Characterization of natural colibactin-nucleobase adducts by tandem MS and isotopic labeling.

Support for DNA alkylation by cyclopropane ring opening.

Mengzhao Xue,[a] Emilee Shine,[b,c] Weiwei Wang,[d,e] Jason M. Crawford,[a,b,c*] and Seth B. Herzon[a,f*]

[a]Department of Chemistry, Yale University, New Haven, Connecticut 06520, United States.

[b]Chemical Biology Institute, Yale University, West Haven, Connecticut 06516, United States.

[c]Department of Microbial Pathogenesis, Yale School of Medicine, New Haven, Connecticut 06536, United States.

[d]Department of Molecular Biophysics and Biochemistry, Yale University School of Medicine, PO Box 208114, New Haven, CT, 06520.

[e]W.M. Keck Biotechnology Resource Laboratory, Yale University School of Medicine, 300 George Street, New Haven, CT 06510.

[f]Department of Pharmacology, Yale School of Medicine, New Haven, Connecticut 06520, United States.

Biochemistry

Supporting Information

Index

Supplementary Figures

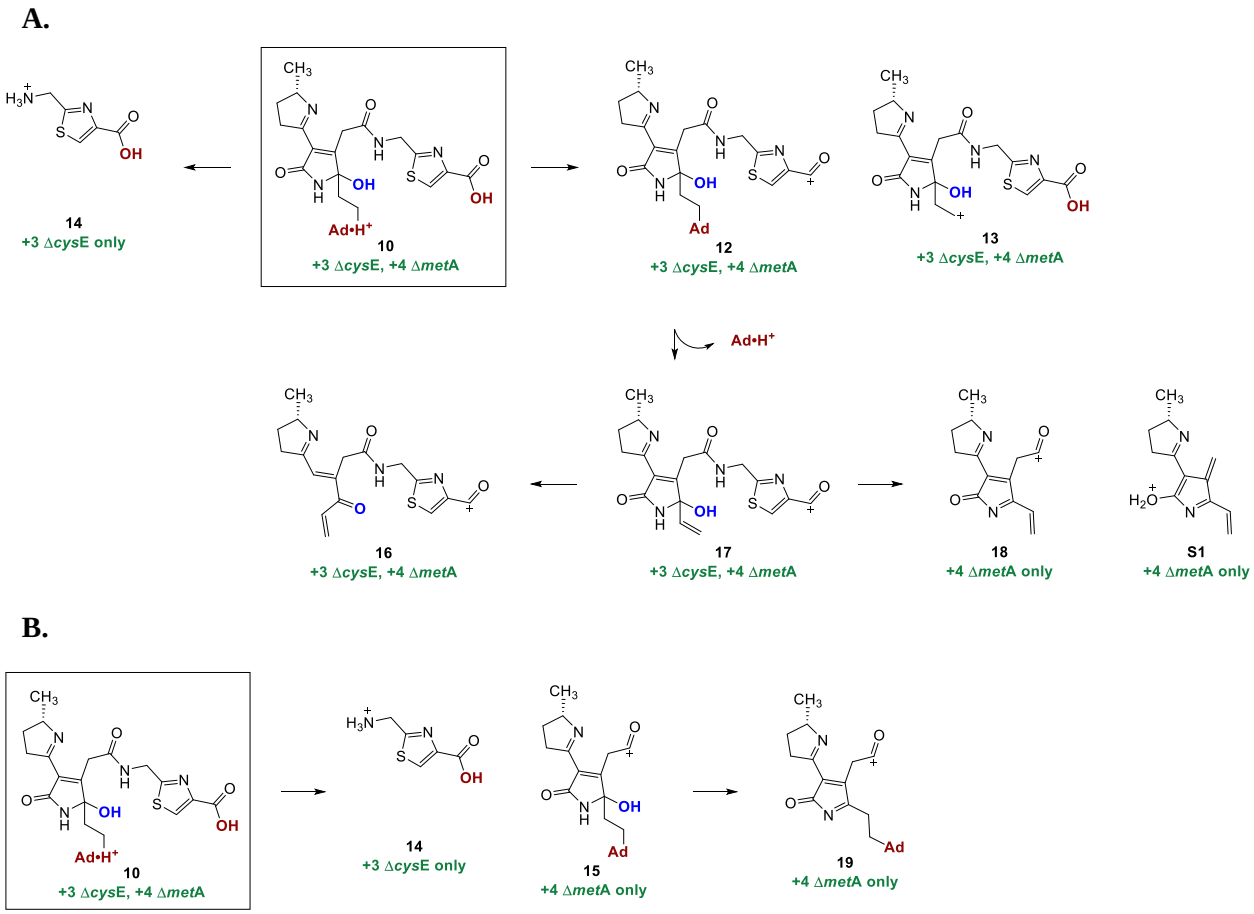


Figure S1. **A**. Fragmentation products derived from MS^2 of **10**, **$^{13}C_3$-10**, and **$^{13}C_4$-10**. B. Fragmentation products derived from daughter ions of **10**, **$^{13}C_3$-10**, and **$^{13}C_4$-10**.

