Duplex interrogation by a direct DNA repair protein in search of base damage

Chengqi Yi1–3, Baoen Chen1, Bo Qi3,5, Wen Zhang3, Guifang Jia3, Liang Zhang3, Charles J Li3, Aaron R Dinner3,5, Cai-Guang Yang4 & Chuan He3

ALKBH2 is a direct DNA repair dioxygenase guarding the mammalian genome against \(N^1\)-methyladenine, \(N^3\)-methylcytosine and \(1,N^6\)-ethenoadenine damage. A prerequisite for repair is to identify these lesions in the genome. Here we present crystal structures of human ALKBH2 bound to different duplex DNAs. Together with computational and biochemical analyses, our results suggest that DNA interrogation by ALKBH2 has two previously unknown features: (i) ALKBH2 probes base-pair stability and detects base pairs with reduced stability, and (ii) ALKBH2 does not have nor need a damage-checking site, which is critical for preventing spurious base cleavage for several glycosylases. The demethylation mechanism of ALKBH2 insures that only cognate lesions are oxidized and reversed to normal bases, and that a flipped, non-substrate base remains intact in the active site. Overall, the combination of duplex interrogation and oxidation chemistry allows ALKBH2 to detect and process diverse lesions efficiently and correctly.

Genomic DNA can be damaged by alkylating agents from both endogenous and exogenous sources, leading to cytotoxicity and carcinogenic mutations. Three mechanisms exist to battle cellular methylation and alkyl damage: one involving DNA glycosylases, one involving the suicidal \(O^6\)-methylguanine-DNA methyltransferases, and direct oxidative repair mediated by the ALKBH2 type iron(II) and 2-ketoglutarate (2KG)-dependent dioxygenases. ALKBH2 is one of the first human homologs identified within the AlkB family of DNA repair enzymes. ALKBH2 discriminates guanine from oxoG (which is bound in its active site) in an alternative damage-checking site, thus representing a ‘gatekeeping’ strategy to prevent incorrect cleavage of normal bases. The human oxoG repair protein OGG1 discriminates guanine from oxoG (which is bound in the active site) in an alternative damage-checking site, thus representing a ‘gatekeeping’ strategy to prevent incorrect cleavage of normal bases. The human oxoG repair protein OGG1 discriminates guanine from oxoG (which is bound in the active site) in an alternative damage-checking site, thus representing a ‘gatekeeping’ strategy to prevent incorrect cleavage of normal bases.
Figure 1  Base pairs with different stability are discernible by ALKBH2. (a) Cartoon of the CG structure. (b) Local view showing the interrogation of the target C8-G8* pair by ALKBH2, with residues Val101 and Phe102 highlighted. (c) Overall view of the AT structure. ALKBH2 is shown in green, DNA in beige, DNA bases from the upper strand in magenta and those from the bottom strand in cyan.

central C-G or A-T pair (referred to hereafter as the CG and AT structures, respectively) (Fig. 1 and Table 1). One prominent feature of the CG structure is that the central C-G pair is severely distorted by ALKBH2 while remaining fully intrahelical (Fig. 1a). The side chain of Val101 wedges into C8 and A9, disrupting the stacking of C8 to A9 and markedly buckling the C8 base; the finger residue Phe102 intercalates between G8* and T9*, leading to a continuous stack, four base pairs wide, in the complementary strand from A7* to T9* (Fig. 1b). As a result, although the flanking pairs of T7-A7* and A9-T9* remain hydrogen-bonded with the duplex structure outside of the area that is not much affected, the central target pair C8-G8* is weakened, and the stacking of C8 with neighboring bases is essentially abolished. Activity assays confirmed the important roles of the Val101 and Phe102 residues (which are in the β3-β4 hairpin): both ALKBH2 V101A and F102A mutations caused reduced repair activity toward m1A and m3C, and introducing a V101G F102A double mutation abrogated ALKBH2 activity toward m1A and m3C, and introducing a V101G F102A double mutation abrogated ALKBH2 activity.

