Supporting Information for

"Tet-mediated Formation of 5-Hydroxymethylcytosine in RNA"

by

Lijuan Fu, Candace R. Guerrero, Na Zhong, Nicholas J. Amato, Yunhua Liu, Shuo Liu, Qian Cai, Debin Ji, Seung-Gi Jin, Laura J. Niedernhofer, Gerd P. Pfeifer, Guo-Liang Xu and Yinsheng Wang

(*J. Am. Chem. Soc.*, 2014)

Materials and Methods

Materials

All chemicals and enzymes unless otherwise specified were from Sigma-Aldrich (St. Luis, MO). Erythro-9-(2-hydroxy-3-nonyl)adenine (EHNA) hydrochloride, [1,3-$^{15}N_2$]-cytidine, and 5-methylcytidine were obtained from Tocris Bioscience (Ellisville, MO), Cambridge Isotope Laboratories (Andover, MA), and Berry & Associates (Dexter, MI), respectively. The HEK293T human embryonic kidney cells, HeLa human cervical cancer cells, and WM-266-4 human melanoma cells, and cell culture reagents were purchased from ATCC (Manassas, VA, USA). All oligodeoxyribonucleotides and oligoribonucleotides were purchased from Integrated DNA Technologies (Coralville, IA). Expression vectors for the catalytic domain of human Tet1 (amino acids 1418-2136) and its corresponding mutant (H1672Y/D1674A) were previously described [1]. Expression vectors for the full-length mouse Tet1 and its corresponding mutant (H1620Y/D1622A) contained amino acids 1-2007. Expression vectors for the catalytic domain of human Tet2 and its catalytically inactive mutant (H1302Y/D1304A) contained amino acids 1129-2002, and those of the full-length human Tet2 and its catalytically inactive mutant (H1382Y/D1384A) contained amino acids 1-2002 [2]. Expression vectors for the catalytic domain of mouse Tet3 contained amino acids 697-1669, and those of the full-length mouse Tet3 and its corresponding mutant (H950Y/D952A) contained amino acids 1-1669.

Mouse embryonic stem (ES) cells and mouse ES cells with the depletion of all three *Tet* genes or *Tdg* gene were described elsewhere [3]. Mouse tissues used in the present study were from 19-21 week old mice. The human brain tissues used here were described previously [4].

Synthesis of [1, 3-$^{15}N_2$]-5-hydroxymethylcytidine

The titled compound was prepared at a microscale from [1,3-$^{15}N_2$]-cytidine following previously described procedures for the synthesis of isotope-labeled 5-hydroxymethyl-2'-deoxycytidine [5].

Biochemical assay of Tet1-mediated oxidation of 5-mrC in RNA and 5-mdC in DNA

The Tet1-mediated RNA oxidation assay was conducted with the use of the 5mC Tet1 Oxidation Kit (Wisegene, IL, USA), where a 12.5-μL reaction mixture contained 200 pmol of 5-mrC-bearing single-stranded RNA, 5'-UUUCAGCUC(5-mrC)GGUCACGCUC-3', the catalytic domain of mouse Tet1 (with amino acids 1367 to 2039) and reagents 1 and 2 included with the kit. The reaction was incubated at 37°C for 30 min and the enzyme was removed immediately afterwards by chloroform extraction. The reaction mixture was subsequently digested with nuclease P1 at 37°C for 4 hrs and then with alkaline phosphatase at 37°C for 2 hrs. The resulting ribonucleoside mixture was then subjected to HPLC analysis to determine the extent of conversion 5-mrC to 5-hmrC, and the HPLC conditions were described below in the HPLC enrichment section.

Tet1-mediated reactions were also performed using a 11mer 5-mrC-containing RNA, AGCUC(5-mC)GGUCA, and the reaction mixtures were analyzed directly by LC-MS and MS/MS to identify the reaction products and to quantify the levels of conversions of 5-mrC to 5-hmrC and 5-forC. To this end, 10 pmol of RNA was incubated with 0.125 μL of mTet1 protein, along with reagents 1 and 2 included with the kit in a total volume of 10 μL, at 37°C for the indicated periods of time. The reaction mixtures were immediately frozen on dry ince, and the enzyme in the mixture was subsequently removed using chloroform extraction. The volume of the aqueous layer was reduced by using a speed-vac, and the remaining aqueous solution was subjected directly to LC-MS analysis on an LTQ linear ion trap mass spectrometer (Thermo Fisher Scientific), where MS/MS and the higher-resolution "ultra-zoom-scan" MS were acquired for the $[M - 3H]^{3-}$ ions of the starting 11mer RNA as well as the corresponding RNA with the 5-mrC being oxidized to 5-hmrC, 5-forC, or 5-carC (Figure 2 and Figures S2-S3). The intensities of the monoisotopic peak and the +1 isotopic peak were employed for calculating the pecentages 5-mrC, 5-hmrC, and 5-forC in the reaction mixtures (5-carC-containing RNA was not detectable under the reaction conditions used). In this regard, the molecular weight of 5-hmrC and 5-forC differ by only 2 Da; thus, the +2 and +3 isotopic peaks of the 5-forC-bearing RNA overlap respectively with the monoisotopic and +1 isotopic peaks of the 5-hmrC-containing RNA. Therefore, we subtracted the intensities of +2 and +3 isotopic peaks of the 5-forC-harboring RNA from the intensities of the monoisotopic and +1 isotopic peaks of the 5-hmrC-containing RNA, considering that the relative abundances of the monoisotopic, +1, +2, and +3 isotopic peaks of the $[M - 3H]^{3-}$ ion of the 5-forC-carrying RNA (with elemental composition of $C_{107}H_{135}N_{41}O_{65}P_{10}$) follow the ratios of 73.9:100:76.8:42.9.

Similar reactions were performed for a single 5-mC-containing duplex DNA with the same sequence context, i.e., d(AGCTC(5-mC)GGTCA)/d(GTGACCGGAGCTG) under identical reaction conditions, and the levels of the conversions of 5-mdC to 5-hmdC, 5-fodC and 5-cadC were determined by LC-MS using similar procedures as described above for the single-stranded RNA.

Cell culture, transfection and RNA extraction

HEK293T cells were cultured in Dulbecco's Modified Eagle Medium (ATCC). HeLa and WM-266-4 cells were cultured in Eagle's Minimum Essential Medium (ATCC). All cells were incubated at 37°C in 5% CO_2 atmosphere. The culture medium was supplemented with 10% fetal

bovine serum and 100 IU/mL penicillin. The HEK293T cells were seeded in 6-well plates at 50-60% confluence level and, 24 hrs later, the cells were transfected with 1.5 μg plasmid for overexpressing the full-length (FL) or catalytic domain (CD), or the corresponding catalytically inactive mutants (FL-m and CD-m), of the three Tet family enzymes, except that 3.0 μg plasmids were employed for full-length Tet2 and its catalytically inactive mutant, using Lipofectamine 2000 (Invitrogen). Control experiments were also conducted by transfecting cells with 1.5 μg pGEM-T Easy plasmid (Pormega). The cells were harvested for RNA extraction 48 hrs after plasmid transfection. Total RNA was isolated from mammalian cells and tissues using TRI Reagent® according to the manufacturer's recommended procedures. The RNA pellet was dissolved in RNase-free water and its concentration measured by UV absorbance at 260 nm.

DNA was also isolated from the same cells and the levels of 5-hmdC in the DNA samples were measured using LC-MS/MS/MS, as described previously [4].

