Supporting Information

Site-specific cross-linking of proteins to DNA via a new bioorthogonal approach employing oxime ligation

Suresh S. Pujari,[a,b] Yi Zhang,[c] Shaofei Ji,[b,c] Mark D. Distefano,[c] and Natalia Y. Tretyakova[a,b*†]

[a] *Department of Medicinal Chemistry, University of Minnesota College of Pharmacy, Minneapolis, Minnesota 55455, USA*

[b] *The Masonic Cancer Center, University of Minnesota, Minneapolis, Minnesota 55455, USA*

[c] *Department of Chemistry, University of Minnesota, Minneapolis, Minnesota 55455, USA*

† Natalia Tretyakova, Email: trety001@umn.edu. Phone: 612-626-3432; Fax: 612-624-3869

Table of Contents

S1. Supplementary methods: Materials 4

S2. Synthesis and characterization of oligodeoxynucleotides 4

S3. DPC formation between 5fdC containing **DNA-1** and peptides 6

S4. DPC formation between 7-oxoethyl-7-deaza-dG containing DNA and peptides 7

S5. Synthesis and characterization of 5-formyl-dC cross-linked to **Pep-4** 7

S6. Expression and purification of VHH7 8

S7. Sortase A Mediated Ligation of peptide into protein 9

S8. DPC formation between 5fdC containing DNA-1 and oxy-Lys conatining VHH protein 9

S9. Identification of the amino acids participating in DNA-protein cross-links 10

Scheme S1. Synthesis of aminooxy lysine (oxy-Lys) 12

Scheme S2. Synthesis of 7-(2-dihydroxypropyl)-7-deaza-deoxy guanosine phophoramidite........ 13

Scheme S3. Oxidation of diol containing **DNA-2** to 7-(oxoethyl)-7-deaza-dG containing DNA 14

Table S1. HPLC-ESI$^+$-MS/MS analysis of tryptic digests from fusion protein **VHH-2** cross-linked to 5fdC containing 12-mer oligodeoxynucleotide **DNA-1** 15

Figure S1. Concentration dependent study of peptide and DNA-**1** 16

Figure S2. Optimization of reaction conditions for DPC formation by **DNA-1** and **Pep-1** 17

Figure S3. Optimization of reaction conditions for DPC formation by **DNA-1** and **Pep-1** 18

Figure S4. DNA-peptide conjugate formation of 5-fdC containing DNA duplex and **pep-2** and **pep-4**.......19

Figure S5. DNA-peptide conjugate formation of 7-oxoethyl-7-deaza-dG containing DNA and different peptides 20

Figure S6. DNA-peptide conjugate formation of 7-(2-oxoethyl)-dG containing DNA duplex and **pep-1** and **pep-2**........21

Figure S7. Mass-spectrum of aldehyde DNA isolated from oxidation of **DNA**-**2** 22

Figure S8. Concentration dependence study for DNA-protein cross-links with 5fdC containing oligodeoxynucleotide **DNA-1** and oxy-Lys containing protein **VHH-2** 23

Figure S9. Denaturing PAGE analysis of DNA-protein cross-links of 7-oxoethyl-7-deaza-dG containing **DNA-2** 24

Figure S10.NanoLC-NSI$^+$-MS/MS analysis of **VHH7-2** tryptic peptide containing 5-formyl-dC cross-link to oxy-Lys.. 25

Figure S11. ^{1}H-NMR spectrum of compound **3**........26

Figure S12. ^{13}C-NMR spectrum of compound **3**........27

Figure S13. ^{1}H-NMR spectrum of compound **3a**........28

Figure S14. ^{13}C-NMR spectrum of compound **3a**........29

Figure S15. ^{1}H-NMR spectrum of compound **4.**........30

Figure S16. ^{13}C-NMR spectrum of compound **4**........31

Supplementary Methods

S1. Materials. Nucleoside phosphoramidites, solid supports, and all other reagents for solid phase synthesis of DNA were purchased from Glen Research (Sterling, VA). Unmodified oligodeoxynucleotides were purchased from Integrated DNA Technologies (Coralville, IA). 5-methyl-2′-deoxycytidine hydrochloride was acquired from Berry and Associates (Dexter, MI). Cell culture media was obtained from Thermo Fisher Scientific (Waltham, MA). Solvents and reagents for synthesis and purification were purchased from Sigma-Aldrich (St Louis, MO) or Thermo Fisher Scientific (Waltham, MA), unless otherwise noted. Phosphodiesterase I, Phosphodiesterase II, and Deoxyribonuclease I were purchased from Worthington Biochemical Corp. (Lakewood, NJ). Sortase A was generously provided by Dr. Mohammad Rashidian from Dr. Hidde Ploegh's lab at MIT.

S2. Synthesis and characterization of oligodeoxynucleotides. 5fC-containing oligodeoxynucleotides (Table 1) were prepared by solid phase synthesis on an ABI 394 DNA synthesizer (Applied Biosystems, Foster City, CA) using 5-formyl dC III phosphoramidite (Glen Research, Sterling, VA) under standard coupling conditions. The deprotection and cleavage of oligodeoxynucleotides from solid support was accomplished by incubating in concentrated ammonium hydroxide at room temperature in the dark for 17 h. Following ammonia removal in vacuo, DNA strands were incubated in 80% glacial acetic acid in water at room temperature for 6 h to remove acetal protection group. All synthetic oligodeoxynucleotides were purified either by semi-preparative HPLC or by 20% (w/v) denaturing-PAGE containing 7 M urea. The purified oligodeoxynucleotides were desalted by Illustra NAP-5 or NAP-10 columns (GE Healthcare, Pittsburgh, PA), characterized by HPLC-ESI--MS on an Agilent ion trap mass spectrometer, and quantified by UV spectrophotometry. All other unmodified oligodeoxynucleotides were purchased from Integrated DNA Techonologies (Coralville, IA) and purified by semi-preparative HPLC before use.

Benzyl (R)-2-(((benzyloxy)carbonyl)amino)-5-hydroxypentanoate (1a): This compound was synthesized as reported in literature.[1] The characterization data matches the reported data. ^{1}H NMR ($CDCl_3$, 500 MHz): δ 7.26 (s, 10H), 5.40-5.41 (d, 1H, *J* = 6.5Hz), 5.06-5.13 (m, 2H), 5.03-5.5.05 (m, 2H), 4.37-4.41 (m, 1H), 3.54 (t, 1H, *J* = 6.1Hz), 1.84-1.1.91 (m, 1H), 1.67-1.74 (m, 1H), 1.42-1.55 (m, 1H). ^{13}C NMR ($CDCl_3$, 125 MHz): δ 172.1, 155.8, 140.9, 136.1135.2, 128.6, 128.5, 128.5, 128.3, 128.2, 128.1, 127.6, 126.967.4, 67.2, 67.1, 67.0, 65.3, 62.0, 60.4, 59.2, 53.6, 49.6, 29.3, 28.0, 25.5, 21.0, 20.9, 14.2.