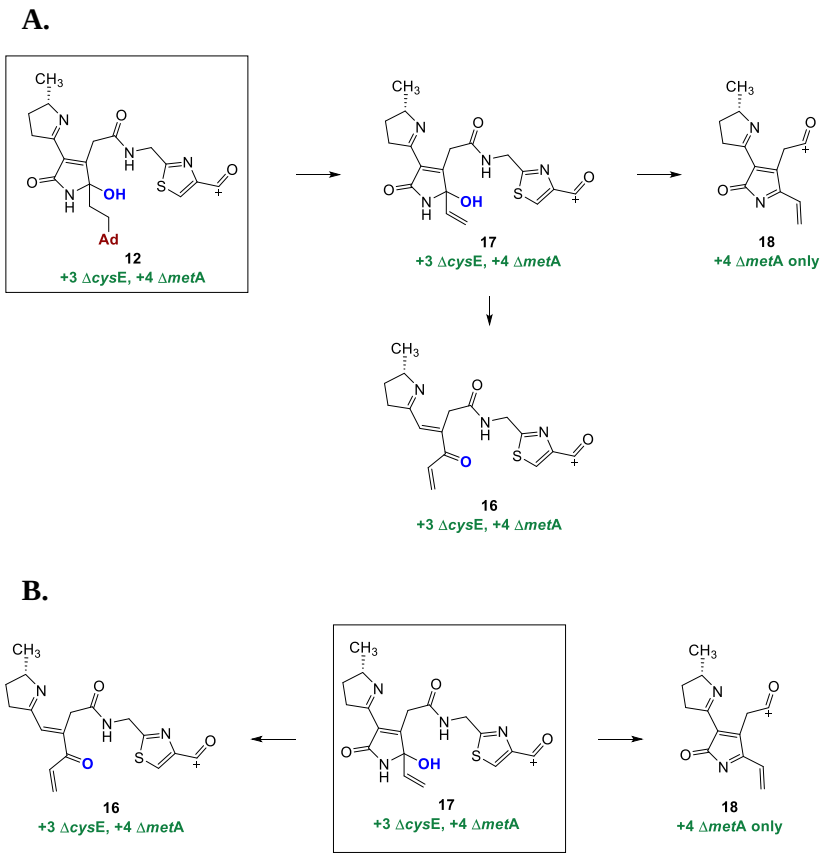
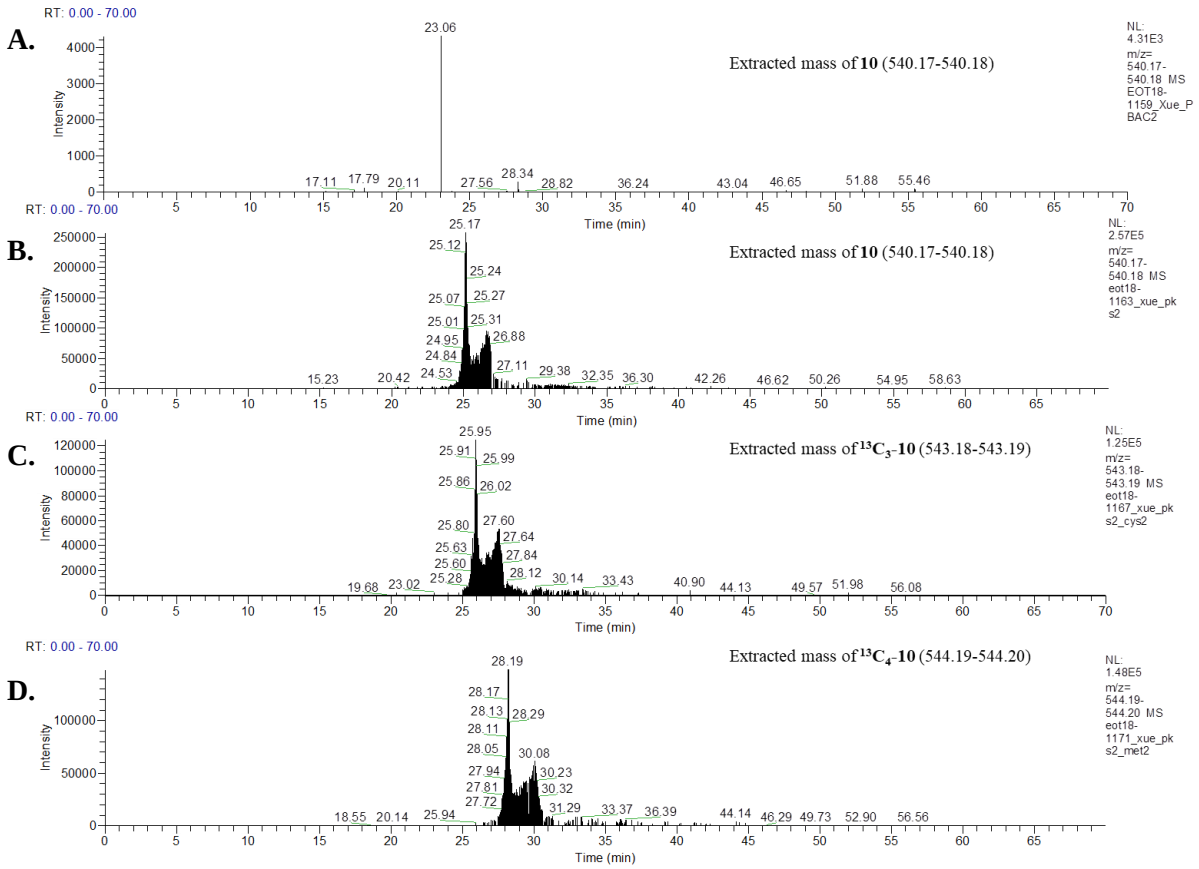


Figure S2. **A**. Fragmentation products derived from MS^3 of **12**, **$^{13}C_3$-12**, and **$^{13}C_4$-12**. **B**. Fragmentation products derived from MS^3 of **17**, **$^{13}C_3$-17**, and **$^{13}C_4$-17**.



Spectra S3. **A**. Trace of extracted mass of **10** (540.17-540.18) in linearized pUC19 DNA that had been exposed to *clb*$^-$ *E. coli* and digested. **B**. Trace of extracted mass of **10** (540.17-540.18) in linearized pUC19 DNA that had been exposed to *clb*$^+$ *E. coli* and digested. **C**. Trace of extracted mass of **$^{13}C_3$-10** (543.18-543.19) in linearized pUC19 DNA that had been exposed to Δ*cys*E *clb*$^+$ *E. coli* and media containing L-[U]-^{13}C-cys and digested. **D**. Trace of extracted mass of **$^{13}C_4$-10** (544.19-544.20) in linearized pUC19 DNA that had been exposed to Δ*met*A *clb*$^+$ *E. coli* and media containing L-[U]-^{13}C-met and digested.

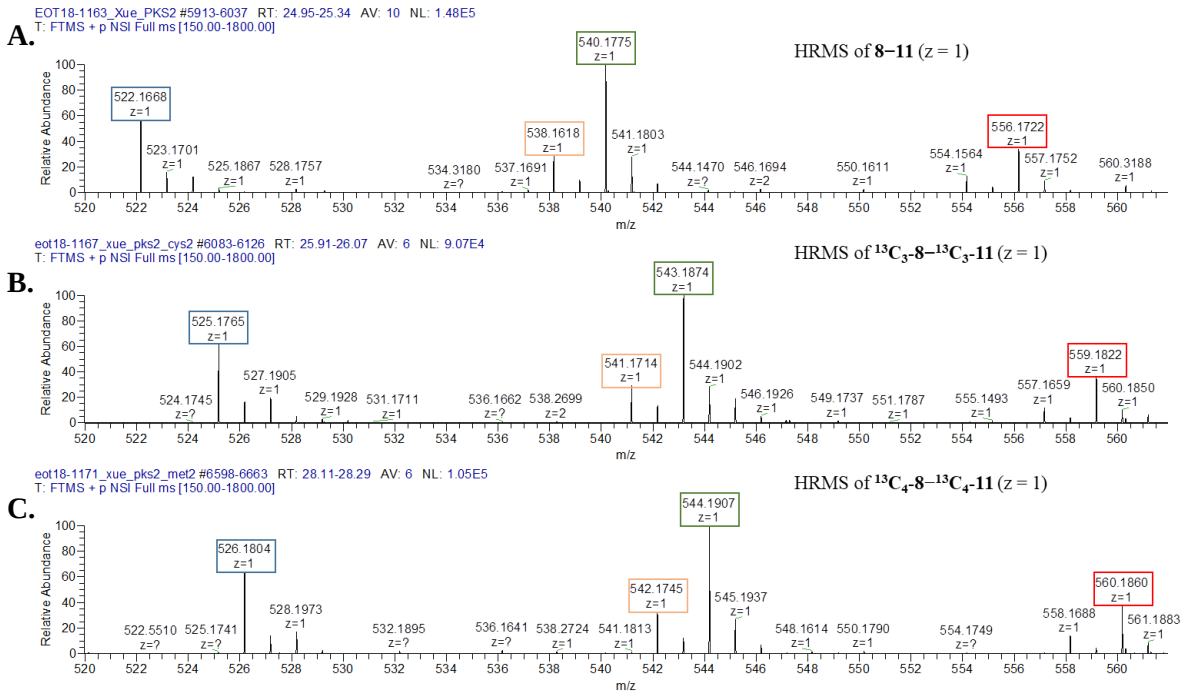


Figure S4. Experimental masses. **A**. Full HRMS spectra of **8** (522.1668, z = 1), **9** (538.1618, z = 1), **10** (540.1775, z = 1), **11** (556.1722, z = 1). **B**. Full HRMS spectra of $^{13}C_3$**-8** (525.1765, z = 1), $^{13}C_3$**-9** (541.1714, z = 1), $^{13}C_3$**-10** (543.1874, z = 1), $^{13}C_3$**-11** (559.1822, z = 1). **C**. Full HRMS spectra of $^{13}C_4$**-8** (526.1804, z = 1), $^{13}C_4$**-9** (543.1745, z = 1), $^{13}C_4$**-10** (544.1907, z = 1), $^{13}C_4$**-11** (560.1860, z = 1).

