**SUPPLEMENTARY INFORMATION**

# Duplex Interrogation by a Direct DNA Repair Protein in Search of Base Damage

Chengqi Yi[1, 3], Baoen Chen[2], Bo Qi[3, 4], Wen Zhang[3], Guifang Jia[3], Liang Zhang[3], Charles J. Li[3], Aaron R. Dinner[3, 4], Cai-Guang Yang[2], and Chuan He[3]

[1]*State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences, Peking University, Beijing 100871, China; Peking-Tsinghua Center for Life Sciences, Beijing, China.*

[3]*Department of Chemistry and Institute for Biophysical Dynamics,* [4]*James Franck Institute, The University of Chicago, 929 East 57th Street, Chicago, Illinois 60637, USA.*

[2]*Shanghai Institute of Materia Medica, Chinese Academy of Sciences, 555 Zuchongzhi Road, Shanghai 201203, China.*

Correspondence and requests for materials should be addressed to C.H. (chuanhe@uchicago.edu) or C.-G.Y. (yangcg@mail.shcnc.ac.cn)

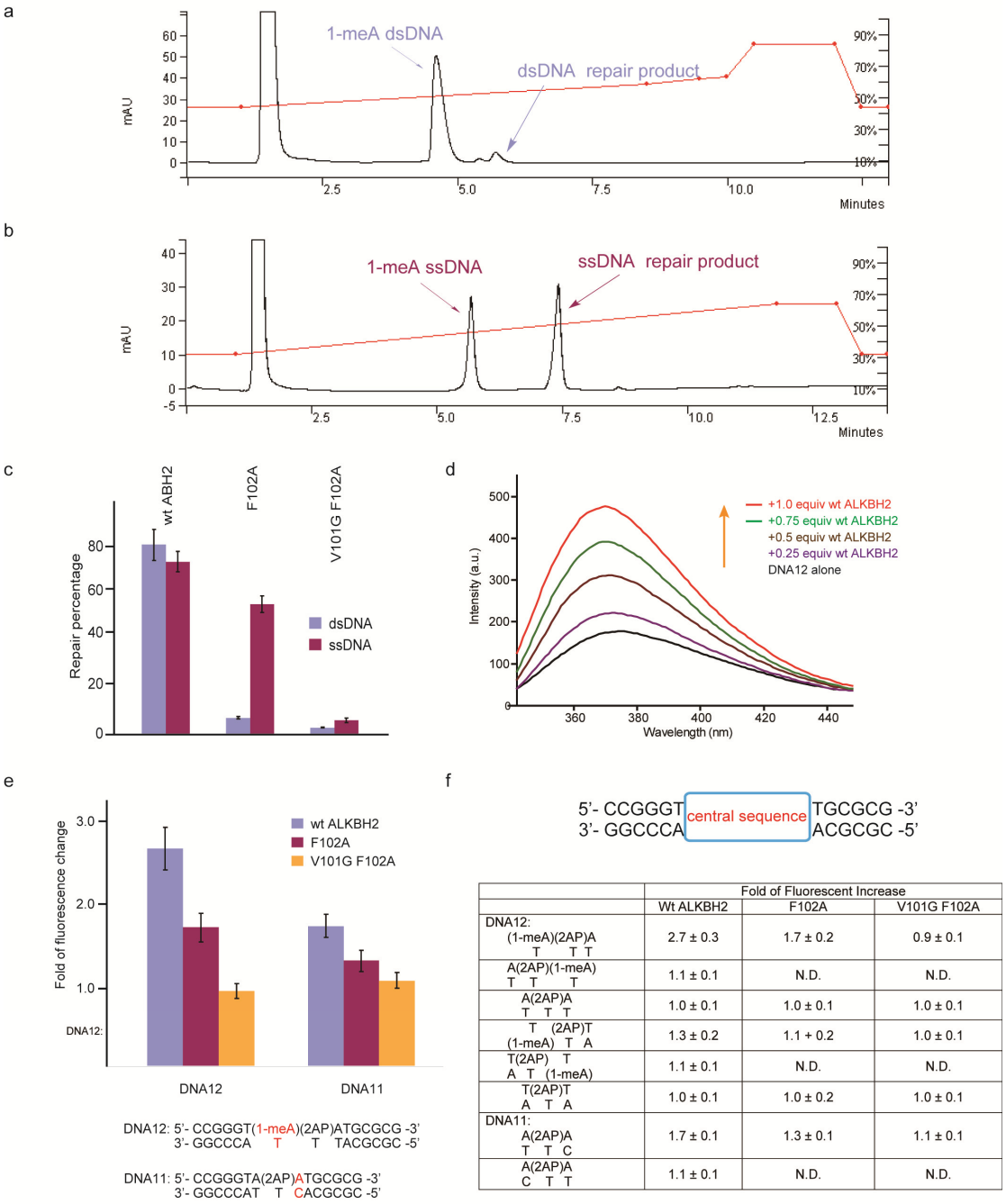
**Supplementary Table 1** Decomposition of the free energy difference into group-by-group contributions: the calculation cycle involving the CG and CI structures (top section), CG and εA structures (middle section) and partial atomic charges of the εA base after parameterization (bottom section) (see Supplementary Note for detailed explanation)