Benzyl (R)-2-(((benzyloxy)carbonyl)amino)-5-((methylsulfonyl)oxy)pentanoate (**2**). This compound was synthesized as reported in literature.[2] The characterization data matches the reported data. ^{1}H NMR ($CDCl_3$, 500 MHz): δ 7.28 (s, 10H), 5.27-5.29 (d, 1H, *J* = 7.5 Hz), 5.05-5.14 (m, 2H), 5.03 (s, 2H), 4.38-4.39 (d, 1H, *J* = 5Hz), 4.12 (s, 2H), 2.88 (s, 3H), 1.93-1.94 (d, 1H, *J* = 5.5Hz), 1.66-1.73 (m, 1H). ^{13}C NMR ($CDCl_3$, 125 MHz): δ 171.7, 155.9, 136.0, 135.0, 128.7, 128.6, 128.5, 128.4, 128.2, 128.1, 68.8, 67.4, 67.1, 60.4, 53.2, 50.8, 37.3, 28.9, 25.0, 21.0, 14.1

Benzyl (R)-2-(((benzyloxy)carbonyl)amino)-5-((1,3-dioxoisoindolin-2-yl)oxy)pentanoate (2a). This compound was synthesized as reported in literature.[2] The characterization data matches the reported data. ^{1}H NMR ($CDCl_3$, 500 MHz): δ 7.73-7.74 (m, 2H), 7.63-7.67 (m, 2H), 7.24-7.28 (m, 10H), 5.42-5.43 (d, *J* = 8 Hz, 1H), 5.12 (s, 2H), 5.02 (s, 2H), 4.39-4.43 (m, 1H), 4.11 (br s, 2H), 2.07-2.09 (m, 1H), 1.89-1.97 (m, 1H), 1.71-1.75 (m, 2H). ^{13}C NMR ($CDCl_3$, 125 MHz): δ 172.1, 163.7, 156.1136.4, 135.4, 134.6, 129.0, 128.7, 128.6, 128.5, 128.4, 128.2, 128.1, 123.6, 67.3, 67.1, 53.8, 28.9, 24.4.

Benzyl (R)-5-(aminooxy)-2-(((benzyloxy)carbonyl)amino)pentanoate (3). To a solution of compound 2a (2.49 g, 4.9 mmol) in dichloromethane (60 mL) was added methylhydrazine (7.44 mmol, 392µL) at 0°C and stirred at 0°C for 1hr. After completion of the reaction (as monitored by TLC), the reraction mixture was passed through celite bed and filterate was evaporated to dryness under reduced pressure and subjected to flash column chromatography (FC) to give compound **3** (1.65 g, 89%)as a colourless solid. TLC (silica gel, CH_2Cl_2/MeOH, 90:10) R_f 0.3. ESI$^+$-MS: Exact mass calc. for $[C_{20}H_{23}N_2O_4]^+$ 355.1652 [M + H]$^+$, Exact mass obs: 355.1656 [M + H]$^+$. ^{1}H NMR ($CDCl_3$, 500 MHz): δ 7.35 (s, 11H), 5.44-5.46 (d, *J* = 8 Hz, 1H), 5.17-5.20 (m, 2H), 5.13 (s, 2H), 4.45-4.49 (m, 1H), 3.66-3.69 (m, 2H), 3.51 (s, 1H), 1.92-1.97 (m, 1H), 1.74-1.86 (m, 1H), 1.60-1.69 (m, 1H). ^{13}C NMR ($CDCl_3$, 125 MHz): δ 172.2, 155.9, 136.2, 135.2, 128.6, 128.5, 128.5, 128.3, 128.2, 128.1, 74.9, 67.2, 67.0, 53.7, 50.8, 49.6, 24.0.

Benzyl (R)-2-(((benzyloxy)carbonyl)amino)-5-(((ter-butoxycarbonyl)amino)oxy)pentanoate (3a). To a solution of compound **3** (100 mg, 0.26 mmol) in dry THF (16 mL) was added triethylamine (54 mg, 0.52 mmol) and di-tert-butyl dicarbonate (116 mg, 0.52 mmol) and the reaction mixture was stirred at room temperature for 3hr. After completion of the reaction (as monitored by TLC), the mixture was evaporated to dryness under reduced pressure and subjected to flash column chromatography (FC) to give compound **3a** (90 mg, 72%) as a colourless oil. TLC (silica gel, EtOAc/Hexanes, 70:30) R_f 0.5. ESI$^+$-MS: Exact mass calc. for $[C_{25}H_{31}N_2O_6]^+$ 455.2177 [M + H]$^+$, Exact mass obs: 455.1215 [M + H]$^+$. ^{1}H NMR ($CDCl_3$, 500 MHz): δ 7.23-7.27 (m, 11H), 7.11 (s, 1H), 5.49-5.51 (d, *J* = 8 Hz, 1H), 5.06-5.12 (m, 2H), 5.02 (s, 2H), 4.32-4.37 (m, 1H), 3.73 (t, 1H), 1.87-1.91 (m, 1H), 1.75-1.79 (m, 1H), 1.55-1.59 (m, 2H). ^{13}C NMR ($CDCl_3$,

125 MHz): δ 172.2, 157.0, 156.1, 136.2, 135.3, 128.6, 128.5, 128.3, 128.1, 81.7, 75.8, 67.1, 66.9, 53.8, 29.0, 28.1, 23.9.

(R)-2-((((9H-fluoren-9-yl)methoxy)carbonyl)amino)-5-(((tert-butoxycarbonyl)amino)oxy)pentanoic acid (4). A solution of compound **3a** (0.6 g, 1.27 mmol) in methanol (10 ml) at room temperature was stirred with 10% Pd/C (0.270 g, 2.53 mmol) under 1 atm H_2 for 2 hr. The Pd/C was removed by filteration and filtrate was concentrated. The residue was dissolved in dioxane/water mixture (1:1, v/v, 10 ml). To this solution was added 9-fluorenylmethylsuccinimidyl carbonate (FmocOsu) (0.51g, 1.51 mmol) and $NaHOC_3$ (0.23 g, 2.75 mmol) and stirred for overnight at room temperature. The reacion mixture was further acidified to pH 3-4 by addition of 1N HCl. After neutralization, the reaction solution was extracted with ethyl acetate (3x 30 ml) and the combined organic extracts were washed with water (30 ml) and brine (30 ml) and dried over anhydrous Na_2SO_4 and evaporated to dryness under reduced pressure. The residue was further purified by FC to obtain as viscous oil. Lyophilization from acetonitrile-water provided colorless solid (0.35 g, 59%). TLC (silica gel, DCM/ MeOH, 90:10) R_f 0.3. ESI^+-MS: Exact mass calc. for $[C_{25}H_{30}N_2NaO_7]^+$ 493.1945 $[M + Na]^+$, Exact mass obs: 493.1946 $[M + Na]^+$. ^1H NMR ($CDCl_3$, 500 MHz): δ 7.65 (bs, 2H), 7.50 (bs, 3H), 7.28 (bs, 2H), 7.12 (bs, 2H), 4.28-4.35 (m, 3H), 4.10 (bs, 1H), 3.76-3.78 (m, 2H), 1.94 (bs, 1H), 1.78 (bs, 1H), 1.62 (bs, 2H), 1.37 (s, 9H). ^{13}C NMR ($CDCl_3$, 125 MHz): δ 143.73, 141.25, 127.67, 127.09, 125.20, 119.92, 53.43, 47.10, 29.71, 28.22.

S3. DPC formation between 5fdC containing DNA-1 and peptides.

Synthetic 12-mer oligodeoxynucleotides containing 5fdC (Table 1, 250 pmol) were radiolabeled by incubating with γ-^{32}P ATP (1 µL, PerkinElmer Life Sciences, Boston, MA) and T4 PNK (20 units) in 20 µL 1× T4 PNK buffer (New England Biolabs, Beverly, MA) at 37 °C for 1 h. The reaction mixture was heated at 65 °C for 10 min to inactivate the enzyme and then filtered through an illustra MicroSpin G-25 column (GE Healthcare, Pittsburgh, PA) to get rid of excess γ-^{32}P ATP. 12-mer oligodeoxynucleotides containing 5fC, (3 pmol) were incubated with 100-fold molar excess of individual polypeptides **1**-**4** (300 pmol, Table 1) in 20 µL ammonium acetate buffer (100 mM, pH 4.5) containing 100mM aniline at 37 °C overnight (~16 h). The reaction mixture was then heated to 90 °C for 10 minutes before loading on geland followed by analysis by denaturing 20% (w/v) PAGE containing 7 M urea. The gel was run at 300 V in 1X TBE buffer until the xylene cyanol band migrated to the bottom (~ 2.5 h). The radiolabeled DNA strands and DPCs were detected with a Typhoon FLA 7000 instrument (GE Healthcare, Pittsburgh, PA) and quantified using ImageQuant TL 8.0 (GE Healthcare, Pittsburgh, PA) (Table 1).

