Supplementary Materials for

The human gut bacterial genotoxin colibactin alkylates DNA

Matthew R. Wilson*, Yindi Jiang*, Peter W. Villalta, Alessia Stornetta, Paul D. Boudreau, Andrea Carrá, Caitlin A. Brennan, Eunyoung Chun, Lizzie Ngo, Leona D. Samson, Bevin P. Engelward, Wendy S. Garrett, Silvia Balbo, and Emily P. Balskus

* = authors contributed equally to this work

correspondence to: balskus@chemistry.harvard.edu, balbo006@umn.edu

This PDF file includes:

Materials and Methods

General materials and methods. All chemicals and solvents were purchased from Sigma Aldrich (St. Louis, MO). Illudin S was purchased from Santa Cruz Biotechnology (Dallas, TX). Isotope-labeled amino acids were purchased from Cambridge Isotope Laboratories (Tewksbury, MA). Calf thymus DNA was purchased from Sigma Aldrich. Dulbecco's phosphate buffered saline (DPBS), Dulbecco's modified eagle medium (DMEM), Antibiotic-Antimycotic (100X), RMPI-1640 with GlutaMAX™, 10,000 U/ml Pen-Strep, 0.25% Trypsin-EDTA with phenol red, UltraPure™ agarose, UltraPure™ low melting point (LMP) agarose, and SYBR™ Gold nucleic acid gel stain (10,000X) were purchased from ThermoFisher Scientific (Waltham, MA). Sylgar™ 184 silicone elastomer kit and bottomless 96-well plates were purchased from VWR (Radnor, PA). GelBond® Film was obtained from Lonza (Portsmouth, NH). Fetal bovine serum (FBS) was obtained from ThermoFisher Scientific (Waltham, MA) or Atlanta Biologicals, Inc. (Flowery Branch, GA). Anhydrous reactions were performed using oven-dried glassware, which were then cooled under vacuum and purged with nitrogen gas. Triethylamine was stored over activated 4 Å molecular sieves pellets. 4 Å molecular sieves were oven-dried overnight and then cooled under high vacuum prior to use. Analytical thin-layered chromatography (TLC) was performed using glass plates pre-coated with silica gel containing a fluorescent indicator (254 nm) (Millipore Sigma, Burlington, MA). TLC plates were analyzed with 254 nm UV light and/or ninhydrin treatment. Silica gel for column chromatography was purchased from Silicycle (SiliaFlash® P60, 230-400 mesh).

Proton nuclear magnetic resonance (^{1}H NMR), carbon nuclear magnetic resonance (^{13}C NMR), and two-dimensional NMR (2D NMR) spectra were recorded on a Varian Inova-500 spectrometer (500 MHz, 125 MHz) or Varian Inova-600 (600 MHz, 150 MHz) NMR spectrometer for the synthetic compounds. 2D NMR spectra of DNA adducts **6** and **7** were obtained on a Bruker AVANCE II 600 MHz spectrometer equipped with a TXO cryoprobe (Bruker-BioSpin Corporation, Billerica, MA). Chemical shifts (δ) are reported in parts per million (ppm) using the solvent resonance as an internal standard for ^{1}H (C_2D_6OS = 2.50 ppm) and ^{13}C (C_2D_6OS = 39.5 ppm). Data are reported as follows: chemical shift, multiplicity (br = broad, s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet), coupling constants in Hertz (Hz), integration, and structural assignment. Proton and carbon data were assigned using 2D NMR spectroscopy techniques [^{1}H-^{1}H COSY (Correlation Spectroscopy); ^{1}H-^{13}C HSQC (Heteronuclear Single Quantum Coherence), and ^{1}H-^{13}C HMBC (Heteronuclear Multiple Bond Connectivity)]. NMR solvents were purchased from Cambridge Isotope Laboratories and Sigma Aldrich. NMR spectra were visualized using MestReNova, version 10.0.0-14411.

Infrared (IR) spectra were recorded on a Bruker Tensor 27 FT-IR spectrophotometer, and absorption frequencies are reported in reciprocal centimeters (cm^{-1}).

Optical rotation data were obtained using a 1 mL cell with a 0.5 dm path length on a Jasco P-2000 polarimeter equipped with a sodium (589 nm, D) lamp.

Optical densities of *E. coli* cultures were determined with a DU 730 Life Sciences UV/Vis spectrophotometer (Beckman Coulter) by measuring absorbance at 600 nm.

DH10B *E. coli* harboring pBeloBAC11-*pks* and the clinical *pks*$^+$ *E. coli* strain (CCR20) were obtained from the Bonnet lab, Laboratoire de Bactériologie Clinique, Centre Hospitalier de Clermont-Ferrand, Clermont-Ferrand F-63003, France. The parent strain *E. coli* BW25113, its methionine auxotrophic derivative (*E. coli* JW3973-1), and its cysteine auxotrophic derivative (*E. coli* JW3582-2) were obtained from the Coli Genetic Stock Center (CGSC, Yale University, New Haven, CT). The *pks*$^+$ *E. coli* uropathogenic strain CFT073 was obtained from the American Type Culture Collection (ATCC, Manassas, VA).

HeLa cells were obtained from ATCC (Manassas, VA). The human colon adenocarcinoma cell line HT29 was obtained from Prof. Christina Woo (Harvard University, Cambridge, MA). The human B-lymphoblastoid cell line TK6 was obtained from Prof. William G. Thilly (Massachusetts Institute of Technology, Cambridge, MA).

Analytical HPLC was performed on a Dionex UltiMate 3000 instrument (Thermo Scientific, Waltham, MA). Preparative HPLC was performed on a Dionex UltiMate 3000 instrument (Thermo Scientific, Waltham, MA).

For synthetic compounds, high-resolution mass spectral data was obtained on an Agilent MicroQTOF-QII fitted with a dual-spray electrospray ionization (ESI) source. The capillary voltage was set to 4.5 kV in the positive-ion mode, the drying gas temperature was maintained at 275 ºC with a flow rate of 8 L/min, and a nebulizer pressure of 35 psi. The liquid chromatography (LC) was performed using an Agilent Technologies 1100 series LC with H_2O and acetonitrile as solvent.

Supplementary Text

Generation of parent and auxotrophic derivative *E. coli* strains harboring colibactin biosynthesis genes. The pBeloBAC11-*pks* BAC was isolated from *E. coli* DH10B harboring pBeloBAC11-*pks* using a BAC DNA Miniprep Kit (Zymo Research, Irvine, CA). The empty pBeloBAC11 BAC was isolated from *E. coli* K12 ER2420 harboring pBeloBAC11 (New England Biolabs, Ipswich, MA) using a Plasmid Miniprep Kit (Qiagen, Hilden, Germany). To generate a parent *E. coli* strain harboring colibactin biosynthesis genes, *E. coli* BW25113 was electroporated with pBeloBAC11-*pks* or pBeloBAC11 and selected on LB agar plates containing 25 μg/mL chloramphenicol. To generate auxotrophic *E. coli* strains harboring colibactin biosynthesis genes, an *E. coli* JW3973-1 or *E. coli* JW3582-2 strain was electroporated with pBeloBAC11-*pks* or pBeloBAC11 and selected on LB agar plates containing 50 μg/mL kanamycin and 25 μg/mL chloramphenicol.

Preparation of *E. coli* strains for HeLa cell infection and DNA adduct detection.

No isotope labeling

A starter culture of an *E. coli* strain (either *E. coli* BW25113 or the methionine auxotrophic BW25113 derivative *E. coli* JW3973-1) harboring pBeloBAC11-*pks* (*pks*$^+$) or empty pBeloBAC11 (*pks*$^-$) was inoculated from a fresh single colony and grown overnight at 37 ºC in sterile LB medium supplemented with 25 µg/mL chloramphenicol. The resulting saturated starter culture was centrifuged (9,000 rpm x 5 min at 4 ºC). The bacterial pellets were washed and resuspended in 1X DPBS. 10 mL of fresh 1X M9 minimal medium containing 25 µg/mL chloramphenicol, 0.1 mM $CaCl_2$, 2 mM $MgSO_4$, 4 g/L glucose, 0.2 g/L Ala, 0.18 g/L Arg, 0.2 g/L Asn, 0.2 g/L Cys, 1.06 g/L Glu, 0.2 g/L Gly, 0.14 g/L His, 0.28 g/L Ile, 0.42 g/L Leu, 0.38 g/L Lys, 0.23 g/L Phe, 0.50 g/L Pro, 0.2 g/L Ser, 0.21 g/L Thr, 0.06 g/L Trp, 0.31 g/L Tyr, 0.25 g/L Val, and 0.2 g/L unlabeled Met was inoculated with a normalized number of cells, such that an OD_{600} of 1 of the overnight starter culture gave a 1:100 volume of inoculum. The culture was incubated at 37 °C with 200 rpm shaking for 15 h, transferred into 50 mL Corning tubes, and centrifuged (4,000 rpm x 10 min at 4 ºC). The cell pellets were resuspended in 12 mL of infection medium (DMEM supplemented with 10% FBS and 25 mM HEPES (1 M stock)) for HeLa cell infection. DNA adduct formation did not differ between *pks*$^+$ *E. coli* BW25113 and *pks*$^+$ *E. coli* JW3973-1. LC-MS data shown in Fig. 2A-C are from experiments with *pks*$^+$ *E. coli* JW3973-1.

Alanine and glycine isotope labeling experiment

A starter culture of methionine auxotrophic *E. coli* JW3973-1 harboring pBeloBAC11-*pks* or empty pBeloBAC11 was inoculated from a fresh single colony and grown overnight at 37 ºC in LB medium supplemented with 50 µg/mL kanamycin and 25 µg/mL chloramphenicol. The starter culture was centrifuged (9,000 rpm x 5 min). The bacterial pellets were washed and resuspended in 1X DPBS. 10 mL of fresh 1 X M9 minimal medium containing 25 µg/mL chloramphenicol, 50 µg/mL kanamycin, 0.1 mM $CaCl_2$, 2 mM $MgSO_4$, 4 g/L glucose, 0.2 g/L Ala, 0.18 g/L Arg, 0.2 g/L Asn, 0.2 g/L Cys, 1.06 g/L Glu, 0.2 g/L Gly, 0.14 g/L His, 0.28 g/L Ile, 0.42 g/L Leu, 0.38 g/L Lys, 0.2 g/L Met, 0.23 g/L Phe, 0.50 g/L Pro, 0.2 g/L Ser, 0.21 g/L Thr, 0.06 g/L Trp, 0.31 g/L Tyr, 0.25 g/L Val, and 1 g/L **L-[2,3-$^{13}C_2$]alanine** or **[1,2-$^{13}C_2$]glycine** was inoculated with a normalized number of cells, such that an OD_{600} of 1 of the overnight starter culture gave a 1:100 volume of inoculum. The culture was incubated at 37 °C with 200 rpm shaking for 15 h, transferred into 50 mL Corning tubes, and centrifuged (4,000 rpm x 10 min at 4 ºC). The cell pellets were resuspended in 12 mL DMEM medium supplemented with 10% FBS, 25 mM HEPES, and 1 g/L **L-[2,3-$^{13}C_2$]alanine** or **[1,2-$^{13}C_2$]glycine** for HeLa cell infection.

Methionine isotope labeling experiment

A starter culture of methionine auxotrophic *E. coli* JW3973-1 harboring pBeloBAC11-*pks* or empty pBeloBAC11 was inoculated from a fresh single colony and grown overnight at 37 ºC in LB medium supplemented with 50 µg/mL kanamycin and 25 µg/mL chloramphenicol. The resulting saturated starter culture was centrifuged (9,000 rpm x 5 min at 4 ºC). The bacterial pellets were washed and resuspended in 1X DPBS. 10 mL of fresh 1X M9 minimal medium containing 25 µg/mL chloramphenicol, 50 µg/mL kanamycin, 0.1 mM $CaCl_2$, 2 mM $MgSO_4$, 4 g/L glucose, 0.2 g/L Ala, 0.18 g/L Arg, 0.2

g/L Asn, 0.2 g/L Cys, 1.06 g/L Glu, 0.2 g/L Gly, 0.14 g/L His, 0.28 g/L Ile, 0.42 g/L Leu, 0.38 g/L Lys, 0.23 g/L Phe, 0.50 g/L Pro, 0.2 g/L Ser, 0.21 g/L Thr, 0.06 g/L Trp, 0.31 g/L Tyr, 0.25 g/L Val, and **0.2 g/L L-[1-^{13}C]methionine** was inoculated with a normalized number of cells, such that an OD_{600} of 1 of the overnight starter culture gave a 1:100 volume of inoculum. The culture was incubated at 37 °C with 200 rpm shaking for 15 h, transferred into 50 mL Corning tubes, and centrifuged (4,000 rpm x 10 min at 4 ºC). The cell pellets were resuspended in 12 mL DMEM medium supplemented with 10% FBS, 25 mM HEPES, and **0.2 g/L L-[1-^{13}C]methionine** for HeLa cell infection.

Cysteine isotope labeling experiment

A starter culture of cysteine auxotrophic *E. coli* JW3582-2 harboring pBeloBAC11-*pks* or empty pBeloBAC11 was inoculated from a fresh single colony and grown overnight at 37 ºC in LB medium supplemented with 50 µg/mL kanamycin and 25 µg/mL chloramphenicol. The resulting saturated starter culture was centrifuged (9,000 rpm x 5 min). The bacterial pellets were washed and resuspended in 1X DPBS. 10 mL of fresh 1 X M9 minimal medium containing 25 µg/mL chloramphenicol, 50 µg/mL kanamycin, 0.1 mM $CaCl_2$, 2 mM $MgSO_4$, 4 g/L glucose, 0.2 g/L Ala, 0.18 g/L Arg, 0.2 g/L Asn, 1.06 g/L Glu, 0.2 g/L Gly, 0.14 g/L His, 0.28 g/L Ile, 0.42 g/L Leu, 0.38 g/L Lys, 0.2 g/L Met, 0.23 g/L Phe, 0.50 g/L Pro, 0.2 g/L Ser, 0.21 g/L Thr, 0.06 g/L Trp, 0.31 g/L Tyr, 0.25 g/L Val, and **0.2 g/L L-[1-^{13}C]cysteine** was inoculated with a normalized number of cells, such that an OD_{600} of 1 of the overnight starter culture gave a 1:100 volume of inoculum. The culture was incubated at 37 °C with 200 rpm shaking for 15 h, transferred into 50 mL Corning tubes, and centrifuged (4,000 rpm x 10 min at 4 ºC). The cell pellets were resuspended in 12 mL DMEM medium supplemented with 10% FBS, 25 mM HEPES, and **0.2 g/L L-[1-^{13}C]cysteine** for HeLa cell infection.

HeLa cell infection. For DNA adduct detection, HeLa cells in a T75 flask were washed with DMEM medium supplemented with 10% FBS (wash medium) (2 x 6 mL) and infected with the appropriate bacterial cell suspension in infection medium for 90 min. HeLa cells were washed with 1X DPBS (3 x 6 mL), trypsinized, and centrifuged (400 g x 2 min at 4 ºC). The HeLa cell pellets were then washed again with 1X DPBS (3 x 6 mL), centrifuged (400 g x 2 min at 4 ºC), and then resuspended in 200 µL 1X DPBS.

For the studies not involving DNA adduct detection, HeLa cells in a 6-well plate were washed twice with the wash medium and infected with bacterial cell suspension in infection medium with a multiplicity of infection (MOI) of 1,000 for 60 min. HeLa cells were washed twice with 1X DPBS and incubated in growth medium (DMEM medium supplemented with 10% FBS and 1X Antibiotic-Antimycotic) supplemented with 200 µg/mL gentamicin for either one day or two days.

HeLa cell treatment with illudin S and synthetic cyclopropanes. HeLa cells were treated with illudin S (100 nM) or cyclopropane **4** (30 µM) in a T75 flask for 24 h at 37 ºC in a CO_2 incubator. After 24 h, the HeLa cells were washed with 1X DPBS (3 x 6 mL), trypsinized, and centrifuged (400 g x 2 min at 4 ºC). The HeLa cell pellets were then washed again with 1X DPBS (3 x 6 mL), centrifuged (400 g x 2 min at 4 ºC), and then resuspended in 200 µL 1X DPBS.

Genomic DNA isolation from HeLa cell pellets. DNA was isolated following a DNA purification from cultured cells protocol (Qiagen, Hilden, Germany), with some modifications. Briefly, the cell culture samples (~15 x 10^6 cells) were centrifuged at 14,000 x g for 5 min to generate a cell pellet at the bottom of the tube. The supernatants were discarded and the cell pellets were re-suspended in 3 mL of Cell Lysis Solution (Qiagen). The cell membranes were disrupted by shaking the tubes at room temperature for 24 h, or longer if necessary to assure full homogenization. RNA was digested by incubating the samples for 2 h at room temperature with 40 µL of RNase-A (Qiagen). Proteins were then precipitated by addition of 1 mL of Protein Precipitation Solution (Qiagen). Protein pellets were generated by centrifugation at 4,500 x g for 3 min, or longer if needed for full precipitation. The supernatants were saved and the pellets discarded. The DNA was precipitated from the supernatants by addition of 4 mL of cold isopropanol (IPA) (100 % v/v, 0 ºC) to each sample. A white solid precipitate consisting of DNA was formed and pelleted by centrifugation (14,000 g x 3 min, 4 ºC), or longer if needed for full precipitation. The supernatants were gently discarded. The DNA samples were further purified by re-suspension in 1 mL of IPA (70 % v/v) and centrifugation (14,000 g x 3 min, 4 ºC). The resulting DNA pellets were saved and the supernatants discarded. Final purification was performed by re-suspending the DNA samples with 1 mL of IPA (100 % v/v), centrifuging at 14,000 g x 3 min at 4 ºC to form the DNA pellet, and discarding the supernatants. The residual solvent of the DNA pellet was removed by evaporation for 60 min at room temperature. Once isolated and dried, the DNA sample was stored at –20 ºC.

DNA enzymatic digestion. The amount of DNA was estimated by dissolving the samples in Tris buffer (Trizma/$MgCl_2$ 10 and 5 mM, pH 7) and measuring the concentration by UV/Vis spectrometry. The spectrophotometer worked in absorbance optical mode, monitoring the 260 and 280 nm wavelengths. The DNA digestion was carried out using a cocktail of enzymes consisting of recombinant DNase I expressed by *Pichia pastoris* (R-DNase, 10000 $U \cdot mg^{-1}$), phosphodiesterase-1 extracted from *Crotalus adamanteus* (PDE-1, 0.4 $U \cdot mg^{-1}$), and recombinant alkaline phosphatase highly active expressed by *Pichia pastoris* (R-ALP, 7000 $U \cdot mg^{-1}$), all purchased from Roche (St. Louis, MO, USA). All enzymes were purified prior to use by passage of the enzyme solution through Amicon Ultra (MilliporeSigma, Darmstadt, Germany) double filtration membranes (0.5 mL, cutoff 10 kDa). The hydrolysis consisted of a two-step process consisting of an initial 24 h treatment with a first aliquot of DNase followed by a 24 h treatment with the full set of enzymes, both performed at room temperature. The first digestion step used 0.5 U/µg DNA of DNase to initiate lysis of the phosphodiester backbone. The second digestion step required approximately 0.5 U/µg DNA, 0.2 U/µg DNA and 0.02 mU/µg DNA of DNase, ALP and PDE-1, respectively, to bring the DNA digestion to completion. To stop the hydrolysis, the enzymes were removed using an Amicon Microcon single filtration membrane (0.5 mL, cutoff 10 kDa). The digestion time course and the yield were assessed by measuring the amount of dG via an LC/UV measurement.

dG quantitation method. Quantitation of dG was performed via LC/UV detection using an Ultimate 3000 HPLC (Thermo Scientific, Waltham, MA) with chromatographic

separation of the four 2′-deoxyribonucleosides carried out using a reversed-phase Luna C18 (Phenomenex, Torrance, CA) column (250 x 0.5 mm, 5 µm, 100 Å) maintained at 40 °C with a flow rate of 15 µL·min^{-1}. Separation was performed using H_2O (A) and MeOH (B) with an initial isocratic step of 5% B (3 min), followed by a first linear gradient of 0.58% B·min^{-1} (12 min), a second linear gradient of 27.67 % B·min^{-1} (3 min), and a final isocratic step at 95% B (3 min). Prior to a subsequent analysis, the column was re-equilibrated (9 min) with an isocratic step of 5% B. The UV detector operated in absorbance optical mode, monitoring the 254 nm wavelength. The analytes of interest (dG and G) were quantified using a calibration curve including 8 standard concentrations (0.0625, 0.125, 0.25, 0.50, 1.00, 2.00, 4.0, 8.0 ng/µL).

Hydrophobic reversed phase fraction collection. Prior to sample analysis, unmodified deoxynucleosides were removed from the enzymatic digestion by off-line HPLC fraction collection, performed using an HPLC (UltiMate 3000, Thermo Scientific, Waltham, MA) equipped with a C18 column (4.6 x 250 mm, Luna C18, 5µm 100 Å, Phenomenex, Torrance, CA) maintained at 25 °C, with a flow rate of 1.0 mL·min^{-1}. The A and B mobile phases were H_2O and MeOH, respectively, and an initial isocratic step of 2% of B (5 min) was used, followed by a linear gradient of 0.7 % B·min^{-1} (25 min) and a second isocratic step at 100% of B (15 min). UV absorption at 190 and 254 nm with 4 Hz detection was used for monitoring the elution of the deoxyribonucleosides. After elution of dA at 33.5 min, the remainder of the liquid was collected, completely dried and stored at –20 °C.

Chromatography for LC-MS (screening and fragmentation). Mass spectrometric data was acquired with the following conditions. Each dried sample was reconstituted in 20 µL of H_2O, and 5 µL of sample, corresponding to DNA amounts ranging from 10 µg (cell samples) or 50 µg (ctDNA samples), were injected onto an UltiMate 3000 RSLCnano UPLC (Thermo Scientific, Waltham, MA) system equipped with a 5 µL injection loop. Separation was performed with a capillary column (75 µm ID, 20 cm length, 10 µm orifice) created by hand packing a commercially available fused-silica emitter (New Objective, Woburn, MA) with 5 µm Luna C18 bonded separation media (Phenomenex, Torrance, CA). The flow rate was 1000 nL/min for 5.5 min, then decreased to 300 nL/min with a 25 min linear gradient from 2 to 50% CH_3CN in 0.05% formic acid aqueous solution, an increase to 98% CH_3CN in 1 min, with a 4 min hold and a 5 min re-equilibration at 1000 nL/min at 2% CH_3CN. The injection valve was switched at 5.5 min to remove the sample loop from the flow path during the gradient.

Chromatography for LC-MS (targeted detection). Mass spectrometric data was acquired with the following conditions. Each dried sample was reconstituted in 3 µL of H_2O and 1 µL of sample, corresponding to DNA amounts ranging from 32 – 81 µg (mouse samples), was injected onto an UltiMate 3000 RSLCnano UPLC (Thermo Scientific, Waltham, MA) system equipped with a 5 µL injection loop. Separation was performed with a capillary column (50 µm ID, 20 cm length, 10 µm orifice) created by hand packing a commercially available fused-silica emitter (New Objective, Woburn, MA) with 5 µm Luna C18 bonded separation media (Phenomenex, Torrance, CA). The flow rate was 500 nL/min for 11 min, then decreased to 150 nL/min with a 7 min linear

gradient from 2 to 25% CH_3CN in 0.05% formic acid aqueous solution, an increase to 98% CH_3CN in 1 min, with a 2 min hold and a 3 min re-equilibration at 500 nL/min at 2% CH_3CN. The injection valve was switched at 12 min to remove the sample loop from the flow path during the gradient.

Mass Spectrometry for DNA adductomics. All mass spectrometric data was acquired with a Fusion mass spectrometer (Thermo Scientific, Waltham, MA). Positive mode electrospray ionization was used under nanospray conditions (150 and 300 nL/min) using a Thermo Scientific Nanoflex ion source with a source voltage of 2.2 kV, and the capillary temperature was 300 °C. The S-Lens RF level setting was 60%.

DNA adduct screening. Data-dependent constant neutral loss (CNL)-MS^3 analysis was performed with repeated full scan detection followed by MS^2 acquisition and constant neutral loss triggering of MS^3 fragmentation. Full scan (300−1000 *m/z*) detection was performed using the Orbitrap detector at a resolution setting of 60,000, automatic gain control (AGC) target settings of 2×10^5, and a maximum ion injection time setting of 118 ms. MS^2 spectra were acquired with quadrupole isolation of 1.5 *m/z* and HCD fragmentation with stepped collision energy of 15% with +/- 10% (5, 15, 25%) of the top 10 most intense full scan ions with Orbitrap detection at a resolution setting of 15000, AGC setting of 5×10^4, maximum ion injection time of 200 ms. Data-dependent parameters were as follows: triggering range of 2.5×10^4 to 1.0×10^7, repeat count of 1, exclusion duration of 15 s, and exclusion mass width of ±5 ppm. Three time-specific targeted mass exclusion lists (0 - 15 min, 15 - 25 min, and 25 - 45 min), each consisting of the 500 most intense ions observed in the full scan ctDNA sample analysis, were included to provide more sensitive adduct detection. MS^3 HCD fragmentation (2.5 *m/z* isolation width, HCD collision energy of 30%) with Orbitrap detection at a resolution setting of 15000 was triggered upon observation of neutral losses (±5 ppm) of 116.0474, 151.0494, 135.0545, 126.0429, and 111.0433 *m/z* between the parent ion and product ions from the MS^2 spectrum, provided a minimum signal of 1×10^4 was observed. The following MS^3 parameters were used: AGC setting of 5×10^4, maximum ion injection time of 200 ms. All spectra were acquired with the EASY-IC lock mass (202.0777 *m/z*) enabled.

DNA adduct fragmentation data acquisition. Targeted MS^2 and MS^3 spectra were acquired (for 540, 541, 542 *m/z*) using an MS^2 quadrupole isolation window of 1.5 *m/z* and an MS^2 HCD fragmentation of 15%. Subsequent MS^3 fragmentations were performed with a 2.5 *m/z* ion trap isolation of the MS^2 fragment ions and an MS^3 HCD fragmentation of 30%. Resolution settings of 15000, AGC settings of 5×10^4, and maximum injection times of 200 ms were used. All spectra were acquired with the EASY-IC lock mass (202.0777 *m/z*) enabled.

DNA adduct targeted detection. Targeted MS^2 detection was acquired with fragmentation of 540.2 *m/z* with MS^2 quadrupole isolation window of 1.5 *m/z* and MS^2 HCD fragmentation using a stepped collision energy of 15% with +/- 10% (5, 15, 25%). Resolution settings of 120,000, AGC settings of 1×10^6, and maximum injection times of

1000 ms were used. Spectra were acquired with the EASY-IC lock mass (202.0777 *m/z*) enabled.

HT29 cell infection. HT29 cells were infected with *pks*$^+$ *E. coli* JW3973-1 using the HeLa cell infection procedure described above. DNA adducts **1** and **2** were detected using the methods described above.

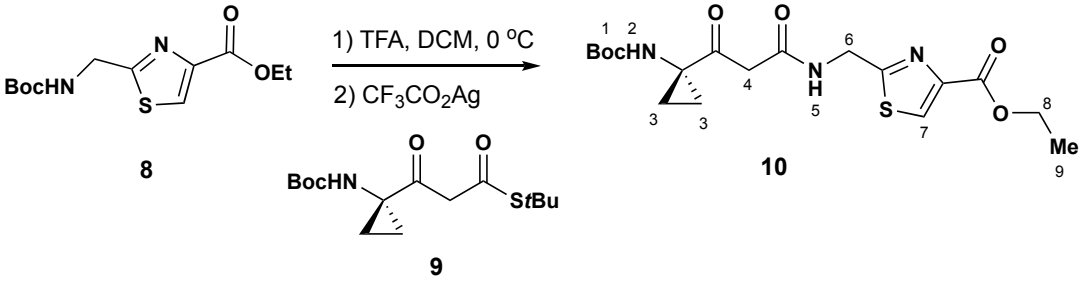
HeLa cell infection native colibactin-producing strains. HeLa cells were infected with native *pks*$^+$ *E. coli* strains CCR20 and CFT073 using the HeLa cell infection procedure described above. DNA adducts **1** and **2** were detected using the methods described above.

Bacterial monocolonization of germ-free mice. Overnight pBelo and *pks*$^+$ *E. coli* BW25113 cultures were labeled as "A" and "B". Testers involved in colonization of germ-free mice, isolation of colonic epithelial cells, DNA extraction from colonic epithelial cells, and DNA adduct detection were blinded to the bacterial genotypes. All samples were decoded after the DNA adduct detection procedure. Germ-free C57BL/6J mice, aged 5 to 10 weeks, were gavaged with ~5 x 10^7 colony-forming units of strains "A" and "B" in a volume of 100 μL of PBS. Mice were maintained for two weeks under a strict 12-hour light cycle in individually-ventilated isolator cages at the Harvard T. H. Chan Gnotobiotic Center for Mechanistic Microbiome Studies. Over the course of the experiment, mice were exposed to 10 μg/mL chloramphenicol in the drinking water, *ad libitum*, to maintain selection of the BAC plasmids harbored by these strains. At the end of the experiment, bacterial colonization was measured by dilution-plating of homogenized stool samples on 25 μg/mL chloramphenicol LB plates (table S1). For mouse experiments, a minimum of three experiments were conducted to generate colonic epithelial cell DNA for targeted LC-MS/MS method development and experimental analysis. All experiments were approved and carried out in accordance with Harvard Medical School's Standing Committee on Animals and the National Institutes of Health guidelines for animal use and care.

Isolation of colonic epithelial cells. Colons were dissected and fat and blood vessels were removed. Colons were cut open longitudinally and washed with PBS to remove feces and debris, then incubated in PBS containing 5 mM EDTA, 0.145 mg/ml dithiothreitol and 2% FBS for 10 min on ice. The tubes were gently inverted every few minutes. Then, colons were placed in pre-warmed PBS containing 5 mM EDTA and 2% FBS and incubated for 15 min at 37 ºC on a shaking platform. After being vortexed for 15 s, colon tissues were passed through 100 μm strainers and the dissociated cells were collected as colonic epithelial cells. After repeating the incubation in pre-warmed PBS containing 5 mM EDTA and 2% FBS and passing through 100 μm strainers, colonic epithelial cells were pooled and resuspended in PBS for DNA extraction.

Experimental procedures for preparing synthetic cyclopropanes.

Preparation of β-ketoamide **10**

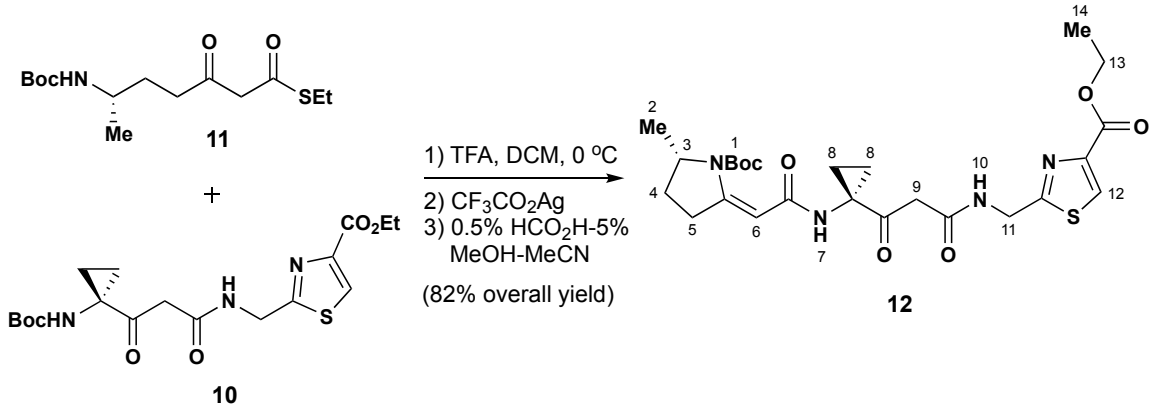


(59% overall yield, 2 steps)

Trifluoroacetic acid (17.5 mL, 228 mmol, 65 equiv) was added dropwise to a cooled solution of the known Boc-protected amine **8** *(40)* (1.00 g, 3.49 mmol, 1 equiv) in dry CH_2Cl_2 (17.5 mL, 0.2 M) at 4 ºC. The resulting reaction mixture was stirred at 4 ºC until TLC indicated the reaction was complete (~1 h). Once the reaction was complete, dry PhMe (20 mL) was added to the reaction mixture. The resulting solution was then concentrated *in vacuo* to give a dark brown/red viscous oil. The crude TFA salt was taken to the next step without further purification.

Dry DMF (18 mL, 0.2 M) was added to the crude TFA salt followed by NEt_3 (1.94 mL, 13.9 mmol, 4.0 equiv) and known β-ketothioester **9** (*40*) (1.31 g, 4.18 mmol, 1.2 equiv). The resulting solution was cooled to 4 ºC and CF_3CO_2Ag (0.921 g, 4.18 mmol, 1.2 equiv) was added in three separate portions every 20 min (total of 1 h). The reaction mixture was then stirred at 4 ºC for an additional 30 min (total reaction time of 1.5 h). After 1.5 h, EtOAc (20 mL) and H_2O (10 mL) were added to the reaction mixture. To remove the resulting precipitates, the combined EtOAc and H_2O solution was transferred to a 50 mL falcon tube and centrifuged (4,000 rpm x 10 min at 23 ºC). The layers were then separated. The aqueous layer was extracted with EtOAc (2 x 20 mL). The combined organic layers were washed with H_2O (2 x 30 mL), dried over Na_2SO_4, and then filtered. The filtrate was then concentrated *in vacuo* to give 1.51 g of a viscous dark red/purple oil. Purification of the crude material by flash chromatography (25% EtOAc/75% Hexanes to 100% EtOAc) on silica gel afforded compound **10** (0.848 g, 59%, 2 steps) as a white solid. (R_f = 0.50; visualized on silica gel; 10% MeOH/90% CH_2Cl_2, UV lamp). ^{1}H NMR (600 MHz DMSO-d_6): δ (ppm) = 8.94 (t, J = 5.9 Hz, 1H, H_5); 8.43 (s, 1H, H_7); 7.75 (s, 1H, H_2); 4.56 (d, J = 5.9 Hz, 2H, H_6); 4.30 (q, J = 7.1 Hz, 2H, H_8); 3.55 (s, 2H, H_4); 1.40 (s, 9H, H_1); 1.37 – 1.35 (m, 2H, H_3); 1.30 (t, J = 7.1 Hz, 3H, H_9); 1.08 – 1.06 (m, 2H, H_3). ^{13}C NMR (125 MHz, DMSO-d_6): δ (ppm) = 204.7, 170.2, 166.8, 160.7, 156.0, 145.5, 129.4, 78.6, 60.7, 46.1, 41.2, 40.5, 28.1, 19.5, 14.2. IR (ATR-FIR), cm^{-1}: 3309, 2979, 2934, 1702, 1663, 1500, 1237, 817. HRMS (ESI): calcd for $C_{18}H_{26}N_3O_6S$ $[M+H]^+$, 412.1537; found, 412.1516.

Preparation of enamide **12**



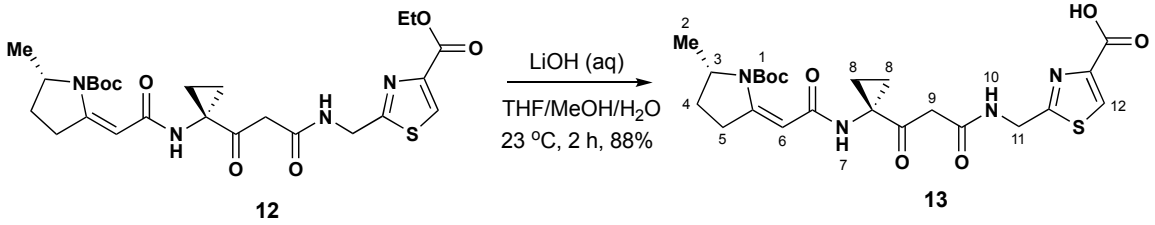
Trifluoroacetic acid (4.50 mL, 58.7 mmol, 68 equiv) was added dropwise to a cooled solution of the Boc-protected amine **10** (0.353 g, 0.858 mmol, 1 equiv) in dry CH_2Cl_2 (4.5 mL, 0.2 M) at 4 ºC. The resulting reaction mixture was stirred at 4 ºC until TLC deemed the reaction was complete (~2.5 h). Once the reaction was complete, dry PhMe (6 mL) was added to the reaction mixture. The resulting solution was then concentrated *in vacuo* to give a dark brown/red viscous oil. The crude TFA salt was taken to the next step without further purification.

Dry DMF (17 mL, 0.05 M) was added to the crude TFA salt followed by NEt_3 (0.478 mL, 3.43 mmol, 4.0 equiv) and CF_3CO_2Ag (0.494 g, 2.24 mmol, 1.2 equiv). The reaction mixture was cooled to 4 ºC using an ice bath and then a solution of the known β-ketothioester **11** (*22*) (0.341 g, 1.12 mmol, 1.2 equiv) in dry DMF (5 mL) was added dropwise. The reaction mixture was then stirred at 4 ºC for 1 h. After 1 h, EtOAc (10 mL) and H_2O (10 mL) were added to the reaction mixture. The resulting layers were then separated. The aqueous layer was extracted with EtOAc (2 x 20 mL). The combined organic layers were washed with H_2O (2 x 30 mL) and sat. aqueous NaCl (20 mL), dried over Na_2SO_4, and then filtered. The filtrate was then concentrated *in vacuo* to give a viscous, light yellow oil. The crude oil was dissolved in 0.5% HCO_2H-5% MeOH-MeCN (600 mL) and concentrated *in vacuo*. This process was repeated two more times. The resulting oil was then purified by flash chromatography (2% MeOH/98% DCM to 10% MeOH/90% DCM) on silica gel to afford compound **12** (0.376 g, 82%, 3 steps) as a viscous, white semi-solid. (R_f = 0.40; visualized on silica gel; 10% MeOH/90% CH_2Cl_2, UV lamp).

Crude **12** could also be dissolved in MeCN and purified by reverse-phase preparative HPLC (Thermo Hypersil Gold aQ C18, 250 x 20 mm, 5 μm) using a gradient separation at a flow rate of 8 mL/min [a linear gradient increasing from 5 to 95% solvent B in solvent A over 15 min, hold at 95% B for 6 min, then re-equilibration with 5% solvent B in solvent A over 5 min, hold at 5% solvent B in solvent A for 4 min (solvent A = H_2O; solvent B = acetonitrile)] (Total run time = 30 min) to afford pure enamide **12** as a white solid (retention time of **12** = 19 min). ^{1}H NMR (600 MHz DMSO-d_6): δ (ppm) = 8.90 (t, J = 5.9 Hz, 1H, H_{10}); 8.45 (br s, 1H, H_7); 8.41 (s, 1H, H_{12}); 6.54 (s, 1H, H_6); 4.55 (d, J =

5.9 Hz, 2H, H_{11}); 4.28 (q, J = 7.1 Hz, 2H, H_{13}); 4.17 (app p, J = 6.5 Hz, 1H, H_3); 3.53 (s, 2H, H_9); 3.41 (dd, J = 18.0, 8.6 Hz, 1H, H_5); 2.85 – 2.79 (m, 1H, H_5); 1.89 (ddd, J = 19.0, 11.3, 7.9 Hz, 1H, H_4); 1.57 – 1.53 (m, 1H, H_4); 1.47 (s, 9H, H_1); 1.37 – 1.35 (m, 2H, H_8); 1.29 (t, J = 7.1 Hz, 3H, H_{14}); 1.15 (d, J = 6.3 Hz, 3H, H_2); 1.02 – 1.00 (m, 2H, H_8). ^{13}C NMR (125 MHz, DMSO-d_6): δ (ppm) = 204.7, 170.4, 168.9, 167.0, 160.7, 153.2, 151.2, 145.5, 129.2, 98.4, 80.8, 60.7, 56.0, 46.4, 40.6, 40.5, 28.8, 27.82, 27.77, 19.48, 19.45, 19.3, 14.2. IR (ATR-FIR), cm^{-1}: 3299, 2978, 2932, 1714, 1654, 1606, 1513, 1213, 1158, 850. HRMS (ESI): calcd for $C_{25}H_{35}N_4O_7S$ $[M+H]^+$, 535.2221; found, 535.2275. $[\alpha]_D^{24}$ = +18.0° (c = 1.0, $CHCl_3$).

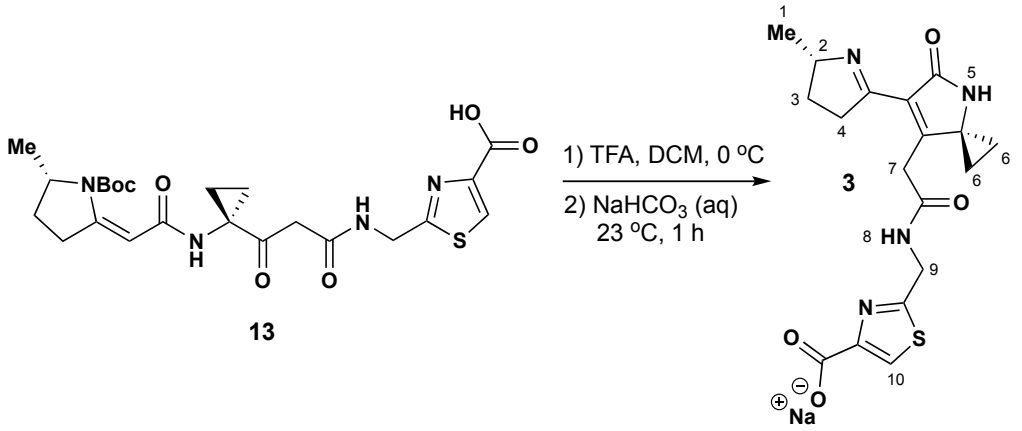
Preparation of enamide **13**



A solution of ester **12** (0.008 g, 0.0149 mmol, 1 equiv) in THF/MeOH/H_2O (3:2:1) (300 μL, 0.05 M) was treated with aqueous 1N LiOH (15.6 μL, 0.0156 mmol, 1.05 equiv), and the reaction mixture was stirred at room temperature for 1 h. After 1 h, an additional portion of aqueous 1N LiOH (25 μL, 0.025 mmol, 1.67 equiv) was added, and the solution stirred for an additional 30 min. Once the reaction was deemed complete by TLC, the reaction mixture was concentrated *in vacuo* to remove THF and MeOH. H_2O (5 mL) was added, and the aqueous layer was extracted with EtOAc (2 x 5 mL). Aqueous 1N HCl (50 μL, 3.35 equiv) was added to the aqueous layer. The acidified aqueous layer (pH ~ 4) was diluted with H_2O (20 mL) and then extracted with EtOAc (3 x 20 mL). The combined organic layers were dried over Na_2SO_4, filtered, and the resulting filtrate concentrated *in vacuo* to give (0.0067 g, 88%) as a viscous, colorless oil. Enamide **13** was used without further purification.

