SUPPORTING INFORMATION

Tet2 catalyzes stepwise 5-methylcytosine oxidation by an iterative and *de novo* mechanism

Daniel J. Crawford, Monica Yun Liu, Christopher S. Nabel, Xing-Jun Cao, Benjamin A. Garcia, Rahul M. Kohli

Supplemental Methods

Enzymatic Preparation of *S*-[$^{13}C^2H_3$-Me]-adenosyl-*L*-methionine ([$^{13}C^2H_3$]-SAM): [$^{13}C^2H_3$]-SAM was enzymatically synthesized using [$^{13}C^2H_3$]-*L*-methionine (Sigma) and recombinant SAM synthetase. The *E. coli* SAM synthetase gene (*metK*) was cloned in a pET41 vector with an N-terminal HIS tag, expressed and purified as described[1]. To prepare [$^{13}C^2H_3$]-SAM, 0.2 mM [$^{13}C^2H_3$]-*L*-methionine was reacted with 2 mM ATP and 0.2 mg/ml SAM synthetase in 1X NEB Buffer 2 (10 mM Tris-Cl, 50 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT, pH 7.9) (Fig S1A). The mixture was incubated at 37°C for 15 min, after which duplexed DNA was added with CpG methyltransferase to generate [$^{13}C^2H_3$]-labeled mC oligos *in situ*, as described below.

Enzymatic Preparation of mC Oligonucleotides: 49.2 µM Oligo 1 and 50.8 µM Oligo 2 (Table S1) (49.2 µM duplex) were annealed in a thermocycler by incubation for 4 min at 95 °C followed by 30 s stepwise decreases of 5 °C to 40 °C. Similar protocols were utilized to generate substrates with specific introduction of (A) ^{14}C-methylation or (B) $^{13}C^2H_3$-methylation (Fig S1A). For (A), final concentrations of 5.8 µM Duplex in a reaction containing 1X NEB Buffer 2, 35 µM *S*-[^{14}C-Me]-adenosyl-*L*-methionine ([^{14}C]-SAM; Perkin Elmer) and 224 U/mL CpG Methyltransferase (NEB) was reacted at 37 °C for 4 h. For (B), final concentrations of 10 µM Duplex in a reaction containing 1X NEB Buffer 2, 1:5 [$^{13}C^2H_3$]-SAM enzymatic reaction mix (above), and 1000 U/mL CpG Methyltransferase (NEB) was reacted at 37° C for 5 h. For (A), a small sample of the reaction was collected to determine specific radioactivity for liquid scintillation counting (LSC). For either (A) or (B), at the end of methylation, the reaction was precipitated with 0.1 volume 3 M Sodium Acetate, pH 5.2 and 2.5 volumes of 100% ethanol on ice for 1 h and pelleted by centrifugation at 13,000 rpm for 30 min. The supernatant was then removed by pipetting, the pellet was washed with 70% ethanol/75 mM sodium acetate pH 5.2 stored at -20 °C and briefly re-centrifuged. The wash supernatant was removed and the pellet dried by air. The dried reaction pellet was dissolved in 0.2 initial reaction volume H_2O and desalted with G-25 spin columns equilibrated in water (CS-901, Princeton Separations). The desalted, methylated oligonucleotide mix was digested with 5,000 U/mL HpaII (NEB) in 1X Cutsmart Buffer (NEB), at 37 °C overnight. The digestion reaction was ethanol precipitated and washed as before, and the dried pellet dissolved in Purification Buffer A (100 mM triethylamine acetate (TEAA), pH 7). This sample was purified by ion-pairing HPLC using a 4.6 x 100 mm Zorbax Eclipse Plus C18 (Agilent) column pre-equilibrated to 65°C in 65% Purification Buffer A/35% Purification Buffer B (50% Methanol/100 mM TEAA, pH 7) and separated over a 20 min gradient from 35-45% Purification Buffer B at 1 mL/min, collecting 0.5 mL fractions (Fig S2). The HPLC conditions separate the 36-FAM-labeled full-length methylated Oligo 1 from Oligo 2 and unmethylated digested strands. Peaks were pooled and lyophilized for further use as [^{14}C]-mC-Oligo 3 or [$^{13}C^2H_3$]-mC-Oligo 3 (Table S1) and [^{14}C]-mC-Oligo 4 or [$^{13}C^2H_3$]-mC-Oligo 4 (Table S1).

