Supplementary Information

Iron-Catalyzed Oxidation Intermediates Captured in a DNA Repair Dioxygenase

Chengqi Yi¹, Guifang Jia¹, Guanhua Hou², Qing Dai³, Wen Zhang¹, Guanqun Zheng¹, Xing Jian¹, Cai-Guang Yang¹,⁴, Qiang Cui², and Chuan He¹

¹Department of Chemistry and Institute for Biophysical Dynamics, ²Department of Biochemistry and Molecular Biology, The University of Chicago, 929 East 57th Street, Chicago, Illinois 60637, USA.

²Department of Chemistry and Theoretical Chemistry Institute, University of Wisconsin, Madison 1101 University Avenue, Madison, Wisconsin 53706, USA.

⁴Shanghai Institute of Materia Medica, Chinese Academy of Sciences, 555 Zuchongzhi Road, Shanghai 201203, China.
**Supplementary Table 1. Data collection and model statistics**

|                  | (Mn/αKG) | (Mn/αKG) | (Mn/αKG) | (Mn/αKG) | (Fe/succinate) | (Fe/succinate) | (Fe/succinate) |
|------------------|----------|----------|----------|----------|----------------|----------------|----------------|
| AlkB-DNA: εA     | P12₁₁    | P12₁₁    | P12₁₁    | P12₁₁    | P12₁₁          | P12₁₁          | P12₁₁          |
| AlkB-DNA: 3-meT  |          |          |          |          |                |                |                |
| AlkB-DNA: 3-meC  |          |          |          |          |                |                |                |
| AlkB-DNA: 3-deazaMeC |      |          |          |          |                |                |                |
| AlkB-DNA: glycol |          |          |          |          |                |                |                |
| AlkB-DNA: hemiaminal |      |          |          |          |                |                |                |
| AlkB-DNA: oxidized |      |          |          |          |                |                |                |
| AlkB-DNA: 3-meC  | P12₁₁    | P12₁₁    | P12₁₁    | P12₁₁    | P12₁₁          | P12₁₁          | P12₁₁          |
|                  |          |          |          |          |                |                |                |
|                  |          |          |          |          |                |                |                |
|                  |          |          |          |          |                |                |                |
|                  |          |          |          |          |                |                |                |
|                  |          |          |          |          |                |                |                |
|                  |          |          |          |          |                |                |                |
|                  |          |          |          |          |                |                |                |
|                  |          |          |          |          |                |                |                |
|                  |          |          |          |          |                |                |                |
|                  |          |          |          |          |                |                |                |
|                  |          |          |          |          |                |                |                |
|                  |          |          |          |          |                |                |                |
|                  |          |          |          |          |                |                |                |
|                  |          |          |          |          |                |                |                |
|                  |          |          |          |          |                |                |                |
|                  |          |          |          |          |                |                |                |
|                  |          |          |          |          |                |                |                |
|                  |          |          |          |          |                |                |                |