^{1}H NMR (600 MHz DMSO-d_6): δ (ppm) = 8.89 (t, J = 6.0 Hz, 1H, H_{10}); 8.46 (br s, 1H, H_7); 8.32 (s, 1H, H_{12}); 6.53 (s, 1H, H_6); 4.54 (d, J = 6.0 Hz, 2H, H_{11}); 4.17 (app p, J = 6.5 Hz, 1H, H_3); 3.53 (s, 2H, H_9); 3.40 (dd, J = 18.2, 8.8 Hz, 1H, H_5); 2.85 – 2.79 (m, 1H, H_5); 1.92 – 1.85 (m, 1H, H_4); 1.57 – 1.53 (m, 1H, H_4); 1.47 (s, 9H, H_1); 1.37 – 1.35 (m, 2H, H_8); 1.15 (d, J = 6.3 Hz, 3H, H_2); 1.02 – 0.99 (m, 2H, H_8). ^{13}C NMR (125 MHz, DMSO-d_6): δ (ppm) = 204.7, 170.0, 168.9, 167.0, 162.0, 153.3, 151.2, 146.6, 128.8, 98.4, 80.9, 56.0, 46.4, 40.6, 40.5, 28.83, 27.82, 27.78, 19.46, 19.43, 19.3. IR (ATR-FIR), cm^{-1}: 3276, 3090, 2976, 2931, 1706, 1650, 1606, 1382, 1290, 1242, 1159, 850. HRMS (ESI): calcd for $C_{23}H_{29}N_4O_7S$ $[M–H]^-$, 505.1762; found, 505.1766. $[\alpha]_D^{23}$ = +16.0° (c = 0.5, DMSO-d_6).

Preparation of acid-containing cyclopropane **3**

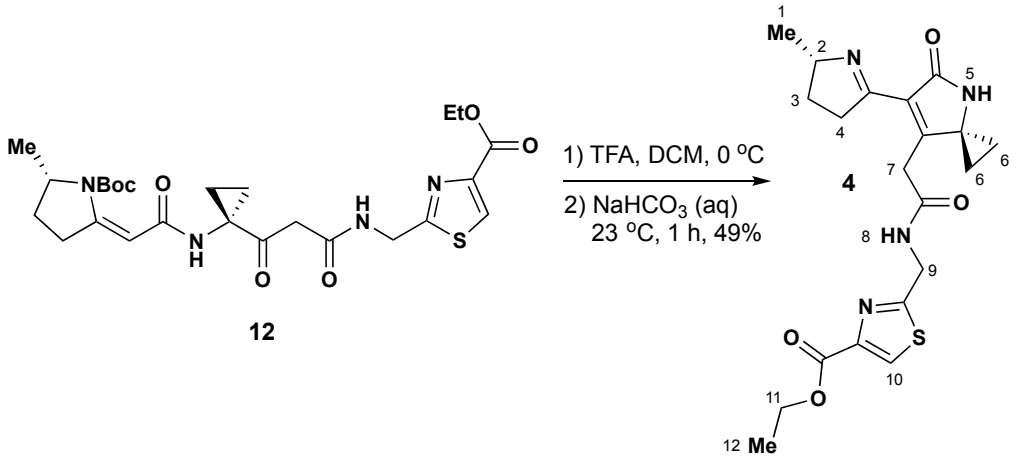


Trifluoroacetic acid (0.223 mL, 2.92 mmol, 133 equiv) was added dropwise to a cooled solution of **13** (0.0112 g, 0.022 mmol, 1 equiv) in dry CH_2Cl_2 (0.440 mL, 0.05 M) at 4 ºC. The resulting reaction mixture was incubated at 4 ºC for 12 h. After stirring overnight, dry PhMe (1 mL) was added to the reaction mixture. The resulting solution was then concentrated *in vacuo* to give a colorless, viscous oil. The crude TFA salt was taken to the next step without further purification.

Saturated aqueous $NaHCO_3$ (0.400 mL) was added to the crude TFA salt at 23 ºC. The resulting reaction mixture was vigorously stirred at 23 ºC for 1 h. After 1 h, the reaction mixture was concentrated *in vacuo* via lyophilization to give a white solid containing both excess $NaHCO_3$ salts and the desired product **3**. DMSO-d_6 (0.300 mL) was added to the white solid and the resulting white slurry vortexed for 1 min. The solution was centrifuged (13,000 rpm x 1 min at 23 ºC) and the resulting DMSO-d_6 supernatant transferred to a Shigemi tube matched to DMSO-d_6 for NMR analysis. Cyclopropane **3** was used without further purification. ^{1}H NMR (600 MHz DMSO-d_6): δ (ppm) = 10.17 (t, J = 5.8 Hz, 1H, H_8); 8.62 (br s, 1H, H_5); 7.70 (s, 1H, H_{10}); 4.52 (app t, J = 6.1 Hz, 2H, H_9); 4.12 – 4.09 (m, 1H, H_2); 3.30 (d, J = 5.9 Hz, 2H, H_7); 3.11 – 3.09 (m, 1H, H_4); 2.86 (dd, J = 18.0, 9.5 Hz, 1H, H_4); 2.11 – 2.05 (m, 1H, H_3); 1.68 – 1.66 (m, 2H, H_6); 1.38 – 1.35 (m, 3H, H_3 and H_6); 1.17 (d, J = 6.7 Hz, 3H, H_1). ^{13}C NMR (125 MHz, DMSO-d_6): δ (ppm) = *, 169.6, 168.3, 168.1, 167.8, 157.6, 157.2, 127.6, 127.3, 66.2, 45.2, 40.2, 36.2, 33.3, 29.4, 21.6, 11.47, 11.46. HRMS (ESI): calcd for $C_{18}H_{20}N_4O_4S$ $[M+H]^+$, 389.1278; found, 389.1298.

* = carboxylate carbon could not be observed via long-range HMBC correlations from H10; this phenomenon was also observed in the HMBC spectrum of cyclopropane **4** (see Fig S16). However, high-resolution LC-MS^2 analyses along with 1D- and 2D-NMR experiments support the structural assignment of cyclopropane **3**.

Preparation of ester-containing cyclopropane **4**



Trifluoroacetic acid (0.412 mL, 5.39 mmol, 120 equiv) was added dropwise to a cooled solution of **12** (0.024 g, 0.045 mmol, 1 equiv) in dry CH_2Cl_2 (1.12 mL, 0.04 M) at 4 ºC. The resulting reaction mixture was incubated at 4 ºC for 12 h. After incubating at 4 ºC overnight, dry PhMe (1.50 mL) was added to the reaction mixture. The resulting solution was then concentrated *in vacuo* to give a colorless, viscous oil. The crude TFA salt was taken to the next step without further purification.

Saturated aqueous $NaHCO_3$ (1.00 mL) was added to the crude TFA salt at 23 ºC. The resulting reaction mixture was vigorously stirred at 23 ºC for 1 h. Once the reaction was complete, EtOAc (5 mL) and H_2O (2 mL) were added to the reaction mixture. The layers were separated and the aqueous layer extracted with EtOAc (4 x 5 mL). The combined organic layers were dried over Na_2SO_4, filtered, and the resulting filtrate concentrated *in vacuo* to give a light yellow semi-solid (0.013 g, 49%). Cyclopropane **4** was used without further purification. ^{1}H NMR (600 MHz DMSO-d_6): δ (ppm) = 10.24 (t, J = 5.9 Hz, 1H, H_8); 8.58 (br s, 1H, H_5); 8.42 (s, 1H, H_{10}); 4.56 (dd, J = 5.9, 4.3 Hz, 2H, H_9); 4.29 (q, J = 7.1 Hz, 2H, H_{11}); 4.14 – 4.10 (m, 1H, H_2); 3.31 (s, 2H, H_7); 3.08 (dddt, J = 16.7, 9.7, 4.6, 2.2 Hz, 1H, H_4); 2.86 (dddd, J = 17.9, 9.8, 7.8, 2.0 Hz, 1H, H_4); 2.12 – 2.05 (m, 1H, H_3); 1.68 – 1.66 (m, 2H, H_6); 1.40 – 1.32 (m, 3H, H_3 and H_6); 1.29 (t, J = 7.1 Hz, 3H, H_{12}); 1.18 (d, J = 5.9 Hz, 3H, H_1). ^{13}C NMR (125 MHz, DMSO-d_6): δ (ppm) = 170.4, 169.6, 168.4, 168.3, 160.7, 158.0, 157.8, 145.7, 129.3, 66.5, 60.7, 45.3, 40.5, 36.7, 33.5, 29.7, 21.9, 14.2, 11.9, 11.8. IR (ATR-FIR), cm^{-1}: 3392, 3149, 2986, 2872, 1716, 1575, 1399, 1059, 840. HRMS (ESI): calcd for $C_{20}H_{25}N_4O_4S$ $[M+H]^+$, 417.1591; found, 417.1599. $[\alpha]_D^{23} = -5.2^o$ (c = 1.0, DMSO-d_6).

Tandem MS^2 fragmentation of synthetic cyclopropanes. High-resolution LC-MS^2 analysis was performed on an Agilent 6530 Accurate-Mass time-of-flight mass spectrometer system equipped with an electrospray ionization source. Liquid chromatography was performed on an Agilent 1260 Infinity HPLC using a Cogent Diamond Hydride column (150 x 3.0 mm, 100 Å, 4 μm particle size; 5 μL injection volume) at a flow rate of 0.500 mL/min and the following elution conditions: gradient separation [(linear gradient of 90-30% solvent B in solvent A over 6 min, hold at 30%

solvent B in solvent A for 2 min, then re-equilibration with 90% solvent B in solvent A for 3 min, and then hold at 90% solvent B in solvent A for 4 min (solvent A = H_2O + 0.1% formic acid; solvent B = acetonitrile + 0.1% formic acid) (Total run time = 15 min)]. Elution before 7 min and after 10 min was diverted to waste. The experiment was performed in positive ionization mode with a mass range of *m/z* 100 to 1700, a scan speed of 1 scan/second, and a product ion collision energy of 15 eV. The capillary voltage was set to 3500 V. The source parameters were set with a gas temperature of 275 ºC, a flow rate of 8 L/min, and nebulizer at 35 psig. LC-MS data were acquired with MassHunter Workstation Data Acquisition (Agilent Technologies) and analyzed using MassHunter Qualitative Analysis software (Agilent Technologies). The precursor product ion for the MS^2 fragmentation of cyclopropane **3** was *m/z* 389.1278 and *m/z* 417.1591 for the tandem fragmentation of cyclopropane **4**. Tandem MS data is shown in Figure S29.

DNA shearing assay. The 4.4 kb pTrcHis2A plasmid was isolated from *E. coli* TOP10 harboring pTrcHis2A (Invitrogen) using Plasmid Midiprep Kit (Qiagen) and digested with BamHI-HF (New England Biolabs) overnight. The linearized plasmid was purified using the DNA Clean & Concentrator kit (Zymo Research). Various concentrations of synthetic compounds **3** and **4** dissolved in DMSO, 100 μM cisplatin (Millipore) dissolved in water, or 500 μM methyl methanesulfonate dissolved in water were added to 140 ng of purified linearized DNA and then diluted to a total volume of 10 μL with H_2O and incubated at 37 ºC. After 16 h, the samples were treated with denaturation buffer containing 6% sucrose and 1% sodium hydroxide for 15 min at room temperature and separated on a 0.9% 1X Tris/Borate/EDTA (TBE) agarose gel prestained with SYBR Safe DNA gel stain (Invitrogen).

Cell cycle analysis. One or two days post bacterial infection (*pks*$^+$ *E. coli* JW3973-1) or synthetic compound treatment, HeLa cells were collected, centrifuged (400 g x 2 min at 4 ºC), washed with 1X DPBS, and resuspended in 70% ice-old ethanol for fixation at 4 °C overnight. For DNA content analysis, cells were centrifuged (800 g x 5 min at 4 ºC), washed with 1X DPBS, and stained with 1X DPBS supplemented with 20 μg/mL propidium iodide and 200 μg/mL RNase A (Thermo Fisher Scientific, Waltham, MA) at room temperature for at least 30 min. Cell cycle was monitored on the BD LSR II flow cytometer (BD Biosciences, Franklin Lakes, NJ) with 10,000 events/determination and analyzed with Flowjo software (Tree Star Inc., Ashland, OR). All cell cycle experiments were replicated three times independently and cell cycle profiles from one representative experiment was shown.

MTT assay. HeLa cells were seeded into 96-well plates at a density of 1×10^5 cells per well in 100 μL growth medium (DMEM medium supplemented with 10% FBS and 1X Antibiotic-Antimycotic) and incubated at 37 ºC in 5% CO_2 incubator for 1 day. Wells containing growth medium only were used as background controls. Cells were treated with various concentrations of synthetic compounds **3** and **4** in quadruplicate. One day or two days post treatment, 20 μL of CellTiter 96® AQ_{ueous} One Solution Reagent (Promega, Madison, WI) was added to each well. The plates were incubated at 37 ºC in 5% CO_2 incubator for 2 h followed by absorbance measurement at 490 nm using a Synergy HTX Multi-Mode Microplate Reader (BioTek, Winooski, VT). To calculate relative cell

viability, the readings for each compound concentration were subtracted from the background controls and normalized to DMSO controls. All MTT assays were replicated three times independently and relative cell viability from one representative experiment is shown.

Immunocytochemical staining. HeLa cells were grown on glass coverslips in 12-well plates. After bacterial infection (*pks*$^+$ *E. coli* JW3973-1) or compound treatment, cells were fixed with ice cold 95% methanol and 5% acetic acid, permeabilized with 0.3% Triton X-100 containing DPBS, and blocked with 0.3% normal goat serum and 0.3% Triton X-100 containing DPBS. Cells were stained by adding the γH2AX primary antibody (1:1000 ratio, Novus Biologicals, Littleton, CO) and then incubated overnight at 4 °C followed by Cy3-conjugated goat anti-rabbit secondary antibody (1:200 ratio, Jackson ImmunoResearch, West Grove, PA) staining for 1 h at room temperature. After DAPI (4',6-diamidino-2-phenylindole) counterstaining of cell nuclei, glass coverslips were mounted on glass slides using ProLong Gold antifade reagents (Thermo Fisher Scientific, Waltham, MA). Images were acquired under the same imaging protocol using a Zeiss LSM700 laser scanning confocal microscope.

Analytical *in vitro* DNA alkylation of cyclopropanes with calf-thymus DNA. Cyclopropane **3** (30 μL, ~10 mM stock in DMSO, ~1 mM final concentration), cyclopropane **4** (15 μL, 10 mM stock in DMSO, 500 μM final concentration), or DMSO (30 μL) was mixed with a homogenous calf thymus DNA (0.002 g, 2792 ng/uL, 698 μg) solution in H_2O (250 μL) and then diluted to a final volume of 300 μL by adding H_2O (50 μL). The resulting reaction mixture was incubated for 19.5 h at 37 ºC. After 19.5 h, the DNA solutions were heated to 100 ºC for 45 min, cooled to room temperature, flash frozen, and then concentrated *in vacuo* via lyophilization. The resulting solids were then extracted with 1-butanol (3 x 500 μL per sample). The samples were centrifuged (13,000 rpm x 5 min) to pellet insoluble DNA particulates, and the 1-butanol layers carefully removed from the microcentrifuge tubes. The combined 1-butanol extracts were concentrated *in vacuo* using a Genevac (High BP, Initial time: 0.5 h; Final stage time: 0.5 h) to yield a white (for cyclopropane **3**) or yellow (for cyclopropane **4**) solid. The resulting solids were dissolved in 100 μL of a 75% MeCN/25% H_2O solution and then diluted 1 to 10 to a final volume of 100 μL. The solutions were then analyzed by high-resolution LC-MS analysis.

High-resolution LC-MS analysis was performed on an Agilent 6530 Accurate-Mass time-of-flight mass spectrometer system equipped with an electrospray ionization source. Liquid chromatography was performed on an Agilent 1260 Infinity HPLC using a Cogent Diamond Hydride column (150 x 3.0 mm, 100 Å, 4 μm particle size; 5 μL injection volume) at a flow rate of 0.500 mL/min and the following elution conditions: gradient separation [(linear gradient of 90-30% solvent B in solvent A over 6 min, hold at 30% solvent B in solvent A for 2 min, then re-equilibration with 90% solvent B in solvent A for 3 min, and then hold at 90% solvent B in solvent A for 4 min (solvent A = H_2O + 0.1% formic acid; solvent B = acetonitrile + 0.1% formic acid) (Total run time = 15 min)]. Elution before 3.5 min and after 12 min was diverted to waste. The experiment was performed in positive ionization mode with a mass range of *m/z* 100 to 1700 and a

scan speed of 1 scan/second. The capillary voltage was set to 3500 V. The source parameters were set with a gas temperature of 275 ºC, a flow rate of 8 L/min, and nebulizer at 35 psig. LC-MS data was acquired with MassHunter Workstation Data Acquisition (Agilent Technologies) and analyzed using MassHunter Qualitative Analysis software (Agilent Technologies). A mass window of ± 10 ppm was used to extract the $[M+H]^+$ ion (*m/z* 552.2136) for detection of **5**, (*m/z* 524.1823) $[M+H]^+$ for the detection of **5'**, (*m/z* 568.2085) $[M+H]^+$ for the detection of **6** and **7**, and (*m/z* 540.1772) $[M+H]^+$ for the detection of **1** and **2**.

Preparative *in vitro* DNA alkylation of cyclopropane 4 with calf thymus DNA. A homogenous calf thymus DNA solution (DNA concentration: 2,638 μg/mL) was prepared by incubating calf thymus DNA fibers (0.032 g) with H_2O (10 mL) at 4 ºC on a nutating rocker overnight. Cyclopropane **4** (50 μL, 100 mM stock in DMSO, 500 μM final concentration) was thoroughly mixed with the calf thymus DNA solution in a 50 mL falcon tube and the resulting solution incubated at 37 ºC for 19.5 h. After 19.5 h, 2M $NaCl_{(aq)}$ (1.50 mL) was added to the reaction mixture followed by 200 proof EtOH (27 mL). The resulting solution was vortexed for 30 s to fully precipitate the DNA and then centrifuged (10,800 rpm x 15 min, 4 ºC) to pellet the DNA. The EtOH supernatant was carefully removed and the DNA pellet allowed to air dry for 10 min. The DNA pellet was then dissolved in H_2O (10 mL) by vigorously vortexing the solution in addition to pipetting the solution up and down with a 25 mL serological pipet.

The above protocol was repeated 10 times (reaction of 10 mL calf thymus DNA with cyclopropane **4** x 10) giving 100 mL of reacted calf thymus DNA solution.

To a 250 mL round bottom flask containing a stir bar was added the combined reacted calf thymus DNA solutions (100 mL). 1-butanol (50 mL) was then added to the calf thymus DNA solution and the resulting mixture heated to 100 ºC for 30 min while stirring slowly (stir setting: 4). After 30 min, the reaction mixture was cooled to room temperature (~15 min) and the 1-butanol layer separated by centrifugation (4,000 rpm x 5 min, 23 ºC). Fresh 1-butanol (50 mL) was added again to the aqueous layer, and the resulting solution heated to 100 ºC for 30 min. The reaction mixture was cooled to room temperature and the 1-butanol layer separated by centrifugation (4,000 rpm x 5 min, 23 ºC). The combined 1-butanol layers were concentrated *in vacuo* using a Genevac (High BP setting: 1 h to final stage; 1 h at final stage for a total of 2 h) giving approximately 0.0035 g of crude DNA adduct material.

Conversion of "non-oxidized" adduct 5 to "oxidized" DNA adducts 6 and 7. The crude DNA adduct material containing the "non-oxidized" DNA adduct **5** (~0.0031 g) was dissolved in 3 mL of a 90% MeCN/10% H_2O solution and incubated for 2 days at 23 ºC. After 2 days, LC-MS and HPLC analysis confirmed that the "non-oxidized" DNA adduct **5** converted completely to the corresponding "oxidized" adducts **6** and **7**. The solution was then concentrated *in vacuo* using a Genevac (High BP, Initial time: 1 h; Final stage time: 1 h) to give a yellow/orange solid (~0.003 g).

HPLC analysis of "non-oxidized" to "oxidized" adduct conversion. At the specified time points (t= 0, 24, and 48 h), a 25 μL aliquot was removed from the DNA adduct containing solution mentioned above and diluted 1 to 2 to a final volume of 50 μL in LC-MS grade MeCN. The solution was then analyzed by analytical HPLC. Analytical HPLC analysis was performed on a Dionex UltiMate 3000 instrument (Thermo Scientific) using a Cogent Diamond Hydride column (150 x 3.0 mm, 100 Å, 4 μm particle size) and the following elution conditions (flow rate = 0.500 mL/min): gradient separation [(90% solvent B in solvent A for 2 min, linear gradient of 90-20% solvent B in solvent A over 6 min, hold at 20% solvent B in solvent A for 3 min, then re-equilibration with 90% solvent B in solvent A over 5 min, hold at 90% solvent B in solvent A for 4 min (Total run time = 21 min) (solvent A = H_2O + 0.1% formic acid; solvent B = acetonitrile + 0.1% formic acid)].

Purification of the *in vitro*-derived DNA adducts 6 and 7. DNA contaminants were removed by Sephadex LH-20 column chromatography (2 x 15 cm, 100% EtOH isocratic elution) to give approximately 0.003 g of a bright yellow solid. (R_f = 0.20; visualized on alumina; 10% MeOH/90% CH_2Cl_2, UV lamp).

The resulting material was subjected to solid phase purification using a Waters SepPak C_{18} cartridge (50 mg) and eluted using a MeCN/H_2O step gradient. The elution steps were 100% H_2O (2 x 1 mL), 25% MeCN/H_2O (1 x 1 mL), 50% MeCN/H_2O (1 x 1 mL), 75% MeCN/H_2O (1 x 1 mL), 100% MeCN (2 x 1 mL). Based on analytical HPLC analysis, the majority of the purified adenine adducts eluted in the 25% MeCN/H_2O fraction. The product containing fraction was concentrated *in vacuo* to afford approximately 0.001 g of the desired adenine adducts **6** and **7**.

Structural characterization of DNA adducts 6 and 7. ^{1}H, gCOSY, gHSQC, gHMBC, and ROESYAD data were obtained on a Bruker AVANCE II 600 MHz spectrometer equipped with a TXO cryoprobe. NMR spectra were obtained in 250 μL DMSO-d_6 (99.9 atom % D) using a symmetrical NMR tube susceptibility matched to DMSO (Shigemi, Inc). ^{1}H and ^{13}C NMR (DMSO-d_6, 600 MHz) (see table S3). IR (film), cm^{-1}: 3293, 2951, 2929, 2867, 1719, 1660, 1560, 1410, 1216, 1098, 761. HRMS (ESI): calcd for $C_{23}H_{26}N_9O_5S^+$ $[M+H]^+$, 540.1772; found, 540.1773. λ_{max} (nm) = 210, 280, 360; $[\alpha]_D^{23}$ = –32.0° (*c* = 0.1, EtOH).

Gaussian NMR calculations for DP4 computational analysis. For each of the eight stereochemical configurations, a truncated product with the ethyl ester replaced by the acid and the free adenine amine replaced by a hydrogen atom was drawn by hand in GaussView version 6.0.16. For each of the configurations three or four rotamers were drawn, and in each case at least one pair of rotamers was included where the carbonyl at C11 was rotated 180º with respect to its bond with C10 to ensure that the initial structures were not biased around any specific orientation of the atoms near this key position. These structures were geometry optimized to reach an energy minimum in Gaussian '09 on the Harvard Odyssey Cluster using the PDDG method with DMSO modeled by SCRF with a polarizable continuum model. To the output files, the acid was then extended to the full ethyl ester and the amine in adenine was installed by hand in GaussView. These

structures were then energy minimized again in Gaussian using the B3LYP method with the 6-31+G(d) basis set, with the same solvent modeling as before. Finally, the output files of these serial energy minimizations were resubmitted to Gaussian for NMR calculation using the mPW1PW91 method and the 6-311+G(2d,p) basis set, again with the same solvent modeling (The B3LYP method with the 6-311+G(2d,p) basis set was also used to calculate NMR shifts, but the resulting ^{13}C values were consistently worse than the mPW1PW91 method and are not discussed). From the output of this operation, ^{13}C isotropic values were extracted for each carbon position and corrected according to literature scaling values available on CHESHIRE (CHEmical SHIft REpository, cheshirenmr.info (*41*)). DP4 calculations were made with the DP4 applet available on the Goodman group website (http://www-jmg.ch.cam.ac.uk/tools/nmr/DP4/ using the second DP4 database with a *t*-distribution (*30*).

Ethyl ester hydrolysis using pig liver esterase. To a solution of porcine liver esterase (1 mg, ~15 units of enzyme) in 100 mM sodium phosphate buffer pH 7.40 (400 μL) was added esters **6** and **7** (~300 μg) as a solution in EtOH (40 μL). The resulting reaction mixture was incubated at 37 ºC with 220 rpm shaking for 19 h. After 19 h, acetone (200 μL) was added to the enzyme reaction and the solution incubated on ice for 15 min to precipitate the protein. The heterogenous solution was then centrifuged (13,000 rpm x 10 min at 4 ºC) and 50 μL of the resulting supernatant transferred to a clean LC-MS vial. The solution was then analyzed by high-resolution LC-MS analysis.

LC-MS analysis was performed on an Agilent 6530 Accurate-Mass time-of-flight mass spectrometer system equipped with an electrospray ionization source. Liquid chromatography was performed on an Agilent 1260 Infinity HPLC using a Thermo Scientific Hypersil Gold aQ analytical column (150 x 3.0 mm, 175 Å, 3 μm particle size; injection volume = 5 μL) at a flow rate of 0.350 mL/min and the following elution conditions: gradient separation [(linear gradient of 0-100% solvent B in solvent A over 10 min, hold at 100% solvent B in solvent A for 3 min, then re-equilibration with 0% solvent B in solvent A for 2 min, hold at 0% solvent B in solvent A for another 2 min (Total run time = 17 min) (solvent A = H_2O + 0.1% formic acid; solvent B = acetonitrile + 0.1% formic acid)]. Elution before 1 min and after 17 min was diverted to waste. The experiment was performed in positive ionization mode with a mass range of *m/z* 100 to 1200 and a scan speed of 1 scan/sec. The capillary voltage was set to 3500 V. The source parameters were set with a gas temperature of 275 ºC, a flow rate of 8 L/min, and nebulizer at 35 psig. LC-MS data was acquired with MassHunter Workstation Data Acquisition (Agilent Technologies) and analyzed using MassHunter Qualitative Analysis software (Agilent Technologies). A mass window of ± 10 ppm was used to extract the $[M+H]^+$ ion (*m/z* 540.1772) for detection of **1** and **2** and (*m/z* 568.2085) $[M+H]^+$ for the detection of **6** and **7**.

Alkaline CometChip. The level of DNA migration was measured using the alkaline CometChip platform as described previously (*33*). Briefly, a solution of molten 1% w/v UltraPure™ agarose was prepared in PBS. A polydimethylsiloxane (PDMS) stamp with micropegs approximately the size of single cells was fabricated using the Sylgar™ 184 silicone elastomer kit. To create microwells, the PDMS stamp was pressed into 1% w/v

molten agarose. The agarose gelated after 15 min at room temperature, and the stamp was lifted to reveal an array of microwells that were ~240 μm spaced apart. Each microwell was ~40 μm in both diameter and depth. A bottomless 96-well plate was pressed on top of the microwell array to create 96 macrowells, each of which contained up to 300 microwells at its base.

Cells were loaded into the microwells via gravity. Specifically, ~2,000 – 200,000 cells in suspension were placed into each macrowell, and the chip was incubated at 37 ºC in the presence of 5% CO_2 for 15 min. After loading, the excess cells were washed off with PBS using sheer force. A layer of 1% w/v Ultrapure™ LMP agarose was placed on top of the chip to encapsulate the cells in microwells. The chip was then incubated for 2 min at room temperature followed by 2 min at 4 ºC to ensure complete gelation.

To measure DNA migration, cells embedded in CometChip were lysed in alkaline lysis buffer (2.5 M NaCl, 100 mM Na_2EDTA, 10 mM Trizma® base, and 1% v/v Triton X-100, pH ~10) at 4 ºC overnight. The DNA was then submerged in alkaline unwinding buffer (0.3 M NaOH and 1 mM Na_2EDTA, pH ~ 13.5) for 40 min at 4 ºC and subsequently electrophoresed for 30 min at 1 V/cm and ~300 mA in the same buffer condition. Afterward, the DNA was stained for 15 min with 1X of SYBR™ Gold. Fluorescent images of the comets were captured at 40X magnification using an epifluorescence microscope (Nikon Eclipse 80i, Nikon Instruments, Inc., Melville, NY) with a 480 nm excitation filter. The images were analyzed using Guicometanalyzer, a custom software developed in MATLAB (The MathWorks Inc., Natick, MA) as previously described (*33*). The median % DNA in tail was obtained from an average of approximately 100 comets per well on a 96-well plate. The median values from triplicate wells were averaged to obtain the final data point for each experiment.

Gamma irradiation. γ-rays from a ^{137}Cesium source at ~1 Gy/min (Gammacell 40 Exactor, Best Theratronics Model C-440) were used to induce DNA strand breaks.

Cisplatin-induced ICLs. TK6, a human B-lymphoblastoid cell line, was cultured in 1X RPMI 1640 medium with GlutaMAX™ supplemented with 10% FBS and 100 U/ml Pen-Strep. TK6 cells were incubated with cisplatin for 1.5 h at 37 ºC in the presence of 5% CO_2 and then loaded directly into CometChip microwells. Residual cisplatin was removed by rinsing the chip three times in PBS. The chip was then placed in fresh culture medium and incubated at 37 ºC in the presence of 5% CO_2. To monitor the level of ICLs over a period of time, a chip was removed from the culture at different time points and exposed to 6 Gy of γ-rays. The cells were immediately lysed with CometChip alkaline lysis buffer, and the level of DNA migration was analyzed as described in the alkaline CometChip procedure.

Colibactin-induced ICLs. HeLa cells were infected with *pks*$^+$ *E. coli* BW25113 for 1 h and washed as described in the HeLa cell infection procedure. The cells were then trypsinized and loaded into CometChip microwells. Residual bacteria were removed by rinsing the chip three times in PBS. The chip was then placed in fresh culture medium and incubated at 37 ºC in the presence of 5% CO_2. To monitor the level of ICLs over a

period of time, a chip was removed from the culture at different time points and exposed to 8 Gy of γ-rays. The cells were immediately lysed with CometChip alkaline lysis buffer, and the level of DNA migration was analyzed as described in the alkaline CometChip procedure.

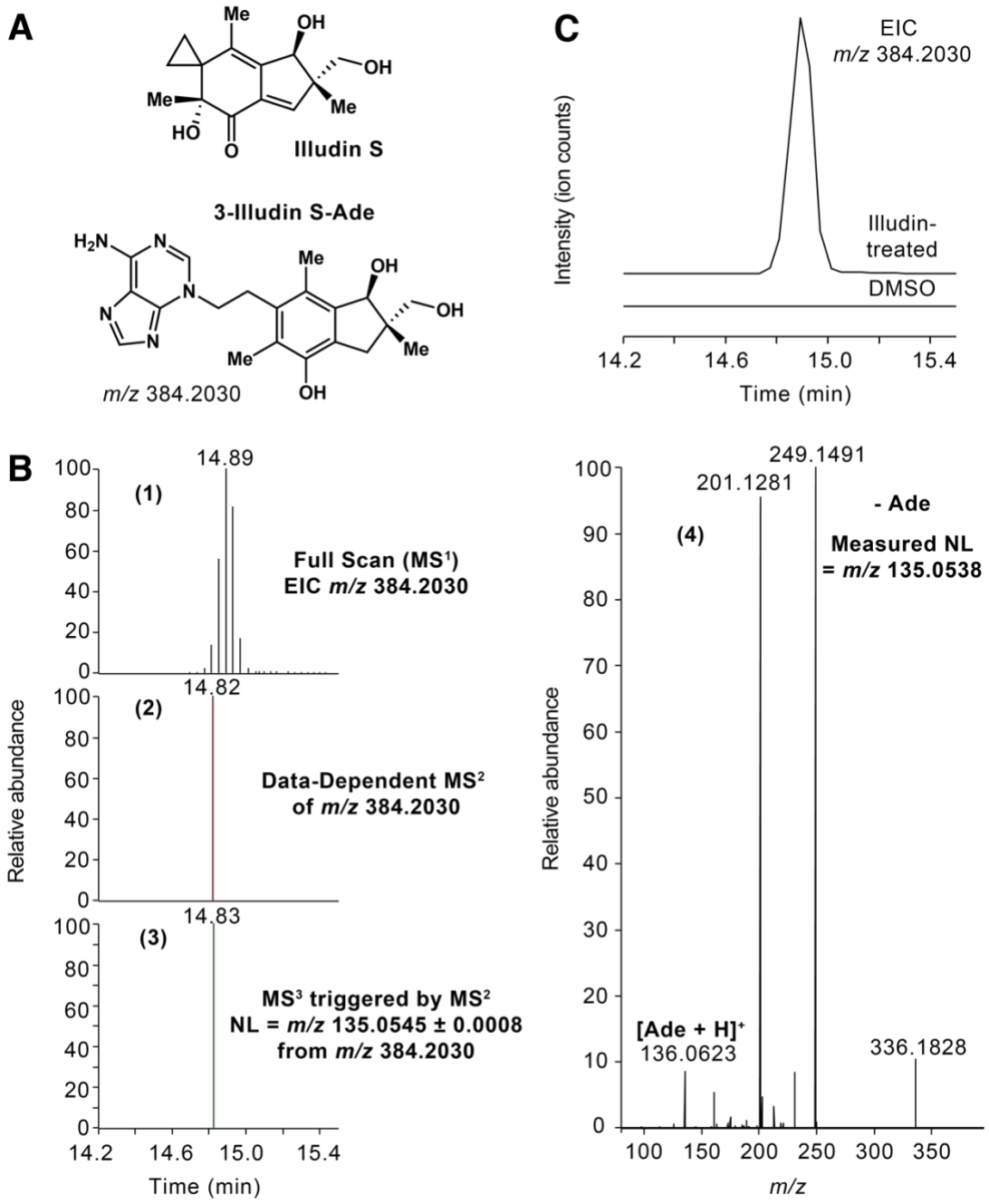


Fig. S1. **High-resolution accurate mass (HRAM) LC-MS^3 DNA adductomic analysis identified illudin S-DNA adducts in illudin S-treated HeLa cells**. (A) Chemical structures of illudin S and a known illudin S adenine adduct. (B) 1. Full scan extracted ion chromatogram (EIC) of the illudin S-DNA adduct (*m/z* 384.2030). 2. Signal corresponding to the data-dependent MS^2 event (RT = 14.82 min) (RT = retention time). 3. Signal corresponding to MS^3 event (RT = 14.83 min) triggered by the neutral loss of adenine. 4. MS^2 mass spectrum resulting from fragmentation of *m/z* 384.2030 which triggered the MS^3 event. (C) EIC chromatogram of DNA adduct in HeLa cells treated with illudin S (100 nM) and negative control sample (DMSO).

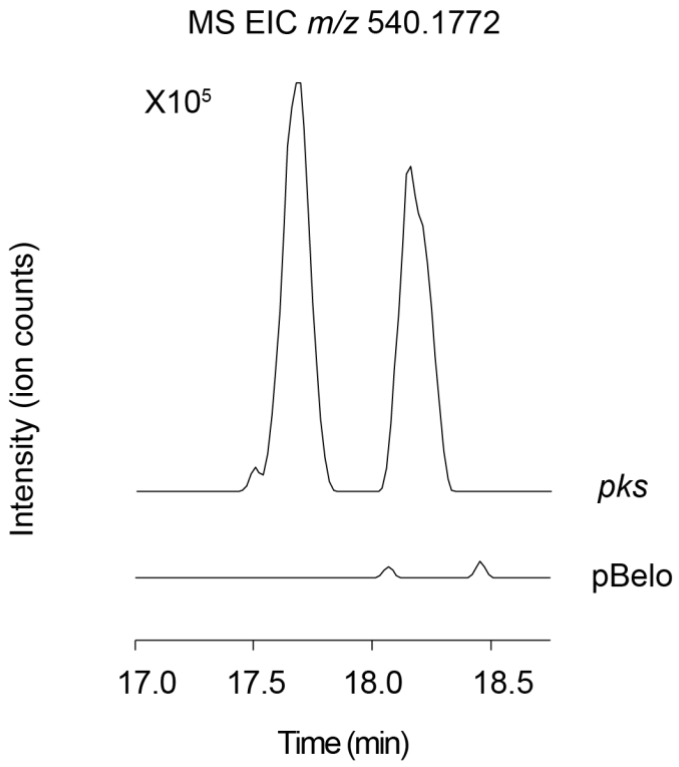


Fig. S2.
Identification of DNA adducts 1 and 2 from colonic epithelial cells exposed to *pks*$^+$ *E. coli*. Extracted ion chromatograms for DNA adducts **1** and **2** (*m/z* 540.1772) in a colonic epithelial cell line HT29 exposed to *pks*$^+$ *E. coli* and HT29 cells exposed to non-colibactin producing pBeloBAC *E. coli*. A mass window of ± 5 ppm was used to extract ions of interest.

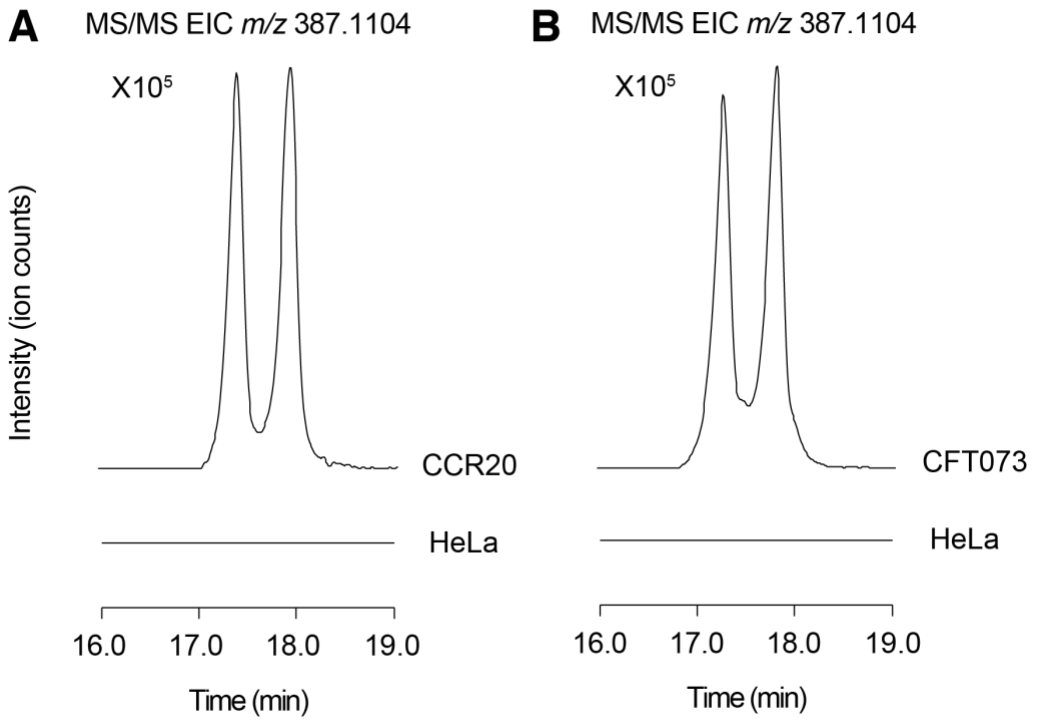


Fig. S3.
Identification of DNA adducts 1 and 2 from HeLa cells exposed to native *pks*$^+$ *E. coli* strains. Extracted ion chromatograms of the most abundant MS^2 fragmentation ion (*m/z* 387.1104) of the adducts **1** and **2** precursor ion (*m/z* 540.1772) in HeLa cells exposed to native colibactin-producing *E. coli* CCR20 (A) and CFT073 (B) and HeLa cells alone. A mass window of ± 5 ppm was used to extract ions of interest.

