Electronic Supplementary Material (ESI) for ChemComm.

Electronic Supplementary Information

Sequencing for oxidative DNA damage at single-nucleotide resolution with click-code-seq v2.0

Songjun Xiao, Aaron M. Fleming, and Cynthia J. Burrows

Experimental Section

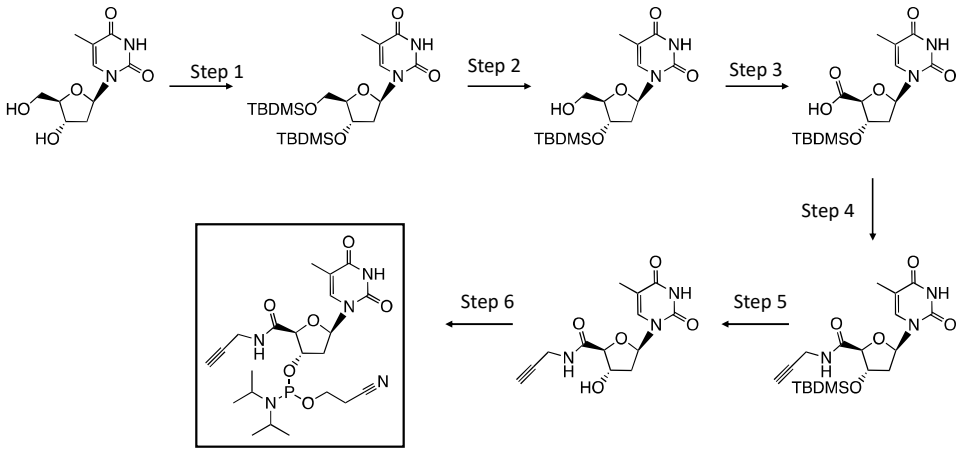
Materials and Instrumentation

All enzymes used in the study were purchased from New England Biolabs. All chemicals were obtained from commercial vendors and used without further purification. The 3′-azido-2′,3′-dideoxyguanosine-5′-triphosphate was obtained from TriLink Biotechnologies. The NMR experiments were conducted on Bruker 500-MHz instrument.

Phosphoramidite preparation

The 5′-alkynylated thymidine phosphoramidite was synthesized following an establish literature protocol.[1-3] The 6-step procedure is outlined in the scheme below. Briefly, the conditions for each step are as follows. Intermediates were verified by ^{1}H-NMR and the final product was more fully characterized with the data provided. Step 1. tert-butyldimethylsilyl chloride, imidazole. Step 2. pyridinium p-toluenesulfonate. Step 3. TEMPO, bis(acetoxy)iodobenzene, 1:1 MeCN/water. Step 4. hydroxybenzotriazole, DIPEA, propargyl amine, 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC). Step 5. NH_4F, CH_3OH. The final product before addition of the phosphoramidite group produced ^{1}H-NMR signals consistent with the literature.[2] ^{1}H-NMR (500 MHz, DMSO-d_6) δ 11.33 (1H, s, thymidine NH), 8.72 (1H, t, 5'-NH, *J* = 5.6 Hz), 8.05 (1H, s, H6), 6.33 (1H, dd, H1', *J* = 8.8, 5.7 Hz), 5.64 (1H, d, 3'-OH, *J* = 4.4 Hz), 4.33 (1H, tt, H3', *J* = 4.9, 2.0 Hz), 4.24 (1H, d, H4', *J* = 1.5 Hz), 3.92 (2H, dd, propargyl CH_2, *J* = 5.6, 2.5 Hz), 3.16 (1H, t, alkyne CH, *J* = 2.5 Hz), 2.16 (2H, m, H2'), and 1.79 (3H, s, CH_3) Step 6. 2-cyanoethyl-*N,N*-diisopropylchlorophosphoramidite, DIPEA. ^{31}P-NMR (202 MHz, DMSO-d_6) δ 147.92 and 147.65. ESI-MS calcd $[M+H]^+$ = 510.5; found 510.3

Scheme for the synthesis of 5′-alkynylated thymidine phosphoramidite



Oligonucleotide synthesis and purification

All DNA strands were synthesized by the DNA/Peptide core facility of the University of Utah following standard protocols. The site-specific introduction of dOG was achieved using a commercially available phosphoramidite (Glen Research). The 5′-alkyne modified DNA strand was synthesized using the 5′-alkynylated thymidine phosphoramidite generated in the previous step. The solid-phase synthetic protocol for the oligonucleotide was not changed from the standard protocol for this custom strand.

After synthesis, the crude DNA strands obtained from the core facility were purified using an analytical anion-exchange HPLC column (DNAPac PA100 4 x 250 mm). The mobile phases are as follows: A = 1:9 MeCN:ddH_2O; B = 1 M sodium chloride, 20 mM sodium acetate at pH 7 in 1:9 MeCN:ddH_2O. The method was initiated at 15% B and increased to 100% B via a linear gradient over 20 min at 1 mL/min while monitoring absorbance at 260 nm. The purified samples were dialyzed against ddH_2O for 1 d while changing the water three times. The dialyzed samples were lyophilized to dryness and then resuspended in a known volume of ddH_2O to make stock solution. The concentrations of the stock DNA samples were determined by measuring the absorbance at 260 nm and using an extinction coefficient calculated based on the primary sequencing using the nearest-neighbor method.