| **Data collection** |
|---------------------|
| Space group         | P12₁₁    | P12₁₁    | P12₁₁    | P12₁₁    | P12₁₁          | P12₁₁          | P12₁₁          |
| Cell dimensions     |
| $a, b, c$ (Å)       | 41.4, 75.9, 51.4 | 41.4, 75.9, 50.8 | 41.4, 76.0, 51.2 | 41.5, 75.8, 52.0 | 41.1, 75.9, 51.2 | 41.1, 75.9, 51.2 | 41.4, 76.0, 51.7 | 41.6, 76.2, 52.4 |
| $\alpha, \beta, \gamma$ (°) | 90, 107.9, 90 | 90, 108.3, 90 | 90, 107.8, 90 | 90, 108.2, 90 | 90, 107.2, 90 | 90, 107.2, 90 | 90, 107.9, 90 | 90, 108.0, 90 |
| Resolution* (Å)    | 50 – 1.51 | 50 – 1.75 | 50 – 1.50 | 50 – 1.77 | 50 – 1.54 | 50 – 1.62 | 50 – 1.48 | 50 – 1.90 |
| ($1.58 – 1.51$)     | ($1.82 – 1.75$) | ($1.55 – 1.50$) | ($1.83 – 1.77$) | ($1.62 – 1.54$) | ($1.73 – 1.62$) | ($1.53 – 1.48$) | ($1.97 – 1.90$) |
| $R_{merge}$        | 0.04 (0.576) | 0.09 (0.401) | 0.04 (0.368) | 0.07 (0.618) | 0.05 (0.351) | 0.07 (0.412) | 0.04 (0.435) | 0.10 (0.642) |
| $I/\sigma I$       | 14.4 (2.2) | 12.1 (2.3) | 19.2 (2.1) | 12.9 (2.1) | 16.8 (2.3) | 12.9 (2.3) | 17.9 (2.0) | 10.6 (2.0) |
| Redundancy | 3.4 (2.9) | 3.3 (3.2) | 3.5 (3.4) | 3.8 (3.8) | 3.4 (3.4) | 3.5 (3.6) | 2.7 (2.4) | 3.3 (3.4) |
|------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| **Refinement** | | | | | | | | |
| Resolution (Å) | 20 – 1.51 | 15 – 1.92 | 20 – 1.75 | 20 – 1.77 | 20 – 1.54 | 15 – 1.62 | 15 – 1.58 | 15 – 1.90 |
| | (1.55 – 1.51) | (1.97 – 1.92) | (1.80 – 1.75) | (1.82 – 1.77) | (1.58 – 1.54) | (1.66 – 1.62) | (1.62 – 1.58) | (1.95 – 1.90) |
| Completeness (%) | 99.4 (100) | 96.5 (85.8) | 98.0 (85.3) | 99.7 (99.4) | 96.2 (100) | 94.6 (70.8) | 95.3 (78.8) | 96.5 (99.5) |
| No. reflections | 38952 | 21307 | 28324 | 28207 | 38818 | 34246 | 37713 | 22436 |
| $R_{work}$/$R_{free}$ | 18.0/21.5 | 17.4/22.2 | 16.6/19.1 | 16.3/21.9 | 19.5/22.9 | 18.8/22.6 | 19.7/24.3 | 19.0/23.8 |
| Complex in asymmetric unit | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Protein residues | 13 - 214 | 14 - 213 | 14 - 214 | 13 - 214 | 13 - 214 | 13 - 214 | 13 - 214 | 14 - 215 |
| Nucleotides | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 |
| Ligands | Mn(II):1; αKG:1; | Mn(II):1; αKG:1; | Mn(II):1; αKG:1; | Mn(II):1; αKG:1; | Iron:1; succinate:1; | Iron:1; succinate:1; | Iron:1; succinate:1; | Iron:1; succinate:1; |
|                        | Ethanethiol:1 | Ethanethiol:1 | Ethanethiol:1 | Ethanethiol:1 | Ethanethiol:1 | Ethanethiol:1 | Ethanethiol:1 | Ethanethiol:1 |
|------------------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| Most favoured# (%)     | 97.9          | 96.9          | 97.9          | 97.9          | 97.9          | 97.9          | 97.9          | 97.9          |
| Additionally allowed (%)| 1.6           | 2.6           | 1.6           | 1.6           | 1.6           | 1.6           | 1.6           | 1.6           |
| Disallowed (%)         | 0.5           | 0.5           | 0.5           | 0.5           | 0.5           | 0.5           | 0.5           | 0.5           |

R.m.s deviations

| Bond lengths (Å) | 0.005 | 0.009 | 0.006 | 0.013 | 0.008 | 0.007 | 0.009 | 0.018 |
|------------------|-------|-------|-------|-------|-------|-------|-------|-------|
| Bond angles (°)  | 1.212 | 1.329 | 1.180 | 1.636 | 1.321 | 1.283 | 1.321 | 2.066 |
| PDB accession code | 3O1P  | 3O1O  | 3O1M  | 3O1R  | 3O1U  | 3O1T  | 3O1S  | 3O1V  |

* Highest resolution shell is shown in parenthesis.

# Values calculated using PROCHECK from CCP4 suite.
Supplementary Table 2. Modifications to the restraints of 3-hydroxymethyl cytosine (3-hmC) dictionary file (the three restraints in the original 3-hmC dictionary file were not used during REFMAC refinement).

| Bond or angle          | Original dictionary | In refined model |
|------------------------|---------------------|-----------------|
| C20-N3                 | Single: 1.465±0.02 Å| 1.50 Å          |
|                        |                     |                 |
| O21-C20                | Single: 1.432±0.02 Å| 1.22 Å          |
|                        |                     |                 |
| C20 vs cytosine        | Coplanar with base atoms | Slightly tilted |

3-hmC with certain atoms labelled

![3-hmC structure](image)
Supplementary Figure 1. Capture and characterization of oxidative demethylation/dealkylation intermediates in the AlkB-mediated DNA repair. Color coding and labels are the same as used in Fig. 1.
**Supplementary Figure 2. Repair of DNA/RNA lesions by AlkB.**

**a.** The three types of lesions that are oxidatively repaired by AlkB, using iron(II), αKG, and dioxygen.

**b.** Proposed reaction mechanism of AlkB. The first phase of the reaction involves the activation of dioxygen and the second phase is the substrate oxidation.