S4. DPC formation between 7-oxoethyl-7-deaza-dG containing DNA and peptides.

Synthetic 23-mer oligodeoxynucleotides containing 7-dihydroxypropyl-7-deaza DNA (Table 1, **DNA-1**, 250 pmol) were radiolabeled by incubating with γ-^{32}P ATP (1 µL, PerkinElmer Life Sciences, Boston, MA) and T4 PNK (20 units) in 20 µL 1× T4 PNK buffer (New England Biolabs, Beverly, MA) at 37 °C for 1 h. The reaction mixture was heated at 65 °C for 10 min to inactivate the enzyme and then filtered through an illustra MicroSpin G-25 column (GE Healthcare, Pittsburgh, PA) to remove excess γ-^{32}P ATP. 23-mer oligodeoxynucleotides containing 7-hydroxypropyl-7-deaza-dG, (40 pmol) were incubated with 10mM $NaIO_4$ for 6h at 4 °C in 15mM sodium phosphate buffer pH 5.6 in dark conditions and then quenched with 100mM Na_2SO_3 to obtain the 7-oxoethyl-7-deaza-dG containing DNA. 7-oxoethyl-7-deaza-dG containing DNA (3 pmol) was further incubated with 100-fold molar excess of individual polypeptides **Pep 1**-**4** (300 pmol, Table 1) in 20 µL phosphate buffer (15 mM, pH 5.6) at 37 °C overnight (~16 h). The reaction mixture was then heated to 90 °C for 10 minutes before loading on gel and followed by analysis by denaturing 20% (w/v) PAGE containing 7 M urea. The gel was run at 300 V in 1X TBE buffer until the xylene cyanol band migrated to the bottom (~ 2.5 h). The radiolabeled DNA strands and DPCs were detected with a Typhoon FLA 7000 instrument (GE Healthcare, Pittsburgh, PA) and quantified using ImageQuant TL 8.0 (GE Healthcare, Pittsburgh, PA) (Table 1).

S5. Synthesis and characterization of 5-formyl-dC cross-linked to Pep-4.

Nucleoside-peptide conjugate was prepared by incubating 5-formyl-dC (16 nmol) with 11-mer peptide (3 nmol, Table 1) in 4.5 mM ammonium phosphate buffer (pH 4.5, 16 µL) at 37 °C for 2 h. The reaction mixture was dried in vacuo, and reconstituted in 60 µL of 2:98 (v/v) acetonitrile: water containing 0.1 % TFA. Samples were desalted by NAP-5 column and analysed by mass spectrometer. In order to obtain MS^2 spectra of nucleoside-peptide conjugates, capillary HPLC-ESI+-MS/MS analysis was performed on an Agilent 1100 HPLC-MSD ion trap mass spectrometer. A Zorbax 300 SB-C18 column (150 × 0.5 mm, 5 µm) was eluted at a flow rate of 15 µL/min using 0.1% TFA in water (A) and 0.1% TFA in acetonitrile (B). Solvent composition was linearly increased from 2% to 50% B in 25 min. The mass spectrometer was operated in the positive mode with the mass range of m/z 400-1800. MS^2 fragmentation was carried out by collision-induced dissociation with an isolation width of 4.0 m/z and fragmentation amplitude of 0.95 V.

S6. Expression and purification of VHH7

The pHEN6 plasmid was generously provided by Dr. Mohammad Rashidian from Dr. Hidde Ploegh's lab at MIT. The gene encoding VHH7 with C-terminal LPETGGHHHHHHGCVIA sequence (VHH7) was synthesized and cloned into the pHEN6 plasmid by DNA 2.0, Inc. The pelB signal sequence was also attached to the protein for periplasmic expression. The confirmed plasmid was transformed (heat shock, 45 s at 42 °C) into *E.coli* WK6 for protein expression. The expression and purification procedure was adopted from an online protocol with some modification. A single colony was used to inoculate 50 mL LB media containing 100 µg/mL ampicillin. The flask was incubated at 37 °C overnight with shaking at 250 rmp. After overnight incubation, 10 mL of the dense culture was added to 1 L LB media with 100 µg/mL ampicillin. The culture was incubated at 37 °C with shaking at 250 rpm until OD_{600} of 0.6 was reached. To induce protein expression, 1 mM IPTG was added and the culture was incubated at 30 °C overnight with shaking at 180 rpm. After expression, cells were collected by centrifugation at 8,000 g for 10 min at 4 °C.

To extract VHH7 from periplasm, the cell pellet was resuspended and incubated in 150 mL TES buffer (200 mM Tris, 0.5 mM EDTA, 500 mM sucrose, pH 8.0) at 4 °C with slow shaking. After 1 h, 300 mL of 4x diluted TES buffer was added to the solution and the flask was incubated at 4 °C overnight with slow shaking. Supernatant was obtained by centrifugation at 13,000 g for 30 min at 4 °C. Ni-NTA column was employed to purify the VHH7 protein from the collected supernatant. Briefly, Ni-NTA column was first equilibrated with buffer containing PBS and 10 mM imidazole. The supernatant was then loaded to the column, followed by extensive washing with buffer containing PBS and 20 mM imidazole. VHH7 protein was eluted with elution buffer containing PBS and 250 mM imidazole. Fractions were collected from the column and their absorbance at 280 nm was tested. The fractions with Abs_{280} higher than 0.1 were combined and concentrated. Buffer-exchange was performed using Amicon filters (Millipore, 3k cut-off) with PBS. Glycerol was added to a final concentration of 20% and the protein solution was stored in -80 °C freezer. SDS-PAGE was used to characterize the purity of the protein and Bradford assay was employed to determine protein concentration.

VHH7 sequence

DNA sequence

CAGGTGCAGCTGCAGGAGTCAGGGGGAGGATTGGTGCAGGCTGGGGACTCTCTGAGACTCTCCTGCGCAGCCTCTGG
ACGCACCTTCAGTCGCGGTGTAATGGGCTGGTTCCGCCGGGCTCCAGGGAAGGAGCGTGAGTTTGTAGCAATCTTTA
GCGGGAGTAGCTGGAGTGGTCGTAGTACATACTATTCAGACTCCGTAAAGGGCCGATTCACCATCTCCAGAGACAAC
GCCAAGAACACGGTGTATCTGCAAATGAACGGCCTGAAACCTGAGGACACGGCCGTTTATTACTGTGCAGCGGGATA
TCCGGAGGCGTATAGCGCCTATGGTCGGGAGAGTACATATGACTACTGGGGCCAGGGGACCCAGGTCACCGTCTCCT
CAGGATCCCTTCCTGAAACTGGTGGCCATCATCATCATCATCATGGTTGCGTGATCGCGTA

Protein sequence

Q V Q L Q E S G G G L V Q A G D S L R L S C A A S G R T F S R G V M G W F R R
A P G K E R E F V A I F S G S S W S G R S T Y Y S D S V K G R F T I S R D N A
K N T V Y L Q M N G L K P E D T A V Y Y C A A G Y P E A Y S A Y G R E S T Y D
Y W G Q G T Q V T V S S G S L P E T G G H H H H H H G C V I A

S7. **Sortase A mediated ligation of peptide into protein.**

Sortase A catalyzed reactions were performed in the buffer containing: 50 mM Tris·HCl, pH 7.5, 10 mM $CaCl_2$, 200 mM NaCl, 1 mM **Pep-4**, 100 µM VHH7 protein and 10 µM sortase A. After incubation at 4 °C for 2 h, Ni-NTA resin was added to the reaction mixture to remove sortase and unreacted VHH7 protein. The solution was centrifuged and the supernatant was collected. Buffer exchange was performed five times with Amicon filters (3k cut-off) using 100 mM NH_4OAc with 400 mM NaCl, pH 4.5. The modified protein was stored at 4 °C for future usage.