| | | | | | |
|---|---|---|---|---|---|
| CG vs CI | Total free energy difference = 2.4 kcalmol$^{-1}$ | | | | |
| | Group-by-group contributions | | $\Delta G_3$ kcalmol$^{-1}$ | $\Delta G_4$ kcalmol$^{-1}$ | $\Delta\Delta G$ kcalmol$^{-1}$ |
| | Solvent and counter ions (Group *I*) | | 0.5 | -4.1 | -3.6 |
| | ALKBH2 (Group *II*) | Residue Gln100 | 1.0 | 0.0 | 1.0 |
| | | Residue Phe102 | 1.0 | 0.1 | 1.1 |
| | Cytosine of C:G/C:I (Group *III*) | | 3.7 | -0.4 | 3.3 |
| | Two flanking base pairs of C:G/C:I (*IV*) | | 1.5 | -0.7 | 0.8 |
| | DNA Backbone of G/I (*V*) | | -1.1 | 1.4 | 0.3 |
| | Rest of the DNA (*VI*) | | 0.2 | -0.2 | 0.0 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| CG vs εA | Group-by-group contributions | $\Delta G_3$ kcalmol$^{-1}$ | Two HSD | | One HSD & HSP | | Two HSP | |
| | | | $\Delta\Delta G_{Total}$ = -2.3kcalmol$^{-1}$ | | $\Delta\Delta G_{Total}$ = 3.5kcalmol$^{-1}$ | | $\Delta\Delta G_{Total}$ = 5.1kcalmol$^{-1}$ | |
| | | | $\Delta G_4$ kcalmol$^{-1}$ | $\Delta\Delta G$ kcalmol$^{-1}$ | $\Delta G_4$ kcalmol$^{-1}$ | $\Delta\Delta G$ kcalmol$^{-1}$ | $\Delta G_4$ kcalmol$^{-1}$ | $\Delta\Delta G$ kcalmol$^{-1}$ |
| | *I* | 4.1 | -7.8 | -3.7 | 10.8 | 14.9 | 10.5 | 14.6 |
| | *II* | -4.2 | -6.8 | -11.0 | -14.4 | -18.6 | -14.2 | -18.4 |
| | *III* | 13.0 | 0.0 | 13.0 | 0.0 | 13.0 | 0.0 | 13.0 |
| | *IV* | 0.6 | 0.5 | 1.1 | 0.4 | 1.0 | 0.5 | 1.1 |
| | *V* | -10.9 | 13.1 | 2.2 | 11.0 | 0.1 | 10.7 | -0.2 |
| | *VI* | 0.5 | -5.0 | -4.5 | -7.9 | -7.4 | -5.0 | -4.5 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Atom | C5 | N7 | C8 | H8 | N9 | N1 | C2 | H2 |
| | Charge | 0.15 | -0.75 | 0.33 | 0.15 | -0.06 | -0.12 | 0.32 | 0.16 |
| | Atom | N3 | C4 | C6 | N6 | C61 | C62 | H61 | H62 |
| | Charge | -0.68 | 0.43 | 0.62 | -0.74 | -0.32 | 0.22 | 0.17 | 0.12 |