= can be all types of substrates in a.
Supplementary Figure 3. Crystal structures of AlkB complexed to a dsDNA containing εA, 3-meT and 3-meC, respectively. a, The overall structures. b, Active site views and the metal-binding site. Protein is colored in green, DNA in orange, damaged bases in magenta, manganese(II) in yellow, water in red, and αKG in blue.
Supplementary Figure 4. Stereo views of AlkB active sites when the protein binds to (a) εA, (b) 3-meT, and (c) 3-meC, respectively. The bases are invariantly sandwiched between Trp 69 and His 131; the cyclic adduct portion of εA also stacks against the side chain of Asp 133. Same color scheme is used as in Supplementary Fig. 3. Hydrogen bonds are illustrated as dotted lines.
Supplementary Figure 5. Activity assays of wild-type AlkB and several mutant proteins (D135A, D135S, and D135N) with trimer DNA substrates. a, Activity with T-(1-meA)-T. b, Activity with T-(1-meG)-T. The D135S mutant protein shows
greatly increased activity towards T-(1-meG)-T. e, Lesions in category (1) have an exocyclic amino group (highlighted in orange) that forms a direct hydrogen bond with the side chain of Asp 135, which contributes to the more favorable recognition of these lesions. Lesions in the other two categories have hydrogen bond acceptors (highlighted in blue) at the equivalent sites instead. For all activity assays, a 100 μL reaction solution containing 50 mM buffer (Tris: pH 7.0), 2 mM ascorbic acid, 1 mM αKG, 0.28 mM (NH₄)₂Fe(SO₄)₂, and 1 nmol DNA substrate was incubated with 0.1 nmol AlkB protein at 37 °C for 3 h. The reaction was terminated by adding EDTA to 5 mM and analyzed by HPLC with a C18 separation column (150 × 4.6 mm) equilibrated with buffer A (HPLC grade aqueous solution containing 50 mM ammonium acetate) and buffer B (50 mM ammonium acetate, 50% of acetonitrile, 50% water and 0.1% trifluoroacetic acid) at a flow rate of 1 ml min⁻¹ at room temperature. Repair percentage was quantified using the area of the corresponding peaks. The standard deviation was generated based on two runs of HPLC analyses and the experiments were duplicated. *In Panel a, we observed complete repair of 1-meA by the wild-type AlkB under the same reaction conditions used for other AlkB mutants; the wild-type AlkB is much more active towards 1-meA over the AlkB mutants tested.
Supplementary Figure 6. NBO charge analysis of 3-meC. a, NBO charges of the 3-meC base are labeled in black; key bond distances are also shown. The formal positive charge on N3 nitrogen atom is fairly well distributed in the conjugated system, and the amino group bears a large portion of the charge. This allows 1-meA and 3-meC to be preferentially recognized by AlkB. b, Leading resonance structures of 3-meC.
Supplementary Figure 7. Comparison between lesions similar in size. 

a, εA vs 1-meA.  
b, 3-meT vs 3-meC. When these structures were superimposed, only the protein portions of the structures were used so as to allow unbiased comparison of the locations of the damaged bases in the active site. As is shown (only the flipped nucleotides of these overlaid structures are shown), lesions similar in size overlap very well with each other regardless of their hydrogen bonding capacities. εA and 3-meT are colored in magenta while 1-meA and 3-meC in grey.
Supplementary Figure 8. Overlay of AlkB-dsDNA structures containing εA and 3-meC. **a**, Local view of the base-flipping region of AlkB. The “pinch” sequence and Loop1 (colored in blue) adjust the positions of DNA backbone 5’ (highlighted in yellow) to the flipped base, and Loop2 and Loop3 (in cyan) anchor the 3’ DNA backbone (labeled in hollow yellow circles). **b**, Closer view of the active site. The direct hydrogen bond between 3-meC and Asp 135 side chain is shown as a dotted line. In both views, black arrows indicate backbone displacement. Red arrows illustrate “buffer effect” in **b** and the orange curve represents the hydrophobic surface formed by the side chains of Met 57, Val 59, Met 61, and Leu 128.
Supplementary Figure 9. Overall views of the oxidized (a) εA and (b) 3-meT structures. Succinate, which is the oxidized product of αKG, is colored in blue and the rest of the color coding is the same as in Supplementary Fig. 3.
Supplementary Figure 10. Three different de-alkylation mechanisms proposed for (a) εA, (b) 3-meT, and (c) 3-meC, based on intermediates observed and calculated in this study. The presence of water molecules (all are involved in indirect H-bonds between protein residues and the bases) and general base represents observed interactions, and the usage of protons (observed as water molecules in the crystal structures) highlights the requirements of protonation of specific atoms in order to release the alkyl adducts.