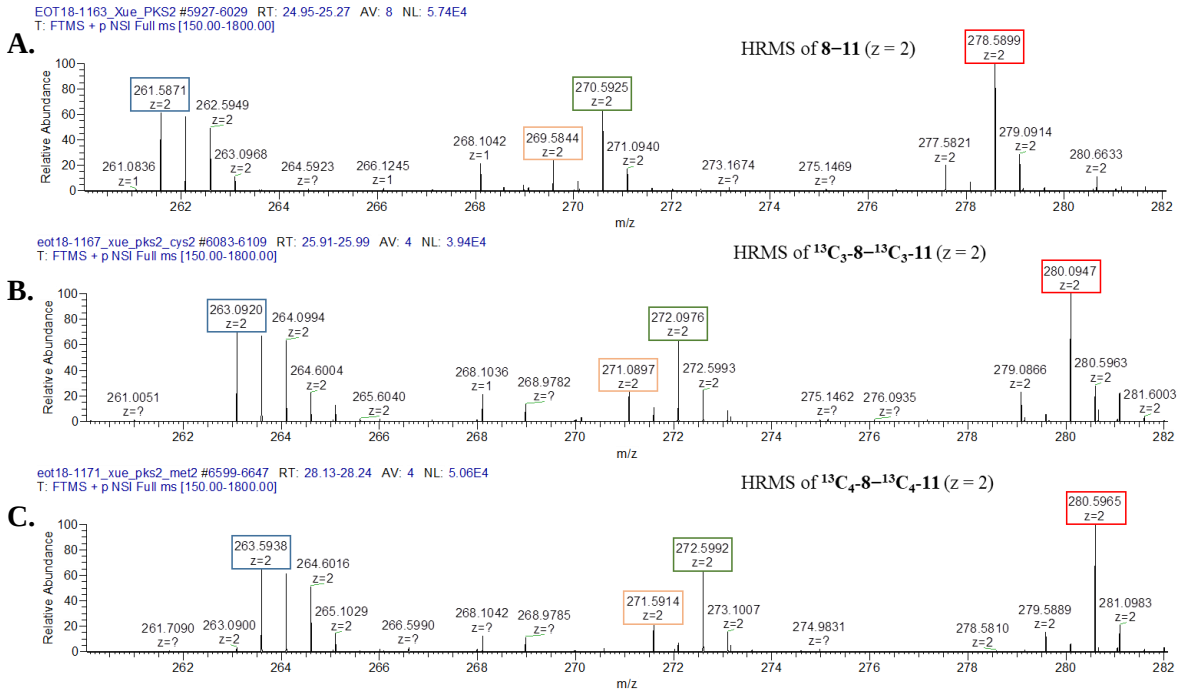


Figure S5. Experimental masses. **A**. Full HRMS spectra of **8** (261.5871, z = 2), **9** (269.5844, z = 2), **10** (270.5925, z = 2), **11** (278.5899, z = 2). **B**. Full HRMS spectra of $^{13}C_3$**-8** (263.0920, z = 2), $^{13}C_3$**-9** (271.0897, z = 2), $^{13}C_3$**-10** (272.0976, z = 2), $^{13}C_3$**-11** (280.0947, z = 2). **C**. Full HRMS spectra of $^{13}C_4$**-8** (263.5938, z = 2), $^{13}C_4$**-9** (271.5914, z = 2), $^{13}C_4$**-10** (272.5992, z = 2), $^{13}C_4$**-11** (280.5965, z = 2).

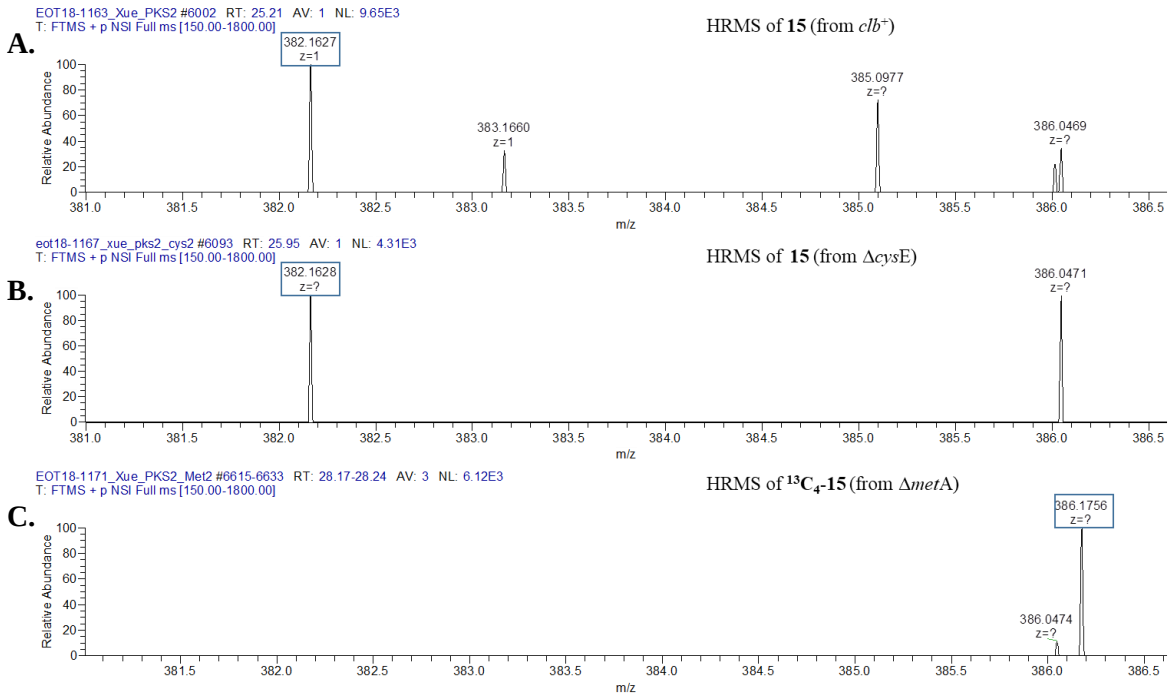


Figure S6. All of the identified mass are experimental values. **A**. HRMS spectrum of **15** from *clb*$^+$ (382.1627, z = 1). **B**. HRMS spectrum of **15** from Δ*cys*E *clb*$^+$ + L-[U]-^{13}C-cys (382.1628). **C**. HRMS spectrum of **$^{13}C_4$-15** from Δ*met*A *clb*$^+$ + L-[U]-^{13}C-met (386.1756).