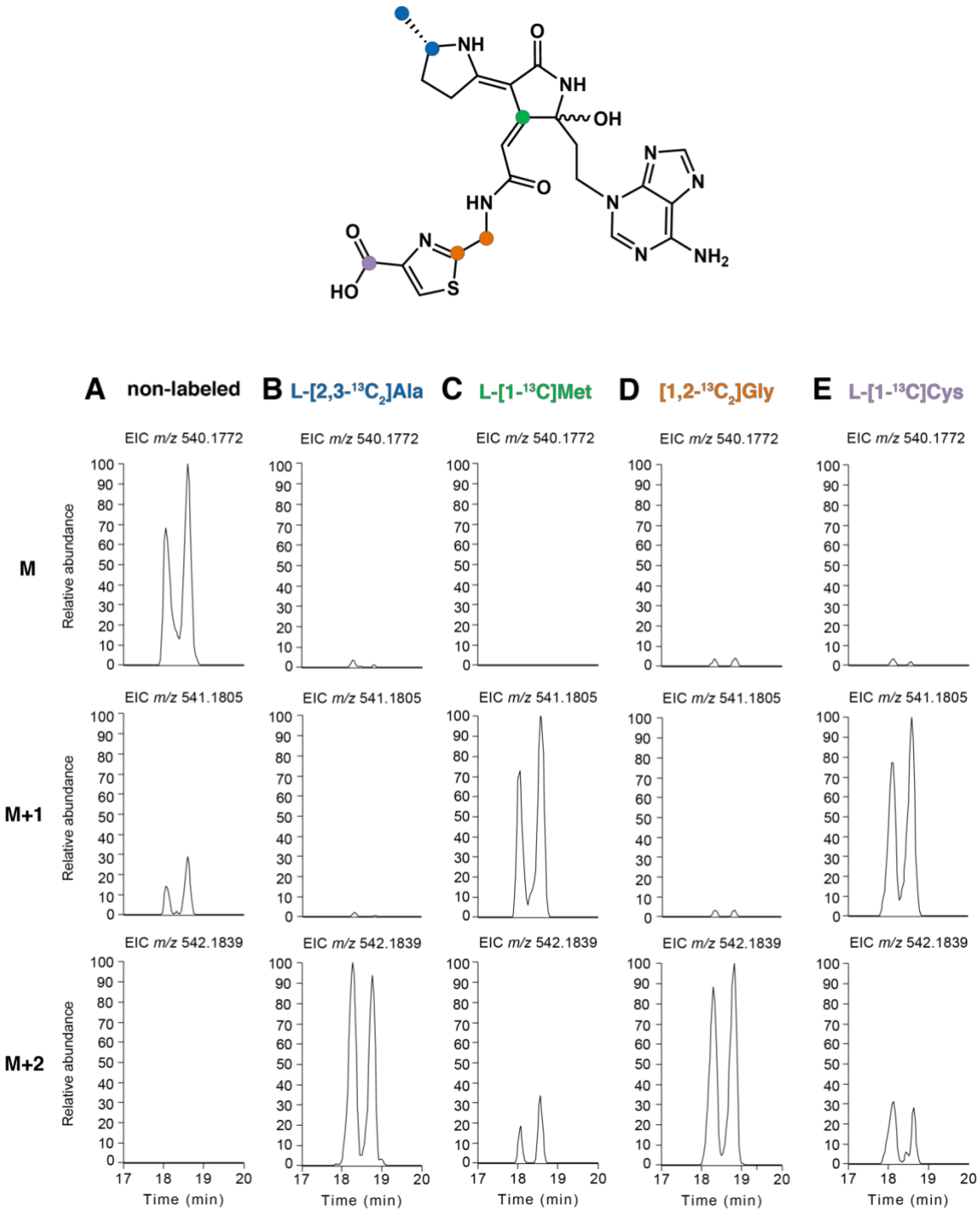
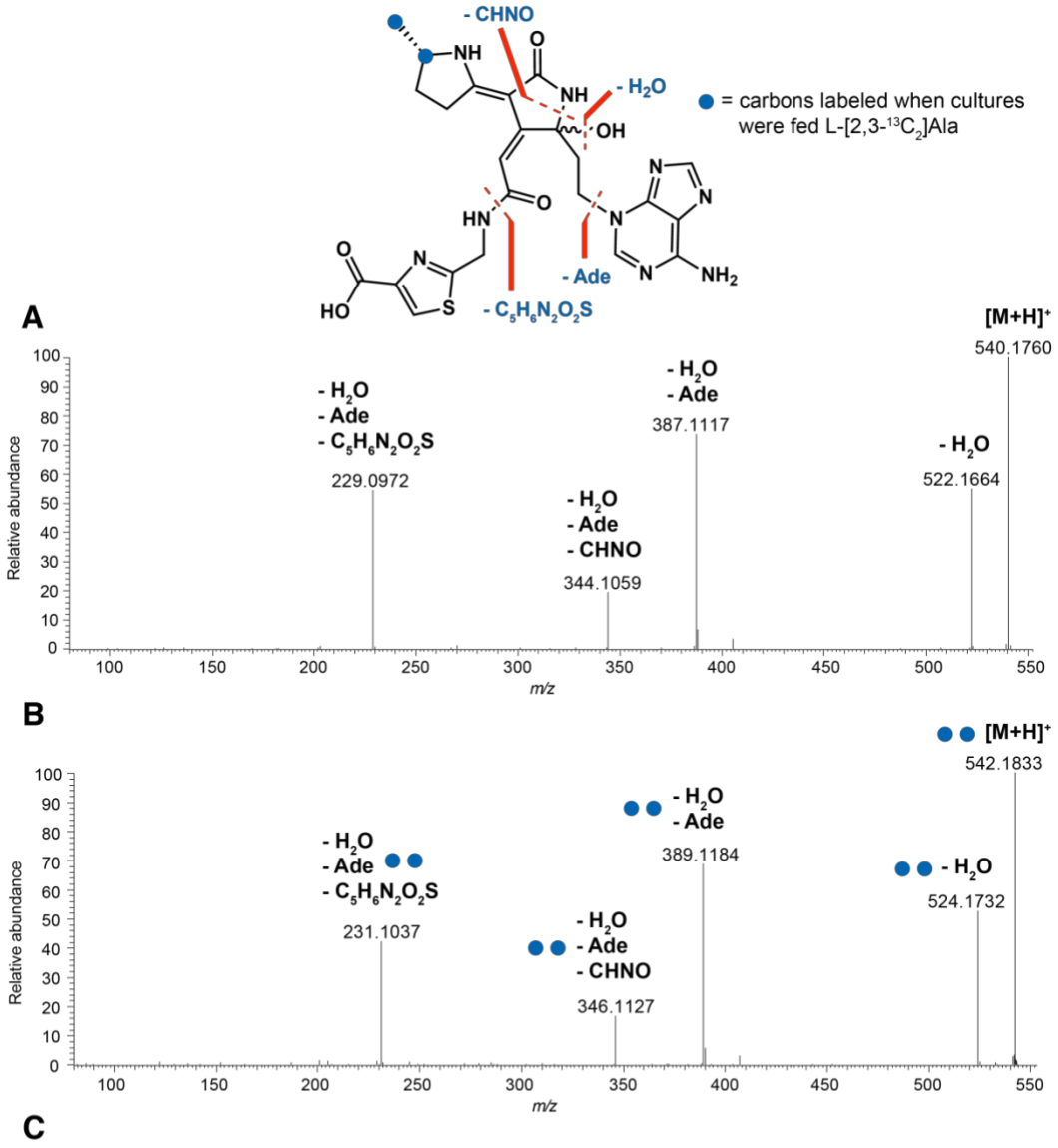


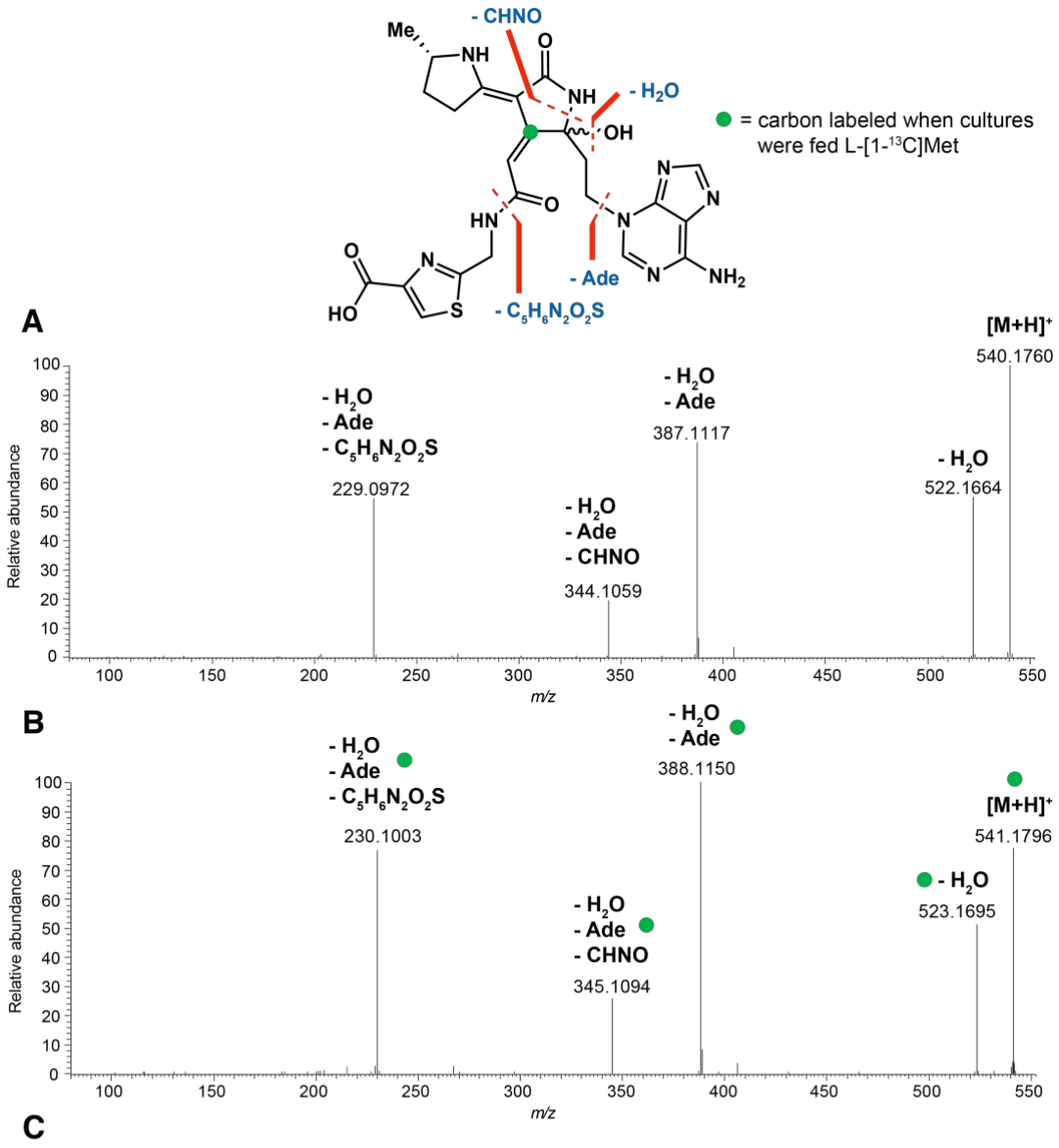
Fig. S4.

Isotopic labeling of DNA adducts 1 and 2 in *pks*$^{+}$ *E. coli*-treated HeLa cells. Amino acids incorporated into DNA adducts **1** and **2** as determined by isotopically labeled feeding experiments. Extracted ion chromatograms are shown for *m/z* 540.1772, 541.1805, and 542.1839 for (A) non-labeled, (B) L-[2,3-$^{13}C_2$]Ala, (C) L-[1-^{13}C]Met, (D) [1,2-$^{13}C_2$]Gly, and (E) L-[1-^{13}C]Cys.



Non-labeled feeding				L-[2,3-$^{13}C_2$]Ala Feeding			
Species	Theoretical *m/z*	Detected *m/z*	Error (Δppm)	Species	Theoretical *m/z*	Detected *m/z*	Error (Δppm)
$C_{23}H_{26}N_9O_5S^+$	**540.1772**	**540.1760**	**2.22**	**$^{13}C_2{}^{12}C_{21}H_{26}N_9O_5S^+$**	**542.1839**	**542.1833**	**1.10**
$C_{23}H_{24}N_9O_4S^+$	522.1666	522.1664	0.38	$^{13}C_2{}^{12}C_{21}H_{24}N_9O_4S^+$	524.1733	524.1732	0.19
$C_{18}H_{19}N_4O_4S^+$	387.1122	387.1117	1.29	$^{13}C_2{}^{12}C_{16}H_{19}N_4O_4S^+$	389.1189	389.1184	1.28
$C_{17}H_{18}N_3O_3S^+$	344.1069	344.1059	2.90	$^{13}C_2{}^{12}C_{15}H_{18}N_3O_3S^+$	346.1136	346.1127	2.60
$C_{13}H_{13}N_2O_2^+$	229.0972	229.0972	0.00	$^{13}C_2{}^{12}C_{11}H_{13}N_2O_2^+$	231.1039	231.1037	0.86

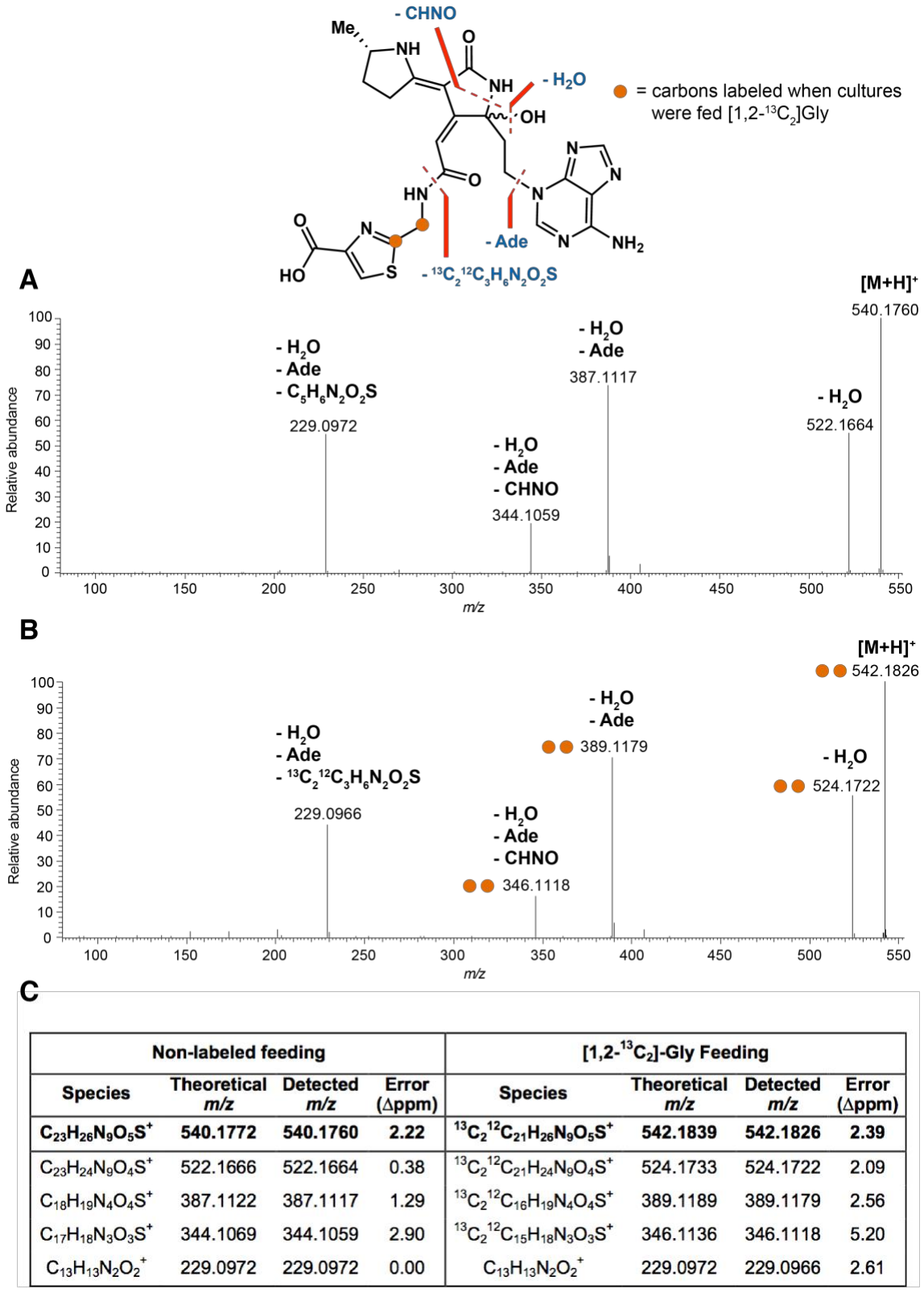
Fig. S5.
High-resolution MS^2 fragmentation of *in vivo* adducts 1 and 2 (*m/z* 540.1772) and the L-[2,3-$^{13}C_2$]Ala labeled species (*m/z* 542.1839). Comparison of the major fragment ions detected in (A) non-labeled and (B) L-[2,3-$^{13}C_2$]Ala labeled feeding experiments along with (C) the tabulated comparison.



Non-labeled feeding				L-[1-^{13}C]Met Feeding			
Species	Theoretical *m/z*	Detected *m/z*	Error (Δppm)	Species	Theoretical *m/z*	Detected *m/z*	Error (Δppm)
$C_{23}H_{26}N_9O_5S^+$	**540.1772**	**540.1760**	**2.22**	**$^{13}C_1{}^{12}C_{22}H_{26}N_9O_5S^+$**	**541.1805**	**541.1796**	**1.66**
$C_{23}H_{24}N_9O_4S^+$	522.1666	522.1664	0.38	$^{13}C_1{}^{12}C_{22}H_{24}N_9O_4S^+$	523.1699	523.1695	0.76
$C_{18}H_{19}N_4O_4S^+$	387.1122	387.1117	1.29	$^{13}C_1{}^{12}C_{17}H_{19}N_4O_4S^+$	388.1155	388.1150	1.28
$C_{17}H_{18}N_3O_3S^+$	344.1069	344.1059	2.90	$^{13}C_1{}^{12}C_{16}H_{18}N_3O_3S^+$	345.1102	345.1094	1.73
$C_{13}H_{13}N_2O_2^+$	229.0972	229.0972	0.00	$^{13}C_1{}^{12}C_{12}H_{13}N_2O_2^+$	230.1005	230.1003	0.86

Fig. S6.

High-resolution MS^2 fragmentation of *in vivo* adducts 1 and 2 (*m/z* 540.1772) and the L-[1-^{13}C]Met labeled species (*m/z* 541.1805). Comparison of the major fragment ions detected in (A) non-labeled and (B) L-[1-^{13}C]Met labeled feeding experiments along with (C) the tabulated comparison.

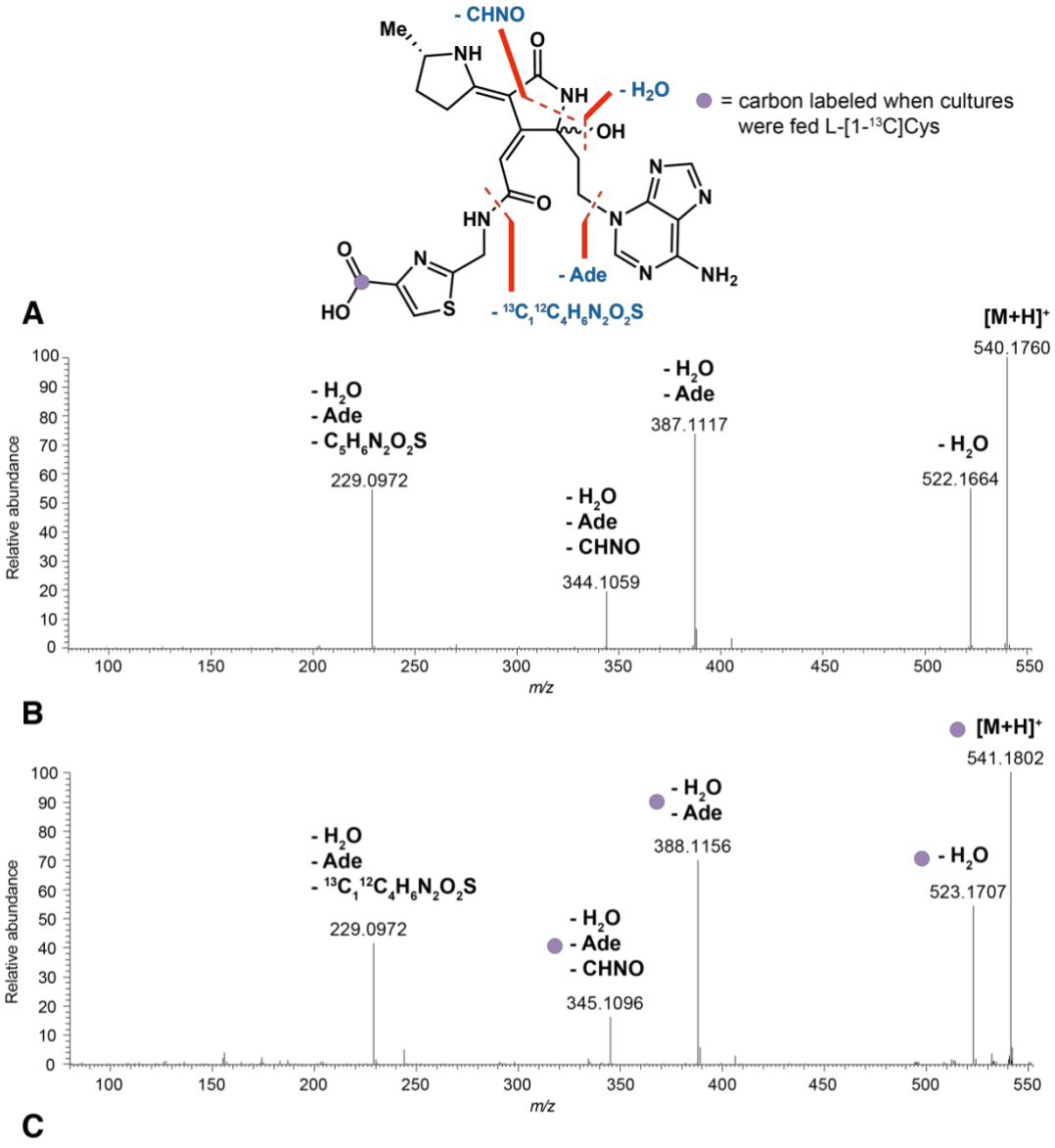


C

Non-labeled feeding				[1,2-$^{13}C_2$]-Gly Feeding			
Species	Theoretical *m/z*	Detected *m/z*	Error (Δppm)	Species	Theoretical *m/z*	Detected *m/z*	Error (Δppm)
$C_{23}H_{26}N_9O_5S^+$	**540.1772**	**540.1760**	**2.22**	**$^{13}C_2{}^{12}C_{21}H_{26}N_9O_5S^+$**	**542.1839**	**542.1826**	**2.39**
$C_{23}H_{24}N_9O_4S^+$	522.1666	522.1664	0.38	$^{13}C_2{}^{12}C_{21}H_{24}N_9O_4S^+$	524.1733	524.1722	2.09
$C_{18}H_{19}N_4O_4S^+$	387.1122	387.1117	1.29	$^{13}C_2{}^{12}C_{16}H_{19}N_4O_4S^+$	389.1189	389.1179	2.56
$C_{17}H_{18}N_3O_3S^+$	344.1069	344.1059	2.90	$^{13}C_2{}^{12}C_{15}H_{18}N_3O_3S^+$	346.1136	346.1118	5.20
$C_{13}H_{13}N_2O_2^+$	229.0972	229.0972	0.00	$C_{13}H_{13}N_2O_2^+$	229.0972	229.0966	2.61

Fig. S7.

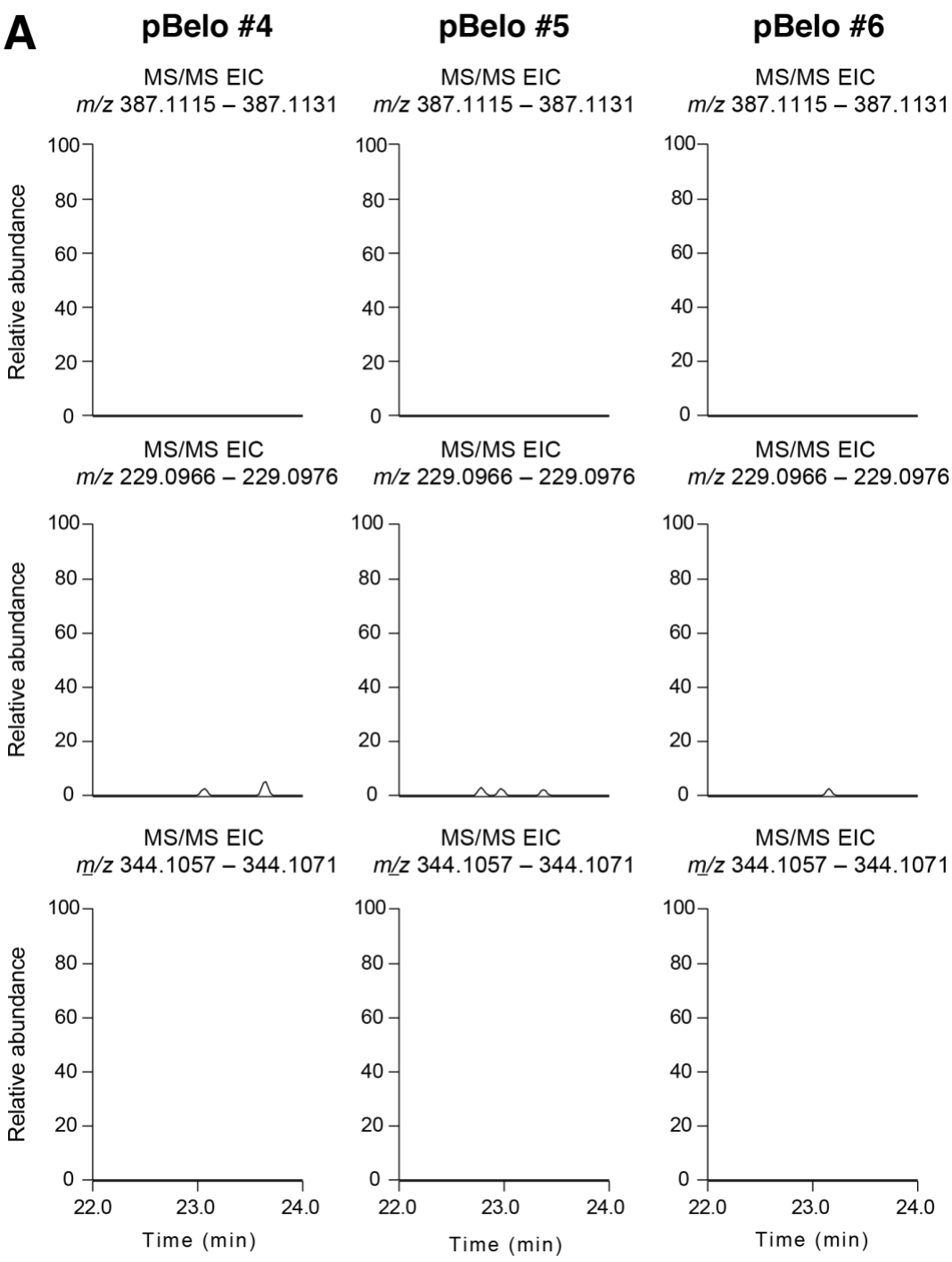
High-resolution MS^2 fragmentation of *in vivo* adducts 1 and 2 (*m/z* 540.1772) and the [1,2-$^{13}C_2$]Gly labeled species (*m/z* 542.1839). Comparison of the major fragment ions detected in (A) non-labeled and (B) [1,2-$^{13}C_2$]Gly labeled feeding experiments along with (C) the tabulated comparison.

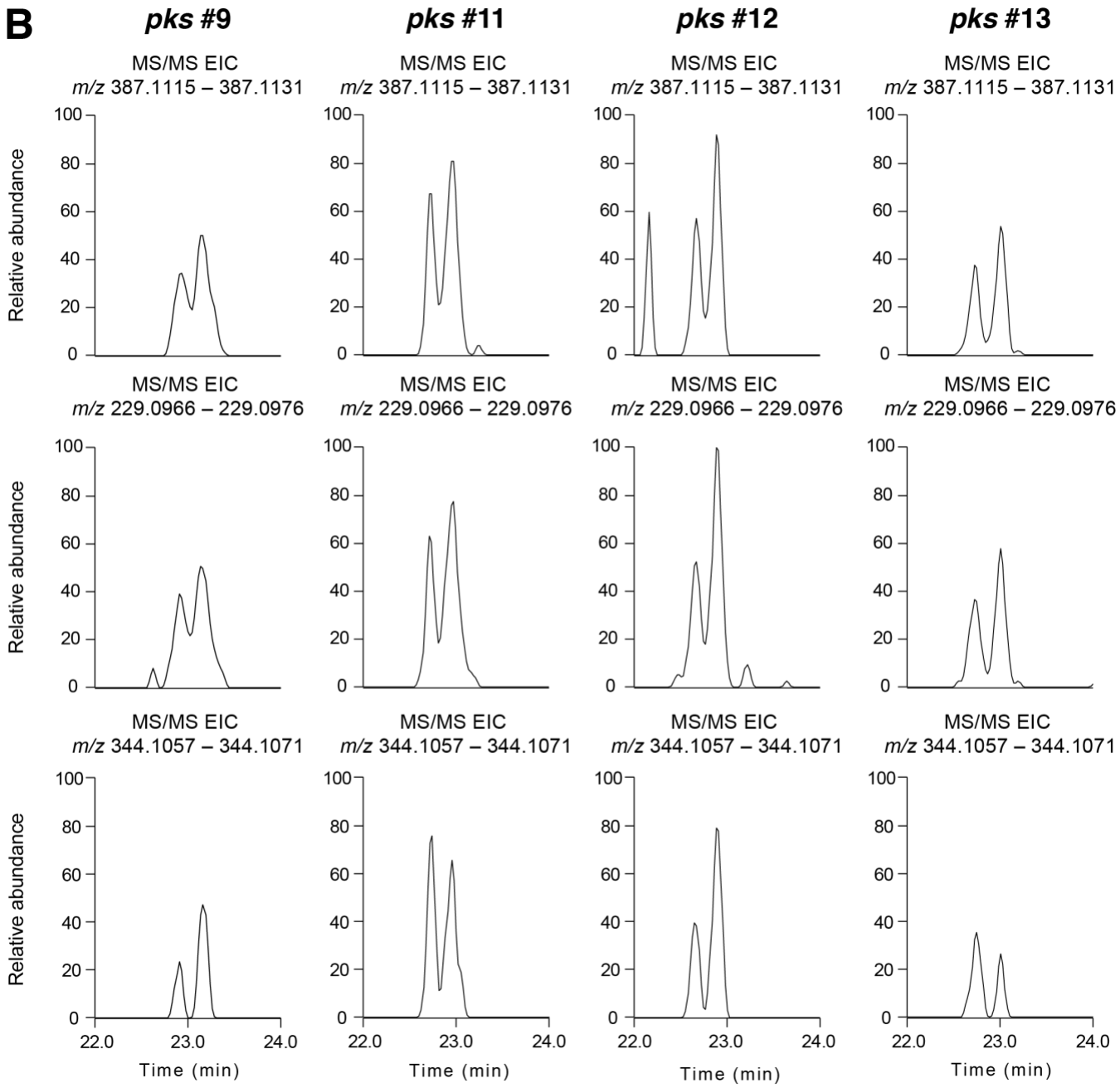


Non-labeled feeding				L-[1-^{13}C]Cys Feeding			
Species	Theoretical *m/z*	Detected *m/z*	Error (Δppm)	Species	Theoretical *m/z*	Detected *m/z*	Error (Δppm)
$C_{23}H_{26}N_9O_5S^+$	**540.1772**	**540.1760**	**2.22**	**$^{13}C_1{}^{12}C_{22}H_{26}N_9O_5S^+$**	**541.1805**	**541.1802**	**0.55**
$C_{23}H_{24}N_9O_4S^+$	522.1666	522.1664	0.38	$^{13}C_1{}^{12}C_{22}H_{24}N_9O_4S^+$	523.1699	523.1707	1.52
$C_{18}H_{19}N_4O_4S^+$	387.1122	387.1117	1.29	$^{13}C_1{}^{12}C_{17}H_{19}N_4O_4S^+$	388.1155	388.1156	0.25
$C_{17}H_{18}N_3O_3S^+$	344.1069	344.1059	2.90	$^{13}C_1{}^{12}C_{16}H_{18}N_3O_3S^+$	345.1102	345.1096	1.73
$C_{13}H_{13}N_2O_2^+$	229.0972	229.0972	0.00	$C_{13}H_{13}N_2O_2^+$	229.0972	229.0972	0.00

Fig. S8.

High-resolution MS^2 fragmentation of *in vivo* adducts 1 and 2 (*m/z* 540.1772) and the L-[1-^{13}C]Cys labeled species (*m/z* 541.1805). Comparison of the major fragment ions detected in (A) non-labeled and (B) L-[1-^{13}C]Cys labeled feeding experiments along with (C) the tabulated comparison.


A
pBelo #4
pBelo #5
pBelo #6
MS/MS EIC
m/z 387.1115 – 387.1131
MS/MS EIC
m/z 229.0966 – 229.0976
MS/MS EIC
m/z 344.1057 – 344.1071
Relative abundance
100
80
60
40
20
0
22.0
23.0
24.0
Time (min)


B
pks #9
pks #11
pks #12
pks #13
MS/MS EIC
m/z 387.1115 – 387.1131
MS/MS EIC
m/z 229.0966 – 229.0976
MS/MS EIC
m/z 344.1057 – 344.1071
Relative abundance
100
80
60
40
20
0
22.0
23.0
24.0
Time (min)

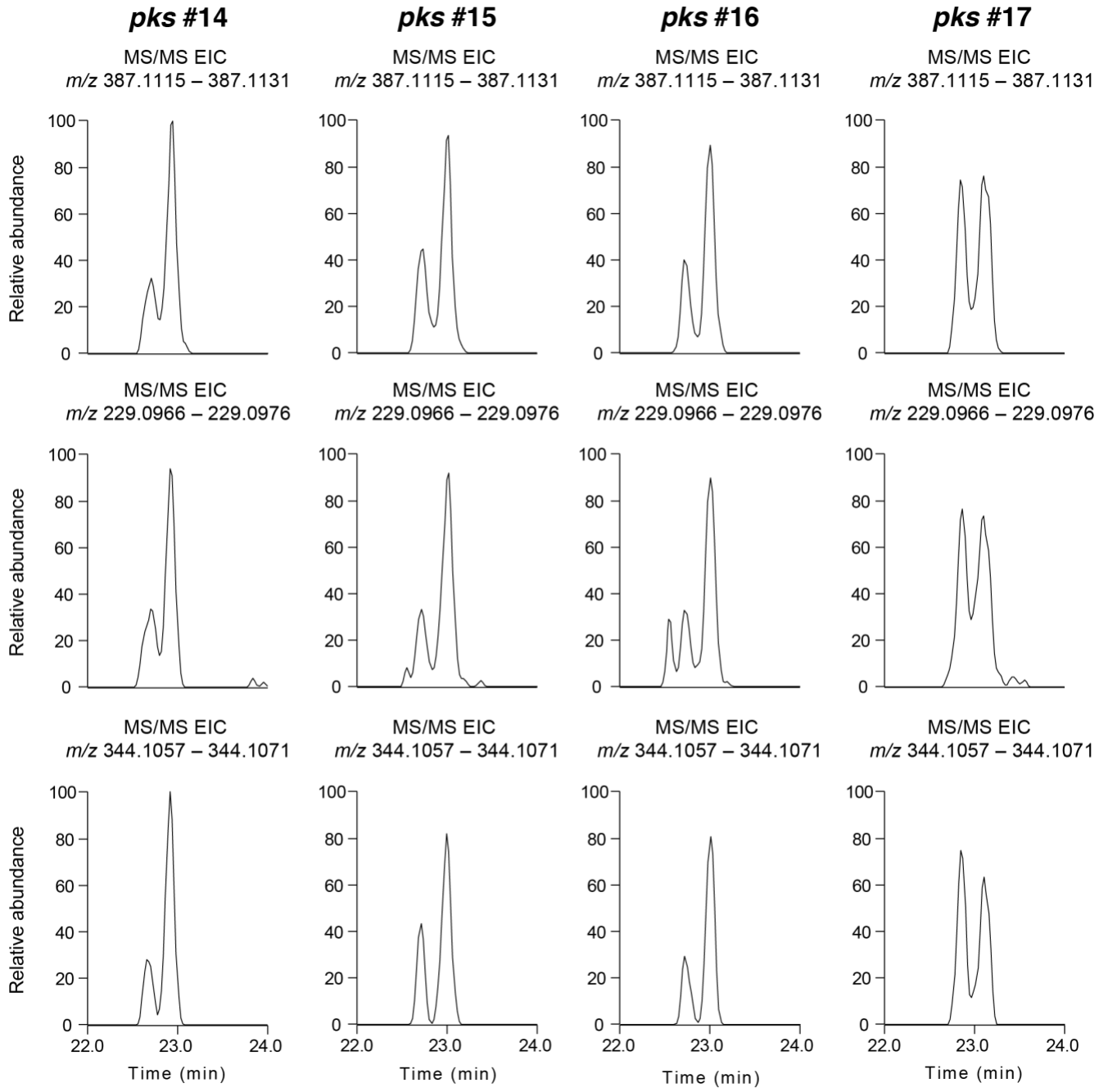


Fig. S9.

Detection of DNA adducts 1 and 2 in colonic epithelial cells from mice exposed to *pks*$^+$ *E. coli*. Extracted ion chromatograms of the MS^2 fragmentation ions (*m/z* 229.0966–229.0976, 344.1071–344.1057, 387.1131–387.1115) of the adducts **1** and **2** precursor ion (*m/z* 540.1772) in colonic epithelial cells isolated from 3 mice colonized with pBelo *E. coli* (A) or 8 mice colonized with *pks*$^+$ *E. coli* (B) for 2 weeks.

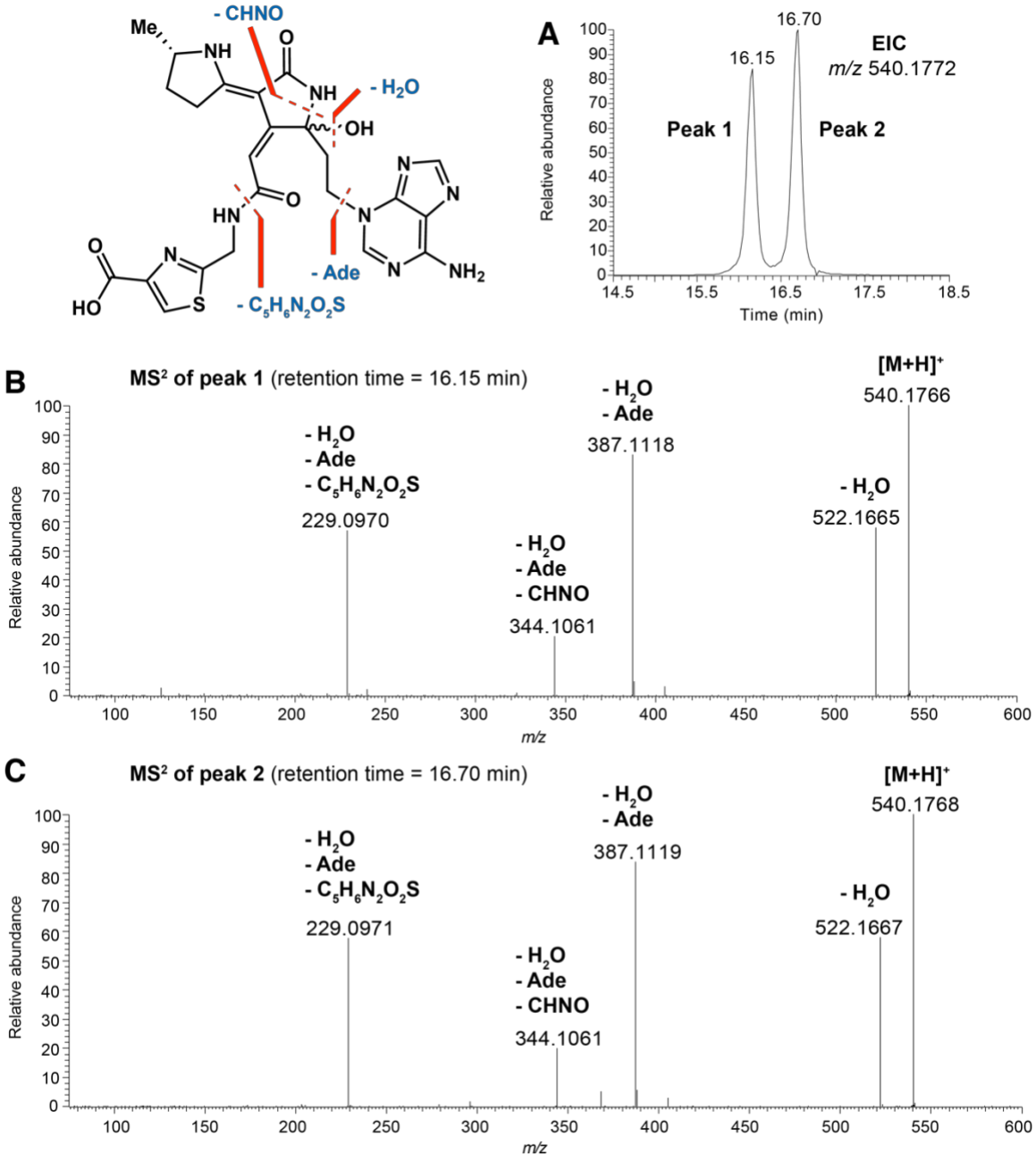
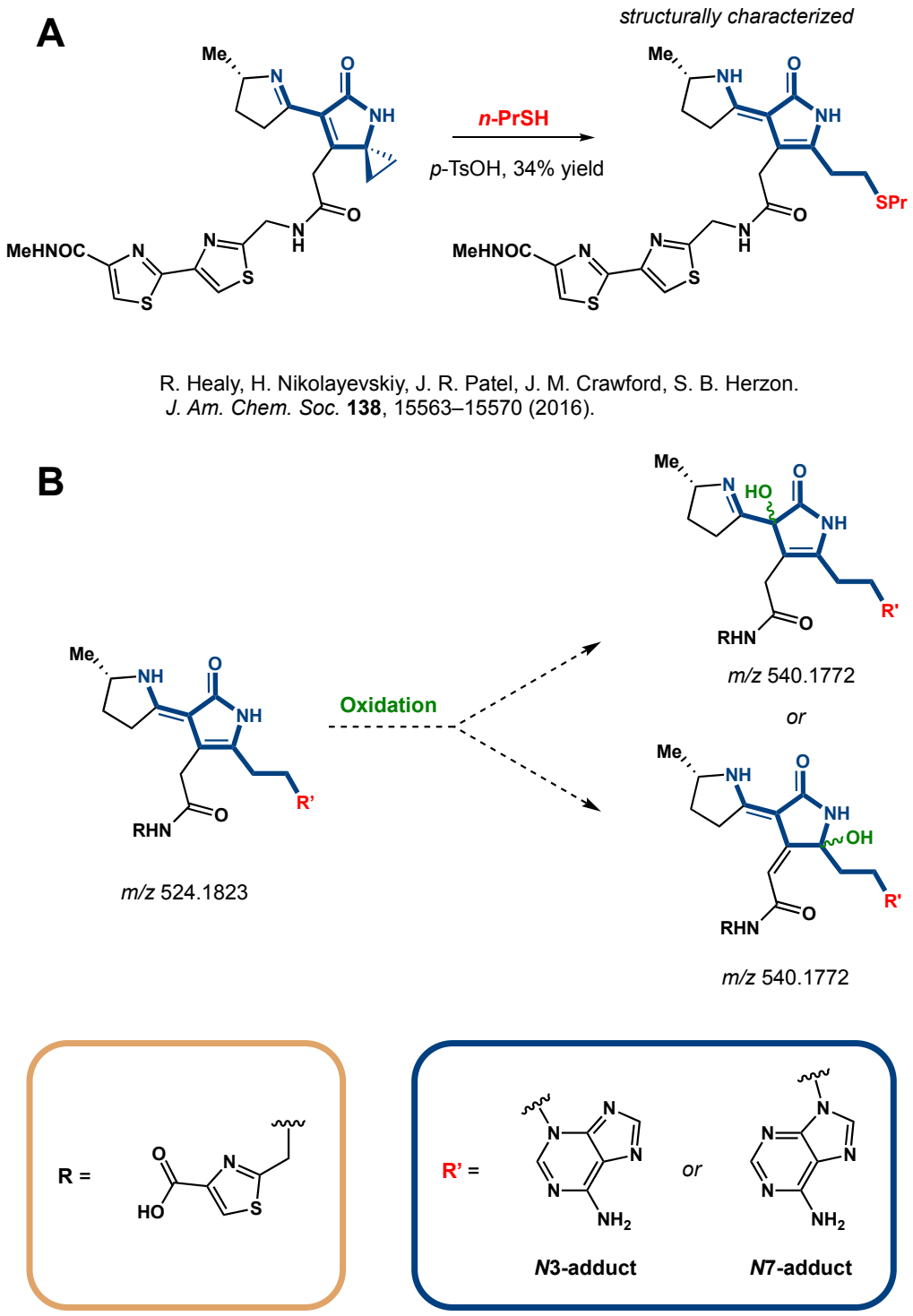


Fig. S10.