Supplementary Figure 11. Geometry of succinate in the glycol 1 and hemiaminal 2 structures. Overlaid structures at the metal-binding sites are shown, with the glycol 1 structure colored as in Supplementary Fig. 3 and the hemiaminal structure 2 colored in grey. The succinate molecules are highlighted in the orange box.
Electron density maps using a 3-meC model

Electron density maps using a 3-hmC model

Electron density maps using cytosine and water
Supplementary Figure 12. Initial attempts to build a 3-meC, 3-hmC, or cytosine (plus a water molecule) model for the oxidized 3-meC crystal. A, Model with a 3-meC resulted in an excess of positive \( F_{\text{obs}} - F_{\text{cal}} \) density (the big green sphere) during refinement. This model was subjected to simulated annealing refinement using Phenix and similar green globular density was also observed on top of the methyl group. B, Then, we tried to use 3-hmC to replace 3-meC during the refinement. Such a 3-hmC model resulted in an excess of negative \( F_{\text{obs}} - F_{\text{cal}} \) density (the two red spheres) at the positions of C20 and O21. Note C20 atom doesn’t fit well into the density either. C, Model with a cytosine and a water molecule. In addition to the green density, the water molecule, which doesn’t sit in the center of the density, is already too close to the N3 nitrogen atom (~2.3 Å) of cytosine. The blue \( 2F_{\text{obs}} - F_{\text{cal}} \) maps are contoured at 1.0\( \sigma \) and the \( F_{\text{obs}} - F_{\text{cal}} \) maps are contoured at 2.8\( \sigma \) (green) and -3.0\( \sigma \) (red). The color scheme of the nucleotides is the same as in Fig. 2.
Supplementary Figure 13. The best fitting model of the oxidized 3-meC structure. 

**a**, Electron density maps of the oxidized base are shown. In this model, all atoms of the base are set to 100% occupied and the temperature factors of C20 and O21 are comparable to the rest of the base atoms. **b**, A mixture model using both 3-meC (~50% occupancy) and cytosine (plus a water molecule, also ~50% occupied). The model doesn’t fit the density as well as the model in **a**. **c**, Overlay of the model shown in **a** (colored) and the mixture model in **b** (grey). The blue 2F_{obs}-F_{cal} maps are contoured at 1.0σ and the green F_{obs}-F_{cal} simulated annealing omit maps are contoured at 3.0σ in **a**. The color scheme of the nucleotides is the same as in Fig. 2.
Supplementary Figure 14. Chemical synthesis of 3-deazameC phosphoramidite. Compound IV was synthesized according to literature methods and was directly isolated from the α-anomer by careful column chromatography. Deprotection of the N-PhOAc group was achieved with 180 microliter 50% cystamine in water at 55 °C for 20 h. For compound XIII: $^{31}$P NMR (202.5 MHz) (CD$_3$CN) $\delta$: 148.3 and 148.4 ppm; High Resolution Mass Spectrometry of XIII $C_{49}H_{57}N_4NaO_9P$, [MNa]$^+$ 899.3761 (calculated); 899.3753 (found).
Supplementary Figure 15. Crystal structure of AlkB-dsDNA containing 3-deazameC. 