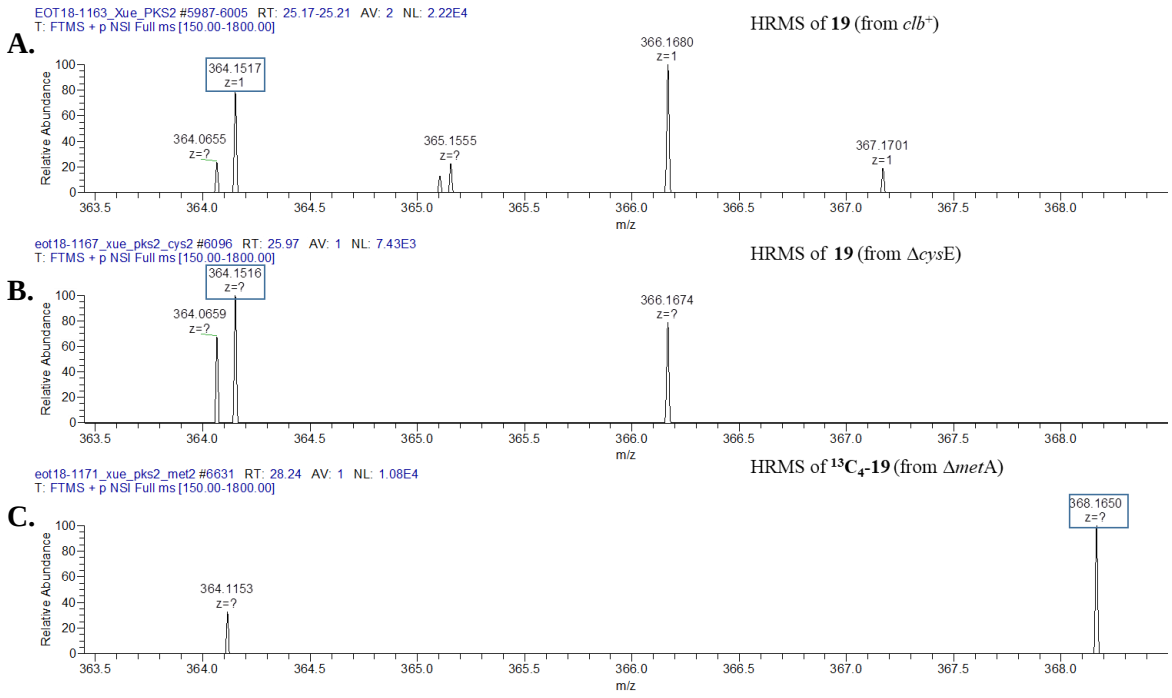


Figure S7. Experimental masses. **A**. HRMS spectrum of **19** from *clb*$^+$ (364.1517, z = 1). **B**. HRMS spectrum of **19** from Δ*cys*E *clb*$^+$ + L-[U]-^{13}C-cys (364.1516). **C**. HRMS spectrum of **$^{13}C_4$-19** from Δ*met*A *clb*$^+$ + L-[U]-^{13}C-met (368.1650).

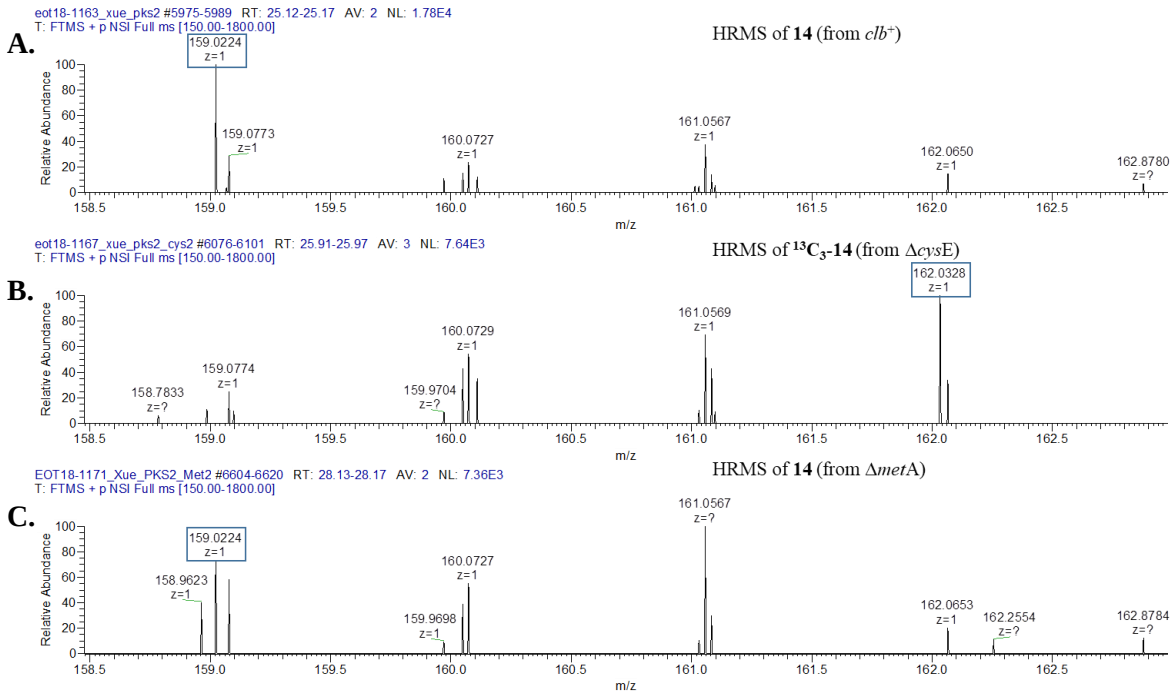


Figure S8. Experimental masses. **A**. HRMS spectrum of **14** from *clb*$^+$ (159.0224, z = 1). **B**. HRMS spectrum of **$^{13}C_3$-14** from Δ*cys*E *clb*$^+$ + L-[U]-^{13}C-cys (162.0328, z = 1). **C**. HRMS spectrum of **14** from Δ*met*A *clb*$^+$ + L-[U]-^{13}C-met (159.0224, z = 1).