Workflow for PAGE analysis

The procedure for ^{32}P radiolabeling of the dOG-containing 30-mer DNA strand was achieved following a literature protocol.[4] The radiolabeled DNA was mixed with unlabeled dOG-containing 30-mer DNA in a 9:1 ratio (total = 50 pmol) followed by mixing the sample with an excess of the complementary strand (62.5 pmol) in 1x NEBuffer 2 to reach a final volume of 50 µL. The mixture was annealed by heating it to 90 °C for 5 min followed by slow cooling to room temperature. The annealed sample was preincubated at 37 °C for 5 min, followed by the addition of 1 µL of Fpg (8 U; NEB) and 1 µL of Endo IV (10 U; NEB) and the reaction was allowed to progress at 37 °C for 30 min. After the gap-forming reaction was completed, the sample was heated to 65 °C for 10 min to thermally denature the enzymes. Next, to the mixture was added 1 µL 3'-N_3-ddGTP (stock conc. = 10 mM) and 5 U (NEB) of Klenow Fragment (3'→5' exo$^-$) and the reaction was incubated at 37 °C for 60 min. The gap-filled reaction was purified with a Bio-Rad Micro Bio-Spin™ 6 Column following the manufacturer's protocol. As a note, the DNA was eluted with ddH_2O to avoid Tris buffer inhibiting the subsequent CuAAC reaction.[5] The CuAAC reaction was conducted by taking 600 pmol of the 5′-alkyne DNA and adding it to the elution mixture followed by lyophilization to dryness. To the residue was added 5 µL of DMSO, 2 µL of 1 M TEAA buffer (pH 7), 1.5 µL of ddH_2O, and 0.5 µL of preformed TBTA-$CuSO_4$ (1:1 TBTA:$CuSO_4$ at a conc = 10 mM in 55% DMSO and 45% ddH_2O). Next, 1 µL of freshly prepared sodium ascorbate solution (5 mM) was added to the mixture followed by vortexing. The reaction was allowed to react for 24 h at 22 °C. The mixture was then diluted with 90 µL of water and dialyzed overnight against ddH_2O to remove the organic solvent and Cu-TBTA complex. For the PAGE in Fig. 2, aliquots were withdrawn after each step and quenched with gel loading dye (0.025% xylene cyanol, 0.025% bromophenol blue in 90% DMF with 10 mM NaOH) before analysis on 20% denaturing PAGE.

Procedure for dOG site specific insertion into a plasmid

The psiCHECK2 plasmid used was previously modified to contain the 5-track *VEGF* potential G-quadruplex sequence that is flanked by Nt.BspQ1 nicking endonuclease recognition sequences.[6] First, the plasmid-containing *E. coli* (DH5α; NEB) were grown aerobically in LB media at 37 ° for 20 h to reach stationary phase. The cells were pelleted by centrifugation (6800 rpm for 10 min), and then the plasmids were extracted using a Qiagen miniprep kit following the manufacturer's protocol. Insertion of a site-specific dOG into the plasmid was achieved following a literature protocol from our lab.[6] Briefly, the reactions were performed on a thermal cycler unless specified. Specifically, to 5 µg of

plasmid in 50 µL of 1x NEBuffer™ r3.1 buffer was added Nt.BspQ1 (5 U; NEB). The reaction was incubated at 50 °C for 60 min followed by heating at 80 °C for 20 min to denature the enzyme. Next, 1 nmol of 5′-phosphorylated DNA with the site-specific dOG modification prepared by standard solid-phase synthesis was added to the reaction. The mixture was thermal cycled by heating at 85 °C for 3 min followed by cooling at 0 °C for 3 min, which was repeated four times. The nicks were then sealed using T4 DNA ligase, by adding 6 µL of 10x ligase buffer and T4 DNA ligase (800 U; NEB). The ligation reaction was left at 22 °C for 20 h. The plasmids containing dOG were cleaned up using GeneJET PCR Purification Kit following manufacturer's protocol (Thermo Scientific).

Workflow for validation of click-code-seq v2.0 on long duplex and plasmids

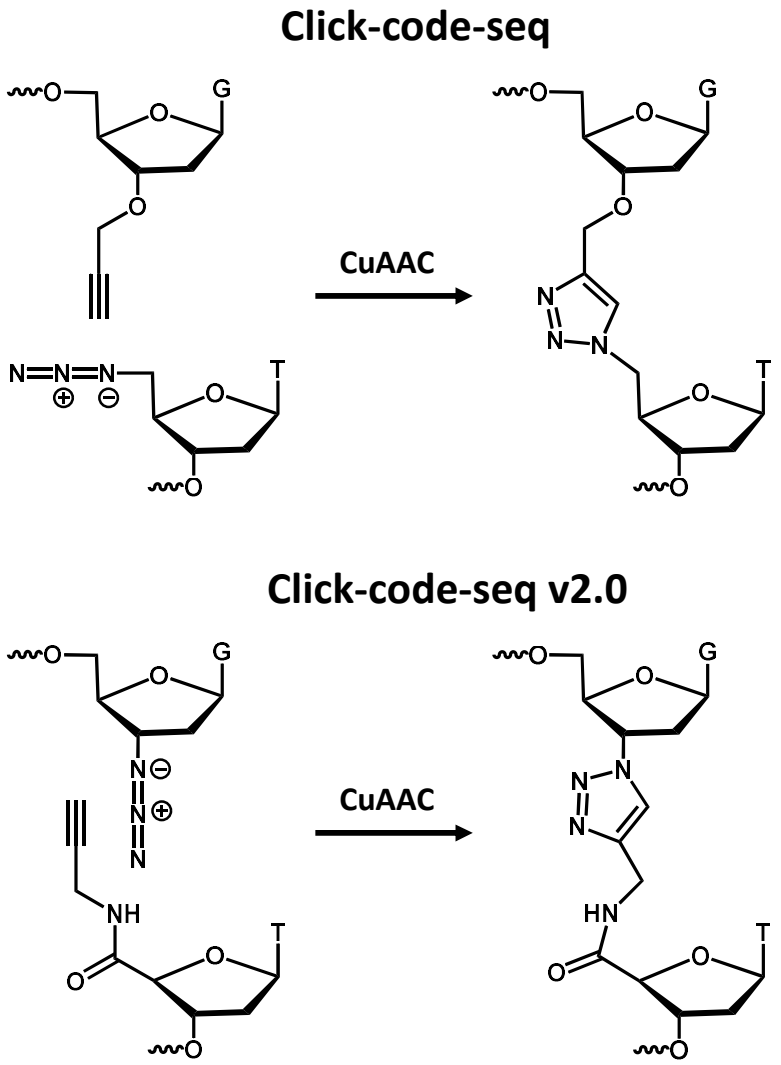
All reactions were performed on a thermal cycler unless specified.

