5522–5529 *Nucleic Acids Research, 2019, Vol. 47, No. 11* *Published online 22 May 2019*
doi: 10.1093/nar/gkz395

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[3,4,5,6], John M. Essigmann[3,5,7], Stacey D. Wetmore[2,*] and Deyu Li[1,*]

[1]Department of Biomedical and Pharmaceutical Sciences, College of Pharmacy, University of Rhode Island, Kingston, RI 02881, USA, [2]Department of Chemistry and Biochemistry, University of Lethbridge, 4401 University Drive West, Lethbridge, Alberta T1K 3M4, Canada, [3]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139, USA, [4]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA, [5]Center for Environmental Health Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA, [6]Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02139, USA and [7]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Received January 23, 2019; Revised April 25, 2019; Editorial Decision April 30, 2019; Accepted May 03, 2019

ABSTRACT

5-Methylcytosine (5mC) in DNA CpG islands is an important epigenetic biomarker for mammalian gene regulation. It is oxidized to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) by the ten-eleven translocation (TET) family enzymes, which are α-ketoglutarate (α-KG)/Fe(II)-dependent dioxygenases. In this work, we demonstrate that the epigenetic marker 5mC is modified to 5hmC, 5fC, and 5caC *in vitro* by another class of α-KG/Fe(II)-dependent proteins—the DNA repair enzymes in the AlkB family, which include ALKBH2, ALKBH3 in huamn and AlkB in *Escherichia coli*. Theoretical calculations indicate that these enzymes may bind 5mC in the *syn*-conformation, placing the methyl group comparable to 3-methylcytosine, the prototypic substrate of AlkB. This is the first demonstration of the AlkB proteins to oxidize a methyl group attached to carbon, instead of nitrogen, on a DNA base. These observations suggest a broader role in epigenetics for these DNA repair proteins.

INTRODUCTION

In mammals, methylation at the 5-position of cytosine (5-methylcytosine, 5mC, Figure 1) is the major form of DNA modification and occurs mainly on CpG dinucleotide sites (1–3). This methylation is achieved and maintained by *S*-adenosylmethionine-dependent methyltransferases (4). The reverse process termed demethylation of 5mC is first carried out by the ten-eleven translocation proteins (TET1/2/3) through iterative oxidation of 5mC to 5-hydroxymethycytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (Figure 1). Subsequently, the oxidation products 5fC and 5caC are removed and repaired back to unmethylated cytosine by thymine DNA glycosylase coupled with base excision repair (1,5–8). The 5mC modification and its three oxidative derivatives play important roles in epigenetic regulations and cell development: they function in transcriptional regulation, gene silencing and reprogramming (1,5). The misregulations of 5mC and derivatives have been associated with cancer and other diseases (1,9,10).

The TET family enzymes belong to the α-ketoglutarate (α-KG)/Fe(II)-dependent dioxygenases; they have been extensively studied in the last decade for their biological functions, biochemical activities and structural features (1,5,6,11). In the structures of the TET enzymes, the highly conserved N-terminal β-hairpin-like element for DNA-base recognition and the C-terminal catalytic domain are also found in another family of α-KG/Fe(II)-dependent nucleic acid-modifying dioxygenases, the AlkB family proteins (5,12,13). In this work, we show that the epigenetic modulator 5mC is modified *in vitro* to 5hmC, 5fC, and 5caC by the DNA repair enzymes in the AlkB family, including hu-

*To whom correspondence should be addressed. Tel: +1 401 874 9361; Email: deyuli@uri.edu
Correspondence may also be addressed to Stacey D. Wetmore. Tel: +1 403 329 2323; Fax: +1 403 329 2057; Email: stacey.wetmore@uleth.ca
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.
Present addresses:
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, CA 94158, USA.

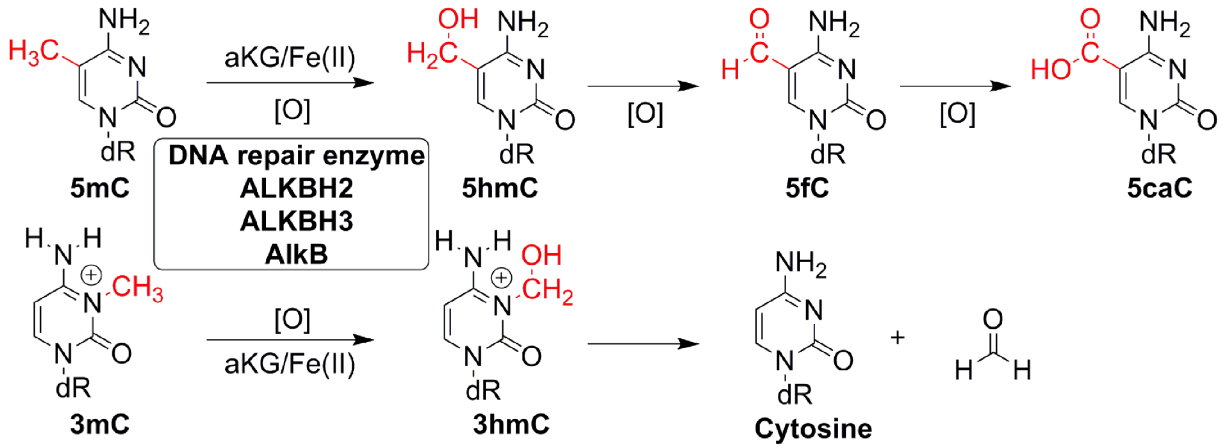

Figure 1. Reaction pathway of the AlkB family DNA repair enzymes modifying 5mC and 3mC.

man ALKBH2, ALKBH3 and its *Escherichia coli* homolog AlkB (Figure 1). The results suggest that these DNA repair enzymes may play a wider role in epigenetic modulation.