a, The overall structure is shown. Mn(II) and αKG were co-crystallized with the complex. 
b, View of the metal binding site. 
c, Side chain of D135 and E136 (through a water molecule) still form hydrogen bonds with the exo-cyclic amino group of 3-deazameC. 
d, Electron density of the 3-deazameC. The blue 2F_{obs}-F_{cal} maps are contoured at 1.5σ. The rest of the color scheme is the same as that of Supplementary Fig. 3.
Supplementary Figure 16. Activity assay of 3-deazameC with AlkB. A 7-mer 3-deazameC-containing DNA was synthesized using solid-state DNA synthesizer and was subjected to repair assay similar to that described in figure caption of Supplementary Fig. 5. The reaction mixture was incubated at 16 °C for 18 h and results at pH =5.0 are shown. a, HPLC analysis of the purified 7-mer DNA, with a C18 column. b, HPLC profile of the repair mixture, which contains 50 mM buffer (sodium acetate: pH 5.0), 2 mM ascorbic acid, 1 mM αKG, 0.28 mM (NH4)2Fe(SO4)2, 1 nmol the 7-mer DNA substrate, and 1 nmol AlkB. c, MALDI-TOF analysis of the synthesized 7-mer DNA, which has a molecular weight (MW) of about 2061. The set of peaks with +22 in MW are from the addition of a Na+ to the 7-mer DNA. d, MALDI-TOF analysis of the repair mixture (which contains both the 7-mer substrate and the 7-mer product containing 3-deazahmC). The presence of a new set of peaks (highlighted with the blue box), with +16 in molecular weight compared to the substrate peaks (MW: 2061), indicates the formation of 3-deazahmC.
Supplementary Figure 17. QM/MM optimized structure of the 3-deazahmC crystal structure. **a**, Additional view of the same optimized structure shown in Fig. 3b, from Set 2 calculations. **b**, Overlay of the optimized 3-deazahmC-containing structure with the crystal structure (yellow). For the optimized structure, carbon atoms are colored in cyan, nitrogen in blue, oxygen in red, iron in pink, and hydrogen in grey. Red NBO charges are labeled for several base atoms and key distances are marked in black (Å).
a

\[
\begin{align*}
\text{zwitterion 5} & \\
& \\
\end{align*}
\]

b

\[
\begin{align*}
\text{zwitterion 5} & \\
& \\
\end{align*}
\]

c

\[
\begin{align*}
\text{Computed 3-deazahmC by N3 to C3 substitution.} & \\
& \\
\end{align*}
\]
**Supplementary Figure 18. Several optimized structures.** a, In the optimized structure of 3-deazahmC from Supplementary Fig. 16a, C3 is replaced by a nitrogen atom and the resulting structure is subjected to QM/MM optimization (Set 2 calculations). The output structure is shown on the left, which contains a zwitterions 5 and water ligand to the iron. An overlay of 5 with 3-deazahmC 4 crystal structure (yellow) is shown on the right. b, Minimized structure using the oxidized 3-meC crystal structure (same structure as shown in Fig. 3c, from Set 1 calculations). On the right is a superimposition of the output minimized structure with the input crystal structure (yellow). c, The N3 nitrogen atom of the minimized structure shown in b is replaced by a carbon atom and the whole structure is again subjected to optimization (Set 1 calculations). The resulting structure (shown on the left) looks very similar to the computed 3-deazameC structure in Supplementary Fig. 16, with the key features—iron(II)-OH and the hydroxymethyl state—reproduced. An overlay of this structure with the crystal structure (yellow) containing 4 is shown on the right. Labels of charge and distance are the same as used in Supplementary Fig. 17.
Supplementary Figure 19. Comparisons between the “same” structures computed from the two independent calculation sets. 

a, 3-deazahmC 4 optimized from Set 2 calculations (left) and 3-deazahmC computed from Set 1 calculations (right). 
b, Zwitterion 3 from Set 1 calculations (left) and Zwitterion 5 from Set 2 calculations (right). As is shown, the “same” species obtained independently from the two sets of calculations are highly similar to each other. Labels of charge and distance are the same as used in Supplementary Fig. 17.
**Supplementary Figure 20. The pKa difference of zwitterion 3 and 3-deazahmC 4.**

Either starting point of the two calculation sets can produce both the two intermediates (the zwitterion and 3-deazahmC) during optimizations. Set 1 calculations are indicated with red arrows and Set 2 calculations with green arrows.
Supplementary Figure 21. Reaction schemes of several iron(II)/αKG-dependent dioxygenases. 