Table 1  Data collection and refinement statistics of CG, AT and CI structures

|                      | CG structure | AT structure | CI structure | (Mn and 2KG) CG structure (Mn and 2KG) CG-DNA2 |
|----------------------|--------------|--------------|--------------|-----------------------------------------------|
| Data collection      |              |              |              |                                               |
| Space group          | P12,1        | P4,2,2       | P2,1,2,1     | P12,1                                        |
| Cell dimensions      |              |              |              |                                               |
| a, b, c (Å)          | 41.2, 61.0, 65.2 | 74.9, 74.9, 169.6 | 55.9, 65.0, 167.9 | 46.0, 60.5, 65.5 |
| a, β, γ (°)          | 90, 89.2, 90 | 90, 90, 90 | 90, 90, 90 | 90, 90, 90 |
| Resolution (Å)       | 50–1.58 (1.65–1.58) | 50–3.06 (3.17–3.06) | 50–2.60 (2.69–2.60) | 50–1.49 (1.54–1.49) |
| Rmerge               | 0.03 (0.408) | 0.07 (0.722) | 0.07 (0.956) | 0.06 (0.308) |
| l / σt               | 20.8 (2.2) | 17.1 (2.2) | 19.8 (2.1) | 35.7 (6.0) |
| Completeness (%)     | 94.3 (93.4) | 99.6 (98.6) | 94.9 (96.5) | 96.2 (89.8) |
| Redundancy           | 2.0 (2.0) | 5.5 (6.0) | 4.9 (5.2) | 7.7 (7.7) |
| Refinement           |              |              |              |                                               |
| Resolution (Å)       | 20–1.62 (1.66–1.62) | 50–3.06 (3.14–3.06) | 20–2.60 (2.67–2.60) | 20–1.60 (1.64–1.60) |
| No. reflections      | 43,430       | 9,236        | 17,484       | 43,675                                        |
| Rmerge / Rfree       | 20.1 / 22.3 | 25.3 / 28.9 | 22.1 / 28.9 | 17.5 / 20.2 |
| No. atoms            |              |              |              |                                               |
| Protein              | 1,695        | 1,641        | 3,210        | 1,727                                         |
| DNA                  | 566          | 527          | 1,052        | 566                                           |
| Ligands/water        | 318          | 4            | 26           | 364                                           |
| β-factors            | 18.6         | 45.6         | 33.5         | 22.8                                          |
| Protein              |              |              |              |                                               |
| DNA                  | 23.8         | 84.5         | 33.8         | 27.8                                          |
| Ligands/water        | 28.0         | 63.7         | 38.9         | 33.4                                          |
| R.m.s. deviations    |              |              |              |                                               |
| Bond lengths (Å)     | 0.006        | 0.007        | 0.014        | 0.008                                         |
| Bond angles (°)      | 1.191        | 1.255        | 1.718        | 1.464                                         |

Values in parentheses are for highest-resolution shell. Each structure was solved using one crystal.
structure, instead of remaining intrahelical as observed in the CG structure, C7 is flipped by ALKBH2 (Fig. 2a).

**Role of ALKBH2 β3-β4 hairpin in stability interrogation**

To evaluate the relative free energies for the protein to flip a cytosine base out of C-G and C·I pair, we carried out ‘alchemical’ molecular dynamics simulations as described in the Supplementary Note 39. Two hypothetical states—one with an intrahelical C·I pair being interrogated and the other with a broken C-G pair—were constructed by computationally mutating the CG and CI crystal structures (Supplementary Fig. 3). The simulation indicates that it is ~2.4 kcal mol⁻¹ less favorable to flip a cytosine from a C-G than from a C·I pair (Fig. 2b). This calculated value can be further evaluated to understand its physical basis (Supplementary Table 1). The solvation energy of the two-aminogroup of guanine (~ −3.6 kcal mol⁻¹) and the penalty of decreased hydrogen bonding capacity (~−3.3 kcal mol⁻¹) almost balance. As a result, the net contribution comes mainly from the protein (~−2.1 kcal mol⁻¹). More specifically, both Gln100 and Phe102 show a preference for the intrahelical C·I pair rather than an intrahelical C·G state, through either electrostatic or van der Waals interactions (Supplementary Fig. 3e,f). Thus, our computational analysis suggests that ALKBH2 applies a moderate mean force to DNA, which can be used to distinguish the stability of different base pairs.

We next did fluorescence studies using 2-aminopurine (2AP) to further investigate how ALKBH2 senses base pair stability in duplex DNA. When 2AP-T was incorporated into the flanking base pair adjacent to an A·C pair also showed increased fluorescence upon protein binding. Thus, a non-cognate mismatched base pair could also be readily detected by ALKBH2 to trigger a substantial conformational change. With the F102A mutant protein, reduced fluorescence increases were observed for DNAs containing m¹A-T and A·C, and no fluorescence enhancement took place with the V101G F102A double mutant ALKBH2, as expected (Supplementary Fig. 1e).

**ALKBH2 detects weakened base pairs to locate its substrates**

ALKBH2 is a promiscuous repair protein whose substrates include both purines (m¹A and 1,N⁶-ethenoadenine (or εA)) and pyrimidines (m³C and m³T). Although structurally diverse, a common feature shared by these N-alkylated lesions is the disrupted Watson-Crick interface, which leads to considerably weakened base pairs compared to their undamaged Watson-Crick counterparts31,32. A superposition of the CG structure with the m¹A structure (PDB 3BTY) (or any other ALKBH2–dsDNA structure with a flipped base) reveals how

---

**Figure 2** Contributions of ALKBH2 to base flipping in duplex DNA. (a) Cartoon view of the CI structure. The same color coding as in Figure 1 is used. (b) Computational analysis of free energy difference for ALKBH2 to break a C-G or C·I pair. States observed in crystal structures are in the orange boxes, and contributions to the free energy difference Δ Gł are plotted.