Different homologs of the *E. coli* AlkB protein exist in prokaryotic and eukaryotic species; nine homologs exist in human cell (ALKBH1-8 and FTO) (12,14). Among the nine homologs, ALKBH2 and ALKBH3 are DNA repair enzymes that protect the informational integrity of the genome.(5,14–18) They use an α-KG/Fe(II)-dependent mechanism to oxidize aberrant alkyl groups, ultimately restoring the undamaged DNA bases (14,17,18). The reported substrate scope of AlkB, ALKBH2 and ALKBH3 includes 3-methylcytosine (3mC, Figure 1), 1-methyladenine (1mA), 3-methylthymine (3mT) and 1-methylguanine (1mG), as well as other nitrogen-attached methyl lesions occurring at the Watson-Crick base pairing interface of DNA bases (15,16,18–20).

Key structural information about the AlkB family enzymes has been obtained from crystal structures of AlkB and ALKBH2 bound to lesion-containing DNA, which reveal that their active sites share several characteristics. Specifically, the AlkB (ALKBH2 in brackets) complex contains a metal center Fe(II) in the wild-type enzymes coordinated to H131 (H171), H187 (H236), D133 (D173), α-KG, and molecular oxygen (13,21). Repair of 3mC or 1mA has been proposed to be further enhanced by interactions between D135 of AlkB (E175 of ALKBH2) and the exocyclic amino groups of the lesion (22). Interestingly, all crystal structures of the AlkB enzymes indicate the lesions are bound in the *anti* glycosidic bond conformation (e.g., for 3mC: χ; $\angle$(O4′C1′N1C2) = 180° ± 90°, Supplementary Figure S18a.)

Although both 5mC and 3mC carry a methyl modification, the methyl groups are on the opposite sides of the pyrimidine ring. It is reasonable to predict that members of the AlkB enzyme family may be able to oxidize 5mC if the methyl group can be positioned near the catalytic center. The *in vitro* experimental results demonstrated here reveal that the AlkB enzymes can not only repair DNA lesions, such as 3mC, but also modify the epigenetic biomarker 5mC and generate its oxidative derivatives. Our theoretical calculations suggest that AlkB enzymes bind 5mC in the *syn* glycosidic conformation ($\chi = 0° \pm 90°$, Supplementary Figure S18b) to align the 5-methyl moiety for oxidation, which is similar to how the TET family enzymes bind 5mC. This paper is the first work to demonstrate the ability of the AlkB family enzymes to oxidize a methyl group that is attached to carbon, instead of nitrogen, on a DNA base.

MATERIALS AND METHODS

Synthesis of oligonucleotides containing 5mC and other modifications

All oligonucleotides used in this study were synthesized by solid-phase synthesis (23–26). The 5mC and other phosphoramidites was purchased from Glen Research. Synthetic oligonucleotides were purified by reverse-phase HPLC and identified by electrospray ionization mass spectrometry (ESI-MS).

Protein expression and purification

The expression and purification of ALKBH2, ALKBH3 and AlkB proteins were described by previous published papers (23,24). ALKBH2 and ALKBH3 in storage buffer containing 50 mM *N*-[tris(hydroxymethyl)methyl]-3-aminopropanesulfonic acid (TAPS), 300 mM NaCl, 10% glycerol and 1 mM 2-mercaptoethanol, pH 8.0, AlkB in similar storage buffer (10 mM Tris, 100 mM NaCl, 1 mM 2-mercaptoethanol, 10% glycerol, pH 8.0), were all stored at –80°C.

Enzymatic reaction

The reactions were performed based on previous published procedures (23,24). A 16mer oligonucleotide containing one 5mC with sequence of 5′-GAAGACCT-5mC-GGCGTCC-3′ was used as the substrate. One hundred pmol 16mer 5mC was mixed into reaction buffer (5 μM $Fe(NH_4)_2(SO_4)_2$, 0.93 mM α-ketoglutarate, 1.86 mM ascorbic acid, and 46.5 mM TAPS, pH 7.0) in a total volume of 16 μl. The reactions were started by adding proteins (170 pmol AlkB, 140 pmol ALKBH2 or 128 pmol ALKBH3), and incubated at 37°C for 1 h. EDTA was employed to quench the

reactions and the reaction mixture was immediately heated up to 80°C for 5min. A 16/23mer duplex DNA was pre-annealed as the substrate for dsDNA reaction (24,25). The 23mer oligonucleotide is complementary to the 16mer sequence plus 7 nucleotides longer with the sequence of 5′-CTGGGACGCCGAGGTCTTCACTG-3′. The rest steps of dsDNA reaction were the same as ssDNA reaction described above. All reactions were carried out in triplicate.

Oligonucleotide digestion

The procedures of oligonucleotide digestion to deoxyribonucleoside were adopted from published procedure (27). A digestion mixture was premade by adding 250 Units Benzonase (Sigma-Aldrich, MO, USA), 300 mU phosphodiesterase I (Sigma-Aldrich, MO) and 200 Units alkaline phosphatase (Sigma-Aldrich, MO) to 5 ml Tris–HCl buffer (20 mM, pH 7.9) containing 100 mM NaCl and 20 mM $MgCl_2$. Reaction of oligonucleotide containing 5mC with AlkB was quenched by heating up to 80°C for 5 min, then digested by adding 50 μl digestion mixture and incubating at 37°C for 6 h. The nucleoside products were analyzed by HPLC-MS. Samples were chromatographed on a Luna Omega Polar C18 column (150 × 4.6 mm, 5 μm, 100 Å, Phenomenex, CA, USA) eluted at 1 ml/min with an acetonitrile gradient (1–15%) to water. ESI triple quadrupole time of flight mass spectrometry was conducted to detect nucleoside signals in the negative ion mode. Standard deoxyribonucleosides (including dA, dT, dC, dG, 5mdC, 5hmdC, 5fdC and 5cadC) were purchased from Berry & Associates (MI).

