TABLE OF CONTENTS

Materials and Experimental Methods

Warning: 1,2-Dibromoethane is toxic and a known carcinogen. It should be handled with extreme caution and protection should be worn at all times while handling it.

Materials. Unless otherwise mentioned, all materials were obtained from Sigma-Aldrich (St. Louis, MO) and were of the highest purity available. Oligonucleotides were purchased from Midland Certified Reagent Co. (Midland, TX). Human AGT was heterogeneously expressed in *Escherichia coli* and purified as described elsewhere.[1,2]

Formation of DPC by AGT and DBE in oligonucleotides and calf thymus DNA. In a typical reaction, 50 mg of sonicated calf thymus DNA in 50 mM potassium phosphate buffer (pH 7.7) containing 5 mM EDTA, 10 nmol of AGT, and 10 mM DBE (in case of ^{14}C-DBE, the specific activity was 6.5 μCi/μmol) was incubated for 3 h at 37 °C. (Note: turbidity may be seen in the DNA+DBE+AGT samples due to precipitation of alkylated AGT.

In case of oligonucleotides, 4 nmol of double-stranded 15-mer GC-rich (5-ACCCGCGTCCGCGCC) or 12-mer $(AT)_6$ oligonucleotide in 50 mM potassium phosphate buffer (pH 7.7) containing 5 mM EDTA, 2 nmol of AGT, and 20 mM DBE was incubated for 3 h at 37 °C.

Detection of DPCs in GC- and AT-rich oligonucleotides. After reaction of the oligonucleotides with AGT and DBE, the samples were loaded onto an Amicon filter (M_r 3,000 cutoff) and centrifuged at 3,000 × *g* for 30 min. After the initial filtration, 200 μL of NH_4HCO_3 buffer (pH 8.0) was added and the samples were centrifuged again. This process was repeated twice to remove excess unreacted DBE. The residue remaining on the filter was then dissolved in 200 μL of NH_4HCO_3 buffer (pH 7.5) containing 0.01% SDS (w/v) and 2.5 mM dithiothreitol (DTT) and heated at 90 °C for 20 min. The solution was cooled and 2 μg of trypsin was added and incubated for 16 h at 37 °C. After incubation the samples were dried and redissolved in 100 μL of H_2O and analyzed by LC-MS following a previously published protocol.[2] In the case of the piperidine workup, samples were dissolved in 200 mM piperidine and heated at 90 °C for 20 min. The samples were subsequently dried in a centrifugal lyophilizer and again dissolved in 50 μL of H_2O and dried. This process was repeated twice prior to LC-MS analysis.

LC-MS analysis of samples was performed in an ESI linear ion trap mass spectrometer (LTQ, Thermo-Fisher, San Jose, CA) connected to a Waters Acquity UPLC system (Waters, Milford, MA) and using an Acquity UPLC BEH octadodecylsilane (C_{18}) column (1.7 μm, 1.0 mm × 100 mm). DNA-peptide crosslinks were separated using buffer A (containing 10 mM $NH_4CH_3CO_2$ in 2% CH_3CN, v/v) and buffer B (containing 10 mM $NH_4CH_3CO_2$ in 95% CH_3CN v/v) following a gradient program with a flow rate of 150 μL min-1: 0-3.0 min, linear gradient from 100% A to 97% A; 3.0-4.5 min, linear gradient to 80% A; 4.5-5.0 min, linear gradient 100% B; 5.0-5.5 min, hold at 100% B; 5.5-6.5 min, linear gradient to 100% A; 6.5-9.5 min, hold at 100% A. The temperature of the column was maintained at 50 °C and samples (10 μL)

were infused with an auto-sampler. ESI conditions were as follows: source voltage 4 kV, source current 100 μA, auxiliary gas flow rate setting 20, sweep gas flow rate setting 5, sheath gas flow setting 34, capillary voltage -49 V, capillary temperature 350 °C, tube lens voltage -90 V. MS/MS conditions were as follows: normalized collision energy 35%, activation Q 0.250, activation time 30 ms. The quadruply-charged species (*m/z* 1452.6 and 1244.5) were used for CID analysis. The *m/z* values of the CID fragments corresponding to the peptide-DNA crosslinks were calculated using a program linked to the Mass Spectrometry Group of the Department of Medicinal Chemistry at the University of Utah (medlib.med.utah.edu/masspec).

Efficiency of Raney nickel-mediated desulfurization. *S*-(2-Chloroethyl)GSH was prepared as described,[3] and 10 equiv was incubated with 2′-deoxyguanosine or 2′-deoxyadenosine in $H_2O/(CH_3)_2SO$ (1:1, v/v) for 2 h at 37 °C. Reactions were subsequently heated for 30 min at 90 °C under neutral conditions to release some of the modified bases (i.e., *N*7-alkyl dG). The products were separated by HPLC as previously described,[4] and *S*-[2-(*N*7-guanyl)ethyl]GSH, *S*-[2-(*N*1-deoxyadenosyl)ethyl]GSH, and *S*-[2-(N^6-deoxyadenosyl)ethyl]GSH adducts were identified by UV spectroscopy[4-6] and LC-MS/MS.

The *N*7-EtG, *N*1-EtdA, and N^6-EtdA adducts obtained after Raney nickel reaction by desulfurization of ethylene-GSH adducts (*vide infra*) were analyzed using [^{18}O]-*N*7-G-$(OH)_3$butane[3] as an internal standard (ISTD) by LC-MS/MS. LC-MS/MS analysis was performed using a Waters Acquity UPLC system (Waters) interfaced to a Thermo-Finnigan LTQ Orbitrap mass spectrometer (ThermoElectron) equipped with an ESI source. Chromatographic separation was achieved with a Waters Acquity UPLC BEH C18 octadecylsilane column (2.1 mm × 100 mm, 1.7 μm). LC conditions were as follows: the mobile phase A was 0.1% CH_3CO_2H in H_2O (v/v) and mobile phase B was 0.1% CH_3CO_2H in CH_3CN (v/v). The following gradient program (v/v) was used with a flow rate of 300 μL/min: the gradient started with 5% B (v/v), increased to 15% B (v/v) at 2 min, to 30% B (v/v) at 6 min, and held at 30% B (v/v) for 1 min. The column was re-equilibrated for 3 min with 5% B (v/v). The temperature of the column was maintained at 40 °C. The MS conditions were as follows: ion spray voltage, 4.5 kV; capillary voltage, 20 V; capillary temperature, 350 °C; tube lens voltage, 40 V.