Purification of Tet2: The plasmid encoding the catalytic domain of mTet2 (1042-1912) (referred to as Tet2 throughout) cloned into pFastBac1 plasmid with an N-terminal FLAG tag was generously provided by Yi Zhang. Tet2 protein was expressed in Sf9 cells as previously described[2]. 6 g of cells were resuspended in lysis buffer (50 mM HEPES, 500 mM NaCl, 0.1% NP-40 pH 7.4 at 4 °C) with cOmplete EDTA-free Protease Inhibitor Cocktail (Roche). Cells were lysed by 3 passes through a microfluidizer. The lysate was then clarified by centrifugation and bound to 1 mL α-Flag M2-affinity gel (Sigma) by gentle agitation for 2 h at 4 °C and applied to a PolyPrep column (BioRad). Bound protein was washed with 4 X

1 mL wash buffer (20 mM HEPES, 150 mM NaCl, 15% glycerol pH 7.4 at 4 °C) and eluted with 100 µg/mL 3X FLAG peptide (Sigma) in wash buffer, collecting 500 µL fractions. Fractions were evaluated for protein by BioRad Protein Assay (BioRad) and by SDS-PAGE (Fig S3A) for estimation of concentration. Fractions were pooled, DTT added to 1 mM, and aliquots frozen at -80 °C.

Tet2 Activity Assay: Duplexed oligonucleotide substrates were annealed in a thermocycler as described above. Optimized assay reaction conditions were used which yielded highest activity and linear turnover with time (Fig S3B). Notably, the conditions which show optimal activity may promote maintenance of active enzyme as the selected pH is associated with slower Fe(II) oxidation relative to higher pH conditions.[3] To a reaction mixture containing 50 mM HEPES pH 6.5, 100 mM NaCl, 1 mM 2-ketoglutarate, 1 mM DTT, 2 mM sodium ascorbate and the pre-annealed oligo, 75 µM freshly prepared ammonium iron(II) sulfate (Sigma) was added, followed by Tet2. The substrate and enzyme concentrations used in each reaction are noted in each figure legend, with typical reaction volumes of 50-200 µL. The reactions were incubated at 37 °C, quenched by addition of 8X 100% ethanol and 2X Oligo Binding Buffer (Zymo), purified over Oligo Clean and Concentrator columns (Zymo) per manufacturer instructions and eluted in H_2O. For the two highly quantitative HPLC and LC-MS/MS assays described below, the eluted reaction products were then degraded to component nucleosides with DNA Degradase Plus (Zymo) per manufacturer instructions at 37 °C overnight and further analyzed (Fig S1B). For semi-quantitative assays (Fig S3C), the purified reaction products were used (without degradation to nucleosides). For the MspI assay to detect most ox-mC products,[4] the purified reaction products were incubated with 2 mM UDP-glucose and 1:25 by volume of T4 β-glucosyltransferase (βGT, New England Biolabs) to glucosylate hmC products and incubated at 37 ºC for 30 min. Subsequently, the products were digested by MspI (3 U/µL) at 37 ºC for 2 hrs, resolved using denaturing PAGE and imaged for FAM fluorescence on a Typhoon scanner. For the TDG assay to detect fC and caC products, the reaction products were incubated with a 10-fold molar excess of TDG (purified as previously described[5]). After incubation at 37 ºC for 2 hrs, an equal volume of 0.3 M NaOH/0.03 M EDTA was added and the reaction incubated at 85 ºC for 15 min to cleave abasic sites. The samples were then processed and imaged as described for the MspI assay.