Enzymatic digestion of total RNA

One unit of nuclease P1, 0.00125 unit of phosphodiesterase 2, 2.5 nmol of EHNA and a 20-μL solution containing 300 mM sodium acetate (pH 5.6) and 10 mM zinc chloride were added to 10 μg of RNA, where EHNA was added to minimize the deamination of adenosine. The above digestion was continued at 37°C for 4 hrs. To the digestion mixture were then added 1 unit of alkaline phosphatase, 0.0025 unit of phosphodiesterase 1, and 40 μL of 0.5 M Tris-HCl buffer (pH 8.9). The digestion mixture was incubated at 37°C for 2 hrs. To the mixture was then added 200 fmol of [1,3-$^{15}N_2$]5-hmrC. The enzymes in the digestion mixture were subsequently removed by chloroform extraction. The resulting aqueous layer was dried, reconstituted in doubly distilled water, and subjected to off-line HPLC separation for the enrichment of the 5-hmrC.

HPLC enrichment

HPLC analysis was performed on a Beckman HPLC system with pump module 125 and a UV detector (module 126). A 4.6×250 mm Alltima HP C18 column (5 μm in particle size, Grace Davison, Deerfield, IL) was used. A solution of 10 mM ammonium formate (solution A) and methanol (solution B) were used as mobile phases, and the flow rate was 0.7 mL/min. A gradient of 42 min 0-0.5% B and 27 min 0.5-8% B was employed. A typical HPLC trace is depicted in Figure S6 of the Supporting Information. A minor DNA contamination (less than 5% for most RNA samples) was often observed for the total RNA isolated from mammalian cells and tissues based on the chromatograms obtained during the HPLC enrichment. Correction for DNA contamination was made for each RNA sample according to the peak areas of the 2'-deoxynucleosides relative to those of ribonucleosides. The HPLC fractions eluting at 12.5-16 min were pooled for 5-hmrC. The collected fractions were dried in the Speed-vac, redissolved in H_2O, and injected for LC-MS/MS/MS analysis.

Under the above HPLC conditions, we were able to resolve 5-mrC from all other ribonucleosides, which also allowed for the determination of the level of 5-mrC, in tissue and cellular RNA, based on the peak areas of mrC and rC as well as the extinction coefficients of these two nucleosides at 260 nm.

LC-MS/MS/MS Analysis of 5-hmrC

A 0.5 mm × 250 mm Zorbax SB-C18 column (particle size, 5 μm, Agilent) was used for the separation of the above-enriched 5-hmrC fraction, and the flow rate was 8.0 μL/min, which was delivered by using an Agilent 1100 capillary HPLC pump (Agilent Technologies). A solution of 0.1% (v/v) formic acid in water (solution A) and a solution of 0.1% (v/v) formic acid in methanol (solution B) were used as mobile phases, and a 30-min linear gradient of 0-70% B was employed. The effluent from the LC column was directed to an LTQ linear ion-trap mass spectrometer, which was set up for monitoring the fragmentation of the labeled and unlabeled 5-hmrC in the positive-ion mode. The temperature for the ion transport tube was maintained at 275 °C, the sheath gas flow rate was 15 arbitrary units, and no auxiliary gas was used. The electrospray, capillary, and tube lens voltages were 5 kV, 16 V, and 60 V, respectively; each MS^3 scan was composed of three microscans, and the maximum time for each microscan was 250 ms. The normalized collision energies were 42% and 27% in MS/MS and MS/MS/MS, respectively, and activation Q was 0.25 in both MS/MS and MS/MS/MS.

References:

(1) Guo, J. U.; Su, Y.; Zhong, C.; Ming, G. L.; Song, H. *Cell* **2011**, *145*, 423-34.

(2) Ko, M.; Huang, Y.; Jankowska, A. M.; Pape, U. J.; Tahiliani, M.; Bandukwala, H. S.; An, J.; Lamperti, E. D.; Koh, K. P.; Ganetzky, R.; Liu, X. S.; Aravind, L.; Agarwal, S.; Maciejewski, J. P.; Rao, A. *Nature* **2010**, *468*, 839-43.

(3) Hu, X.; Zhang, L.; Mao, S. Q.; Li, Z.; Chen, J.; Zhang, R. R.; Wu, H. P.; Gao, J.; Guo, F.; Liu, W.; Xu, G. F.; Dai, H. Q.; Shi, Y. G.; Li, X.; Hu, B.; Tang, F.; Pei, D.; Xu, G. L. *Cell Stem Cell* **2014**, *14*, 512-22.

(4) Liu, S.; Wang, J.; Su, Y.; Guerrero, C.; Zeng, Y.; Mitra, D.; Brooks, P. J.; Fisher, D. E.; Song, H.; Wang, Y. *Nucleic Acids Res.* **2013**, *41*, 6421-6429.

(5) LaFrancois, C. J.; Fujimoto, J.; Sowers, L. C. *Chem Res Toxicol* **1998**, *11*, 75-83.

(6) McLuckey, S. A.; Van Berkel, G. J.; Glish, G. L. *J. Am. Soc. Mass Spectrom.* **1992**, *3*, 60-70.

Table S1. The levels of 5-hmrC in RNA samples isolated from HEK293T cells overexpressing catalytic domain or full-length (FL) and their corresponding catalytically inactive mutants (CD-m and FL-m), or from HEK293T cells transfected with the control pGEM-T EASY vector (pGEM-T). The data represent the results from three independent transfection, enrichment, and LC-MS/MS/MS measurement results. "S.D." represents standard deviation.

		Replicate 1	Replicate 2	Replicate 3	Mean±S.D.
5-hmrC (modifications per 10^6 nucleosides)	pGEM-T	1.92	1.86	1.89	1.89±0.03
	Tet1-CD	10.89	11.83	13.06	11.9±1.1
	Tet1-CD-m	1.84	2.17	1.96	2.0±0.2
	Tet1-FL	1.93	1.53	1.61	1.7±0.2
	Tet1-FL-m	1.53	1.41	1.46	1.46±0.06
	pGEM-T	1.35	1.41	1.58	1. 5±0.1
	Tet2-CD	4.36	4.27	4.35	4.33±0.05
	Tet2-CD-m	1.58	1.22	1.43	1.4±0.2
	Tet2-FL	1.61	1.64	1.59	1.62±0.03
	Tet2-FL-m	1.47	1.35	1.28	1.37±0.03
	pGEM-T	1.94	1.45	1.53	1.6±0.3
	Tet3-CD	3.67	3.36	3.77	3.6±0.2
	Tet3-FL	4.27	4.24	3.87	4.1±0.2
	Tet3-FL-m	1.81	1.74	1.75	1.77±0.04

Table S2. The levels of 5-hmdC in DNA samples isolated from HEK293T cells overexpressing catalytic domain or full-length (FL) and their corresponding catalytically inactive mutants (CD-m and FL-m), or transfected with the control pGEM-T EASY vector (pGEM-T). The data represent the results from three independent transfection, enrichment, and LC-MS/MS/MS measurement results. "S.D." represents standard deviation.

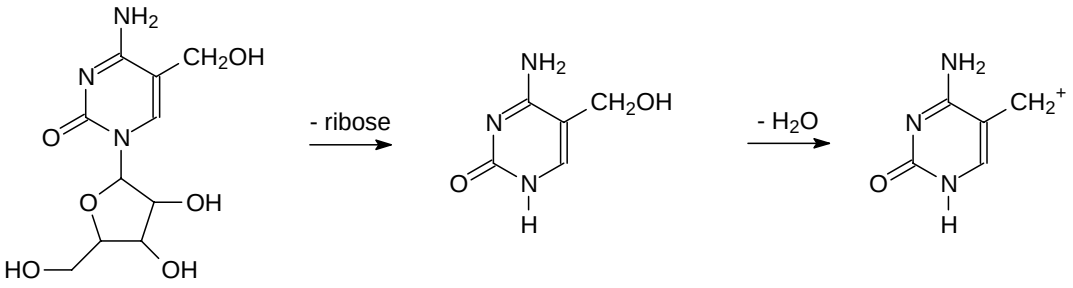
		Replicate 1	Replicate 2	Replicate 3	Mean ± S.D.
5-HmdC (modifications per 10^6 nucleosides)	pGEM-T	36.60	40.46	38.78	39 ± 2
	Tet1-CD	624.18	860.28	608.49	700 ± 140
	Tet1-CD-m	35.13	38.13	36.63	37 ± 2
	Tet1-FL	360.81	353.32	350.99	355 ± 5
	Tet1-FL-m	34.64	36.76	40.34	37 ± 3
	pGEM-T	15.18	14.94	12.99	14.4 ± 1.2
	Tet2-CD	149.33	162.23	197.60	170 ± 25
	Tet2-CD-m	14.92	17.88	18.48	17 ± 2
	Tet2-FL	132.21	134.56	150.35	139 ± 10
	Tet2-FL-m	31.14	29.45	26.31	29 ± 2
	pGEM-T	36.60	40.46	38.78	39 ± 2
	Tet3-CD	127.47	153.25	150.63	140 ± 10
	Tet3-FL	131.86	185.48	191.00	170 ± 30
	Tet3-FL-m	47.19	36.02	41.85	42 ± 6

Table S3. The levels of 5-hmrC and 5-mrC in RNA samples isolated from tissues of three animals or from three separate preparations of cells. "S.D." represents standard deviation.