a, Oxidative demethylation of 3-meC by AlkB. 
b, Oxidative demethylation of a trimethyl-lysine residue by JHDM3 proteins. 
c, Oxidative demethylation of 3-meU by FTO. 
d, Oxidation of 5-meC by TET1 yields 5-hmC, which is a stable alcohol. Intermediates proposed in the brackets may also exist in these oxidative demethylation processes.
Supplementary Figure 22. AlkB-dsDNA complex containing 3-meC was co-crystallized with Mn(II) and αKG, and was treated (air exposure) the same way as that of Fe(II)/αKG-containing crystals. a, A crystal with Mn(II) and αKG was exposed to air for 2 h and its structure was solved to 1.92 Å. The αKG cofactor molecule was omitted to calculate the simulated annealing omit map; it has good electron density, indicating that Mn(II) is incapable of oxidation catalysis. b, View of the 3-meC nucleotide from the same structure. No additional difference map is observed in close vicinity of the methyl group when a 3-meC base is modeled. The entire base was omitted during simulated annealing procedure. c, Another crystal with Mn(II) and αKG was exposed to air for 6 h and its structure was solved to 1.98 Å. d, View of the 3-meC nucleotide from the 6 h-air-exposed structure. The blue 2F_{obs} - F_{cal} maps are all contoured at 1.5σ and the green F_{obs} - F_{cal} simulated annealing omit maps are contoured at 3.0σ. The color scheme of the nucleotides is the same as in Fig. 2.
Supplementary Notes

Enzymatic recognition of substrates with different base-dimensions. Least-square superpositions using only the protein portion of the four structures reveal that damaged bases similar in size overlap perfectly when bound in the active site, regardless of their hydrogen bonding capacities (Supplementary Fig. 7). However, the positions of damaged purine (1-meA or εA) and pyrimidine (3-meT or 3-meC) are consistently different (Supplementary Fig. 8). When AlkB is presented with a damaged pyrimidine, the DNA backbone (lesion containing strand) immediately 5′ to the flipped base is lifted further up towards the enzyme (~0.4-0.5Å) by the “pinch” sequence (Thr 51 to Gly 53) and a flexible loop (Loop1), so that smaller substrates can still be inserted deep into the active site (Supplementary Fig. 8a). Similar displacement was also observed in an earlier report1. Meanwhile, the 3′ DNA backbones overlay perfectly among all these structures (Supplementary Fig. 8a), suggesting that the small but consistent displacement at 5′ backbones is part of the enzymatic solution to recognize damaged bases with varying size. In the active site, AlkB uses Tyr 78, Arg 210, and a substrate recognition lid (Lys 134 to Glu 136) to contact the damaged base, and the positions of these residues can vary in order to accommodate the flipped base distinct in size. Such a “buffer” effect has also been proposed for the AlkB human homologue ABH32. In addition, side chains of Met 57, Val 59, Met 61, and Leu 128 constitute a hydrophobic fence so that the damaged base is anchored properly in the direction perpendicular to that of the insertion (Supplementary Fig. 8a). Through a collection of enzymatic contacts on both the DNA backbone and the flipped base, the final position of the substrate in the binding pocket is controlled so that the aberrant alkyl group is well positioned towards the metal site, allowing efficient catalysis of substrates that differ in base-dimension.

Discussions for in crystallo oxidation reactions. The first phase of the oxidative reactions (activation of dioxygen molecules, see Supplementary Fig. 2b) seems to
proceed very quickly in our single crystals. We have yet to be able to trap any dioxygen-activation intermediate. In a previous study, AlkB crystals with a T-(1-meA)-T trimer DNA were exposed to air with no oxidation intermediate reported\(^1\). In our system, we have a much longer DNA complexed to the enzyme, which is known to enhance the activity of the enzyme\(^3\). The presence of a disulfide cross-link locked the flipped base in the enzymatic pocket, providing confined nano-environment and restricted protein dynamics so that certain intermediates become long-lived enough to be structurally characterized. In addition, one notable feature of our crystals is that they are very two-dimensional, with a typical size of 300 micron × 300 micron × 5 micron. Such “plate” shape may allow oxygen to fully diffuse into a crystal for complete oxidation.

**Stability of the intermediates in crystallo.** Although formation of intermediates 1, 2, and 4 are fast in the crystals, they seem to be quite stable in the single crystals of the cross-linked AlkB-dsDNA complexes. The occupancy of the extra atoms of the intermediates is all set to 100\%, and the temperature factors of these atoms are similar to those of the DNA base atoms nearby. In fact, the intermediates seem to be stable for up to several days. After initial air exposure, the same intermediates were also observed crystallographically in such crystals after several days as well as after 2 h of air exposure. For all the intermediates described in this paper, we have collected at least 4 independent datasets and the same intermediates are repeatedly observed.