High-resolution MS^2 fragmentation of the two observed extracted ion chromatogram peaks with *m/z* 540.1772. (A) Extracted ion chromatogram of DNA adducts **1** and **2** (*m/z* 540.1772) showing the observed two peaks, (B) MS^2 fragmentation of the first peak (retention time = 16.15 min), and (C) MS^2 fragmentation of the second peak (retention time = 16.70 min).



R. Healy, H. Nikolayevskiy, J. R. Patel, J. M. Crawford, S. B. Herzon.
J. Am. Chem. Soc. **138**, 15563–15570 (2016).

Fig. S11.
Our initially proposed structures of DNA adducts 1 and 2. (A) Known cyclopropane ring opening of a synthetic 'colibactin mimic' with propanethiol to give a structurally characterized conjugated enamine compound (*22*). (B) Initially proposed chemical structures of DNA adducts **1** and **2** based on MS^2 fragmentation data and the known reactivity of synthetic 'colibactin mimics' with nucleophiles.

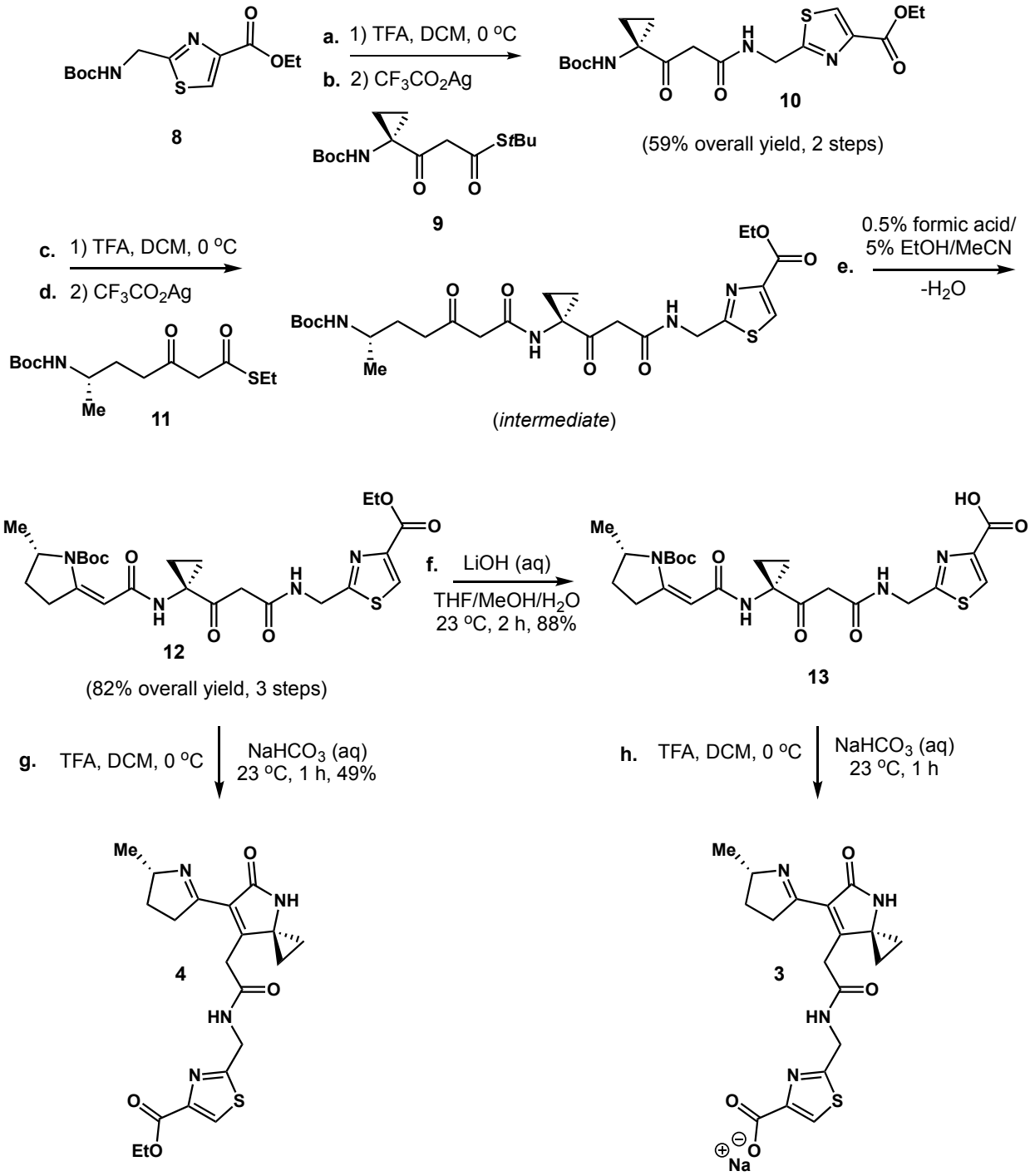


Fig. S12.

Overall synthetic scheme used to access 'colibactin mimics' 3 and 4. (a) Trifluoroacetic acid (TFA), CH_2Cl_2, 0 ºC; (b) silver trifluoroacetate (CF_3CO_2Ag), triethylamine (NEt_3), DMF, 0 ºC (59%, 2 steps); (c) TFA, CH_2Cl_2, 0 ºC; (d) CF_3CO_2Ag NEt_3, DMF, 0 ºC; (e) concentrate from 0.5% HCO_2H-5% EtOH-MeCN (82%, 3 steps); (f) Aqueous lithium hydroxide (LiOH), THF/MeOH/H_2O, 23 ºC, 2 h, 88%; (g) TFA, CH_2Cl_2, 0 ºC; aqueous $NaHCO_3$, 49%; (h) TFA, CH_2Cl_2, 0 ºC; aqueous $NaHCO_3$.

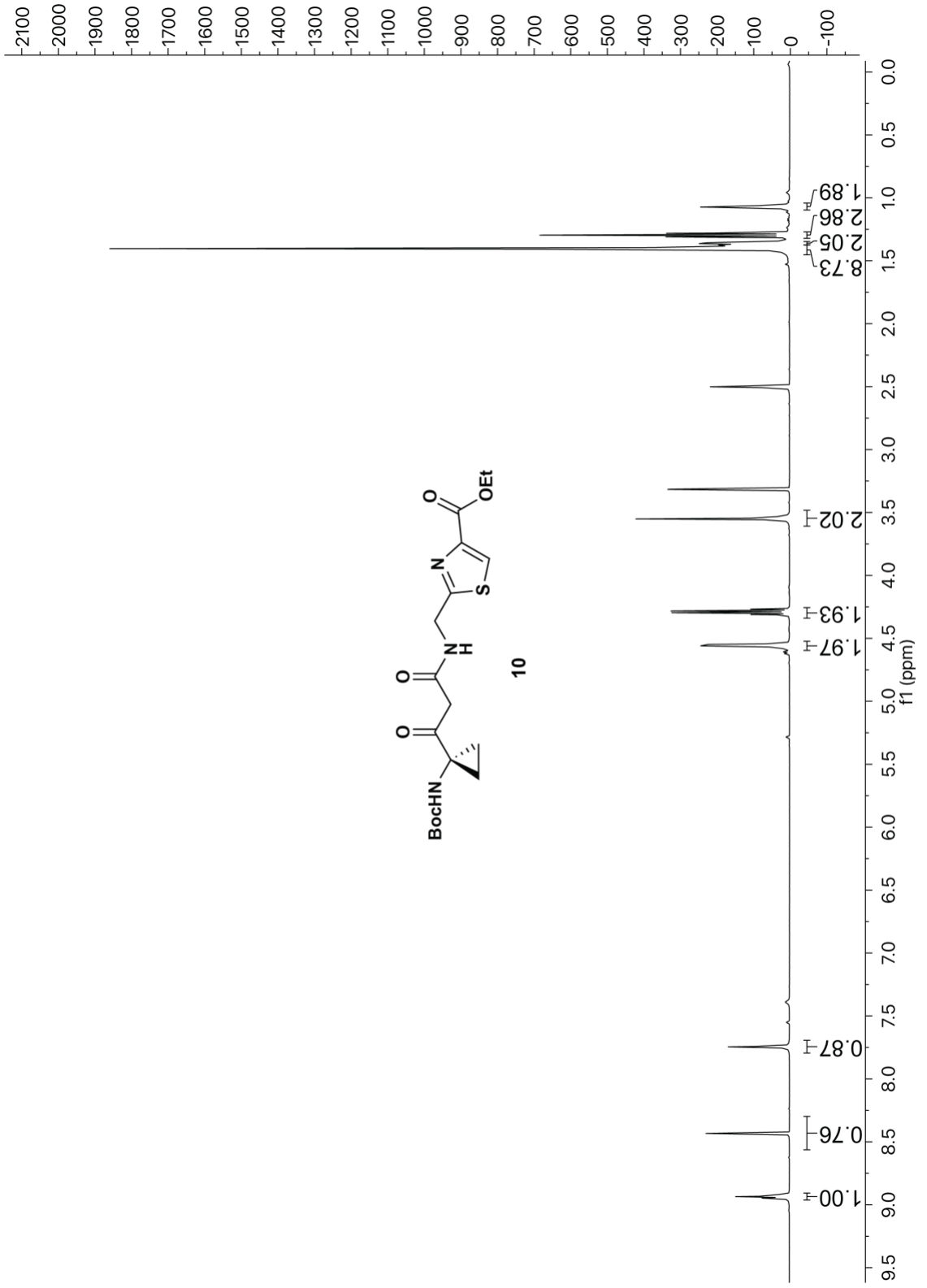


Fig. S13.
^{1}H spectrum of compound **10** (recorded in DMSO-d_6 at 600 MHz).

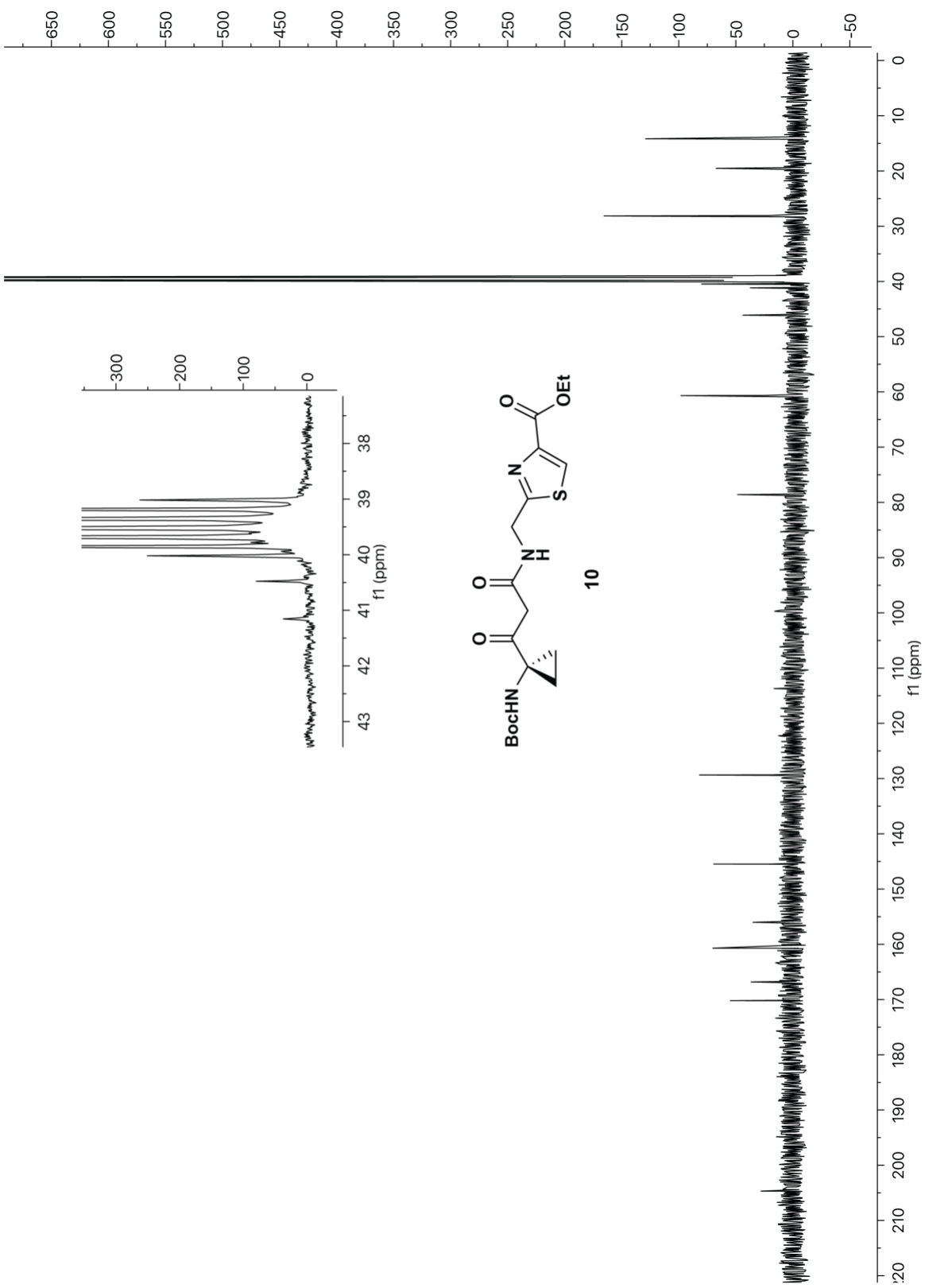


Fig. S14.
^{13}C spectrum of compound **10** (recorded in DMSO-d_6 at 600 MHz).

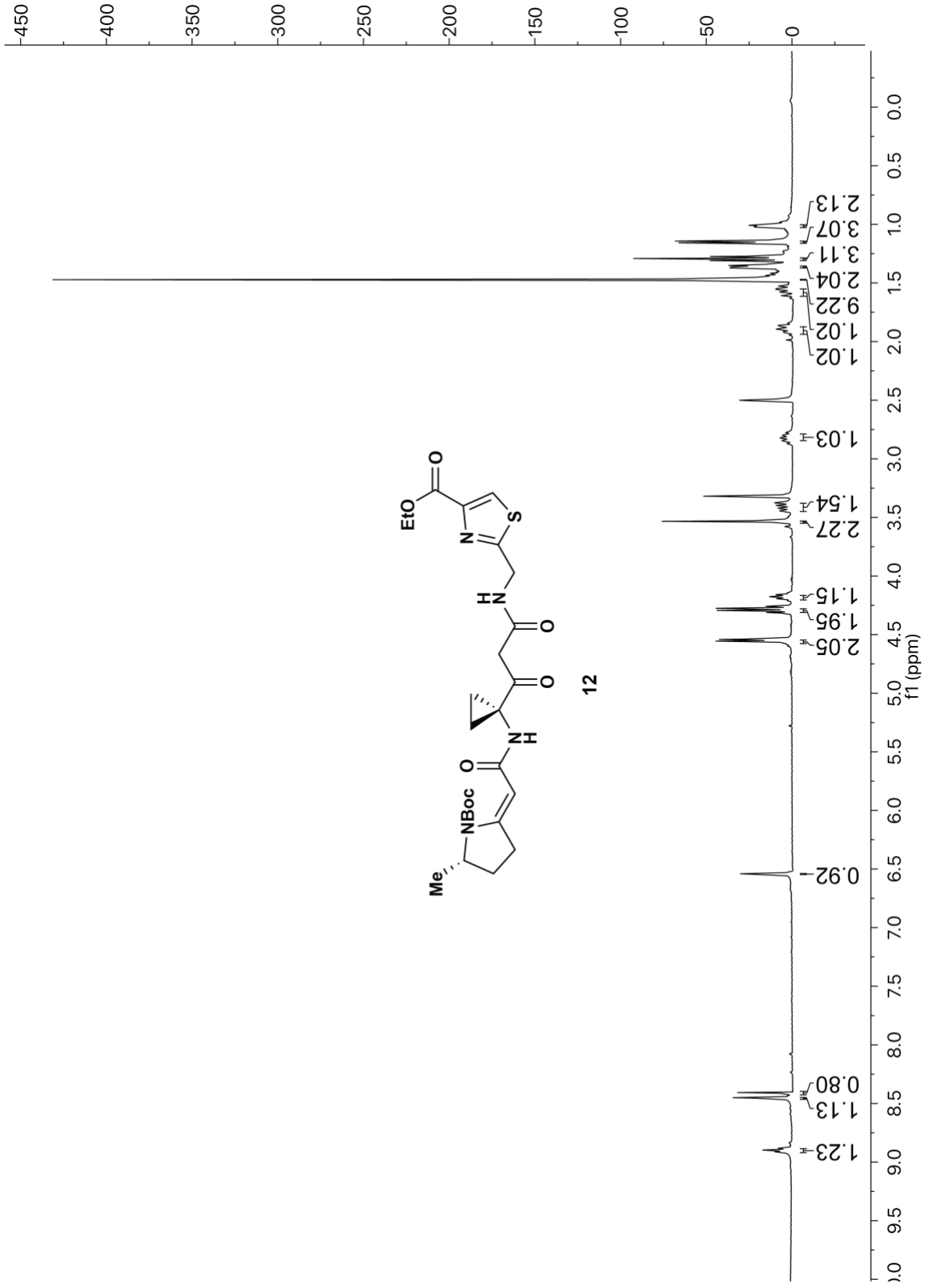


Fig. S15.
^{1}H spectrum of compound **12** (recorded in DMSO-d_6 at 600 MHz).

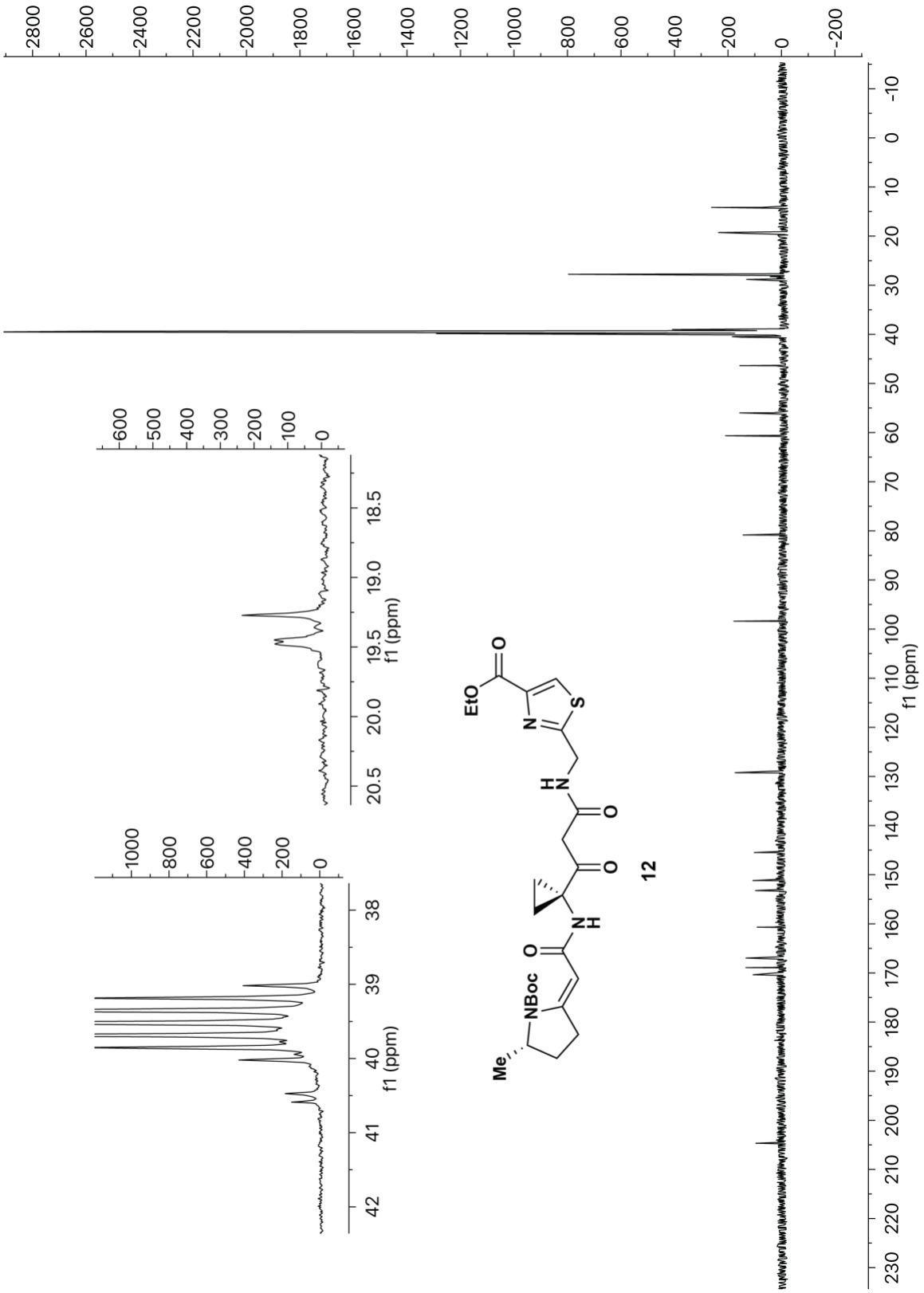


Fig. S16.
^{13}C spectrum of compound **12** (recorded in DMSO-d_6 at 600 MHz).

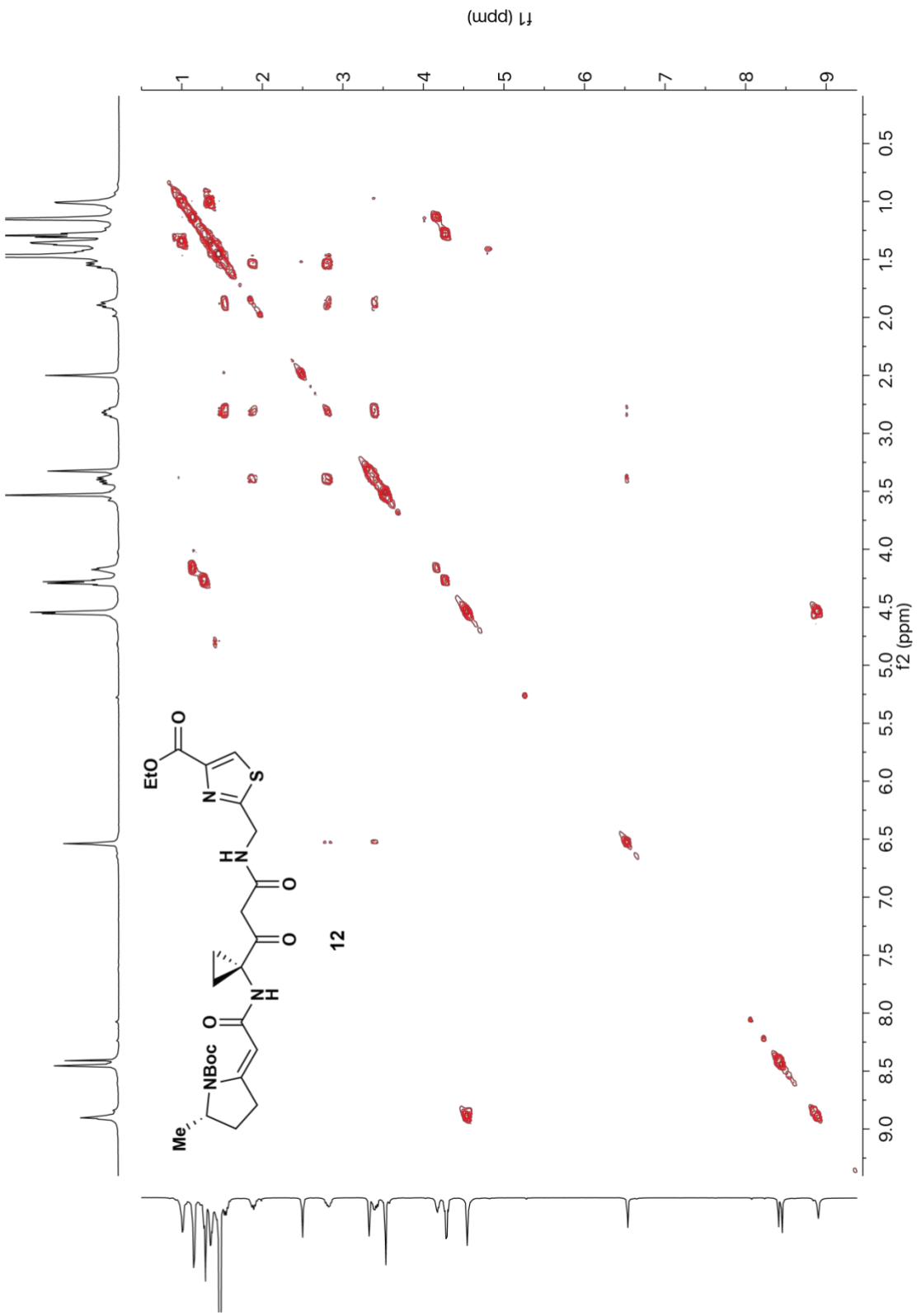


Fig. S17.
gCOSY spectrum of compound **12** (recorded in DMSO-d_6 at 600 MHz).

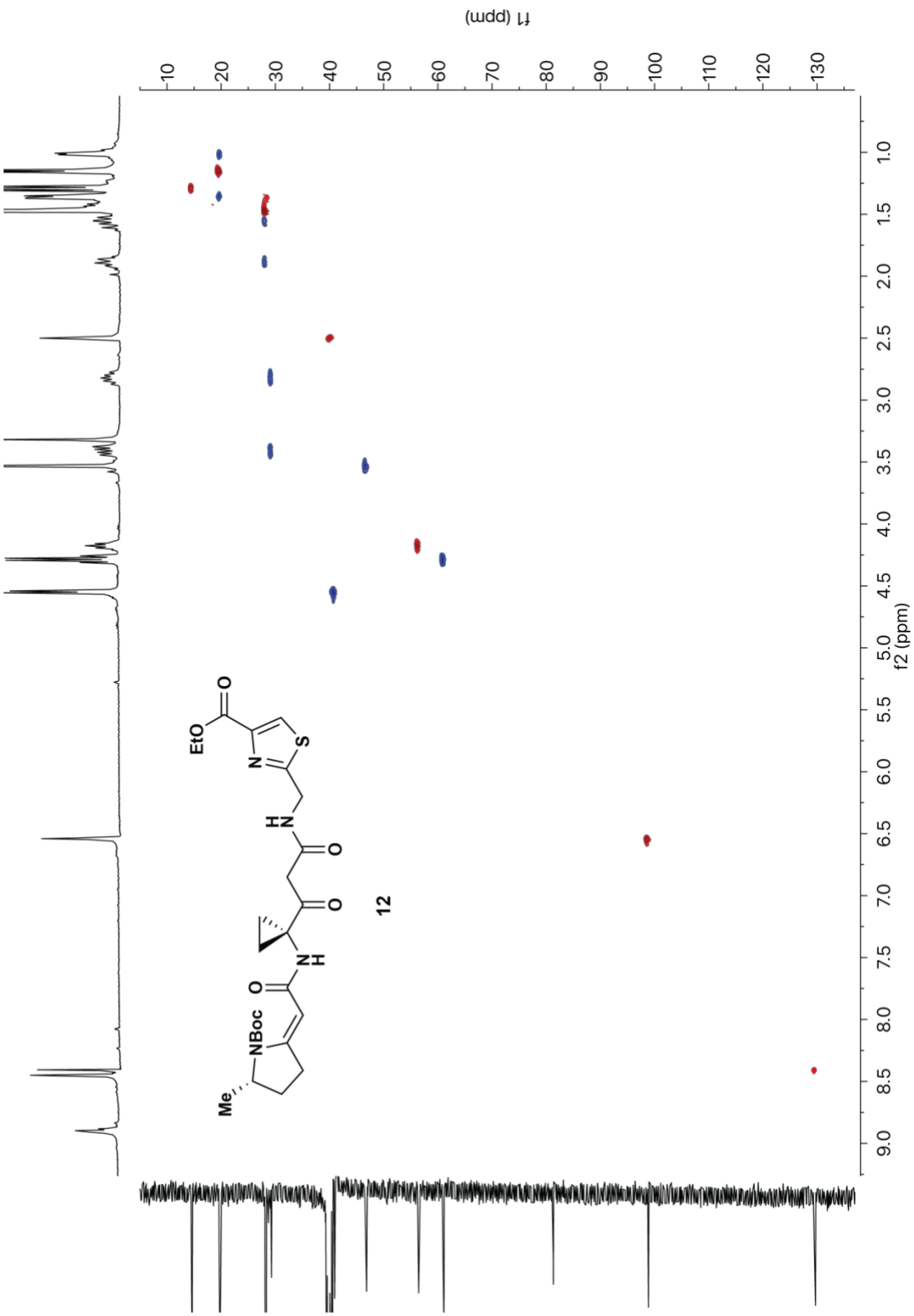


Fig. S18.
gHSQCAD spectrum of compound **12** (recorded in DMSO-d_6 at 600 MHz).

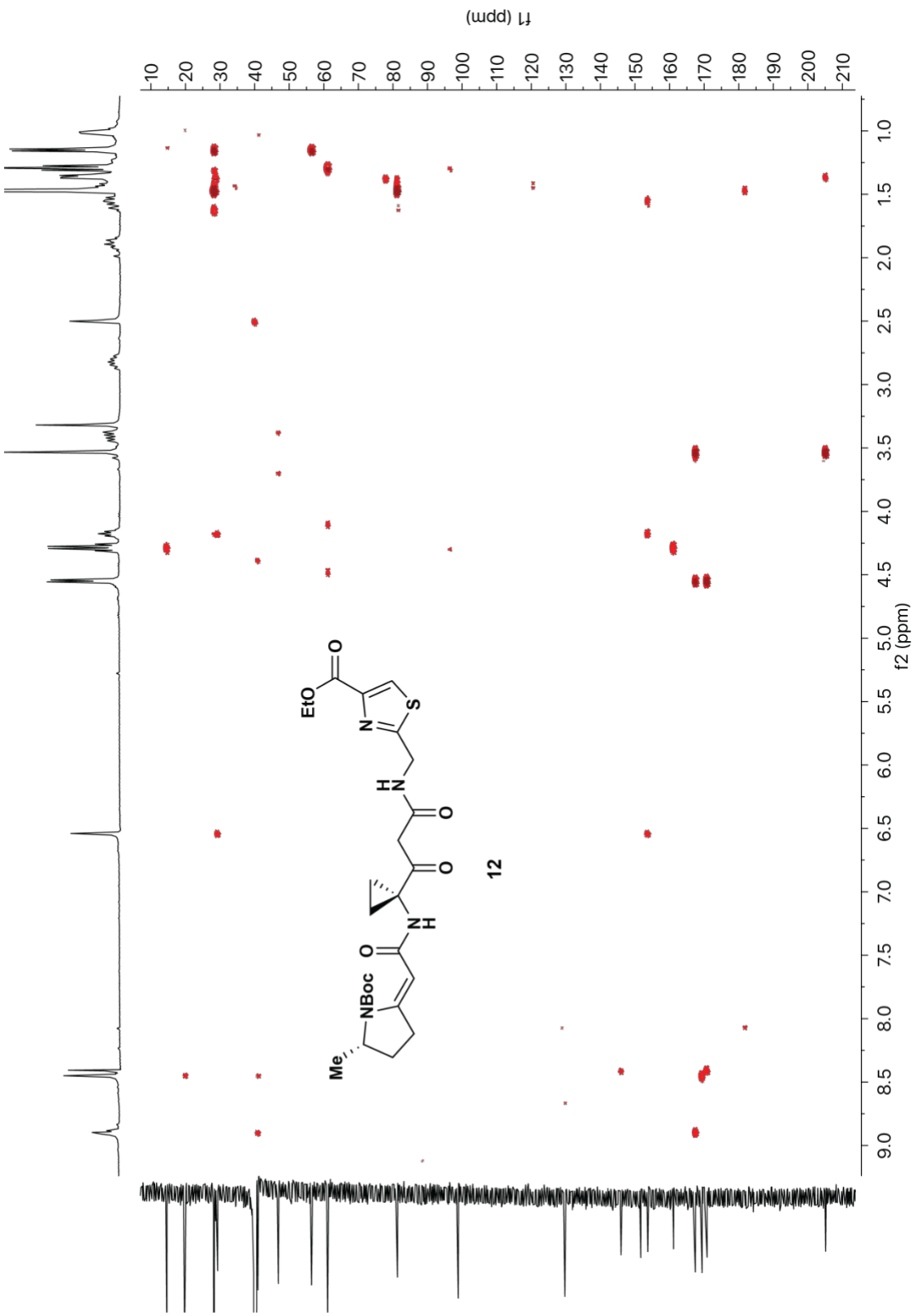


Fig. S19.
gHMBCAD spectrum of compound **12** (recorded in DMSO-d_6 at 600 MHz).

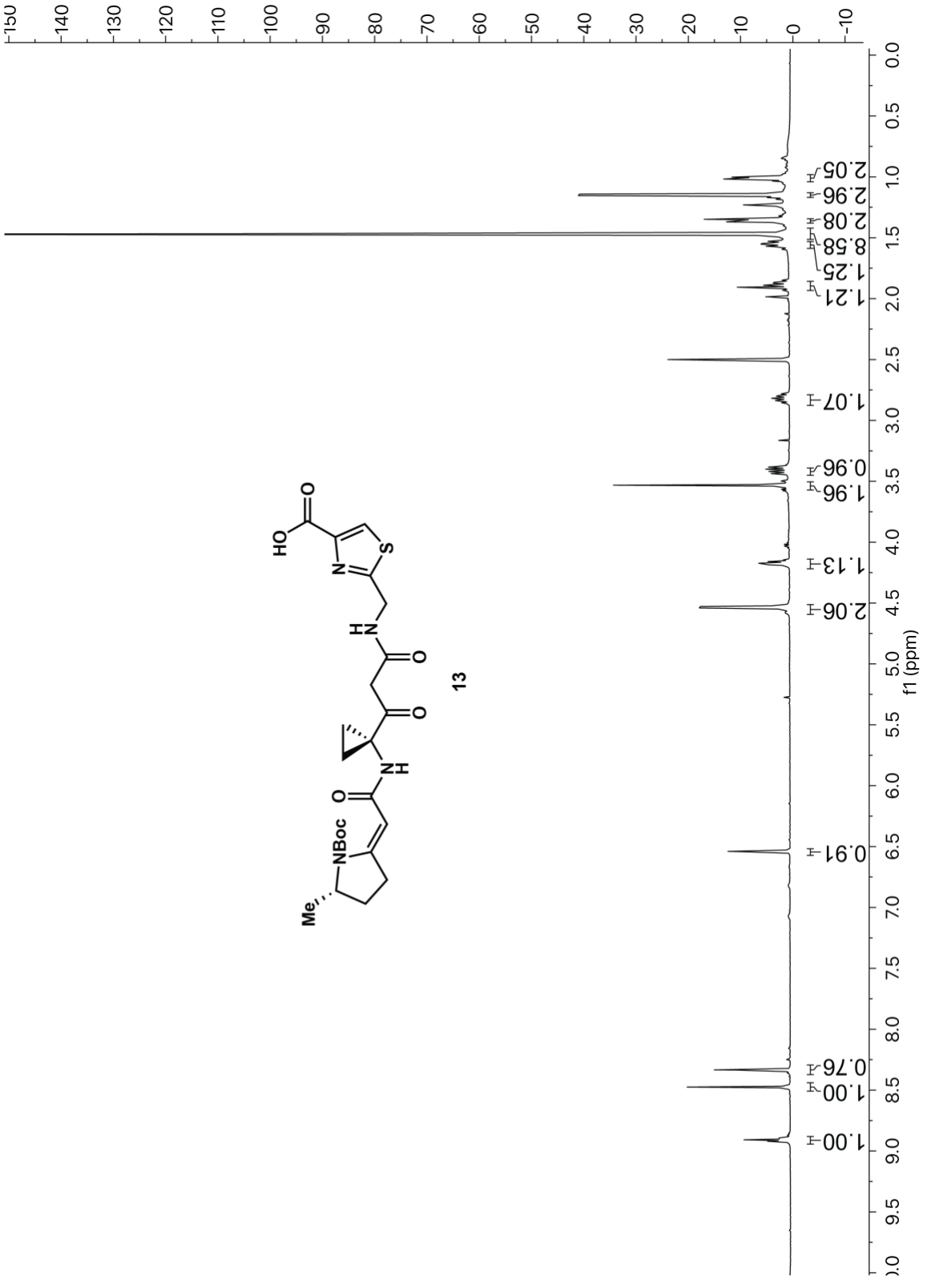


Fig. S20.
^{1}H spectrum of compound **13** (recorded in DMSO-d_6 at 600 MHz).

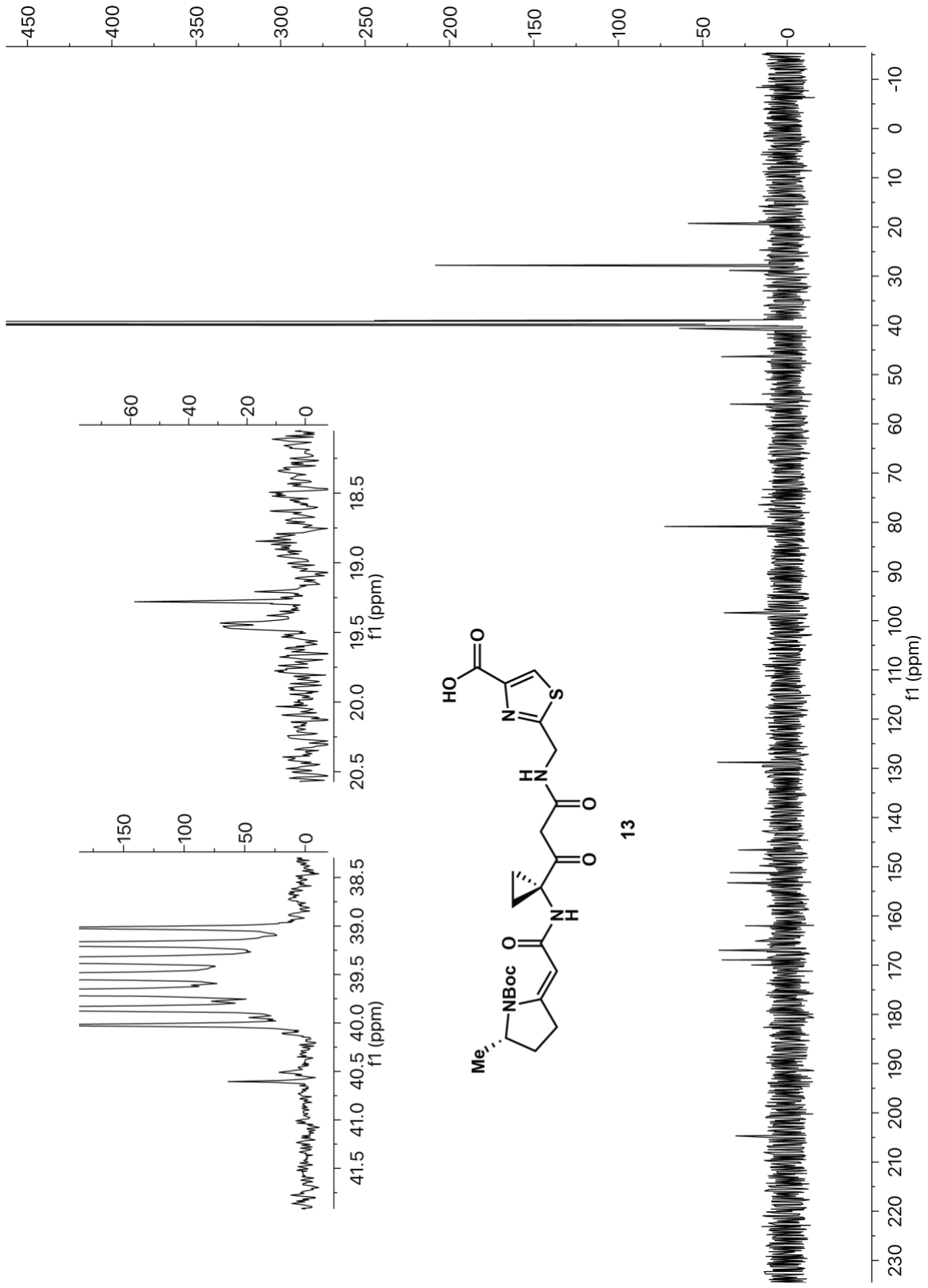


Fig. S21.
^{13}C spectrum of compound **13** (recorded in DMSO-d_6 at 600 MHz).

