DNA Repair Enzymes ALKBH2, ALKBH3, and AlkB Oxidize 5-Methylcytosine to 5-Hydroxymethylcytosine, 5-Formylcytosine, and 5-Carboxylcytosine *in Vitro*

Ke Bian,[1,±] Stefan A.P. Lenz,[2,±] Qi Tang,[1] Fangyi Chen,[1] Rui Qi,[1] Marco Jost,[3] Catherine L. Drennan,[3,4,6,7] John M. Essigmann,[3,5,6] Stacey D. Wetmore,[2,*] and Deyu Li[1,*]

[1]Department of Biomedical and Pharmaceutical Sciences, College of Pharmacy, University of Rhode Island, Kingston, Rhode Island 02881, United States

[2]Department of Chemistry and Biochemistry, University of Lethbridge, 4401 University Drive West, Lethbridge, Alberta T1K 3M4, Canada

[3]Department of Chemistry, [4]Department of Biology, [5]Department of Biological Engineering, [6]Center for Environmental Health Sciences, and [7]Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States

[±]The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

* To whom correspondence should be addressed. Deyu Li, Tel: +1(401) 874-9361; Email: deyuli@uri.edu; and Stacey D. Wetmore, Tel: +1(403) 329-2323; Fax: +1(403) 329-2057; Email: stacey.wetmore@uleth.ca.

Present Address: [Fangyi Chen], Xiamen University, 4221 Xiang An South Road, Xiang An District, Xiamen, Fujian 361102, P.R. China. [Marco Jost], Department of Cellular and Molecular Pharmacology, University of California, San Francisco, San Francisco, California 94158, United States.

Table of Contents

Table S1. Calculated and observed monoisotopic molecular weight and m/z value of modified oligonucleotides.

Table S2. Calculated and observed monoisotopic molecular weight and m/z value of peptide fragments digested by trypsin.

Table S3. Summary of important hydrogen bonds formed during MD simulations of *anti*-3mC, *anti*-5mC, *syn*-5mC, *syn*-5hmC, and *syn*-5fC in the ALKBH2 complex.

Table S4. Summary of important hydrogen bonds formed during MD simulations of *anti*-3mC, *anti*-5mC*, syn*-5mC, *syn*-5hmC, and *syn*-5fC in the AlkB complex.

Table S5. Summary of important distances and dihedral angles adopted during MD simulations of *anti*-3mC, *anti*-5mC*, syn*-5mC, *syn*-5hmC, and *syn*-5fC in the ALKBH2 complex.

Table S6. Summary of important distances and dihedral angles adopted during MD simulations of *anti*-3mC, *anti*-5mC*, syn*-5mC, *syn*-5hmC, and *syn*-5fC in the AlkB complex.

Table S7. Time to half maximum conversion ($T_{1/2}$, min) for the oxidation of different alkyl DNA modifications by the three AlkB family enzymes.

Table S8. Comparison of important distances and dihedral angles adopted during MD simulations of *anti*-3mC, *anti*-5mC, and *syn*-5mC in the ALKBH2 complex.

Table S9. Comparison of important distances and dihedral angles adopted during MD simulations of *syn*-5hmC *and syn*-5fC in the ALKBH2 complex.

Table S10. Comparison of important distances and dihedral angles adopted during MD simulations of *anti*-3mC, *anti*-5mC, and *syn*-5mC in the AlkB complex.

Table S11. Comparison of important distances and dihedral angles adopted during MD simulations of *syn*-5hmC *and syn*-5fC in the AlkB complex.

Product oligonucleotides analyses

It was important to assure that the observed new oligonucleotide species indeed contain the proposed oxidative products. The product oligonucleotides were digested with benzonase into single nucleosides, which were analyzed by LC-MS and compared with standard nucleosides for retention time and molecular weight. In the digestion of the product generated from the AlkB reaction, a nucleoside was eluted out at 7.3 min in HPLC (Figure S2) and had an observed m/z 240.2386 in the negative mode of MS analysis (theoretical m/z 240.2392, Figure S4); these data agree well with the analysis of the standard sample of 5mC nucleoside (7.3 min in HPLC and 240.2385 in m/z, Figure S3). Another nucleoside from digestion had a retention time at 4.9 min in HPLC (Figure S2) and had an m/z 256.2387 in the MS analysis (Figure S6); these data are comparable to the standard sample of 5hmC nucleoside (5.0 min in HPLC and 256.2409 in m/z, theoretical m/z 256.2382, Figures S2 and S5). Similarly, nucleosides 5fC and 5caC were also discovered in the mixture of the digested reaction product and their identities were confirmed by comparing to standard nucleosides (Figures S2, S7 to S10). These observations support that the oxidative products generated from enzymatic reactions are indeed 5hmC, 5fC, and 5caC.

Product distribution for the oxidation of 5mC

For the oxidation of 5mC, the formation of products 5hmC, 5fC, and 5caC had different distribution patterns for the three enzymes; and the three enzymes had different preferences of oxidation in ss- or ds-DNA reactions (Figure 3). The relative amount of each product in the final reaction mixture was quantified according to the abundance of its corresponding ion at -3 charge state in the MS analysis. The product oligonucleotides containing 5hmC was the dominant species for the reactions of ALKBH2 and 3 in both ds- and ss-DNA (Figures 3a and 3b). For example, 5hmC represents 78.7% of the total amount of the three oxidative products in the reaction of ALKBH2 with 5mC in ds-DNA; 5fC and 5caC are quantified as 10.9% and 10.4% correspondingly (Figure 3a). For the oxidation of 5mC by AlkB, 5hmC is the major species (51.2%) in ds-DNA reaction; however, 5fC is the most abundant species (65.3%) in ss-DNA reaction (Figures 3a and 3b). We were also able to quantify the yield of the oxidative products by comparing them to the total oligonucleotide species. The ratios of the oxidative products show ALKBH2 prefers to oxidize 5mC in ds-DNA (8.7%) over in ss-DNA (5.5%), and ALKBH3 prefers to modify 5mC in ss-DNA (22.1%) over in ds-DNA (4.2%) (Figure 3c). These results support the strand preference in repair reactions of ALKBH2 (preferring ds-DNA) and ALKBH3 (preferring ss-DNA) reported in the literature.(1) For AlkB, the enzyme oxidized 5mC in ss-DNA (23.5%) more efficiently than in ds-DNA (13.3%) (Figure 3c); these results are similar to the preference of AlkB repairing 3mC in ss-DNA.(2)

Modifications on different alkyl substrates by the AlkB enzymes

We compared the conversion efficiency of 5mC by the three AlkB family enzymes to other known methyl substrates of the AlkB proteins. Those substrates include 3mC, 1mA, 3mT, and 1mG; and all of them have been demonstrated to be repaired by AlkB both *in vitro* and *in vivo*.(1, 3) The results show the conversion of 5mC to the corresponding oxidative products are comparable to the

demethylation of 3mT and 1mG; and the reactions of these three methyl modifications are slower than 3mC and 1mA (Figure S23 and Table S7). For example, $T_{1/2}$ of the reactions for 5mC (16.9 min) is slightly shorter than 3mT (17.4 min) but longer than 1mG (1.7 min) in the AlkB reactions. For the ALKBH2 reactions, $T_{1/2}$ for 5mC are comparable to 3mT and 1mG (all between 16.4 to 18.7 min). For the ALKBH3 protein, 5mC is oxidized faster than 3mT and 1mG with about 5 minutes shorter in $T_{1/2}$ than the other two (Figure S23 and Table S7). Those biochemical results may indicate similar modification could happen in cell: the modification of 5mC is comparable to the weaker AlkB substrates 3mT and 1mG but less efficiently than the stronger substrates 3mC and 1mA.(3)

Simulations of AlkBH2 and AlkB bound to 3mC, 5mC, 5hmC, and 5fC

To validate that the simulated structures of the ALKBH2 or AlkB complexes are consistent with the observed catalytic activity, we examined the active site conformations adopted upon binding of both enzymes to *anti*-3mC- *syn*-5mC-, *syn*-5hmC-, or *syn*-5fC-containing DNA. Specifically, we performed two 100 ns simulations with different initial velocities on each system, and extended one replicate to 500 ns. For each simulation on the same system, the measured distances and angles between the substrate and key active site features are within ~0.3 Å and ~40° (Table S8-S11). Therefore, only the data from the 500 ns production simulations are discussed below and in the main text.

Similar DNA–protein interactions form when 3mC and 5mC are bound. Specifically, the 3mC N^4 amino group forms direct hydrogen bonds with the D174 or E175 sidechains when bound to ALKBH2 (Figure S16b and Table S3), and direct or water-mediated hydrogen bonds with D135 and E136 when bound to AlkB (Table S4 and Figure S17). As a result, the N3 methyl group of 3mC occupies an equivalent active site position as the C5 methyl group of 5mC for both enzymes (Figures S16b and S17b), with a distance between the Fe(IV)–oxo and the 3mC methyl moiety of ~ 3.3 Å for both enzymes (Tables S5-S6), and the *syn*-5mC moiety of ~3.6 Å for AlkB and 3.7 Å for ALKBH2. Thus, the simulation data is consistent with the proposal that the AlkB family of enzymes is able to oxidize 5mC in the *syn* orientation.

Similar to ALKBH2 or AlkB bound to 5mC, *syn*-5hmC is stabilized by direct or water-mediated hydrogen bonds between the N^4 amino group of the nucleobase and the carboxylate moieties of D174 and E175 (D135 and E136 in AlkB; Tables S3-S4 and Figure S20). For 5hmC bound by ALKBH2 or AlkB, hydrogen bonds are formed between the C5 substituent hydroxy group and the Fe(IV)-oxo moiety (occupancy = 14% for ALKBH2 and 64% for AlkB; Tables S3-S4 and Figure S21). In ALKBH2, an additional hydrogen bond is formed between the C5 substituent hydroxy group and D173 (59%, Table S3), which results in one orientation of the C5 substituent throughout the simulation (Figure S21a). In contrast, the C5 substituent hydroxy group does not interact with D133 in AlkB, which results in two conformations of 5hmC within the active site (Figure S21b-d). More importantly, the C5 substituent is at an optimum distance from the Fe(IV)-oxo moiety in AlkB (~3.4 Å), which matches the prototypic substrate (3mC; ~3.4 Å), while the equivalent distance is longer in the ALKBH2–5hmC complex (~4.5 Å; Tables S5-S6 and Figure 3), which would impede catalysis. This helps explain the

observed higher abundance of 5fC for AlkB compared to ALKBH2 catalyzed oxidation on 5hmC, although several other factors could also be significant such as DNA binding and unique base flipping mechanism for each enzyme.

As discussed for 5mC and 5hmC, the *syn*-conformation of 5fC is stabilized by hydrogen bonds between the N^4 amino group of the nucleobase and carboxylate sidechains of active site residues (Tables S3-S4 and Figure S20). When bound to either ALKBH2 or AlkB, 5fC is planar due to an intramolecular hydrogen bond between the N^4 amino group and the carbonyl of the C5 substituent. A hydrogen bond also forms between the carbonyl of the C5 substituent and an active site arginine, which is notably more persistent for AlkB (R210, occupancy = 100.0%) than ALKBH2 (R254, occupancy = 20.3%, Tables S3-S4 and Figure S20). As a result, the distance between the hydrogen atom of the C5 substituent and the Fe(IV)-oxo group is longer for AlkB (~3.6 Å) compared to ALKBH2 (3.3 Å; Tables S5-S6 and Figure S22). Although the difference in distance for 5fC bound by ALKBH2 and AlkB is not as significant as observed for 5hmC, QM/MM studies on TET2-catalyzed oxidation of 5mC, 5hmC, and 5fC reveal that the initial hydrogen atom abstraction step is rate limiting, with the barrier increasing as 5mC < 5hmC < 5fC.(4) This suggests that the position of the substituent relative to the Fe(IV)-oxo group is even more crucial in the case of 5fC. Thus, our predicted structures correlate with the lower abundance of 5caC relative to 5fC for AlkB compared to ALKBH2 (Figure 3). Nevertheless, as discussed for 5hmC, nucleotide recognition and the base flipping mechanism could also play different roles in oxidative conversion of 5fC to 5caC.