S8. DPC formation between 5fdC containing DNA-1 and oxy-Lys conatining VHH protein.

Synthetic 12-mer oligodeoxynucleotides containing 5fdC (Table 1, 300 pmol) were radiolabeled by incubating with γ-^{32}P ATP (2 µL, PerkinElmer Life Sciences, Boston, MA) and T4 PNK (20 units) in 20 µL 1× T4 PNK buffer (New England Biolabs, Beverly, MA) at 37 °C for 1 h. The reaction mixture was heated at 65 °C for 10 min and filtered through an illustra MicroSpin G-25 column (GE Healthcare, Pittsburgh, PA) to remove excess of γ-^{32}P ATP. 12-mer oligodeoxynucleotide containing 5fdC **DNA-1** (3 pmol) was incubated with 100-fold molar excess of **VHH-2** protein (300 pmol) in 20 µL ammonium acetate buffer containing aniline (100 mM, pH 4.5) at 37 °C overnight (~16 h). The reaction mixture was dried and then reconstituted in gel loading buffer (20 mM EDTA in 95% formamide containing 0.05% bromophenol blue and xylene cyanol), and followed by analysis by denaturing 20% (w/v) PAGE containing 7 M urea heating

10-15 min at 90 °C before loading onto the gel. The gel was run at 300 V in 1X TBE buffer until the xylene cyanol band migrated to the bottom (~ 2.5 h). The radiolabeled DNA strands and DPCs were detected with a Typhoon FLA 7000 instrument (GE Healthcare, Pittsburgh, PA) and quantified using ImageQuant TL 8.0 (GE Healthcare, Pittsburgh, PA). Identical procedure was employed for other reactions containing **VHH-1** and **DNA-3** (unmodified 12-mer DNA identical to **DNA-1** except the 5-fdC modification) as well.

S9. Identification of the amino acids participating in DNA-protein cross-links.

12-mer oligodoxynucleotide containing 5fdC (300 pmol, Table 1, **DNA-1**) was cross-linked to fusion protein **VHH-2** (600 pmol, 2-fold molar excess) in 100 mM ammonium acetate in aniline buffer as described above. The reaction mixtures were purified by 4-12% SDS PAGE and stained with SimplyBlue stain. The DPC-containing gel bands were diced into 1 mm cubes, subjected to DTT reduction for 20 min at 50 °C (25 mM ammonium bicarbonate, 10 mM DTT at pH 7.9), and alkylated with saturated iodoacetamide (10 µL) for 20 min at room temperature. Gel pieces were dehydrated with acetonitrile (until slices turned white), dried in vacuo and incubated at 37°C overnight with PDE I (120 mU), PDE II (105 mU), Dnase I (35 U) and alkaline phosphate (22 U) in Tris buffer (100 µL, pH 7.9) containing 10 mM Tris, 15 mM $MgCl_2$. The sample was subjected to tryptic digestion to get tryptic peptides (2.5 µg, MS grade Trypsin Gold, Promega, Madison, WI), followed by StageTip desalting, and stored at -20 °C prior to MS analysis.

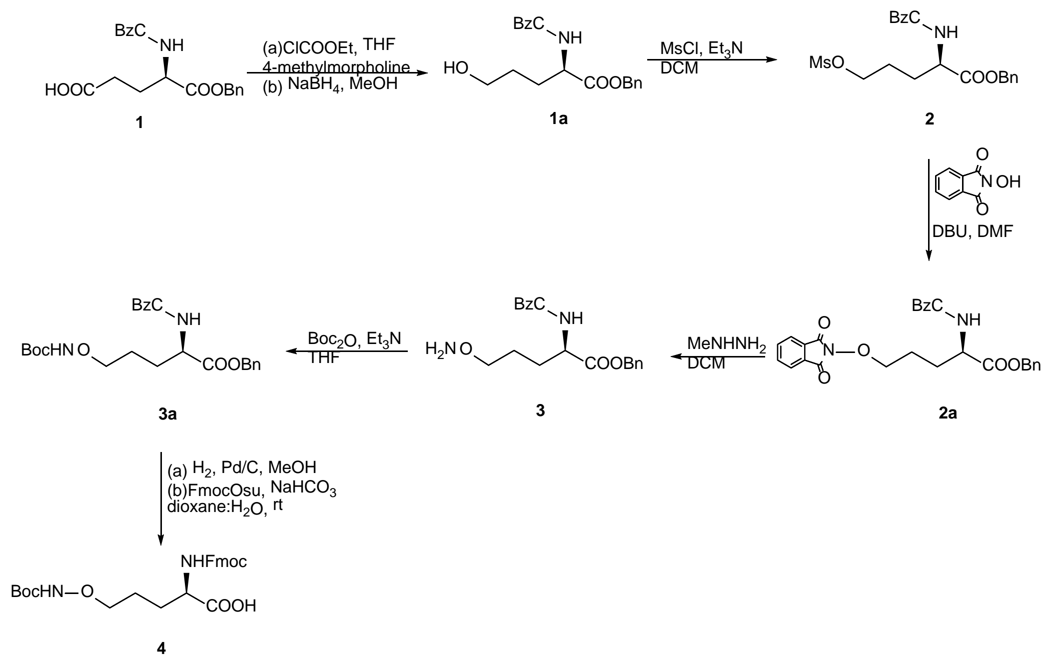
HPLC-ESI$^+$-MS/MS analyses of tryptic peptides were conducted using a Thermo Dionex UltiMate3000 HPLC coupled to a LTQ Orbitrap Elite mass spectrometer (Thermo Scientific, Watham, MA). Peptide mixtures (5 µL) were separated on a nano-HPLC column (75 µm i.d., 10 cm packed bed, 15 µm emitter, New Objective, Woburn, MA) packed in house with Luna C18, 5 µm, separation media (Phenomenex, Torrance, CA). Liquid chromatography was conducted at ambient temperature with a flow rate of 0.3 µL/min using 0.1% formic acid in water (A) and 0.1% formic acid in acetonitrile (B). The gradient program was kept at 2% B for 6 min, increased to 70% B over 90 min and further to 95% B in 2 min. Mass spectrometry was performed with the Orbitrap mass analyzer operating at a resolution of 60000, a scan range of m/z 400-2000 and a 20 V in-source fragmentation. MS/MS spectra were collected with the ion trap mass analyzer in a data-dependent manner, in which one full scan was followed by eight MS/MS spectra with an isolation width of 1.0 m/z, activation time of 30 ms, activation Q of 0.25, normalized CID collision energy of 35%, 1 repeat count, exclusion duration of 20 s and a exclusion mass width of 5 ppm. Spectral data were analyzed using Thermo Proteome Discoverer 1.4 (Thermo Scientific, Watham, MA)

that combined raw data extraction, database searching and probability scoring. Cysteine carbamidomethylation (+ 57.021 Da) was set as a fixed modification, and lysine or arginine residues were set as having dynamic modification sites oxy-Lys modified (+123.006 Da) and cross-linked to 5fC (+ 239.504 Da). The search parameters included trypsin specificity and up to 2 missed cleavage sites with 10 ppm as precursor mass tolerance (Table S1).

References.

1. *S. Jiang, P. Li, C. C. Lai, J. A. Kelly, P. P. Roller. J. Org. Chem*. **2006**, *71*, 7307-7314.
2. *F. Liu, J. Thomas, T. R. Burke Jr. Synthesis* **2008**, 2432-2438

Scheme S1. Synthesis of aminooxy lysine (oxy-Lys).