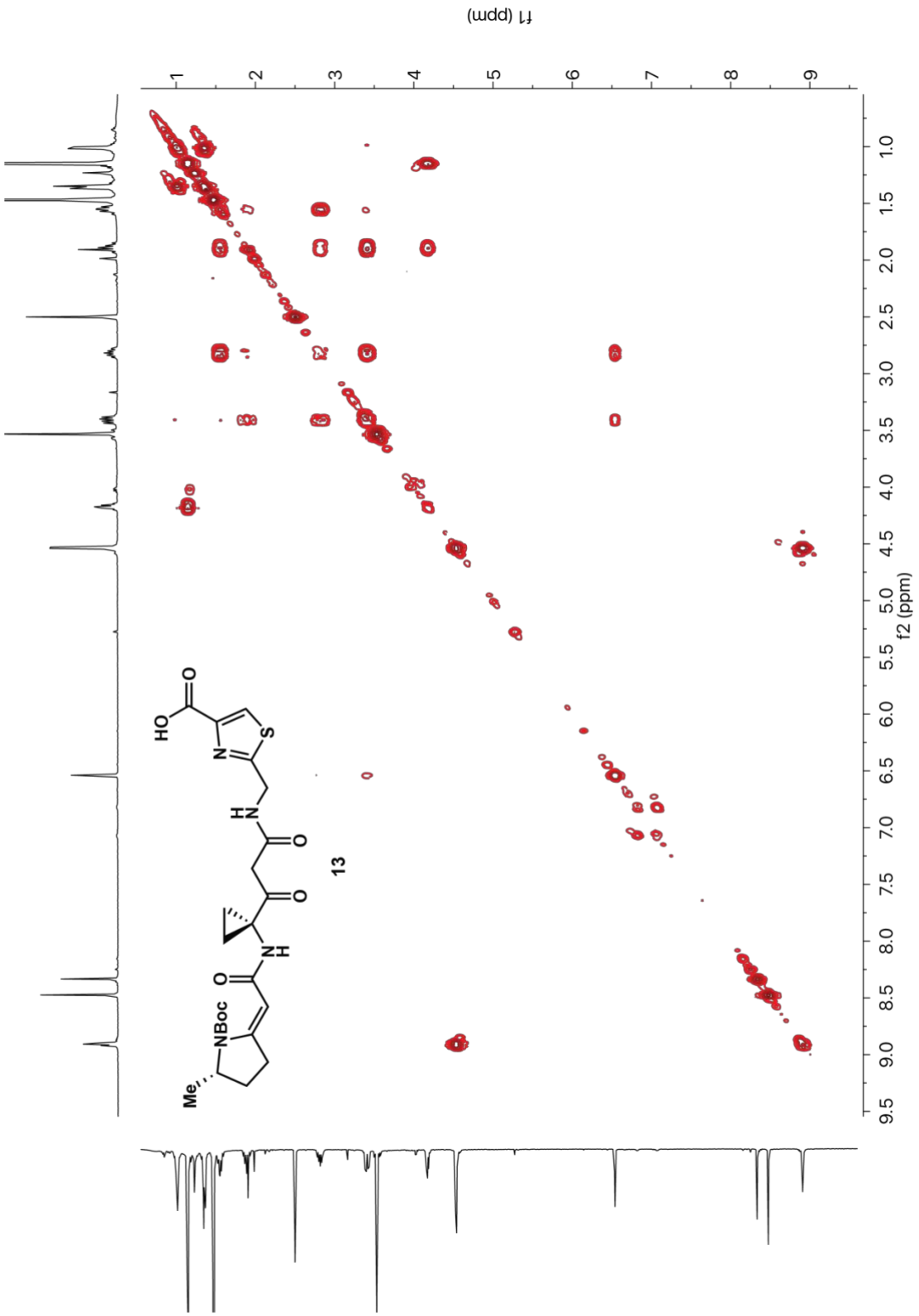


Fig. S22.
gCOSY spectrum of compound **13** (recorded in DMSO-d_6 at 600 MHz).

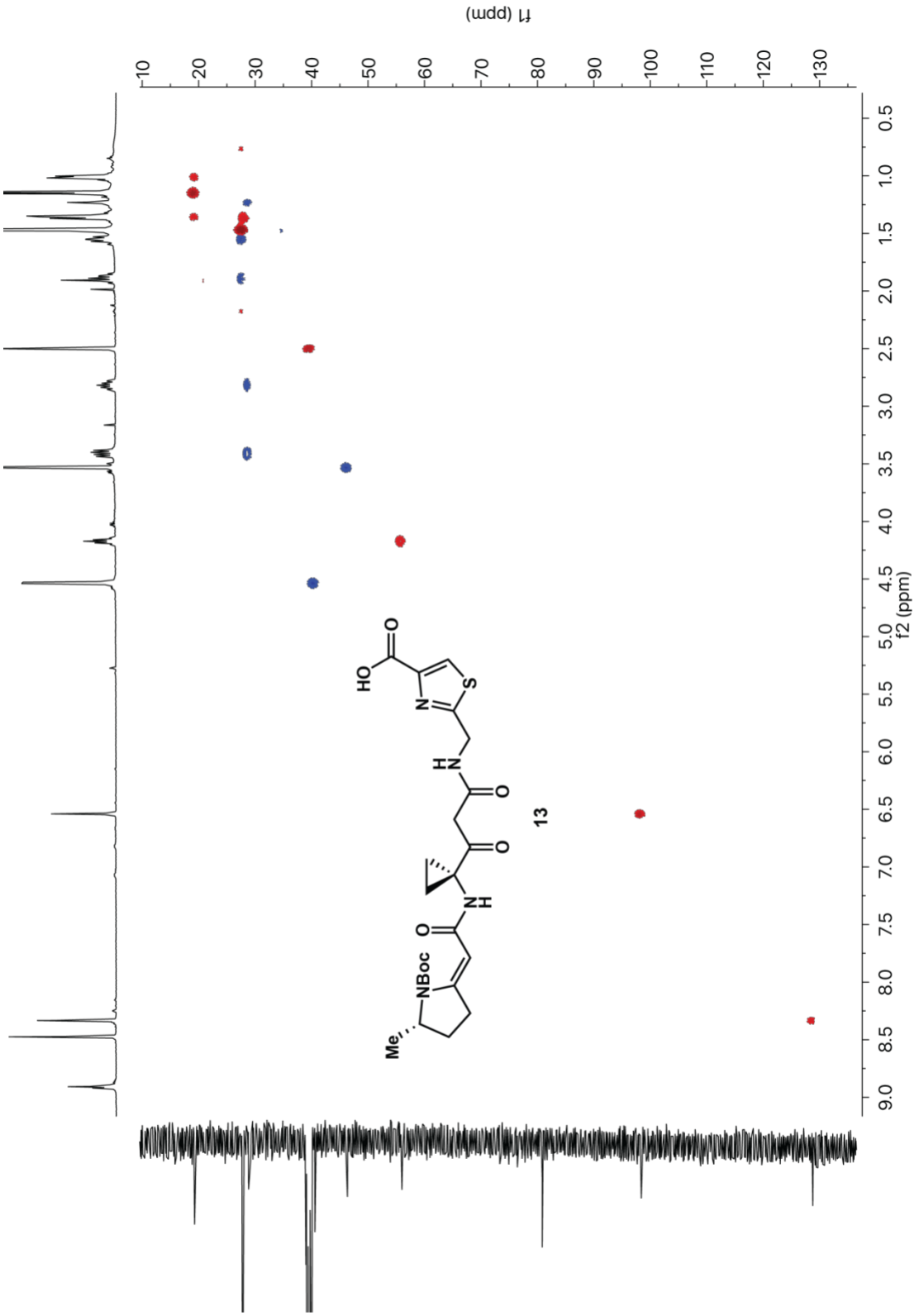


Fig. S23.
gHSQCAD spectrum of compound **13** (recorded in DMSO-d_6 at 600 MHz).

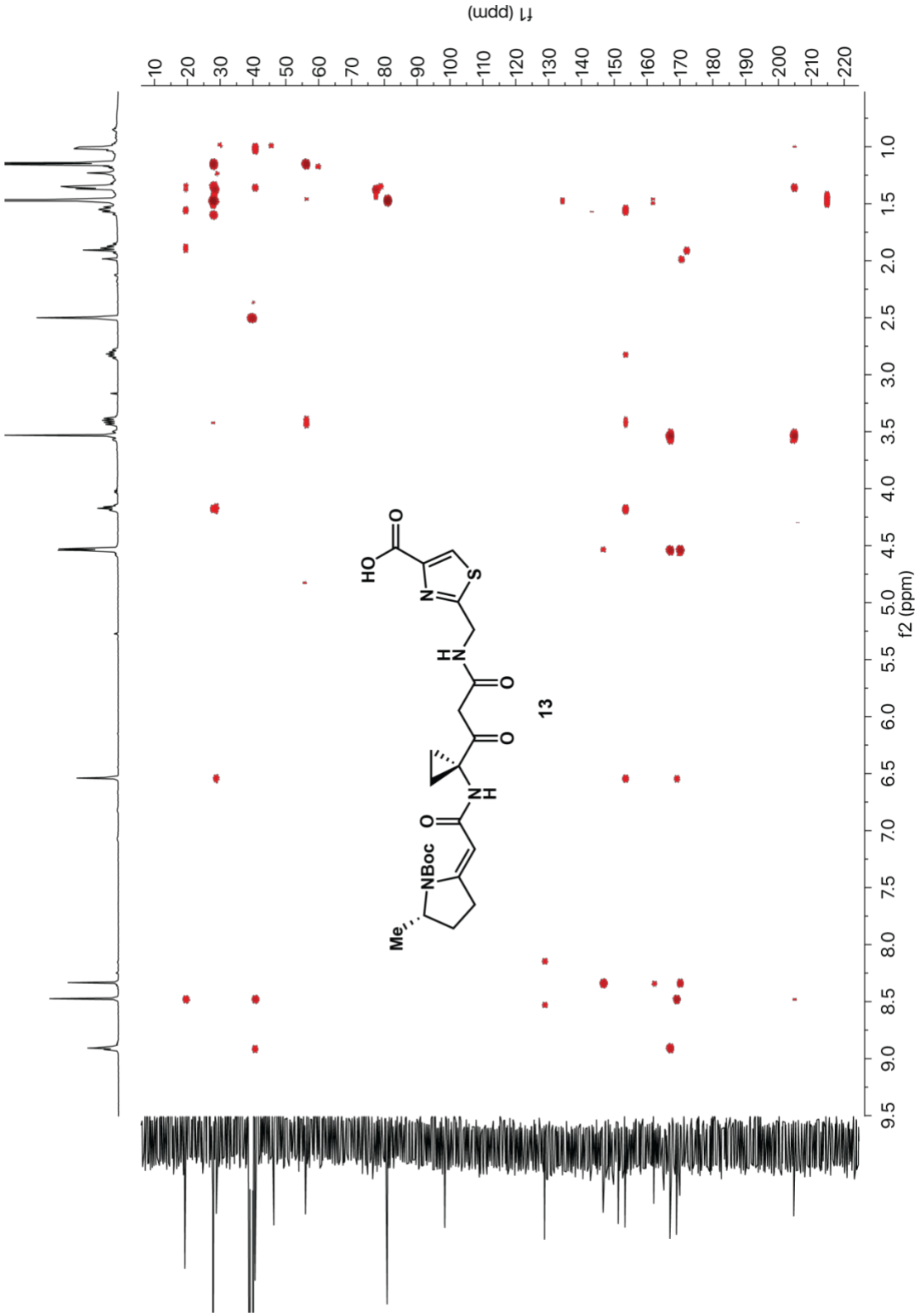


Fig. S24.
gHMBCAD spectrum of compound **13** (recorded in DMSO-d_6 at 600 MHz).

Fig. S25.
^{1}H spectrum of compound **3** (recorded in DMSO-d_6 at 600 MHz).

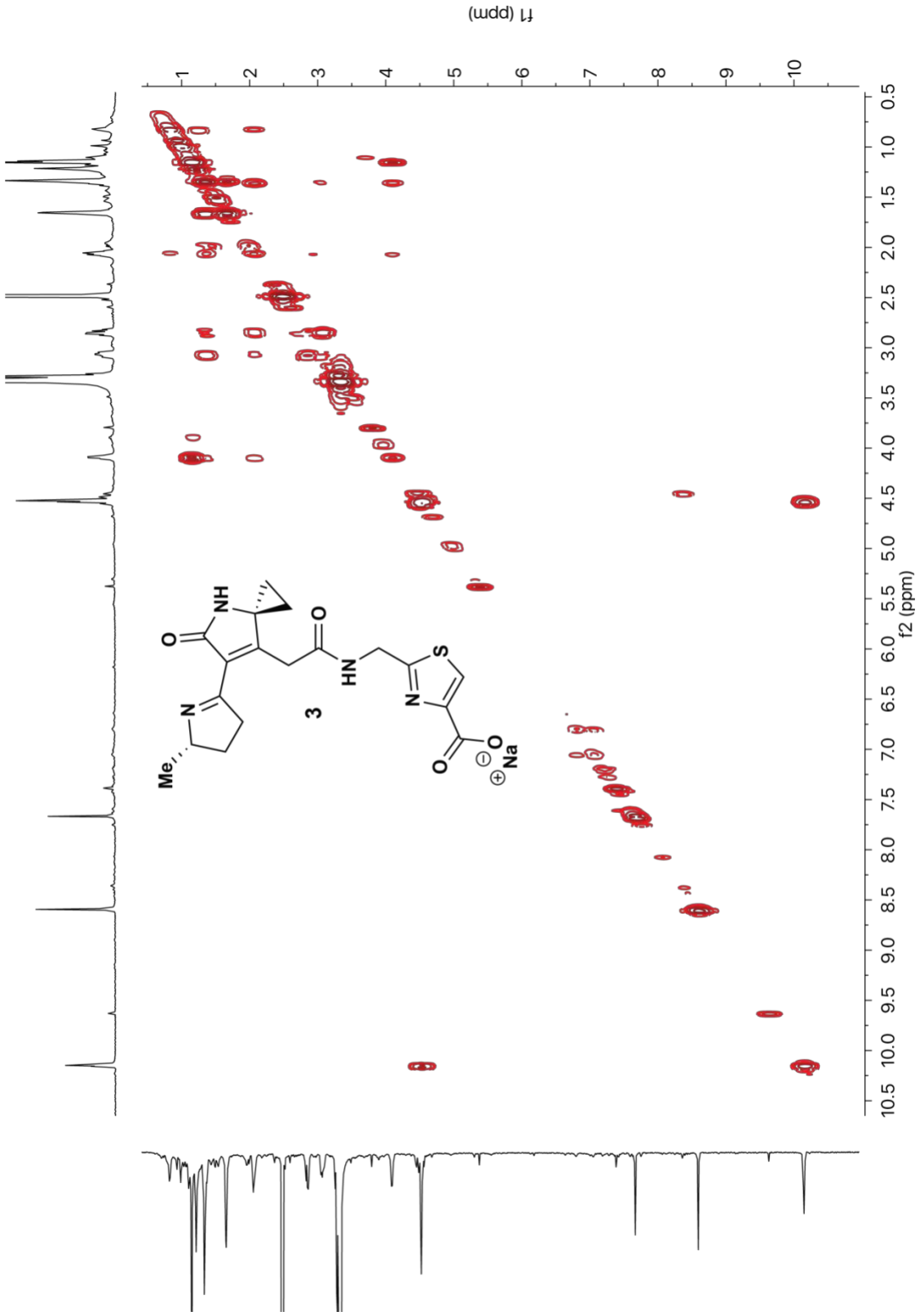


Fig. S26.
gCOSY spectrum of compound **3** (recorded in DMSO-d_6 at 600 MHz).

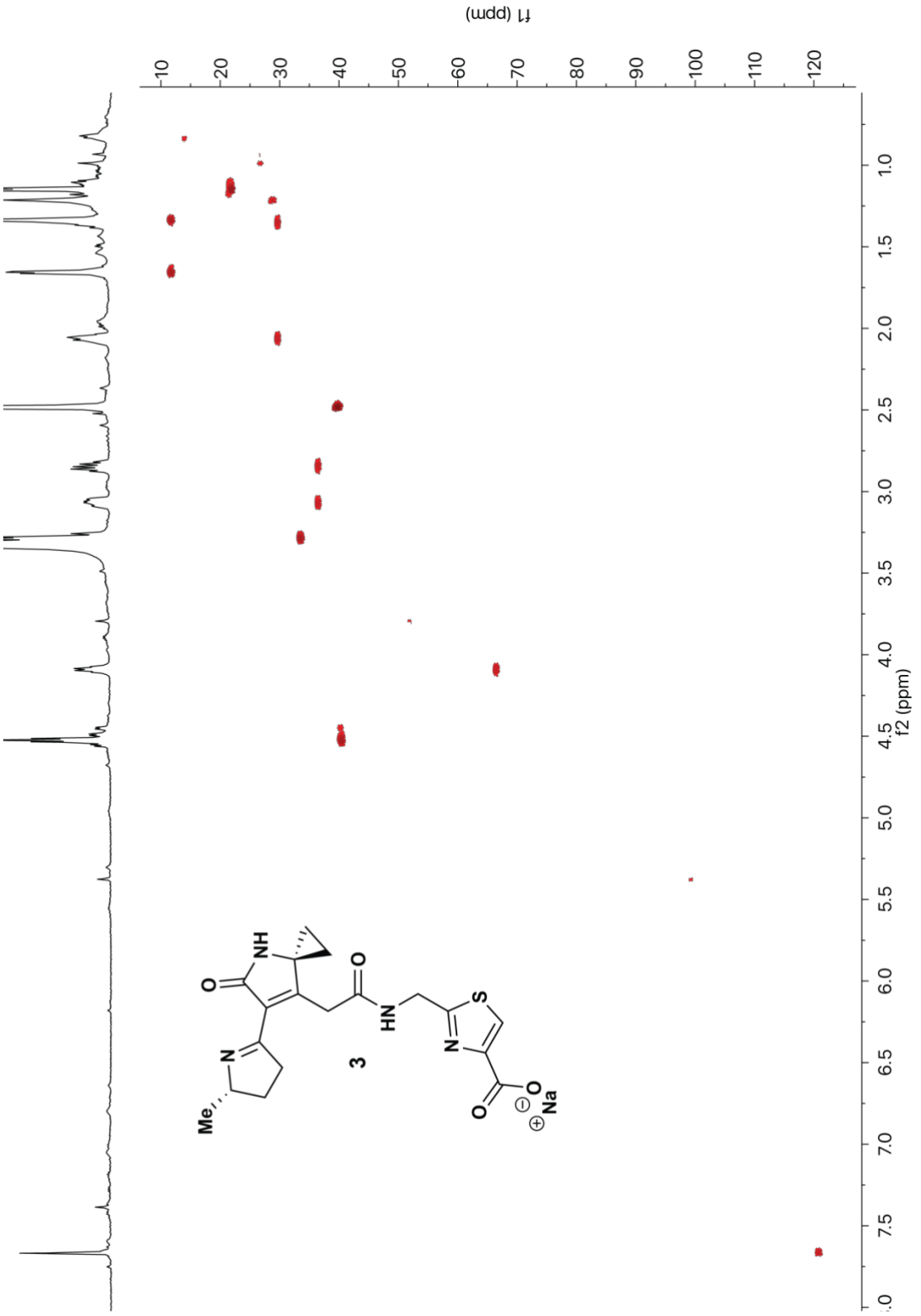


Fig. S27.
gHSQCAD spectrum of compound **3** (recorded in DMSO-d_6 at 600 MHz).

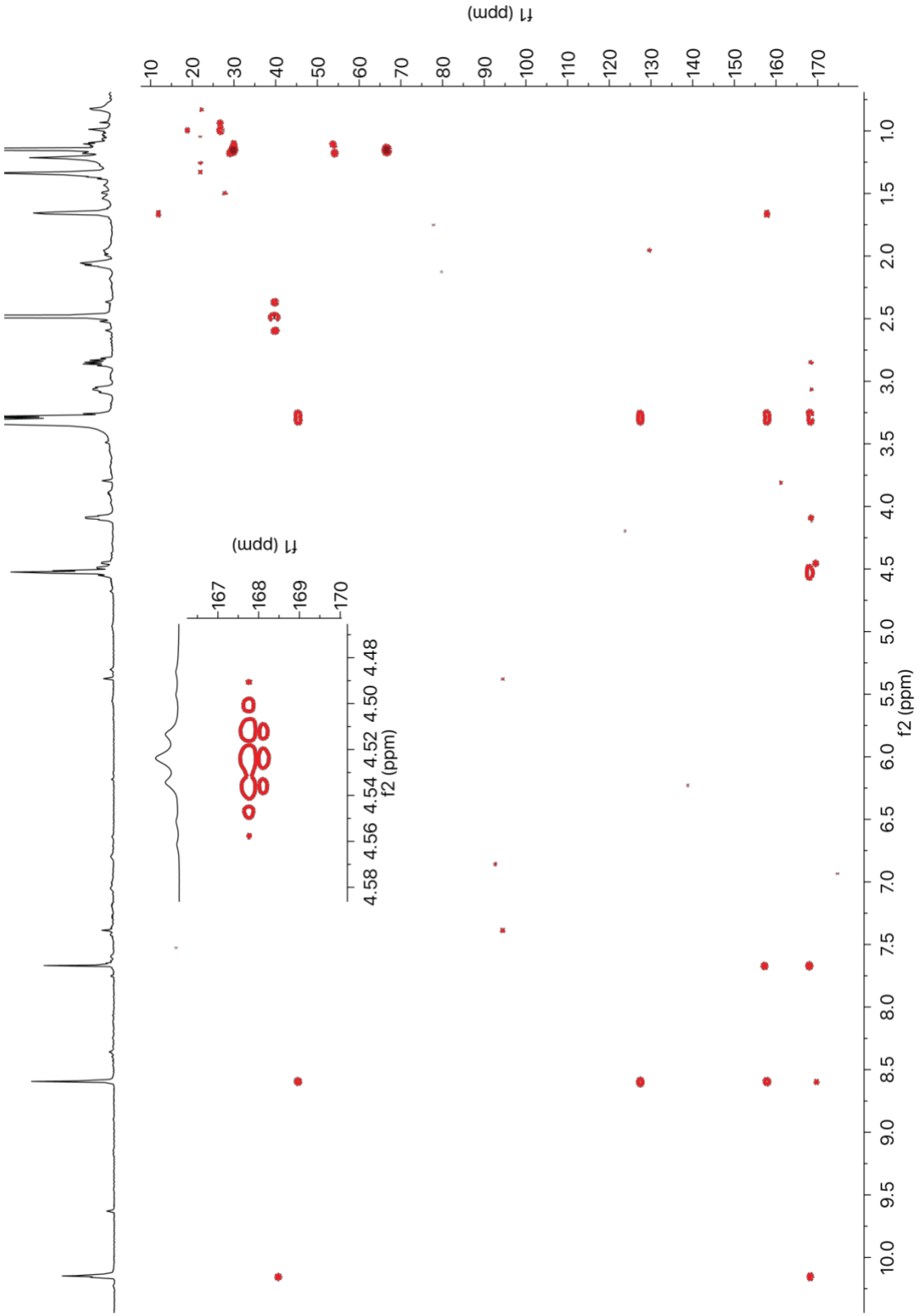


Fig. S28.
gHMBCAD spectrum of compound **3** (recorded in DMSO-d_6 at 600 MHz).

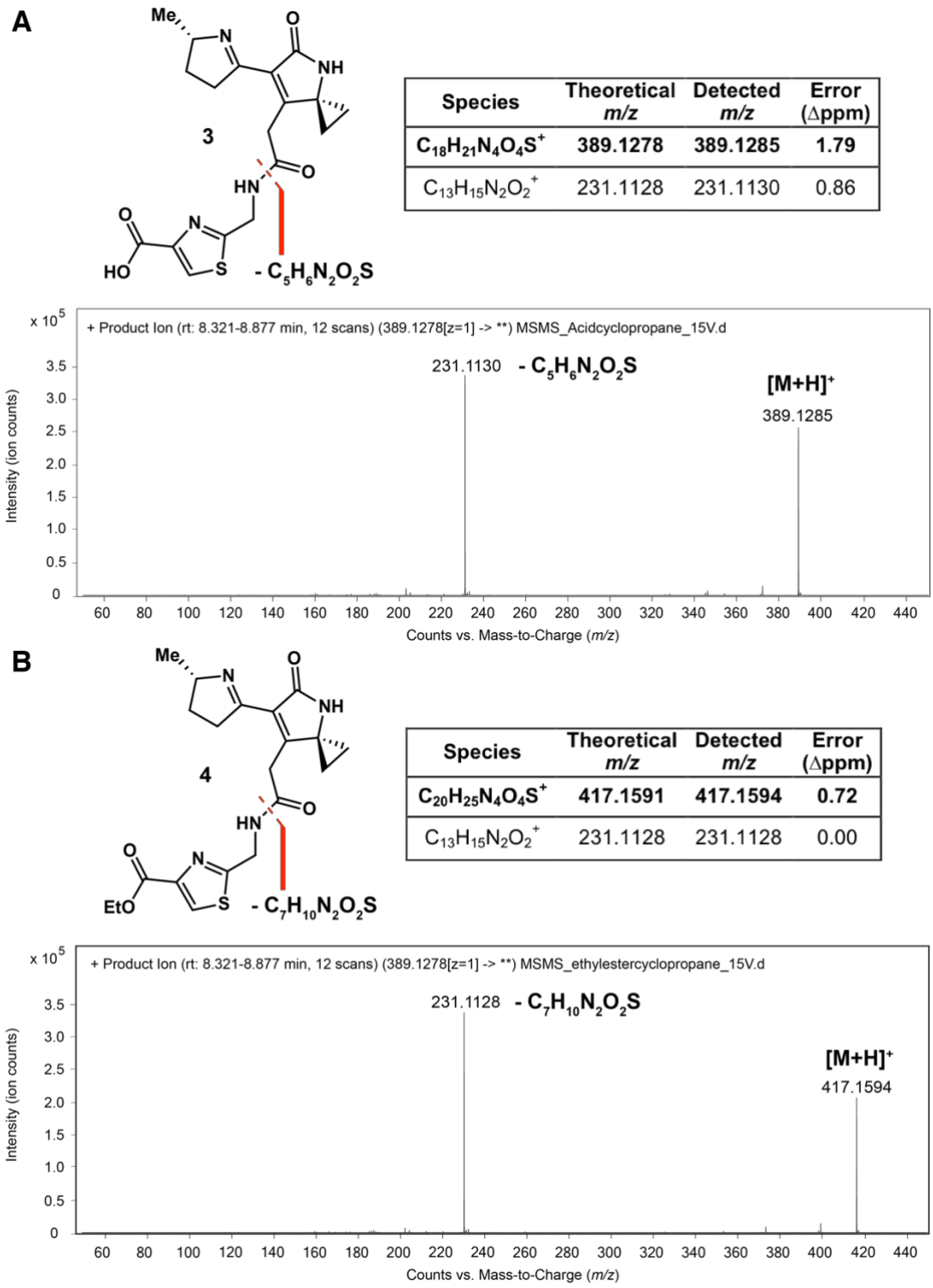


Fig. S29.

High-resolution LC-MS^2 fragmentation of synthetic cyclopropanes 3 and 4. (A) MS^2 fragmentation pattern of cyclopropane **3** (*m/z* 389.1278); fragmentation was acquired with a collision energy of 15 eV. (B) MS^2 fragmentation pattern of cyclopropane **4** (*m/z* 417.1591); fragmentation acquired with a collision energy of 15 eV.

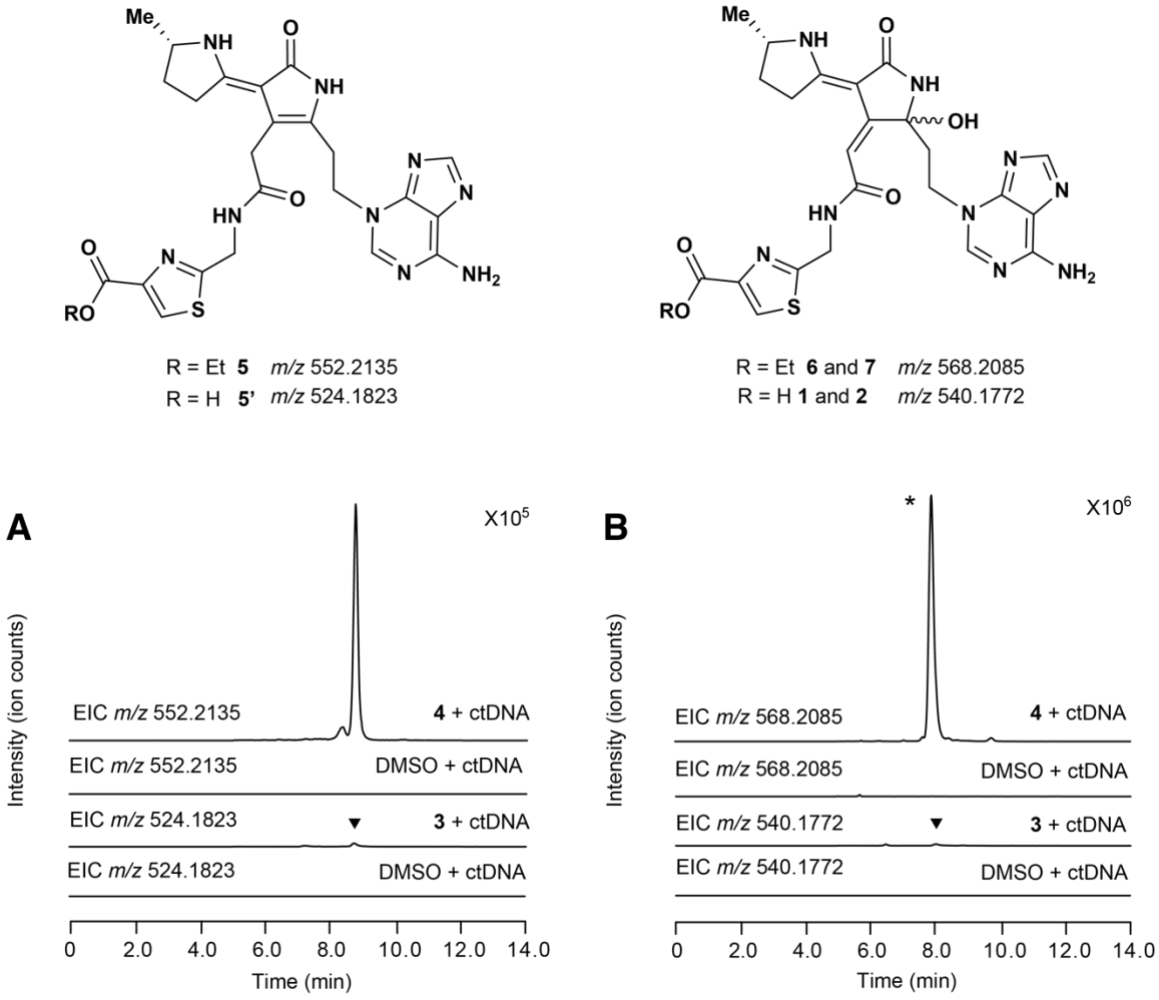


Fig. S30.

In vitro **DNA alkylation reaction between calf thymus DNA (ctDNA) and cyclopropanes 3 and 4.** (A) Extracted ion chromatograms for DNA adducts **5** (*m/z* 552.2135) and **5'** (*m/z* 524.1823). (B) Extracted ion chromatograms for diastereomeric DNA adducts **6** and **7** (*m/z* 568.2085) and adducts **1** and **2** (*m/z* 540.1772). A mass window of ± 5 ppm was used to extract ions of interest. * = the observed peak resolves into two peaks on longer HPLC methods (see Fig. S39C).

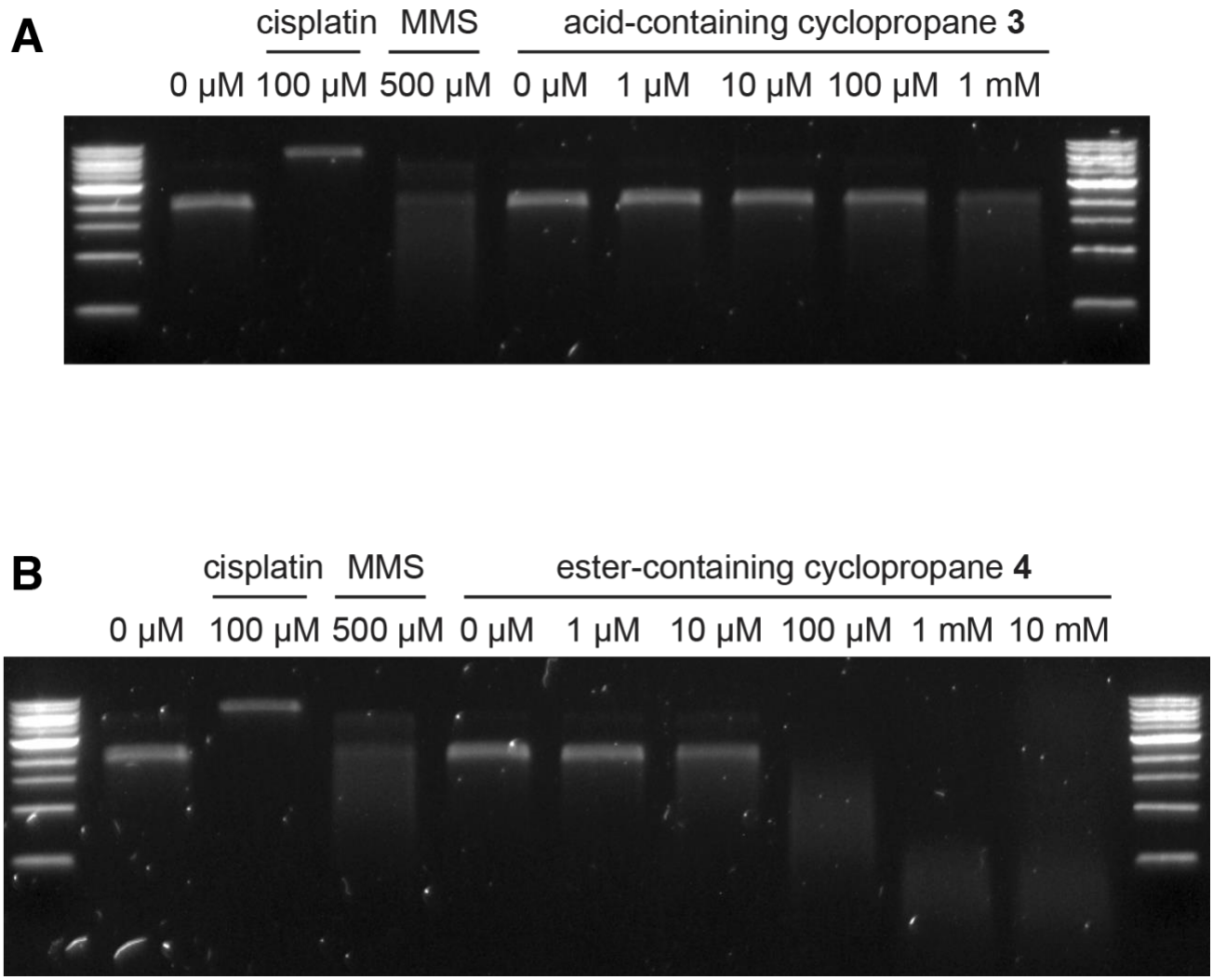


Fig. S31.
DNA shearing assay. (A) DNA shearing assay employing linearized pTrcHis2A plasmid DNA and the synthetic cyclopropane **3**. (B) DNA shearing assay employing linearized pTrcHis2A plasmid DNA and the synthetic cyclopropane **4**. Conditions: linearized pTrcHis2A plasmid DNA (140 ng), various concentrations of **3** (0 μM, 1 μM, 10 μM, 100 μM, and 1 mM) or **4** (0 μM, 1 μM, 10 μM, 100 μM, 1 mM, and 10 mM), 37 ºC, 16 h. Cisplatin (100 μM) and methyl methanesulfonate (MMS, 500 μM) were used as positive controls for DNA cross-linking and alkylation, respectively. DNA was visualized using SYBR Safe.

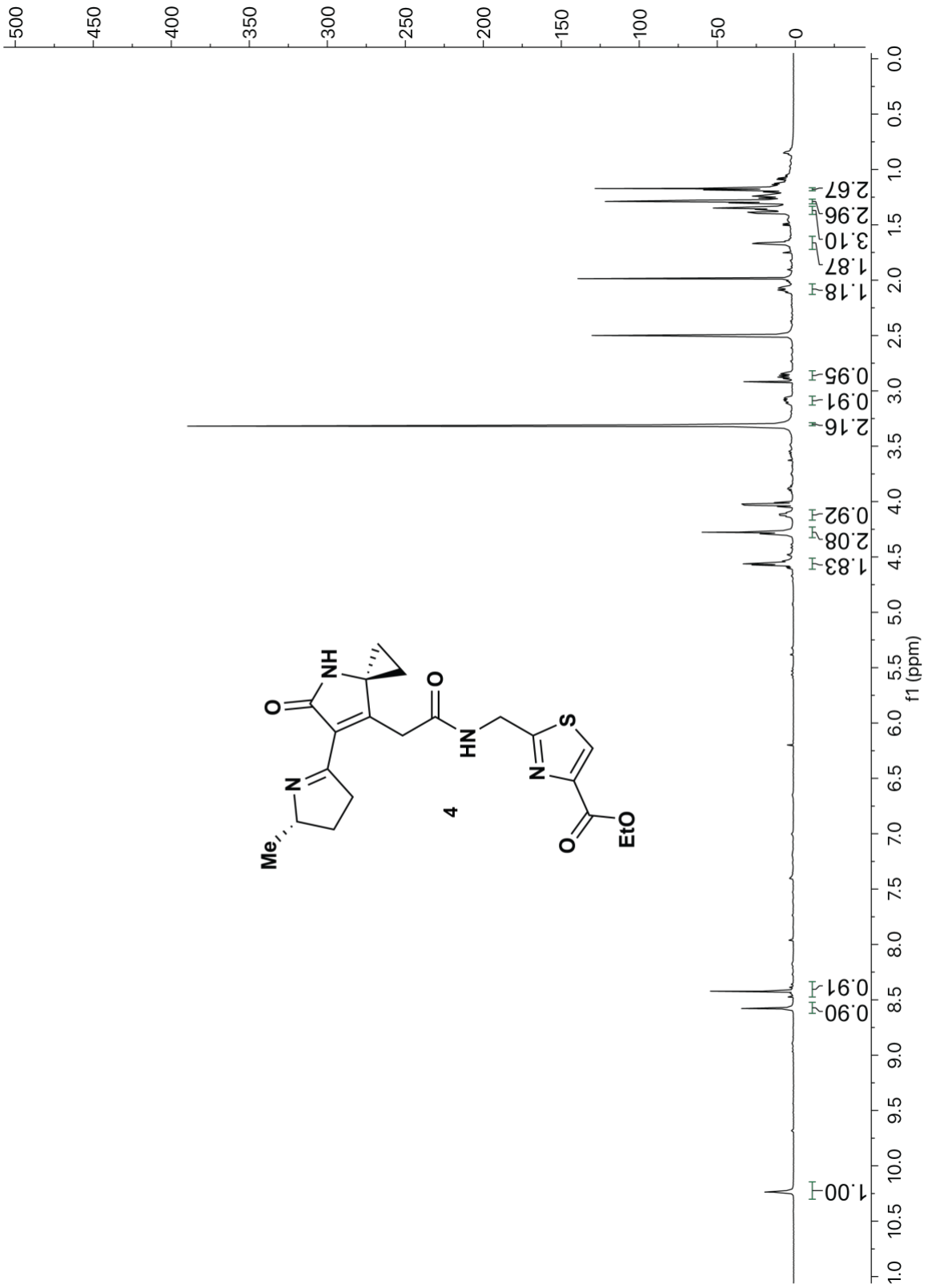


Fig. S32.
^{1}H spectrum of compound **4** (recorded in DMSO-d_6 at 600 MHz).