**Ligand modeling.** All the intermediate structures captured were modeled initially as regular damaged bases (εA, 3-meT, and 3-meC). Further modifications (bond lengths, angle, and the presence of additional atoms) were built into the density from \(F_{\text{obs}} - F_{\text{cal}}\) maps, and the presence of extra atoms on the damaged base was verified by the presence of positive \(F_{\text{obs}} - F_{\text{cal}}\) density from simulated annealing omit maps.
**ab initio QM/MM calculation set-up and protocols.** QM/MM minimizations were carried out for the ternary complex consisting of AlkB, succinate, and the dsDNA that contains either the oxidized 3-meC 3 or 3-deazahmC 4. To better contrast the intrinsic structural features of the two systems, two different starting points were adopted: (1) the crystal structure with the oxidized 3-meC, model 3, was used as the starting point for optimization; then the N3 in the optimized zwitterion was replaced by a carbon atom for 3-deazahmC calculations; (2) the starting model was the crystal structure that contains 4. After obtaining a stable model that overlapped well with the crystal structure, C3 was replaced by a nitrogen atom so as to mimic the situation for 3-meC oxidation (Supplementary Figs. 17-19). In the following, these two independently performed optimizations will be referred to as Set 1 and Set 2 calculations. In both cases, starting from the PDB structures, hydrogen atoms were added with the HBUILD module in CHARM 4.5. All basic and acidic amino acids were kept in their physiological protonation state. Water molecules were added following the standard protocol of superimposing the system with a water sphere of 16 Å radius and removing water molecules within 2.8 Å from any atom resolved in the crystal structure 6. The calculations were carried out using the QChem program interfaced with CHARMM (c36a2 version) 5.7.

During minimization, all residues that contain any atom within 5 Å of the iron were fully flexible while the rest were fixed at the positions in the crystal structure; moreover, most of the nucleotides of the DNA in the crystal structure were deleted except the flipped bases and their two nearby neighboring nucleotides. This leaves 194/204 mobile atoms among 3,548/3561 atoms in the Set 1/Set 2 system; despite simplicity, this set up is appropriate for the current purpose of investigating active-site geometry.

The QM region includes all key fragments in the active site: the iron site, side-chains of His 131, His 187, Asp 133, Arg 210, the oxidized co-factor succinate, the modified bases, and all water molecules within 5 Å of the iron. Link atoms were
added between Cα and Cβ of all protein side chains and between the pyrimidine/pyridine and deoxyribose of the modified bases. The treatment of the QM/MM frontier follows the EXGR scheme in CHARMM, in which the QM atoms interact with all MM atoms electrostatically except for the MM group at the QM/MM boundary; previous benchmark calculations have shown that this scheme generally gives reliable results for structure and energetics in QM/MM calculations\(^8\). The QM atoms were treated with the B3LYP functional and the 6-31G(d,p) basis set\(^9\)-\(^12\), while the MM atoms were treated with the all-atom CHARMM22 force field\(^13\). The convergence criteria for geometry minimization was carefully chosen so that the root-mean-square force on the mobile atoms was smaller than 0.3 kcal/(mol·Å) and the maximum force was smaller than 0.45 kcal/(mol·Å).

To analyze the electronic structure of the active site, NBO analysis was carried out at the minimized structure using the same QM/MM method with NBO 5.0\(^14\).

**Choice of charge and spin states.** For Set 1 calculations, based on the proposed reaction mechanism for AlkB family proteins\(^15\), the O209 and iron in the relevant intermediate should together bear a +2 charge. The most likely state is Fe(II)+H\(_2\)O, although the state of Fe(III)+OH\(^-\) cannot be completely ruled out. We have tested both possibilities, and quintet and quartet spin states were studied, respectively. As shown in Supplementary Fig. 18b, the QM/MM optimized structure for Fe(II)+H\(_2\)O agrees very well with the crystal structure; the optimized structure (not shown) for the Fe(III)+OH\(^-\) state has fairly large structural deviations, especially for iron-ligand distances.

To illustrate the \(pK_a\) difference between 3 and its analog 4, we started from the optimized structure for the zwitterion 3 and replaced N3 by a carbon atom. During QM/MM optimization, a spontaneous proton transfer occurred from the iron-bound water (water 209) to the -CH\(_2\)O\(^-\) group, while other structural properties remained
largely constant (see Supplementary Figs. 18c, 19). With the -CH₂O⁻ group protonated in 4 and Fe(II)+H₂O, QM/MM optimization also led to a stable structure (not shown).