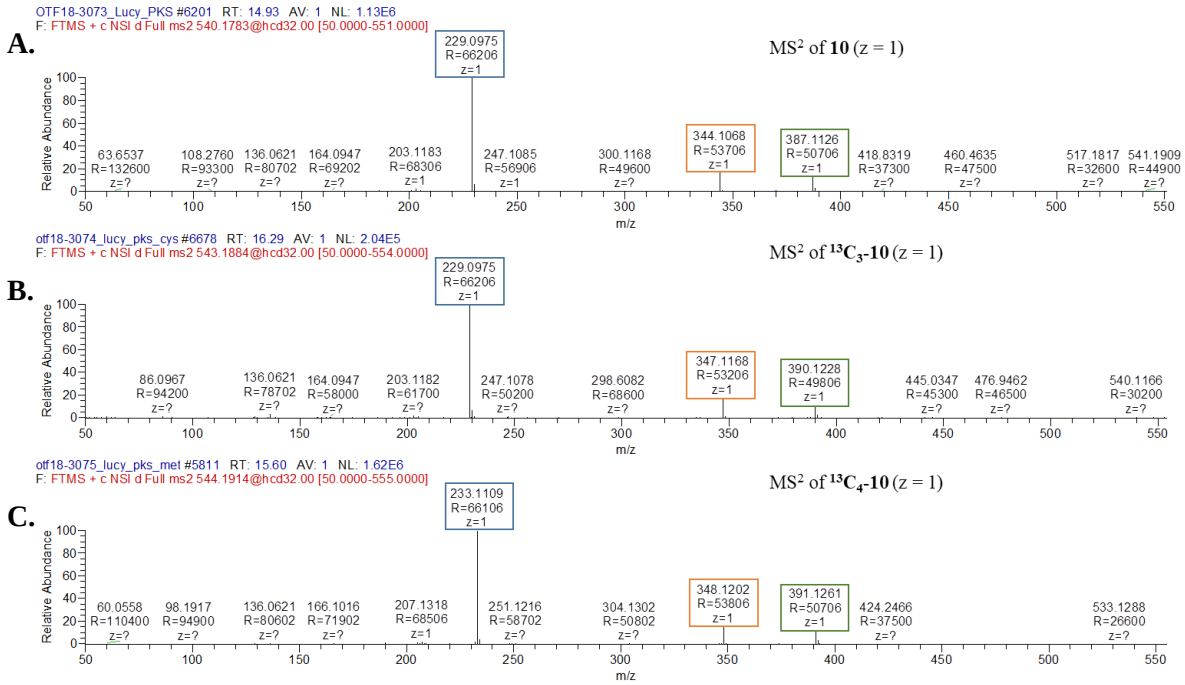


Figure S9. Experimental masses. **A**. MS^2 spectra-1 of **10** (z = 1), **18** (229.0975, z = 1), **16** (344.1068, z = 1), **17** (387.1126, z = 1) were detected. **B**. MS^2 spectra-1 of $^{13}C_3$**-10** (z = 1), **18** (229.0975, z = 1), $^{13}C_3$**-16** (347.1168, z = 1), $^{13}C_3$**-17** (390.1228, z = 1) were detected. **C**. MS^2 spectra-1 of $^{13}C_4$**-10** (z = 1), $^{13}C_4$**-18** (233.1109, z = 1), $^{13}C_4$**-16** (348.1202, z = 1), $^{13}C_4$**-17** (391.1261, z = 1) were detected.

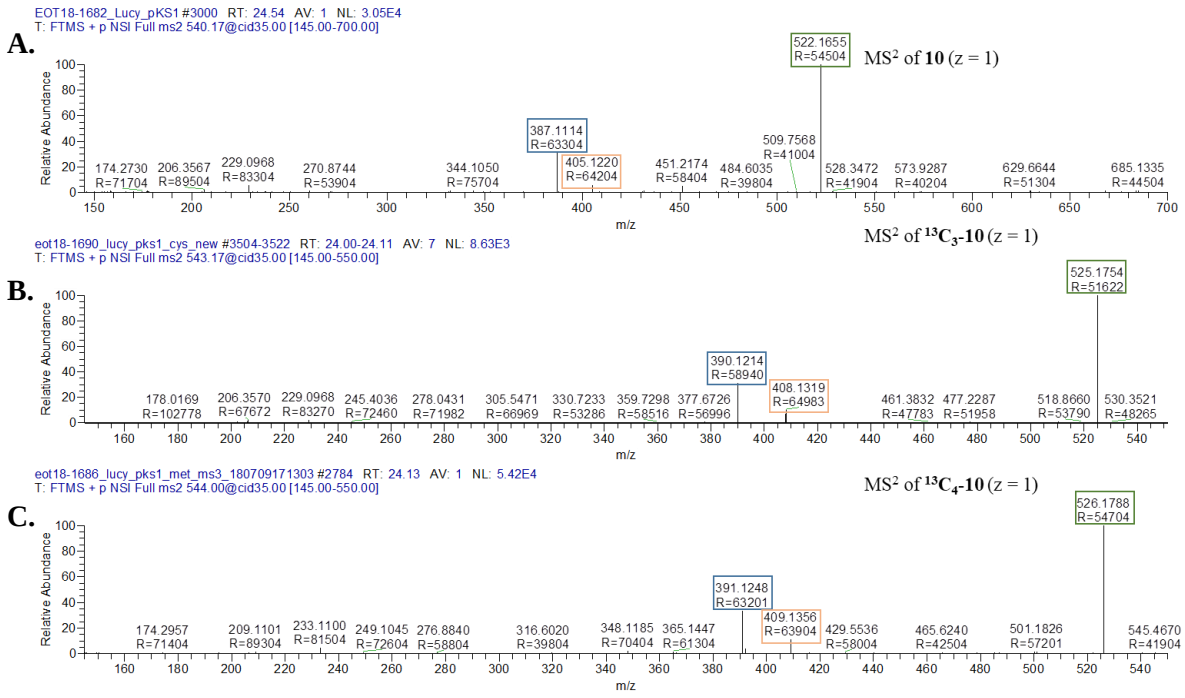


Figure S10. Experimental masses. **A**. MS^2 spectra-2 of **10** (z = 1), **17** (387.1114), **13** (405.1220), **12** (522.1655) were detected. **B**. MS^2 spectra-2 of $^{13}C_3$**-10** (z = 1), $^{13}C_3$**-17** (390.1214), $^{13}C_3$**-13** (408.1319), $^{13}C_3$**-12** (525.1754) were detected. **C**. MS^2 spectra-2 of $^{13}C_4$**-10** (z = 1), $^{13}C_4$**-17** (391.1248), $^{13}C_4$**-13** (409.1356), $^{13}C_4$**-12** (526.1788) were detected.