Protein digestion

The trypsin digestion of wide type AlkB and its variants were set up according to Trypsin-ultraTM kit (New England Biolabs, MA, USA). Ten microgram of AlkB proteins were mixed with 250 ng trypsin in trypsin-ultra buffer, incubating at 37°C overnight. The protein fragments were analyzed by HPLC-MS (24,26). Standard protein fragments (including LFYHGIQPLK and LSLHQDK sequences) were purchased from New England Peptide (MA).

Preparing catalytically inactive variants of AlkB

The AlkB H131A, D133A and H187A variants were generated by QuikChange Mutagenesis (Agilent), using pET24a-AlkBΔN11 as the PCR template and primer pairs encoding the desired mutations. All three variants were purified essentially as described (28). Briefly, One Shot BL21 Star (DE3)pLysS *E. coli* cells (Invitrogen) transformed with an AlkB variant construct were grown at 37°C until OD_{600} had reached ∼0.4, at which point the temperature was lowered to 30°C and protein production was induced by addition of 1 mM IPTG. Cells were harvested after 4 h and stored at −80°C until use. For purification, cell pellets were resuspended in lysis buffer (10 mM Tris, pH 7.3, 300 mM NaCl, 2 mM $CaCl_2$, 10 mM $MgCl_2$, 5% (v/v) glycerol, 1 mM 2-mercaptoethanol) and lysed by sonication. After clarification by centrifugation, the lysate was loaded onto a Ni-NTA column (Qiagen), the column was washed twice with lysis buffer supplemented with 10 and 20 mM imidazole, and bound protein was eluted with lysis buffer supplemented with 70 and 250 mM imidazole. Elution fractions containing AlkB, as assessed by SDS-PAGE, were combined and dialyzed for 16 h against 50 mM 2-[[1,3-dihydroxy-2-(hydroxymethyl)propan-2-yl]amino]ethanesulfonic acid (TES), pH 7.1 with or without 20 mM NaCl, and loaded onto a 5 mL HiTrap SP cation exchange column (GE Healthcare). Bound AlkB was eluted with a linear gradient of 0.02–1 M NaCl over 12 column volumes (60 mL). Fractions containing pure AlkB were pooled and purity was established by SDS-PAGE.

Computational methodology

X-ray crystal structures of AlkB (PDB ID: 3O1S) (22) and ALKBH2 (PDB ID: 3RZJ) (21) bound to lesion-containing DNA were chosen as initial structures. The 3O1S crystal structure contains AlkB bound to an oxidized 3mC intermediate-containing DNA, Fe(II) ion, and succinate. The oxidized intermediate was reverted to 3mC, and a water molecule bound to Fe(II) was modelled as an oxo ligand to generate the Fe(IV)–oxo complex. To generate the active Fe(IV)–oxo ALKBH2 complex, the cofactors (Fe(IV), succinate, and oxo ligand) and metal-binding amino acids (H131, D133 and H187) of the AlkB complex were super imposed onto the cofactors (Mn(II) and α-KG) and metal-binding amino acids (H171, D173, and H236) of the ALKBH2 complex. Hydrogen atoms were assigned using the tLEaP AMBER utility (29). Protonation states of ionizable amino acids were initially assigned using H++ (30) and adjusted using chemical intuition. For AlkB, H66, H97, H172 and H197 were assigned epsilon protonation, while H72, H131, and H187 were assigned delta protonation. For ALKBH2, H55, H59, H106, H144 and H228 were assigned epsilon protonation, H199 and H220 were assigned delta protonation, and H167 was modelled as cationic. All crosslinks induced to facilitate crystallization between DNA and protein were removed for both complexes. The G169C, C67S, C165S and C192S mutations to ALKBH2 were reverted to generate the wild-type enzyme, and the overhanging DNA ends (residues 259, 260, and 284) were truncated. To generate the AlkB–5mC or ALKBH2–5mC complexes, the *anti* and *syn* conformations of the 5mC nucleotide were overlaid onto the 3mC substrate in the ALKBH2- or AlkB-DNA complex using chemical intuition to minimize steric clashes between the bound substrate and active site amino acids.

Each complex was modelled using the AMBER parm14SB force field. Parameters assigned to the nonstandard 5mC and 3mC nucleotides were supplemented with GAFF parameters, and Restrained Electrostatic Potential (RESP) charges. For Fe(IV) and the iron-ligating residues, the Metal Center Parameter Builder (31) was used to assign RESP charges, and bonding, angle, dihedral and non-bonding parameters based on B3LYP/LANL2DZ (Fe(IV)) and B3LYP/6-31G(d,p) (H, C, N and O) optimized structures (Gaussian 09 revision D.01) (32) of the iron-binding site using the Seminario method (33). The

complexes were neutralized with Na^+ counter ions, and solvated in a water box such that at least 10.0 Å of water exists between the DNA–protein complex and water box boundary.