**Figure 3** ALKBH2 probes the stability of a base pair to detect DNA damage. (a) Side view of the CG structure. The approximate location of a flipped base is indicated with a magenta box and that of the orphaned base (which could have multiple conformations) is indicated with a dashed cyan box. (b) Overlay of the m¹A structure (PDB 3BTY) and the CG structure. A clear shift of the hairpin loop is highlighted. The protein portion of 3BTY is shown in cyan, and the DNA part is omitted for clarity.

**Table 2** Data collection and refinement statistics of εA and m³C structures

| Data collection | εA structure | m³C structure | (Mn and 2KG) |
|----------------|--------------|---------------|-------------|
| Space group    | P6₃22       | P6₃22         | P6₃22      |
| Cell dimensions|              |               |             |
| a, b, c (Å)    | 78.6, 78.6, 228.9 | 78.1, 78.1, 230.0 | 77.8, 77.8, 229.0 |
| α, β, γ (°)    | 90, 90, 120 | 90, 90, 120 | 90, 90, 120 |
| Resolution (Å) | 50–2.78     | 50–2.25       | 50–2.50    |
| Rmerge         | 0.07 (0.829) | 0.07 (0.821) | 0.06 (0.872) |
| Completeness (%) | 99.9 (100) | 99.7 (98.4) | 99.3 (98.2) |
| Redundancy     | 13.2 (13.0) | 17.8 (16.9) | 15.5 (15.3) |
| Refinement     |              |               |             |
| Resolution (Å) | 20–2.78     | 20–2.25       | 50–2.50    |
| No. reflections | 10,604      | 19,430        | 14,048     |
| Rwork / Rfree | 22.8 / 25.9 | 22.4 / 25.6 | 22.7 / 25.9 |
| No. atoms      |              |               |             |
| Protein        | 1,626       | 1,636         | 1,636      |
| DNA            | 527         | 526           | 526        |
| Ligands/water  | 27          | 75            | 59         |
| B-factors      |              |               |             |
| Protein        | 48.9        | 28.0          | 31.2       |
| DNA            | 45.4        | 28.7          | 30.8       |
| Ligands/water  | 52.8        | 27.7          | 34.4       |
| R.m.s. deviations |          |               |             |
| Bond lengths (Å) | 0.007       | 0.011         | 0.006      |
| Bond angles (°) | 1.359       | 1.541         | 1.112      |

Values in parentheses are for highest-resolution shell. Each structure was solved using one crystal.
the interrogation mechanism of ALKBH2 could facilitate detection
of these lesions when they are hidden in different contexts. The β3-β4
hairpin motif inserts itself much deeper into the duplex in the m1A
structure (Fig. 3). This motif also shifts up substantially toward the
lesion-containing strand in the m1A structure, compared to the CG
structure (Fig. 3b), which allows the finger residue Phe102 to fully
occupy the vacant space created by base flipping in the lesion-specific
complex. On the other hand, before the location of a damaged base,
the shallow insertion of the hairpin tip, observed in the CG inter-
rogation complex, would exert force on the target base pair. In the
case of a weakened base pair that often harbors DNA damage, inter-
rogation by the hydrophobic hairpin tip enables ALKBH2 to readily
detect the lesioned base and initiate base flipping. The exact identity
of the lesioned base does not affect the probing mechanism used by
ALKBH2. By contrast, the class of glycosylases that have been sug-
gested to use active interrogation (for instance AlkA and MutM) rec-
ognizes unique characteristics of their cognate lesions17,20. Whether
or not such indiscriminate stability interrogation can also be used by
other DNA glycosylases to directly probe an intrahelical base pair
remains to be determined.

Unlike many glycosylases, ALKBH2 does not bend or induce sub-
stantial global distortion to duplex DNA in the process of both base
interrogation and damaged base recognition in the active site. It uses
several structural motifs, especially the finger-bearing β3-β4 hairpin
and the long loop (between β11 and β12), to clamp a dsDNA from
two sides (Fig. 3a). These motifs also represent the major elements
involved in DNA duplex interrogation. A few other base-modification
enzymes do not cause marked disruption to the duplex structure
either33. In fact, these duplex-interacting elements are unique for
ALKBH2 among the other human homolog proteins; swapping the
hairpin sequence of ALKBH2 with the corresponding sequence from
ABH3, a single-stranded nucleic acid demethylase, leads to a switched
preference of ALKBH2 toward single-stranded DNA (ssDNA)28,29.

Structures of ALKBH2 bound to εA- and m3C-containing DNA

Once ALKBH2 locates a weakened base pair in the genome, it
then needs to recognize the potential base lesion. To investigate
how ALKBH2 accommodates various bases, we also crystallized
ALKBH2–dsDNA complexes containing εA with Mn(II) and 2KG
(εA structure), m3C in the absence of cofactors, and m3C with
Mn(II) and 2KG (m3C structure), respectively (Supplementary
Fig. 2c–e and Table 2). ALKBH2 was shown to use at least five active
site residues for m1A recognition11, but base recognition for εA or
m3C is less tight (Supplementary Fig. 2d). ALKBH2, similarly to its
E. coli homolog protein AlkB and despite the differences in hydro-
gen bonding capacities and base dimensions, can well accommo-
date the flipped lesion in its active site to allow efficient catalysis
(Supplementary Fig. 2f, g)11,34–36.