		Replicate 1	Replicate 2	Replicate 3	Mean ± S.D.
5-HmrC (modifications per 10^6 nucleosides)	Mouse Pancreas	1.03	1.20	0.95	1.1 ± 0.1
	Mouse Spleen	2.16	1.88	2.07	2.0 ± 0.1
	Mouse Heart	4.27	3.15	4.21	3.9 ± 0.6
	Mouse Brain	2.47	2.56	2.65	2.56 ± 0.09
	Human Brain	1.45	1.47	1.39	1.44 ± 0.04
	WM-266-4	1.64	1.58	1.48	1.57 ± 0.08
	Hela	0.67	0.59	0.63	0.63 ± 0.04
	ES-WT	1.57	1.49	1.38	1.48 ± 0.09
	ES-Tet$^{-/-}$	0.83	0.79	0.85	0.82 ± 0.03
	ES-Tdg$^{-/-}$	1.51	1.56	1.36	1.5 ± 0.1
5-mrC (% of rC)	Mouse Pancreas	0.59	0.62	0.59	0.60 ± 0.02
	Mouse Spleen	0.90	0.77	0.88	0.85 ± 0.07
	Mouse Heart	2.02	2.09	2.27	2.1 ± 0.1
	Mouse Brain	0.99	1.04	1.05	1.03 ± 0.03
	Human Brain*	0.72	0.76	0.75	0.74 ± 0.02
	WM-266-4	0.70	0.74	0.72	0.72 ± 0.02
	Hela	0.47	0.47	0.47	0.47 ± 0.00
	ES-WT	1.08	1.06	1.09	1.08 ± 0.02
	ES-Tet$^{-/-}$	0.98	0.98	0.92	0.96 ± 0.03
	ES-Tdg$^{-/-}$	1.00	1.03	1.00	1.01 ± 0.01

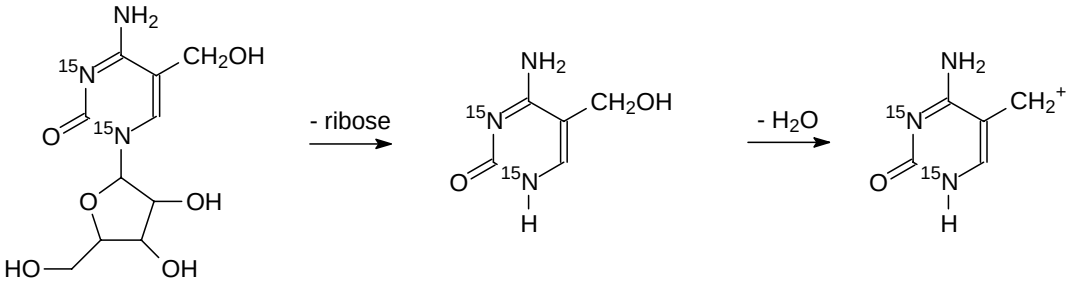
Scheme 1. Proposed structures of fragment ions observed in the MS/MS of the $[M + H]^+$ ions of 5-mrC, 5-hmrC and [1,3-$^{15}N_2$]-5-hmrC as well as MS/MS/MS of the $[M + H]^+$ ions of the nucleobase portions of 5-mrC, 5-hmrC and [1,3-$^{15}N_2$]-5-hmrC (spectra shown in Figure 1 c&d, Figure S1 and Figure S7).

a. 5-hmrC



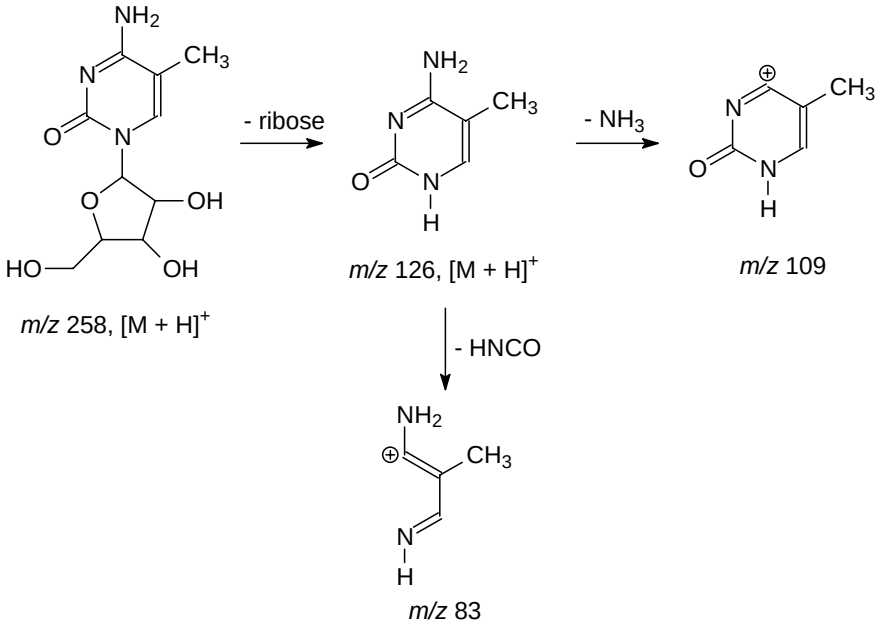
m/z 274, $[M + H]^+$ *m/z* 142, $[M + H]^+$ *m/z* 124

b. [1,3-$^{15}N_2$]-5-hmrC



m/z 276, $[M + H]^+$ *m/z* 144, $[M + H]^+$ *m/z* 126

c. 5-mrC



m/z 258, $[M + H]^+$ *m/z* 126, $[M + H]^+$ *m/z* 109

m/z 83

Figure S1. Positive-ion ESI-MS/MS (a) and MS/MS/MS (b) of the 5-mrC fraction from the HPLC separation of the nucleoside mixture of RNA isolated from the *in-vitro* Tet1-oxidation assay. The inset in (a) gives the high-resolution "ultra-zoom scan" MS for 5-mrC.