Detection and quantitation of DPCs formed with calf thymus DNA. Following DPC formation (*vide supra*), the reaction mixtures were subjected to C_2H_5OH precipitation using 70% cold C_2H_5OH (v/v) and 0.3 M sodium acetate. The precipitate was pelleted by centrifugation at 10,000 × *g*, washed three times with cold 80% C_2H_5OH to remove unreacted excess DBE, and redissolved in 50 mM Tris buffer (pH 7.4) containing 1 mM $CaCl_2$. In case of the radioactive samples, 2 μL (from a total volume of 200 μL) of the reaction mixture was used for scintillation counting after adding 4 mL of scintillation cocktail. The reaction mixtures were digested with proteinase K for 3 h (to avoid depurination of the labile adducts) and C_2H_5OH-precipitated to remove the digested peptides. The precipitate were dissolved in 50 mM potassium phosphate buffer (pH 7.4), heated at 90 °C for 30 min to cause depurination, and again C_2H_5OH-precipitated. The supernatant (following centrifugation) was separated and 10 μL (of a total volume 1.0 mL) was counted (in a liquid scintillation counter) to quantify labile adducts.

The precipitate was dissolved in 50 mM potassium phosphate buffer (200 μL total volume) and 2 μL was subjected to scintillation counting (for non-labile adducts). The supernatant (containing the labile crosslinks) was dried under a stream of N_2 and redissolved in 50 mM potassium phosphate buffer (pH 7.4, 200 μL total volume). Both the solutions (labile and non-labile) were separately treated with Raney Ni (5 μL) at 60 °C for 30 min. The reactions were cooled, centrifuged to remove the Ni, and purified using a C_{18} SPE column (Agilent, 3 mL). In the case of labile modifications, the reaction mixtures were analyzed by HPLC-flow counting. For the non-labile DPCs, the SPE-purified products were digested with nucleases and phosphatases following standard literature procedures.[7]

After digestion, the reaction mixture was passed through an Amicon filter (M_r 3,000 cutoff) and analyzed by LC-MS/MS and LC-MS^3. In case of ^{14}C-DBE, the nucleoside-ethyl adducts were HPLC-purified after spiking with authentic standards and monitoring UV absorbance (260 nm). Radioactivity in the fractions corresponding to the various ethyl dG and dA adducts was counted using a scintillation counter for quantitation.

For LC-MS analysis, an ESI linear trap ion mass spectrometer (LTQ, Thermo-Fisher, San Jose, CA) was used, connected to a Waters Acquity UPLC system (Waters, Milford, MA) with an Acquity UPLC BEH octadodecylsilane (C_{18}) column (1.7 μm, 2.1 mm × 100 mm). DNA- peptide crosslinks were separated using solvent A (containing 0.1% HCO_2H and 5% CH_3CN, v/v) and solvent B (containing 0.1% HCO_2H and 95% CH_3CN, v/v) following a gradient program with a flow rate of 300 μL min^{-1}: 0-5.0 min, linear gradient from 100% A to 75% A (v/v); 5.0-5.5 min, linear gradient to 100% B; 5.5-6.5 min, hold at 100% B (v/v); 6.5-7.0 min, linear gradient to 100% A (v/v); 7-10 min, hold at % A (v/v). The temperature of the column was maintained at 40°C and samples (20 μL) were infused with an auto-sampler. The mass spectrometer was tuned using authentic N^2-EtdG.

Figure S1. ESI-LC-MS and ESI-LC-MS/MS chromatograms and spectra of an (AGT+DBE)-crosslinked GC-rich oligonucleotide fragment. AGT (10 μM) + DBE (20 mM)-modified DNA fragment (20 μM) was digested with trypsin (2 μg) at 37 °C for 16 h, heated at 90 °C for 30 min, and analyzed by LC-MS. After removal of unbound DBE, it was heated at 90 °C for 15 min, digested with trypsin, and subjected to LC-MS analysis in the negative ESI mode. The LC chromatogram (Figure S1b) showed a collection of peaks at retention time (t_R) ~6 min with *m/z* values of 1452.6 and 1937.5 that corresponds to the -4 and -3 charged species of the 12-mer tryptic peptide (GNPVPILIPCHR) of AGT covalently linked to the 15-mer GC-rich oligonucleotide (5'-ACCCGCGTCCGCGCC) through an ethylene linkage (Figure S1c). The collection of peaks at t_R ~6 min observed in the chromatogram is probably due to crosslinks of AGT at various positions within the oligonucleotide, generating multiple isomers of the DPCs. Collision-induced dissociation (CID) of the *m/z* 1452.6 species gave a-B and w ions that are consistent with the presence of the above 12-mer peptide-ethylene-(15-mer, GC-rich) oligonucleotide complex (Figure S1d and S1e). To further confirm that it is indeed a non-labile DPC, the trypsin-digested reaction mixture was subjected to hot piperidine treatment to remove any residual labile crosslinks. Subsequent LC-MS analysis gave a similar -4 charged *m/z* 1452.6 peak. a) Calculated *m/z* values of the expected a-B and w ions of the 12-mer peptide-ethylene-(GC-rich) oligonucleotide crosslink. b) LC-MS chromatogram of (AGT+DBE)-treated GC-rich oligonucleotide reaction mixture. c) LC-MS spectra of a region of the chromatogram at t_R ~6 min in Part a showing the presence of the *m/z* 1452.6 ion. d) LC-MS/MS chromatogram of the *m/z* 1452.6 ion. e) CID spectra of *m/z* 1452.6(-4) species from the 12-mer peptide-ethylene-GC rich oligonucleotide.