εA was previously found to form base pairs with thymine and
guanine in B-form DNA32,37, whereas m3C is not able to pair with
an opposite guanine base11. Because the stability of εA-T was thus
expected to be much closer to C-G, we focused our quantitative
analysis on the εA-T case. To evaluate the energetics for ALKBH2
to flip an εA lesion out of a weak εA-T pair, we again used alchemi-
cal molecular dynamics simulations (Supplementary Fig. 4). These
simulations were based on the εA and the CG structures as discussed
in the Supplementary Note. The calculations show that flipping an
εA lesion from the εA-T pair is ∼3.5 kcal mol−1 more favorable than
flipping a cytosine from a C-G pair (Supplementary Table 1); there-
fore, the εA-T pair is calculated to be ∼1.1 kcal mol−1 weaker than
the C-T pair discussed above. Overall, the contributions of individual
parts of the system are larger in magnitude than in the CI-to-CG
case, and the difference in stability arises from a balance of many
compensating interactions.

ALKBH2 does not have nor need a damage-checking site

When the complex structures with each of the five bases (m1A, εA,
adine, m3C and cytosine) is flipped by ALKBH2, the location of an
undamaged base and its derived damage can be compared. An overlay
using m1A, εA and AT structures reveals that all these bases reside at
a similar position in the active site of ALKBH2 (Fig. 4a); moreover,
m3C and cytosine are also found to be located at the identical spot
(Fig. 4b, c). Such observations are in sharp contrast to previous ones
with regard to DNA glycosylases: to prevent accidental glycosidic

![Figure 4](image-url)

**Figure 4** ALKBH2 does not have nor need a damage-checking site; its oxidation chemistry insures that non-substrate bases are not modified.
(a) Superposition of the AT structure, εA structure and 3BTY to show that the three flipped bases, damaged or not, are bound to the same site of ALKBH2. The view on the left is shown at the same angle as in Figure 3; a 90° clockwise rotation of the left structure gives the view on the right. (b) Overlay of the CI and m3C structures using the protein part of the complex. A zoomed-in view on the right shows the four active site residues that interact with the m3C base (Tyr122 and Glu175 form hydrogen bonds to N4 of m3C; Phe124 and His171 stack against m3C). (c) Final positions of m3C (magenta) and cytosine (pale cyan) in the active site of ALKBH2 are the same. The identical view as in b is shown, and all protein residues are omitted for clarity. (d) Stereo view (which is a 90° clockwise rotation of c) of the metal site of the m3C structure. The aberrant methyl group (highlighted with yellow background) is precisely positioned by ALKBH2 for efficient oxidation. The approximate location of the putative iron(IV)–oxo
species, after 2KG is converted to succinate, is colored in blue.
bond cleavages that generate potential DNA breaks, DNA glycosylases use different and strict strategies to insure that a normal base cannot be flipped into the catalytic site to initiate glycosylation, and ALKBH2 must accommodate and accurately position the aberrant alkyl group on different base lesions in the active site in order to carry out the oxidative dealkylation. Even if bases from weakened pairs (substrates or not) are flipped to the same position in the active site, only the damaged bases possessing the alkyl adduct can be oxidatively modified, owing to the proximity of the extra alkyl portion to the catalytic metal site (Fig. 4d). Therefore, flipping a non-substrate base into the enzymatic pocket of ALKBH2 does not run the risk of unwanted modification of the base (Supplementary Fig. 5a,b), and a damage-checking site, which has been shown to be critical for DNA glycosylases to function properly, is not necessary for ALKBH2.

Similarly, the recently discovered TET family proteins, which use the same oxidation mechanism to convert 5-methylcytosine (m5C) to 5-hydroxymethyl cytosine in genomic DNA, should oxidatively modify only m5C but not cytosine, even if a cytosine base is accidently flipped, assuming the TET family proteins also flip m5C out of the duplex for oxidation.

DISCUSSION
DNA damage constitutes only a very small percentage of the human genome, yet the lesions must be promptly repaired to avoid cytotoxic and/or mutagenic consequences. To search for these sporadic sites, repair proteins have evolved different strategies to assure their efficiencies in damage location. We report here that the direct repair dioxygenase ALKBH2 interrogates DNA using mainly a hydrophobic hairpin motif, which is partially inserted into the DNA duplex. This interrogation mechanism enables ALKBH2 to probe base pair stability irrespectively of base pair identity, thus restricting its structurally diverse cognate lesions to being only within the weakened base pairs (Supplementary Fig. 5c).