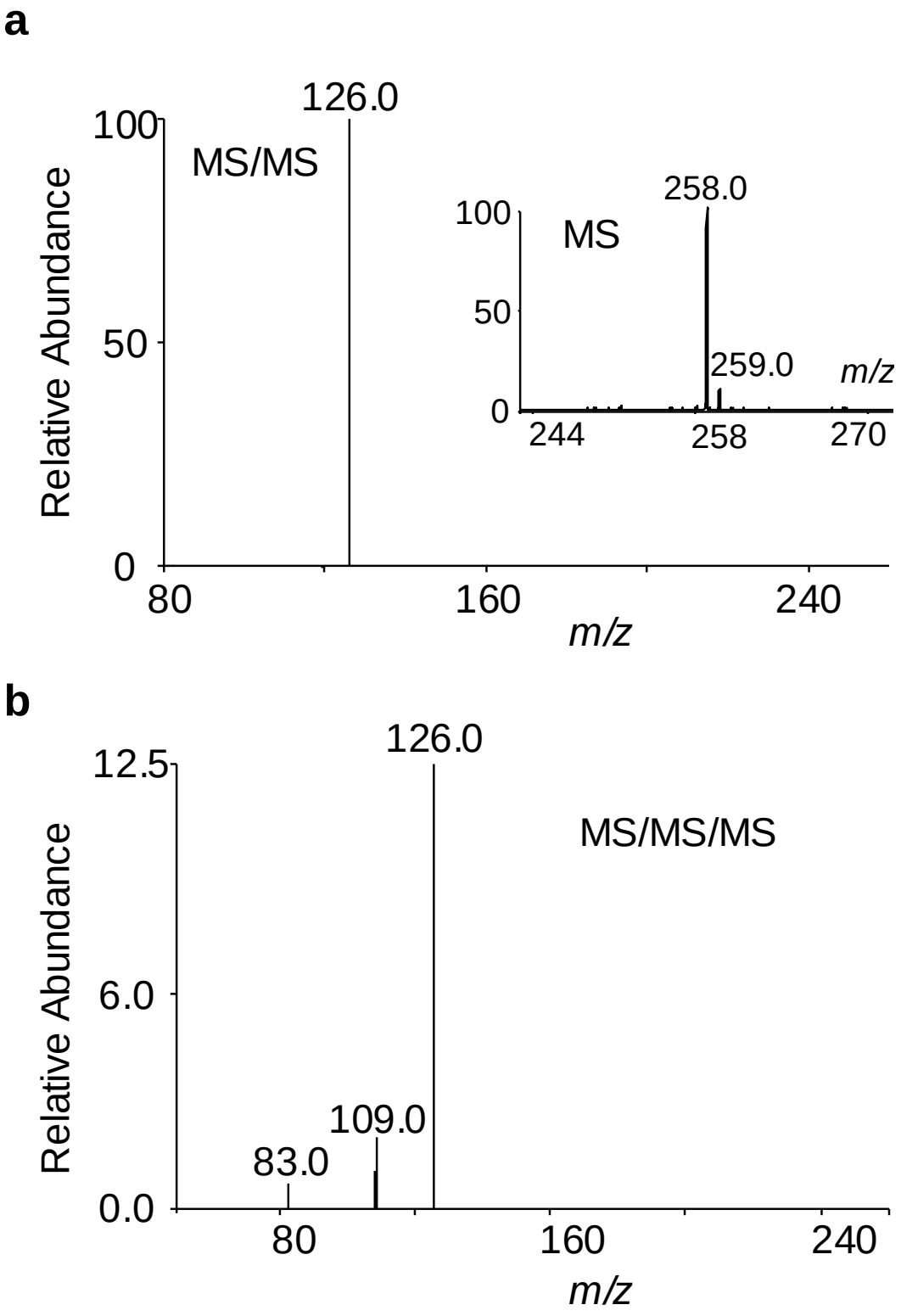


Figure S2. ESI-MS/MS for the $[M - 3H]^{3-}$ ions of AGCUCXGGUCA, where 'X' is a 5-mrC (a) or 5-hmrC (b) found in the Tet1-catalyzed reaction mixture of the 11-mer single-stranded RNA. The *m/z* values of fragment ions for RNA were calculated using the Mongo Oligo Mass Calculator v2.06 (http://mods.rna.albany.edu/masspec/Mongo-Oligo). The mass difference between the neighboring $[d_n - H_2O]$, w_n, or y_n ions defines the identity of the nucleotide flanked by the two neighboring ions. In particular, the mass difference between the w_5 and w_6 ions, between the y_5 and y_6 ions, or between $[d_5 - H_2O]$ and $[d_6 - H_2O]$ ions, corresponds to the residue mass of 5-mrC-5'-monophosphate (a) or 5-hmrC-5'-monophosphate (b).

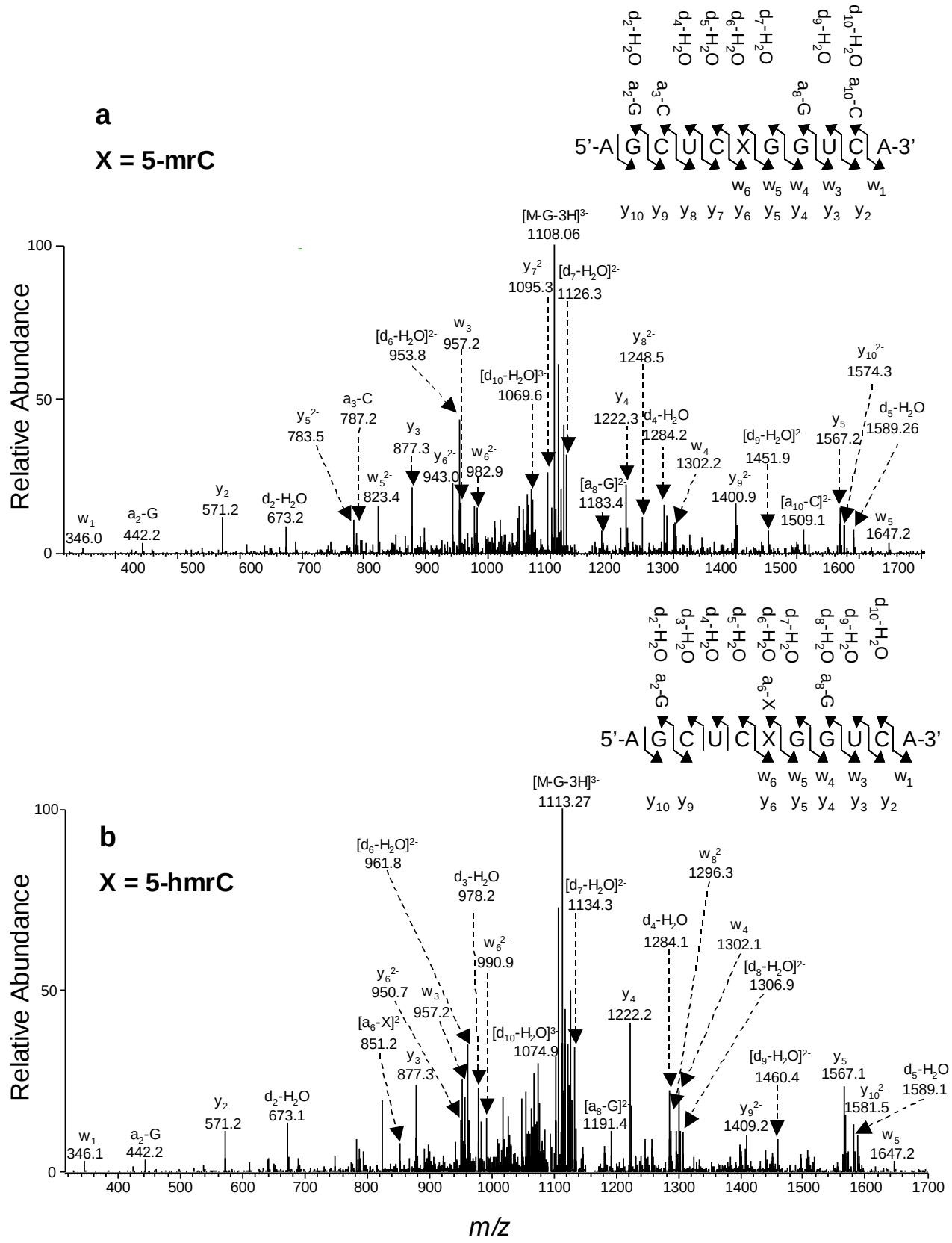

a
X = 5-mrC
5'-AGCUCXGGUCA-3'
Relative Abundance
b
X = 5-hmrC
5'-AGCUCXGGUCA-3'
Relative Abundance
m/z

Figure S3. ESI-MS/MS for the $[M - 3H]^{3-}$ ions of d(AGCTCXGGTCA) found in Tet1-catalyzed reaction mixture of the 11-mer duplex DNA, where 'X' is a 5-mdC (a), 5-hmdC (b), 5-fodC (c), or 5-cadC (d) (on next page). Collisional activation of deprotonated ions of ODNs led to the loss of nucleobases (A, C, or G) and subsequent cleavages of the 3' C-O bond of the same nucleoside to give $[a_n - Base]$ and its complementary w_n ions [6]; the mass difference between the neighboring $[a_n - Base]$ or w_n ions defines the identity of the nucleotide flanked by the two neighboring ions. For instance, the mass difference between the w_5 and w_6 ions, or between $a_6 - X$ and $a_7 - G$ ions, corresponds to the residue mass of 5-mdC-5'-monophosphate (a) or its corresponding oxidized derivatives (b-d).