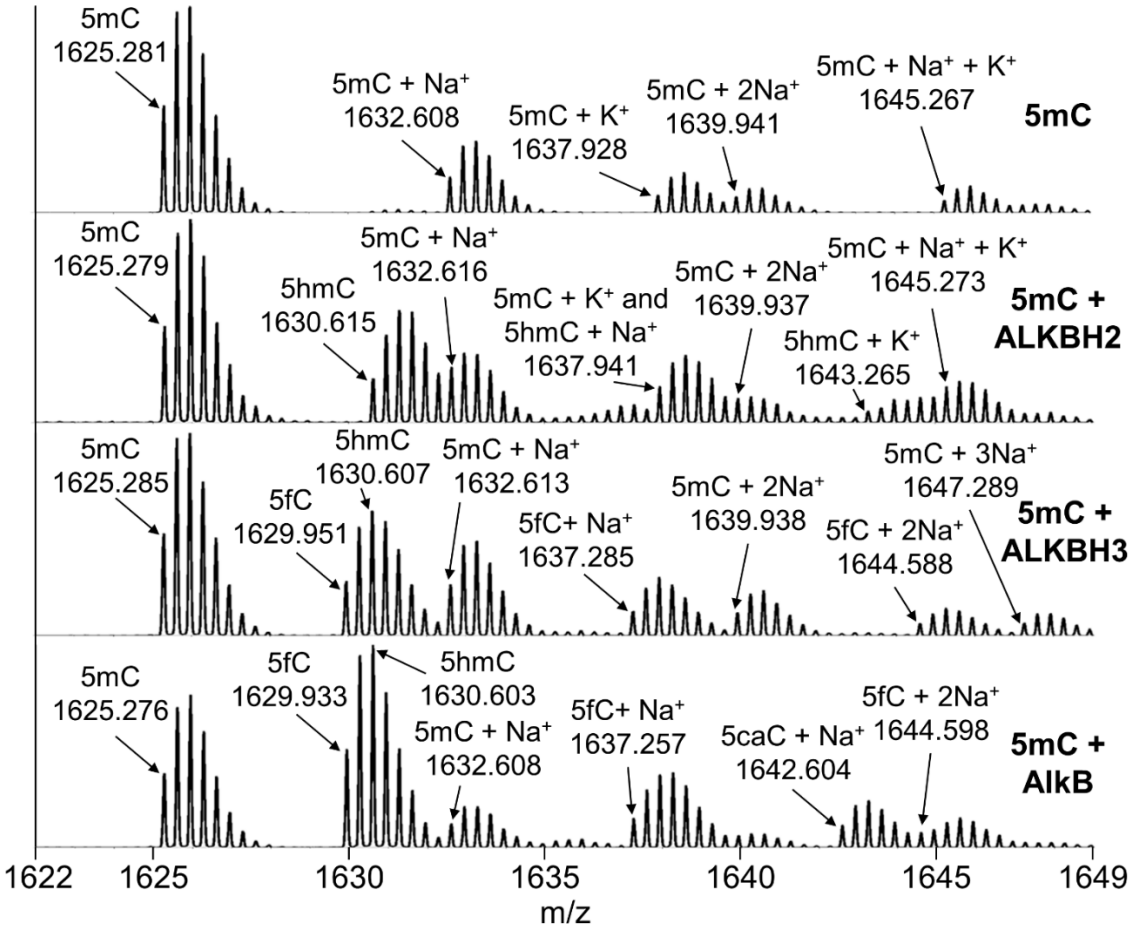

Figure S1. ESI-TOF analyses of standard oligonucleotide containing 5mC and reaction mixtures. The observed m/z values represent the oligonucleotides under their -3 charge state. The theoretical m/z values of the corresponding species are listed in Table S1.

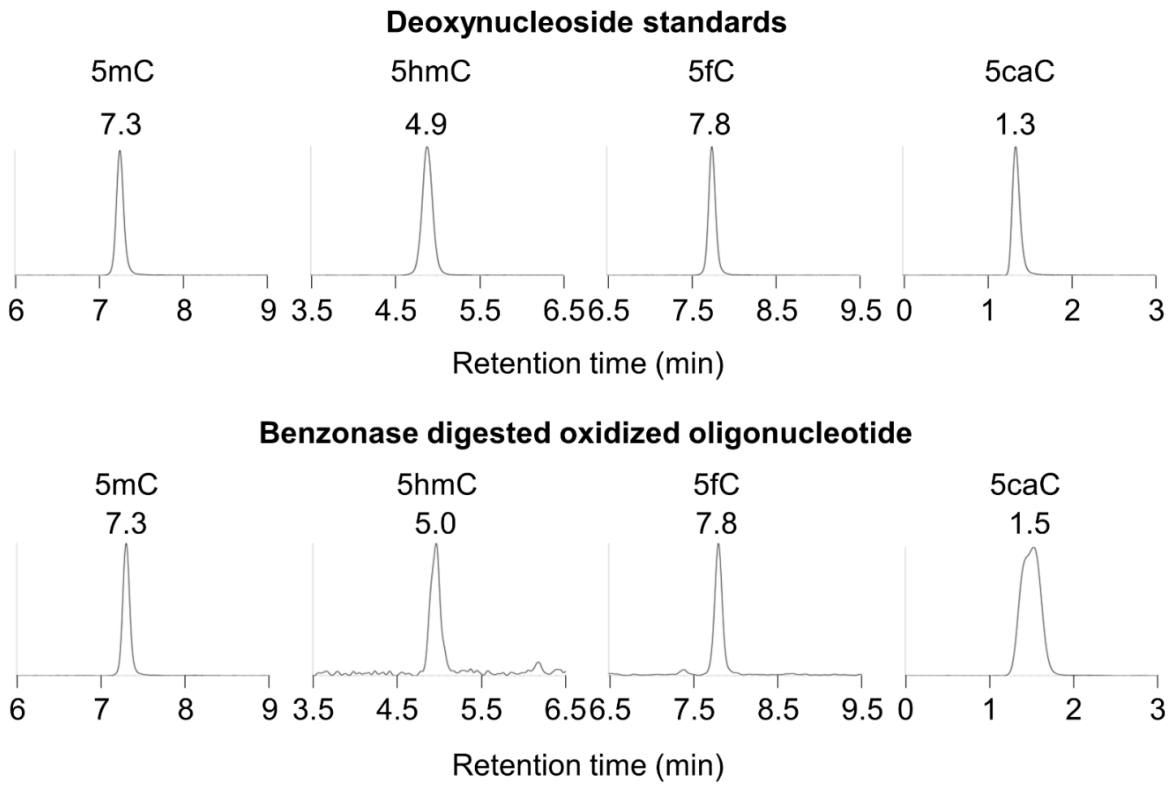

Figure S2. HPLC analyses of deoxyribonucleoside standards and benzonase digested product oligonucleotides.

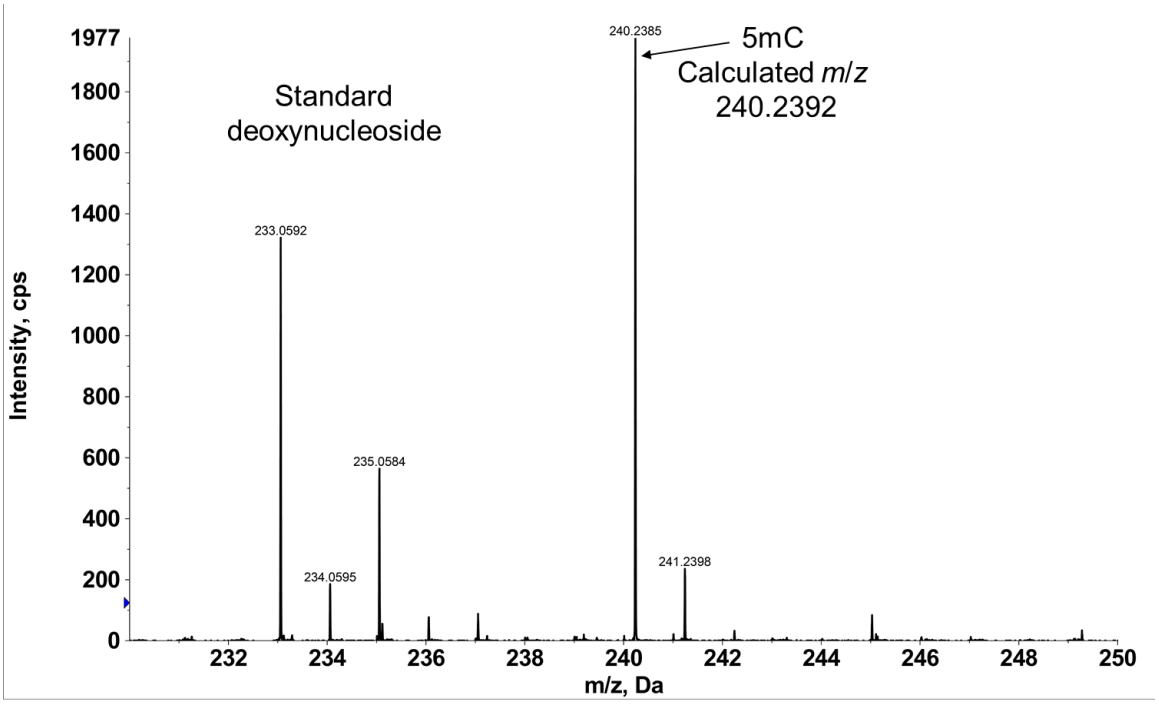

Figure S3. MS analysis of standard 5mC deoxyribonucleoside under negative ion mode. The theoretical m/z value of the corresponding species is listed next to the observed peak.

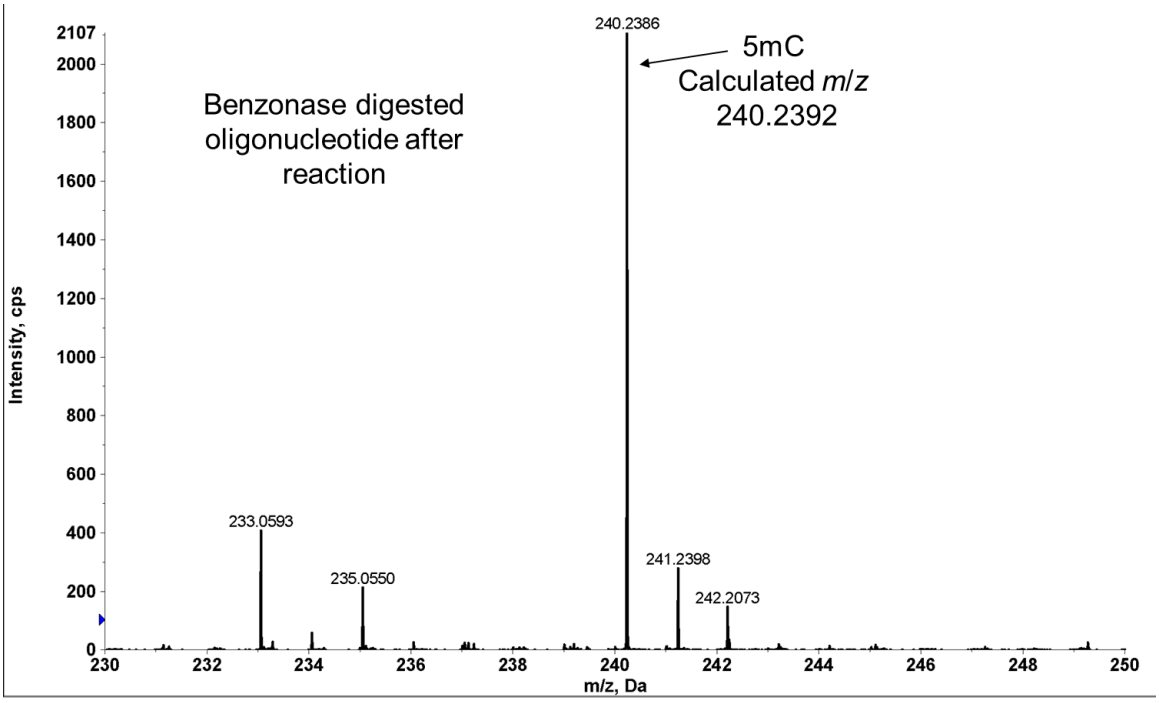

Figure S4. MS analysis of digested oligonucleotide containing 5mC deoxyribonucleoside under negative ion mode. The theoretical m/z value of the corresponding species is listed next to the observed peak.