Invasive interrogation of undamaged DNA has previously been shown in the case of MutM. Human OGG1 could also break a G-C base pair when it is bound to undamaged DNA. Although results reported in this study suggest that ALKBH2 can discern base pairs with different stabilities, we do not think that ALKBH2 flips every base it encounters during the damage-searching process. Both sliding (in which protein diffuses along DNA and makes continuous contact) and hopping (in which protein undergoes microscopic dissociation and re-association) have been suggested to be effective mechanisms for DNA repair proteins to locate lesions in the genome, for example, the time scale of hopping and sliding of UDG has been established and studied. Recent kinetic experiments have shown that lesion searching is a multistep process that could involve a non-specific ‘search complex’, an ‘interrogation complex’ capable of base discrimination and, eventually, a ‘substrate-recognition complex’ featuring a fully flipped base. As for ALKBH2, duplex distortion occurs locally in both the CG structure and structures with a flipped substrate (Fig. 1a and Supplementary Fig. 2c).

It is tempting to speculate that such minimal disturbance would reduce the energetic cost for short-range sliding. Further biophysical and biochemical studies can help to provide mechanistic insights and kinetic characteristics of the nature of the ALKBH2-mediated damage-searching process. It is noteworthy that ALKBH3, a homolog of ALKBH2, relies on partner proteins to open the DNA duplex for damage searching, presenting a notable example of DNA damage searching and repair.

An immediate consequence of the non-selective lesion-detection process is that ALKBH2 may flip a non-cognate lesion into its active site, or even a normal base from a mismatched base pair. Our structures show that ALKBH2 does not have nor need a damage-checking site, which is demonstrated to be critical for assuring correct enzymatic activity in the cases of the MutM and UDG DNA glycosylases. Whether or not other DNA glycosylases also possess such a damage-checking site requires further investigation. For ALKBH2, a flipped base is always at the same position in the active site; the oxidative demethylation chemistry of ALKBH2 must then ensure that a non-cognate base does not get modified in its active site.

Previous studies on both model compounds and iron(II) and 2KG dioxygenases (including a density functional theory study of AlkB) have yielded useful knowledge on dioxygen activation in the context of the conserved 2-His–1-Asp( or Glu) motif (metal-chelating ligand of AlkB and ALKBH2). The spatial relationship of the modified base to the putative, catalytic iron(IV)–oxo species has been shown to be essential for efficient catalysis. Thus, when a non-cognate base is flipped into the active site, owing to the absence of a geometrically optimal chemical group, the flipped base should not be oxidized by ALKBH2 (Supplementary Fig. 5a,b).

On the other hand, when ALKBH2 detects weakened base pairs that do not contain cognate lesions (for example, base pairs with other types of lesions or mismatched pairs), it remains to be seen whether ALKBH2 carries signaling functions, as demonstrated by the ATL protein and perhaps the AlkD protein (Supplementary Fig. 5c). Overall, the combination of DNA interrogation, non-discriminative base flipping and oxidation chemistry by the dioxygenase ALKBH2 can improve the efficiency of DNA repair by increasing the odds of finding hot spots with structurally diverse base lesions while not endangering the integrity of the genome by accidental and unwanted DNA modification.

METHODS
Methods and any associated references are available in the online version of the paper.

Accession codes. Protein Data Bank: Coordinates and structure factors for the ALKBH2–dsDNA complex have been deposited under accession codes 3RZG (CG structure), 3RZM (AT structure), 3RZL (CI structure), 3RZK (Mn and 2KG; m3C structure), 3RZJ (Mn and 2KG; m3C structure), 3S5Y (Mn and 2KG; CG structure) and 3S5A (Mn and 2KG; CG-DNA2 structure).

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS
This study was supported by the US National Institutes of Health (GM071440 to C.H.), the Hundred Talent Program of the Chinese Academy of Sciences (to C.-G.Y.), the State Key Development Program of Basic Research of China (2009CB918502 to C.-G.Y.), the special grant for Stem Cell and Regenerative Medicine (XDA01040305 to C.-G.Y.), the National Science Foundation (MCB-0547854 to A.R.D.), Beamlines 23ID-B (General Medicine and Cancer Institutes Collaborative Access Team, GM/CA-CAT), 21ID-D (Life Sciences Collaborative Access Team, LS-CAT) and 24ID-E (The Northeastern Collaborative Access Team, NE-CAT) at the Advanced Photon Source at Argonne National Laboratory, the Shanghai Synchrotron Radiation Facility (BL17U1) and the US Department of Energy.

AUTHOR CONTRIBUTIONS
C.Y., C.-G.Y. and C.H. designed the experiments. Experiments were conducted by C.Y., B.C., W.Z., G.J., L.Z. and C.J.L.; computational analyses were carried out by B.Q. and A.R.D. C.Y. and C.H. wrote the paper, and B.Q., A.R.D. and C.-G.Y. contributed to editing the manuscript.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.
1. Sedgwick, B. Repairing DNA-methylation damage. Nat. Rev. Mol. Cell Biol. 5, 148–157 (2004).

2. Sedgwick, B., Bates, P.A., Paik, J., Jacobs, S.C. & Lindahl, T. Repair of alkylated DNA: Recent advances. DNA Repair (Amst.) 6, 429–442 (2007).