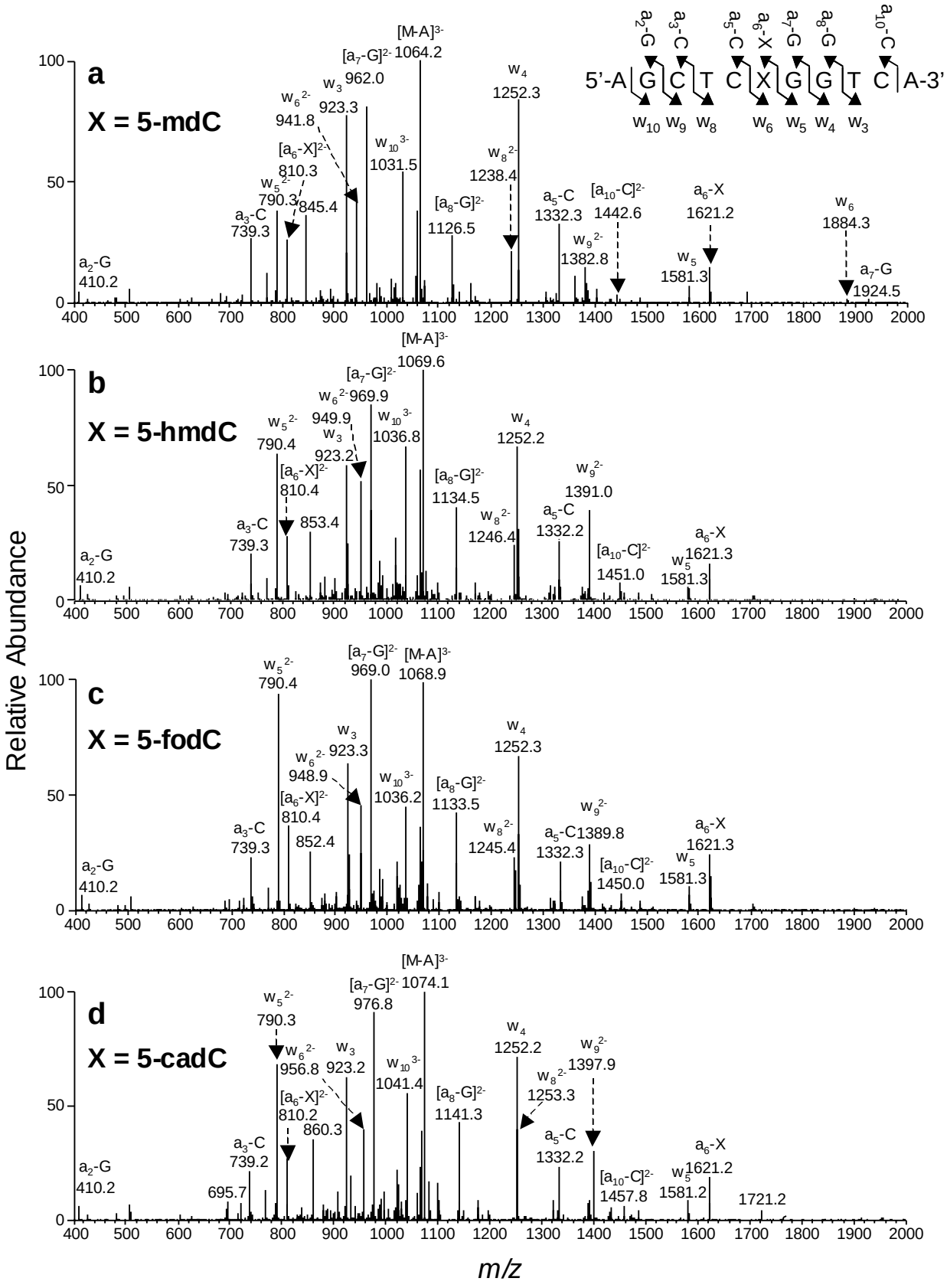

a
X = 5-mdC
b
X = 5-hmdC
c
X = 5-fodC
d
X = 5-cadC
5'-A G C T C X G G T C A-3'
Relative Abundance
m/z

Figure S4. LC-MS for monitoring the Tet1-mediated oxidation of 5-mrC in a single-stranded RNA, AGCUC(5-mrC)GGUCA in complete Tet1 reaction buffer (a), or in the same buffer without the addition of Fe^{2+} (b) or 2-oxoglutarate (c). Shown are the higher-resolution "ultra-zoom-scan" MS results for monitoring the $[M-3H]^{3-}$ ions of the initial 5-mC-bearing 11mer RNA, together with their oxidation products, where the 5-mrC is oxidized to 5-HmrC or 5-ForC.