To the pre-annealed duplex DNA (10 pmol) or plasmid (5 µg) each containing dOG at a known site were placed in a 50-µL reaction containing 1x NEBuffer™ 2, Fpg (8 U; NEB), Endo IV (10 U; NEB) followed by allowing the reaction to progress at 37 °C for 60 min. After the 60 min incubation the reaction was heated at 65 °C for 10 min to denature the enzymes. Next, 1 µL 3′-N_3-ddGTP (stock conc. = 10 mM) and Klenow Fragment (3'→5' exo$^-$; 5 U; NEB) were added to the reaction that was then incubated at 37 °C for 60 min. The reaction was purified using a GeneJET PCR Purification Kit following the manufacture's protocol (Thermo Scientific). After the reaction was purified, the 5′-alkynyl code DNA strand (600 pmol) was added and the mixture was lyophilized to dryness. To the residue remaining after lyophilization was added 5 µL of DMSO, 2 µL of 1 M TEAA buffer (pH 7), 1.5 µL of ddH_2O, and 0.5 µL of TBTA-$CuSO_4$ (1:1 TBTA:$CuSO_4$ at a stock conc. Of 10 mM, in 55% DMSO and 45% ddH_2O) followed by vortexing. To the mixture was added 1 µL of freshly prepared sodium ascorbate solution (5 mM) followed by vortexing. The reaction was allowed to react for 24 h at 22 °C. Following the CuAAC reaction, the mixture was diluted with 90 µL of ddH_2O and then purified using GeneJET PCR Purification Kit (Thermo Scientific) following the manufacturer's protocol.

The purified product with the 1,2,3-triazole linkage was then submitted to exponential PCR amplification. To a 50 µL reaction was added 10 µL of 5x Phusion GC Buffer (NEB), 1 µL of 10 mM stock dNTP mix (NEB), 2.5 µL of 10 µM Forward Primer, 2.5 µL of 10 µM Reverse Primer, 1.5 µL of DMSO, 0.5 µL of Phusion DNA Polymerase (NEB), 10-50 ng of the 1,2,3-triazole linked DNA, and the appropriate amount of ddH_2O to achieve the 50 µL final volume. The PCR thermal cycler procedure consisted of initial denaturation at 98 °C for 30 s, followed by 35 cycles of amplification. Each amplification cycle consisted of three steps: (1) denaturation at 98 °C for 30 s, (2) anneal at 55 °C for 30 s, and (3) extension at 72 °C for 1 min. The final extension was conducted at 72 °C for 10 min.

The PCR products were analyzed by agarose gel electrophoresis (0.5-3.5%) using 1x TAE buffer to confirm the desired length when compared against a molecular weight ladder. The resulting DNA was purified using a GeneJET PCR Purification Kit following the manufacturer's protocol (Thermo Scientific). The Sanger sequencing was conducted by the DNA Core Facility at the University of Utah using a BigDye Terminator v3.1 Cycle Sequencing Kit. The sequencing results were visualized using Snapgene software.

Figure S1. Structures for the different 1,2,3-triazole linkages



A comparison of the two different 1,2,3-triazole linkages in click-code-seq (top)[7] vs. click-code-seq v2.0 (bottom).

Figure S2. The sequences studied

Oligonucleotides used in this study

name	sequence
dOG-30-mer	5'-GCCACATCTTCAAT[OG]TATGCTACAAAAGAT
30-mer-comp	5'-ATCTTTTGTAGCATACATTGAAGATGTGGC
KRAS 3'-dOG	5'-GCTAATTCAGAATCATTTTGTGGACGAATATGATCCAACAATAGAGGCCTGCTGAAAATGACTGAATATAAACTTGTGGTAGTTGGAGCTG[OG]TGGCGTAGGCAAGAGTGCCTTGACGATAC
KRAS dOG-comp	5'-GTATCGTCAAGGCACTCTTGCCTACGCCACCAGCTCCAACTACCACAAGTTTATATTCAGTCATTTTCAGCAGGCCTCTATTGTTGGATCATATTCGTCCACAAAATGATTCTGAATTAGC
KRAS dOG-primer	5'-CAGAATCATTTTGTGGACGAATATG
Code DNA 1 (mass spec/gel)	5'-[alkyneT] GCCTACGCCAGGAGCTCC
Code DNA 2 (*KRAS* dOG)	5'-[alkyneT] ACCTCACCGCTTGGTTCGA
Code DNA 2 comp	5'-TCGAACCAAGCGGTGAGGT
VEGF OG pos1	5'-pGGGGCGGGCCGGGGGC[OG]GGGTCCCGGCGGGGCGCTCTT CT
DB-code-2	5'-[alkyneT] GACTCACCTAGGTCTGACTG [3'-desthiobiotin]
DB-code-4	5'-[alkyneT] ACCTCACCGCTTGGTTCGA [3'-desthiobiotin]
DB-code-2 comp polyT	5'-TTTTTTTTTTTTTTTTTTTTCAGTCAGACCTAGGTGAGTC
plasmid-f-primer	5'- CTTTATAGTCCTGTCGGGTTTCGC