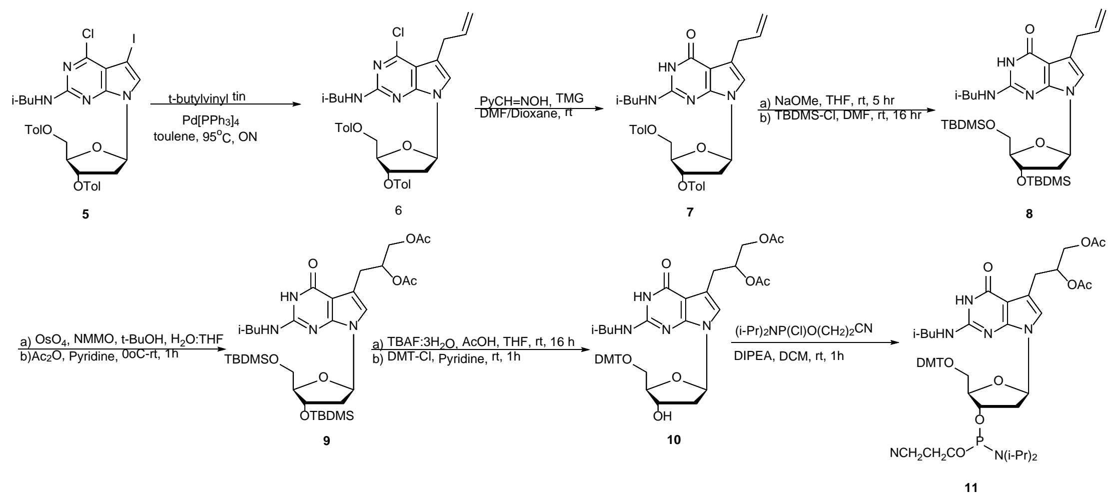


References.

1. *S. Jiang, P. Li, C. C. Lai, J. A. Kelly, P. P. Roller. J. Org. Chem*. **2006**, *71*, 7307-7314.

2. *F. Liu, J. Thomas, T. R. Burke Jr. Synthesis* **2008**, 2432-2438

Scheme S2. Synthesis of 7-(2-dihydroxypropyl)-7-deaza-deoxy guanosine phophoramidite.[3]



3. *T. Angelov, A. Guainazzi, O. D. Scharer. Org. Lett.*, **2009**, *11*, 661-664.

Scheme S3. Oxidation of diol containing **DNA-2** to 7-(oxoethyl)-7-deaza-dG containing DNA.

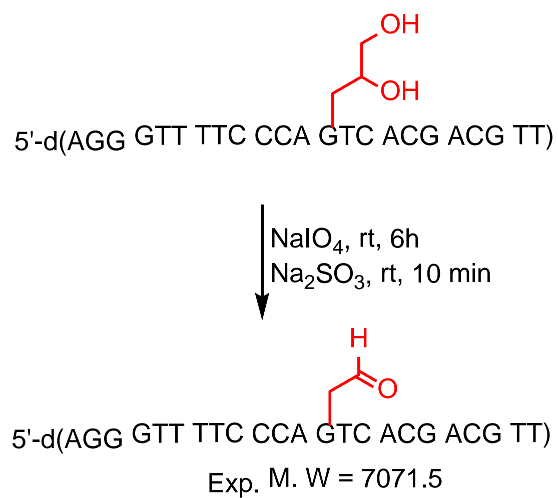


Table S1. HPLC-ESI$^+$-MS/MS analysis of tryptic digests from fusion protein **VHH-7-2** cross-linked to 5fdC containing 12-mer oligodeoxynucleotide **DNA-1**.

Sequences	Modification	Charge	m/z	M(Da)
EFVAIFSGSSWSGR		2	765.3724	1528.7308
ESTYDYWGQGTQVTVSSGSLPETGGGK		3	931.0930	2790.2568
ESTYDYWGQGTQVTVSSGSLPETGGGK		2	1396.1320	2790.2508
EREFVAIFSGSSWSGRSTYYSDSVKGR		4	765.3733	3057.4638
EREFVAIFSGSSWSGR		3	605.6315	1813.8728
APGKEREFVAIFSGSSWSGRSTYYSDSVKGR		5	683.1370	3410.6488
ESTYDYWGQGTQVTVSSGSLPETGGGK		3	931.0956	2790.2648
EREFVAIFSGSSWSGRSTYYSDSVKGRFTISR		5	733.3693	3661.8098
APGKEREFVAIFSGSSWSGRSTYYSDSVK		4	800.3943	3197.5478
APGKEREFVAIFSGSSWSGRSTYYSDSVK		5	640.5164	3197.5458
RAPGKEREFVAIFSGSSWSGRSTYYSDSVKGR		5	714.3589	3566.7588
RAPGKEREFVAIFSGSSWSGR		4	581.8041	2323.1868
EFVAIFSGSSWSGRSTYYSDSVKGR		3	925.1030	2772.2868
LScAASGR	C3(Carbamidomethyl)	2	411.2005	820.3864
EREFVAIFSGSSWSGRSTYYSDSVKGR		5	612.4967	3057.4470
APGKEREFVAIFSGSSWSGRSTYYSDSVK		4	800.3998	3197.5770
STYYSDSVK		2	525.2425	1048.4708
APGKEREFVAIFSGSSWSGRSTYYSDSVK		5	640.5154	3197.5410
EFVAIFSGSSWSGRSTYYSDSVK		3	854.0707	2560.1908
FTISRDNAKNTVYLQMNGLkPEDTAVYYcAAGYPEAYSAYGR	K20(aminoxy-dC-dR), C29(Carbamidomethyl)	4	1223.5600	4890.2160
LScAASGRTFSRGVMGWFRR	C3(Carbamidomethyl)	4	576.2909	2301.1438
GLGkGGA	K4(aminoxy-dC-dR)	2	341.6668	681.3190
APGKEREFVAIFSGSSWSGRSTYYSDSVKGRFTISR		5	804.0077	4015.0018
STYYSDSVKGR		2	631.8036	1261.5928
LScAASGRTFSRGVmGWFRRAPGKER	C3(Carbamidomethyl), M15(Oxidation)	6	493.5861	2955.4728
GLGkGGA	K4(aminoxy-dC-dR)	2	341.6664	681.3182
STYYSDSVkGRFTISRDNAk	K9(aminoxy-dC-dR), K20(aminoxy-dC)	2	1329.0990	2656.1838
GVmGWFR	M3(Oxidation)	2	434.7107	867.4067
APGKEREFVAIFSGSSWSGRSTYYSDSVkGRFTISR	K29(aminoxy-dC-dR)	6	690.6776	4138.0218
FTISRDNAKNTVYLQmNGLkPEDTAVYYcAAGYPEAYSAYGR	M16(Oxidation), K20(aminoxy-dC-dR), C29(Carbamidomethyl)	6	818.7070	4906.1978

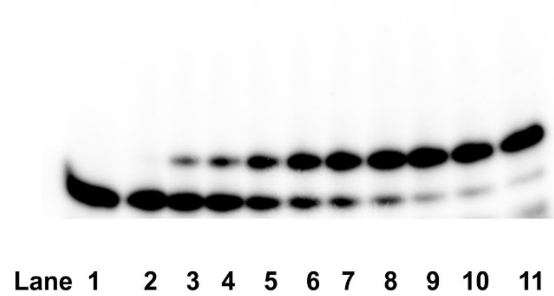


Figure S1. Concentration dependent study of peptide and DNA-**1**. Study was conducted with increasing concentartion of peptide to DNA. The reaction was carried out in 100 mM ammonium acetate buffer pH 4.5 containing 100mM aniline at 37°C for overnight. DNA was radiolabeled with [γ-^{32}P] ATP and the 20% PAGE gel was run at 300 volt for 2.5h. Lane 1. DNA Lane 2. DNA:peptide ratio being, 1:2 (4.9%); Lane 3. 1:5 (23.8%); Lane 4. 1:10 (31.5%); Lane 5. 1:20 (46.4%); Lane 6. 1:30 (59.5%); Lane 7. 1:40 (67.8%); Lane 8. 1:50 (75.4%); Lane 9. 1:70 (84.9%); Lane 10. 1:85 (85.1%); Lane 11. 1:100 (91.2%).