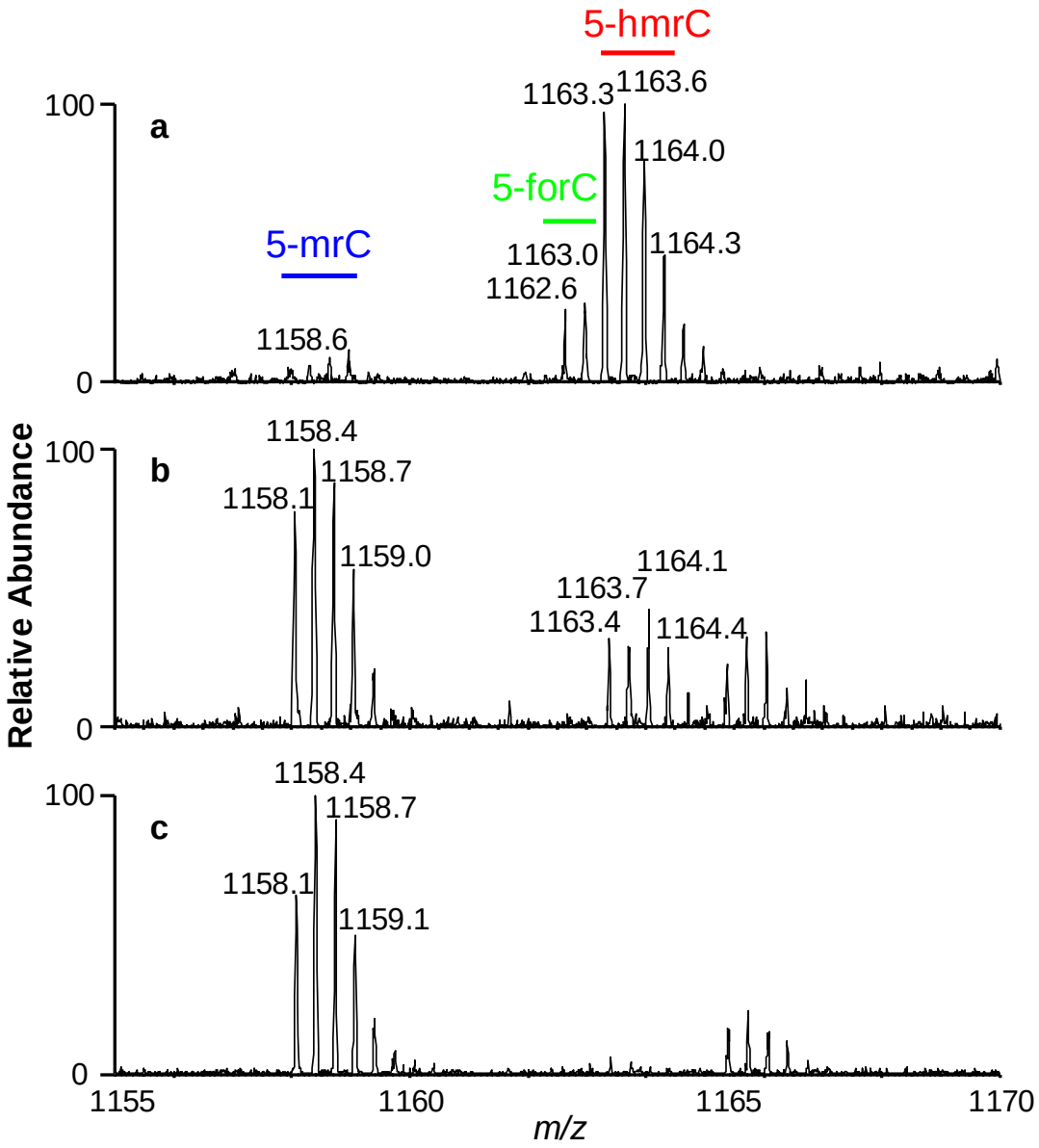


Figure S5. Time-dependent formation of oxidation products of 5-mrC in single-stranded RNA, AGCUC(5-mrC)GGUCA (a), and of 5-mdC in single-stranded DNA, d(AGCTC(5-mdC)GGTCA) (b). The products were quantified from LC-MS analyses using the similar method as described in Figure 2 of the main text.

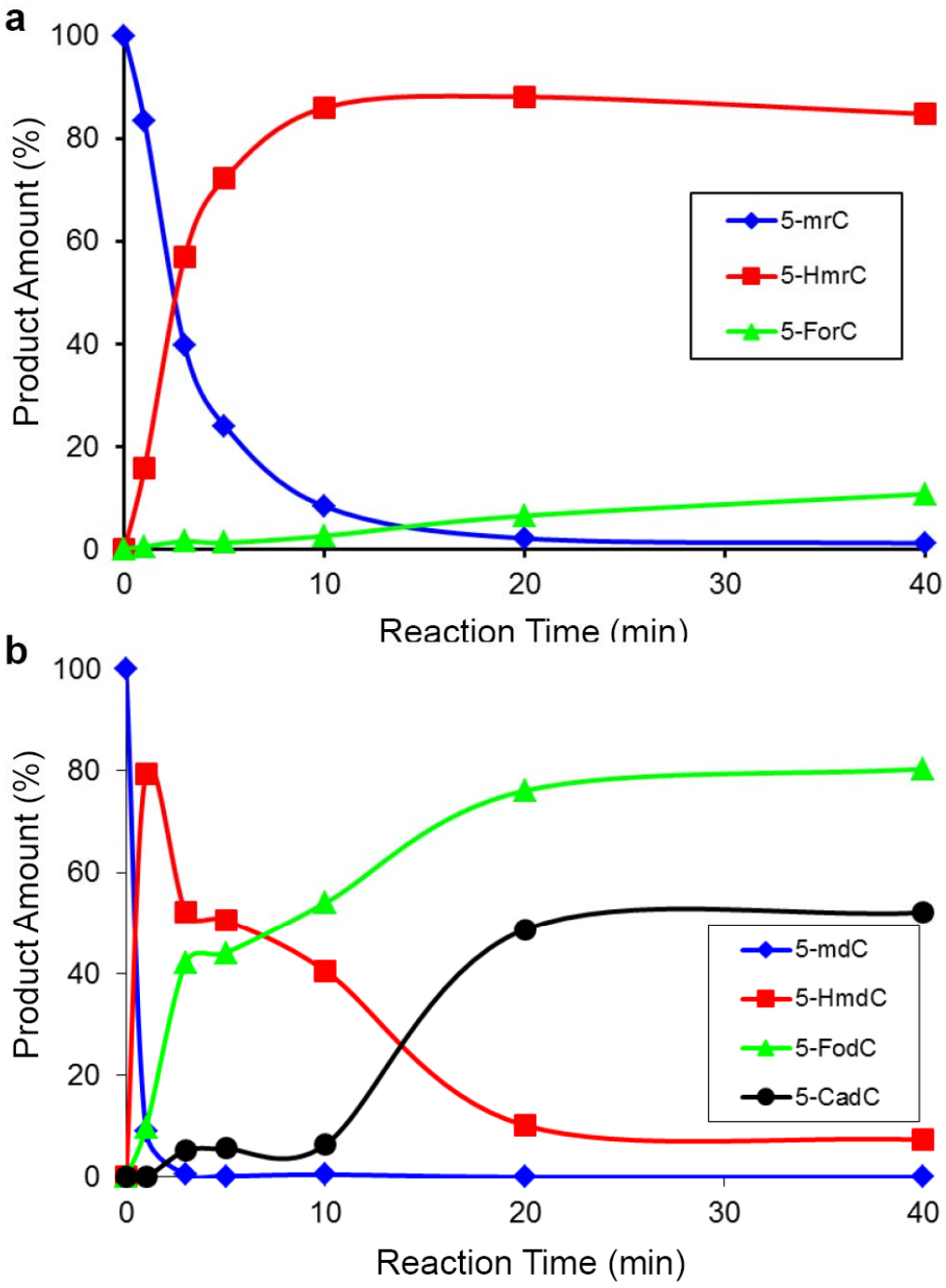


Figure S6. A representative HPLC trace for the enrichment of 5-hmrC and for the quantification of 5-mrC from the enzymatic digestion mixture of total RNA isolated from cells or tissues. Shown is the trace for the nucleoside mixture of a RNA sample isolated from mouse brain.