3. Yi, C., Yang, C.G. & He, C. A non-heme iron-mediated demethylation in DNA and RNA. Acc. Chem. Res. 42, 519–529 (2009).

4. Duncan, T. et al. Reversal of DNA alkylation damage by two human deoxyguaninases. Proc. Natl. Acad. Sci. USA 99, 16660–16665 (2002).

5. Aas, P.A. et al. Human and bacterial oxidative deoxyguaninases repair alkylation damage in both RNA and DNA. Nature 421, 859–863 (2003).

6. Ringov, J. et al. Repair deficient mice reveal mABH2 as the primary oxidative demethylase for repairing 1methyladenine and 3methylcytosine in DNA. EMBO J. 25, 2189–2198 (2006).

7. Ringov, J. et al. AlkB homologue 2-mediated repair of etheno-adenine lesions in mammalian DNA. Cancer Res. 68, 4142–4149 (2008).

8. Cetica, V. et al. Pediatric brain tumors: mutations of two deoxyguaninases (hABH2 and hABH3) that directly repair DNA alkylation damage. J. Neurooncol. 94, 195–201 (2009).

9. Wu, S.S. et al. Down-regulation of ALKBH2 increases cisplatin sensitivity in H1299 lung cancer cells. Acta Pharmacol. Sin. 32, 393–398 (2011).

10. Gilljam, K.M. et al. Identification of a novel, widespread, and functionally important PCNA-binding motif. J. Cell Biol. 186, 645–654 (2009).

11. Yang, C.G. et al. Crystal structures of DNA/RNA repair enzymes AlkB and ABH2 bound to dSNA. Nature 452, 961–965 (2008).

12. Huffman, J.L., Sundheim, O. & Tainer, J.A. DNA base damage recognition and removal: new twists and grooves. Mutat. Res. 577, 55–76 (2005).

13. Yang, C.G., Garcia, K. & He, C. Damage detection and base flipping in direct DNA repair. Nature 468, 303–333 (2010).

14. Yu, B. et al. Crystal structures of catalytic complexes of the oxidative DNA/RNA repair enzyme AlkB. Nature 439, 879–884 (2006).

15. Yi, C. et al. Iron-catalysed oxidation intermediates captured in a DNA repair dimer. Nature 468, 303–333 (2010).

16. Yu, B. & Hart, J.F. Enzymological and structural studies of the mechanism of promiscuous substrate recognition by the oxidative DNA repair enzyme AlkB. Proc. Natl. Acad. Sci. USA 106, 14315–14320 (2009).

17. Leonard, G.A. et al. Guanine-1,N2-ethenoadenine base pairs in the crystal structure of dGCGGAATT(epsilon-dGCGG). Biochemistry 33, 4755–4761 (1994).

18. Blainey, P.C., van Oijen, A.M., Banerjee, A., Verdine, G.L. & Xie, X.S. A base-excision repair DNA protein finds intrahelical lesion bases by fast sliding in contact with DNA. Proc. Natl. Acad. Sci. USA 103, 5752–5757 (2006).

19. Blainey, P.C. et al. Nonspecifically bound proteins spin while diffusing along DNA. Nat. Struct. Mol. Biol. 16, 1224–1229 (2009).

20. Lin, Y. et al. Using the bias from flow to elucidate single DNA repair protein sliding and interactions with DNA. Biophys. J. 96, 1191–1197 (2009).

21. Hedglin, M. & O’Brien, P.J. Hopping enables a DNA repair glycosylase to search both strands and bypass a bound protein. ACS Chem. Biol. 5, 427–436 (2010).

22. Scholfield, C.J. & Zhang, Z.H. Structural and mechanistic studies on 2-oxoglutarate-dependent oxygenases and related enzymes. Annu. Rev. Biochem. 79, 731–760 (2010).

23. Lido, F. et al. Unwinding of DNA/RNA damage in a repair-dependent DNA alkylation repair and cancer cell proliferation. Mol. Cell 44, 373–384 (2011).

24. Gao, Z. et al. Rapid freeze-quench 57Fe Mossbauer spectroscopy: monitoring of the repair activity of cell-free bacteriophage T4 DNA polymerase. J. Phys. Chem. B 115, 4887–4898 (2011).

25. Costas, M., Mehn, M.P., Jensen, M.P. & Que, L. DNA oxidation in mononuclear nonheme iron active sites:enzymes, models, and intermediates. Annu. Rev. Biochem. 80, 377–413 (2011).

26. Liao, X., Liao, J., Li, L., Scholfield, C.J. & Zhang, Z.H. Structural and mechanistic studies on 2-oxoglutarate-dependent oxygenases and related enzymes. Annu. Rev. Biochem. 80, 377–413 (2011).

27. Huang, H., Chopra, R., Verdone, G.L. & Harrison, S.C. Structure of a covalently trapped catalytically competent ubiquitously expressed tumor-transcriptase: implications for drug resistance. Science 282, 1669–1675 (1998).