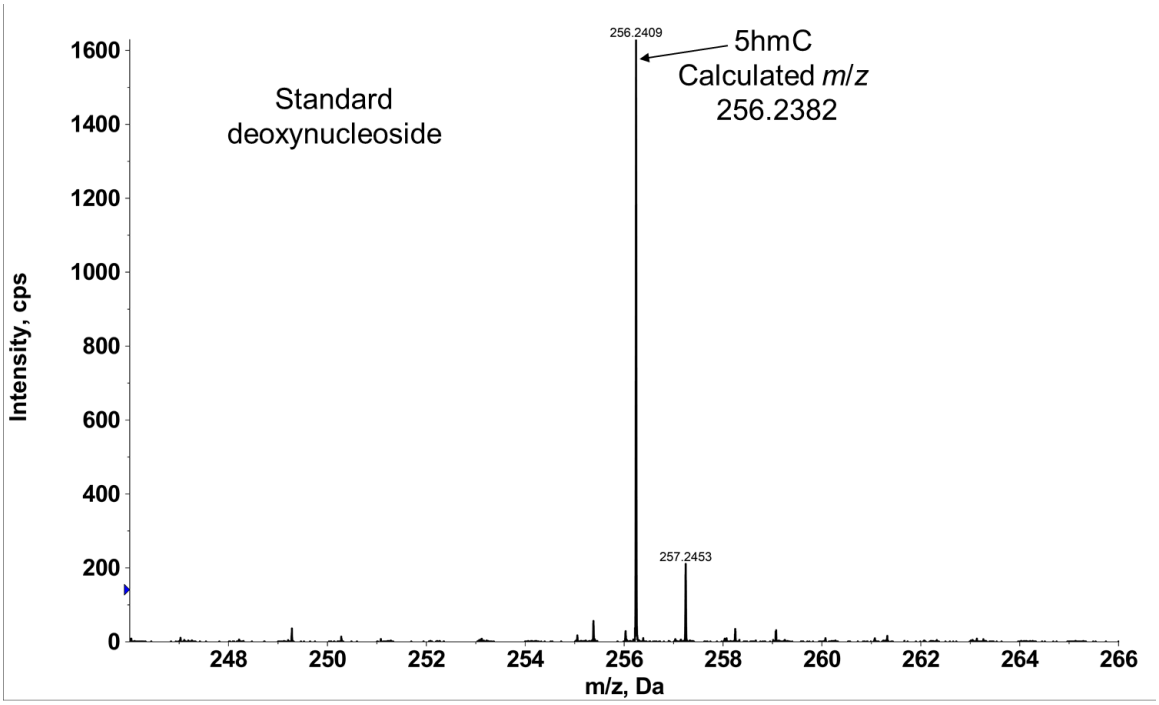

Figure S5. MS analysis of standard 5hmC deoxyribonucleoside under negative ion mode. The theoretical m/z value of the corresponding species is listed next to the observed peak.

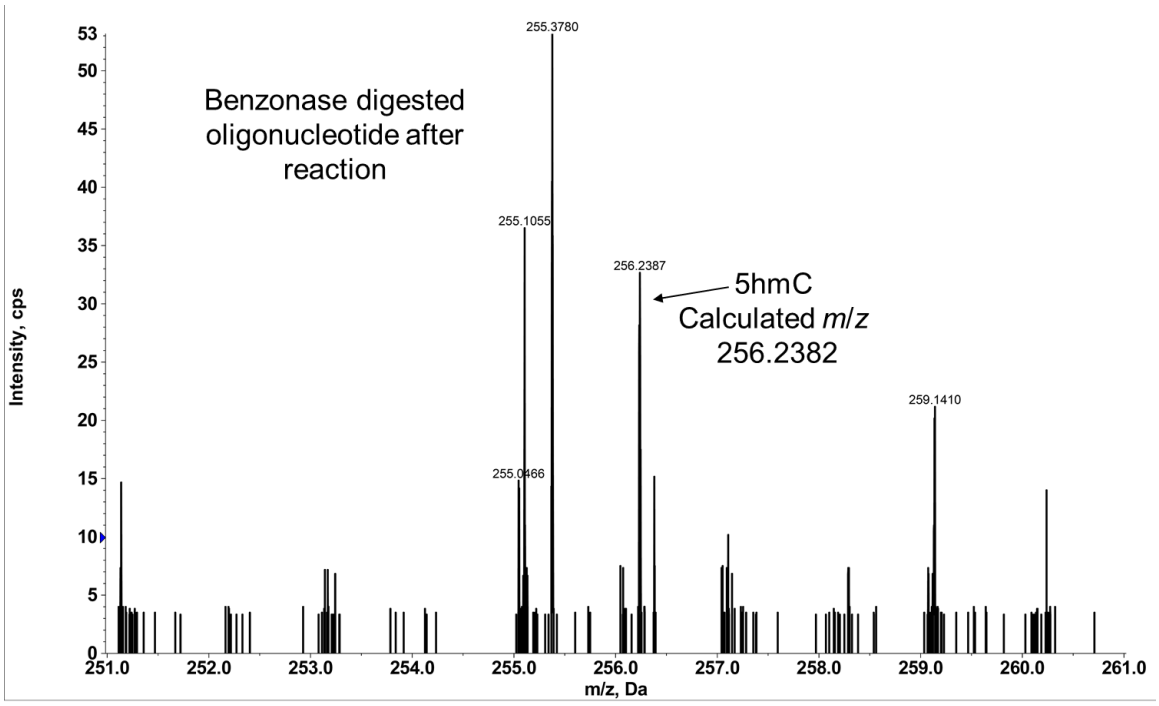

Figure S6. MS analysis of digested oligonucleotide containing 5hmC deoxyribonucleoside under negative ion mode. The theoretical m/z value of the corresponding species is listed next to the observed peak.

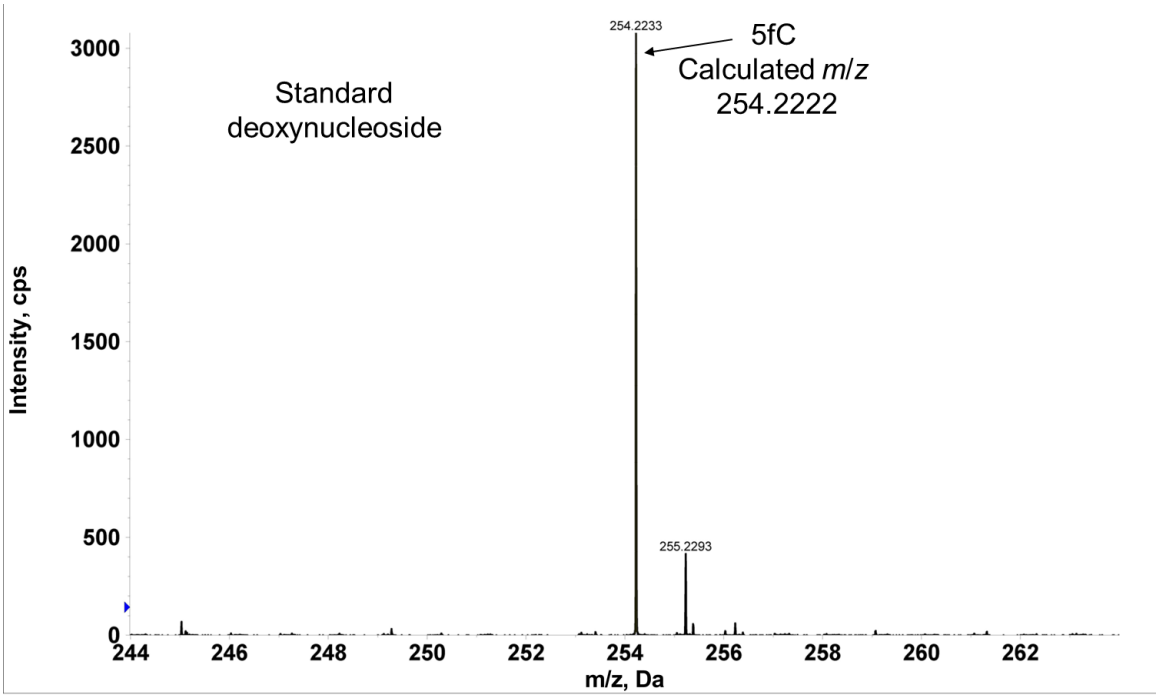

Figure S7. MS analysis of standard 5fC deoxyribonucleoside under negative ion mode. The theoretical m/z value of the corresponding species is listed next to the observed peak.

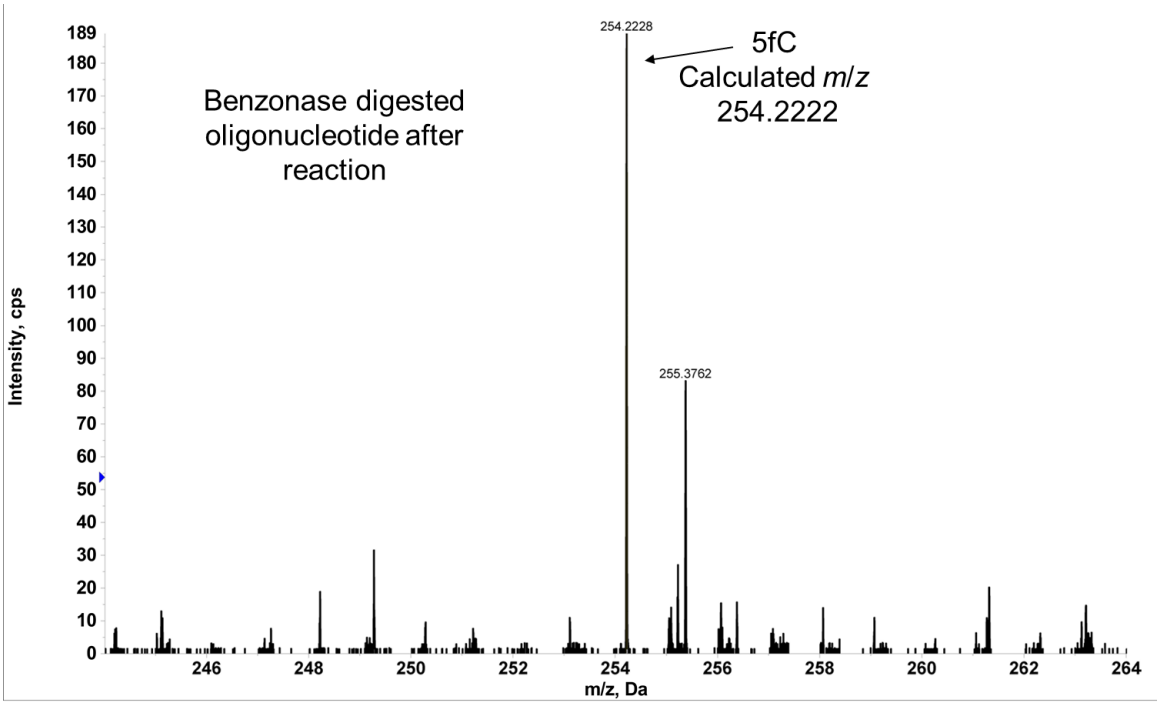

Figure S8. MS analysis of digested oligonucleotide containing 5fC deoxyribonucleoside under negative ion mode. The theoretical m/z value of the corresponding species is listed next to the observed peak.

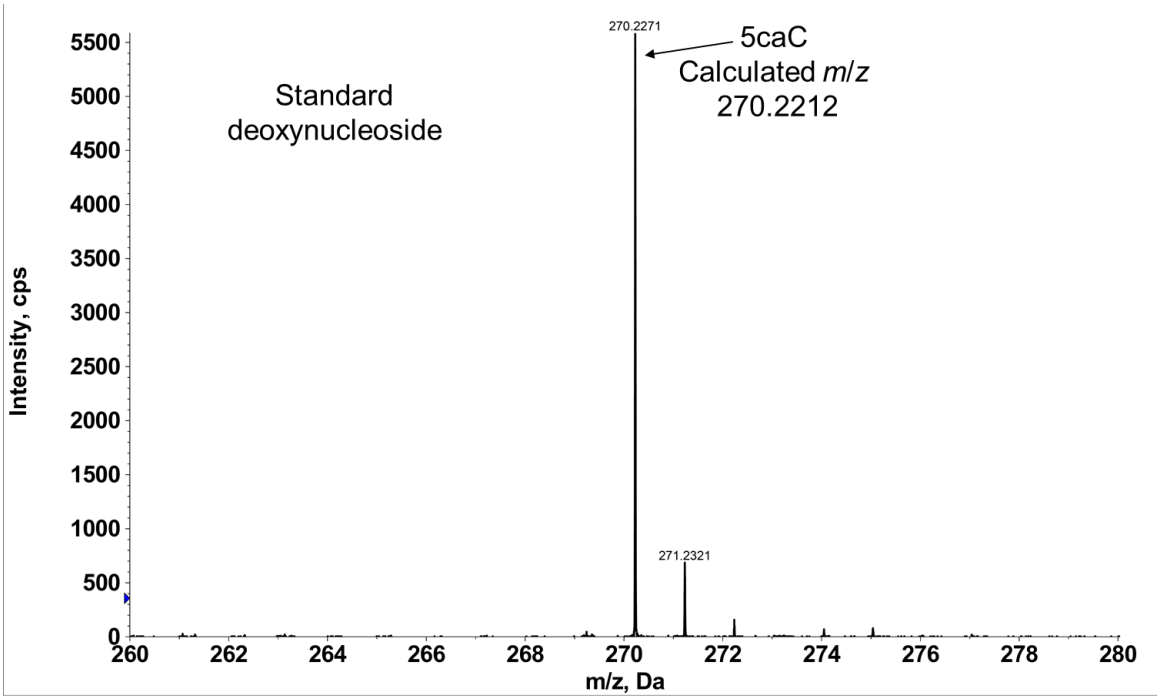

Figure S9. MS analysis of standard 5caC deoxyribonucleoside under negative ion mode. The theoretical m/z value of the corresponding species is listed next to the observed peak.