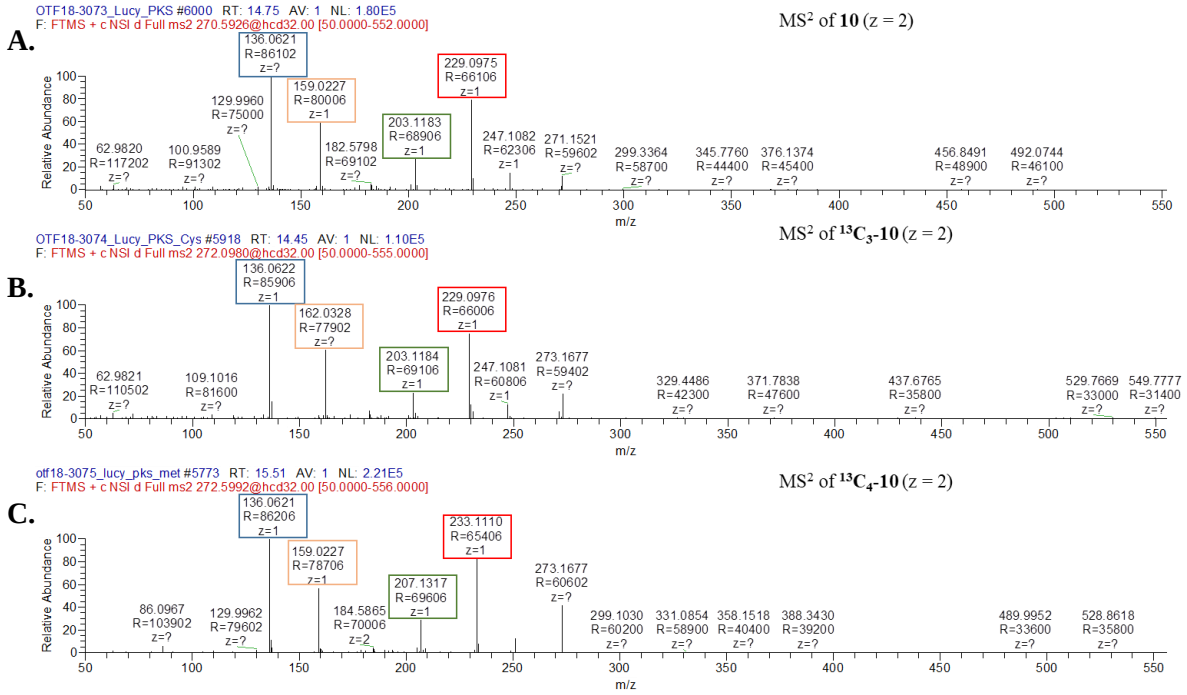


Figure S11. Experimental masses. **A**. MS^2 spectrum of **10** (z = 2), **Ad·H**$^+$ (136.0621), **14** (159.0227, z = 1), **S1** (203.1183, z = 1), **18** (229.0975, z = 1) were detected. **B**. MS^2 spectrum of ^{13}**C$_3$-10** (z = 2), **Ad·H**$^+$ (136.0622, z = 1), ^{13}**C$_3$-14** (162.0328, z = 1), **S1** (203.1184, z = 1), **18** (229.0976, z = 1) were detected. **C**. MS^2 spectrum of ^{13}**C$_4$-10** (z = 2), **Ad·H**$^+$ (136.0621, z = 1), **14** (159.0227, z = 1), ^{13}**C$_4$-S1** (207.1317, z = 1), ^{13}**C$_4$-18** (233.1109, z = 1) were detected. **Ad** = adenine.

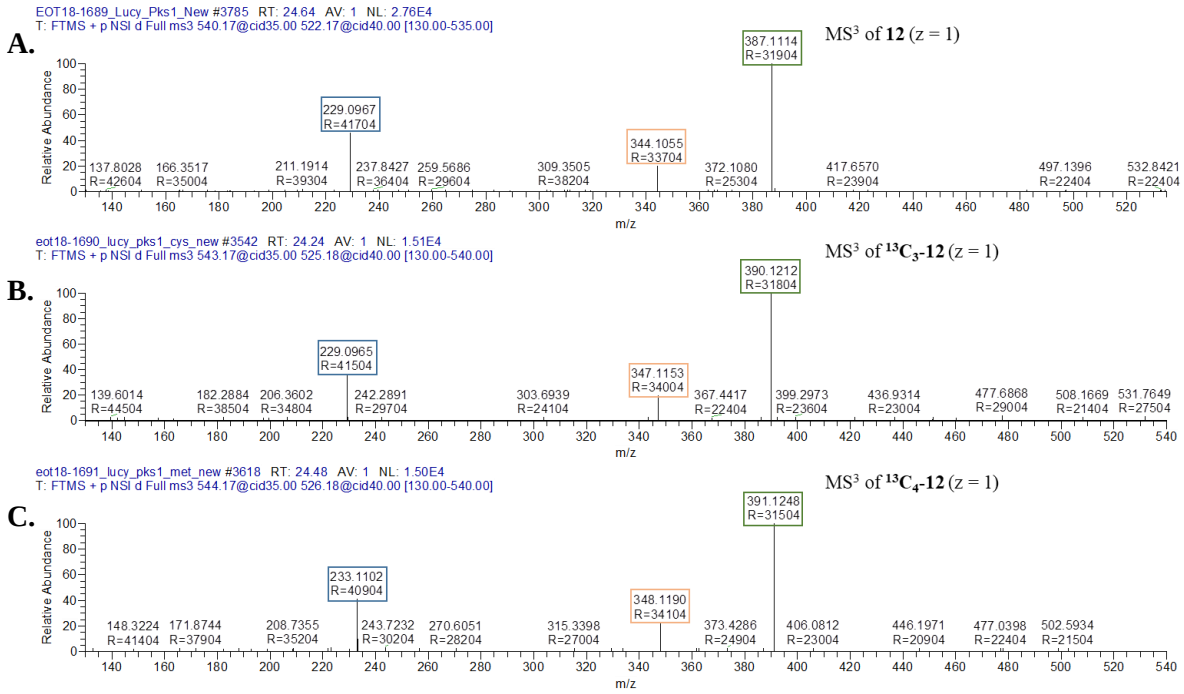


Figure S12. Experimental masses. **A**. MS^3 spectrum of **12** (z = 1), **18** (229.0967), **16** (344.1055), **17** (387.1114) were detected. **B**. MS^3 spectrum of **$^{13}C_3$-12** (z = 1), **18** (229.0965), **$^{13}C_3$-16** (347.1153), **$^{13}C_3$-17** (390.1212) were detected. **C**. MS^3 spectrum of **$^{13}C_4$-12** (z = 1), **$^{13}C_4$-18** (233.1102), **$^{13}C_4$-16** (348.1190), **$^{13}C_4$-17** (391.1248) were detected.

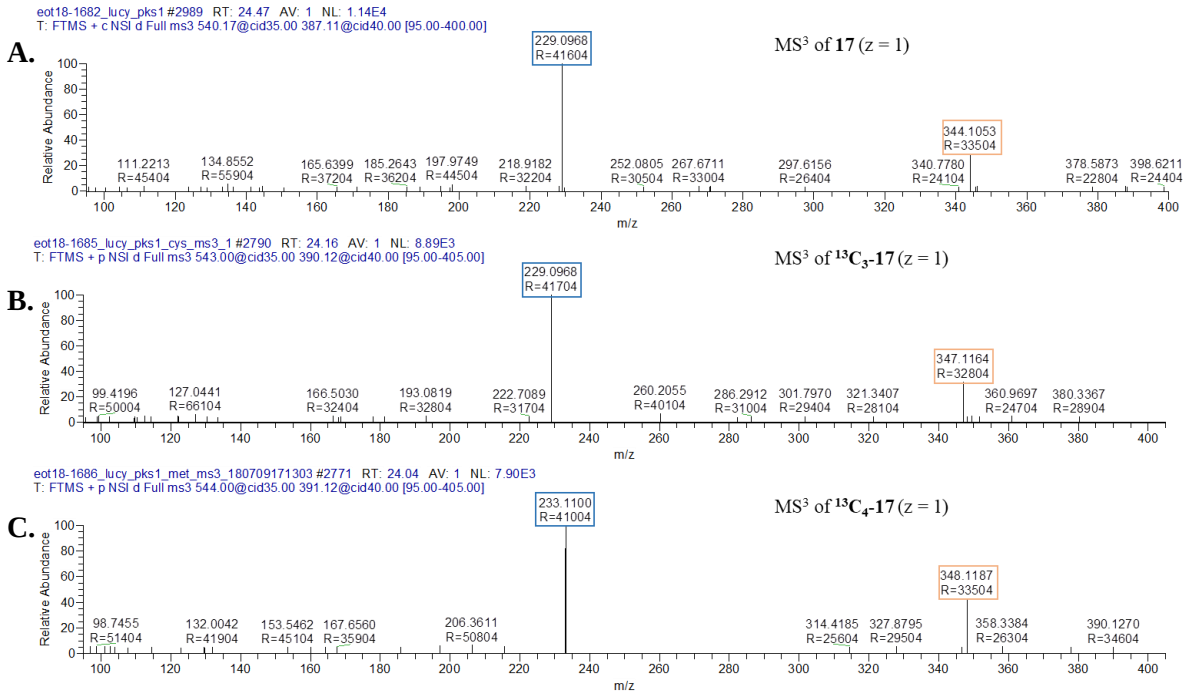


Figure S13. Experimental masses. **A**. MS^3 spectrum of **17** (z = 1), **18** (229.0968), **16** (344.1053) were detected. **B**. MS^3 spectrum of **$^{13}C_3$-17** (z = 1), **18** (229.0968), **$^{13}C_3$-16** (347.1164) were detected. **C**. MS^3 spectrum of **$^{13}C_4$-17** (z = 1), **$^{13}C_4$-18** (233.1100), **$^{13}C_4$-16** (348.1187) were detected.