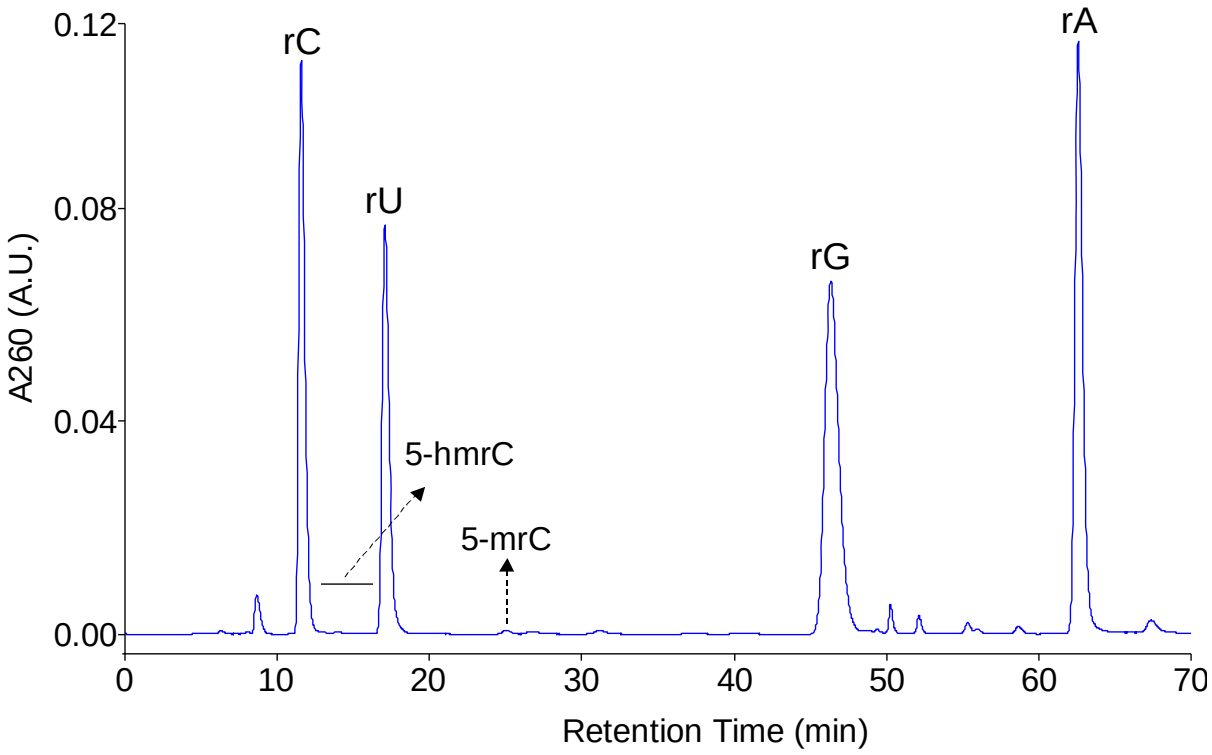


Figure S7. Representative LC-MS/MS/MS results for the quantification of 5-hmrC in cellular and tissue RNA. Shown are the selected-ion chromatograms for monitoring the indicated transitions for the analyte (a) and the isotope-labeled standard (b), and the insets give the corresponding MS/MS/MS for the analyte and internal standard. The RNA sample used for this analysis was from mouse brain.

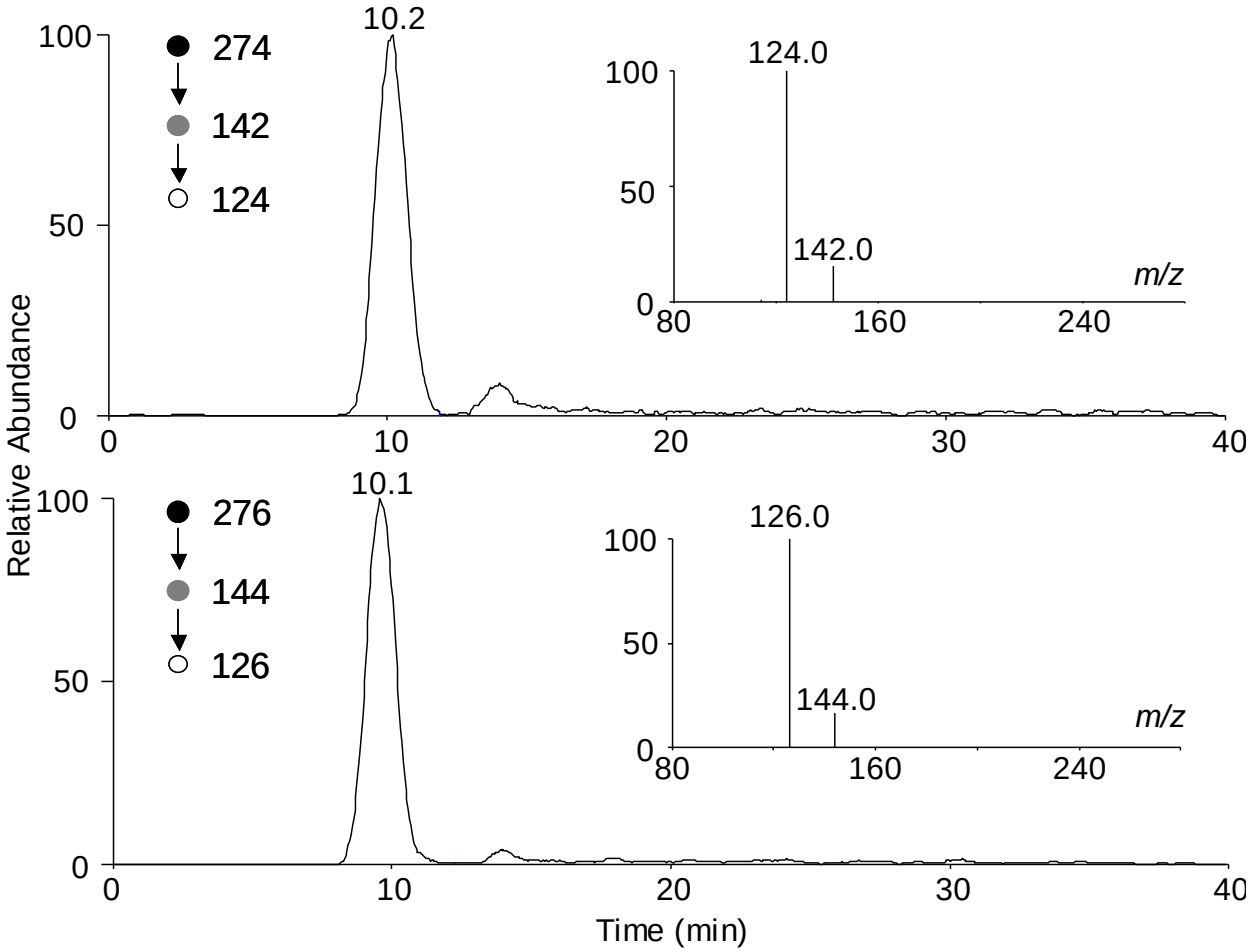


Figure S8. Calibration curve for the quantification of 5-hmrC in RNA. The amount of internal standard was 200 fmol, and the amount of unlabeled 5-hmrC ranged from 5 to 1280 fmol.

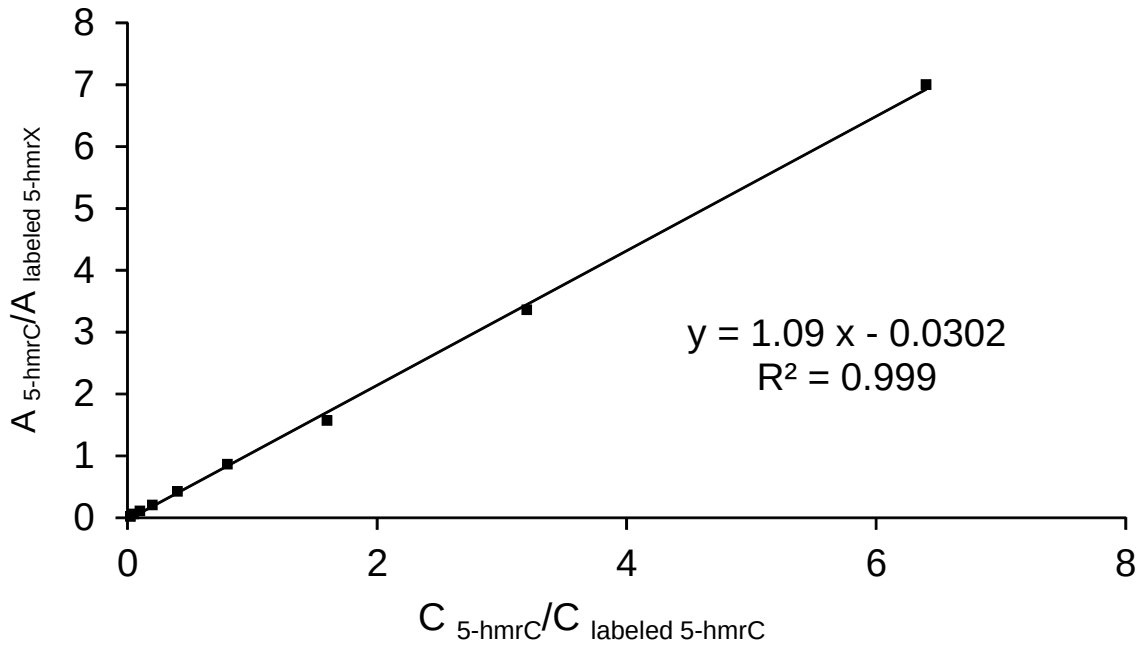


Figure S9. The levels of 5-hmrC and 5-hmdC in HEK293T cells overexpressing individually the catalytic domain (CD) of Tet proteins, or their catalytically inactive mutants (CD-m). 'pGEM-T' refers to DNA samples from HEK293T cells transfected with the control pGEM-T Easy plasmid. The data represent the means and standard deviations of three independent transfection and measurement results. The *p* values were calculated using unpaired two-tailed Student's *t*-test.