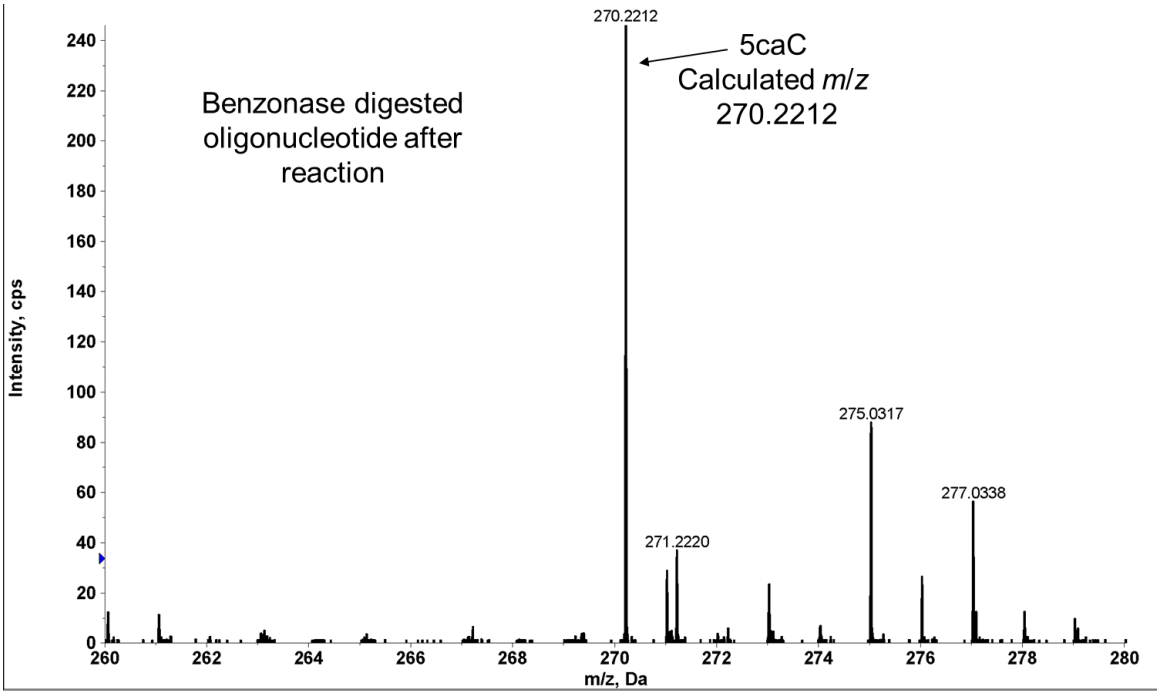

Figure S10. MS analysis of digested oligonucleotide containing 5caC deoxyribonucleoside under negative ion mode. The theoretical m/z value of the corresponding species is listed next to the observed peak.

AlkB sequence

MLDLFADAEP WQEPLAAGAV ILRRFAFNAA EQLIRDINDV ASQSPFRQMV TPGGYTMSVA
MTNCGHLGWT THRQGYLYSP IDPQTNKPWP AMPQSFHNLC QRAATAAGYP DFQPDACLIN
RYAPGAKLSL HQDKDEPDLR APIVSVSLGL PAIFQFGGLK RNDPLKRLLL EHGDVVVWGG
ESRLFYHGIQ PLKAGFHPLT IDCRYNLTFR QAGKKE

Peptide fragments from trypsin digestion

Mass	Position	Peptide sequence
3415.6230	74-102	QGYLYSPIDPQTNKPWPAMPQSFHNLCQR
2835.2713	48-73	QMVTPGGYTMSVAMTNCGHLGWTTHR
2525.2988	1-23	MLDLFADAEPWQEPLAAGAV ILR
2013.1662	141-160	APIVSVSLGLPAIFQFGGLK
1992.9363	103-121	AATAAGYPDFQPDACLINR
1764.9158	168-183	LLLEHGDVVVWGGESR
1347.6419	36-47	DINDVASQSPFR
1278.6720	25-35	FAFNAAEQLIR
1228.6022	194-204	AGFHPLTIDCR
1214.6811	184-193	LFYHGIQPLK
839.4501	128-134	LSLHQDK
812.4181	205-210	YNLTFR
743.3450	135-140	DEPDLR
605.3173	122-127	YAPGAK
585.3122	162-166	NDPLK

MS analyses of digested protein fragments

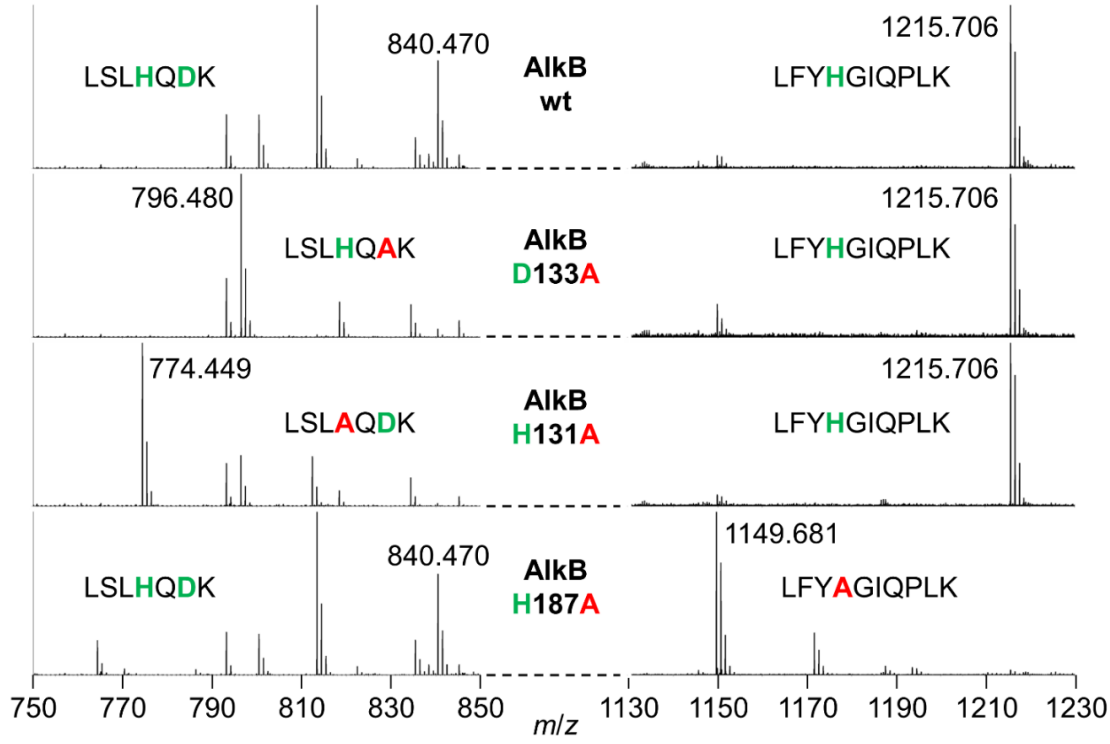

Figure S11. MS analyses of trypsin digested proteins including the wild type and variant AlkB enzymes.

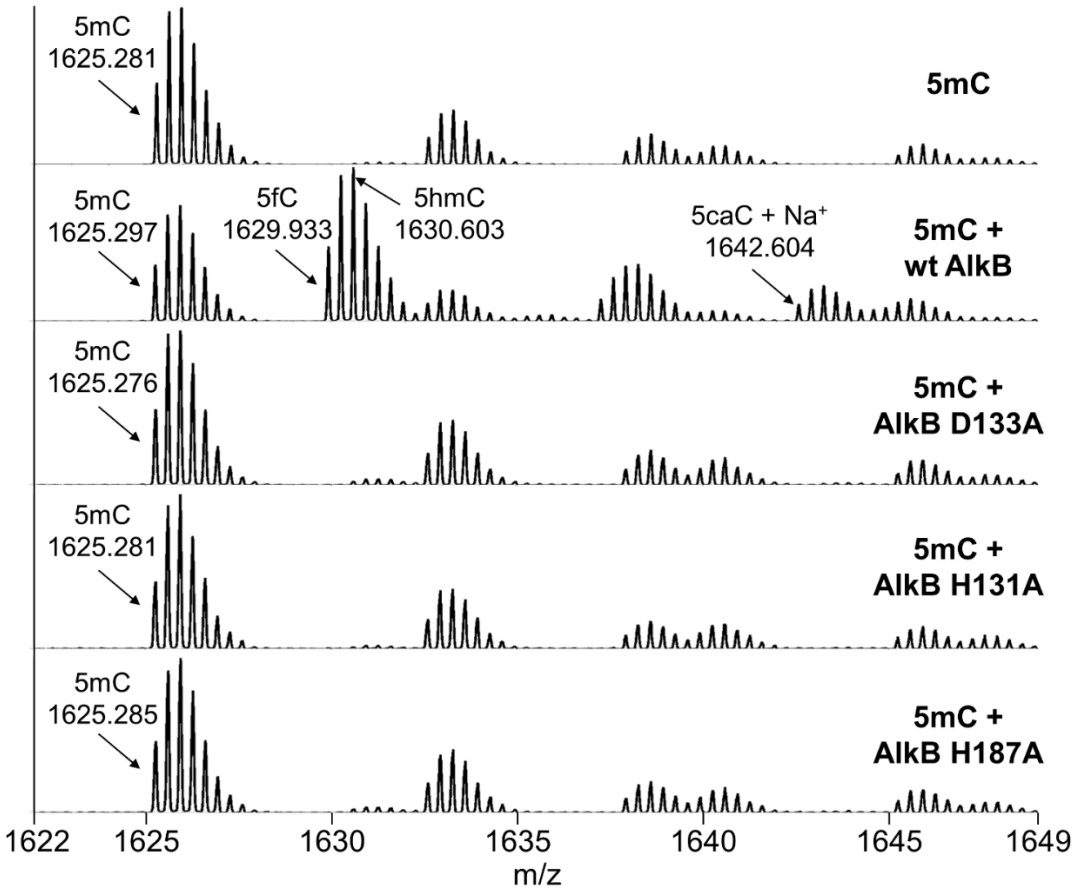

Figure S12. ESI-TOF analyses of standard oligonucleotide containing 5mC and reaction mixtures with wild type and variant proteins. The observed m/z values represent the oligonucleotides under their -3 charge state. The theoretical m/z values of the corresponding species are listed in Table S1.

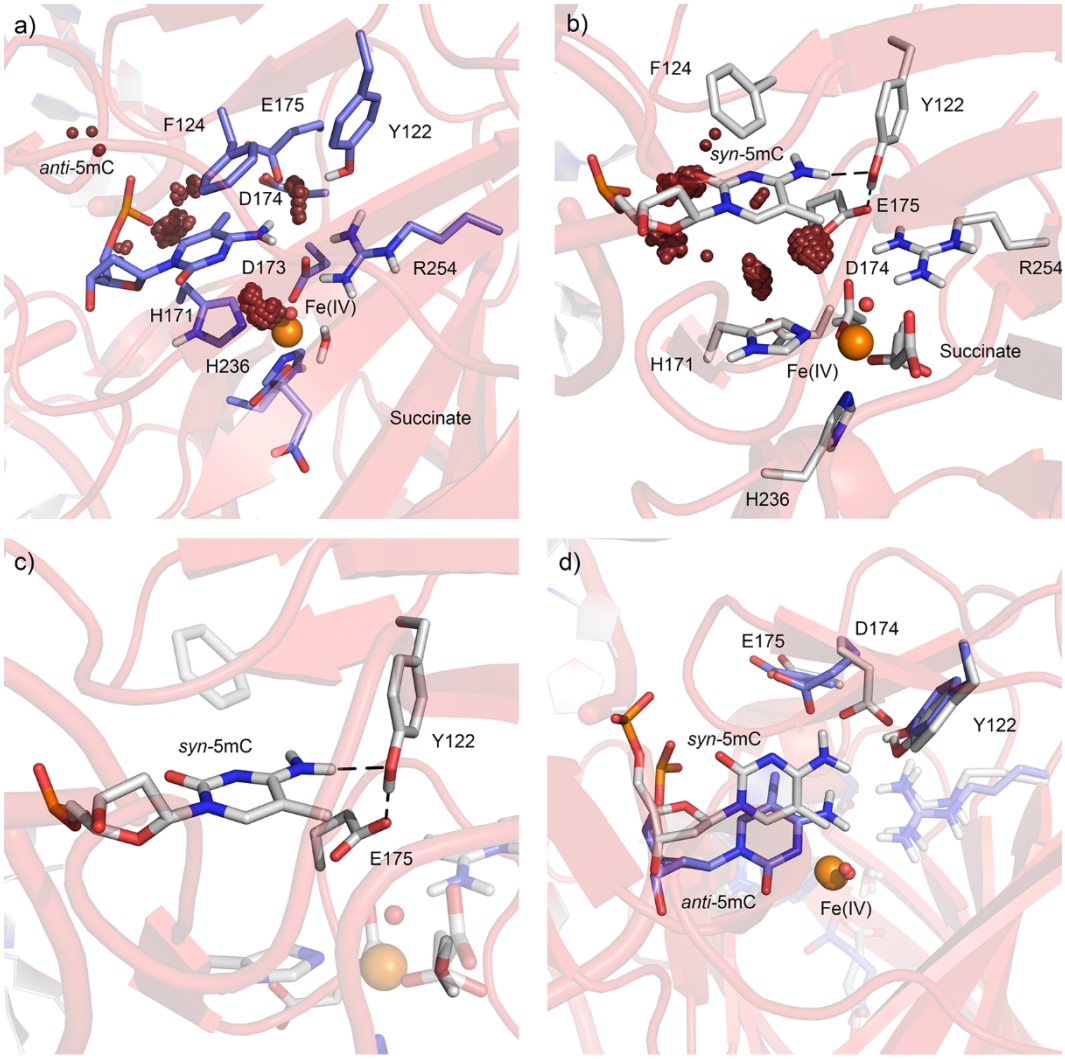

Figure S13. Distribution of active site water (red spheres) during MD simulations of ALKBH2 bound to a) anti-5mC or b) syn-5mC. c) ALKBH2 bound to syn-5mC highlighting key hydrogen-bonding interactions. d) Overlay of MD representative structures of ALKBH2 bound to anti- (blue) or syn- (gray) 5mC.