All minimization, heating, equilibration, and production steps were performed with the GPU-accelerated PMEMD module available in AMBER 14 or AMBER 16 (34–36). For each step, the particle mesh Ewald method was used to employ a 10.0 Å electrostatic cutoff and a 10.0 Å cutoff was also applied to the van der Waals interactions. To minimize each system, 1000 cycles of steepest descent (SD) and subsequent 1000 cycles of conjugate gradient (CG) minimization were first applied to the solvent and Na^+ counter ions, followed by the complex hydrogen atoms, and finally the complex heavy/hydrogen atoms. For the final minimization step, the entire system was subjected to 1000 cycles of SD minimization followed by 2000 cycles of CG minimization. Subsequently, each system was heated to 310 K over 120 ps with a 25.0 kcal mol^{-1} $Å^{-2}$ restraint placed on the solute using the Langevin thermostat with a time step of 1 fs under NVT conditions. The 25.0 kcal mol^{-1} $Å^{-2}$ restraint was reduced to 0.0 kcal mol^{-1} $Å^{-2}$ over 100 ps under constant temperature and pressure conditions (Berendsen barostat) using a 2 fs time step and SHAKE to constrain bonds involving hydrogen, which was followed by 1 ns of unrestrained equilibration. Several short (5–10 ns) pre-production simulations were performed for each system to ensure complex stability and resolve any minor steric clashes between the substrate and active site residues of ALKBH2 or AlkB. Representative structures based on a standardized clustering methodology (described below) were chosen from these pre-production simulations to initiate two 100 ns replicate simulations (run with different random seeds). Since both replicates lead to similar active site geometries (see Supplementary Data), one replica for each system was extended to yield the final 500 ns production simulations, which is discussed throughout the main text. Coordinates were saved from each simulation every 5 ps and analyzed over the same interval.

The CPPTRAJ (37) module was utilized for analysis of each trajectory. The occupancy of hydrogen bonds is reported for the duration of the simulation that the heavy atom distance is <3.4 Å and the hydrogen-bonding angle is $> 120°$. To examine the placement of water within the active site, a three dimensional 20 $Å^3$ grid centered on the bound substrate was generated. Dark red spheres are shown on representative structures and reflect an oxygen atom of a water molecule located in a 0.25 $Å^3$ grid space for at least 40% of the simulation. The pre-production and production trajectories were clustered according to the positions of key active site residues (bound substrate, W69, H131, D135 and E136 for AlkB, and bound substrate, F124, H171, D174 and E175 for ALKBH2). A hierarchical agglomerative clustering methodology was used to obtain a 3.0 Å minimum cutoff between clusters or four clusters were obtained. While the representative structures are static snapshots, we report key structural parameters over the entire trajectory to ensure that conclusions are representative of each simulation trajectory. All distances are reported in Å and all angles are reported in degrees.

RESULTS AND DISCUSSION

A 16mer DNA oligonucleotide (5′-GAAGACCTXGGC GTCC-3′, X = 5mC) containing 5mC in a CpG dinucleotide context was prepared through solid phase DNA synthesis with the phosphoramidite of 5mC (23–26). High resolution electrospray ionization time-of-flight (ESI-TOF) MS analysis of the oligonucleotide exhibited an m/z of 1625.281 at its –3 charge state, which is in good agreement with the theoretical m/z 1625.281 expected of the product oligonucleotide (Figure 2A and Supplementary Table S1). In the same -3 charge envelope, we also observed the ions of 5mC + Na^+ (1632.608) and 5mC + K^+ (1637.928). The observed m/z values of these species are consistent with the corresponding calculated m/z values (Supplementary Table S1).

Previously, our lab has purified the three AlkB family enzymes mentioned above and tested their repair efficiency for 3mC, 1mA, 3mT and 1mG in both ss- and ds-DNA (23,24). Similar procedures were adopted for the modification reactions of 5mC. The reaction conditions and the oligonucleotide-enzyme ratios are similar to those observed in the conversion of 5mC to its oxidative derivatives by the TET proteins reported in the literature (6–8). For each enzyme, experiments were conducted in triplicate both in the presence and absence of the enzyme with all necessary cofactors at 37°C under both ds- and ss-DNA conditions, and the reaction products were analyzed by high resolution MS to ensure the differentiation of reaction products that have very similar m/z values (24).

First, we tried to identify the new oxidative products that appeared after the enzymatic reactions. The three oxidative products, 5hmC, 5fC and 5caC, all formed in the reactions with all three enzymes, but each enzyme had a preference to generate a certain oxidative derivative. Below we use typical examples to demonstrate the formation of a certain product. The MS results of ALKBH2 oxidizing 5mC in ds-DNA (Figure 2B) showed a new oligonucleotide species that has an m/z for the monoisotopic peak at 1630.615 at –3 charge state, which corresponds well to the theoretical m/z value of the 16mer oligonucleotide containing 5hmC (1630.612 calculated, Supplementary Table S1). In the reaction of 5mC with ALKBH3 in ss-DNA, another oligonucleotide envelope appeared at m/z 1629.951 (Figure 2C), which agrees with the 5fC base in the 16mer oligonucleotide (1629.940 calculated). For the oxidation of 5mC by the AlkB protein in ss-DNA, we observed the peak envelopes of 5hmC (1630.603), 5fC (1629.933) together with a new species with an m/z value of 1642.604, which is close to the 16mer oligonucleotide containing the sodium salt of 5caC (1642.600 calculated, Figure 2D and Supplementary Table S1). Other oligonucleotide species containing metal ions, such as Na^+ and K^+, were also observed. The complete assignments of the major species generated from MS analyses are summarized in Supplementary Figure S1 and Supplementary Table S1.

The product oligonucleotides were digested into nucleosides, analyzed by LC–MS, and compared with standard nucleosides to confirm the oxidative products generated from enzymatic reactions are indeed 5hmC, 5fC and 5caC (see the Product oligonucleotides analyses section in SI,