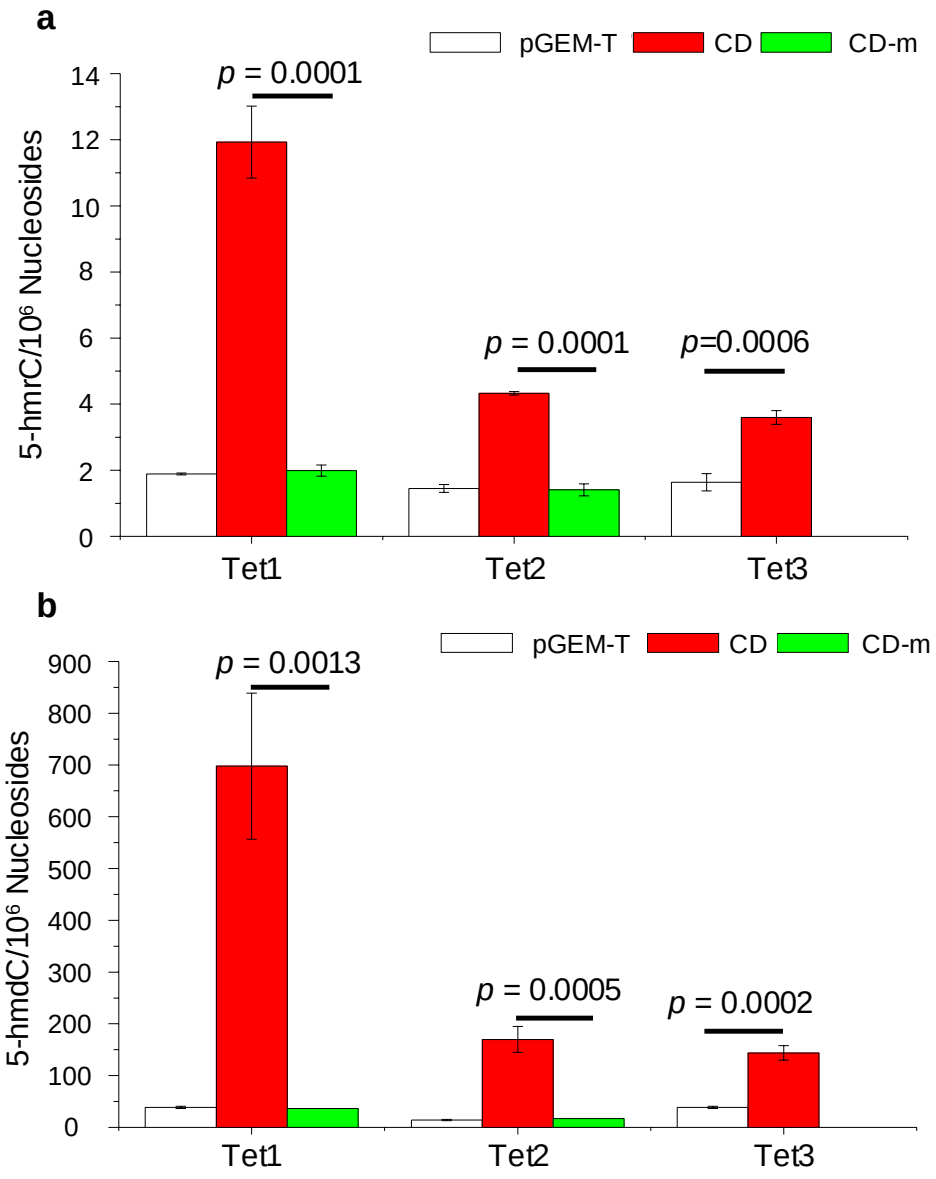


Figure S10. The levels of 5-hmrC in mammalian cells and tissues. Displayed are the quantification results for the levels of 5-hmrC in: (a) wild-type, *Tet*-null and *Tdg*$^{-/-}$ mouse ES cells; (b) HeLa and WM-266-4 cells; and (c) human brain and in different mouse tissues. The data represent the mean and standard deviation of the measurement results (n = 3). The *p* values in (a) were calculated using unpaired two-tailed Student's *t*-test.

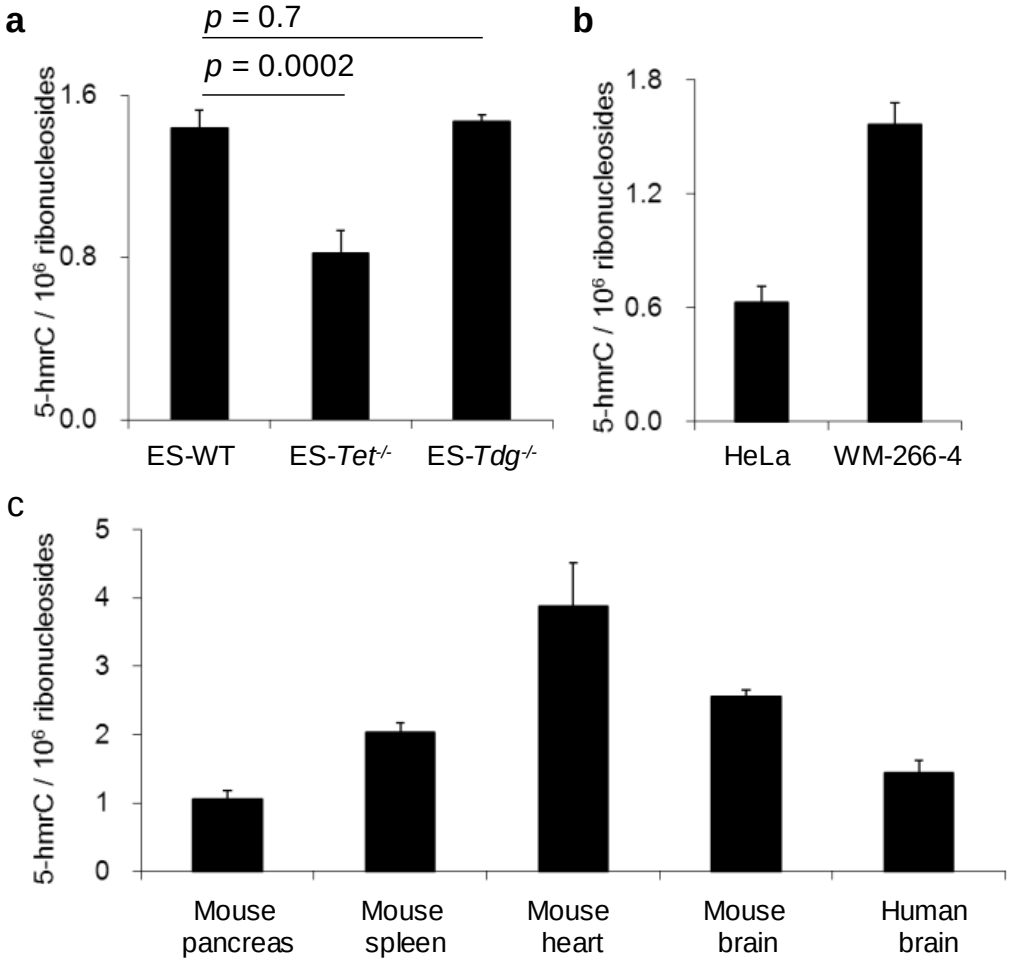


Figure S11. Quantification results for the level of 5-mrC in human brain and different mouse tissues (a, n = 3) as well as in cultured human cancer cells (b, n = 3). The level of 5-mrC was calculated from the HPLC peak areas of 5-mrC and rC with the consideration of the extinction coefficients of the two nucleosides at 260 nm.