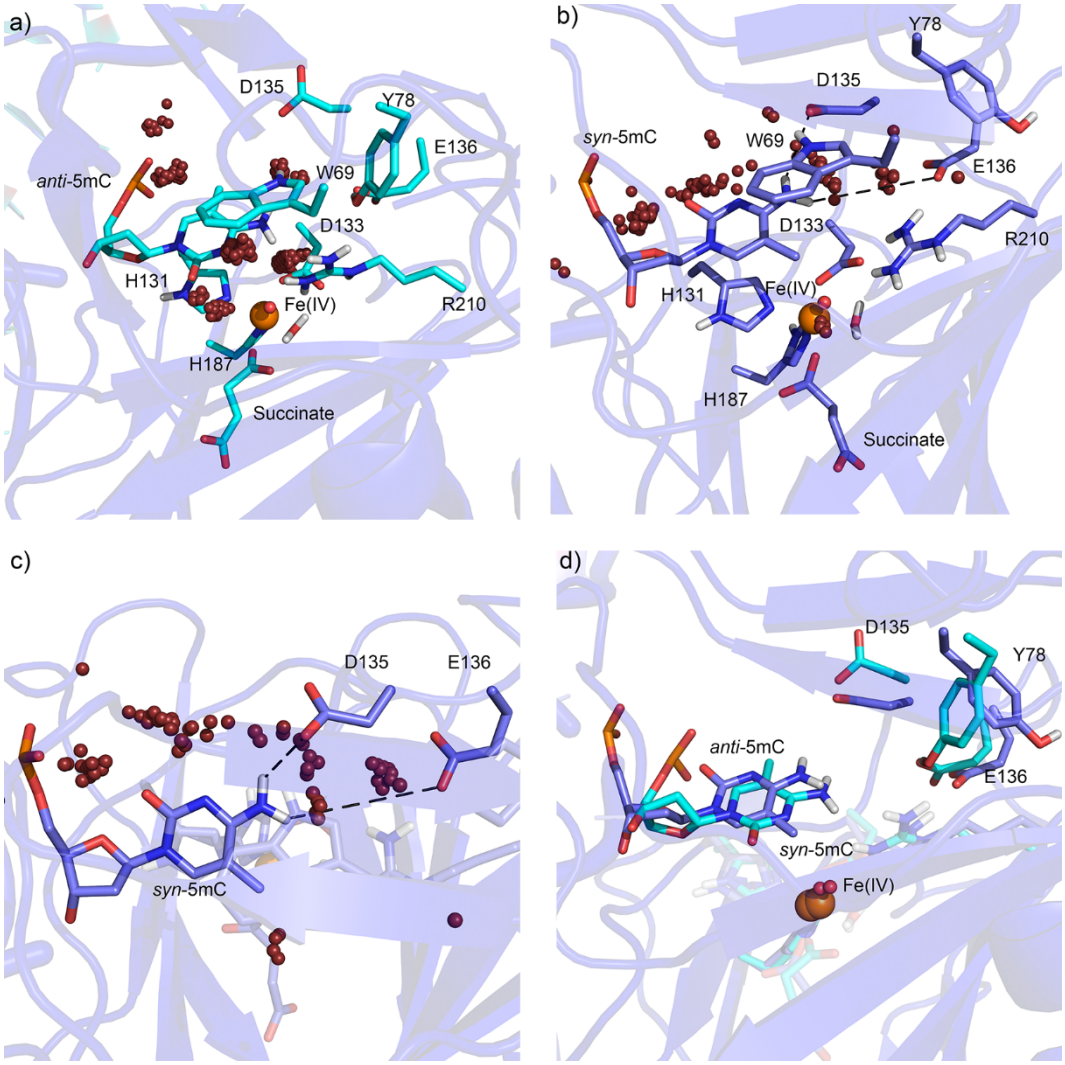

Figure S14. Distribution of active site water (red spheres) during MD simulations of AlkB bound to a) anti-5mC or b) syn-5mC. c) AlkB bound to syn-5mC highlighting key water mediated hydrogen-bonding interactions. d) Overlay of MD representative structures of AlkB bound to anti- (cyan) or syn- (blue) 5mC.

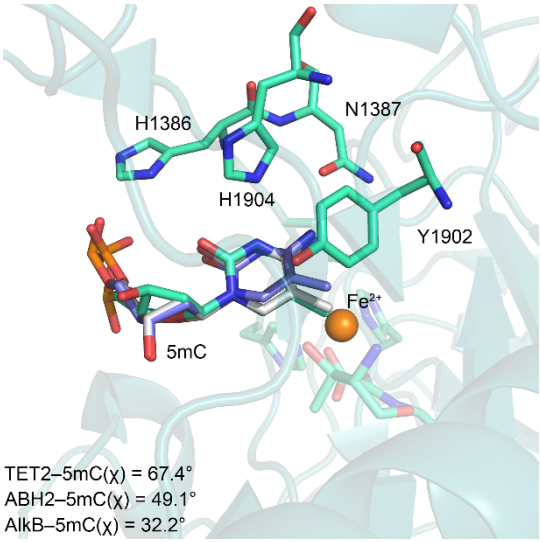

Figure S15. Overlay of the 5mC nucleotide isolated from MD representative structures of AlkB–5mC (blue) and ALKBH2–5mC (gray) complexes onto a crystal structure of TET2 bound to 5mC-containing DNA, Fe(II), and α-KG (teal; PDB ID: 4NM6),(5) highlighting the similarity in the 5mC glycosidic bond conformation (χ).

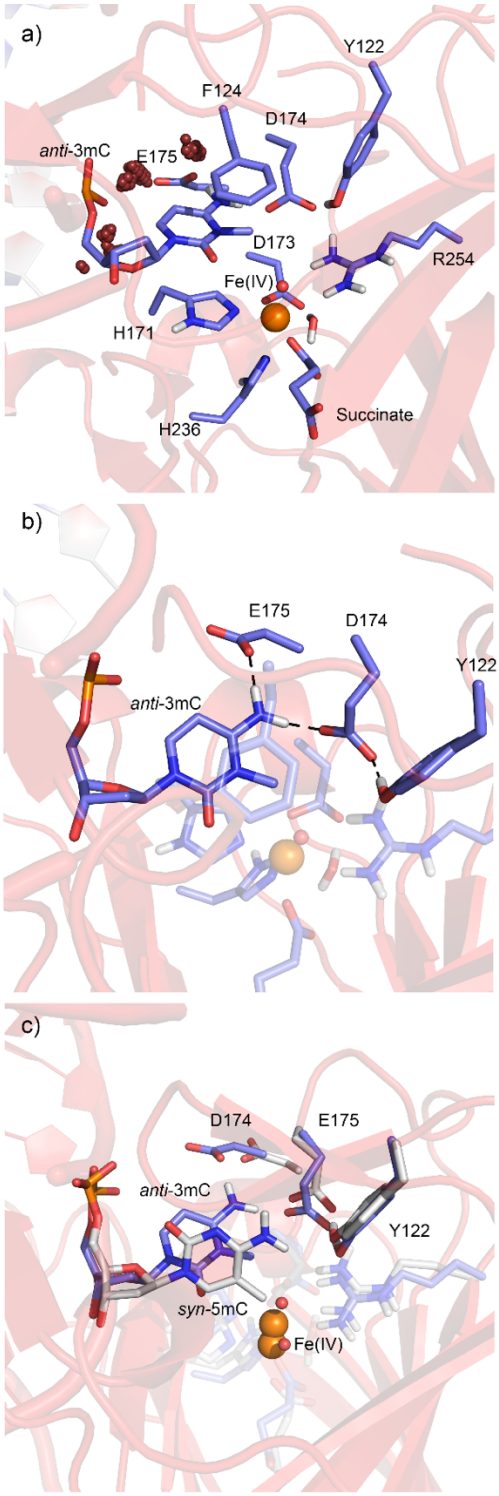

Figure S16. Distribution of active site water during MD simulations of ALKBH2 bound to anti-3mC and overlay of MD representative structures of ALKBH2 bound to syn-5mC and anti-3mC. a) Distribution of active site water (red spheres) during MD simulations of ALKBH2 bound to anti-3mC. b) ALKBH2 bound to anti-3mC, highlighting key hydrogen-bonding interactions. c) Overlay of MD representative structures of ALKBH2 bound to syn-5mC (gray) and anti-3mC (blue).

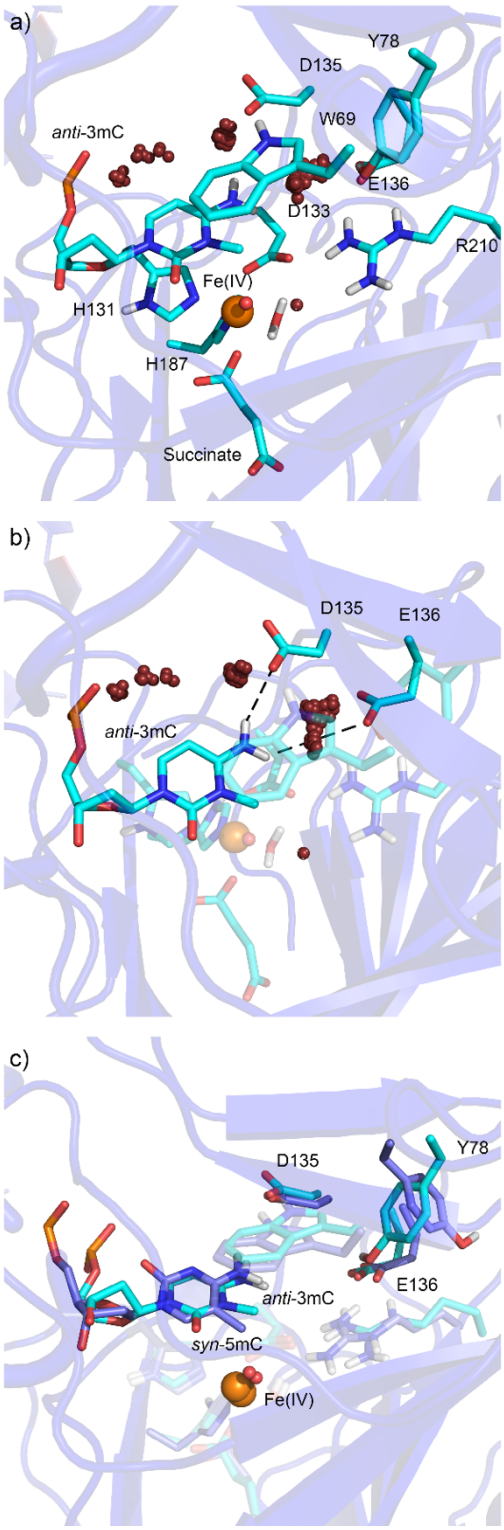

Figure S17. Distribution of active site water during MD simulations of AlkB bound to anti-3mC and overlay of MD representative structures of AlkB bound to syn-5mC and anti-3mC. a) Distribution of active site water (red spheres) during MD simulations of AlkB bound to anti-3mC. b) AlkB bound to anti-3mC, highlighting key water mediated hydrogen-bonding interactions. c) Overlay of MD representative structures of AlkB bound to syn-5mC (blue) and anti-3mC (cyan).