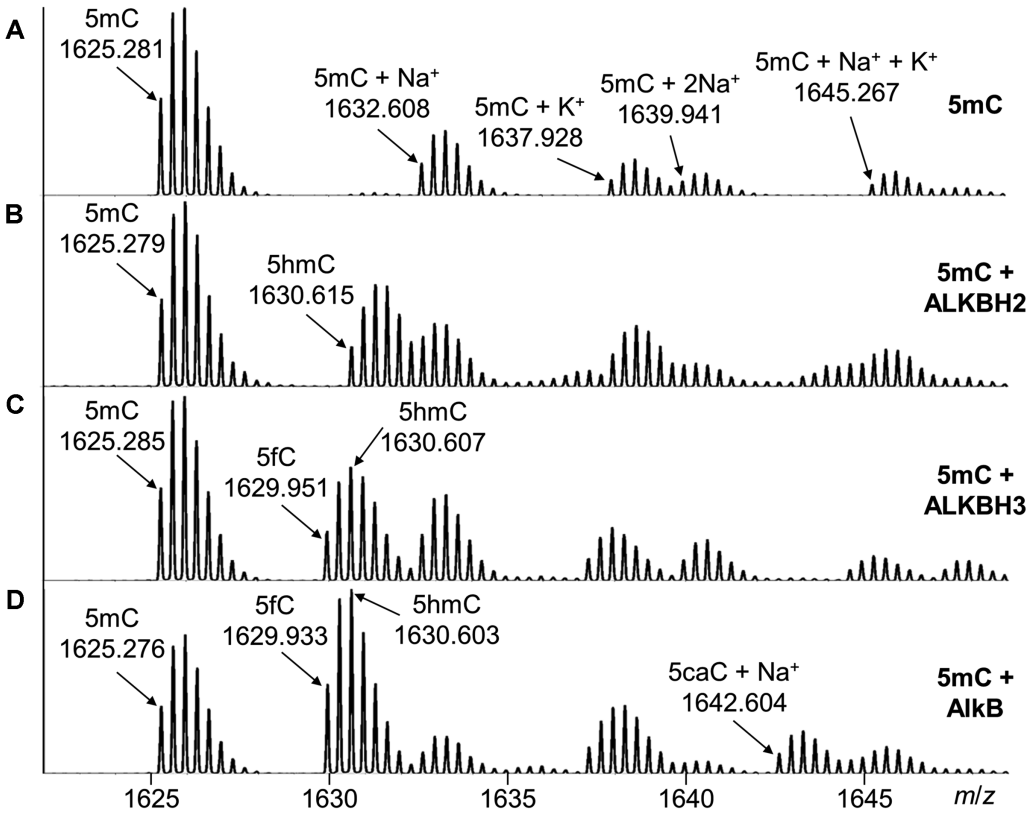

Figure 2. High resolution ESI-TOF MS analyses of 16mer DNA oligonucleotides containing 5mC and oxidized products. The observed *m/z* values represent the oligonucleotides under their –3 charge state. (**A**) 5mC; (**B**) 5mC (in ds-DNA) + ALKBH2; (**C**) 5mC (in ss-DNA) + ALKBH3 and (**D**) 5mC (in ss-DNA) + AlkB.

Supplementary Figures S2–S10). Also, to make sure the oxidations were carried out by AlkB and its homologs, we generated the catalytically inactive protein variants of AlkB: H131A, D133A and H187A (Supplementary Figure S11). The three substituted amino acids in AlkB are the key residues that coordinate the Fe(II) ion (13,38). The sequences of wild type and variant proteins were confirmed by trypsin digestion with MS analyses (Supplementary Figure S11 and Supplementary Table S2). None of the AlkB variants showed any detectable oxidative product when reacted with 5mC (Supplementary Figure S12); these observations suggest that the oxidative modifications were carried out by AlkB and not by a contaminating enzyme.

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 for oxidation in ss- or ds-DNA reactions (Figure 3 and the Product distribution for the oxidation of 5mC section in SI). The overall activities of the three AlkB proteins on oxidizing 5mC are similar to the proficiencies of the TET enzymes on modifying 5mC reported in the literature (6–8) and are generally consistent with theoretical calculations that report higher barriers for each successive oxidation steps by TET2 (39). The conversion of 5mC to the corresponding oxidative intermediates by the AlkB proteins were comparable to their repair of other known substrates (modifications on different alkyl substrates by the AlkB enzymes section in SI). In all of the enzymatic reactions, we only observed the oxidation of 5mC, but not the thymine DNA base, which naturally has a 5-methyl group. The same finding was reported for the TET family enzymes (8).

To probe the molecular basis by which the AlkB enzymes are able to oxidize 5mC, we performed molecular dynamics (MD) simulations to examine how 5mC is accommodated in the active sites of ALKBH2 and AlkB (Figure 4). We found that the 5-methyl of *anti*-5mC in our model is far from the Fe(IV)–oxo moiety (~6.3 Å for ALKBH2 in Figure 4A and ~8.2 Å for AlkB in Figure 4D; also in Supplementary Figures S13–S14, and Supplementary Tables S3–S6). In contrast, the distances between the Fe(IV)–oxo moiety and the 5-methyl group in the models with *syn*-5mC bound to ALKBH2 or AlkB are much shorter (~3.8 Å for ALKBH2 in Figure 4B and ~3.8 Å for AlkB in Figure 4E) and similar to the distances in the structures of ALKBH2 or AlkB bound to their prototypic substrate 3mC in the *anti*-conformation (~3.3 Å for ALKBH2 in Figure 4C and ~3.4 Å for AlkB in Figure 4F.; See also the Simulations of ALKBH2 and AlkB bound to 3mC, 5hmC and 5fC section in SI). With *syn*-5mC bound to ALKBH2 and AlkB, hydrogen bonds appear possible between the N^4 amino group of 5mC and active site residues. Specifically, *syn*-5mC interacts with Asp and Glu residues (D174/E175 for ALKBH2, and D135/E136 for AlkB) through water (Supplementary Tables S3–S4 and Supplementary Figures S13–S14), as well as the Y122 hydroxy group in ALKBH2 (Supplementary Table S3). These interactions likely facilitate oxidative cataly-