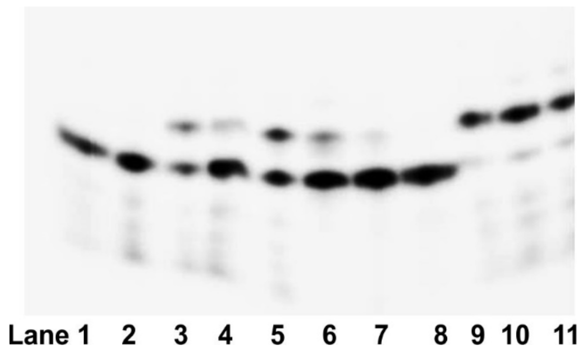


Figure S2. Optimization of reaction conditions for DPC formation by **DNA-1** and **Pep-1**. All the reactions were carried out in NH_4OAc buffer with 100 mM aniline. DNA was radiolabeled with [γ-^{32}P] ATP and the gel was run at 300 volt for 2.5hr with DNA:peptide ratio being Lane 1. 1:5, rt, pH = 4.5 (yield, 0%); Lane 2. 1:10, rt, pH = 4.5 (0%); Lane 3. 1:100, rt, pH = 4.5 (45%); Lane 4. 1:1000, rt, pH = 4.5 (14%); Lane 5. 1:100, rt, pH = 5.6 (43%); Lane 6. 1:100, rt, pH = 7.0 (17%); Lane 7. 1:100, rt, pH = 8.5 (trace); Lane 8. 1:100, rt, pH = 10 (0%); Lane 9. 1:100, 37°C, pH = 4.5 (90%); Lane 10. 1:100, 50°C, pH = 4.5 (89%); Lane 11. 1:100, 65°C, pH = 4.5 (89%).

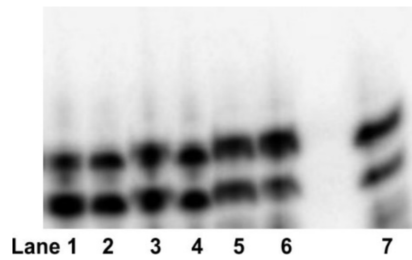


Figure S3. Optimization of reaction conditions for DPC formation by **DNA-1** and **Pep-1** with DNA:peptide ratio being 1:100. All the reactions were carried out in NH_4OAc buffer pH=4.5 with 100 mM aniline at 37°C. DNA was radiolabeled with [γ-^{32}P] ATP and the gel was run at 300 volt for 2.5hr. Lane 1. 0.5 h (28%); Lane 2. 1 h (40%); Lane 3. 1.5 h (48%); Lane 4. 2 h (52%); Lane 5. 3 h (52%); Lane 6. 5 h (77%); Lane 7. 8 h (77%).

The 23-mer 5fdC containing DNA 5'-d(AGG GTC TTC C**C**A GTC ACG ACG TT)-3' (where **C** represents 5-fdC) was used for duplex studies. Upon addition of complementary strand 3'-d(TCC CAG AAG GGT CAG TGC TGC AA)-5', the mixture was heated to 90°C for 10 min and cooled down slowly to room temperature overnight and conjugation reaction was performed.

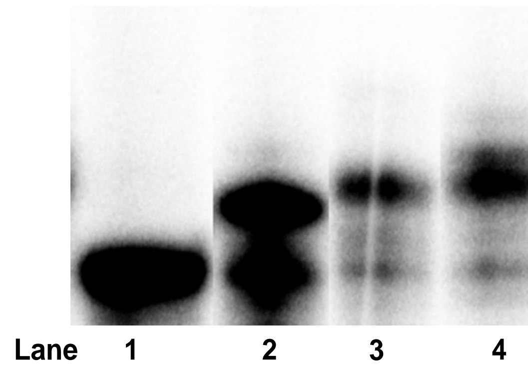


Figure S4. DNA-peptide conjugate formation of 5-fdC containing DNA duplex and **pep-2** and **pep-4**. Lane 1. 5'-d(AGG GTC TTC C**C**A GTC ACG ACG TT)-3'; Lane 2. ss-DNA 5'-d(AGG GTC TTC C**C**A GTC ACG ACG TT)-3' conjugated with **Pep**-**2**; Lane 3. Preformed duplex of 5'-d(AGG GTC TTC C**C**A GTC ACG ACG TT)-3' and 3'-d(TCC CAG AAG GGT CAG TGC TGC AA)-5' reacted with **Pep**-**2. Lane 4.** Preformed duplex of 5'-d(AGG GTC TTC C**C**A GTC ACG ACG TT)-3' and 3'-d(TCC CAG AAG GGT CAG TGC TGC AA)-5' reacted with **Pep**-**4**. DNA was radiolabeled with [γ-^{32}P] ATP and the 20% PAGE gel was run at 300 volt for 2.5hr.

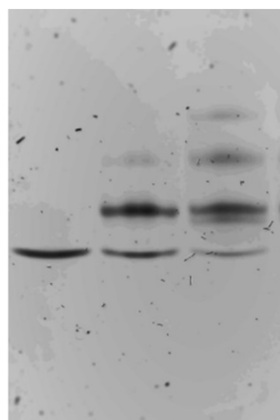


Figure S5. DNA-peptide conjugate formation of 7-oxoethyl-7-deaza-dG containing DNA and different peptides. Lane 1. DNA -**2**; Lane 2. Peptide-**2**; Lane 3. **Pep**-**4** with DNA:peptide ratio being 1:100. All the reactions were carried out in phosphate buffer pH=5.6 with no catalyst at 37°C. DNA was radiolabeled with [γ-^{32}P] ATP and the gel was run at 300 volt for 2.5hr.

The 23-mer 7-(2-dihydroxypropyl)-dG containing **DNA-2** was used for duplex studies. Upon addition of complementary strand 3'-d(TCC CAG AAG GGT CAG TGC TGC AA)-5', the mixture was heated to 90°C for 10 min and cooled down slowly to room temperature overnight and consequently oxidation and conjugation reaction was performed.

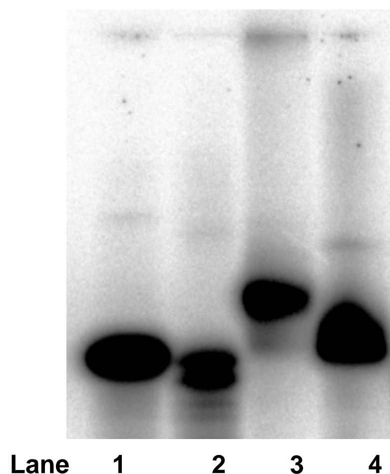
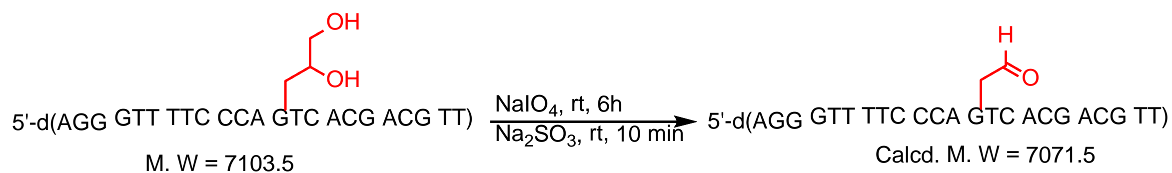


Figure S6. DNA-peptide conjugate formation of 7-(2-oxoethyl)-dG containing **DNA-2** duplex and **pep-1** and **pep-2**. Lane 1. **DNA-2**; Lane 2. Duplex of **DNA-2 and** 3'-d(TCC CAG AAG GGT CAG TGC TGC AA)-5'; Lane 3. Preformed duplex of **DNA-2** and 3'-d(TCC CAG AAG GGT CAG TGC TGC AA)-5' reacted with **Pep-2. Lane 4.** Preformed duplex of **DNA-2** and 3'-d(TCC CAG AAG GGT CAG TGC TGC AA)-5' reacted with **Pep**-**1**. DNA was radiolabeled with [γ-^{32}P] ATP and the 20% PAGE gel was run at 300 volt for 2.5hr.