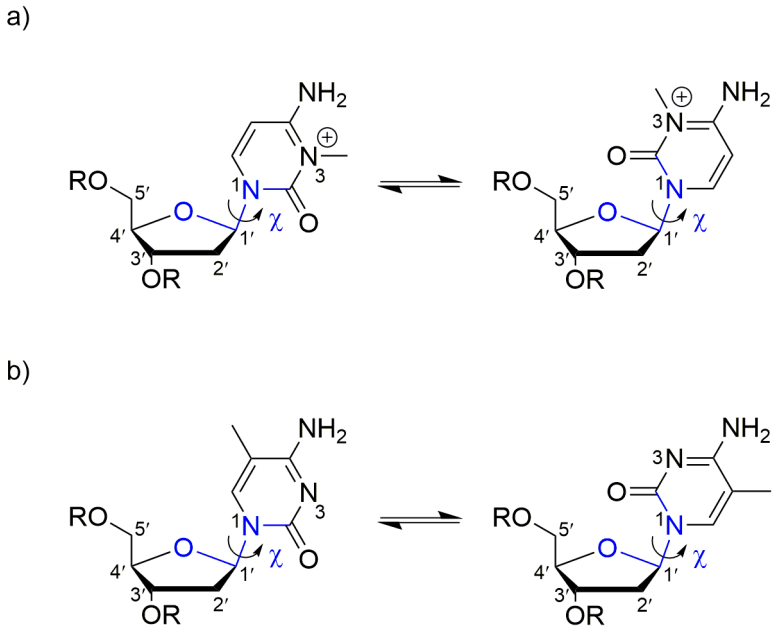

Figure S18. Structures and chemical numbering of a) 3mC and b) 5mC. The glycosidic bond orientation is defined by dihedral angle χ (blue, ∠O4′C1′N1C2) as either anti (left, 180 ± 90°) or syn (right, 0 ± 90°).

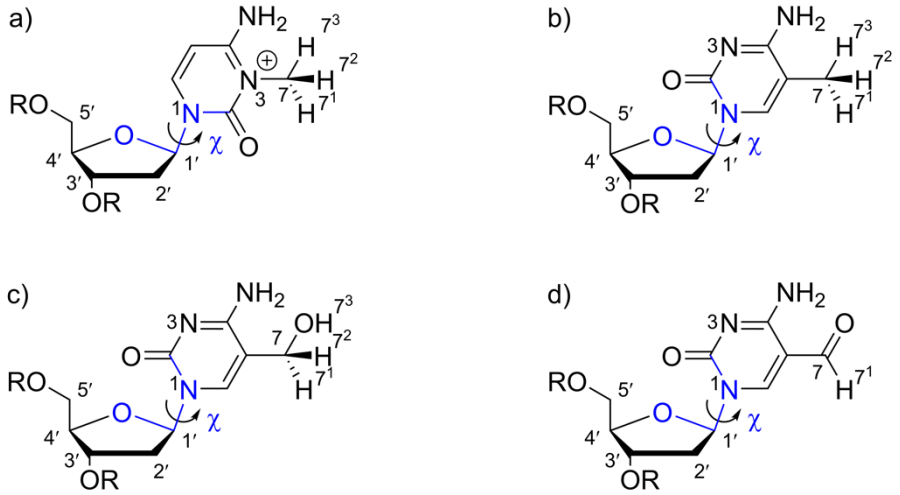

Figure S19. Structures and chemical numbering of a) anti-3mC, b) syn-5mC, c) syn-5hmC, and d) syn-5fC. The glycosidic bond orientation is defined by the χ dihedral angle (blue, ∠(O4′C1′N1C2)) as anti (180 ± 90°) or syn (0 ± 90°).

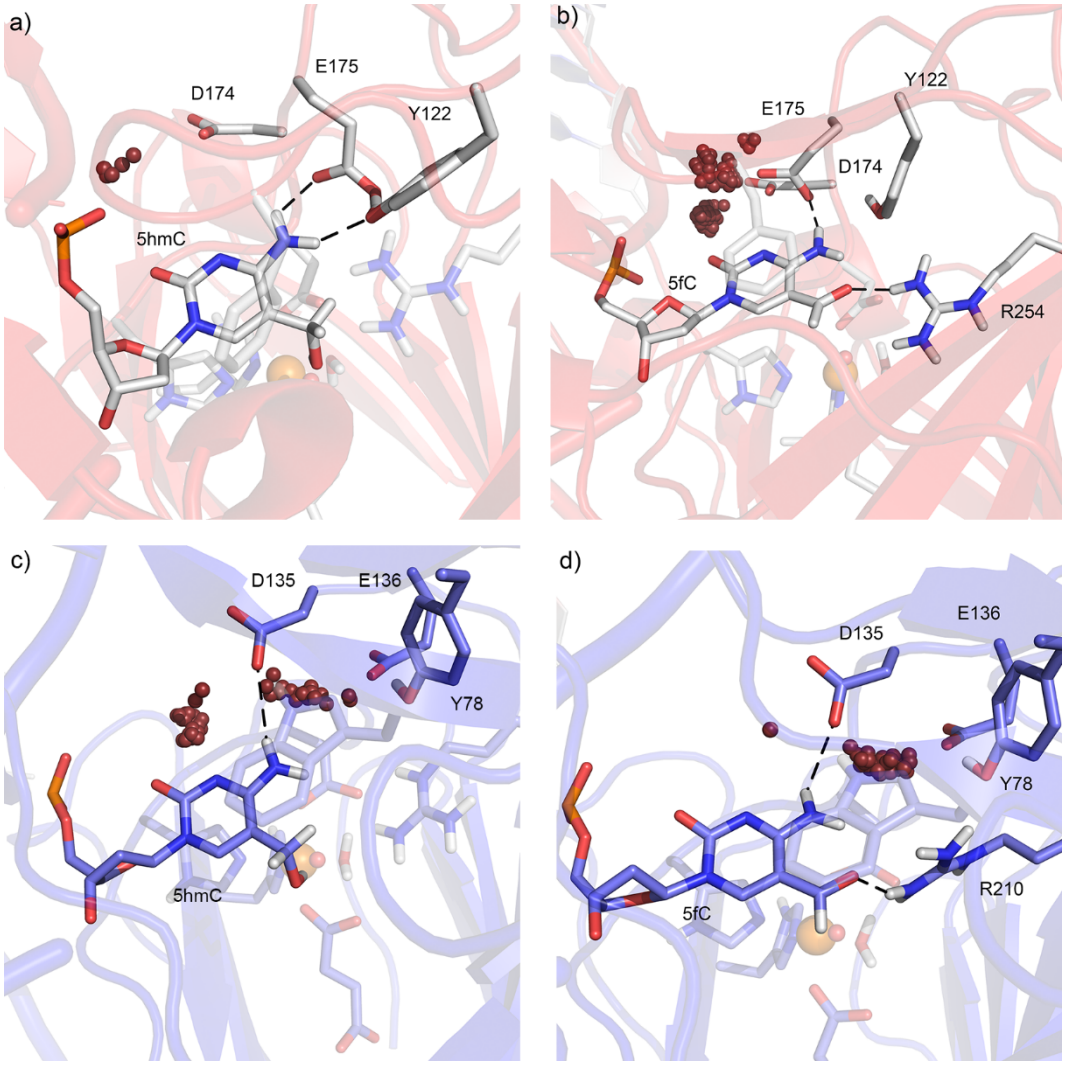

Figure S20. Distribution of active site water (red spheres) during MD simulations of a) ALKBH2 bound to syn-5hmC, b) ALKBH2 bound to syn-5fC, c) AlkB bound to syn-5hmC, or d) AlkB bound to syn-5fC. Key hydrogen-bonding interactions are highlighted with dashed lines.

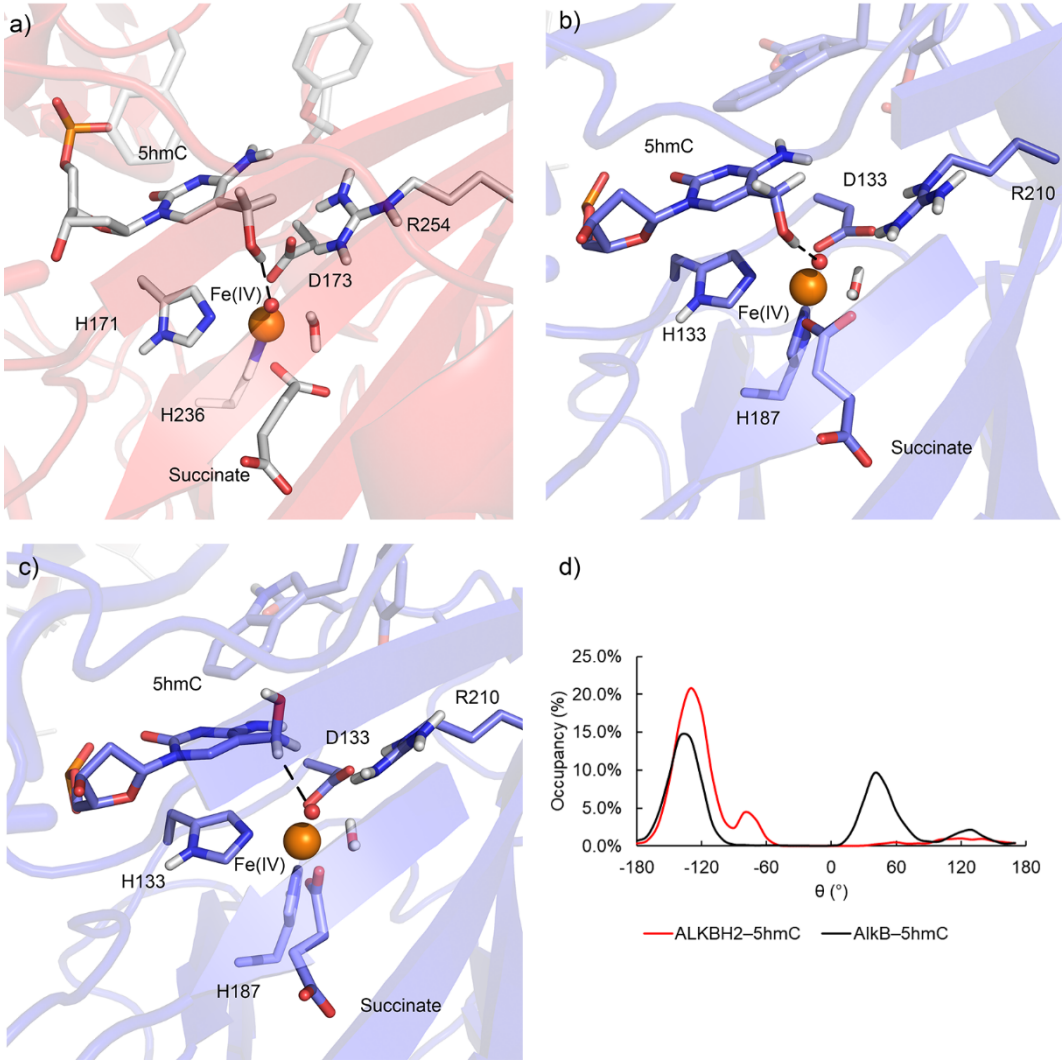

Figure S21. Simulations of ALKBH2 or AlkB bound to 5hmC. a) ALKBH2 bound to syn-5hmC, and AlkB bound to syn-5hmC with b) θ (∠(C4C5C7O7)) ~ –132° and c) θ ~ 50°, highlighting different orientations of the C5 substituent. d) Histogram displaying the occupancy of θ (degree) during simulations of ALKBH2 or AlkB bound to 5hmC.

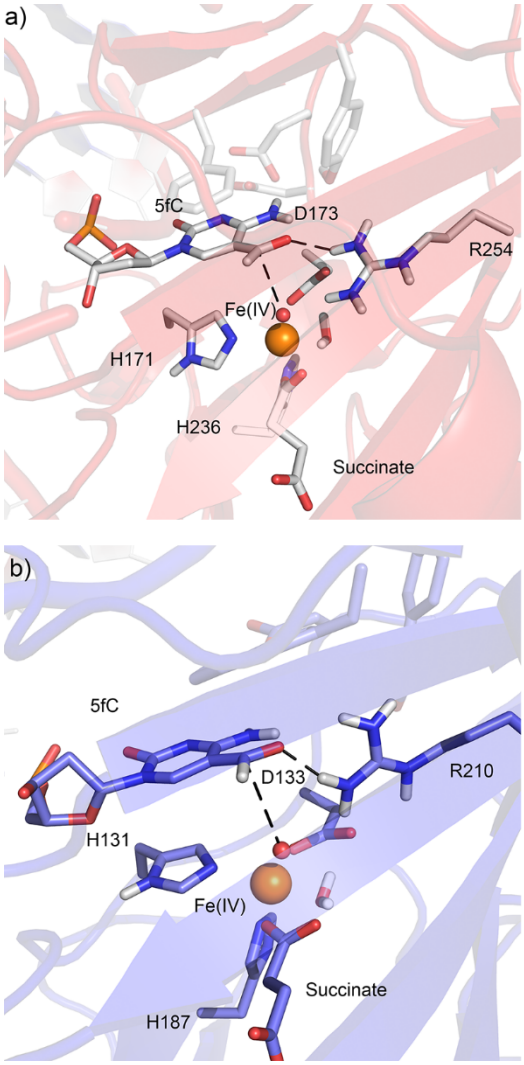

Figure S22. syn-5fC Bound by a) ALKBH2 or b) AlkB, highlighting the distance between the C5 substituent and the Fe(IV)-oxo moiety.