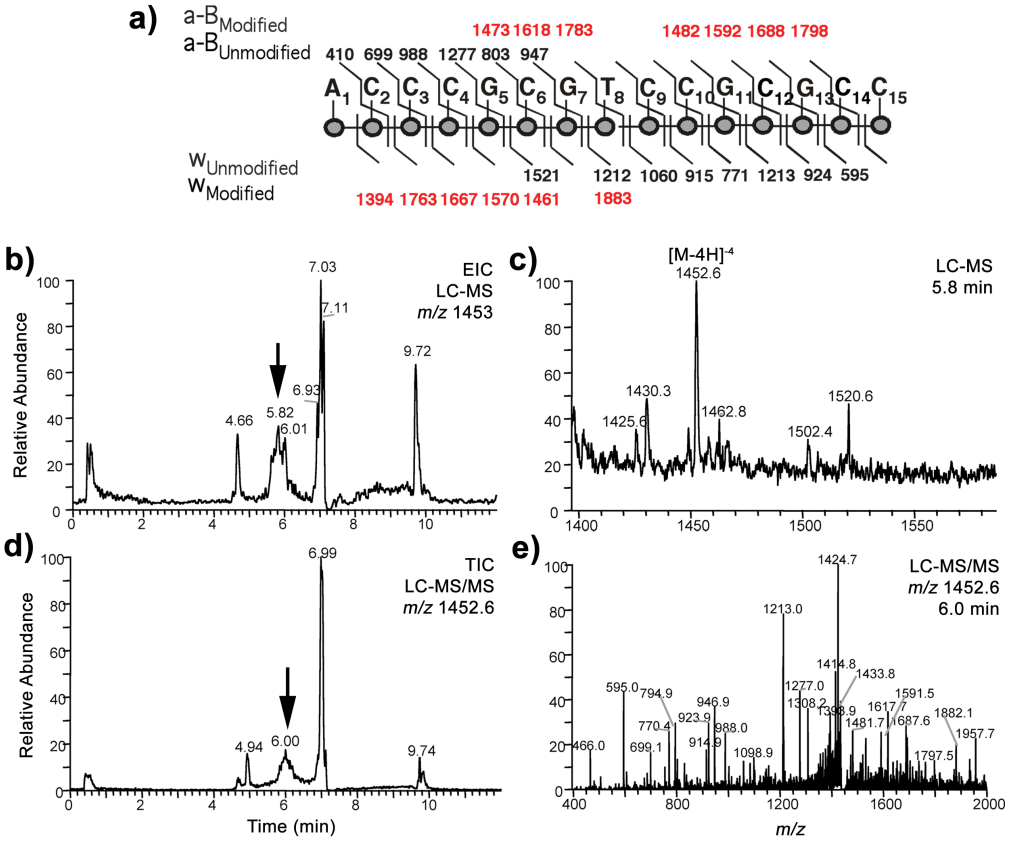

Figure S2. LC-MS chromatogram and spectrum of 12-mer peptide-ethylene-[GC-rich 15-mer oligonucleotide] crosslinks after piperidine treatment to remove labile adducts. The AGT (10 μM)+DBE (20 mM)-modified DNA fragment (20 μM) was digested with trypsin (2 μg) at 37 °C for 16 h, treated with hot piperidine for 30 min, and analyzed by LC-MS. a) LC-MS chromatogram of AGT+DBE treated GC-rich 15- mer oligonucleotide reaction mixture. The arrow indicates peaks corresponding to the crosslinks. b) LC-MS spectra of a region of the chromatogram at t_R ~5.4 min showing presence of the *m/z* 1452.6 ion, which corresponds to the -4 charged crosslink species.

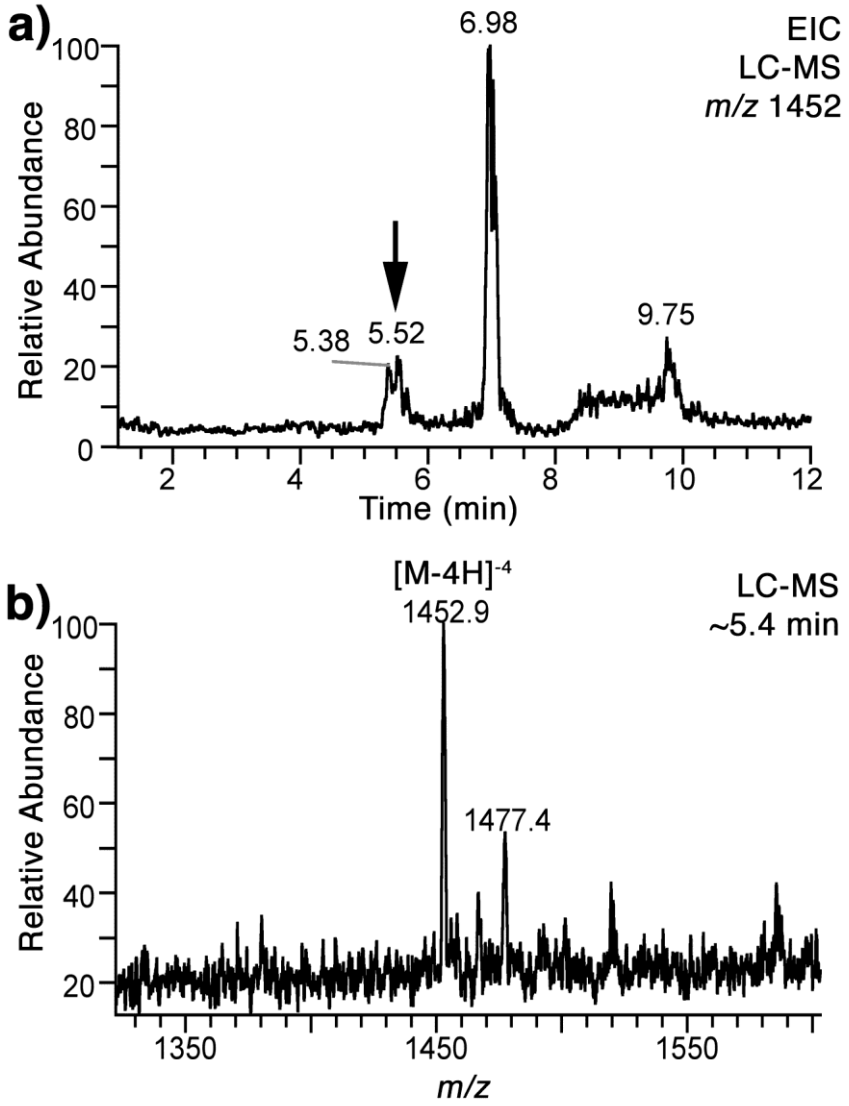

Figure S3. LC-MS/MS chromatogram and spectrum of 12-mer peptide-ethylene-$T_5G_2T_4$ oligonucleotide crosslink. The AGT (10 μM)+DBE (20 mM)-modified DNA fragment (20 μM) was digested with trypsin (2 μg) at 37 °C for 16 h, heated at 90 °C for 30 min, and analyzed by LC- MS/MS. a) Calculated *m/z* values of the expected a-B and w ions of the 12-mer peptide-ethylene-$T_5G_2T_4$ oligonucleotide crosslink. b) LC-MS/MS chromatogram of the 12-mer peptide-ethylene-$T_5G_2T_4$ oligonucleotide crosslinks. c) CID spectrum of a region of the chromatogram at t_R ~4.9 min showing presence of the various a-B and w ions, which corresponds to the 12-mer peptide-ethylene-$T_5G_2T_4$ oligonucleotide crosslinks.