28. Monsen, V.T. et al. Divergent β-hairpins determine double-strand versus single-strand substrate recognition of human AikB homologues 2 and 3. Nucleic Acids Res. 39, 4415–4425 (2010).

29. Chen, B., Liu, H., Sun, X. & Yang, C.G. Mechanistic insight into the recognition of single-stranded and double-stranded DNA substrates by ABH2 and ABH3. Mol. Bioyst. 6, 2143–2149 (2010).

30. Ma, A., Hu, J., Karplus, M. & Dinner, A.R. Implications of alternative substrate binding modes for catalysis by uracil-DNA glycosylase: an apparent discrepancy resolved. Biochemistry 45, 13687–13696 (2006).

31. Lu, L., Li, C., Ji, X., Zheng, G. & He, C. Structure determination of DNA repair enzymes AlkB and ABH3 bound to ends of duplex DNA: a system for the structure determination of lesion-containing DNA. Structure 16, 1166–1174 (2008).

32. Blainey, P.C. et al. Structure determination of lesion-containing DNA. Structure 16, 341–350 (2008).

33. Blainey, P.C. et al. Repair DNA/RNA protein finds intrahelical lesion bases by fast sliding in contact with DNA. Proc. Natl. Acad. Sci. USA 103, 5752–5757 (2006).

34. Blainey, P.C. et al. Unwinding of DNA/RNA damage in a repair-dependent DNA alkylation repair and cancer cell proliferation. Mol. Cell 44, 373–384 (2011).

35. Chang, X. & Blumenthal, R.M. Mammalian DNA methyltransferases: a structural perspective. Structure 16, 341–350 (2008).

36. Blainey, P.C. et al. Nonspecifically bound proteins spin while diffusing along DNA. Nat. Struct. Mol. Biol. 16, 1224–1229 (2009).

37. Lin, Y. et al. Using the bias from flow to elucidate single DNA repair protein sliding and interactions with DNA. Biophys. J. 96, 1191–1197 (2009).

38. Hedglin, M. & O’Brien, P.J. Hopping enables a DNA repair glycosylase to search both strands and bypass a bound protein. ACS Chem. Biol. 5, 427–436 (2010).

39. Scholfield, C.J. & Stowers, J.T. Timing facilitated site transfer of an enzyme on DNA. Nat. Chem. Biol. 8, 205–210 (2012).

40. Sun, Y., Friedman, J.L. & Stivers, J.T. Complete paramagnetic relaxation enhancements detect transient conformations of human uracil DNA glycosylase (UNG). Biochemistry 50, 10724–10731 (2011).

41. Dango, S. et al. DNA unwinding by a complex helicase is coupled to AlkB-mediated DNA alkylation repair and cancer cell proliferation. Nat. Cell Biol. 44, 373–384 (2011).

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

43. Krobs, C. et al. Rapid freeze-quench 57Fe Mossbauer spectroscopy: monitoring changes of an iron-containing active site during a biochemical reaction. Inorg. Chem. 44, 742–752 (2005).

44. Scholfield, C.J. & Zhang, Z.H. Structural and mechanistic studies on 2-oxoglutarate-dependent oxygenases and related enzymes. Annu. Rev. Biochem. 80, 377–413 (2011).

45. Costas, M., Mehn, M.P., Jensen, M.P. & Que, L. DNA oxidation in mononuclear nonheme iron active sites:enzymes, models, and intermediates. Chem. Rev. 109, 939–986 (2004).

46. Liu, H., Liao, J. & Gauld, J.W.A. DFT study of nucleobase dealkylation by the DNA repair enzyme AlkB. J. Phys. Chem. B 113, 4887–4898 (2009).

47. Tubbs, J.L. et al. Flipping of alkylated DNA damage bridges base and nucleotide excision repair. Nature 459, 808–813 (2009).

48. Robinson, E.H., Gowda, A.S., Spratt, T.E., Gold, B. & Eichman, B.F. An unprecedented nucleic acid capture mechanism for excision of DNA damage. Nature 468, 406–411 (2010).
ONLINE METHODS

Expression and purification of ALKBH2-ΔN55. The gene encoding ALKBH2-ΔN55 was cloned between the Ndel and HindIII sites of a pET28a+ vector (Novagen). This plasmid was transformed into BL21 (DE3) E. coli cells (Stratagene) for protein overexpression. The protein was purified using Ni-NTA chromatography (GE Healthcare) with a binding buffer of 50 mM sodium phosphate (pH 8.0), 300 mM NaCl and 50 mM imidazole, and an elution buffer of 50 mM sodium phosphate (pH 8.0), 300 mM NaCl and 400 mM imidazole. The N-terminal His tag was removed by an overnight thrombin digestion at 4 °C. About 4–6 mg of protein was obtained from 1 l of bacterial cells.

Oligonucleotide synthesis. Oligonucleotides containing modified bases were prepared using solid-phase synthesis. Phosphoramidites were all purchased from Glen Research Corporation and ChemGenes. All synthetic oligonucleotides were purified with denaturing PAGE and verified with MALDI-TOF MS.