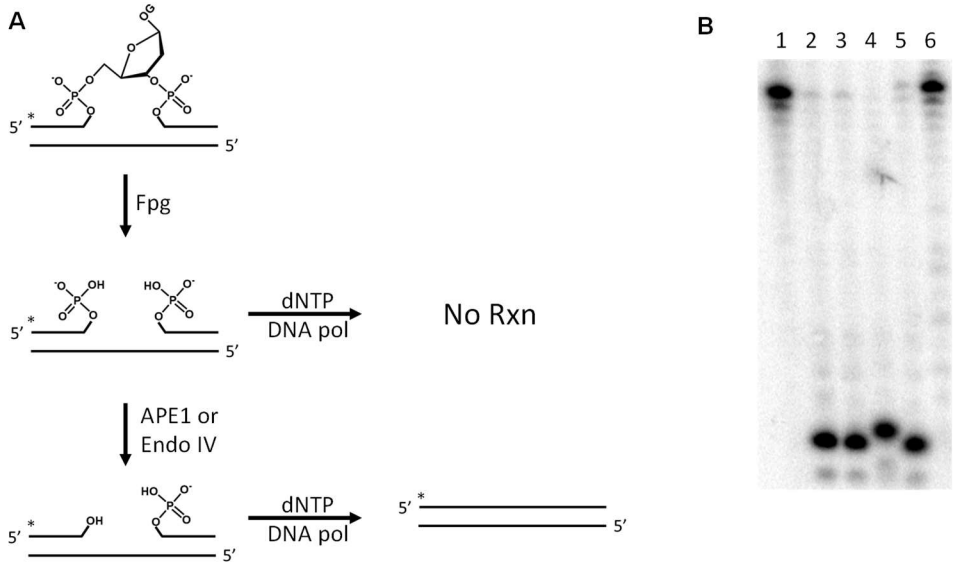
Plasmid sequence with 5 track *VEGF* inserted (sequence highlighted in red and dOG locations are in blue)

5'-
AGATCTGCGCAGCACCATGGCCTGAAATAACCTCTGAAAGAGGAACTTGGTTAGGT
ACCTTCTGAGGCGGAAAGAACCAGCTGTGGAATGTGTGTCAGTTAGGGTGTGGAA
AGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCA
GCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCA
TGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCC
CTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGAC**GCTCTTCCGGGG**
CGGGCCGGGGGCGGGGTCCCGGCGGGGCGCTCTTCTGCAGAGGCCGAGGCCGC
CTCGGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCT
TTTGCAAAAAGCTTGATTCTTCTGACACAACAGTCTCGAACTTAAGCTGCAGAAGTT
GGTCGTGAGGCACTGGGCAGGTAAGTATCAAGGTTACAAGACAGGTTTAAGGAGA
CCAATAGAAACTGGGCTTGTCGAGACAGAGAAGACTCTTGCGTTTCTGATAGGCAC
CTATTGGTCTTACTGACATCCACTTTGCCTTTCTCTCCACAGGTGTCCACTCCCAGT
TCAATTACAGCTCTTAAGGCTAGAGTACTTAATACGACTCACTATAGGCTAGCCACC
ATGGCTTCCAAGGTGTACGACCCCGAGCAACGCAAACGCATGATCACTGGGCCTC
AGTGGTGGGCTCGCTGCAAGCAAATGAACGTGCTGGACTCCTTCATCAACTACTAT
GATTCCGAGAAGCACGCCGAGAACGCCGTGATTTTTCTGCATGGTAACGCTGCCTC
CAGCTACCTGTGGAGGCACGTCGTGCCTCACATCGAGCCCGTGGCTAGATGCATC
ATCCCTGATCTGATCGGAATGGGTAAGTCCGGCAAGAGCGGGAATGGCTCATATC
GCCTCCTGGATCACTACAAGTACCTCACCGCTTGGTTCGAGCTGCTGAACCTTCCA
AAGAAAATCATCTTTGTGGGCCACGACTGGGGGGCTTGTCTGGCCTTTCACTACTC
CTACGAGCACCAAGACAAGATCAAGGCCATCGTCCATGCTGAGAGTGTCGTGGAC
GTGATCGAGTCCTGGGACGAGTGGCCTGACATCGAGGAGGATATCGCCCTGATCA
AGAGCGAAGAGGGCGAGAAAATGGTGCTTGAGAATAACTTCTTCGTCGAGACCATG
CTCCCAAGCAAGATCATGCGGAAACTGGAGCCTGAGGAGTTCGCTGCCTACCTGG
AGCCATTCAAGGAGAAGGGCGAGGTTAGACGGCCTACCCTCTCCTGGCCTCGCGA
GATCCCTCTCGTTAAGGGAGGCAAGCCCGACGTCGTCCAGATTGTCCGCAACTAC
AACGCCTACCTTCGGGCCAGCGACGATCTGCCTAAGATGTTCATCGAGTCCGACC
CTGGGTTCTTTTCCAACGCTATTGTCGAGGGAGCTAAGAAGTTCCCTAACACCGAG
TTCGTGAAGGTGAAGGGCCTCCACTTCAGCCAGGAGGACGCTCCAGATGAAATGG
GTAAGTACATCAAGAGCTTCGTGGAGCGCGTGCTGAAGAACGAGCAGTAATTCTAG
GCGATCGCTCGAGCCCGGGAATTCGTTTAAACCTAGAGCGGCCGCTGGCCGCAAT
AAAATATCTTTATTTTCATTACATCTGTGTGTTGGTTTTTTGTGTGAGGATCTAAATG
AGTCTTCGGACCTCGCGGGGGCCGCTTAAGCGGTGGTTAGGGTTTGTCTGACGCG
GGGGGAGGGGGAAGGAACGAAACACTCTCATTCGGAGGCGGCTCGGGGTTTGGT
CTTGGTGGCCACGGGCACGCAGAAGAGCGCCGCGATCCTCTTAAGCACCCCCCC
GCCCTCCGTGGAGGCGGGGGTTTGGTCGGCGGGTGGTAACTGGCGGGCCGCTGA
CTCGGGCGGGTCGCGCGCCCCAGAGTGTGACCTTTTCGGTCTGCTCGCAGACCCC
CGGGCGGCGCCGCCGCGGCGGCGACGGGCTCGCTGGGTCCTAGGCTCCATGGG
GACCGTATACGTGGACAGGCTCTGGAGCATCCGCACGACTGCGGTGATATTACCG
GAGACCTTCTGCGGGACGAGCCGGGTCACGCGGCTGACGCGGAGCGTCCGTTGG
GCGACAAACACCAGGACGGGGCACAGGTACACTATCTTGTCACCCGGAGGCGCGA
GGGACTGCAGGAGCTTCAGGGAGTGGCGCAGCTGCTTCATCCCCGTGGCCCGTT
GCTCGCGTTTGCTGGCGGTGTCCCCGGAAGAAATATATTTGCATGTCTTTAGTTCTA