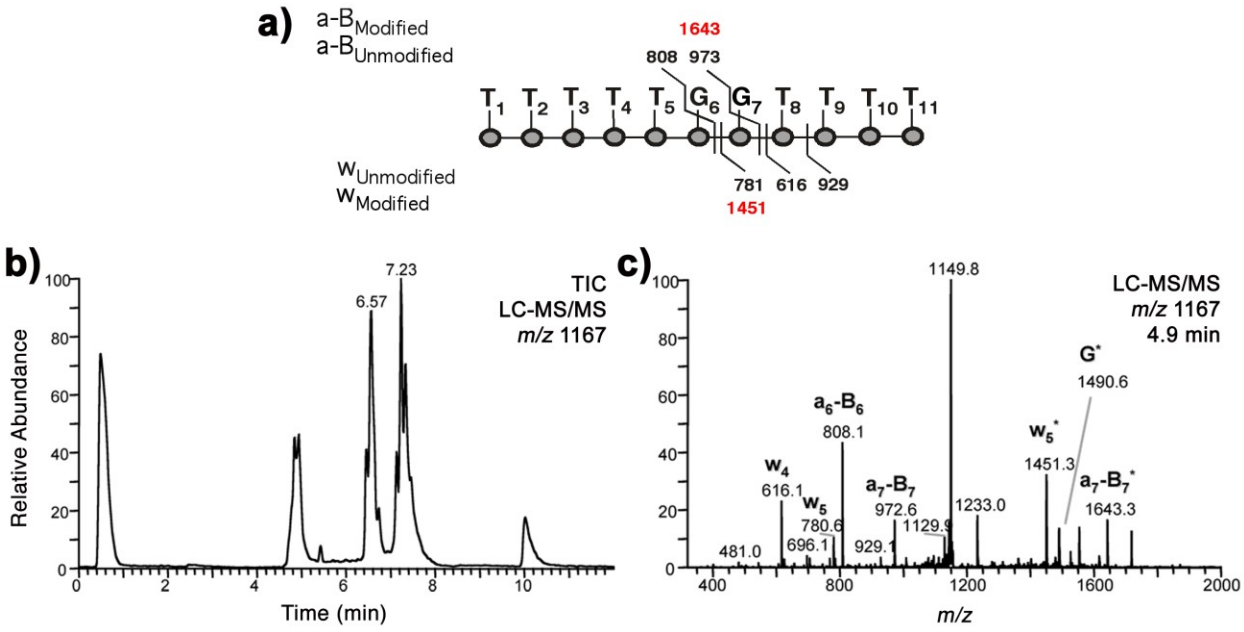

Figure S4. LC-MS and LC-MS/MS chromatograms and spectra of 12-mer peptide-ethylene-$(AT)_6$ oligonucleotide crosslinks. AGT (10 μM)+DBE (20 mM)-modified DNA fragment (20 μM) was digested with trypsin (2 μg) at 37 °C for 16 h, heated at 90 °C for 30 min, and analyzed by LC-MS/MS. The LC-MS data showed a collection of peaks at t_R ~6 min with *m/z* values of 1244.5 and 1659.9 that corresponds to the -4 and -3 charged species of the 12-mer peptide (GNPVPILIPCHR) of AGT covalently linked to the $(AT)_6$ oligonucleotide through an ethylene linker. CID of the *m/z* 1244.5 species gave a-B and w ions that are consistent with the presence of the 12-mer peptide-ethylene-$(AT)_6$ complex. a) Calculated *m/z* values of the expected a-B and w ions of the 12-mer peptide-ethylene-$(AT)_6$ oligonucleotide crosslink. b) LC-MS chromatogram of AGT+DBE treated $(AT)_6$ oligonucleotide reaction mixture. c) LC-MS spectra of a region of the chromatogram at t_R ~6 min in a showing presence of the *m/z* 1244.5(-4) ion. d) LC-MS/MS chromatogram of the *m/z* 1244.5 ion. e) CID spectra of *m/z* 1244.5(-4) species from 12-mer peptide-ethylene-$(AT)_6$ oligonucleotide.