Cross-linking and crystallization of the ALKBH2–dsDNA complexes. Cross-linking reactions between ALKBH2 and dsDNA were done as follows: to set up distal cross-linking, a G169C mutation was introduced into ALKBH2–ΔN55 GI69C. This mutant protein was used to cross-link with different synthetic oligonucleotides: 5′-CTGTCACTGCAAGGC-3′ paired with 5′-TCGCGAGTTGAGACA-3′ for the CG structure; 5′-ATGTATAACTGGCG-3′ for the AT structure; 5′-ATGTATACGTCG-3′ paired with 5′-TCGCGTTATACAGA-3′ for the CI structure; 5′-CTGTCT(mC)AC*TGCG-3′ paired with 5′-TGAGGCTACGAGAGA-3′ for the mG structure; 5′-CTGTCA(εA)ACTGGCG-3′ paired with 5′-TGAGGCTACGAGAGA-3′ for the εA structure. The cross-linking reactions were done at 4 °C for 16 h, and the covalently linked ALKBH2–dsDNA complexes were purified using Mono-Q anion exchange chromatography, and were then buffer exchanged to 100 mM NaCl, 10 mM Tris-HCl (pH = 8.0). The complexes were then concentrated to 5 mg ml–1 for crystallization trials. Rod-shaped crystals grew in 1–4 weeks at 4 °C, in hanging drops containing 1 µL of complex solution and 1 µL of a reservoir solution of 100 mM NaCl, 50 mM MgCl2, and 100 mM cacodylate (pH 6.3) and 8% 8K PEG. Cofactor-containing ALKBH2–DNA structures were obtained by soaking apo crystals in well solutions containing MnCl2 (~2.5 mM) and 2KG (~5 mM) for 2–4 h to give the (Mn and 2KG) ALKBH2–DNA structures.

Experimental details of activity assays. A 100-µL reaction mixture containing 50 mM Tris (pH 8.0), 2 mM ascorbic acid, 1 mM K2G, 0.28 mM (NH4)2Fe(SO4)2 and 1 nmol dsDNA or ssDNA substrate was incubated at 4 °C for ~60 min in the presence of 0.5 nmol wild-type ALKBH2 protein and different ALKBH2 mutants (F102A and V101G F102A double mutant). The reaction was terminated by adding EDTA to 5 mM to quench the enzymatic activity. The reaction mixture was run on a DNA-Pac PA200 analytical column (4×250 mm; Dionex) connected to a Dionex HPLC system, with a flow rate of 1 ml min–1 at room temperature. Buffer A contains 50 mM sodium citrate (pH ~5.3), and buffer B contains 50 mM sodium citrate (pH ~5.3), 1 M NaCl. The red lines in Supplementary Figure 1a,b show the exact gradient of the HPLC program. Each experiment was conducted in triplicate.

Experimental details of the 2AP fluorescence studies. For a typical fluorescence study, 1 µM 2AP-containing dsDNA was prepared in a buffer containing 50 mM NaCl, 10 mM Tris-HCl (pH = 8.0) and 1 mM DTT. ALKBH2 stock (~0.12 mM) was also prepared in the same buffer, and the appropriate amount (from 0.25 equiv to up to 1.25 equiv of dsDNA) was added to the same dsDNA solution (~20 µL). Fluorescence traces (fluorescence from ALKBH2 had been subtracted) obtained when increasing amounts of stock ALKBH2 protein were added to the dsDNA-containing solution are shown in Supplementary Figure 1d–f. The fold fluorescence change (the fluorescence when 1 equiv of ALKBH2 was added to DNA divided by that of free DNA) is plotted. Emission spectra (340–420 nm) were recorded at an excitation wavelength λex = 320 nm with excitation and emission bandwidths of 5 nm. All fluorescence measurements (conducted at least in triplicate) were acquired at 4 °C, using a Varian Cary Bio 300 UV–Vis Spectrophotometer.

Structure determination and refinement. All X-ray data were integrated and scaled using HKL200054 and converted to structure factors with the CCP4 program. The ALKBH2–DNA complex structure was phased by molecular replacement (with Phaser)55, using the previously published ALKBH2 structure as a search model. The model was built by using COOT56, and refinement was carried out with REFMAC5 from the CCP4 suite57. Data collection and refinement parameters for all structures are given in Table 1. Molecular graphics figures were prepared with PyMOL (http://www.pymol.org/).

53. Mishina, Y., Chen, L.X. & He, C. Preparation and characterization of the native iron(ii)-containing DNA repair AlkB protein directly from Escherichia coli. J. Am. Chem. Soc. 126, 16930–16936 (2004).
54. Otwinski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. 276, 307–326 (1997).
55. Read, R.J. Pushing the boundaries of molecular replacement with maximum likelihood. Acta Crystallogr. D Biol. Crystallogr. 57, 1373–1382 (2001).
56. Emsley, P. & Cowtan., K. Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. 60, 2126–2132 (2004).
57. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D Biol. Crystallogr. 50, 760–763 (1994).