HPLC analysis of [^{14}C]-labeled nucleosides: 10 µM each mC, hmC, fC and caC nucleosides (Berry and Associates) were added to the degraded Tet reactions (as chromatographic controls), and this sample was injected onto a 2.1 x 250 mm Supelcosil LC-18S analytical column (Sigma) equilibrated to 50°C in 100% Analysis Buffer A (5 mM ammonium formate, pH 6.0). The nucleosides were separated in a gradient of 0-30% Analysis Buffer B (4 mM ammonium formate pH 6.0, 20% methanol) over 20 min at a flow rate of 0.5 mL/min. Fractions (0.25 mL in target areas, 1.0 mL for everywhere else) were collected and mixed with Opti-Fluor liquid scintillant (Perkin Elmer) for liquid scintillation counting (LSC) on a Tri-Carb 2910 TR (Perkin Elmer). Each vial was counted for 10 min using the ^{14}C DPM setting, with an automatic background correction made from the DPM measurement of a vial containing an appropriate liquid scintillant/HPLC buffer mixture. The outputted DPM measurements were corrected for inputted volume to reflect the total DPM of each fraction and plotted against the HPLC UV trace from that run, which shows the nucleoside standards to confirm identities (Fig 2A). The total radioactivity of each peak on the chromatogram was then analyzed to calculate percentage of each product, and normalized to the known inputted concentration to convert to molar quantities of products.

LC-MS/MS analysis of [$^{13}C^2H_3$]-labeled nucleosides: A 150 µm x 17 cm precolumn and 100 µm x 26 cm analytical reverse phase column were made by first preparing a Kasil frit (a 1:3 Formamide:Kasil 1624 (PQ corporation) mixture was drawn into column using capillarity, polymerized at 100 °C overnight and trimmed to ~5 mm), and packed with Supelcosil LC-18S resin (Sigma). Nano LC-MS chromatography was performed using an Easy-nLC 1000 (Thermo) with a two-column setup. The sample was bound to the precolumn and desalted by 5 min of isocratic flow of 0.1% formic acid, then separated on a gradient of

0-30% of acetonitrile into 0.1% formic acid over 30 min at a flow rate of 600 nL/min. Nucleosides were subjected to positive ion mode electrospray ionization in a Q Exactive hybrid quadrupole-orbitrap mass spectrometer (Thermo), with a capillary temperature of 275 °C and spray voltage of 2.9 kV. Total ion count and tandem MS transitions were collected, and the ratios of peak areas of heavy and light nucleosides were compared for each individual nucleotide modification (Fig S4).

References for Supplementary Methods:

1. Ottink, O. M.; Nelissen, F. H.; Derks, Y.; Wijmenga, S. S.; Heus, H. A. Enzymatic stereospecific preparation of fluorescent S-adenosyl-L-methionine analogs. *Anal. Biochem.* **2010***, 396*, 280-283.

2. Ito, S.; D'Alessio, A. C.; Taranova, O. V.; Hong, K.; Sowers, L. C.; Zhang, Y. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **2010***, 466*, 1129-1133.

3. Morgan, B.; Lahav, O. The effect of pH on the kinetics of spontaneous Fe(II) oxidation by O_2 in aqueous solution – basic principles and a simple heuristic description. *Chemosphere* **2007**, *68*, 2080-2084.

4. Ito, S.; Shen, L.; Dai, Q.; Wu, S. C.; Collins, L. B.; Swenberg, J. A.; He, C.; Zhang, Y. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **2011***, 333*, 1300-1303.

5. Maiti, A.; Drohat, A. C. Thymine DNA Glycosylase Can Rapidly Excise 5-Formylcytosine and 5-Carboxylcytosine: Potential Implications for Active Demethylation of CpG Sites. *J. Biol. Chem.* **2011***, 286*, 35334-35338.