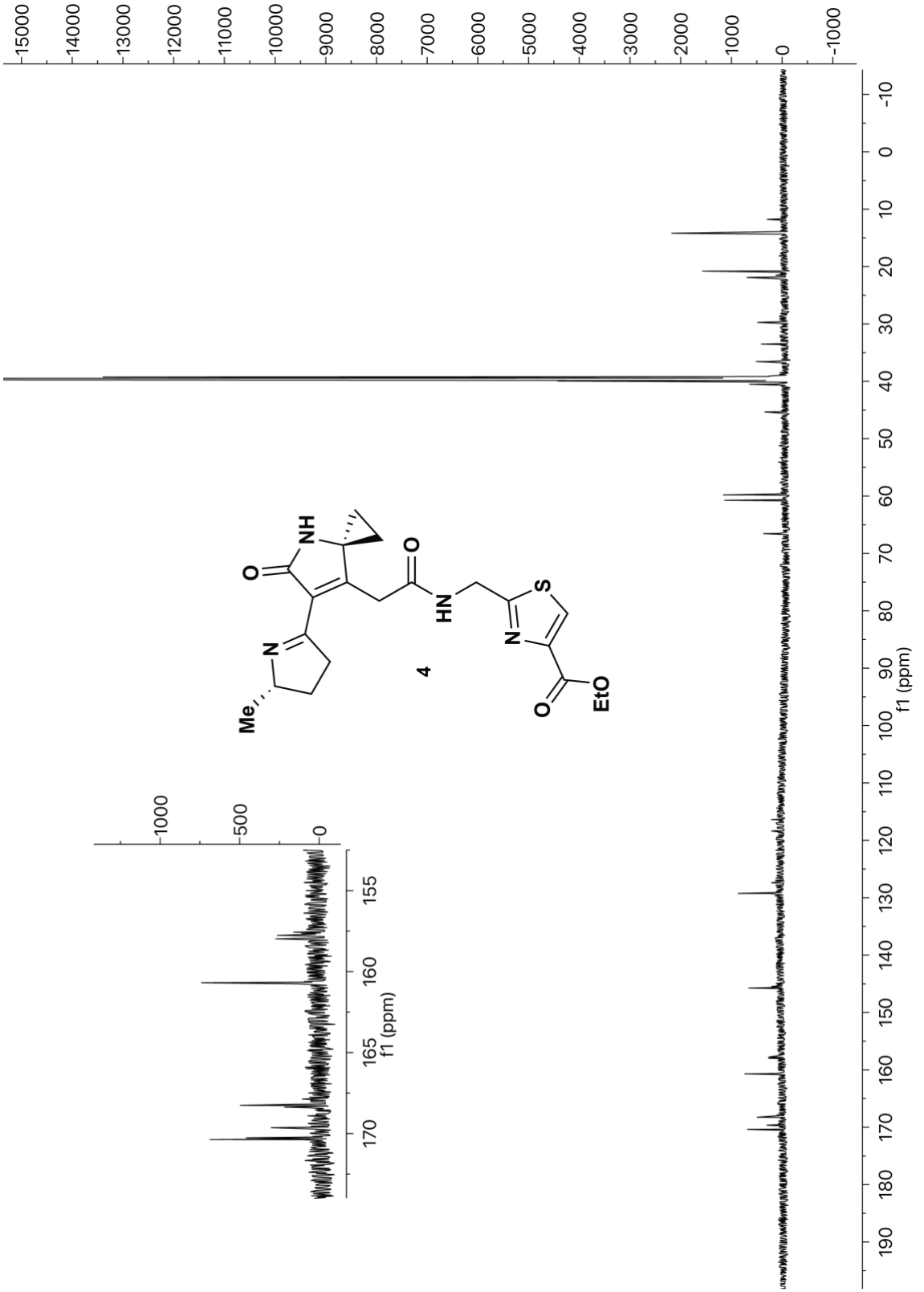


Fig. S33.
^{13}C spectrum of compound **4** (recorded in DMSO-d_6 at 600 MHz).

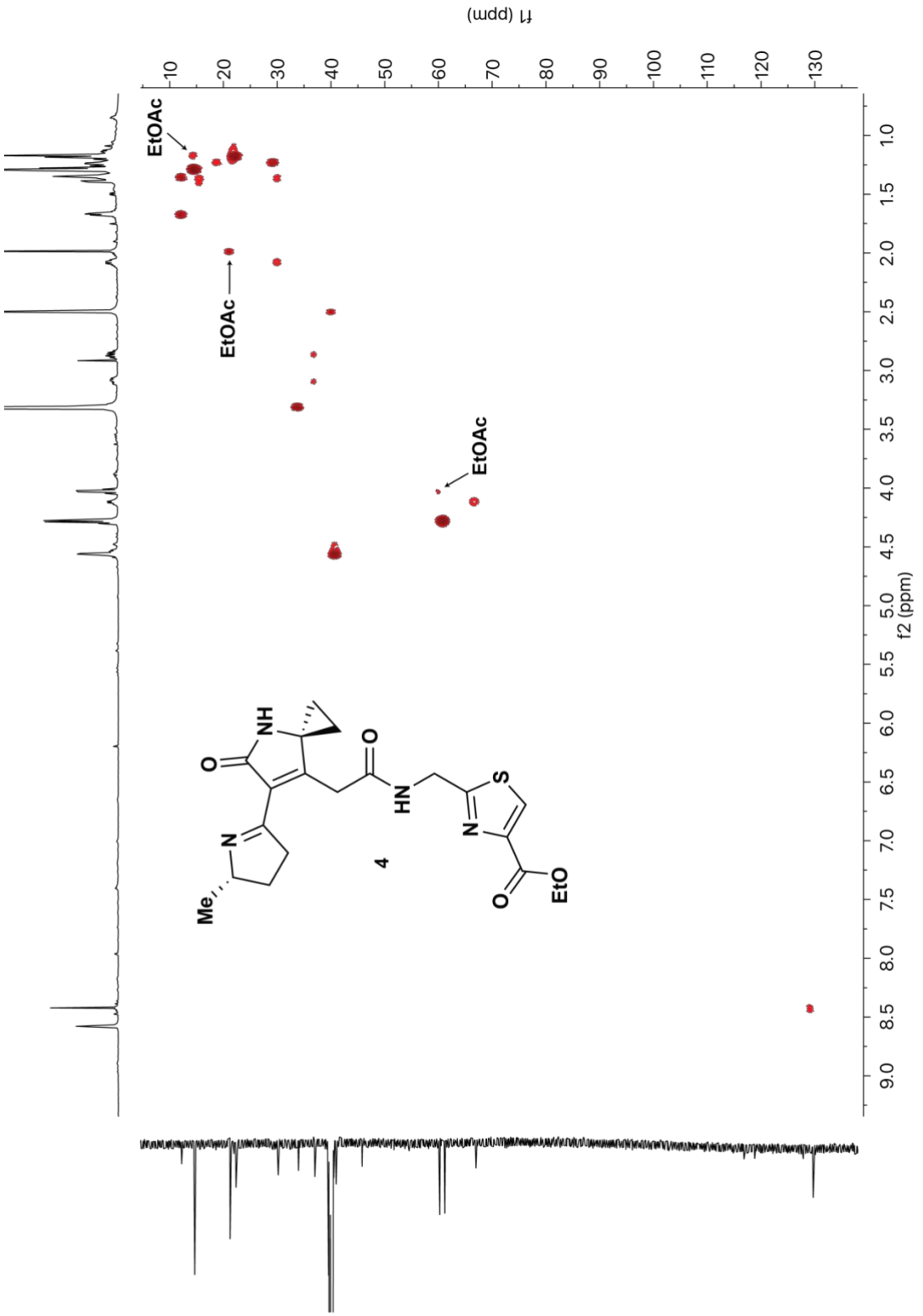


Fig. S34.
gHSQCAD spectrum of compound **4** (recorded in DMSO-d_6 at 600 MHz) (EtOAc = ethyl acetate).

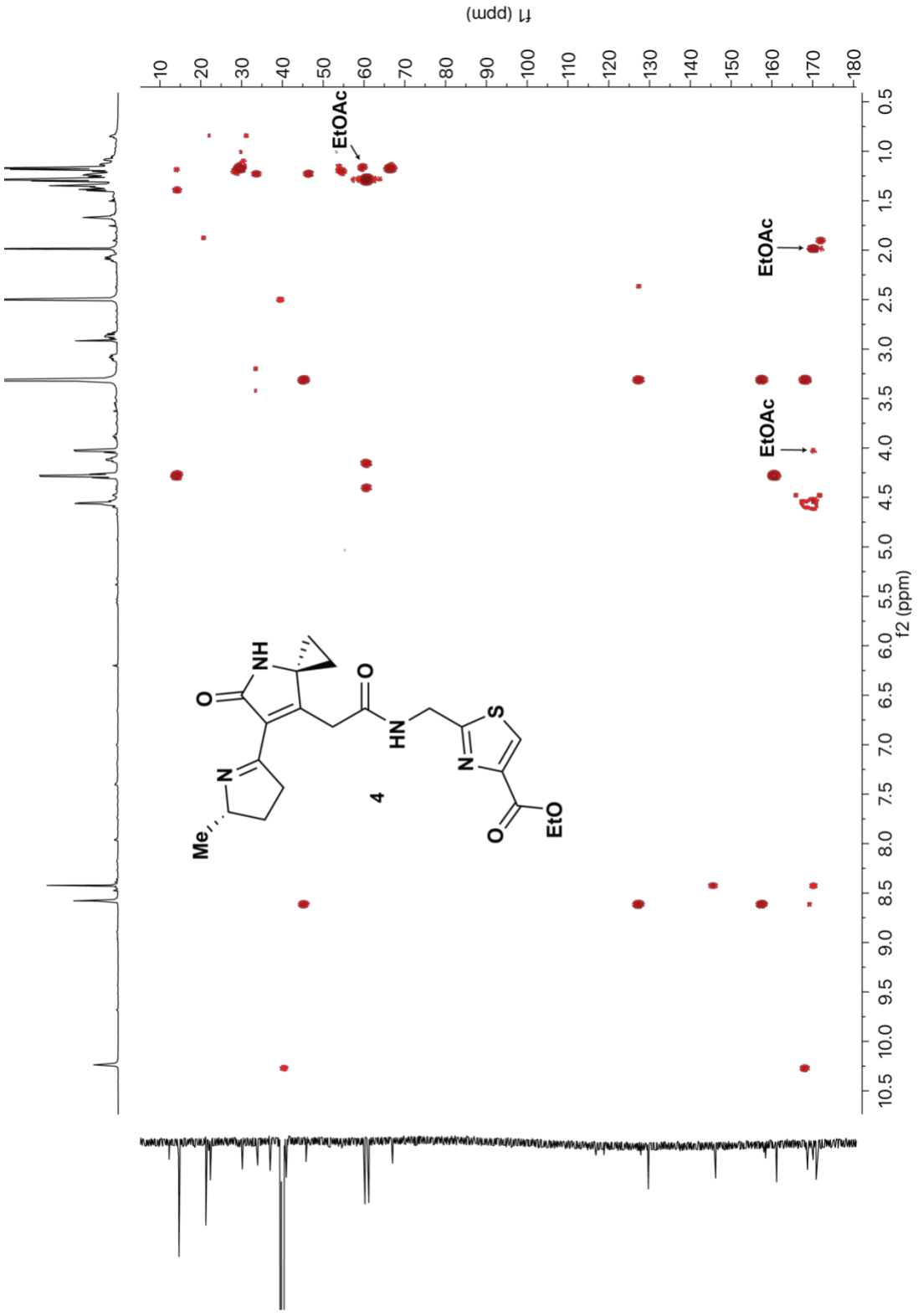


Fig. S35.
gHMBCAD spectrum of compound **4** (recorded in DMSO-d_6 at 600 MHz) (EtOAc = ethyl acetate).

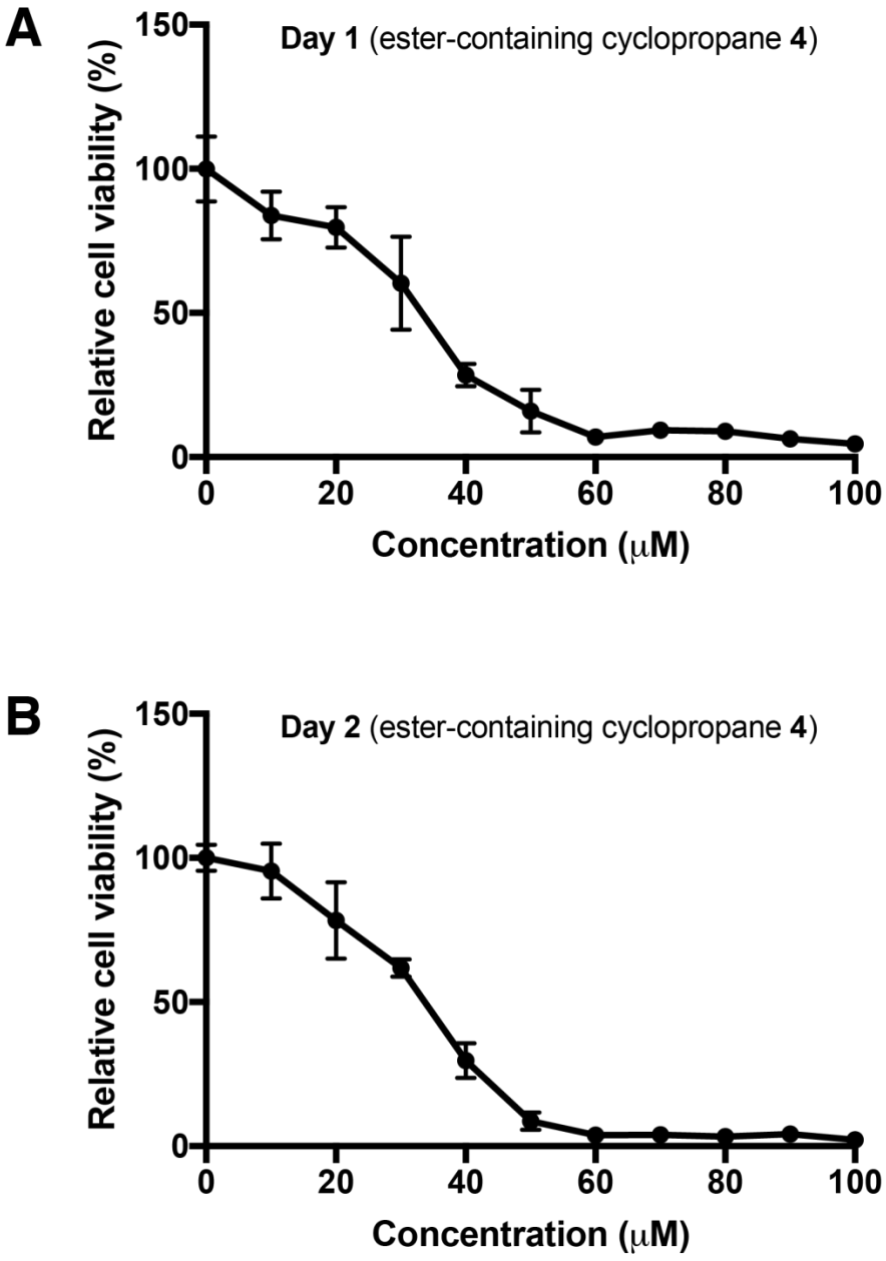


Fig. S36.
Effects of synthetic cyclopropane 4 on HeLa cell viability. MTT assay was used to measure HeLa cell viability after treatment with various concentrations of **4** for (A) 1 day and (B) 2 days.

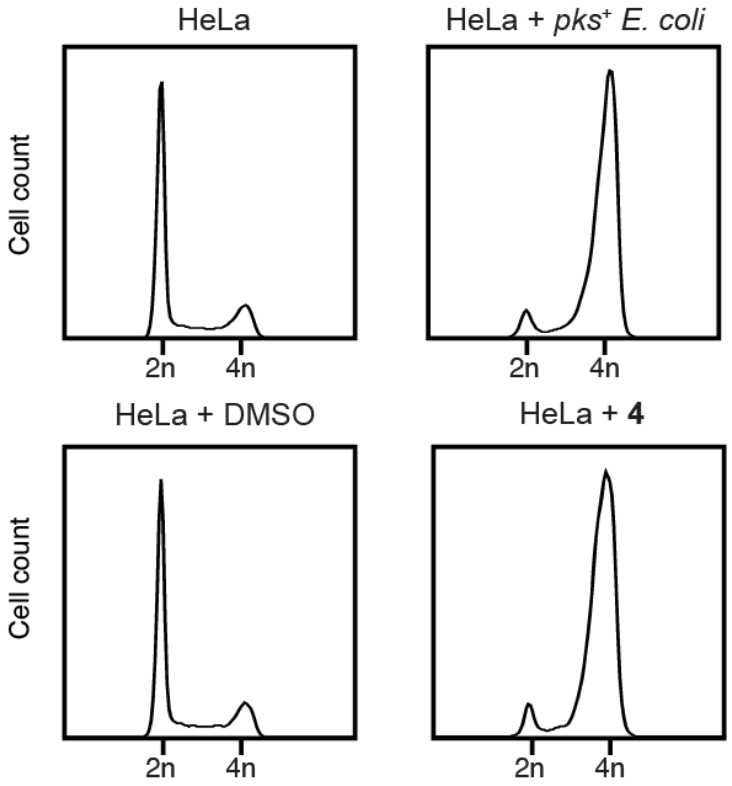


Fig. S37.

***pks*$^+$ *E. coli* infection and compound 4 treatment induced cell cycle arrest**. As reported previously (*2*), HeLa cells were infected with *pks*$^+$ *E. coli* [multiplicity of infection (MOI) = 1,000] for 1 h or were left uninfected. After 1 h, the cells were washed to remove *E. coli* and then cultured in gentamicin containing media for 23 h. HeLa cells were also treated with DMSO or cyclopropane **4** (20 μM) for 24 h. Cell cycle progression was monitored by flow cytometry after infection or compound treatment.

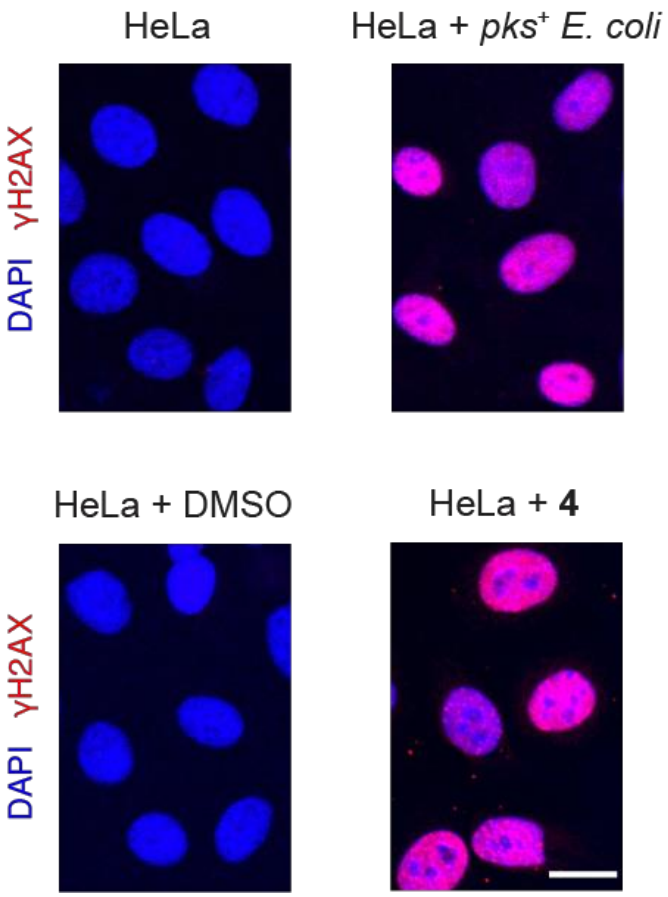


Fig. S38.
***pks*$^+$ *E. coli* infection and compound 4 treatment induced DSBs**. As reported previously (*2*), HeLa cells were infected with *pks*$^+$ *E. coli* [multiplicity of infection (MOI) = 1,000] for 1 h or were left uninfected. After 1 h, the cells were washed to remove *E. coli* and then cultured in gentamicin-containing media for 23 h. HeLa cells were also treated with DMSO or cyclopropane **4** (20 μM) for 24 h. Cells were examined by confocal microscopy for DNA (blue) and phosphorylated H2AX (γH2AX, red) (Scale bar, 20 μm).

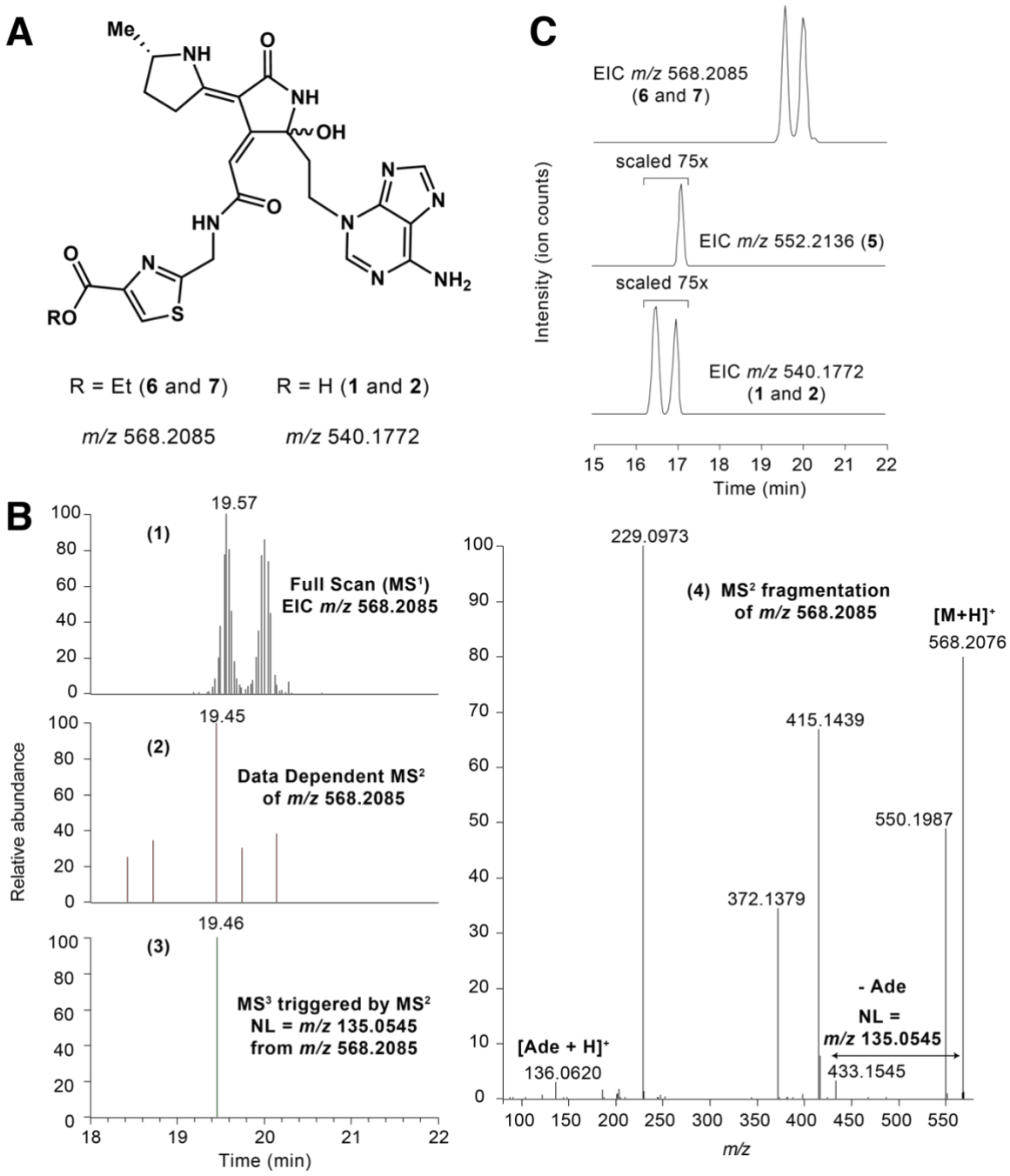
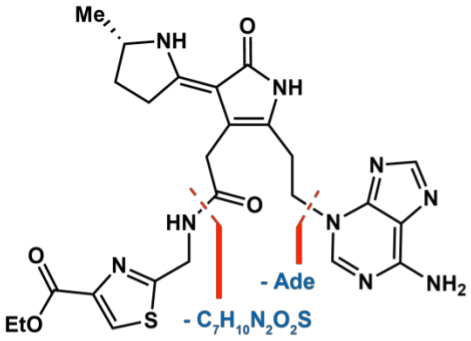


Fig. S39.

High-resolution accurate mass (HRAM) LC-MS^3 DNA adductomic analysis identified DNA adducts in 'colibactin mimic'-treated HeLa cells. (A) Chemical structures of DNA adducts **6** and **7** and the observed *in vivo* adducts **1** and **2**. (B) 1. Full scan extracted ion chromatogram (EIC) of adducts **6** and **7** (*m/z* 568.2085). 2. Signal corresponding to the data dependent MS^2 event (RT = 19.45 min). 3. Signal corresponding to MS^3 event (RT = 19.46 min) triggered by the neutral loss of adenine. 4. MS^2 mass spectrum resulting from fragmentation of *m/z* 568.2085 which triggered an MS^3 event. (C) EIC chromatograms of DNA adducts **6** and **7**, **5**, and **1** and **2** in HeLa cells treated with synthetic cyclopropane **4**.



Species	Theoretical *m/z*	Detected *m/z*	Error (Δppm)
$C_{25}H_{30}N_9O_4S^+$	**552.2136**	**552.2125**	**1.99**
$C_{20}H_{25}N_4O_4S^+$	417.1591	417.1588	0.72
$C_{13}H_{15}N_2O_2^+$	231.1128	231.1127	0.43

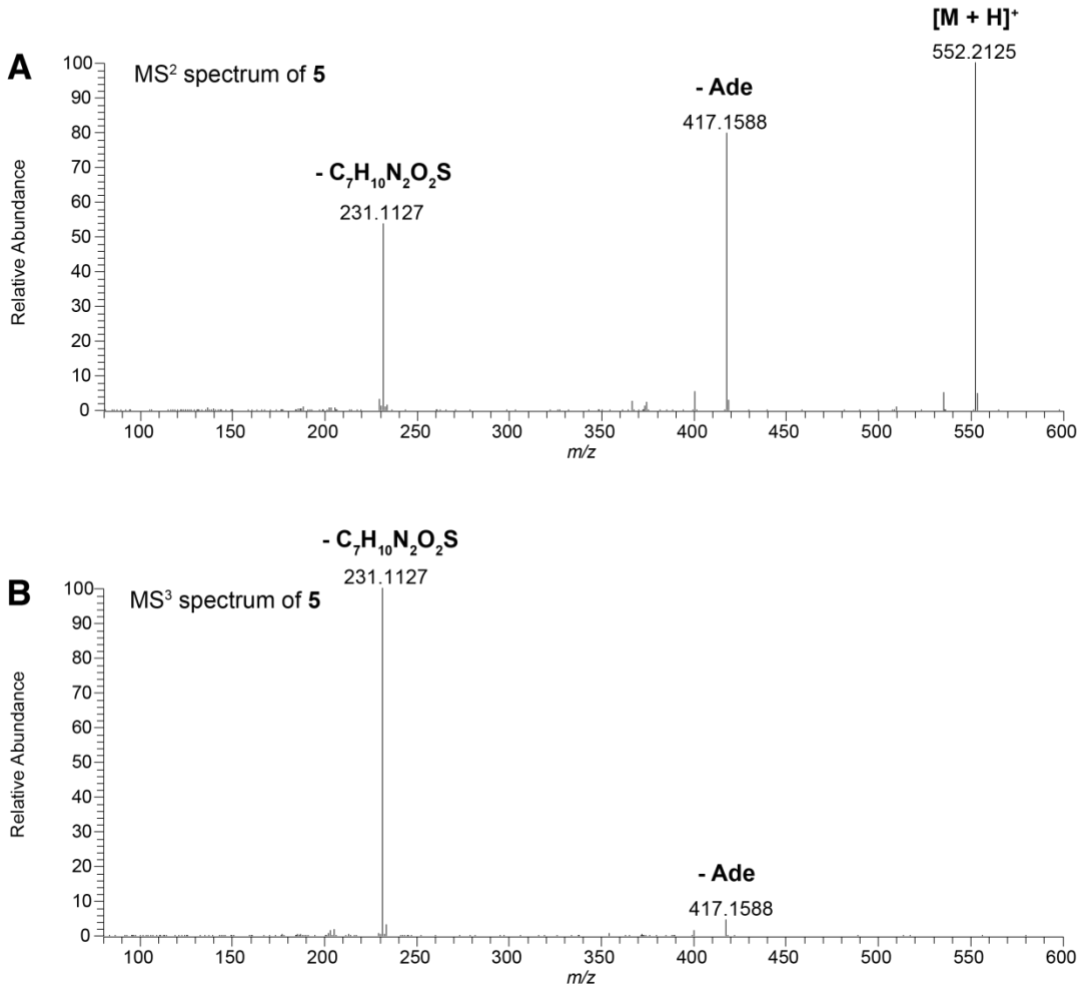
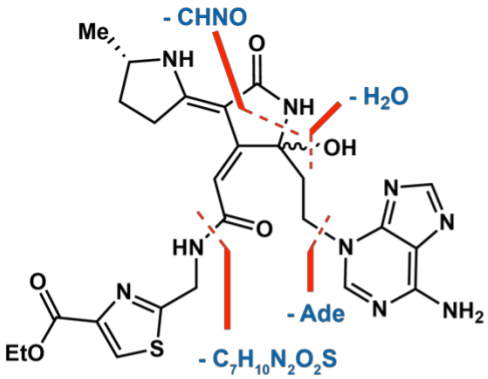
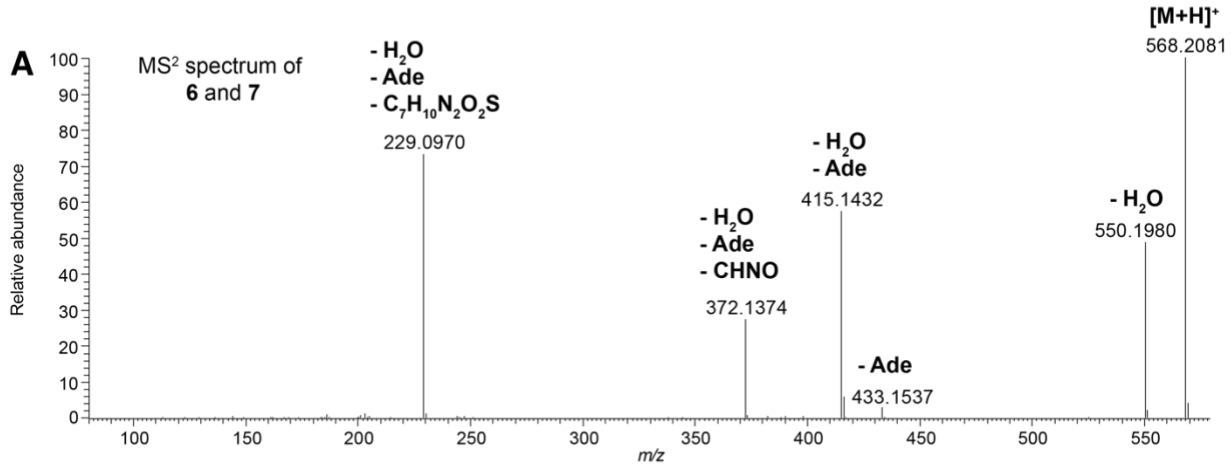


Fig. S40.
MS^2 and MS^3 fragmentation of 5 (*m/z* 552.2136), a DNA adduct formed in an *in vitro* DNA alkylation reaction between synthetic cyclopropane 4 and calf thymus DNA. (A) MS^2 fragmentation spectrum of **5** (*m/z* 552.2136); fragmentation acquired with an HCD collision energy of 15%. (B) MS^3 fragmentation spectrum of **5** (*m/z* 417.1591); fragmentation acquired with an HCD collision energy of 30%.



Species	Theoretical *m/z*	Detected *m/z*	Error (Δppm)
$C_{25}H_{30}N_9O_5S^+$	**568.2085**	**568.2081**	**0.70**
$C_{25}H_{28}N_9O_4S^+$	550.1979	550.1980	0.18
$C_{20}H_{25}N_4O_5S^+$	433.1540	433.1537	0.69
$C_{20}H_{23}N_4O_4S^+$	415.1435	415.1432	0.72
$C_{19}H_{22}N_3O_3S^+$	372.1382	372.1374	2.14
$C_{13}H_{15}N_2O_3^+$	247.1077	247.1073	1.61
$C_{13}H_{13}N_2O_2^+$	229.0972	229.0970	0.87



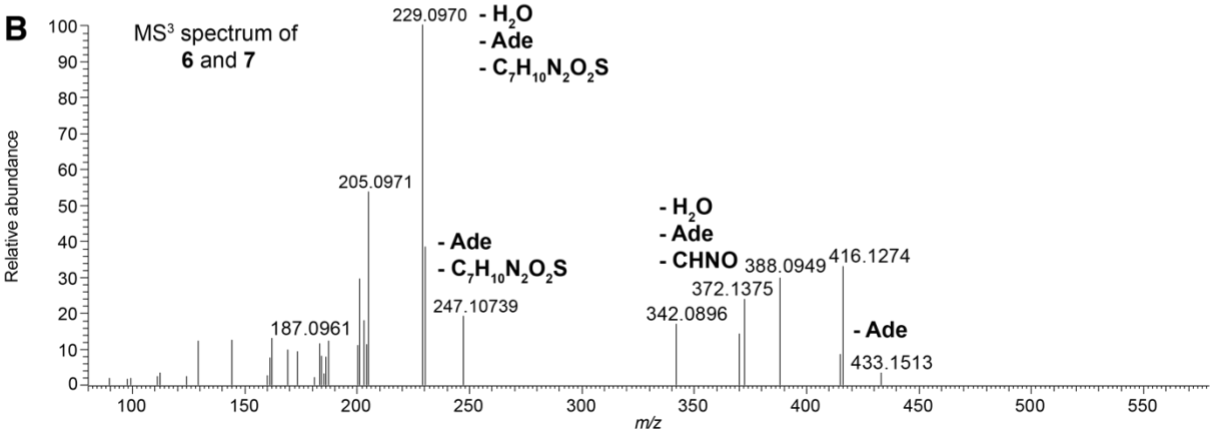
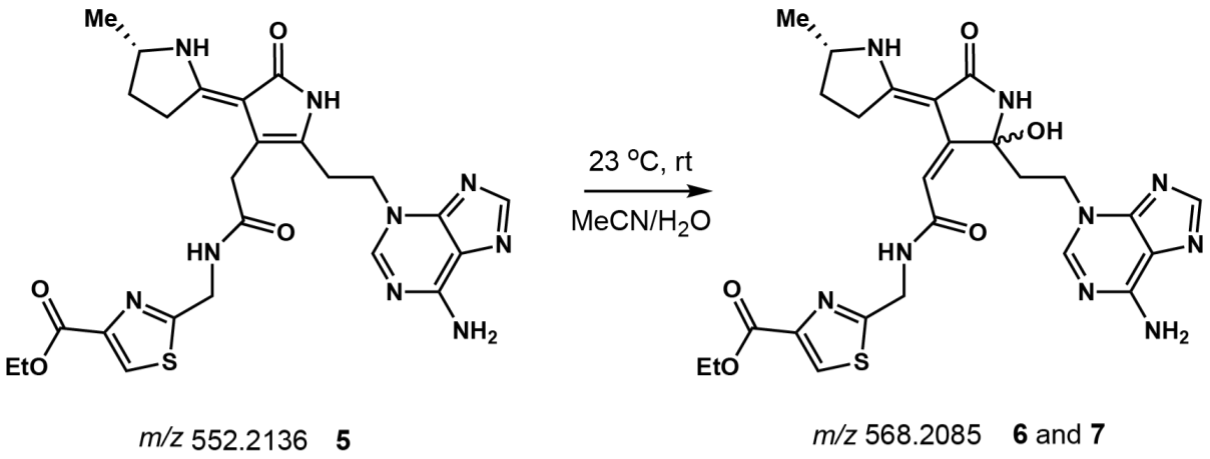


Fig. S41.
High-resolution LC-MS^2 and MS^3 fragmentation of 6 and 7 (*m/z* 568.2085), DNA adducts formed in an *in vitro* DNA alkylation reaction between synthetic cyclopropane 4 and calf thymus DNA. (A) MS^2 fragmentation spectrum of **6** and **7** (*m/z* 568.2085); fragmentation acquired with an HCD collision energy of 15%. (B) MS^3 fragmentation spectrum of **6** and **7** (*m/z* 433.1540); fragmentation acquired with an HCD collision energy of 30%.



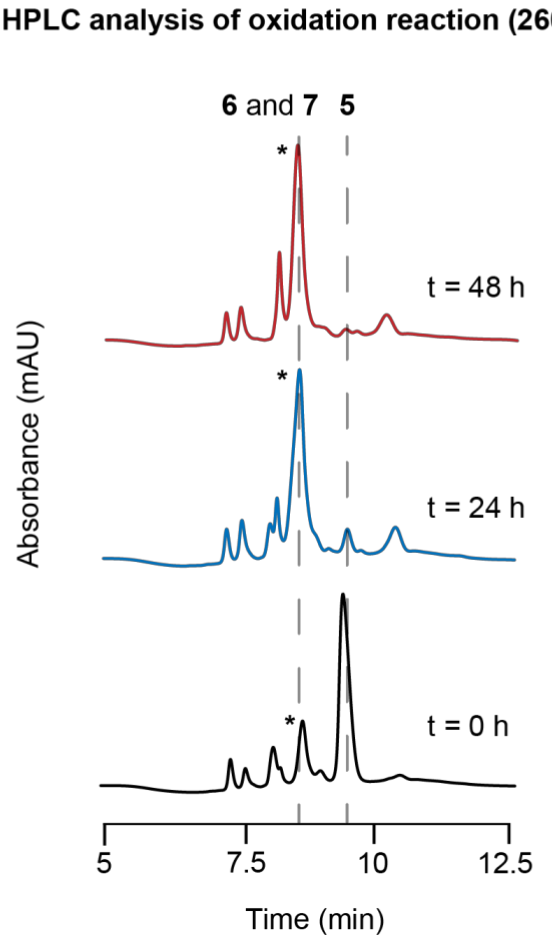


Fig. S42.
HPLC analysis of the oxidation reaction. HPLC traces monitoring the conversion of unstable adduct 5 to the diastereomeric adducts 6 and 7. * = the observed peaks resolve into two peaks on longer HPLC methods (see Fig. S39C).

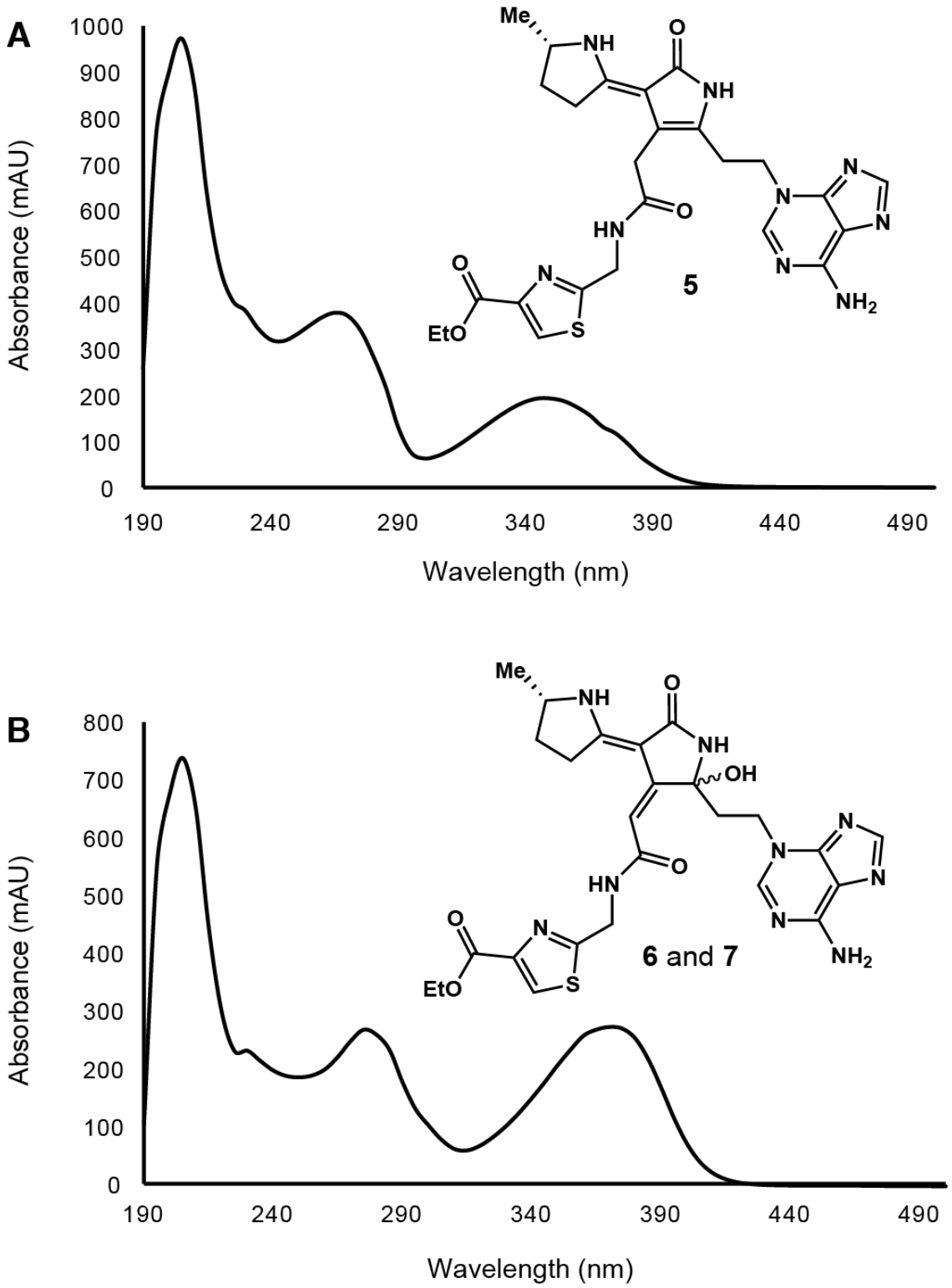


Fig. S43.