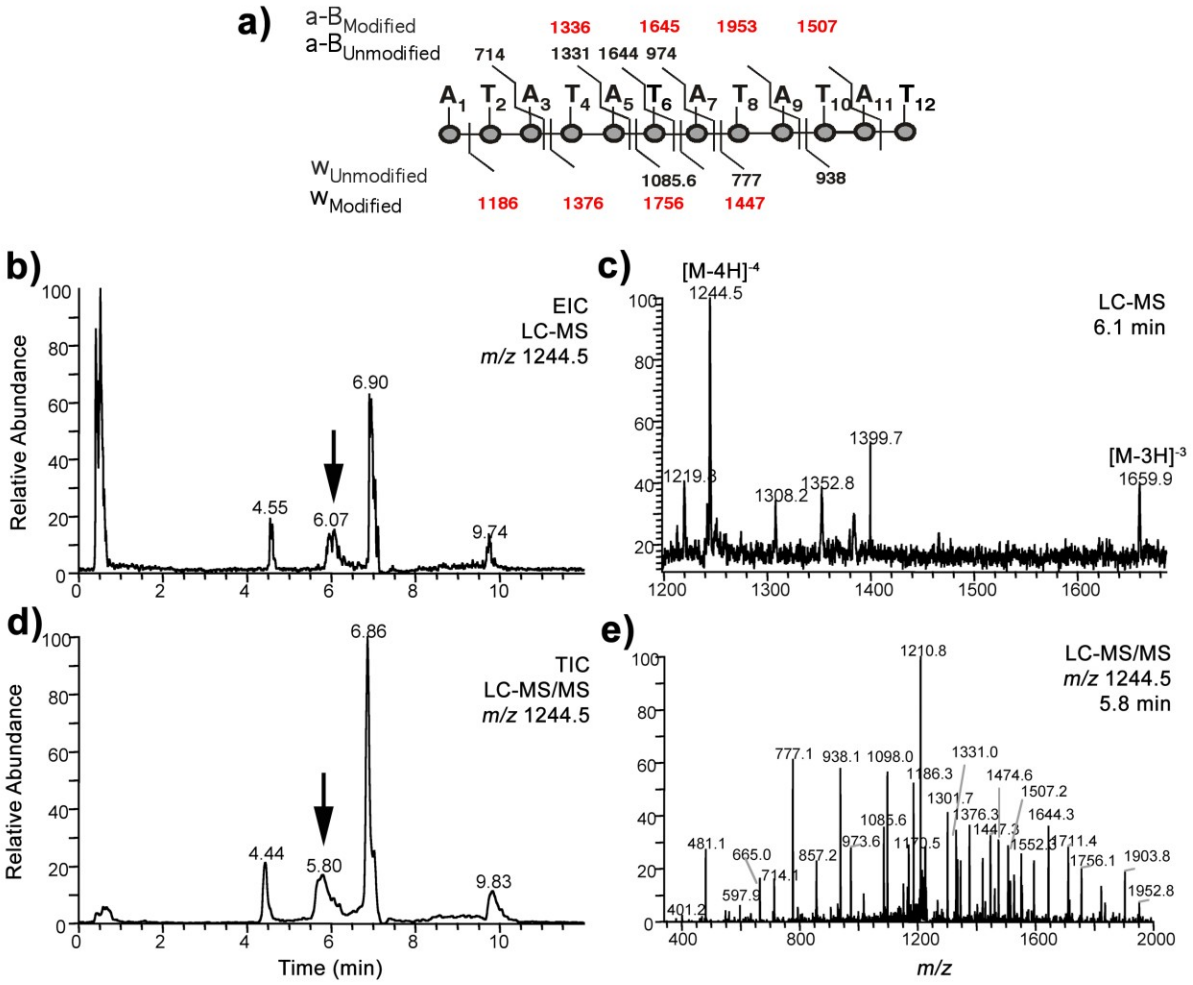

Figure S5. LC-MS/MS and LC-MS^3 chromatograms and spectra showing the presence of N^6-EtdA adducts from calf thymus DNA. Calf thymus DNA was treated with AGT (20 μM) and DBE (10 mM) to form the DPCs. The DPCs formed were converted to EtdA adducts by Raney Ni treatment and analyzed by LC-MS/MS and LC-MS^3 using specific transitions. a) LC-MS/MS chromatogram using the *m/z* 280→164 transition. b) LC-MS^3 chromatogram using the *m/z* 280→164→136 transition. c) LC-MS/MS spectra of the region of the chromatogram at t_R ~3.6 min in a showing presence of the *m/z* 164 ion, which indicates presence of the ethyl dA adduct. d) LC-MS^3 spectra of the region of the chromatogram at t_R ~3.6 min in b showing the presence of the *m/z* 136 ion, again indicating the presence of the EtdA adduct.

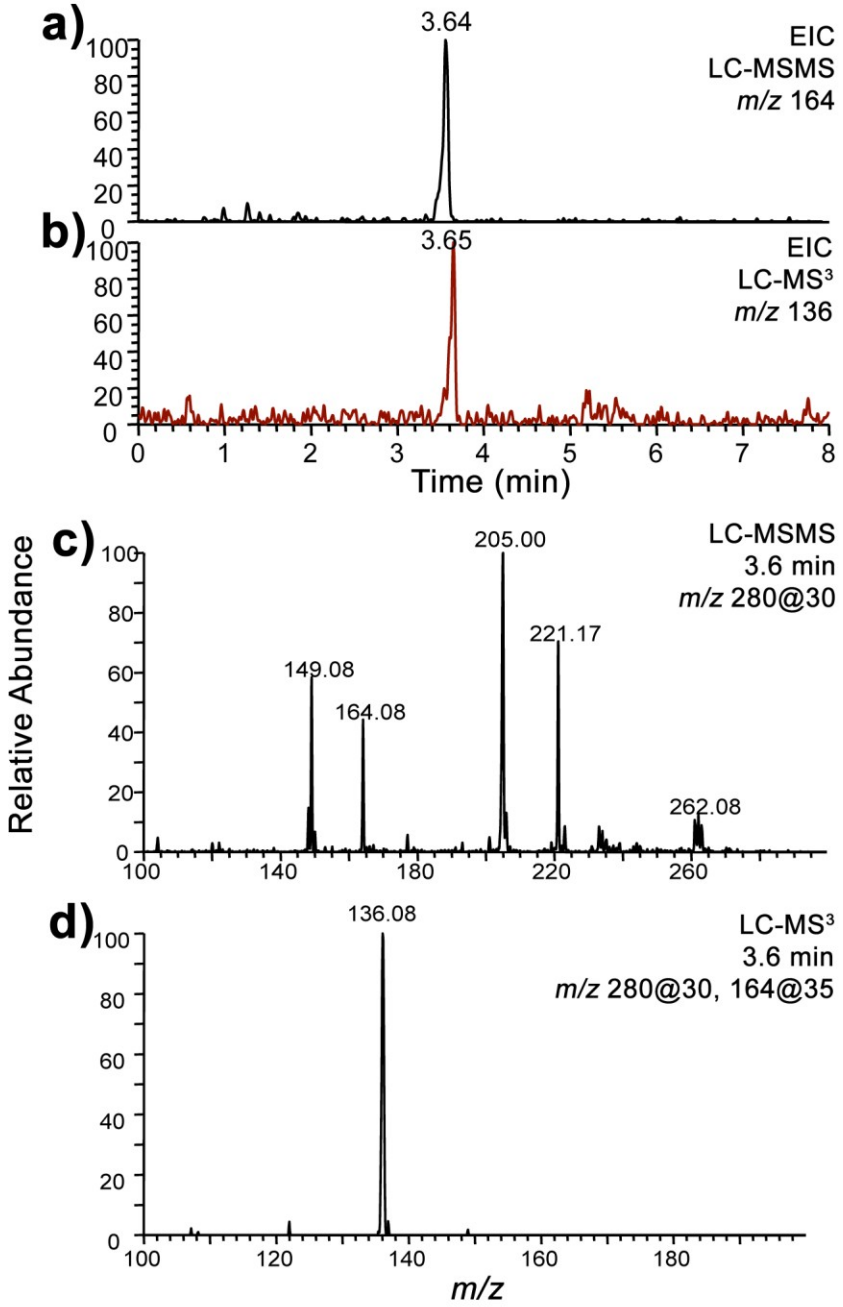

Figure S6. LC-MS/MS and LC-MS^3 chromatograms and spectra showing the presence of EtdG adducts in calf thymus DNA. Calf thymus DNA was treated with AGT (20 μM) and DBE (10 mM) to form the DPCs. The DPCs formed were converted to EtdG adducts by Raney Ni treatment and analyzed by LC-MS/MS and LC-MS^3 using specific transitions. a) LC-MS/MS chromatogram using the *m/z* 296→180 transition. b) LC-MS^3 chromatogram using the *m/z* 296→180→163 transition (for N^2-EtdG adducts). c) LC-MS^3 chromatogram using the *m/z* 296→180→152 transition (for *N*1-EtdG and O^6-EtdG adducts). d) LC-MS^3 spectra of the region of the chromatogram at t_R ~3.6 min in b) showing the presence of the *m/z* 163 ion, which indicates the presence of the N^2-EtdG adduct. e) LC-MS^3 spectra of the region of the chromatogram at t_R ~4.1 min in c) showing presence of the *m/z* 152 ion, which indicates the presence of the O^6-EtdG adduct.