Table S1: Oligonucleotides used in this study

Name	Sequence (all bases are 2'-deoxynucleotides)	Source/Notes
Oligo 1	5'-GTA TCT AGT TCA ATC CGG TTC ATA GCA-(36-FAM)-3'	Synthesized by Integrated DNA Technologies (IDT)
Oligo 2	5'-TGC TAT GAA CCG GAT TGA ACT AGA TAC-3'	IDT
Oligo 3	5'-GTA TCT AGT TCA ATC mCGG TTC ATA GCA-(36-FAM)-3'	Chemoenzymatic Preparation (isotopically labeled) or IDT
Oligo 4	5'-TGC TAT GAA CmCG GAT TGA ACT AGA TAC-3'	Chemoenzymatic Preparation (isotopically labeled) or IDT
Oligo 5	5'-TGC TAT GAA ChmCG GAT TGA ACT AGA TAC-3'	Synthesized*
Oligo 6	5'-TGC TAT GAA CfCG GAT TGA ACT AGA TAC-3'	Synthesized by TriLink
Oligo 7	5'-TGC TAT GAA CcaCG GAT TGA ACT AGA TAC-3'	Synthesized*

* Oligos 5 and 7 were prepared in house using standard phosphoramidite chemistry (reagents from Glen Research) on an ABI394 synthesizer (Applied Biosystems).

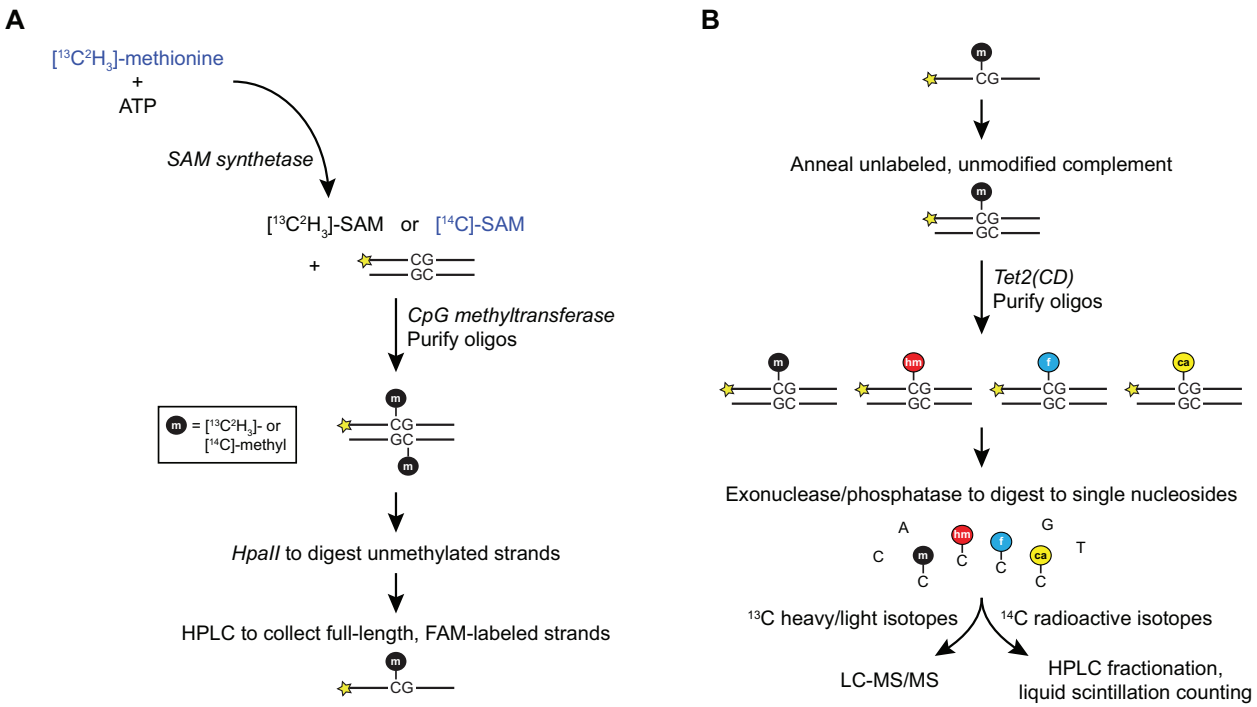

Figure S1: Schematic of experimental setup and analysis. **A)** Chemoenzymatic preparation of isotopically-labeled mC-containing substrates, as described in Supplementary Methods. The starting reagents purchased from commercial sources are indicated in blue: [^{14}C]-SAM or [$^{13}C^{2}H_{3}$]-methionine, which was enzymatically converted into [$^{13}C^{2}H_{3}$]-SAM. **B)** Tet activity assays. Duplexed oligonucleotide substrates were incubated with Tet2. The purified reaction products were degraded to nucleoside mixtures, which were then analyzed by either HPLC-LSC or LC-MS/MS.