Comparison of the UV spectrums of DNA adduct 5 and the diastereomeric adducts 6 and 7. (A) UV spectrum of compound **5** (B) UV spectrum of diastereomeric adducts **6** and **7**.

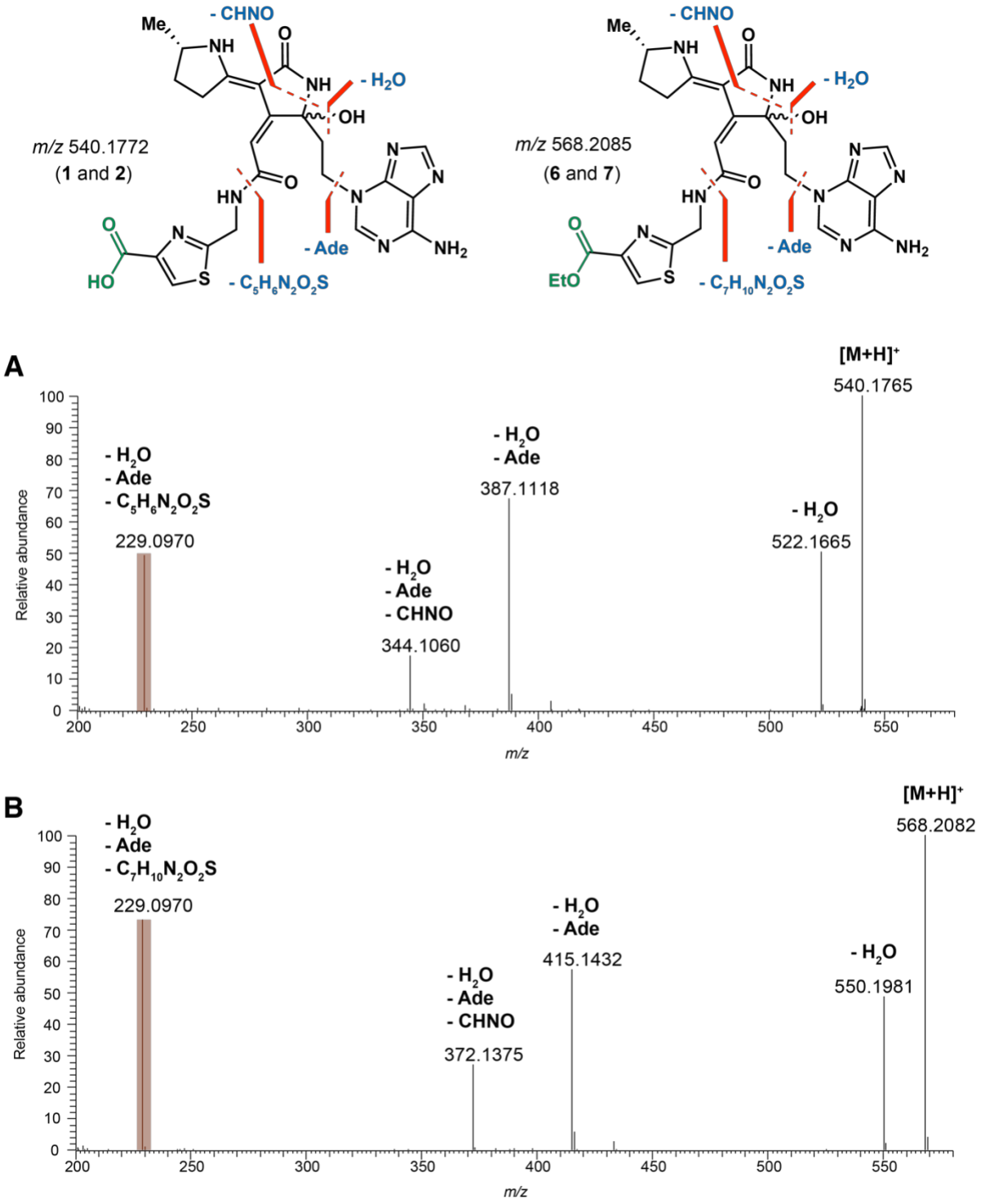


Fig. S44.
High-resolution MS^2 fragmentation of DNA adducts 1 and 2 (*m/z* 540.1772) and 6 and 7 (*m/z* 568.2085) highlighting the shared fragment ion (*m/z* 229.0972). (A) MS^2 fragmentation spectrum of **1** and **2** (*m/z* 540.1772); fragmentation acquired with an HCD collision energy of 15%. (B) MS^2 fragmentation spectrum of **6** and **7** (*m/z* 568.2085); fragmentation acquired with an HCD collision energy of 15%.

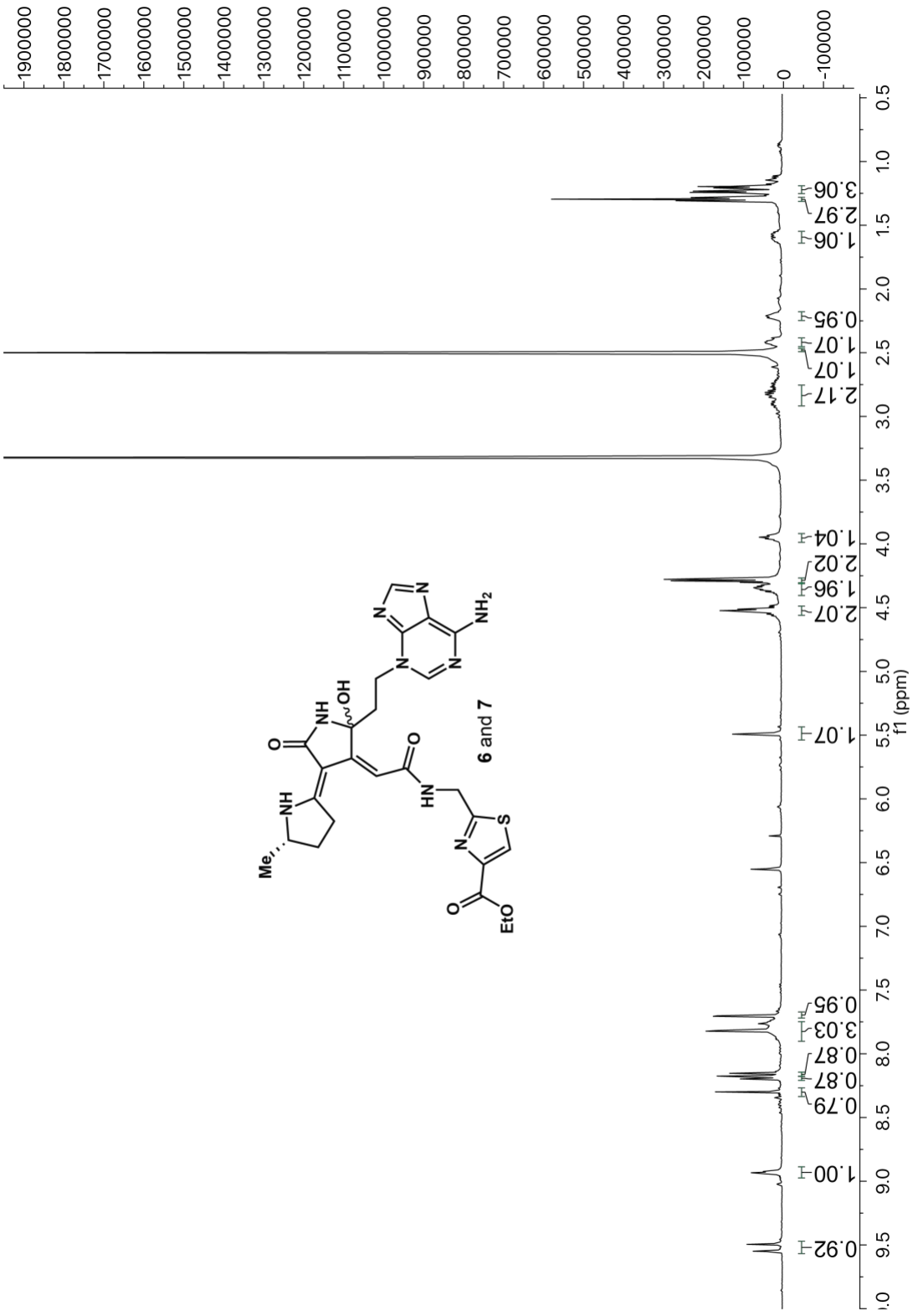


Fig. S45.
^{1}H spectrum of diastereomeric DNA adducts **6** and **7** (recorded in DMSO-*d*$_6$ at 600 MHz).

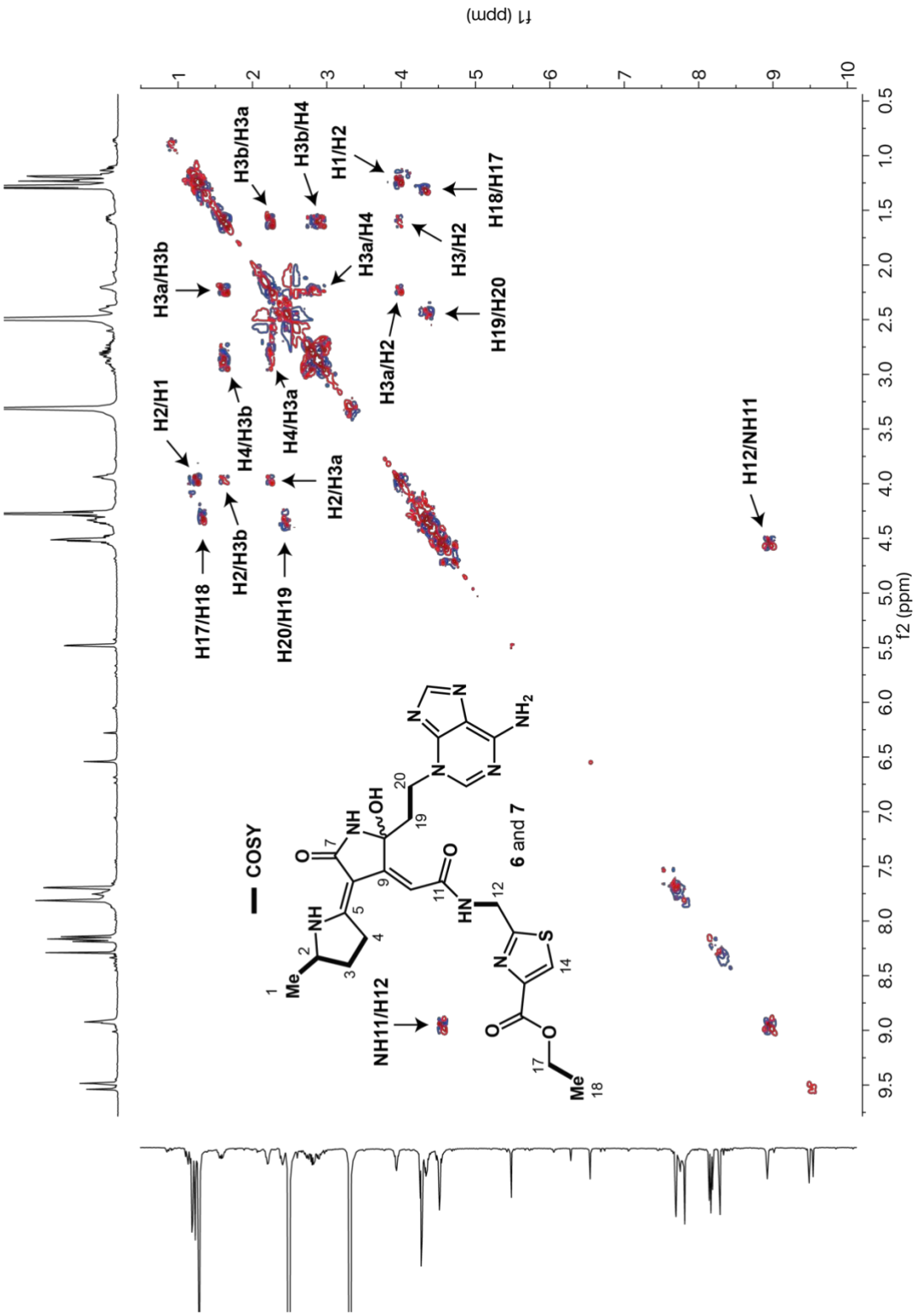


Fig. S46.
gCOSY spectrum of diastereomeric adducts **6** and **7** (recorded in DMSO-d_6 at 600 MHz). ^{1}H-^{1}H correlations are highlighted in structure (bolded bonds) and spectra.

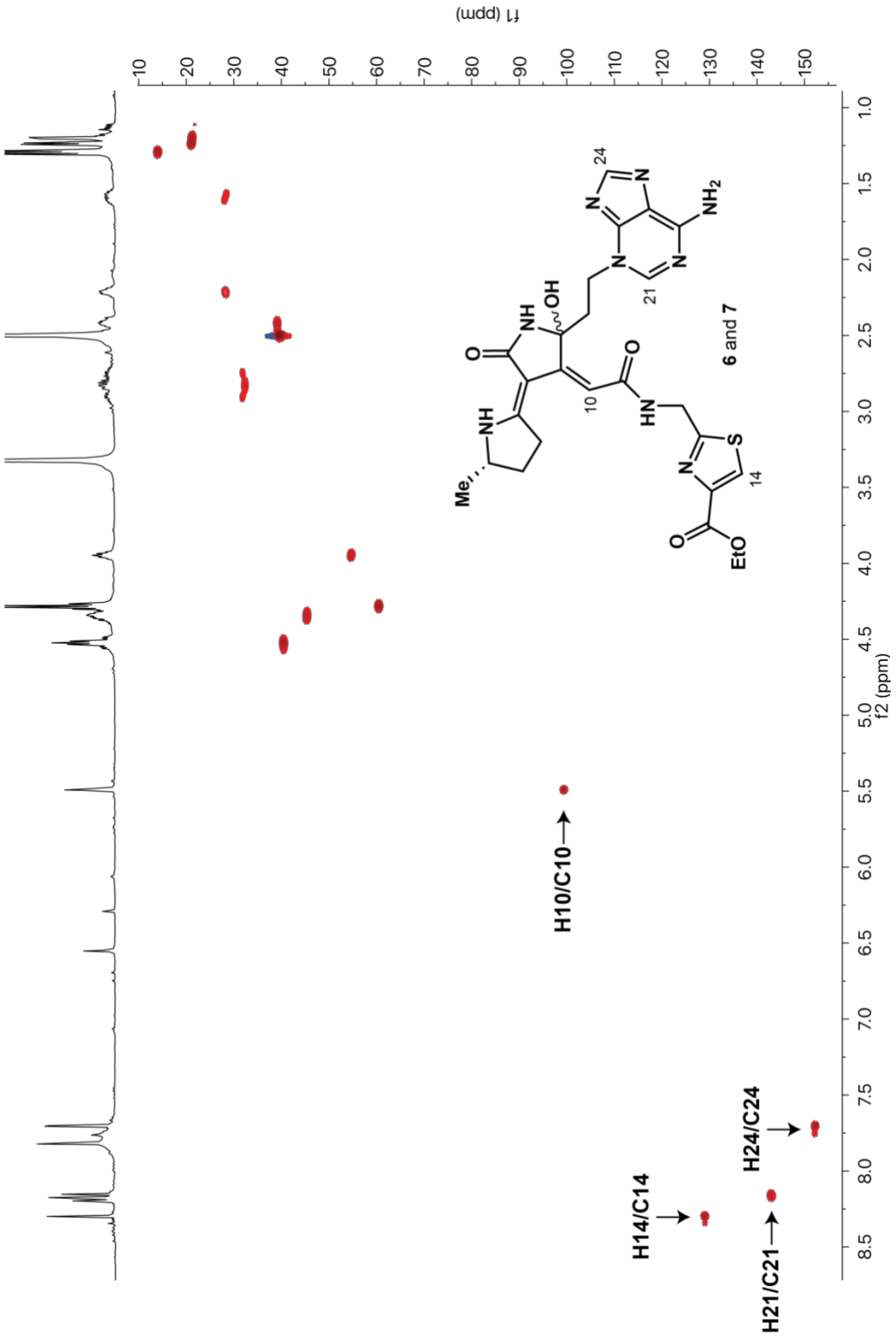


Fig. S47.
gHSQCAD spectrum of diastereomeric adducts **6** and **7** (recorded in DMSO-*d*$_6$ at 600 MHz). Key ^{1}H-^{13}C correlations are highlighted in the spectra.

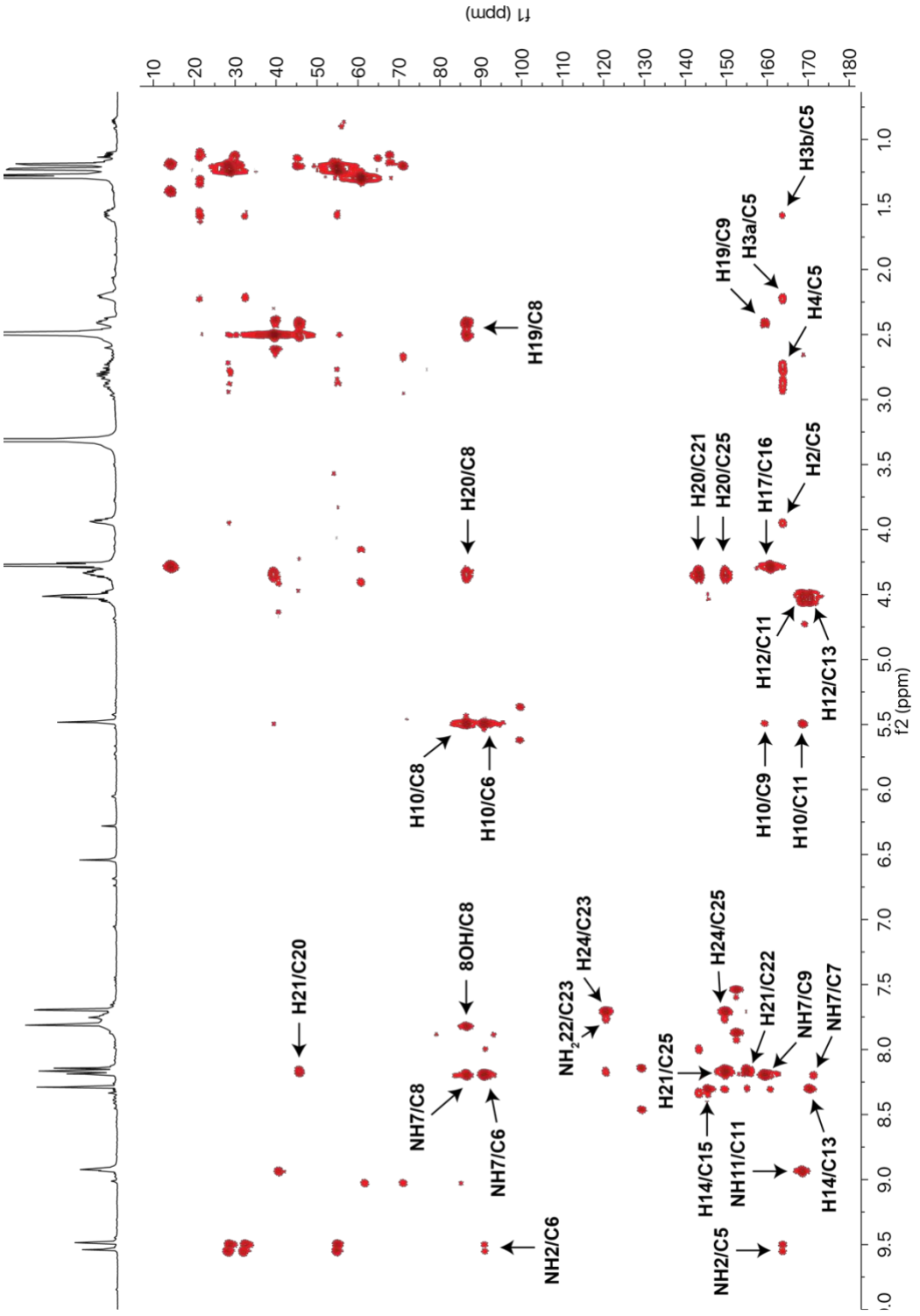


Fig. S48.
gHMBCAD spectrum of diastereomeric adducts **6** and **7** (recorded in DMSO-d_6 at 600 MHz). Key HMBC correlations are highlighted in the spectra.

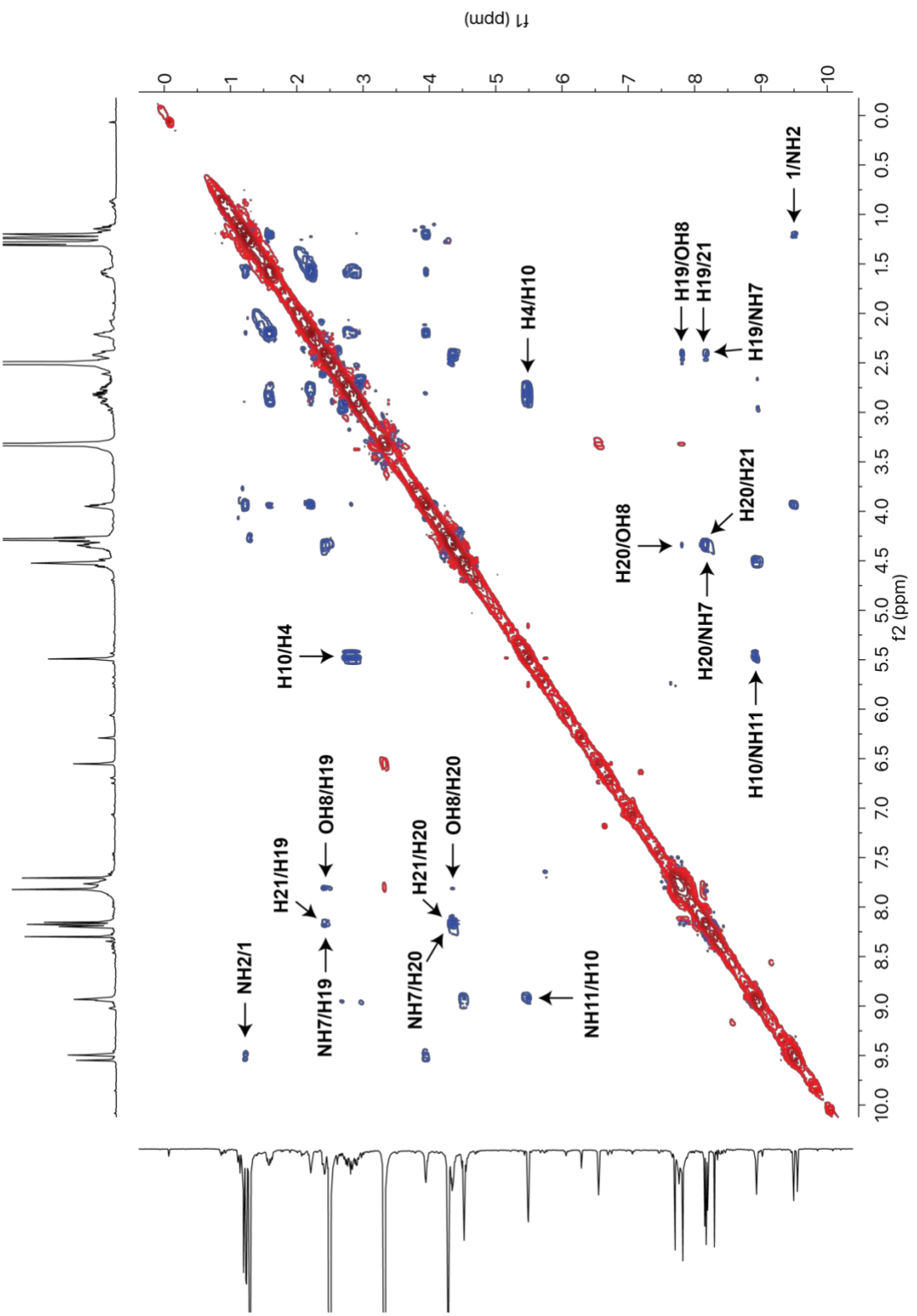


Fig. S49.
ROESYAD spectrum of diastereomeric adducts **6** and **7** (recorded in DMSO-d_6 at 600 MHz). Key through-space ^{1}H-^{1}H correlations are highlighted in the spectra.

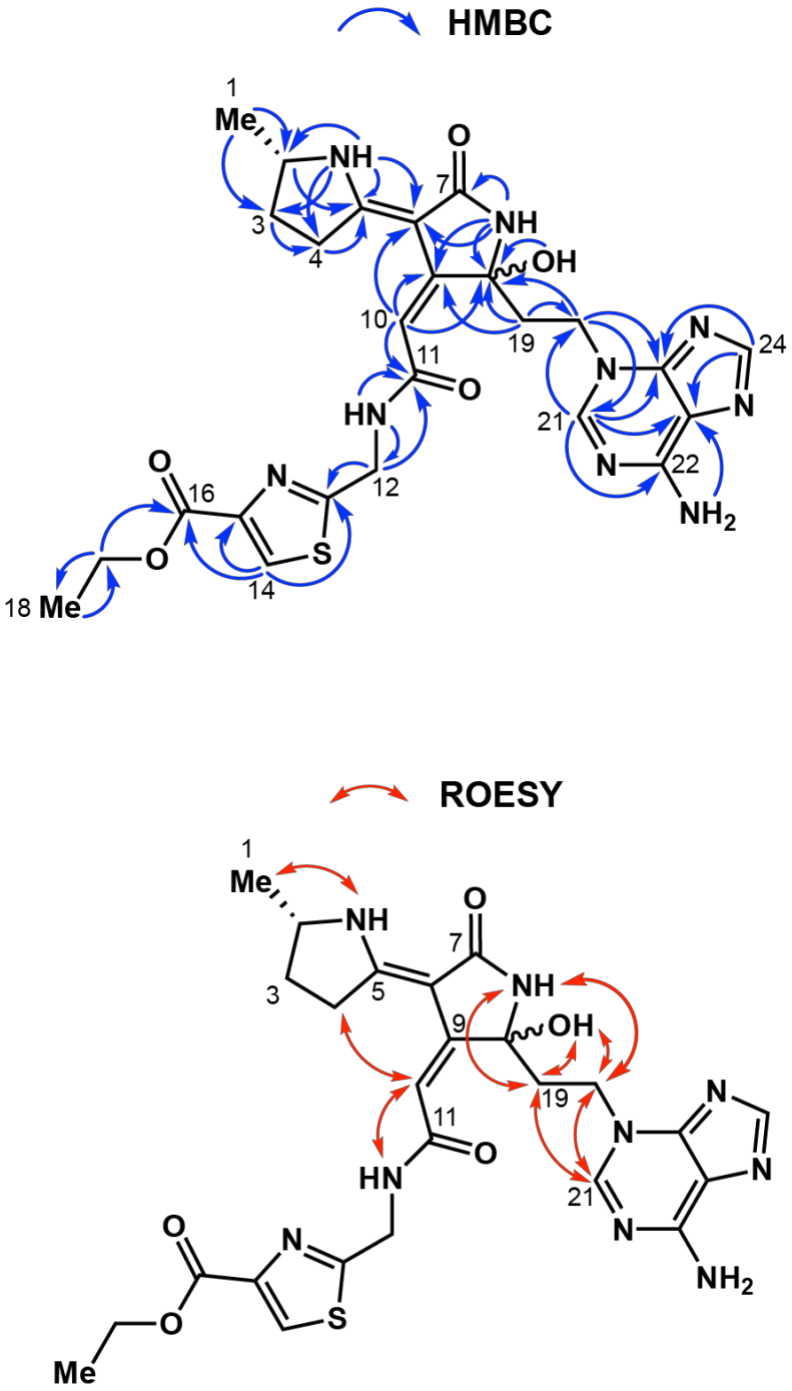


Fig. S50.
Key HMBC and ROESY correlations used to structurally assign diastereomeric adducts **6** and **7**.

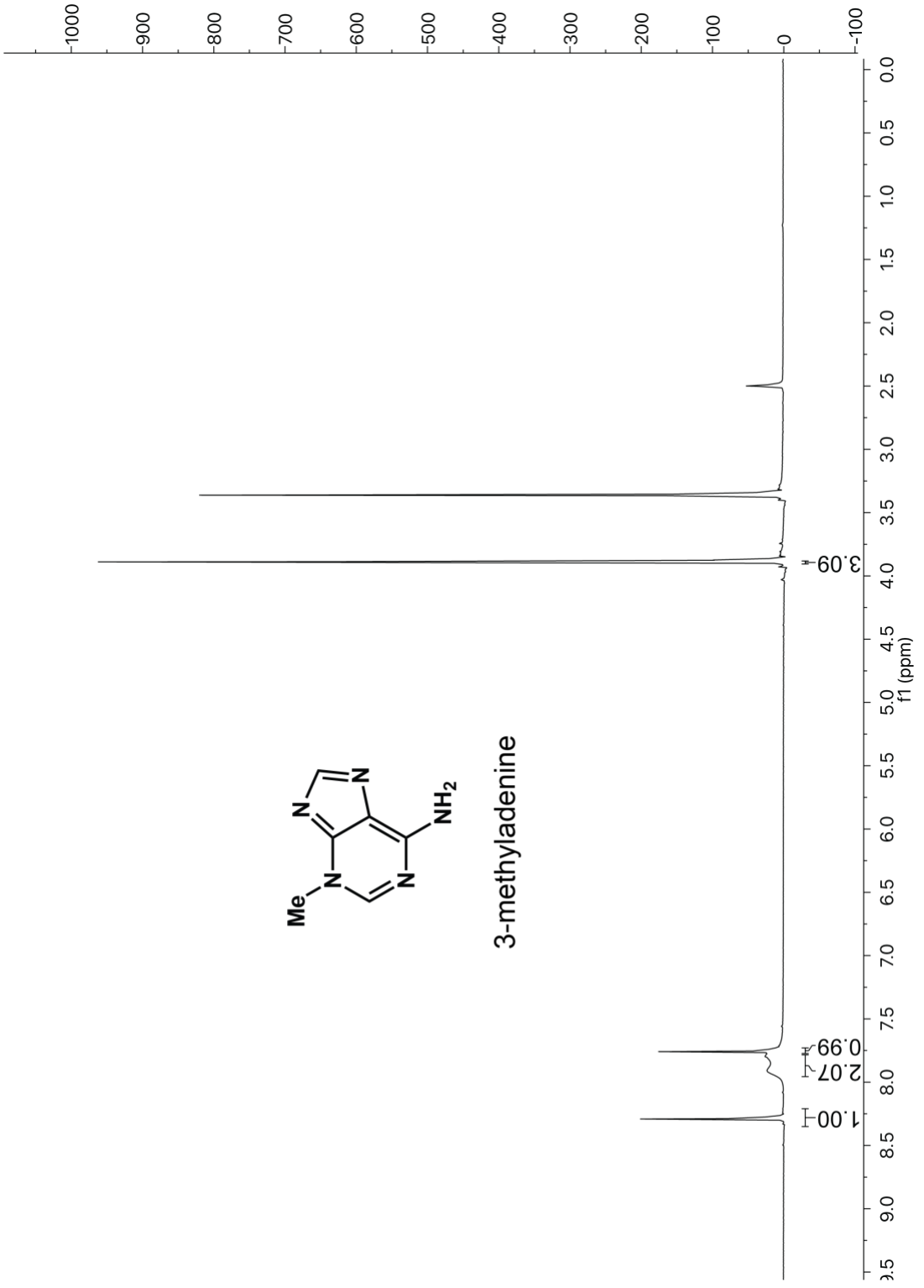


Fig. S51.
^{1}H spectrum of 3-methyladenine (recorded in DMSO-d_6 at 600 MHz).

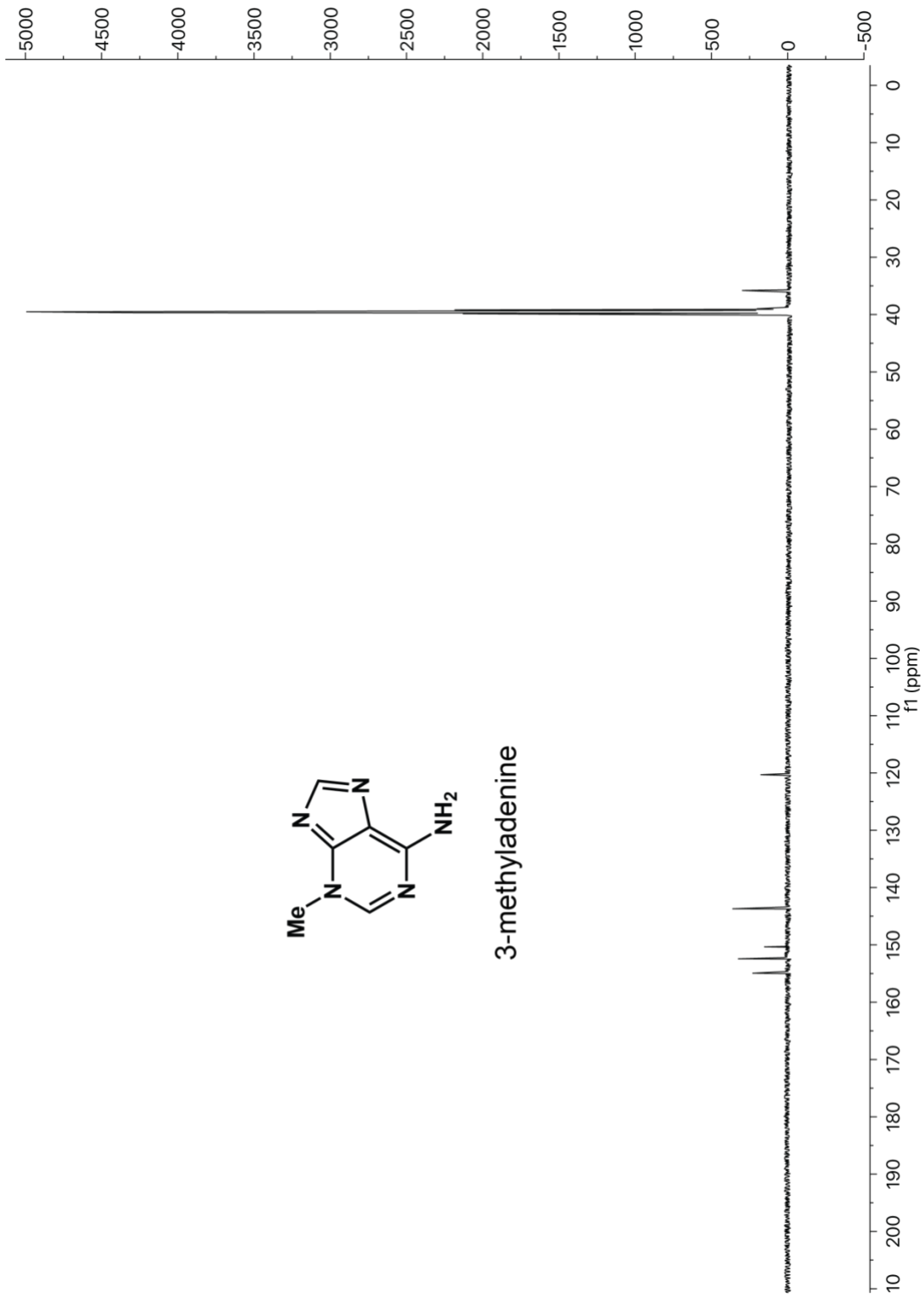


Fig. S52.

^{13}C spectrum of 3-methyladenine (recorded in DMSO-d_6 at 600 MHz).

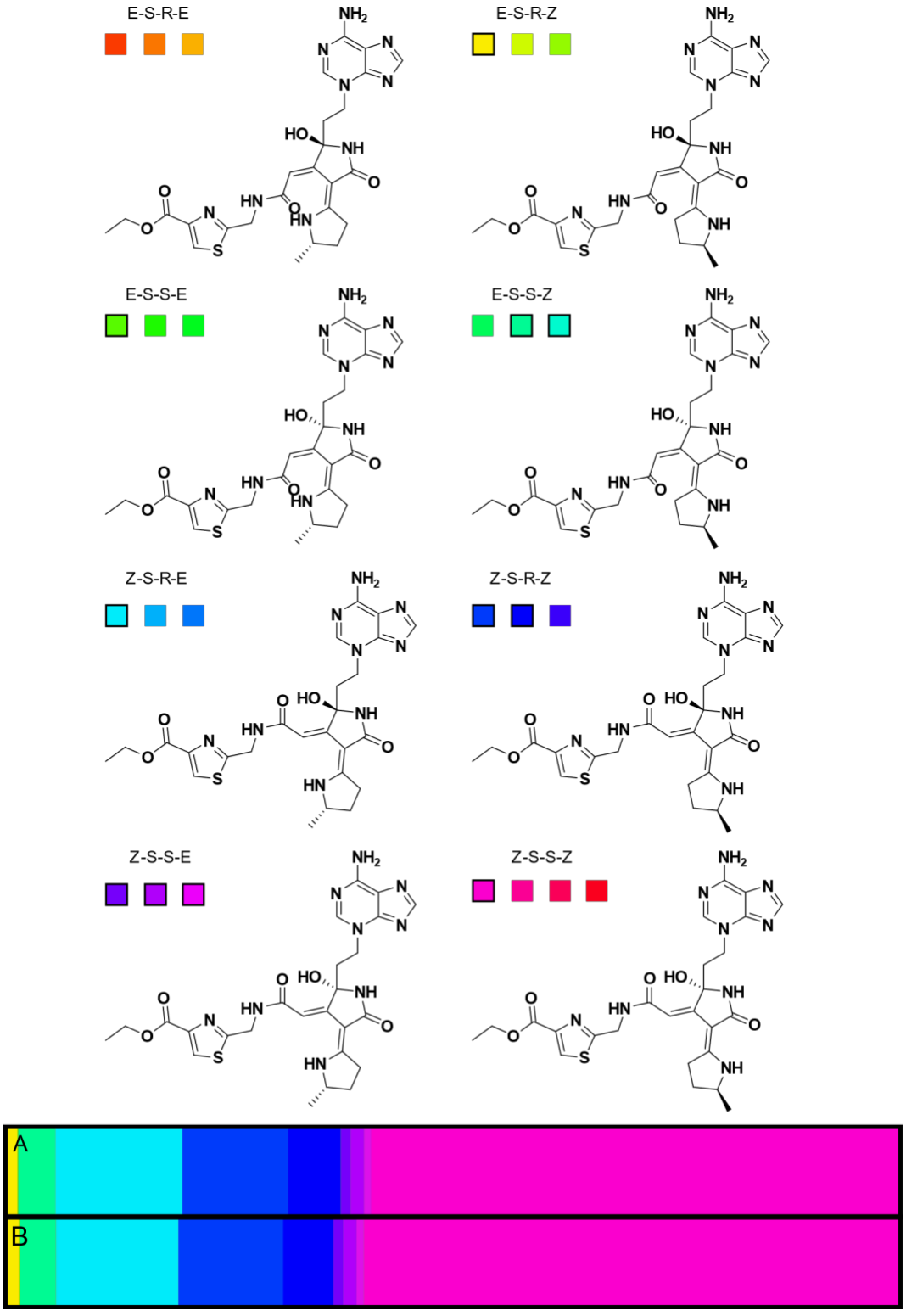


Fig. S53.

Bar plot of DP4 score for each rotamer of the calculated spectra compared to the diasteromeric pair A/B of the experimental data set. Contributing rotamers are bolded below their structure configuration.