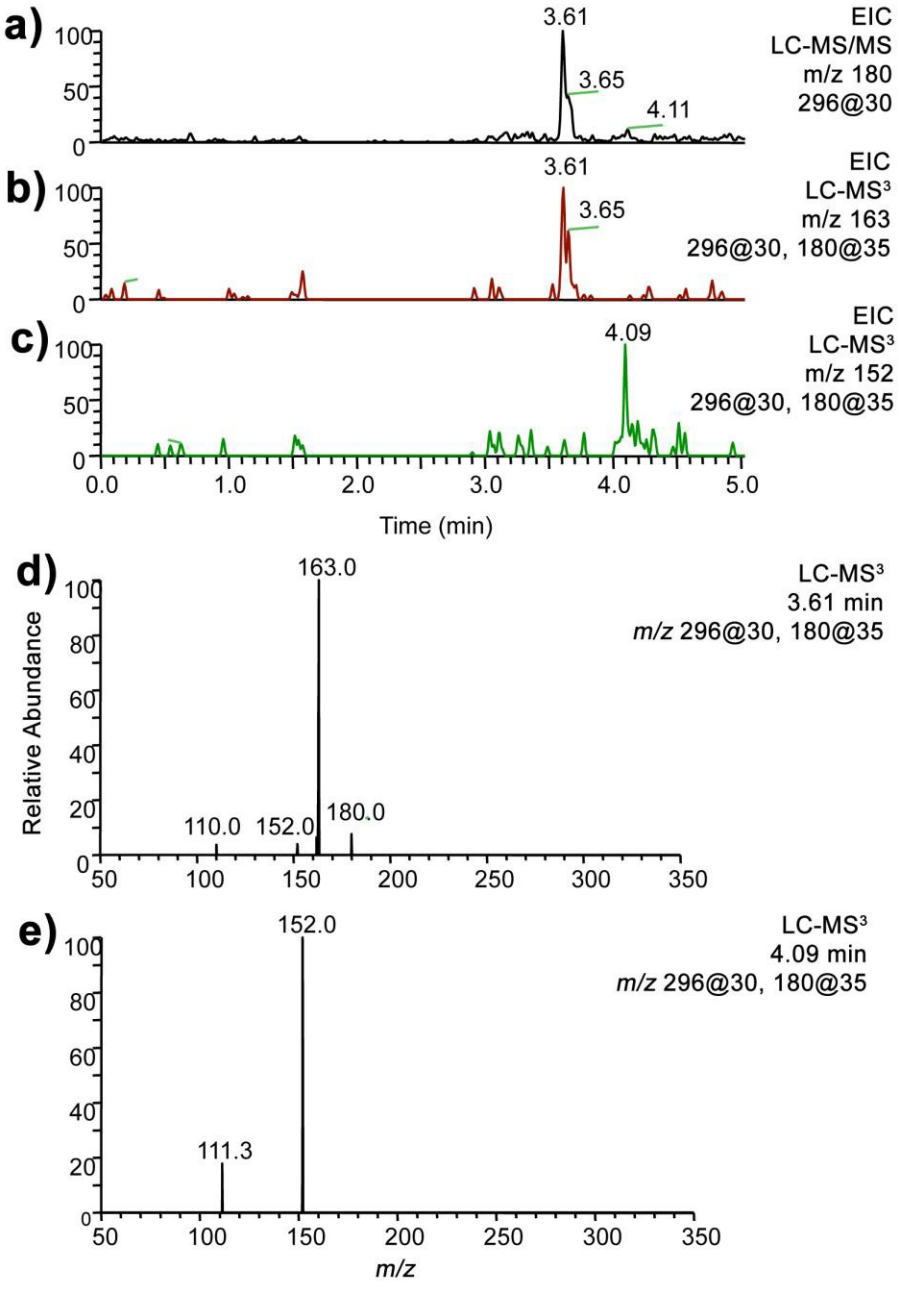

Figure S7. LC-MS/MS and LC-MS^3 chromatograms of synthesized authentic EtdG adducts. a) Extracted ion chromatogram (EIC) using the *m/z* 296→180 transition. b) LC-MS^3 chromatogram using CID of *m/z* 296@30,180@30. Fragmentation of the daughter ion *m/z* 180 gave two major fragments at *m/z* 163 (N^2-EtdG) and *m/z* 152 (*N*1-EtdG and O^6-EtdG adducts). Based on the UV-vis spectra, the peak at t_R~3.2 min (with a major ion at *m/z* 152) was assigned to the *N*1-EtdG and the peak at t_R~4.0 min (with a major ion at *m/z* 152) was assigned to the O^6-EtdG adduct.