strain	compound	z	experimental	theoretical	error (ppm)	Originated from
clb$^+$	**10**	1	540.1775	540.1772	0.56	full HRMS
	12	1	522.1655	522.1666	2.11	MS^2 of **10** (z = 1)
	13	1	405.1220	405.1227	1.73	MS^2 of **10** (z = 1)
	17	1	387.1126	387.1122	1.03	MS^2 of **10** (z = 1)
	17	1	387.1114	387.1122	2.07	MS^3 of **12**
	16	1	344.1068	344.1063	1.45	MS^2 of **10** (z = 1)
	16	1	344.1055	344.1063	2.32	MS^3 of **12**
	16	1	344.1053	344.1063	2.91	MS^3 of **17**
	18	1	229.0975	229.0972	1.31	MS^2 of **10** (z =1 & 2)
	18	1	229.0967	229.0972	2.18	MS^3 of **12**
	18	1	229.0966	229.0972	2.62	MS^3 of **17**
	14	1	159.0224	159.0223	0.63	daughter ion of **10**
	14	1	159.0227	159.0223	2.52	MS^2 of **10 (**z = 2)
	15	1	382.1627	382.1622	1.31	daughter ion of **10**
	19	1	364.1517	364.1516	0.27	daughter ion of **10**
	S1	1	203.1183	203.1179	1.97	MS^2 of **10** (z = 2)
	Ad·H$^+$	1	136.0621	136.0618	2.20	MS^2 of **10 (**z = 2)
	Ad	neutral loss	135.0541	135.0545	2.96	neutral loss from **12** to **17**

Table S1. Full MS^2 and MS^3 data of **10**, derived from exposure of linearized pUC19 DNA to *clb*$^+$ *E. coli*. The experimental values, theoretical values, and errors in ppm are listed in the table.

strain	compound	z	experimental	theroratical	error (ppm)	Originated from
clb$^+$Δ*cys*E	**$^{13}C_3$-10**	1	543.1874	543.1873	0.28	full HRMS
	$^{13}C_3$-12	1	525.1754	525.1767	2.38	MS^2 of **$^{13}C_3$-10** (z = 1)
	$^{13}C_3$-13	1	408.1319	408.1328	2.08	MS^2 of **$^{13}C_3$-10** (z = 1)
	$^{13}C_3$-17	1	390.1228	390.1223	1.41	MS^2 of **$^{13}C_3$-10** (z = 1)
	$^{13}C_3$-17	1	390.1212	390.1223	2.69	MS^3 of **$^{13}C_3$-12**
	$^{13}C_3$-16	1	347.1168	347.1164	1.30	MS^2 of **$^{13}C_3$-10** (z = 1)
	$^{13}C_3$-16	1	347.1153	347.1164	3.02	MS^3 of **$^{13}C_3$-12**
	$^{13}C_3$-16	1	347.1153	347.1164	3.02	MS^3 of **$^{13}C_3$-17**
	18	1	229.0975	229.0972	1.31	MS^2 of **$^{13}C_3$-10** (z = 1 & 2)
	18	1	229.0965	229.0972	3.06	MS^3 of **$^{13}C_3$-12**
	18	1	229.0968	229.0972	1.75	MS^3 of **$^{13}C_3$-17**
	$^{13}C_3$-14	1	162.0328	162.0324	2.78	daughter ion of **$^{13}C_3$-10**
	$^{13}C_3$-14	1	162.0328	162.0324	2.78	MS^2 of **$^{13}C_3$-10** (z = 2)
	15	1	382.1626	382.1622	1.05	daughter ion of **$^{13}C_3$-10**
	19	1	364.1516	364.1516	0.00	daughter ion of **$^{13}C_3$-10**
	S1	1	203.1184	203.1179	2.46	MS^2 of **$^{13}C_3$-10** (z = 2)
	Ad·H$^+$	1	136.0622	136.0618	2.94	MS^2 of **$^{13}C_3$-10** (z = 2)
	Ad	neutral loss	135.0542	135.0545	2.22	neutral loss from **$^{13}C_3$-12** to **$^{13}C_3$-17**

Table S2. Full MS^2 and MS^3 of **$^{13}C_3$-10**, derived from exposure of linearized pUC19 DNA to Δ*cys*E *clb*$^+$ *E. coli* and M9 media supplemented with L-[U]-^{13}C-cys. The experimental values, theoretical values, and errors in ppm are listed in the table.

strain	compound	z	experimental	theroratical	error (ppm)	Originated from
clb$^+$Δ*met*A	**$^{13}C_4$-10**	1	544.1906	544.1906	0.00	full HRMS
	$^{13}C_4$-12	1	526.1788	526.1800	2.28	MS^2 of **$^{13}C_4$-10** (z = 1)
	$^{13}C_4$-13	1	409.1356	409.1361	1.22	MS^2 of **$^{13}C_4$-10** (z = 1)
	$^{13}C_4$-17	1	391.1261	391.1256	1.28	MS^2 of **$^{13}C_4$-10** (z = 1)
	$^{13}C_4$-17	1	391.1248	391.1256	2.05	MS^3 of **$^{13}C_4$-12**
	$^{13}C_4$-16	1	348.1202	348.1197	1.44	MS^2 of **$^{13}C_4$-10** (z = 1)
	$^{13}C_4$-16	1	348.1190	348.1197	2.01	MS^3 of **$^{13}C_4$-12**
	$^{13}C_4$-16	1	348.1187	348.1197	2.87	MS^3 of **$^{13}C_4$-17**
	$^{13}C_4$-18	1	233.1109	233.1106	1.29	MS^2 of **$^{13}C_4$-10** (z = 1 & 2)
	$^{13}C_4$-18	1	233.1102	233.1106	1.72	MS^3 of **$^{13}C_4$-12**
	$^{13}C_4$-18	1	233.1100	233.1106	2.57	MS^3 of **$^{13}C_4$-17**
	14	1	159.0224	159.0223	0.63	daughter ion of **$^{13}C_4$-10**
	14	1	159.0227	159.0223	2.52	MS^2 of **$^{13}C_4$-10** (z = 2)
	$^{13}C_4$-15	1	386.1756	386.1756	0.00	daughter ion of **$^{13}C_4$-10**
	$^{13}C_4$-19	1	368.1650	368.1650	0.00	daughter ion of **$^{13}C_4$-10**
	$^{13}C_4$-S1	1	207.1317	207.1313	1.93	MS^2 of **$^{13}C_4$-10** (z = 2)
	Ad·H$^+$	1	136.0621	136.0618	2.20	MS^2 of **$^{13}C_3$-10** (z = 2)
	Ad	neutral loss	135.0540	135.0545	3.70	neutral loss from **$^{13}C_4$-12** to **$^{13}C_4$-17**

Table S3. Full MS^2 and MS^3 of **$^{13}C_4$-10**, derived from exposure of linearized pUC19 DNA to Δ*met*A *clb*$^+$ *E. coli* and M9 media supplemented with L-[U]-^{13}C-met. The experimental values, theoretical values, and errors in ppm are listed in the table.