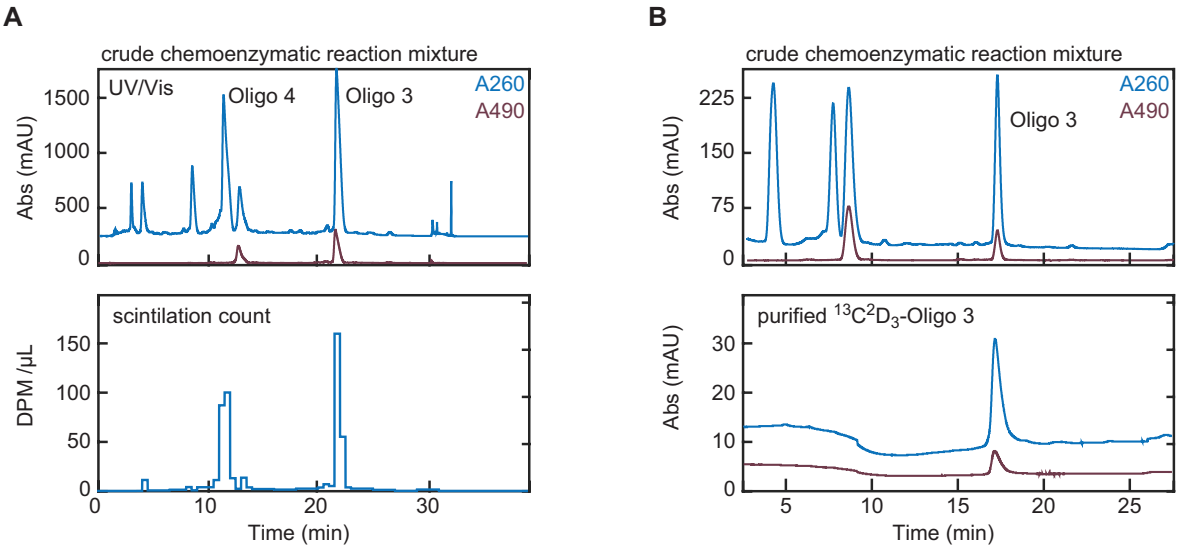

Figure S2: A) Chemoenzymatic generation of $^{14}CH_3$-labeled substrate. Shown is an HPLC chromatogram for the crude of the enzymatically-generated substrates, highlighting the A260 (blue) and FAM strand only (purple trace). At bottom is the associated scintillation counts, showing that Oligo 3 is radiolabeled with $^{14}CH_3$. Unlabeled peaks represent digested, unmethylated oligonucleotides. Purified radiolabeled Oligo 3 was used in all experiments involving LSC. **B)** Chemoenzymatic generation of $^{13}C^{2}D_3$-labeled substrate. Shown at top is the crude chemoenzymatic mixture from generation of the labeled substrate. The peak corresponding to Oligo 3 was purified and used in all experiments involving isotope dilution. The trace at bottom demonstrates the homogeneity of the HPLC purified oligonucleotide.