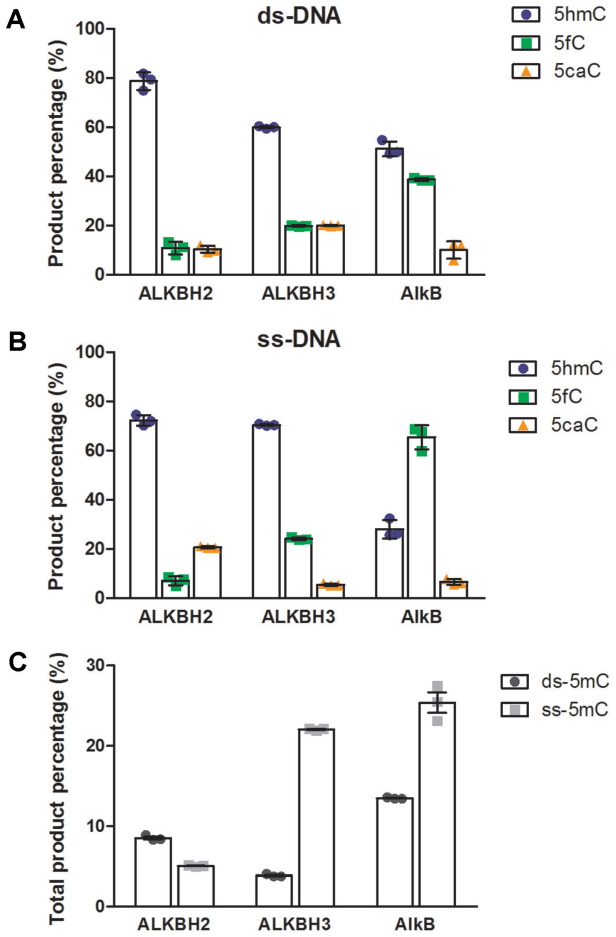

Figure 3. Product distribution and strand preference (in univariate scatterplot) for the oxidation of 5mC by the AlkB family enzymes. Reaction products (5hmC, 5fC and 5caC) generated from the reactions of 5mC with the AlkB family enzymes in (**A**) ds-DNA and (**B**) ss-DNA. (**C**) Total product percentage from reactions of the AlkB family enzymes oxidizing 5mC in ds- and ss-DNA.

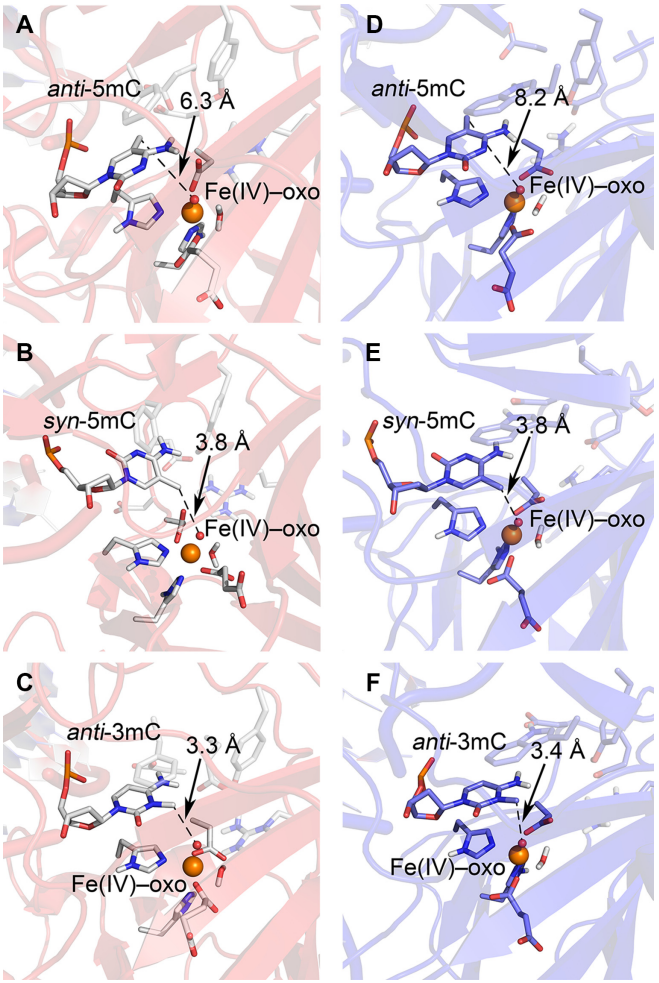

Figure 4. Representative molecular dynamics structures of the ALKBH2 (**A**–**C**) or AlkB (**D**–**F**) complex bound to *anti*-5mC (A, D), *syn*-5mC (B, E), or *anti*-3mC (C, F). The distance between the oxo-moiety and methyl groups is highlighted with dashed lines.

sis by positioning the C5 methyl group near the Fe(IV)–oxo moiety (∼3.8 Å; Supplementary Tables S5 and S6).

Interestingly, a crystal structure of TET2 co-crystallized with 5mC-containing DNA reveals *syn*-5mC in the active site (40), and the reported χ torsion angle is consistent with that predicted for *syn*-5mC in ALKBH2/AlkB (Supplementary Figure S15 and Supplementary Tables S5 and S6). Based on our combined experimental and theoretical data, we propose that the AlkB family enzymes are able to oxidize 5mC bound only in the *syn*-conformation.

To provide insight into the variable activities of ALKBH2 and AlkB for subsequent nucleobase oxidation (Figure 3), MD simulations were performed on *syn*-5hmC and 5fC bound in the active sites. For 5hmC, the C5 substituent is further from the Fe(IV)–oxo moiety for ALKBH2 (∼4.5 Å, Supplementary Figure S16 and Supplementary Tables S5) compared to AlkB (3.4 Å, Supplementary Figure S17 and Supplementary Table S6), which is consistent with the relative low abundance of the 5fC product for ALKBH2 (Figure 3, Supplementary Figures S19 to S21 and Supplementary Tables S3–S6). For 5fC, increased flexibility of the bound nucleobase may permit enhanced catalysis and generate more 5caC for ALKBH2 comparing to AlkB (Figure 3A, 3B, and Supplementary Figure S22). In addition to the insights provided by the MD simulations (see SI for detailed discussions of these calculations), several other factors could influence the product distributions including DNA binding, the *anti*/*syn* conformational preference of the substrate, oxidative reactivity, and different base flipping mechanisms used by each enzyme.