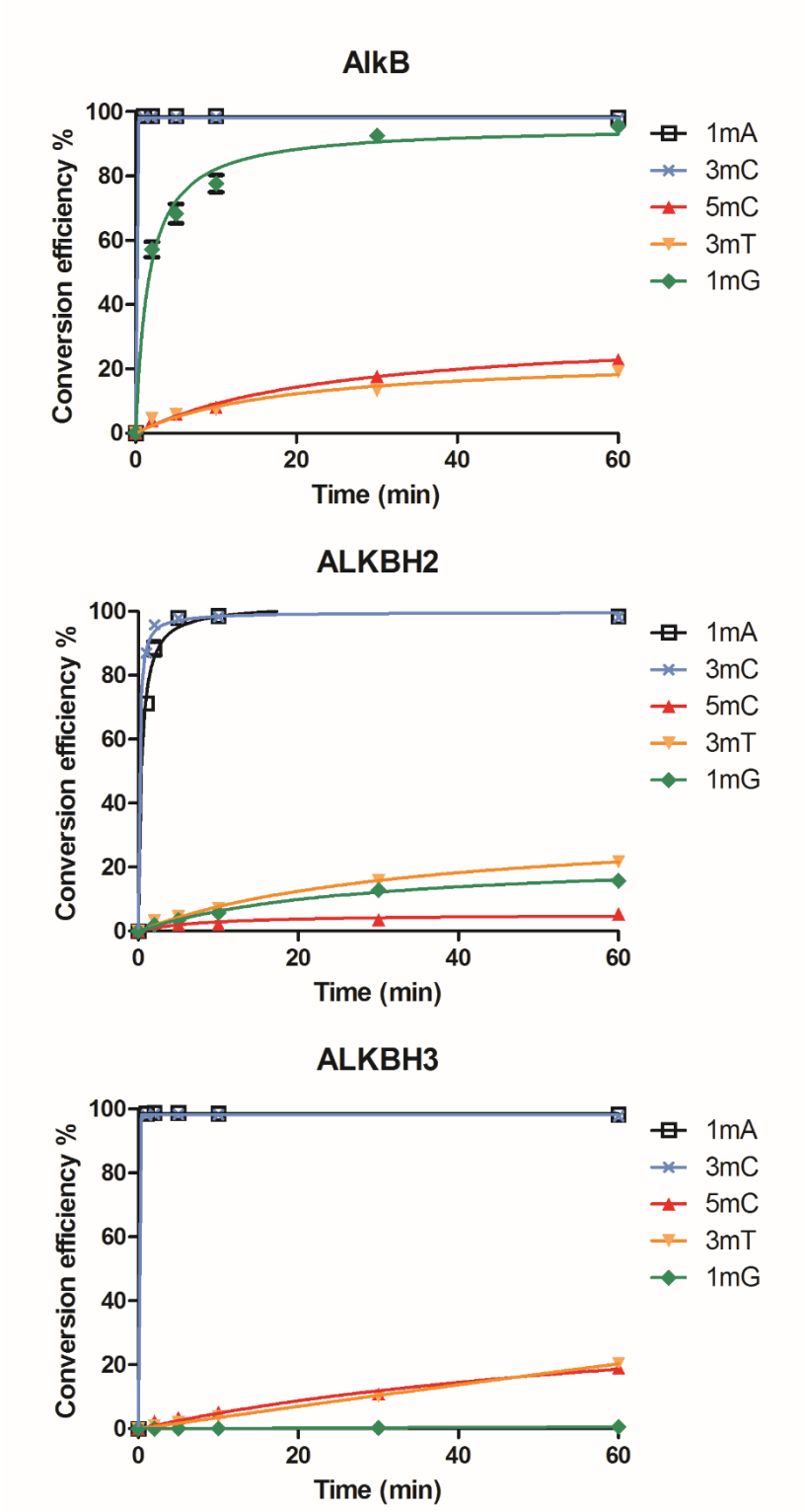

Figure S23. Different alkyl DNA modifications oxidized by the three AlkB family enzymes at various time points. Top: AlkB; middle: ALKBH2; and bottom: ALKBH3.

Table S1. Calculated and observed monoisotopic molecular weight and m/z value of modified oligonucleotides. The sequence of the 16mer was 5'-GAAGACCTXGGCGTCC-3', where X indicates the position of 5mC or oxidized bases.

modification	MW (calculated) of neutral species	*m*/*z* (calculated) -3 charge peak	*m*/*z* (observed) -3 charge peak			
			5mC control	ALKBH2	ALKBH3	AlkB
5mC	4878.866	1625.281	1625.281	1625.279	1625.285	1625.276
5mC + Na^+	4900.848	1632.608	1632.608	1632.616	1632.613	1632.608
5mC + K^+	4916.822	1637.933	1637.928	1637.941	1637.942	1637.931
5mC + 2Na^+	4922.830	1639.935	1639.941	1639.937	1639.938	1639.931
5mC + Na^+ + K^+	4938.803	1645.260	1645.267	1645.273	1645.261	1645.263
5hmC	4894.861	1630.612	-	1630.615	1630.607	1630.603
5hmC + Na^+	4916.843	1637.940	-	1637.941	1637.942	1637.931
5hmC + K^+	4932.816	1643.264	-	1643.265	1643.259	1643.262
5fC	4892.845	1629.940	-	1629.936	1629.951	1629.933
5fC + Na^+	4914.827	1637.268	-	1637.288	1637.285	1637.257
5fC + 2Na^+	4936.809	1644.595	-	1644.605	1644.588	1644.598
5caC + Na^+	4930.814	1642.600	-	1642.616	1642.611	1642.604

Table S2. Calculated and observed monoisotopic molecular weight and m/z value of peptide fragments digested by trypsin.

peptide sequence	MW (calculated) of neutral species	*m*/*z* (calculated) +1 charge peak	*m*/*z* (observed) +1 charge peak			
			AlkB **wt**	AlkB **D133A**	AlkB **H131A**	AlkB **H187A**
LSL**H**Q**D**K	839.450	840.457	840.470	-	-	840.470
LSL**H**Q**A**K	795.460	796.468	-	796.480	-	-
LSL**A**Q**D**K	773.428	774.436	-	-	774.449	-
LFY**H**GIQPLK	1214.681	1215.688	1215.706	1215.706	1215.706	-
LFY**A**GIQPLK	1148.659	1149.667	-	-	-	1149.681

Table S3. Summary of important hydrogen bonds formed during 500 ns MD simulations of *anti*-3mC, *anti*-5mC, *syn*-5mC, *syn*-5hmC, and *syn*-5fC in the ALKBH2 complex.

	3mC		***anti*-5mC**		***syn*-5mC**		***syn*-5hmC**		***syn*-5fC**	
	%[a,b]	**Dist (Angle)[c,d]**	**%[a,b]**	**Dist (Angle)[c,d]**	**%[a,b]**	**Dist (Angle)[c,d]**	**%[a,b]**	**Dist (Angle)[c,d]**	**%[a,b]**	**Dist (Angle)[c,d]**
E175(Oδ)···mC(N4H)	48.7%	2.8 (158)	0.0%	0.0 (0)	10.0%	3.0 (145)	31.3%	2.8 (157)	95.2%	2.9 (160)
D174(Oδ)···mC(N4H)	79.1%	2.8 (162)	0.0%	0.0 (0)	0.0%	N/A	8.6%	2.9 (164)	0.0%	N/A
Y122(O_H)···mC(N4H)	0.0%	3.1 (149)	18.1%	3.0 (132)	60.0%	3.0 (146)	40.9%	3.1 (156)	36.0%	3.1 (135)
H_2O(O)···mC(N4H)	82.8%	3.1 (151)	66.1%	3.1 (141)	54.3%	3.1 (137)	82.6%	3.1 (141)	25.0%	3.0 (130)
mC(O2)···H_2O(OH)	0.1%	3.2 (128)	4.8%	3.1 (143)	>100%	2.8 (152)	74.5%	2.9 (149)	13.0%	2.8 (152)
mC(N3)···H_2O(OH)	0.0%	0.0 (0)	68.8%	3.0 (151)	60.4%	3.0 (147)	49.8%	3.0 (152)	0.4%	3.2 (139)
E175(Oδ)···Y122(OH)	82.8%	2.7 (165)	51.0%	2.9 (151)	98.8	2.8 (158)	36.7%	2.8 (161)	21.4%	2.8 (155)
H_2O(O)···Y122(OH)	0.6%	3.3 (134)	>100%	3.0 (153)	33.4%	2.9 (53)	>100.0%	3.0 (149)	0.0%	N/A
E175(Oε)···H_2O(OH)	5.5%	2.8 (162)	>100%	2.8 (158)	>100%	2.7 (160)	>100.0%	2.8 (158)	>100.0%	2.7 (158)
D174(Oδ)···H_2O(OH)	>100%	2.8 (158)	>100%	2.8 (157)	>100%	2.8 (160)	>100.0%	2.8 (158)	>100.0%	2.8 (154)
mC(O5)···R254(NηH)	N/A	N/A	N/A	N/A	N/A	N/A	50.3%	3.1 (131)	>100.0%	3.0 (143)
Fe(IV)-oxo(O)···mC(O5H)	N/A	N/A	N/A	N/A	N/A	N/A	20.3%	3.2 (140)	10.7%	3.1 (132)
D173(Oδ)···mC(O5H)	N/A	N/A	N/A	N/A	N/A	N/A	14.4%	3.0 (142)	0.2%	3.3 (127)

[a]Percent occupancy of the hydrogen bond, which is defined using cutoffs of a distance less than 3.4 Å and an angle greater than 120°. [b]Greater than 100% occupancy indicates the presence of more than one water hydrogen bonding with an acceptor site. [c]Hydrogen bonding distance in Å. [d]Angle in ° (in parentheses).

Table S4. Summary of important hydrogen bonds formed during 500 ns MD simulations of *anti*-3mC, *anti*-5mC, *syn*-5mC, *syn*-5hmC, and *syn*-5fC in the AlkB complex.

	3mC		***anti*-5mC**		***syn*-5mC**		***syn*-5hmC**		***syn*-5fC**	
	%[a,b]	**Dist (Angle)[c,d]**	**%[a,b]**	**Dist (Angle)[c,d]**	**%[a,b]**	**Dist (Angle)[c,d]**	**%[a,b]**	**Dist (Angle)[c,d]**	**%[a,b]**	**Dist (Angle)[c,d]**
D133(Oδ)···mC(N4H)	28.0%	2.9 (147)	3.2%	3.1 (150)	0.0%	3.0 (123)	28.4%	3.0 (139)	0.4%	3.2 (133)
H_2O(O)···mC(N4H)	>100%	3.0 (145)	>100%	3.1 (139)	>100%	3.1 (142)	>100.0%	3.1 (140)	>100.0%	3.0 (139)
mC(O2)···H_2O(OH)	42.9%	2.9 (152)	96.5%	2.9 (153)	>100%	2.9 (152)	64.4%	3.0 (140)	>100.0%	2.8 (153)
mC(N3)···H_2O(OH)	0.0%	3.2 (127)	38.4%	2.9 (150)	78.1%	3.0 (152)	>100.0%	3.0 (154)	64.6%	2.9 (152)
E136(Oδ)···Y78(OH)	29.7%	2.7 (161)	74.0%	2.7 (162)	43.1%	2.7 (162)	89.4%	2.7 (163)	95.5%	2.7 (160)
H_2O(O)···Y78(OH)	88.9%	3.1 (150)	25.8%	3.1 (137)	1.2%	3.1 (145)	69.0%	3.0 (139)	>100.0%	3.0 (157)
E136(Oε)···H_2O(OH)	>100%	2.8 (159)	>100%	2.8 (156)	>100%	2.9 (153)	>100.0%	2.8 (153)	>100.0%	2.9 (158)
D135(Oδ)···H_2O(OH)	>100%	2.8 (158)	>100%	2.8 (157)	>100%	2.8 (157)	>100.0%	2.9 (148)	>100.0%	2.9 (158)
mC(O5)···R210(NηH)	N/A	N/A	N/A	N/A	N/A	N/A	>100.0%	2.9 (150)	>100.0%	2.9 (158)
Fe(IV)-oxo(O)···mC(O5H)	N/A	N/A	N/A	N/A	N/A	N/A	20.1%	2.9 (155)	>100.0%	2.9 (149)
D133(Oδ)···mC(O5H)	N/A	N/A	N/A	N/A	N/A	N/A	64.3%	2.8 (153)	N/A	N/A

[a]Percent occupancy of the hydrogen bond, which is defined using cutoffs of a distance less than 3.4 Å and an angle greater than 120°. [b]Greater than 100% occupancy indicates the presence of more than one water hydrogen bonding with an acceptor site. [c]Hydrogen-bonding distance in Å. [d]Angle in ° (in parentheses).

