. Soc.*, 2017, **139**, 10617-10620.
9. C. Z. Zhou and M. M. Greenberg, Histone-catalyzed cleavage of nucleosomal DNA containing 2-deoxyribonolactone, *J. Am. Chem. Soc.*, 2012, **134**, 8090-8093.
10. L. S. Pidugu, J. W. Flowers, C. T. Coey, E. Pozharski, M. M. Greenberg and A. C. Drohat, Structural Basis for Excision of 5-Formylcytosine by Thymine DNA Glycosylase, *Biochemistry*, 2016, **55**, 6205-6208.