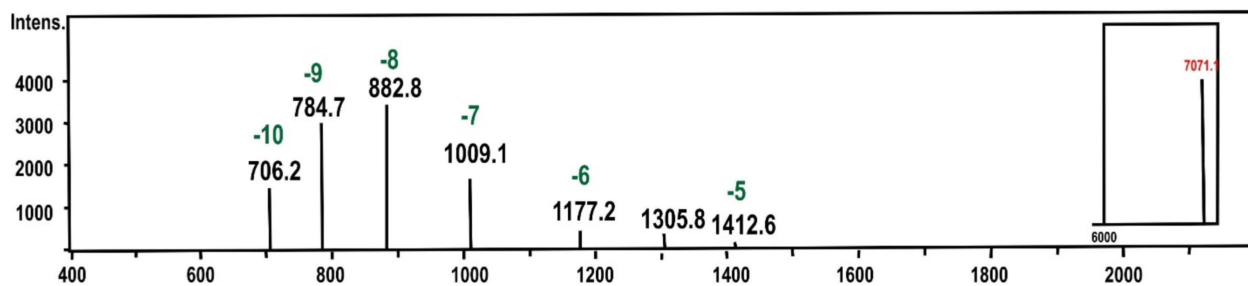


Figure S7. Mass-spectrum of 7-deaza-7-(2-oxoethyl)-dG containing DNA obtained upon oxidation of **DNA-2 (Table 1)**.

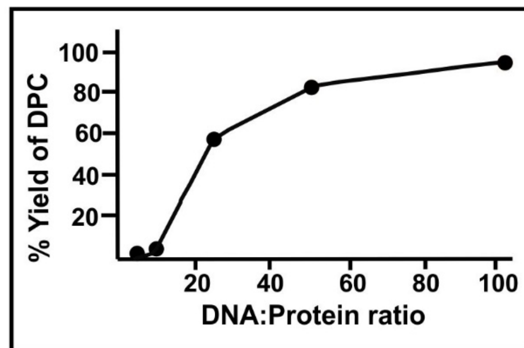


Figure S8. Concentration dependence study for DNA-protein cross-links with 5fdC containing oligodeoxynucleotide **DNA-1** and oxy-Lys containing protein **VHH-2**.

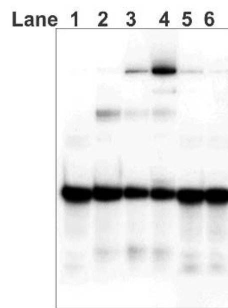


Figure S9. Denaturing PAGE analysis of DNA-protein cross-links of 7-oxoethyl-7-deaza-dG containing **DNA-2**. Lane 1. **DNA-4**. 2. **DNA-2**. 3. DPC of **VHH-1** and **DNA -2**. 4. DPC of **VHH-2** and **DNA-2**. 5. DPC of **VHH-1** and **DNA-4**. 6. DPC of **VHH-2** and **DNA-4**.

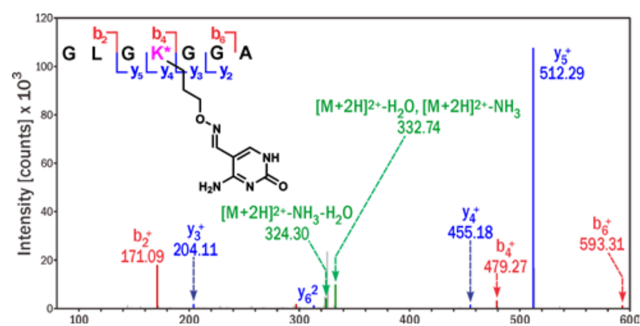


Figure S10. NanoLC-NSI$^+$-MS/MS spectrum of VHH7-2 tryptic peptide containing 5-formyl-dC cross-link to oxy-Lys. See SI for complete protein sequence.

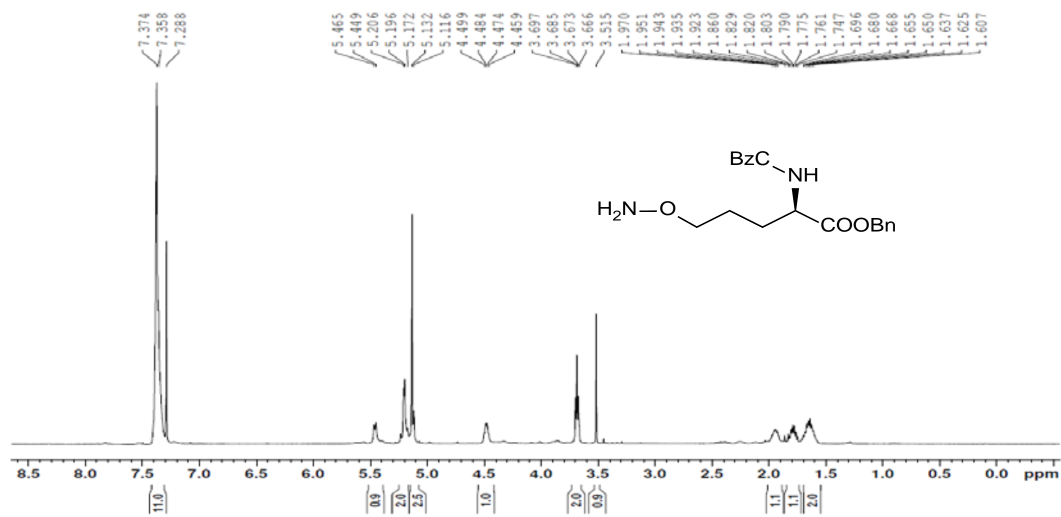


Figure S11. ^{1}H-NMR spectrum of compound **3**.

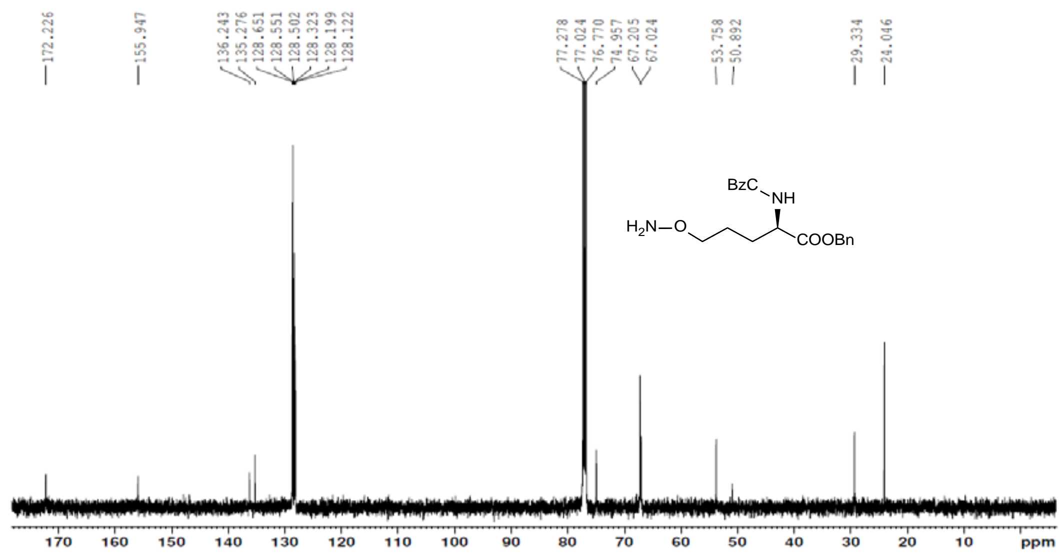


Figure S12. ^{13}C-NMR spectrum of compound **3**.