TGATGACACAAACCCCGCCCAGCGTCTTGTCATTGGCGAATTCGAACACGCAGATG
CAGTCGGGGCGGCGCGGTCCCAGGTCCACTTCGCATATTAAGGTGACGCGTGTGG
CCTCGAACACCGAGCGACCCTGCAGCGACCCGCTTAAAAGCTTGGCATTCCGGTA
CTGTTGGTAAAGCCACCATGGCCGATGCTAAGAACATTAAGAAGGGCCCTGCTCCC
TTCTACCCTCTGGAGGATGGCACCGCTGGCGAGCAGCTGCACAAGGCCATGAAGA
GGTATGCCCTGGTGCCTGGCACCATTGCCTTCACCGATGCCCACATTGAGGTGGA
CATCACCTATGCCGAGTACTTCGAGATGTCTGTGCGCCTGGCCGAGGCCATGAAG
AGGTACGGCCTGAACACCAACCACCGCATCGTGGTGTGCTCTGAGAACTCTCTGC
AGTTCTTCATGCCAGTGCTGGGCGCCCTGTTCATCGGAGTGGCCGTGGCCCCTGC
TAACGACATTTACAACGAGCGCGAGCTGCTGAACAGCATGGGCATTTCTCAGCCTA
CCGTGGTGTTCGTGTCTAAGAAGGGCCTGCAGAAGATCCTGAACGTGCAGAAGAA
GCTGCCTATCATCCAGAAGATCATCATCATGGACTCTAAGACCGACTACCAGGGCT
TCCAGAGCATGTACACATTCGTGACATCTCATCTGCCTCCTGGCTTCAACGAGTAC
GACTTCGTGCCAGAGTCTTTCGACAGGGACAAAACCATTGCCCTGATCATGAACAG
CTCTGGGTCTACCGGCCTGCCTAAGGGCGTGGCCCTGCCTCATCGCACCGCCTGT
GTGCGCTTCTCTCACGCCCGCGACCCTATTTTCGGCAACCAGATCATCCCCGACAC
CGCTATTCTGAGCGTGGTGCCATTCCACCACGGCTTCGGCATGTTCACCACCCTGG
GCTACCTGATTTGCGGCTTTCGGGTGGTGCTGATGTACCGCTTCGAGGAGGAGCT
GTTCCTGCGCAGCCTGCAAGACTACAAAATTCAGTCTGCCCTGCTGGTGCCAACCC
TGTTCAGCTTCTTCGCTAAGAGCACCCTGATCGACAAGTACGACCTGTCTAACCTG
CACGAGATTGCCTCTGGCGGCGCCCCACTGTCTAAGGAGGTGGGCGAAGCCGTG
GCCAAGCGCTTTCATCTGCCAGGCATCCGCCAGGGCTACGGCCTGACCGAGACAA
CCAGCGCCATTCTGATTACCCCAGAGGGCGACGACAAGCCTGGCGCCGTGGGCAA
GGTGGTGCCATTCTTCGAGGCCAAGGTGGTGGACCTGGACACCGGCAAGACCCTG
GGAGTGAACCAGCGCGGCGAGCTGTGTGTGCGCGGCCCTATGATTATGTCCGGCT
ACGTGAATAACCCTGAGGCCACAAACGCCCTGATCGACAAGGACGGCTGGCTGCA
CTCTGGCGACATTGCCTACTGGGACGAGGACGAGCACTTCTTCATCGTGGACCGC
CTGAAGTCTCTGATCAAGTACAAGGGCTACCAGGTGGCCCCAGCCGAGCTGGAGT
CTATCCTGCTGCAGCACCCTAACATTTTCGACGCCGGAGTGGCCGGCCTGCCCGA
CGACGATGCCGGCGAGCTGCCTGCCGCCGTCGTCGTGCTGGAACACGGCAAGAC
CATGACCGAGAAGGAGATCGTGGACTATGTGGCCAGCCAGGTGACAACCGCCAAG
AAGCTGCGCGGCGGAGTGGTGTTCGTGGACGAGGTGCCCAAGGGCCTGACCGGC
AAGCTGGACGCCCGCAAGATCCGCGAGATCCTGATCAAGGCTAAGAAAGGCGGCA
AGATCGCCGTGTAATAATTCTAGAGTCGGGGCGGCCGGCCGCTTCGAGCAGACAT
GATAAGATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATG
CTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATA
AACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGGTGT
GGGAGGTTTTTTAAAGCAAGTAAAACCTCTACAAATGTGGTAAAATCGATAAGGATC
CAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAA
TACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATA
TTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTT
GCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGAT
GCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCG
GTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTA
AAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTC
GGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGA
AAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCA

TGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGA
GCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGG
AACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGT
AGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTC
CCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGC
GCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCG
TGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATC
GTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGAT
CGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACT
CATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAG
ATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAG
CGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCG
TAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCG
GATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGAT
ACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGT
AGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTG
GCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCG
CAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACG
ACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCC
CGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGA
GCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGG
TTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAG
CCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGC
CTTTTGCTCACATGGCTCGAC

Figure S3. Optimization of the endonuclease reaction



A PAGE-based reaction analysis was devised to monitor the efficiency of phosphate removal from the 5′ fragment for APE1 vs. Endo IV (panel A for the scheme). The dOG-containing duplex DNA with a 5′-^{32}P label (*) was analyzed by PAGE (panel B lane 1). This duplex was cleaved with Fpg to yield a faster migrating strand that retained a 3′ phosphate because this glycosylase catalzyes the β,δ-elimination of the phosphates (panel B lane 2). The cleaved strand was then treated with APE1 (panel B lane 3) or Endo IV (panel B lane 4) for removal of the 3′ phosphate from the 5′-radiolabeled fragment. When the 3′ phosphate is removed the strand migrates a little slower. Confirmation that the 3′ phosphate has been removed was achieved by treating the strand with Kf exo$^-$ and a dNTP mix (panel B lane 5 = APE1 and lane 6 = Endo IV). If the phosphate is removed, the polymerase can extend the fragmented strand back to full length, otherwise, the 3′ phosphate blocks the polymerization reaction. This approach found Endo IV catalyzed nearly quantitative removal of the 3′ phosphate while APE1 gave yields <10%.

Figure S4. Characterization data for the 5′-alkynyl-dT phosphoramidite

^{31}P-NMR

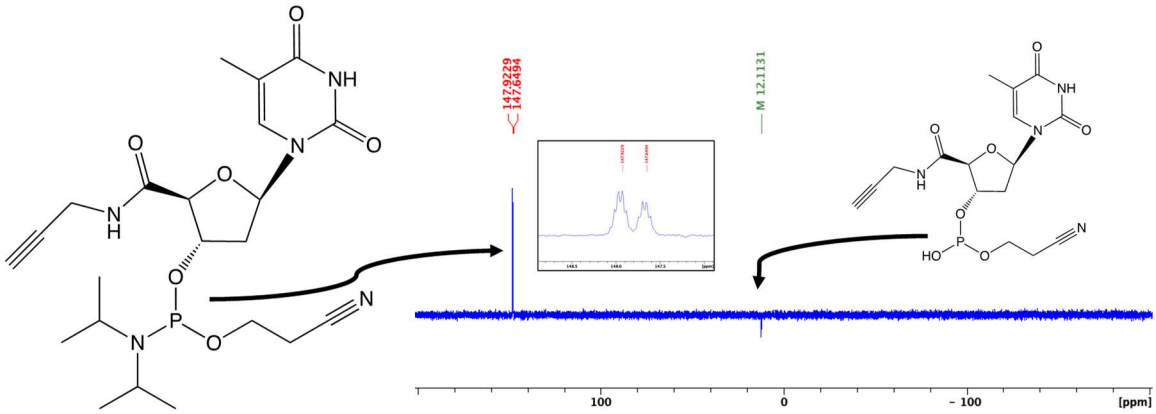
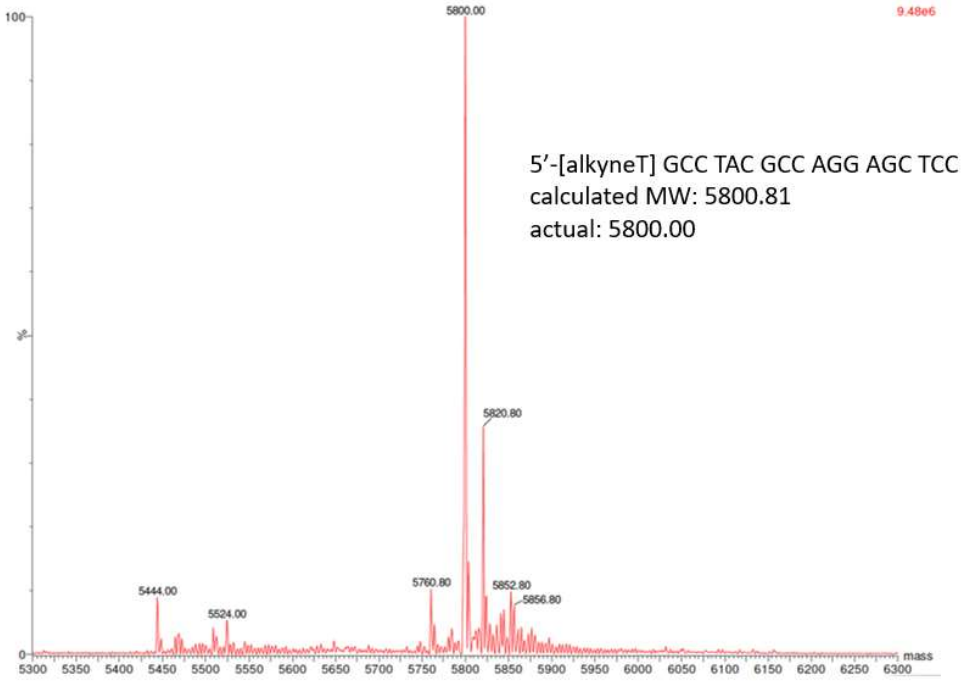