Table S5. Summary of important distances and dihedral angles adopted during 500 ns MD simulations of *anti*-3mC, *anti*-5mC, *syn*-5mC, *syn*-5hmC, and *syn*-5fC in the ALKBH2 complex.

	3mC **Avg (σ)**	***anti*-5mC** **Avg (σ)**	***syn*-5mC** **Avg (σ)**	***syn*-5hmC** **Avg (σ)**	**syn-5fC** **Avg (σ)**
Fe(IV)–oxo···C(C7)[a]	3.3 (0.2)	6.3 (1.1)	3.8 (0.3)	4.5 (0.5)	3.3 (0.3)
Fe(IV)–oxo···C(H7^1)[a]	3.2 (0.6)	6.6 (1.3)	3.6 (0.7)	4.7 (0.7)	3.6 (0.5)
Fe(IV)–oxo···C(H7^2)[a]	3.2 (0.6)	6.6 (1.3)	3.6 (0.7)	7.2 (0.6)	N/A
Fe(IV)–oxo···C(H7^3)[a]	3.2 (0.6)	6.6 (1.3)	3.6 (0.7)	3.0 (0.8)	N/A
C(χ)[b]	–142 (14)	–138 (18)	33.4 (30)	32 (32)	–24 (16)

[a]Modification of either N3 (3mC; Figure S19) or C5 (5mC, 5hmC, or 5fC; Figure S19). [b]χ of bound pyrimidine defined as ∠(O4′C1′N1C2) (Figures S18 and S19).

Table S6. Summary of important distances and dihedral angles adopted during 500 ns MD simulations of *anti*-3mC, *anti*-5mC, *syn*-5mC, *syn*-5hmC, and *syn*-5fC in the AlkB complex.

	3mC **Avg (σ)**	***anti*-5mC** **Avg (σ)**	***syn*-5mC** **Avg (σ)**	***syn*-5hmC** **Avg (σ)**	***syn*-5fC** **Avg (σ)**
Fe(IV)–oxo···C(C7)[a]	3.4 (0.3)	8.2 (1.7)	3.8 (0.6)	3.4 (0.3)	3.2 (0.3)
Fe(IV)–oxo···C(H7^1)[a]	3.4 (0.7)	8.5 (1.8)	3.7 (0.8)	3.1 (0.6)	3.3 (0.5)
Fe(IV)–oxo···C(H7^2)[a]	3.4 (0.7)	8.5 (1.8)	3.7 (0.8)	5.9 (0.4)	N/A
Fe(IV)–oxo···C(H7^3)[a]	3.4 (0.7)	8.5 (1.8)	3.7 (0.8)	3.0 (1.3)	N/A
C(χ)[b]	–123 (14)	–111 (13)	36.3 (36)	52 (13)	50 (10)

[a]Modification of either N3 (3mC; Figure S19) or C5 (5mC, 5hmC, or 5fC; Figure S19). [b]χ of bound pyrimidine defined as ∠(O4′C1′N1C2) (Figures S18 and S19).

Table S7. Time to half maximum conversion (T1/2, min) for the oxidation of different alkyl DNA modifications by the three AlkB family enzymes.

	1mA	**3mC**	**5mC**	**3mT**	**1mG**
AlkB	< 1.0	< 1.0	16.9	17.4	1.7
ALKBH2	< 1.0	< 1.0	17.2	18.7	16.4
ALKBH3	< 1.0	< 1.0	25.2	30.0	30.7

Table S8. Comparison of important distances and dihedral angles adopted during MD simulations of *anti*-3mC, *anti*-5mC, and *syn*-5mC in the ALKBH2 complex.

	3mC			***anti*-5mC**			***syn*-5mC**		
	500 ns R1[c]	100 ns R1[d]	100 ns R2[e]	500 ns R1[c]	100 ns R1[d]	100 ns R2[e]	500 ns R1[c]	100 ns R1[d]	100 ns R2[e]
Fe(IV)–oxo···mC(C7)[a]	3.3 (0.2)	3.3 (0.2)	3.4 (0.3)	6.3 (1.1)	5.3 (0.5)	5.7 (0.3)	3.8 (0.3)	3.8 (0.3)	4.0 (0.4)
Fe(IV)–oxo···mC(H7^1)[a]	3.2 (0.6)	3.2 (0.7)	3.4 (0.7)	6.6 (1.3)	5.5 (0.8)	5.8 (0.7)	3.6 (0.7)	3.6 (0.6)	3.9 (0.7)
Fe(IV)–oxo···mC(H7^2)[a]	3.2 (0.6)	3.2 (0.6)	3.4 (0.7)	6.6 (1.3)	5.5 (0.8)	5.8 (0.7)	3.6 (0.7)	3.6 (0.6)	3.9 (0.7)
Fe(IV)–oxo···mC(H7^3)[a]	3.2 (0.6)	3.2 (0.7)	3.4 (0.7)	6.6 (1.3)	5.5 (0.8)	5.9 (0.7)	3.6 (0.7)	3.6 (0.6)	3.9 (0.7)
mC(χ)[b]	–142 (14)	–147 (15)	–146 (9)	–138 (18)	–143 (19)	–148 (9)	34 (30)	49 (11)	42 (15)

[a]Modification of either N3 (3mC; Figure S19) or C5 (5mC, 5hmC, or 5fC; Figure S19). [b]χ of bound pyrimidine defined as ∠O4′C1′N1C2 (Figure S18). [c]Distances and dihedral angles calculated over the entire 500 ns of the first replicate simulation. [d]Distances and dihedral angles calculated over the first 100 ns of the first replicate simulation. [e]Distances and dihedral angles calculated for the 100 ns second replicate simulation using different initial velocities, but same starting configuration as the first replicate simulation.

Table S9. Comparison of important distances and dihedral angles adopted during MD simulations of *syn*-5hmC *and syn*-5fC in the ALKBH2 complex.

	***syn*-5hmC**			***syn*-5fC**		
	500 ns R1[c]	100 ns R2[d]	100 ns R2[e]	500 ns R1[c]	100 ns R1[d]	100 ns R2[e]
Fe(IV)–oxo···mC(C7)[a]	4.5 (0.5)	4.5 (0.8)	4.0 (0.3)	3.3 (0.3)	3.2 (0.4)	3.1 (0.6)
Fe(IV)–oxo···mC(H7^1)[a]	4.7 (0.7)	4.7 (1.1)	4.2 (0.8)	3.3 (0.5)	3.3 (0.4)	3.6 (0.7)
Fe(IV)–oxo···mC(H7^2)[a]	7.2 (0.6)	6.9 (0.6)	6.3 (0.6)	N/A	N/A	N/A
Fe(IV)–oxo···mC(H7^3)[a]	3.0 (0.8)	3.2 (1.4)	3.2 (1.5)	N/A	N/A	N/A
mC(χ)[b]	32 (32)	–9 (17)	9 (22)	–24 (16)	–19 (19)	–38 (24)

[a]Modification of either N3 (3mC; Figure S19) or C5 (5mC, 5hmC, or 5fC; Figure S19). [b]χ of bound pyrimidine defined as ∠O4′C1′N1C2 (Figure S18). [c]Distances and dihedral angles calculated over the entire 500 ns of the first replicate simulation. [d]Distances and dihedral angles calculated over the first 100 ns of the first replicate simulation. [e]Distances and dihedral angles calculated over the 100 ns second replicate simulation using different initial velocities, but same starting configuration as the first replicate simulation.

Table S10. Comparison of important distances and dihedral angles adopted during MD simulations of *anti*-3mC, *anti*-5mC, and *syn*-5mC in the AlkB complex.

	3mC			***anti*-5mC**			***syn*-5mC**		
	500 ns R1[c]	100 ns R1[d]	100 ns R2[e]	500 ns R1[c]	100 ns R1[d]	100 ns R1[e]	500 ns R1[c]	100 ns R1[d]	100 ns R2[e]
Fe(IV)–oxo···mC(C7)[a]	3.4 (0.3)	3.3 (0.2)	3.3 (0.2)	7.9 (2.0)	8.2 (1.7)	6.7 (1.5)	3.8 (0.6)	3.6 (0.4)	4.0 (0.6)
Fe(IV)–oxo···mC(H7^1)[a]	3.4 (0.7)	3.5 (0.7)	3.3 (0.7)	8.2 (2.0)	8.5 (1.8)	7.0 (1.6)	3.7 (0.8)	3.6 (0.7)	3.9 (0.7)
Fe(IV)–oxo···mC(H7^2)[a]	3.4 (0.7)	3.4 (0.7)	3.4 (0.7)	8.2 (2.0)	8.5 (1.8)	7.0 (1.7)	3.7 (0.8)	3.5 (0.7)	3.9 (0.7)
Fe(IV)–oxo···mC(H7^3)[a]	3.4 (0.7)	3.4 (0.7)	3.4 (0.7)	8.2 (2.0)	8.5 (1.8)	7.0 (1.6)	3.7 (0.8)	3.6 (0.7)	3.9 (0.7)
mC(χ)[b]	–123 (14)	–128 (11)	–129 (14)	–116 (13)	–111 (13)	–123 (12)	36 (36)	32 (17)	43 (18)

[a]Modification of either N3 (3mC; Figure S19) or C5 (5mC, 5hmC, or 5fC; Figure S19). [b]χ of bound pyrimidine defined as ∠O4′C1′N1C2 (Figure S18). [c]Distances and dihedral angles calculated over the entire 500 ns of the first replicate simulation. [d]Distances and dihedral angles calculated over the first 100 ns of the first replicate simulation. [e]Distances and dihedral angles calculated for the 100 ns second replicate simulation using different initial velocities, but same starting configuration as the first replicate simulation.

Table S11. Comparison of important distances and dihedral angles adopted during MD simulations of *syn*-5hmC *and syn*-5fC in the AlkB complex.

	***syn*-5hmC**			***syn*-5fC**		
	500 ns R1[c]	100 ns R1[d]	100 ns R2[e]	500 ns R1[c]	100 ns R1[d]	100 ns R2[e]
Fe(IV)–oxo···mC(C7)[a]	3.4 (0.3)	3.4 (0.2)	3.3 (0.2)	3.2 (0.3)	3.5 (0.4)	3.7 (0.4)
Fe(IV)–oxo···mC(H7^1)[a]	3.1 (0.6)	3.3 (0.6)	3.0 (0.5)	3.6 (0.5)	3.8 (0.5)	4.0 (0.4)
Fe(IV)–oxo···mC(H7^2)[a]	5.9 (0.4)	5.8 (0.4)	5.7 (0.5)	N/A	N/A	N/A
Fe(IV)–oxo···mC(H7^3)[a]	3.0 (1.3)	3.3 (1.5)	3.5 (1.4)	N/A	N/A	N/A
mC(χ)[b]	52 (13)	44 (18)	47 (23)	50 (10)	51 (10)	52 (10)

[a]Modification of either N3 (3mC; Figure S19) or C5 (5mC, 5hmC, or 5fC; Figure S19). [b]χ of bound pyrimidine defined as ∠O4′C1′N1C2 (Figure S18). [c]Distances and dihedral angles calculated over the entire 500 ns of the first replicate simulation. [d]Distances and dihedral angles calculated over the first 100 ns of the first replicate simulation. [e]Distances and dihedral angles calculated for the 100 ns second replicate simulation using different initial velocities, but same starting configuration as the first replicate simulation.

References

1. Chen,F., Tang,Q., Bian,K., Humulock,Z.T., Yang,X., Jost,M., Drennan,C.L., Essigmann,J.M. and Li,D. (2016) Adaptive response enzyme AlkB preferentially repairs 1-methylguanine and 3-methylthymine adducts in double-stranded DNA. *Chem. Res. Toxicol.*, **29**, 687–693.

2. Fedeles,B.I., Singh,V., Delaney,J.C., Li,D. and Essigmann,J.M. (2015) The AlkB family of Fe(II)/α-ketoglutarate-dependent dioxygenases: repairing nucleic acid alkylation damage and beyond. *J. Biol. Chem.*, **290**, 20734–20742.

3. Delaney,J.C. and Essigmann,J.M. (2004) Mutagenesis, genotoxicity, and repair of 1-methyladenine, 3-alkylcytosines, 1-methylguanine, and 3-methylthymine in alkB Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.*, **101**, 14051–14056.

4. Lu,J., Hu,L., Cheng,J., Fang,D., Wang,C., Yu,K., Jiang,H., Cui,Q., Xu,Y. and Luo,C. (2016) A computational investigation on the substrate preference of ten-eleven-translocation 2 (TET2). *Phys. Chem. Chem. Phys. PCCP*, **18**, 4728–4738.

5. Hu,L., Li,Z., Cheng,J., Rao,Q., Gong,W., Liu,M., Shi,Y.G., Zhu,J., Wang,P. and Xu,Y. (2013) Crystal structure of TET2-DNA complex: insight into TET-mediated 5mC oxidation. *Cell*, **155**, 1545–1555.