Since methylation/demethylation of 5mC is an important epigenetic process, the body may utilize several redundant or complementary pathways to control this modulation. For the oxidative modification of 5mC, ALKBH2 and 3 may work as an addition to the TET family enzymes to fulfill a similar biological role. The TET family enzymes have been reported to prefer oxidizing 5mC and its derivatives in ds-DNA (1,11). In this paper, we observed ALKBH2 and 3 are able to modify 5mC and derivatives in both ss- and ds-DNA. Especially, ALKBH3 has a much stronger preference on ss-DNA (∼6-fold) to ds-DNA (Figure 3C). Besides the oxidative transformations in ds-DNA similar to the TET

proteins, these observations suggest ALKBH2 and 3 may carry out modifications under different situations, such as 5mC in ss-DNA during DNA replication and transcription. These hypotheses certainly warrants further study.

In this paper, we demonstrated the *in vitro* oxidative modification of 5mC to 5hmC, 5fC and 5caC by the three AlkB DNA repair enzymes. Thus, the AlkB proteins are not only able to repair DNA adducts, such as 3mC and 3mT, but also can edit the epigenetic modification 5mC and generate the corresponding oxidative derivatives. These observations suggest a possible connection between DNA repair and epigenetic gene modulation. Future investigation includes analyzing the kinetic parameters of the three AlkB enzymes acting on 5mC, confirming the oxidation of 5mC by the AlkB enzymes in cell, and probing whether other α-KG/Fe(II)-dependent dioxygenases can oxidize 5mC.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGMENTS

The authors want to thank the RI-INBRE program, its directors Prof. Zahir Shaikh and Prof. Bongsup Cho, and staff Dr. Al Bach, Kim Andrews, and Patricia Murray for their kind help. We also want to thank Dr. Ang Cai, Dr. Jeremy Setser, Ms. Kerri Bradshaw, and Michael Vittori for helpful contributions.

FUNDING

Institutional Development Award from the National Institute of General Medical Sciences of the National Institutes of Health [P20 GM103430]; National Institutes of Health [R15 CA213042 and R01 ES028865 to D.L.]; Natural Sciences and Engineering Research Council of Canada [NSERC, Discovery 2016-04568, S.D.W., and CGS-D/MSFSS, S.A.P.L.]; Canada Foundation for Innovation [22770, S.D.W.]; Alberta Innovates–Technology Futures (AI-TF) (to S.A.P.L.); National Cancer Institute or National Institute of Environmental Health Sciences grants [P01 CA026731, R01 CA080024, and P30 ES002109 to J.M.E.]; C.L.D. is an HHMI investigator. Funding for open access charge: National Institutes of Health.
Conflict of interest statement. None declared.