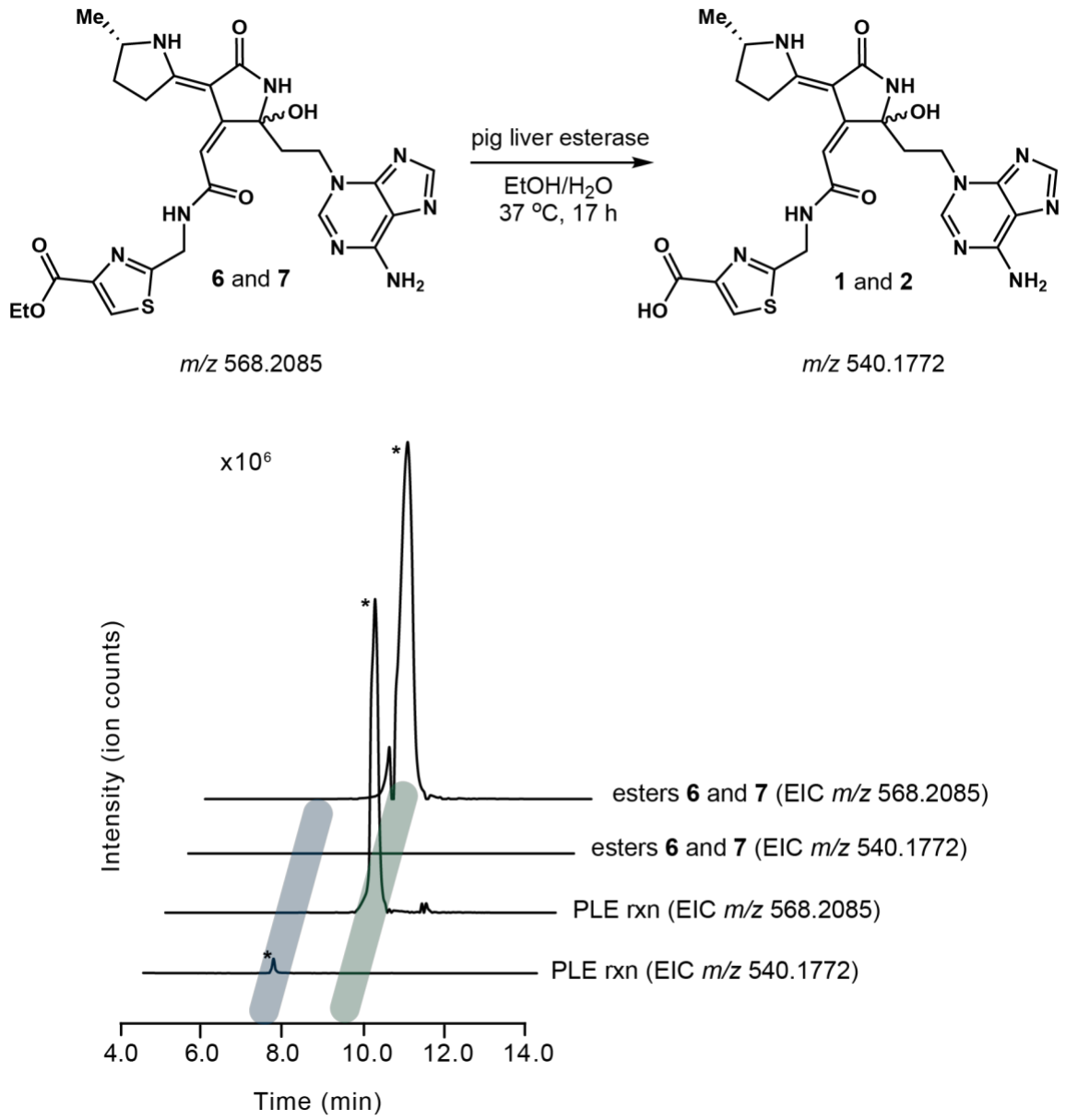


Fig. S54.
Enzyme-catalyzed hydrolysis of the ethyl ester-containing DNA adducts 6 and 7. Extracted ion chromatograms for the hydrolysis reaction of diastereomeric esters **6** and **7** with pig liver esterase enzyme (~15 units of enzyme). * = the observed peaks resolve into two peaks on longer HPLC methods (see main text Fig. 3C for synthetic standard trace; see Fig. S39C for the resolution of adducts **6** and **7**).

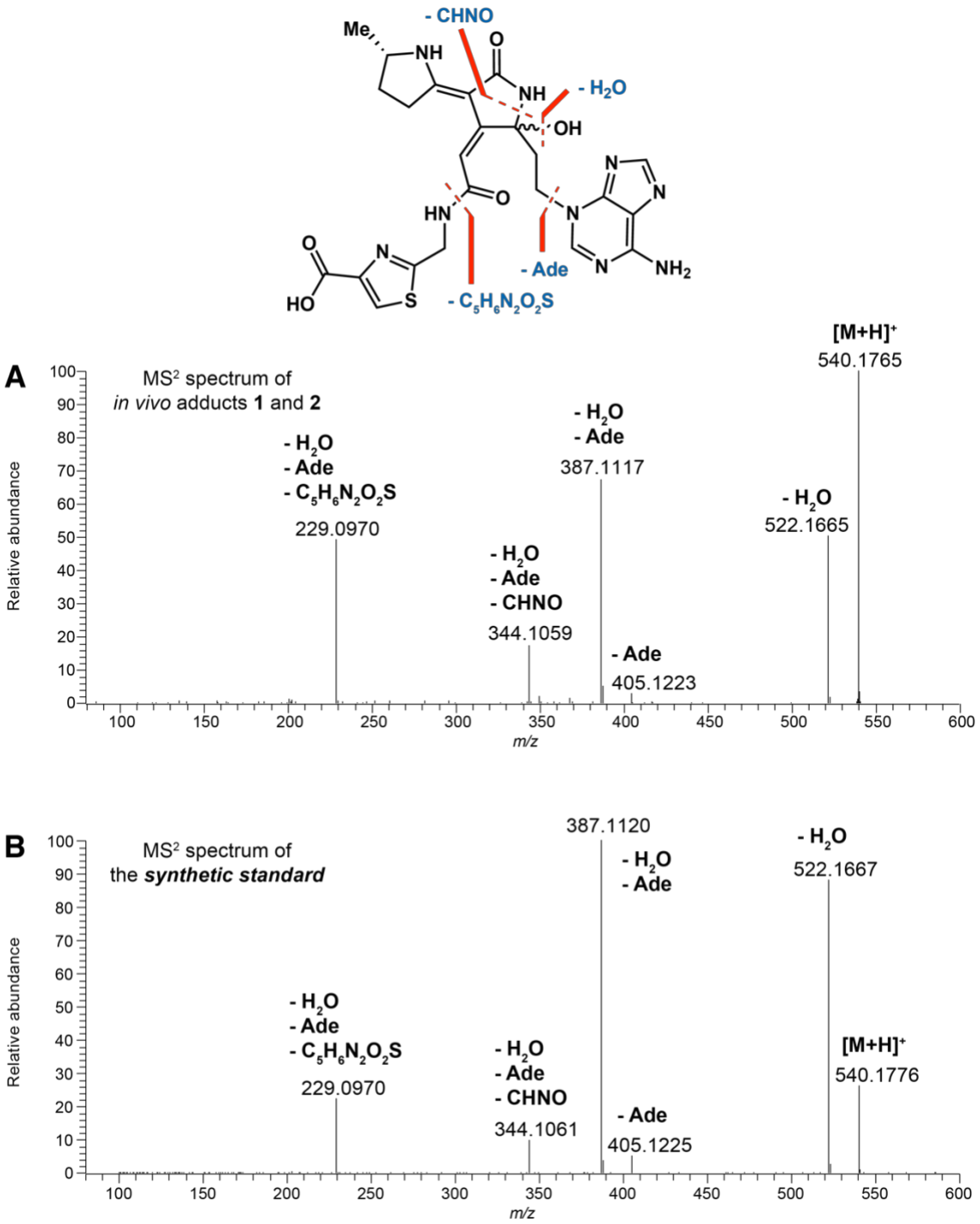


Fig. S55.

High-resolution MS^2 fragmentation of the observed *in vivo* adducts 1 and 2 (*m/z* 540.1772) and the synthetic standard (*m/z* 540.1772). (A) MS^2 fragmentation spectrum of the observed *in vivo* adducts (**1** and **2**) detected in HeLa cells exposed to *pks*$^+$ *E. coli*; fragmentation acquired with an HCD collision energy of 15%. (B) MS^2 fragmentation spectrum of the synthetic standard (*m/z* 540.1772); fragmentation acquired with an HCD collision energy of 15%.

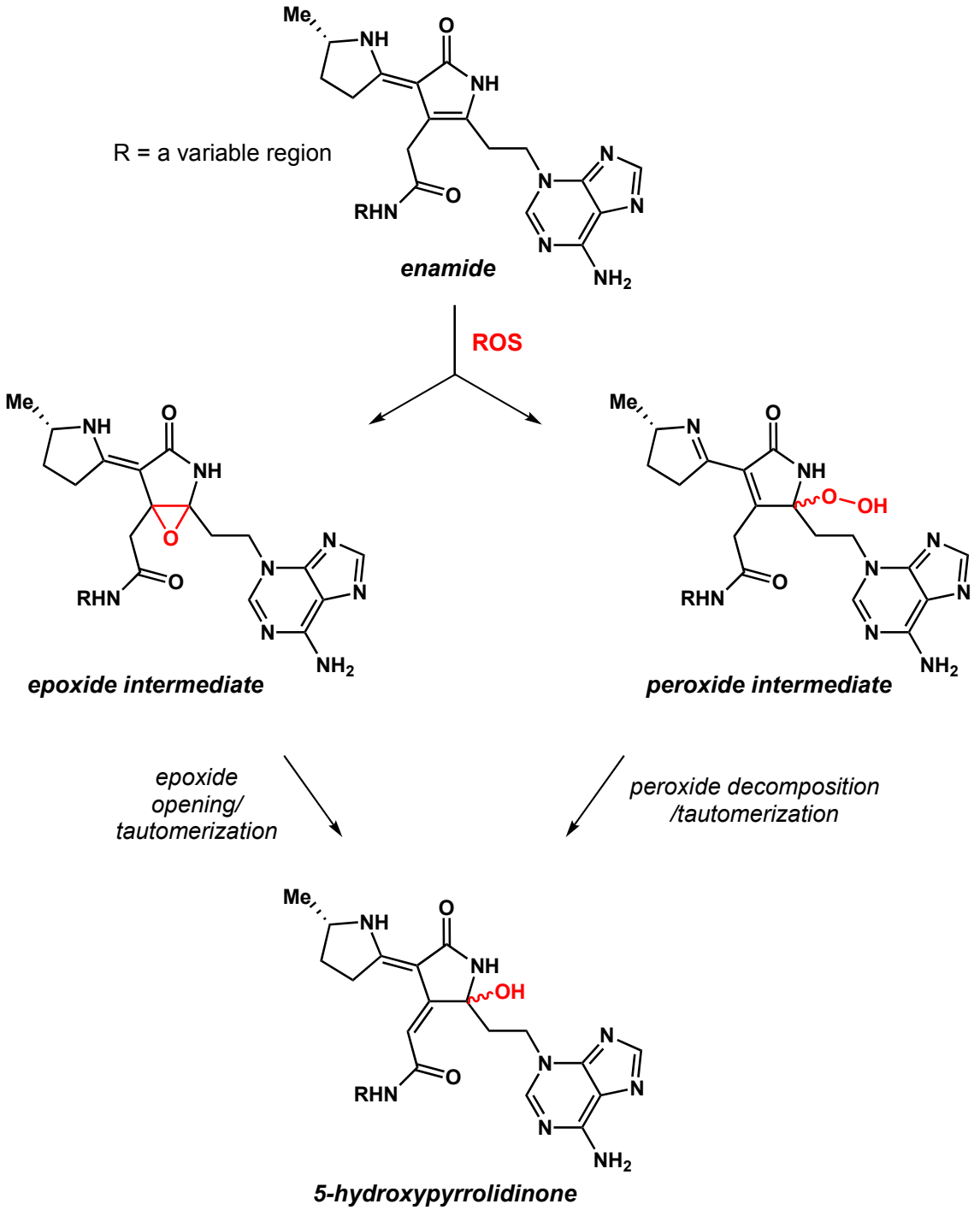


Fig. S56.
Proposed transformation of the unstable enamide to the 5-hydroxypyrrolidinone ring. The enamide could react with reactive oxygen species (ROS) to give an epoxide or peroxide intermediate. Ring opening and tautomerization of the epoxide intermediate could yield the 5-hydroxypyrrolidinone ring system. The peroxide intermediate could undergo decomposition and tautomerization to give the same ring system.

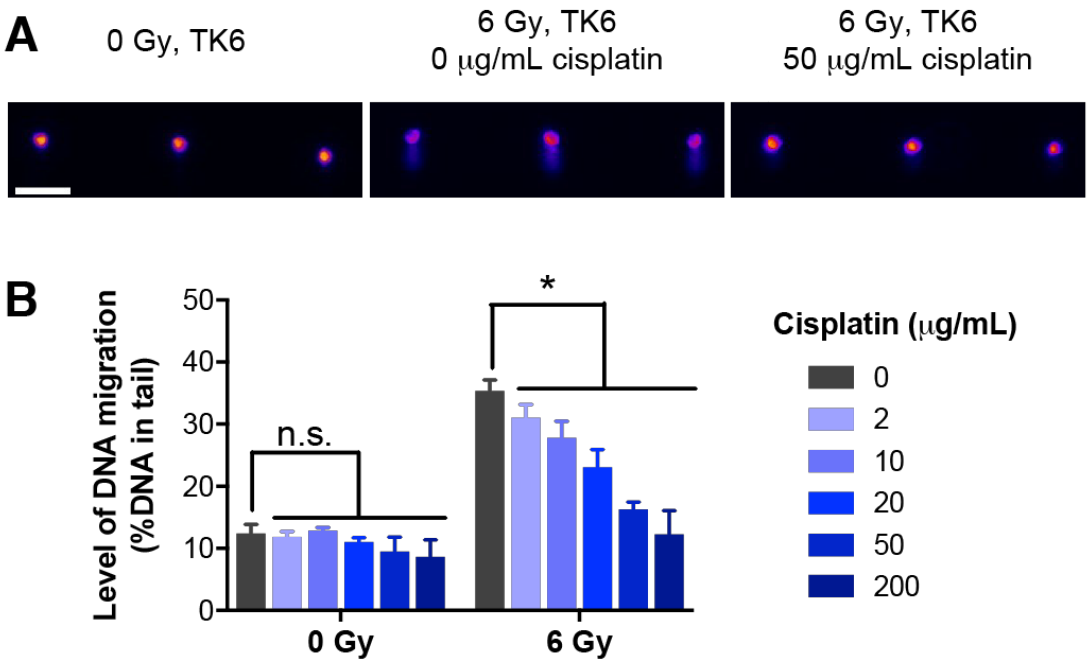


Fig. S57.
Cisplatin treatment induces DNA cross-links in TK6 cells. (A) Arrayed microwell comets from untreated TK6 cells and TK6 cells treated with 6 Gy γIR, with and without cisplatin (0 or 50 μg/mL). Scale bar indicates 100 μm. (B) Quantification of % DNA in tail from untreated TK6 cells and TK6 cells treated with 6 Gy γIR. Unirradiated and irradiated cells were treated with varying concentrations of cisplatin (0–200 μg/mL). Data points and error bars represent mean +/- SEM, respectively, of three independent experiments. Student's *t*-test was performed to compare each treated dose to the corresponding negative controls (n.s. = not significant; * $P < 0.05$).

Table S1.

CFU (colony forming units) counts from C57BL/6J mice monocolonized with pBelo or *pks*$^{+}$ *E. coli* for 2 weeks (samples colored red were used for targeted LC-MS/MS method development).

pBelo samples	**CFU per feces (dry weight)**	***pks*$^{+}$ samples**	**CFU per feces (dry weight)**
pBelo #1	3.284E+08	*pks* #7	1.298E+09
pBelo #2	8.053E+08	*pks* #8	9.094E+08
pBelo #3	1.344E+09	*pks* #9	1.578E+09
pBelo #4	8.081E+08	*pks* #10	5.299E+08
pBelo #5	9.484E+08	*pks* #11	1.083E+09
pBelo #6	5.844E+08	*pks* #12	1.192E+09
		pks #13	7.909E+08
		pks #14	9.166E+08
		pks #15	1.049E+09
		pks #16	7.141E+08
		pks #17	1.239E+09

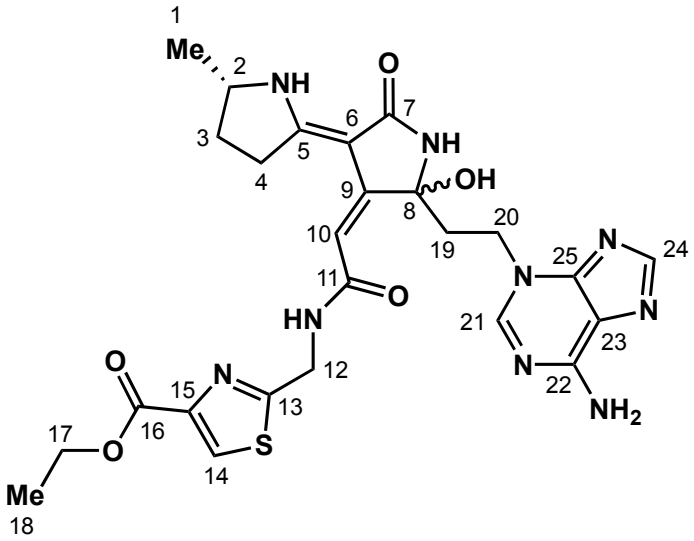
Table S2.

Relative amounts (expressed as ion signals) of adducts **1** and **2** in DNA isolated from colonic epithelial cells of C57BL/6J mice monocolonized with pBelo or *pks*$^+$ *E. coli* for 2 weeks. Amounts determined by area under the curve integrations of the extracted ion chromatograms (EICs) of the most abundant MS^2 fragmentation ion (*m/z* 387.1118 ± 0.008) of the adducts **1** and **2** precursor ion (*m/z* 540.1772).

Samples	**DNA Amount (µg)**	**DNA Equivalent Analyzed (Injected) (µg)**	**EIC *m/z* 387.1118 ion counts (cts)**	**Relative Quantity (cts/µg DNA Injected)**
pBelo #4	95.4	31.8	0	0
pBelo #5	244.9	81.6	0	0
pBelo #6	158.4	52.8	0	0
pks #9	232.3	77.4	11934	154
pks #11	242.8	80.9	16093	199
pks #12	214.0	71.3	13557	190
pks #13	198.1	66.0	9338	141
pks #14	215.9	72.0	13614	189
pks #15	196.0	65.3	15543	238
pks #16	190.4	63.5	12737	201
pks #17	199.7	66.6	17071	257

Table S3.

NMR spectroscopic data for diastereomeric adducts **6** and **7** in DMSO-d_6.



* = denotes diastereotopic peak

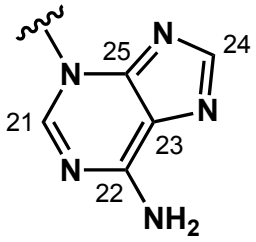
Carbon	δC (type)	δH, multiplicity, *J* (Hz)	COSY	HMBC	ROESY
1	21.2 (CH_3) 21.0 (CH_3)*	1.24, d, (6.2) 1.20, d, (6.2)*	2	2, 3	-
2	54.7 (CH)	3.95, dddd, (13.5, 6.5)	1, 3a, 3b	5	-
	-	NH, 9.55, s 9.50, s*	-	2, 3, 4, 5, 6	1
3	28.3 (CH_2)	a. 2.23-2.19, m b. 1.63-1.55, m	2, 3b, 4 2, 3a, 4	1, 4, 5 1, 2, 4, 5	1 1
4	32.2 (CH_2)	2.95-2.71, m	3a, 3b	2, 3, 5	10
5	163.7 (C)	-	-	-	-
6	90.8 (C)	-	-	-	-
7	171.1 (C)	-	-	-	-
	-	NH, 8.20, s 8.18, s*	-	6, 7, 8, 9	19, 20
8	86.4 (C)	-	-	-	-
	-	OH, 7.82, s 7.81, s*	-	8	19, 20 (weak)
9	159.3 (C)	-	-	-	-
10	99.4 (CH)	5.49, s 5.48, s*	-	6, 8, 9, 11	4, 11 NH
11	168.6 (C)	-	-	-	-
	-	NH, 8.93, t (5.5)	12	11, 12	10
12	40.4 (CH_2)	4.56-4.48, m	11 NH	11, 13	-
13	170.3 (C)	-	-	-	-
14	129.1 (CH)	8.302, s 8.299, s*	-	13, 15, 16 (weak)	-

15	145.4 (C)	-	-	-	-
16	160.7 (C)	-	-	-	-
17	60.5 (CH_2)	4.29, q (7.1)	18	16, 18	-
18	14.0 (CH_3)	1.29, t (7.1)	17	17	-
19	39.1 (CH_2)	2.43-2.40, m	20	8, 9, 20	7 NH, 8 OH, 21
20	45.3 (CH_2)	4.37-4.32, m	19	8, 19, 21, 25	7 NH (weak), 8 OH, 21
21	143.0 (CH)	8.17, s 8.15, s*	-	20, 22, 23, 25	19, 20
22	154.8 (C)	-	-	-	-
	-	NH_2, 7.88-7.72, br s	-	23, 25	-
23	120.5 (C)	-	-	-	-
24	152.2 (C)	7.706, s 7.702, s*	-	23, 25	-
25	150.0 (C)	-	-	-	-

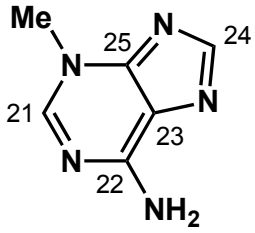
Table S4.

Comparison of the ^{1}H and ^{13}C chemical shifts of the adenine ring in adducts **6** and **7** and 3-methyladenine in DMSO-d_6.

Adenine ring in adducts **6** and **7**



3-Methyladenine



* = denotes diastereotopic peak

Carbon	**δC (ppm)** (DNA adducts **6** and **7**)	**δH (ppm), multiplicity,** (DNA adducts **6** and **7**)	**δC (ppm)** (3-Methyladenine)	**δH (ppm), multiplicity,** (3-Methyladenine)
21	143.0	8.17, s 8.15, s*	143.7	8.29, s
22	154.8	-	154.9	-
	-	NH_2, 7.88-7.72, br s	-	NH_2, 7.92-7.81, br s
23	120.5	-	120.3	
24	152.2	7.706, s 7.702, s*	152.4	7.76, s
25	150.0	-	150.3	-

Table S5.

Gaussian calculated ^{13}C δ for the ester standard, Part A.

Position	Observed	ChemDraw	ESRE 1	ESRE 2	ESRE 3	ESRZ 1	ESRZ 2[b]	ESRZ 3[b]	ESSE 1	ESSE 2	ESSE 3	ESSZ 1	ESSZ 2	ESSZ 3
C1[a]	21.2[a]	21.3	19.9	19.7	19.7	20.5	19.4	19.4	20.3	19.3	19.3	19.6	20.7	20.4
C2	54.7	57.4	58.3	58.5	58.8	57.4	57.4	57.4	58.8	56.6	58.4	57.5	57.4	57.2
C3	28.3	32.3	30.8	31.0	31.4	29.4	31.7	31.7	28.8	30.7	31.3	32.1	29.4	29.0
C4	32.2	28.9	36.0	35.8	35.9	34.7	37.3	37.3	34.5	33.9	37.2	36.9	34.3	36.3
C5	163.7	160.4	168.1	167.5	168.3	167.0	167.3	167.3	167.6	160.3	168.6	165.9	166.5	168.4
C6	90.8	107.5	92.7	92.7	93.1	91.9	91.3	91.3	94.1	89.1	93.4	92.0	92.1	93.2
C7	171.1	174.7	172.5	172.7	172.6	171.4	173.2	173.2	170.3	171.0	172.1	170.6	172.5	169.7
C8	86.4	90.3	87.4	87.6	87.4	87.5	88.3	88.3	89.8	88.1	87.9	90.8	88.5	86.4
C9	159.3	157.4	159.5	157.2	158.8	160.4	156.7	156.7	158.4	153.1	160.4	159.0	159.3	157.4
C10	99.4	119.1	93.2	96.7	93.6	92.7	98.2	98.2	96.0	108.0	94.6	97.1	97.6	98.3
C11	168.6	166.4	168.8	166.4	167.0	167.2	168.4	168.4	168.1	165.9	168.4	166.9	168.7	167.9
C12	40.4	40.7	47.0	42.2	44.0	46.4	43.6	43.6	45.0	42.1	47.1	46.1	45.2	45.4
C13	170.3	165.5	172.5	175.6	164.8	173.0	169.1	169.1	171.7	170.9	172.2	174.4	176.5	177.0
C14	129.1	127.5	131.0	128.9	138.0	135.0	137.3	137.3	135.8	137.6	135.8	130.7	133.6	133.6
C15	145.4	146.8	141.3	142.5	144.7	143.9	143.8	143.8	143.0	142.9	143.3	142.3	145.5	144.9
C16	160.7	160.8	163.9	164.1	163.2	161.5	163.1	163.1	162.0	162.1	161.9	163.7	162.1	162.2
C17	60.5	60.9	68.8	68.8	69.4	61.9	69.3	69.3	62.8	62.8	62.9	68.9	62.8	62.9
C18	14.0	14.1	15.4	15.6	15.4	14.4	15.3	15.3	13.5	13.5	13.5	15.5	13.5	13.6
C19	39.1	39.9	40.3	39.9	39.9	39.1	38.9	38.9	39.5	38.3	39.9	39.0	39.2	40.0
C20	45.3	34.4	46.8	46.7	46.6	46.0	46.7	46.7	46.8	47.8	48.0	46.6	47.1	46.2
C21	143.0	148.4	141.8	141.7	141.9	141.8	141.9	141.9	141.8	141.8	141.5	141.5	141.7	141.6
C22	154.8	155.8	150.6	150.1	150.2	150.6	150.3	150.3	150.4	150.2	150.6	150.1	150.2	150.1
C23	120.5	120.5	120.3	119.6	120.5	120.8	119.3	119.3	119.6	119.9	119.4	119.7	118.3	118.9
C24	152.2	149.1	152.9	153.1	152.9	153.0	153.1	153.1	153.0	178.8	153.0	153.0	153.1	152.9
C25	150.0	155.8	149.0	148.9	149.9	149.5	149.1	149.1	148.4	149.0	149.7	148.7	148.0	148.9

[a]The first position appears split as two equally intense diasterotopic peaks, one at ^{1}H δ 1.20 d (6.2) / ^{13}C δ 21.2 and the other at ^{1}H δ 1.24 d (6.2) / ^{13}C δ 21.0.
[b]The calculated values for these rotamer were not equivalent, but appear as such when rounded to the tenth decimal place.

Table S6.

Gaussian calculated ^{13}C δ for diastereomeric adducts **6** and **7**, Part B.

Position	Observed	ChemDraw	ZSRE 1	ZSRE 2	ZSRE 3	ZSRZ 1	ZSRZ 2	ZSRZ 3	ZSSE 1	ZSSE 2	ZSSE 3	ZSSZ 1	ZSSZ 2	ZSSZ 3	ZSSZ 4
C1[a]	21.2[a]	21.3	19.8	20.1	19.8	19.6	20.7	20.6	19.4	20.2	20.1	19.4	19.4	19.5	19.3
C2	54.7	57.4	57.7	58.2	57.6	57.5	57.4	57.2	57.8	57.6	58.0	57.5	56.5	56.5	57.2
C3	28.3	32.3	28.2	28.7	28.6	29.9	28.8	28.8	30.5	28.4	28.7	30.3	30.6	30.2	30.8
C4	32.2	28.9	33.5	33.1	33.9	35.1	32.8	32.0	35.7	33.1	32.7	35.6	34.0	34.2	34.7
C5	163.7	160.4	163.5	162.6	162.9	164.8	164.9	163.6	163.8	164.0	163.4	164.8	162.2	163.4	165.1
C6	90.8	107.5	92.9	91.0	91.8	91.7	92.1	91.2	92.9	91.3	90.9	92.9	89.3	89.7	93.7
C7	171.1	174.7	167.1	167.3	168.7	168.9	168.7	169.2	167.3	167.6	168.2	168.9	167.9	168.6	169.6
C8	86.4	90.3	86.5	87.0	87.3	88.2	88.8	90.3	86.9	86.5	86.3	86.7	87.5	87.8	87.3
C9	159.3	157.4	163.6	153.2	151.8	162.7	163.1	154.2	164.2	156.0	155.1	160.4	150.5	151.9	163.3
C10	99.4	119.1	99.3	109.5	108.5	97.1	96.9	107.4	99.4	105.5	106.3	98.2	100.6	99.7	93.6
C11	168.6	166.4	166.3	163.9	163.5	165.6	166.0	166.4	166.3	172.8	172.2	166.2	177.1	177.1	167.4
C12	40.4	40.7	42.2	39.4	38.8	40.8	43.2	43.9	41.9	42.5	42.0	41.7	47.3	47.9	45.1
C13	170.3	165.5	171.5	170.4	171.7	171.9	175.3	176.1	172.0	171.5	171.6	171.8	170.5	170.8	172.9
C14	129.1	127.5	134.6	135.1	134.2	133.8	133.3	133.2	135.4	134.6	134.5	134.9	134.6	134.6	135.8
C15	145.4	146.8	143.2	144.5	146.0	144.5	144.7	143.4	143.1	143.5	143.8	143.0	144.1	143.8	143.0
C16	160.7	160.8	161.7	163.1	160.8	162.2	161.6	162.2	161.7	161.6	161.6	161.7	161.9	162.0	162.0
C17	60.5	60.9	62.4	69.6	61.8	62.7	61.9	62.7	61.7	61.9	62.0	61.8	62.9	63.0	62.5
C18	14.0	14.1	13.3	15.4	14.4	13.7	14.4	13.5	14.4	14.5	14.5	14.6	13.6	13.6	14.3
C19	39.1	39.9	39.4	36.9	37.5	37.3	36.8	36.1	37.4	39.2	41.6	39.4	38.7	39.2	39.9
C20	45.3	34.4	45.5	45.5	44.7	46.6	46.3	45.2	47.5	47.5	46.5	47.7	47.4	47.1	47.2
C21	143.0	148.4	141.5	142.1	141.6	143.5	143.4	143.5	141.7	141.2	141.7	141.6	141.2	141.6	141.8
C22	154.8	155.8	150.1	149.9	149.9	150.2	150.1	150.3	150.6	150.6	150.3	150.3	150.5	150.7	150.3
C23	120.5	120.5	119.8	119.7	119.6	119.5	119.5	119.5	120.4	120.6	119.4	119.9	120.4	120.0	119.8
C24	152.2	149.1	152.9	153.0	153.0	152.5	152.5	152.9	153.1	153.0	152.9	153.0	153.0	153.2	153.0
C25	150.0	155.8	148.5	149.2	148.7	148.7	148.6	148.9	148.6	148.7	148.8	148.5	149.4	149.0	148.6

[a]The first position appears split as two equally intense diasterotopic peaks, one at ^{1}H δ 1.20 d (6.2) / ^{13}C δ 21.2 and the other at ^{1}H δ 1.24 d (6.2) / ^{13}C δ 21.0.

Table S7.
Scoring of calculated spectra, based on absolute deviation and DP4 calculation.

Position	Mean A.D.	Max A.D.	DP4 %*
ChemDraw	3.88	19.7	N/A
ESRE_1	2.57	8.3	0/0
ESRE_2	2.47	8.3	0/0
ESRE_3	2.79	8.9	0/0
ESRZ_1	1.94	6.7	1.4/1.4
ESRZ_2	2.47	8.8	0/0
ESRZ_3	2.47	8.8	0/0
ESSE_1	2.16	6.7	0.1/0.1
ESSE_2	3.48	26.6	0/0
ESSE_3	2.51	6.7	0/0
ESSZ_1	2.54	8.4	0/0
ESSZ_2	1.94	6.2	4.1/4.0
ESSZ_3	2.11	6.7	0.1/0.1
ZSRE_1	1.71	5.5	14.1/13.6
ZSRE_2	2.57	10.1	0/0
ZSRE_3	2.20	9.1	0/0
ZSRZ_1	1.85	4.7	11.8/11.7
ZSRZ_2	1.87	5.0	5.9/5.6
ZSRZ_3	2.23	8.0	0/0
ZSSE_1	2.01	6.3	1.1/1.1
ZSSE_2	1.89	6.1	1.5/1.5
ZSSE_3	1.94	6.9	0.8/0.8
ZSSZ_1	1.82	5.8	59.1/60.1
ZSSZ_2	2.46	8.8	0/0
ZSSZ_3	2.32	8.5	0/0
ZSSZ_4	2.33	6.7	0/0

*DP4 scores were calculated separately for the experimental shifts of each diastereomer at the C1 position. In both cases the first ZSSZ configuration is clearly best, and we attribute this to our carbon shifts being determined by HSQC and HMBC projection where it was not possible to detect distinct diasterotopic carbon shifts for any carbon other than C1.

References and Notes:

1. S. V. Lynch, O. Pedersen, The human intestinal microbiome in health and disease. *N. Engl. J. Med*. **375**, 2369–2379 (2016).

2. J.-P. Nougayrède, S. Homburg, F. Taieb, M. Boury, E. Brzuszkiewicz, G. Gottschalk, C. Buchrieser, J. Hacker, U. Dobrindt, E. Oswald, *Escherichia coli* induces DNA double-strand breaks in eukaryotic cells. *Science* **313**, 848–851 (2006).

3. G. Cuevas-Ramos, C. R. Petit, I. Marcq, M. Boury, E. Oswald, J.-P. Nougayrède, *Escherichia coli* induces DNA damage *in vivo* and triggers genomic instability in mammalian cells. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 11537–11542 (2010).

4. J. C. Arthur, E. Peréz-Chanona, M. Mühlbauer, S. Tomkovich, J. M. Uronis, T. J. Fan, B. J. Campbell, T. Abujamel, B. Dogan, A. B. Rogers, J. M. Rhodes, A. Stintzi, K. W. Simpson, J. J. Hansen, T. O. Keku, A. A. Fodor, C. Jobin, Intestinal inflammation targets cancer-inducing activity of the microbiota. *Science* **338**, 120–123 (2012).

5. E. Buc, D. Dubois, P. Sauvanet, J. Raisch, J. Delmas, A. Darfeuille-Michaud, D. Pezet, R. Bonnet, High prevalence of mucosa-associated *E. coli* producing cyclomodulin and genotoxin in colon cancer. *PLoS ONE* **8**, e56964 (2013).

6. J. C. Arthur, R. Z. Gharaibeh, M. Mühlbauer, E. Perez-Chanona, J. M. Uronis, J. McCafferty, A. A. Fodor, C. Jobin, Microbial genomic analysis reveals the essential role of inflammation in bacteria-induced colorectal cancer. *Nat. Commun*. **5**, 4724 (2014).

7. V. Eklof, A. Löfgren-Burström, C. Zingmark, S. Edin, P. Larsson, P. Karling, O. Alexeyev, J. Rutegård, M. L. Wikberg, R. Palmqvist, Cancer-associated fecal microbial markers in colorectal cancer detection. *Int. J. Cancer*. **141**, 2528–2536 (2017).

8. C. M. Dejea, P. Fathi, J. M. Craig, A. Boleij, R. Taddese, A. L. Geis, X. Wu, C. E. DeStefano Shields, E. M. Hechenbleikner, D. L. Huso, R. A. Anders, F. M. Giardiello, E. C. Wick, H. Wang, S. Wu, D. M. Pardoll, F. Housseau, C. L. Sears, Patients with familial adenomatous polyposis harbor colonic biofilms containing tumorigenic bacteria. *Science* **359**, 592–597 (2018).

9. M. I. Vizcaino, P. Engel, E. Trautman, J. M. Crawford, Comparative metabolomics and structural characterizations illuminate colibactin pathway-dependent small molecules. *J. Am. Chem. Soc*. **136**, 9244–9247 (2014).

10. M. I. Vizcaino, J. M. Crawford, The colibactin warhead crosslinks DNA. *Nat. Chem*. **7**, 411–417 (2015).

11. C. A. Brotherton, M. R. Wilson, G. Byrd, E. P. Balskus, Isolation of a metabolite from the *pks* island provides insights into colibactin biosynthesis and activity. *Org. Lett*. **17**, 1545–1548 (2015).

12. X. Y. Bian, A. Plaza, Y. M. Zhang, R. Müller, Two more pieces of the colibactin genotoxin puzzle from *Escherichia coli* show incorporation of an unusual 1-aminocyclopropanecarboxylic acid moiety. *Chem. Sci*. **6**, 3154–3160 (2015).

13. Z. R. Li, Y. Li, J. Y. Lai, J. Tang, B. Wang, L. Lu, G. Zhu, X. Wu, Y. Xu, P. Y. Qian, Critical intermediates reveal new biosynthetic events in the enigmatic colibactin pathway. *ChemBioChem* **16**, 1715–1719 (2015).

14. L. Zha, M. R. Wilson, C. A. Brotherton, E. P. Balskus, Characterization of polyketide synthase machinery from the *pks* island facilitates isolation of a candidate precolibactin. *ACS Chem. Biol*. **11**, 1287–1295 (2016).

15. Z. R. Li, J. Li, J. P. Gu, J. Y. Lai, B. M. Duggan, W. P. Zhang, Z. L. Li, Y. X. Li, R. B. Tong, Y. Xu, D. H. Lin, B. S. Moore, P. Y. Qian, Divergent biosynthesis yields a cytotoxic aminomalonate-containing precolibactin. *Nat. Chem. Biol*. **12**, 773–775 (2016).

16. D. Dubois, O. Baron, A. Cougnoux, J. Delmas, N. Pradel, M. Boury, B. Bouchon, M. A. Bringer, J. P. Nougayrède, E. Oswald, R. Bonnet, ClbP is a prototype of a peptidase subgroup involved in biosynthesis of nonribosomal peptides. *J. Biol. Chem*. **286**, 35562–35570 (2011).

17. C. A. Brotherton, E. P. Balskus, A prodrug resistance mechanism is involved in colibactin biosynthesis and cytotoxicity. *J. Am. Chem. Soc*. **135**, 3359–3362 (2013).

18. X. Bian, J. Fu, A. Plaza, J. Herrmann, D. Pistorius, A. F. Stewart, Y. Zhang, R. Müller, *In vivo* evidence for a prodrug activation mechanism during colibactin maturation. *ChemBioChem* **14**, 1194–1197 (2013).

19. M. Tanasova, S. J. Sturla, Chemistry and biology of acylfulvenes: sesquiterpene-derived antitumor agents. *Chem. Rev*. **112**, 3578–3610 (2012).

20. N. Ghosh, H. M. Sheldrake, M. Searcey, K. Pors, Chemical and biological explorations of the family of CC-1065 and the duocarmycin natural products. *Curr. Top. Med. Chem*. **9**, 1494–1524 (2009).

21. D. Fu, J. A. Calvo, L. D. Samson, Balancing repair and tolerance of DNA damage caused by alkylating agents. *Nat. Rev. Cancer* **12**, 104–120 (2012).

22. R. Healy, H. Nikolayevskiy, J. R. Patel, J. M. Crawford, S. B. Herzon, A mechanistic model for colibactin-induced genotoxicity. *J. Am. Chem. Soc.* **138**, 15563–15570 (2016).

23. J. J. Truglio, D. L. Croteau, B. Van Houten, C. Kisker, Prokaryotic nucleotide excision repair: the UvrABC system. *Chem. Rev*. **106**, 233–252 (2006).

24. N. Bossuet-Greif, D. Dubois, C. Petit, S. Tronnet, P. Martin, R. Bonnet, E. Oswald, J.-P. Nougayrède, *Escherichia coli* ClbS is a colibactin resistance protein. *Mol. Microbiol.* **99**, 897–908 (2016).

25. P. Tripathi, E. E. Shine, A. R. Healy, C. S. Kim, S. B. Herzon, S. D. Bruner, J. M. Crawford, ClbS is a cyclopropane hydrolase that confers colibactin resistance. *J. Am. Chem. Soc*. **139**, 17719–17722 (2017).

26. S. Balbo, S. S. Hecht, P. Upadhyaya, P. W. Villalta, Application of a high-resolution mass-spectrometry-based DNA adductomics approach for identification of DNA adducts in complex mixtures. *Anal. Chem*. **86**, 1744–1752 (2014).

27. A. Stornetta, P. W. Villalta, S. S. Hecht, S. J. Sturla, S. Balbo, Screening for DNA alkylation mono and cross-linked adducts with a comprehensive LC-MS^3 adductomic approach. *Anal. Chem*. **87**, 11706–11713 (2015).

28. K. E. Pietsch, P. M. van Midwoud, P. W. Villalta, S. J. Sturla, Quantification of acylfulvene- and Illudin S-DNA adducts in cells with variable bioactivation capacities. *Chem. Res. Toxicol*. **26**, 146–155 (2013).

29. M. Xue, E. Shine, W. Wang, J. M. Crawford, S. B. Herzon, Characterization of natural colibactin-nucleobase adducts by tandem mass spectrometry and isotopic labeling. Support for DNA alkylation by cyclopropane ring opening. *Biochemistry*. **57**, 6391–6394 (2018).

30. S. G. Smith, J. M. Goodman, Assigning stereochemistry to single diastereomers by GIAO NMR calculation: the DP4 probability. *J. Am. Chem. Soc*. **132**, 12946–12959 (2010).

31. D. L. Boger, T. Ishizaki, H. Zarrinmayeh, Isolation and characterization of the duocarmycin-adenine DNA adduct. *J. Am. Chem. Soc.* **113**, 6645–6649 (1991).

32. N. Bossuet-Greif, J. Vignard, F. Taieb, G. Mirey, D. Dubois, C. Petit, E. Oswald, J.-P. Nougayrède, The colibactin genotoxin generates DNA interstrand cross-links in infected cells. *mBio* **9**, e02393–17 (2018).

33. P. L. Olive, J. P. Banáth, Kinetics of H2AX phosphorylation after exposure to cisplatin. *Cytometry B Clin. Cytom.* **76**, 79-90 (2009).

34. D. K. Wood, D. M. Weingeist, S. N. Bhatia, B. P. Engelward, Single cell trapping and DNA damage analysis using microwell arrays. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 10008–10013 (2010).

35. J. Ge, D. N. Chow, J. L. Fessler, D. M. Weingeist, D. K. Wood, B. P. Engelward, Micropatterned comet assay enables high throughput and sensitive DNA damage quantification. *Mutagenesis* **30**, 11–13 (2014).

36. D. Wenkert, K. M. Eliasson, D. Rudisill, Hydrogen peroxide oxidation of α-(*N*,*N*-dialkyl)aminoketones. *J. Chem. Soc., Chem Commun.* **7**, 392–393 (1983).

37. W. M. Grady, J. M. Carethers, Genomic and epigenetic instability in colorectal cancer pathogenesis. *Gastroenterology* **135**, 1079–1099 (2008).

38. M. C. Poirier, Chemical-induced DNA damage and human cancer risk. *Nat. Rev. Cancer* **4**, 630–637 (2004).

39. T. M. Wassenaar, *E. coli* and colorectal cancer: a complex relationship that deserves a critical mindset. *Crit. Rev. Microbiol.* **44**, 619–632 (2018).

40. A. R. Healy, M. I. Vizcaino, J. M. Crawford, S. B. Herzon, Convergent and modular synthesis of candidate precolibactins. Structural revision of precolibactin A. *J. Am. Chem. Soc*. **138**, 5426–5432 (2016).

41. G. K. Pierens, ^{1}H and ^{13}C NMR scaling factors for the calculation of chemical shifts in commonly used solvents using density functional theory. *J. Comput. Chem*. **35**, 1388–1394 (2014).