In Set 2 calculation, starting from the crystal structure of 3-deazahmC 4, O209 was first assigned as a water molecule. Both quintet and triplet states were studied. For both cases, there were substantial deviations from the crystal structure especially for the Fe-O209 distance and the total coordination number of the iron. By contrast, when O209 was assigned as hydroxide for the quintet state, the optimized structure fits well with the crystal structure, as shown in Supplementary Fig. 17. With this set up, we again tested the $pK_a$ difference between 3-hmC and 3-deazahmC 4 by replacing C3 with a nitrogen atom starting from the optimized structure for 3-deazahmC. We observed a reverse spontaneous proton transfer from the –CH₂OH group to the iron-bound hydroxide, while other structural properties remain largely constant, as shown in Supplementary Fig. 18a. In other words, both Set 1 and Set 2 calculations support that the intermediate involving the oxidized 3-meC is a zwitterionic species.

**ab initio calculations for model systems in the gas-phase.** Several small model systems were studied to better understand the electronic structure of the modified base in the oxidized 3-meC 3 and 3-deazahmC 4 with different protonation states (not shown). The pyrimidine ring was terminated by a hydrogen atom that replaces the deoxyribose. Geometry was optimized at the B3LYP/6-31G(d,p) level using the Gaussian03 program, since the zwitterionic species is not stable in the gas-phase, several bonds (see Supplementary Figs. 18b) were constrained to their values in the QM/MM minimized structure during the geometry optimization. To analyze the electronic structure, NBO analysis (including the Natural Resonance Theory Analysis) was carried out at the minimized structures with NBO 5.0.
1 Yu, B. *et al.* Crystal structures of catalytic complexes of the oxidative DNA/RNA repair enzyme AlkB. *Nature* **439**, 879-884 (2006).

2 Sundheim, O. *et al.* Human ABH3 structure and key residues for oxidative demethylation to reverse DNA/RNA damage. *EMBO J* **25**, 3389-3397 (2006).

3 Koivisto, P., Duncan, T., Lindahl, T. & Sedgwick, B. Minimal methylated substrate and extended substrate range of Escherichia coli AlkB protein, a 1-methyladenine-DNA dioxygenase. *J Biol Chem* **278**, 44348-44354 (2003).

4 Brunger, A. T. & Karplus, M. Polar hydrogen positions in proteins: empirical energy placement and neutron diffraction comparison. *Proteins* **4**, 148-156 (1988).

5 Brooks, B. R. *et al.* Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J Comput Chem* **4**, 187-217 (1983).

6 Brooks, C. L. & Karplus, M. Deformable Stochastic Boundaries in Molecular-Dynamics. *J Chem Phys* **79**, 6312-6325 (1983).

7 Shao, Y. *et al.* Advances in methods and algorithms in a modern quantum chemistry program package. *Phys Chem Chem Phys* **8**, 3172-3191 (2006).

8 Konig, P. H., Hoffmann, M., Frauenheim, T. & Cui, Q. A critical evaluation of different QM/MM frontier treatments with SCC-DFTB as the QM method. *Journal of Physical Chemistry B* **109**, 9082-9095 (2005).

9 Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic-Behavior. *Phys Rev A* **38**, 3098-3100 (1988).

10 Becke, A. D. Density-Functional Thermochemistry .3. The Role of Exact Exchange. *J Chem Phys* **98**, 5648-5652 (1993).

11 Lee, C. T., Yang, W. T. & Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron-Density. *Phys Rev B* **37**, 785-789 (1988).

12 Petersson, G. A. *et al.* A Complete Basis Set Model Chemistry .1. The Total Energies of Closed-Shell Atoms and Hydrides of the 1st-Row Elements. *J Chem Phys* **89**, 2193-2218 (1988).

13 MacKerell, A. D. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *Journal of Physical Chemistry B* **102**, 3586-3616 (1998).

14 Glendening, E. D. *et al.* NBO 5.0 Program. *NBO 5.0 Program* (2001).

15 Yi, C., Yang, C. G. & He, C. A non-heme iron-mediated chemical demethylation in DNA and RNA. *Acc Chem Res* **42**, 530-541 (2009).

16 Frisch, M. J. *et al.* Gaussian 03. (2003).

17 Glendening, E. D. & Weinhold, F. Natural resonance theory: I. General formalism. *J Comput Chem* **19**, 593-609 (1998).

18 Glendening, E. D. & Weinhold, F. Natural Resonance Theory: II. Natural bond order and valency. *J Comput Chem* **19**, 610-627 (1998).
Glendening, E. D., Badenhoop, J. K. & Weinhold, F. Natural resonance theory: III. Chemical applications. *J Comput Chem* **19**, 628-646 (1998).