Figure S5. Characterization data for the 5′-alkynyl-code DNA strand

The 5′-alkynyl strand used in the studies found in Fig. 2.



The 5′-alkynyl strand DB=code-4.

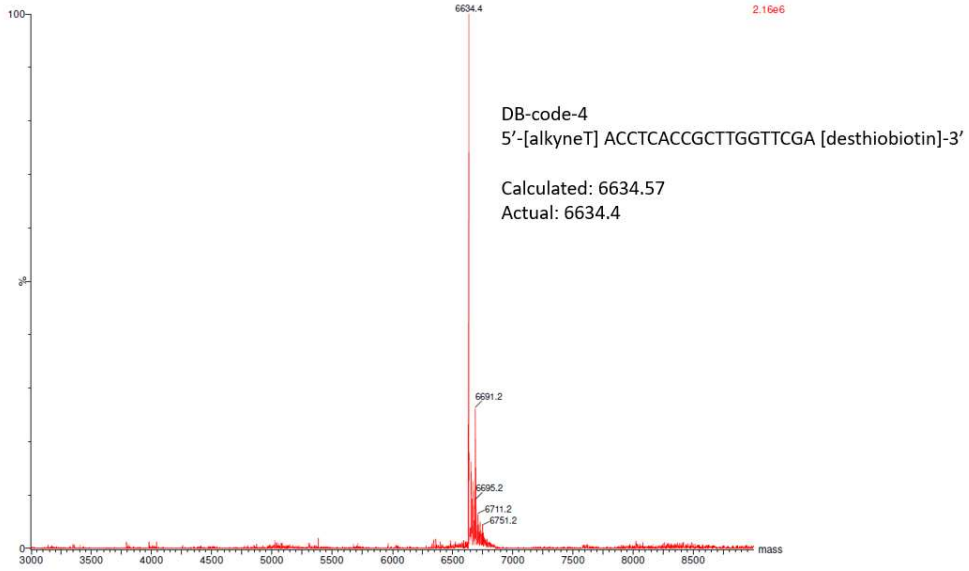
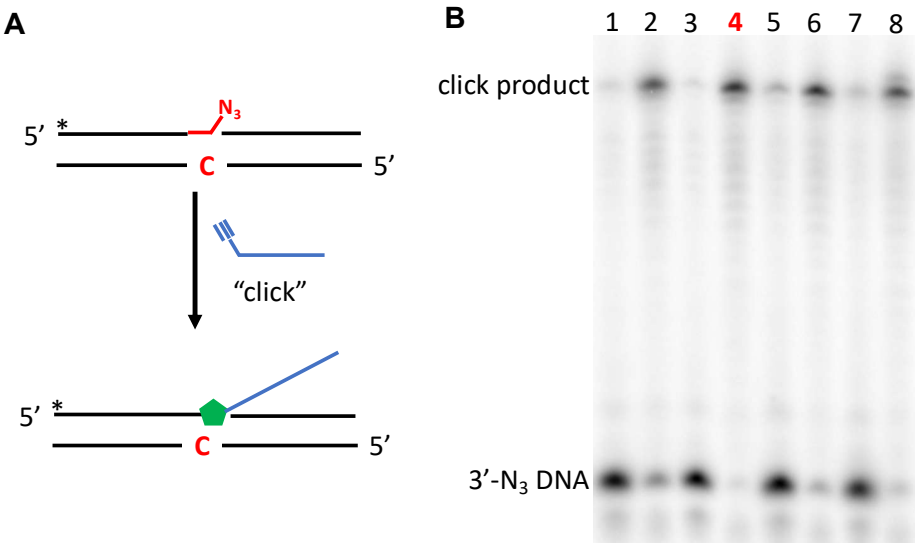