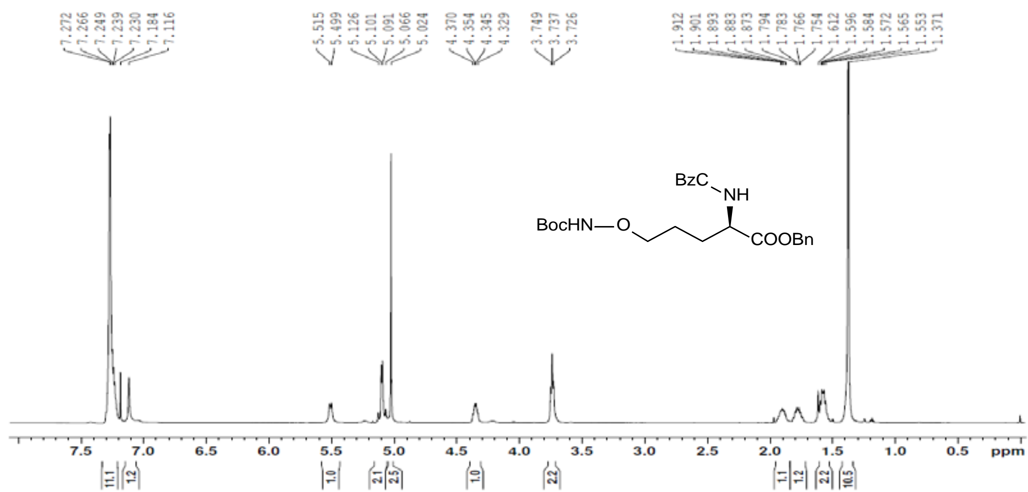


Figure S13. ^{1}H-NMR spectrum of compound **3a**.

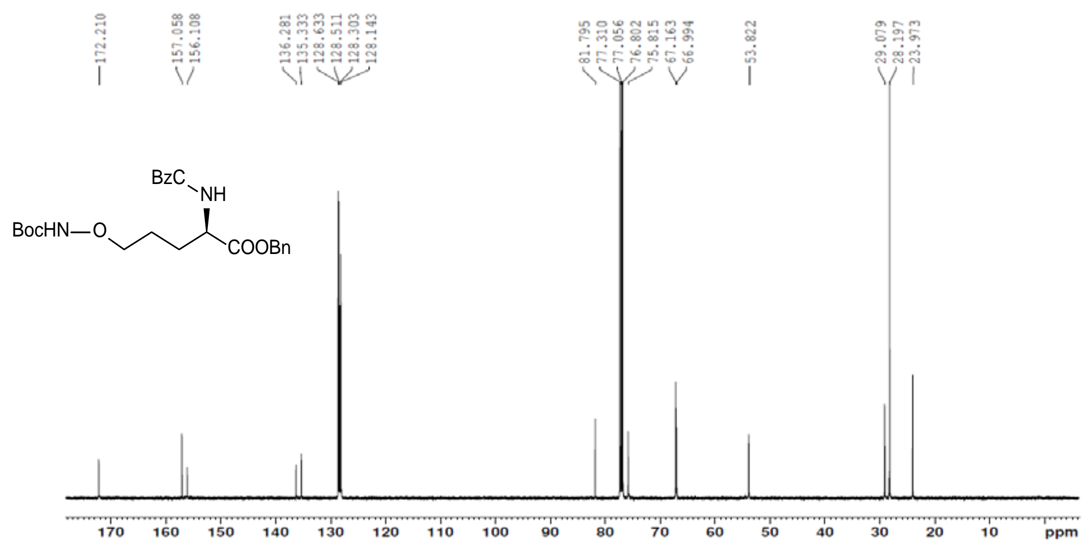


Figure S14. ^{13}C-NMR spectrum of compound **3a**.

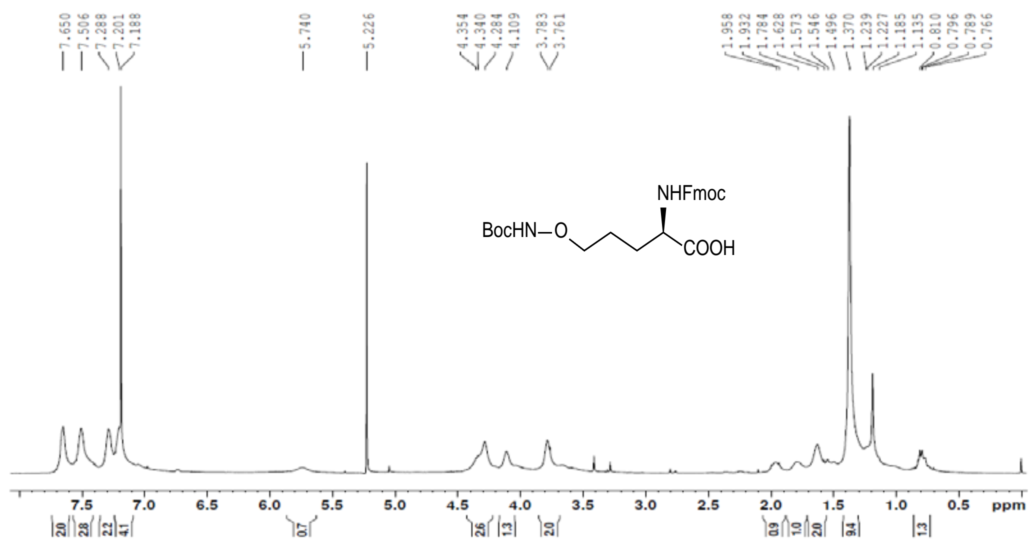


Figure S15. ^{1}H-NMR spectrum of compound **4**.

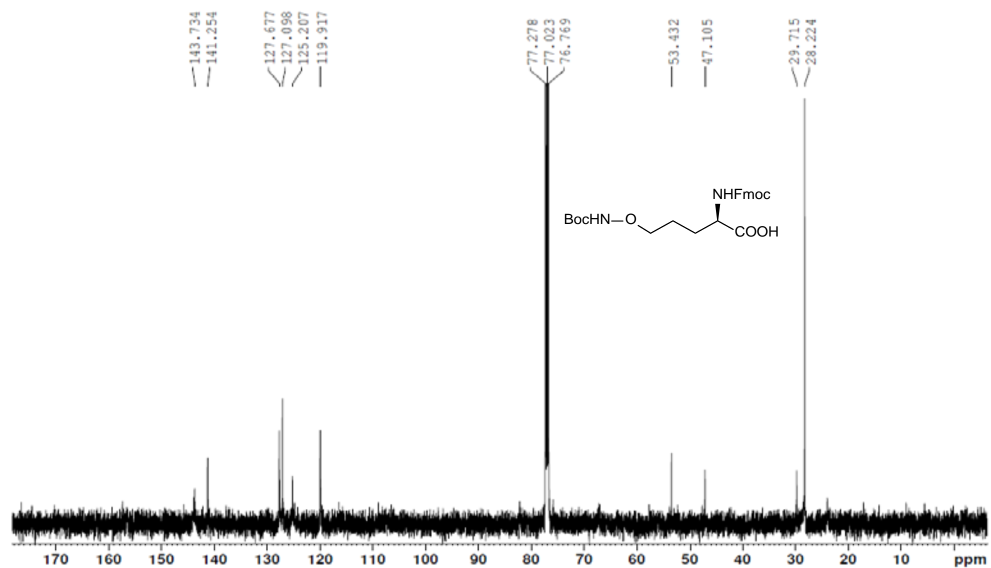


Figure S16. ^{13}C-NMR spectrum of compound **4**.