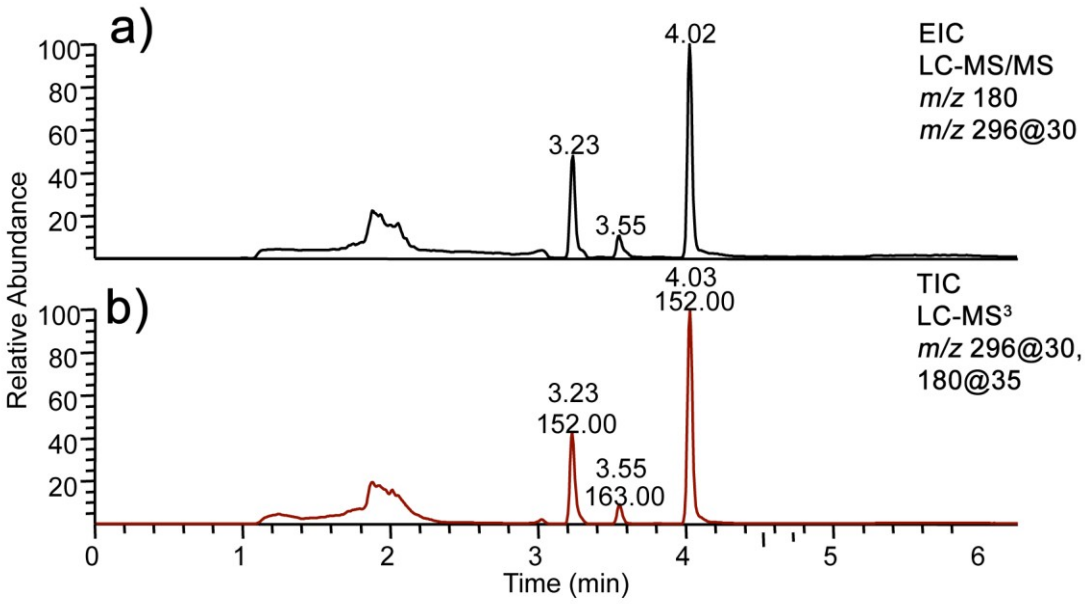

Figure S8. UV-vis spectra of *N*1- and O^6-EtdG adducts

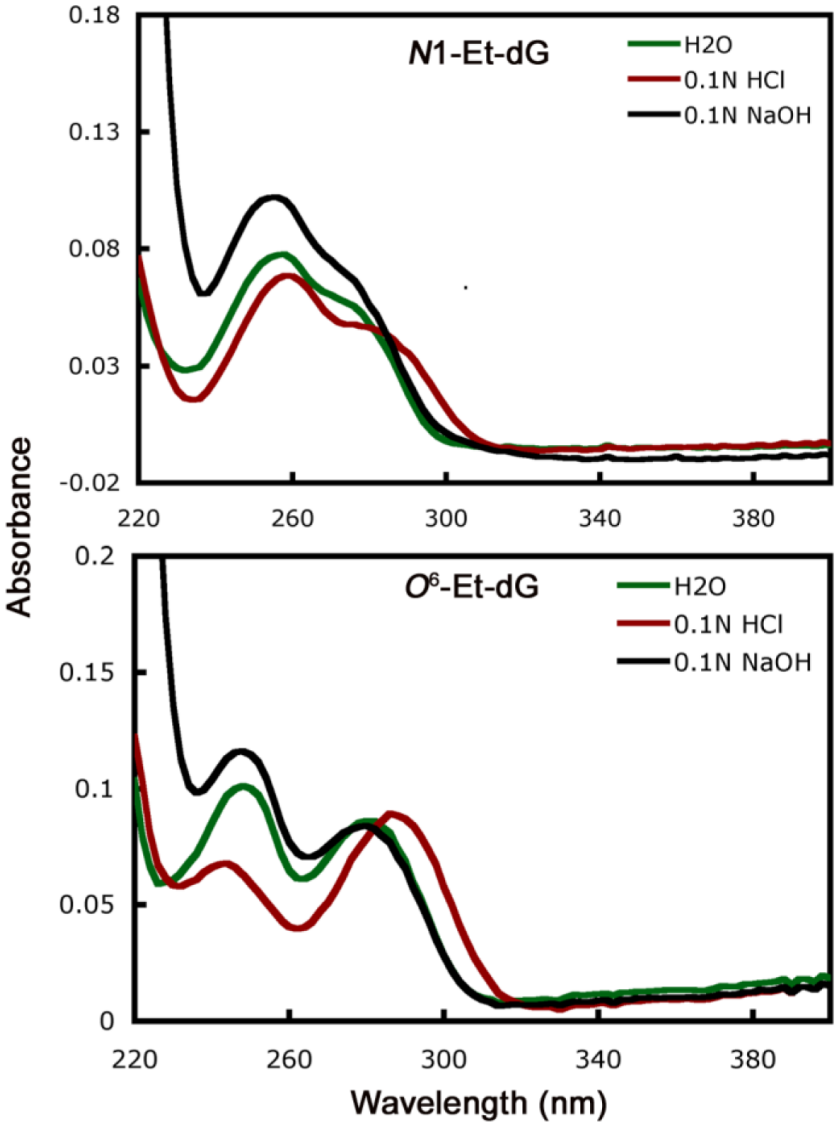

Figure S9. Yields of the various types of DPCs formed in the presence of AGT and ^{14}C-DBE, including control experiments. Yields are shown for various non-labile DPCs (*N*1-dG, N^2-dG, and O^6-dG) measured as their corresponding Et adducts (*n* = 3, ± SD).

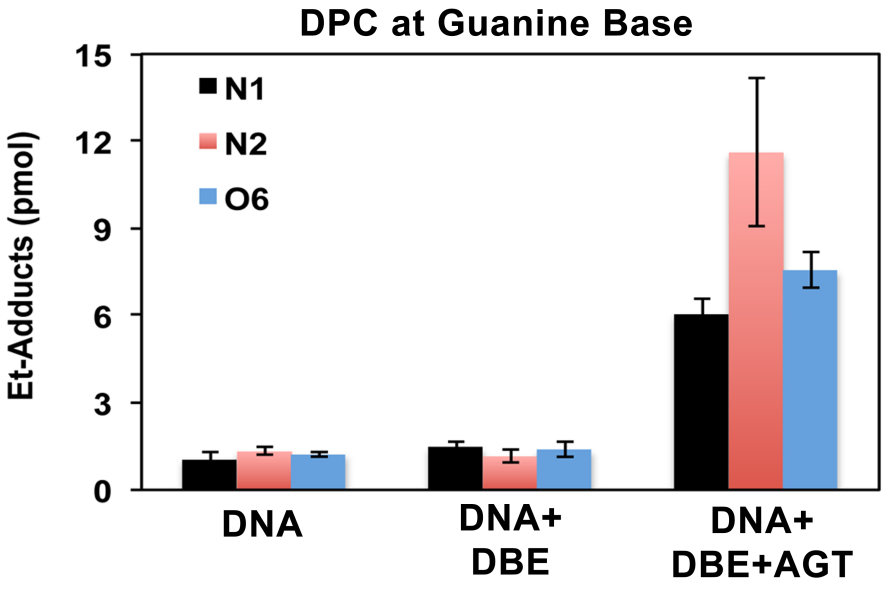

Figure S10. HPLC chromatograms showing the presence of labile adducts (*N*7-EtG) in calf thymus DNA. Calf thymus DNA was treated with AGT (20 μM) and ^{14}C-DBE (2 mM) to form DPCs. The DPCs were purified and converted to *N*7-EtG (or possibly *N*3-EtA) adducts by neutral thermal hydrolysis. The ethyl adducts were analyzed by HPLC-flow counting (flow counter coupled in-line to an HPLC). a) Control (DNA+^{14}C-DBE). b) Sample (DNA+^{14}C-DBE+AGT).