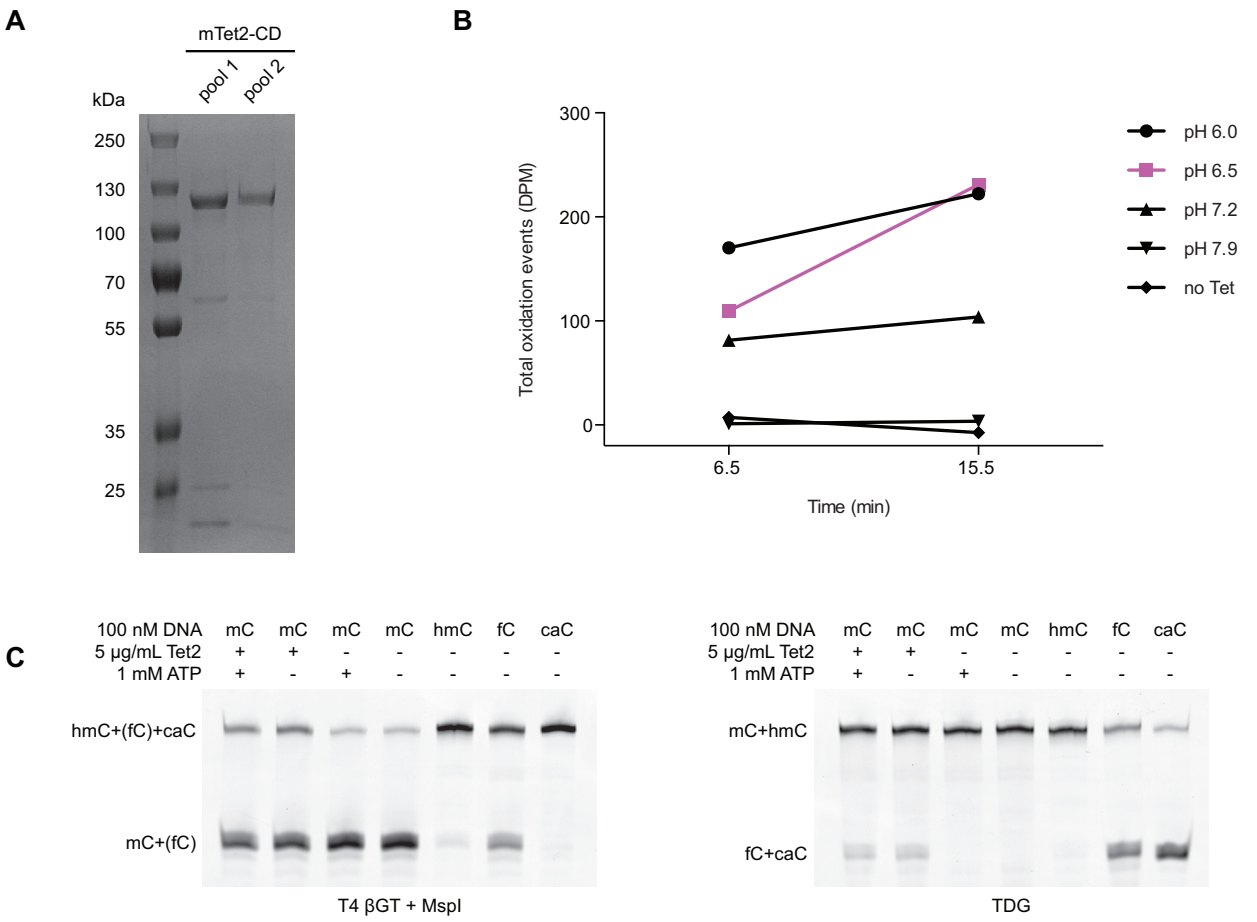

Figure S3: A) Coomassie-stained SDS-PAGE gel showing purified Tet2. Elution fractions from purification were combined into two pools. **B)** pH-dependence of Tet activity was assessed using 200 nM [^{14}C]-labeled mC substrate (Oligo3/Oligo 4 duplex) reacted with 4 µg/ml of Tet2 for 6.5 or 15.5 min. The total oxidation events are plotted as the disintegrations per minute (DPM) of hmC + 2*DPM of fC + 3*DPM of caC, after background subtraction of the controls without Tet2. At pH 6.5, the enzyme activity was linear with time. **C)** Under optimal reaction conditions, Tet2 was reacted with DNA in the presence or absence of 1 mM ATP. The products were analyzed by protection from MspI digestion (left gel). T4 βGT treatment protects hmC-containing DNA from cleavage by MspI, while caC is not cleaved and fC is a poor substrate (reaction with control oligonucleotides lanes 5-8). Additionally, fC and caC formation were detected by treatment with TDG (right gel). Both assays indicate that ATP does not significantly enhance activity (comparison of lanes 1 and 2) and additionally demonstrate that no products are detected in the absence of enzyme (lanes 3 and 4).