REFERENCES

1. Wu,X. and Zhang,Y. (2017) TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat. Rev. Genet.*, **18**, 517–534.
2. Smith,Z.D. and Meissner,A. (2013) DNA methylation: roles in mammalian development. *Nat. Rev. Genet.*, **14**, 204–220.
3. Li,E. and Zhang,Y. (2014) DNA methylation in mammals. *Cold Spring Harb. Perspect. Biol.*, **6**, a019133.
4. Law,J.A. and Jacobsen,S.E. (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.*, **11**, 204–220.
5. Hausinger,R.P. and Schofield,C.J. eds. (2015) *2-Oxoglutarate-Dependent Oxygenases*. Royal Society of Chemistry. Cambridge.
6. Tahiliani,M., Koh,K.P., Shen,Y., Pastor,W.A., Bandukwala,H., Brudno,Y., Agarwal,S., Iyer,L.M., Liu,D.R., Aravind,L. *et al.* (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, **324**, 930–935.
7. Ito,S., D'Alessio,A.C., Taranova,O.V., Hong,K., Sowers,L.C. and Zhang,Y. (2010) Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature*, **466**, 1129–1133.
8. Zhang,L., Chen,W., Iyer,L.M., Hu,J., Wang,G., Fu,Y., Yu,M., Dai,Q., Aravind,L. and He,C. (2014) A TET homologue protein from Coprinopsis cinerea (CcTET) that biochemically converts 5-methylcytosine to 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxylcytosine. *J. Am. Chem. Soc.*, **136**, 4801–4804.
9. Klungland,A. and Robertson,A.B. (2017) Oxidized C5-methyl cytosine bases in DNA: 5-Hydroxymethylcytosine; 5-formylcytosine; and 5-carboxycytosine. *Free Radic. Biol. Med.*, **107**, 62–68.
10. Cimmino,L. and Aifantis,I. (2017) Alternative roles for oxidized mCs and TETs. *Curr. Opin. Genet. Dev.*, **42**, 1–7.
11. Hashimoto,H., Zhang,X., Vertino,P.M. and Cheng,X. (2015) The mechanisms of generation, recognition, and erasure of DNA 5-methylcytosine and thymine oxidations. *J. Biol. Chem.*, **290**, 20723–20733.
12. Aravind,L. and Koonin,E.V. (2001) The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases. *Genome Biol.*, **2**, RESEARCH0007.
13. Yu,B., Edstrom,W.C., Benach,J., Hamuro,Y., Weber,P.C., Gibney,B.R. and Hunt,J.F. (2006) Crystal structures of catalytic complexes of the oxidative DNA/RNA repair enzyme AlkB. *Nature*, **439**, 879–884.
14. Sedgwick,B., Bates,P.A., Paik,J., Jacobs,S.C. and Lindahl,T. (2007) Repair of alkylated DNA: recent advances. *DNA Repair*, **6**, 429–442.
15. Trewick,S.C., Henshaw,T.F., Hausinger,R.P., Lindahl,T. and Sedgwick,B. (2002) Oxidative demethylation by *Escherichia coli* AlkB directly reverts DNA base damage. *Nature*, **419**, 174–178.
16. Falnes,P.Ø., Johansen,R.F. and Seeberg,E. (2002) AlkB-mediated oxidative demethylation reverses DNA damage in *Escherichia coli*. *Nature*, **419**, 178–182.
17. Zheng,G., Fu,Y. and He,C. (2014) Nucleic acid oxidation in DNA damage repair and epigenetics. *Chem. Rev.*, **114**, 4602–4620.
18. Fedeles,B.I., Singh,V., Delaney,J.C., Li,D. and Essigman,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.
19. Delaney,J.C. and Essigman,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.
20. Li,D., Fedeles,B.I., Shrivastav,N., Delaney,J.C., Yang,X., Wong,C., Drennan,C.L. and Essigman,J.M. (2013) Removal of N-alkyl modifications from N(2)-alkylguanine and N(4)-alkylcytosine in DNA by the adaptive response protein AlkB. *Chem. Res. Toxicol.*, **26**, 1182–1187.
21. Yi,C., Chen,B., Qi,B., Zhang,W., Jia,G., Zhang,L., Li,C.J., Dinner,A.R., Yang,C.-G. and He,C. (2012) Duplex interrogation by a direct DNA repair protein in search of base damage. *Nat. Struct. Mol. Biol.*, **19**, 671–676.
22. Yi,C., Jia,G., Hou,G., Dai,Q., Zhang,W., Zheng,G., Jian,X., Yang,C.-G., Cui,Q. and He,C. (2010) Iron-catalysed oxidation intermediates captured in a DNA repair dioxygenase. *Nature*, **468**, 330–333.
23. Chen,F., Tang,Q., Bian,K., Humulock,Z.T., Yang,X., Jost,M., Drennan,C.L., Essigman,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.
24. Chen,F., Bian,K., Tang,Q., Fedeles,B.I., Singh,V., Humulock,Z.T., Essigman,J.M. and Li,D. (2017) Oncometabolites d- and l-2-Hydroxyglutarate inhibit the AlkB family DNA repair enzymes under physiological conditions. *Chem. Res. Toxicol.*, **30**, 1102–1110.
25. Bian,K., Chen,F., Humulock,Z.T., Tang,Q. and Li,D. (2017) Copper inhibits the AlkB family DNA repair enzymes under Wilson's disease condition. *Chem. Res. Toxicol.*, **30**, 1794–1796.
26. Tang,Q., Cai,A., Bian,K., Chen,F., Delaney,J.C., Adusumalli,S., Bach,A.C., Akhlaghi,F., Cho,B.P. and Li,D. (2017) Characterization

of byproducts from chemical syntheses of oligonucleotides containing 1-Methyladenine and 3-Methylcytosine. *ACS Omega*, **2**, 8205–8212.
27. Quinlivan,E.P. and Gregory,J.F. (2008) DNA digestion to deoxyribonucleoside: a simplified one-step procedure. *Anal. Biochem.*, **373**, 383–385.
28. Frick,L.E., Delaney,J.C., Wong,C., Drennan,C.L. and Essigman,J.M. (2007) Alleviation of 1,N6-ethanoadenine genotoxicity by the *Escherichia coli* adaptive response protein AlkB. *Proc. Natl. Acad. Sci. U. S. A.*, **104**, 755–760.
29. Case,D.A., Betz,R.M., Cerutti,D.S., Cheatham,T.E. III, Darden,T.A., Duke,R.E., Giese,T.J., Gohlke,H., Goetz,A.W., Homeyer,N. *et al.* (2016) *AmberTools16*. University of California, San Francisco.
30. Anandakrishnan,R., Aguilar,B. and Onufriev,A.V. (2012) H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.*, **40**, W537–W541.
31. Li,P. and Merz,K.M. (2016) MCPB.py: A Python Based Metal Center Parameter Builder. *J. Chem. Inf. Model.*, **56**, 599–604.
32. Frisch,M.J., Trucks,G.W., Schlegel,H.B., Scuseria,G.E., Robb,M.A., Cheeseman,J.R., Scalmani,G., Barone,V., Petersson,G.A., Nakatsuji,H. *et al.* (2016) *Gaussian 09*. Revision D.01 Gaussian, Inc., Wallingford CT.
33. Seminario,J.M. (1996) Calculation of intramolecular force fields from second-derivative tensors. *Int. J. Quantum Chem.*, **60**, 1271–1277.
34. Salomon-Ferrer,R., Götz,A.W., Poole,D., Le Grand,S. and Walker,R.C. (2013) Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald. *J. Chem. Theory Comput.*, **9**, 3878–3888.
35. Götz,A.W., Williamson,M.J., Xu,D., Poole,D., Le Grand,S. and Walker,R.C. (2012) Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.*, **8**, 1542–1555.
36. Le Grand,S., Götz,A.W. and Walker,R.C. (2013) SPFP: Speed without compromise—a mixed precision model for GPU accelerated molecular dynamics simulations. *Comput. Phys. Commun.*, **184**, 374–380.
37. Roe,D.R. and Cheatham,T.E. (2013) PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.*, **9**, 3084–3095.
38. Yu,B. and Hunt,J.F. (2009) Enzymological and structural studies of the mechanism of promiscuous substrate recognition by the oxidative DNA repair enzyme AlkB. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 14315–14320.
39. 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.
40. 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.