Figure S6. Optimization of the click reaction conditions



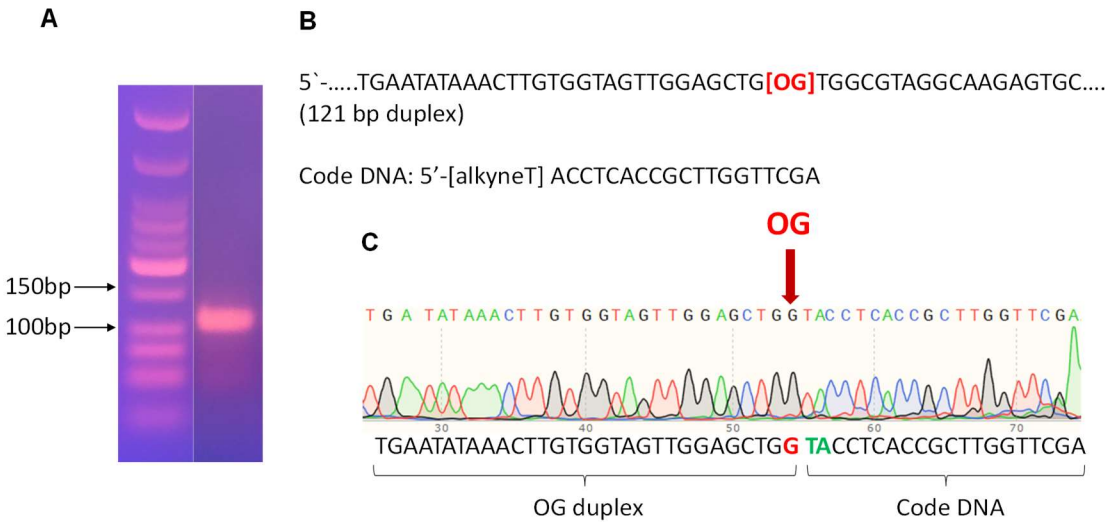
C

	1	2	3	4	5	6	7	8
DMSO%	5%	55%	5%	55%	5%	55%	5%	55%
temperature(°C)	22	22	22	22	37	37	37	37
time	1h	1h	1d	1d	1h	1h	1d	1d
Click yield	10%	70%	6%	94%	16%	85%	13%	87%

The reaction monitored is illustrated in panel A. The denaturing PAGE analysis is shown in panel B for each of the reaction conditions outlined in panel C. In panel C, are the final DMSO concentration, reaction temperature, reaction time, and yields.

The CuAAC reaction was conducted by taking 600 pmol of the 5′-alkyne DNA and adding it to the elution mixture followed by lyophilization to dryness. To the residue was added 5 µL of DMSO, 2 µL of 1 M TEAA buffer (pH 7), 1.5 µL of ddH_2O, and 0.5 µL of preformed TBTA-$CuSO_4$ (1:1 TBTA:$CuSO_4$ at a conc = 10 mM in 55% DMSO and 45% ddH2O). Next, 1 µL of freshly prepared sodium ascorbate solution (5 mM) was added to the mixture followed by vortexing. The reaction was allowed to react for 24 h at 22 °C. The mixture was then diluted with 90 µL of water and dialyzed overnight against ddH_2O to remove the organic solvent and Cu-TBTA complex before 20% denaturing PAGE analysis.

Figure S7. Application of click-code-seq v2.0 to long DNA



The data provided illustrate application of click-code-seq v2.0 on a 121-bp synthetic DNA with dOG at a known position. The agarose gel analysis in panel A verifies success in the click-code-seq v2.0 reaction after the final PCR step. In panel B are the sequences for the target dOG-containing DNA strand and the 5′-alkynyl code DNA strand that replaced dOG for its identification. Panel C is a Sanger chromatogram that verifies success in replacing dOG with the code sequence.

References

(1) Zhang, W.; Robins, M. J. Removal of silyl protecting groups from hydroxyl functions with ammonium fluoride in methanol. *Tetrahedron Lett.* **1992**, *33*, 1177-1180.

(2) El-Sagheer, A. H.; Brown, T. Synthesis and polymerase chain reaction amplification of DNA strands containing an unnatural triazole linkage. *Journal of the American Chemical Society* **2009**, *131*, 3958-3964.

(3) Montevecchi, P. C.; Manetto, A.; Navacchia, M. L.; Chatgilialoglu, C. Thermal decomposition of the tert-butyl perester of thymidine-5′-carboxylic acid. Formation and fate of the pseudo-C4′ radical. *Tetrahedron* **2004**, *60*, 4303-4308.

(4) Fleming, A. M.; Burrows, C. J. G-quadruplex folds of the human telomere sequence alter the site reactivity and reaction pathway of guanine oxidation compared to duplex DNA. *Chem. Res. Toxicol.* **2013**, *26*, 593-607.

(5) Hong, V.; Presolski, S. I.; Ma, C.; Finn, M. G. Analysis and optimization of copper-catalyzed azide-alkyne cycloaddition for bioconjugation. *Angew Chem Int Ed Engl* **2009**, *48*, 9879-9883.

(6) Fleming, A. M.; Ding, Y.; Burrows, C. J. Oxidative DNA damage is epigenetic by regulating gene transcription via base excision repair. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 2604-2609.

(7) Wu, J.; McKeague, M.; Sturla, S. J. Nucleotide-resolution genome-wide mapping of oxidative DNA damage by click-code-seq. *J. Am. Chem. Soc.* **2018**, *140*, 9783-9787.