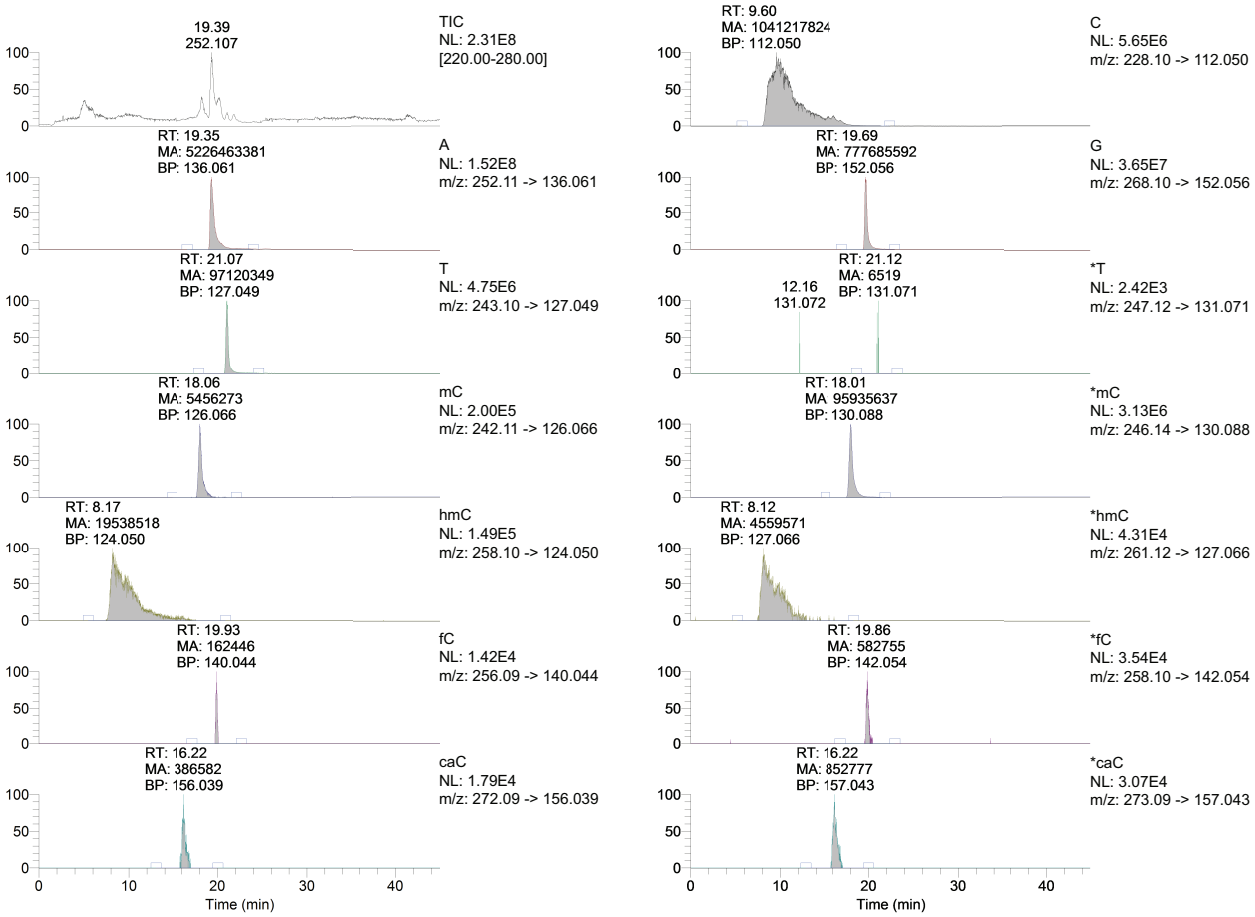

Figure S4: Representative mass chromatograms showing total ion count and the mass transitions for each of the degraded nucleosides in their light and heavy forms (* denotes heavy isotopes).