LC-MS/MS Instrumentation.

LC-MS/MS data were obtained at the Mass Spectrometry and Proteomics Resource of the W.M. Keck Foundation Biotechnology Resource Laboratory at Yale University (New Haven, CT). All HRMS/MS samples were prepared in 1-mL screw neck total recovery vials (Waters, Milford, MA). The concentration of the digested nucleosides was adjusted to 50 ng/μL before injection. 5 μL of sample was injected at 4 ºC. UPLC analysis was performed on an AcQuity M-Class Peptide BEH C18 column (130 Å pore size, 1.7 μm particle size, 75 μm × 250 mm) equipped with an M-Class Symmetry C18 trap column (100 Å pore size, 5 μm particle size, 180 μm × 20 mm) at 37 ºC. Trapping was performed starting at 5 μL/min at 99.5% of aqueous mobile phase (0.1% formic acid in water) for 3 min, and the gradient for separation started at 3% organic mobile phase (0.1% formic acid in acetonitrile), and increased to 5% over 1 min, 25% over 32 min, 50% over 5 min, 90% over 5 min and then maintained at 90% for 50 min and then 3% over 2 min, and equilibrated for an additional 20 min. Mass spectrometry was acquired on an Orbitrap Elite FTMS (Thermo Scientific) or on an Orbitrap Fusion FTMS (Thermo Scientific). The Orbitrap Elite FTMS (Thermo Scientific) was set as full scan from m/z = 150–1800 at a resolution ranging from 30,000 to 60,000, and the data-dependent MS^2 and MS^3 scans were collected with CID (collision-induced dissociation) at collision energy ranging from 35 eV to 40 eV. The Orbitrap Fusion FTMS (Thermo Scientific) was set as full scan from m/z = 150−1000 with a resolution of 60,000, and the data-dependent MS^2 scans were collected with HCD (higher-energy collisional dissociation) at 32 eV collision energy using quadrupole isolation. Data was analyzed using the Thermo Xcalibur Qual Browser software (version 2.2).

Experimental Procedures

Cell lines.

E. coli strains with incorporated BAC genes used in this study were prepared by the Crawford laboratory at Yale University (New Haven, CT). This includes the *E.coli* K-12 BW25113 parent strain with the single gene knock out from BW25113: JW3973-1 (Δ*metA780::kan*) and JW3582-2 (Δ*cysE720::kan*). The isolated BAC DNA (pBAC *clb*$^+$ and *clb*$^-$) were separately incorporated into the BW25113 parent strain, the JW3582-2 cysteine auxotroph, and the JW3973-1 methionine auxotroph.

DNA and nucleic acids.

The 2686 bp plasmid pUC19 was purchased from New England Biolabs and linearized with the endonuclease EcoRI (New England Biolabs, 5 U/μg DNA). The linearized plasmid was purified using the Monarch® PCR and DNA Cleanup Kit (New England Biolabs) and eluted with 10 mM Tris–1 mM EDTA pH 8.0 buffer.

Preparation of media.

Universally-labeled ([U-^{13}C]) cysteine ($^{13}C_3$ Cys, 99% ^{13}C) and methionine ($^{13}C_5$ Met, 99% ^{13}C) were purchased from Cambridge Isotope Laboratories. The labeled amino acids were separately incorporated into modified M9 minimum media for culturing the corresponding auxotrophs JW3582-2 (cysteine) and JW3973-1 (methionine). Natural abundance cysteine and methionine were incorporated into modified M9 minimum media for culturing the BW25113 parent strain as a control. To prepare the media, the M9 minimum media (Difco) was supplemented with 0.4% glucose, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 12.5 μg/ml chloramphenicol, and the following L-amino acid mass composition (5 g/L total): 3.6% Arg, 21.1% Glu, 2.7% His, 5.6% Ile, 8.4% Leu, 7.5% Lys, 4.6% Phe, 9.9% Pro, 4.2% Thr, 1.1% Trp, 6.1% Tyr, 5% Val, 4% Asn, 4% Ala, 4% Met, 4% Gly, 4% Cys, and 4% Ser.

Preparation of DNA cross-links.

For each DNA cross-link preparation, 3200 ng of linearized plasmid DNA was added to 800 μL of modified M9 media (containing the appropriate isotopically labeled amino acid for each auxotroph) inoculated with 2.4×10^7 bacteria growing in exponential phase . The DNA–bacteria mixture was incubated for 4.5 h at 37 ºC. The bacteria were then pelleted. The DNA was isolated from the supernatant using the Monarch® PCR and DNA Cleanup Kit (New England Biolabs) and eluted using ultra purified water (Invitrogen). The isolated DNA was stored at −20 ºC. To verify the presence of a DNA cross-link, a small quantity of DNA was analyzed by denaturing electrophoresis. To prepare the positive control for cross-linked DNA, 200 ng of linearized pUC19 DNA was treated with 100 μM of cisplatin (Biovision) in 10 mM sodium citrate pH 5 buffer with 5% final DMSO concentration. Cross-linking with cisplatin was conducted for 3 h at 37 °C.

Denaturing gel electrophoresis.

The concentration of each DNA sample was adjusted to 10 ng/mL using water. 5 μL (50 ng) of the DNA sample was removed and mixed with 15 μL of 0.4% denaturing buffer (0.53% sodium hydroxide, 10% glycerol, 0.013% bromophenol blue) or 1% denaturing buffer (1.33% sodium hydroxide, 10% glycerol, 0.013% bromophenol blue). The DNA was denatured for 10 min at 4 ºC and then immediately loaded onto a 1% agarose Tris Borate EDTA (TBE) gel. The samples were run in TBE buffer for 1.5 h at 90 V. The DNA was visualized by staining with Sybr® Gold (Thermo Fisher) for 2 h.

Evaluation of hydrolytic stability of clb$^+$-derived DNA cross-links.

A small quantity of the Nucleoside Digestion Mix (New England Biolabs) enzyme was inactivated by heating to 95 ºC for 20 min. For each reaction, 200 ng of non-labeled *clb*$^+$ crosslinked linearized pUC19 DNA was separately mixed with 1 µL of active Nucleoside Digestion Mix, inactivated Nucleoside Digestion mix, or ultra-pure water. 2 µL of the 10× Nucleoside Digestion Mix Reaction Buffer (500 mM sodium acetate, 1 mM zinc chloride, pH 5.4 at 25 ºC) was added into each reaction. The final volume was adjusted to 20 µL using ultra-pure water, and reactions were conducted in 50 mM sodium acetate, 1 mM zinc chloride, pH 5.4 buffer. Reactions proceeded at 37 ºC for 1 h, and then were analyzed by 0.4% and 1% denaturing gel electrophoresis.

Digestion of clb$^+$ cross-linked DNA.

Following gel verification of the DNA cross-link, 2000 ng of the remaining DNA was digested using the Nucleoside Digestion Mix (New England Biolabs) for 1 h at 37 ºC. The digested DNA was stored at −80 ºC.