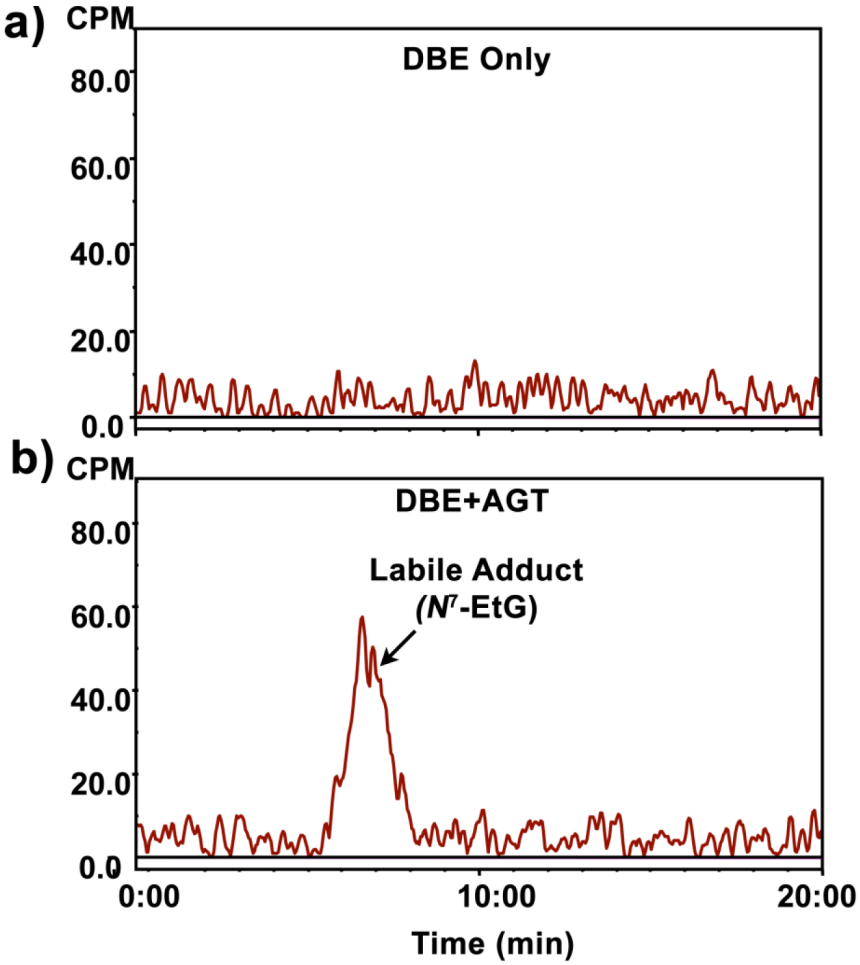

Table S1. Calculated CID fragments expected from the 12-mer peptide (GNPVPILIPCHR)-ethylene-GC rich oligonucleotide crosslinks. Fragments detected in the experimental data (Figure S1) are shown in red.

Sequence :ACC CGC GTC CGC GCC 5OH - DNA[15mer] - 3´OH
C:9 T:1 A:1 G:4
Linker: CH_2CH_2, 28
Peptide: GNPVPILIPCHR, 1413.7
monoisotopic mass, negative mode

		Unmodified		Modified	
n	charge	a-B	w	a-B	w
1	-1		306.048		1648.012
2	-1	410.085	595.094	1752.038	1937.058
	-2				968.025
3	-1	699.131	924.146	2041.084	2266.110
	-2	349.061	461.569	1020.038	1132.551
4	-1	988.177	1213.192	2330.130	2555.156
	-2	493.584	606.092	1164.561	1277.074
5	-1	1277.223	1542.244	2619.176	2884.209
	-2	638.107	770.618	1309.084	1441.600
6	-1	1606.275	1831.290	2948.229	3173.255
	-2	802.634	915.141	1473.610	1586.123
	-3			982.071	1057.079
7	-2	947.157	1059.664	1618.133	1730.646
	-3	631.102	706.106	1078.419	1153.428
8	-2	1111.683	1211.687	1782.659	1882.669
	-3	740.786	807.455	1188.103	1254.777
9	-2	1263.705	1376.213	1934.682	2047.195
	-3	842.134	917.139	1289.452	1364.461
10	-2	1408.228	1520.736	2079.205	2191.718
	-3	938.483	1013.488	1385.098	1460.809
	-4	703.610	759.864	1039.098	1095.355

11	-2	1552.751	1685.262	2223.728	2356.244
	-3	1034.832	1123.172	1482.149	1570.493
	-4	775.872	842.127	1111.360	1177.618
12	-2	1717.278	1829.785	2388.254	2500.767
	-3	1144.516	1219.521	1591.833	1666.842
	-4	858.135	914.388	1193.623	1249.879
13	-2	1861.801	1974.308	2532.777	2645.290
	-3	1240.864	1315.869	1688.182	1763.191
	-4	930.396	986.650	1265.884	1322.141
14	-3	1350.548	1412.218	1797.866	1859.539
	-4	1012.659	1058.911	1384.148	1394.402
	-5	809.926	846.927	1078.316	1115.320

Table S2. Calculated CID fragments expected from the 12-mer peptide (GNPVPILIPCHR)-ethylene-$(AT)_6$ crosslinks. Fragments detected in the experimental data (Figure S4) are shown in red.

Sequence : **ATA TAT ATA TAT** 5′OH - DNA[12mer] – 3′OH
T:6 A:5 N:1
Linker: CH_2CH_2, 28
Peptide: GNPVPILIPCHR, 1413.7
monoisotopic mass, negative mode

		Unmodified		Modified	
n	charge	a-B	w	a-B	w
1	-1		321.047		1662.012
2	-1	410.085	634.105	1752.038	1975.070
3	-1	**714.131**	**938.150**	2056.084	2279.115
	-2	356.561	468.571	1027.538	1139.053
4	-1	1027.188	1251.208	2369.141	2592.173
	-2	513.090	625.100	1184.067	1295.582
5	-1	**1331.234**	1555.253	2673.187	2896.218
	-2	**665.113**	**777.123**	**1336.089**	**1447.605**
6	-1	1644.291	1868.311	2986.244	3209.276
	-2	821.641	933.651	1492.618	1604.134
7	-2	**973.664**	**1085.674**	**1644.641**	1756.156
	-3	648.773	723.447	1096.091	**1170.435**
8	-2	1130.193	1242.203	1801.170	1912.685
	-3	753.126	827.799	1200.444	1274.787
9	-2	1282.216	1394.226	**1953.192**	2064.708
	-3	854.474	929.148	**1301.792**	**1376.136**
10	-2	1438.744	1550.754	2109.721	2221.237
	-3	958.827	1033.500	1406.145	1480.488
11	-2	1590.767	1702.777	2261.744	2373.259
	-3	1060.175	1134.849	**1507.493**	1581.873
	-4	794.879	850.884	1130.368	**1186.126**

Table S3. Yields of conversion of ethylene-GSH adducts to ethyl adducts by Raney nickel reaction.

Et adduct	Yield (%)[a]
N^7-Et G	90.3 ± 3.1
*N*1-Et dG	88.7 ± 4.0
N^6-Et dA	91.0 ± 6.1

[a] Mean ± SD obtained from triplicate assays

References (for Supporting Information section)

1. S. Edara, S. Kangula, K. Goodtzova, and A. E. Pegg, *Cancer Res.* **1996**, *56*. 5571-5575.

2. Q. Fang, A. M. Noronha, S. P. Murphy, C. J. Wilds, J. I. Tubbs, J. A. Tainer, G. Chowdhury**,** F. P. Guengerich, A. E. Pegg, *Biochemistry* **2008,** *47*, 10892-10903.

3. W. G. Humphreys, D.-H. Kim, J. L. Cmarik, T. Shimada, F. P. Guengerich, *Biochemistry* **1990**, *29*, 10342-10350.

4. S.-H. Cho, F. P. Guengerich, *Chem. Res. Toxicol*. **2012**, *25*, 706-712.

5. D.-H. Kim, W. G. Humphreys, F. P. Guengerich, *Chem. Res. Toxicol.* **1990**, *3*, 587-594.

6. N. Ozawa, F. P. Guengerich, F. P., *Proc. Natl. Acad. Sci. U. S. A.* **1983***, 80*, 5266-5270.

7. B . Chen, X. Zhou, K. Taghizadeh K, J. Chen, J. Stubbe, P. C. Dedon, *Chem. Res. Toxicol.* **2007**, *20*, 1701-1708.