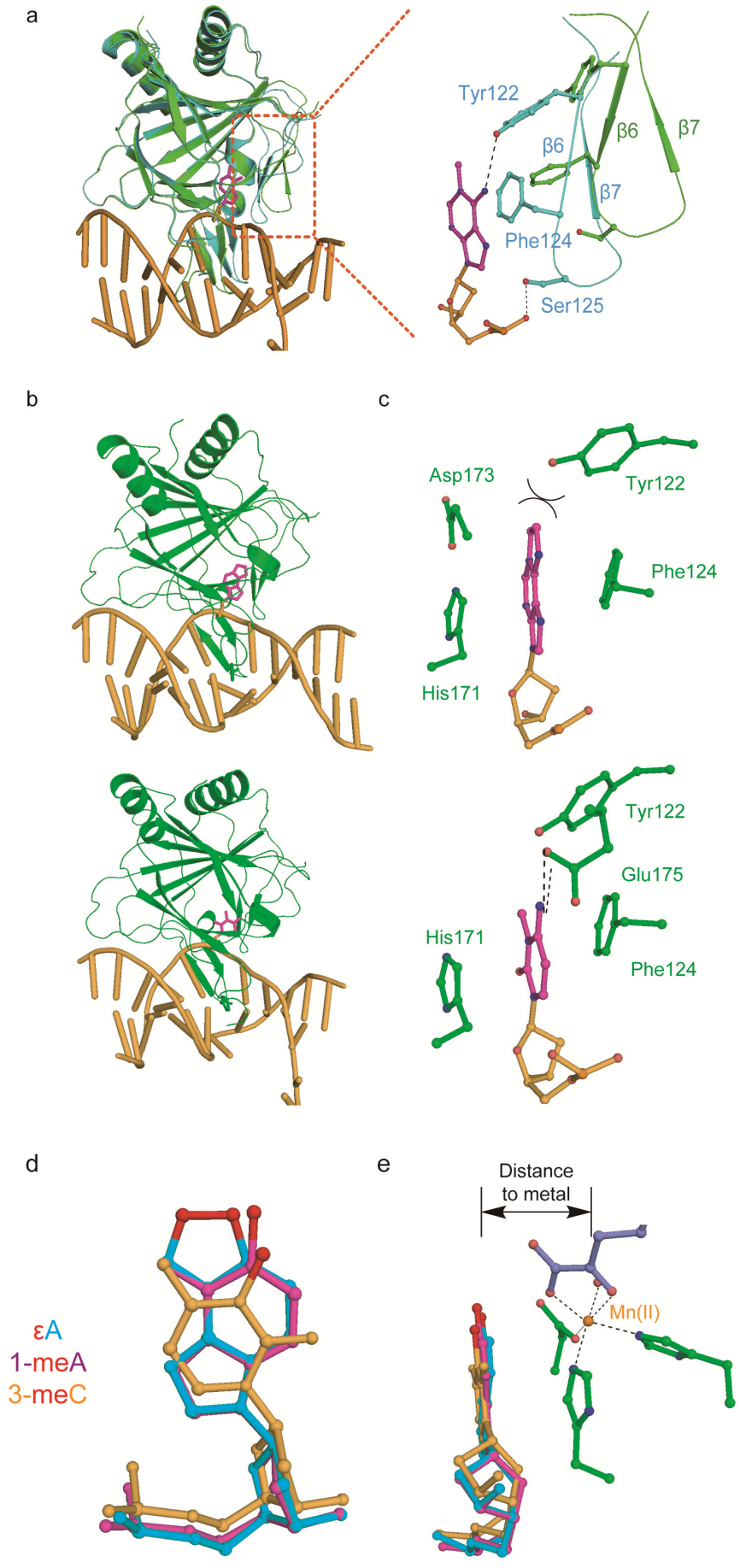
# SUPPLEMENTARY FIGURES



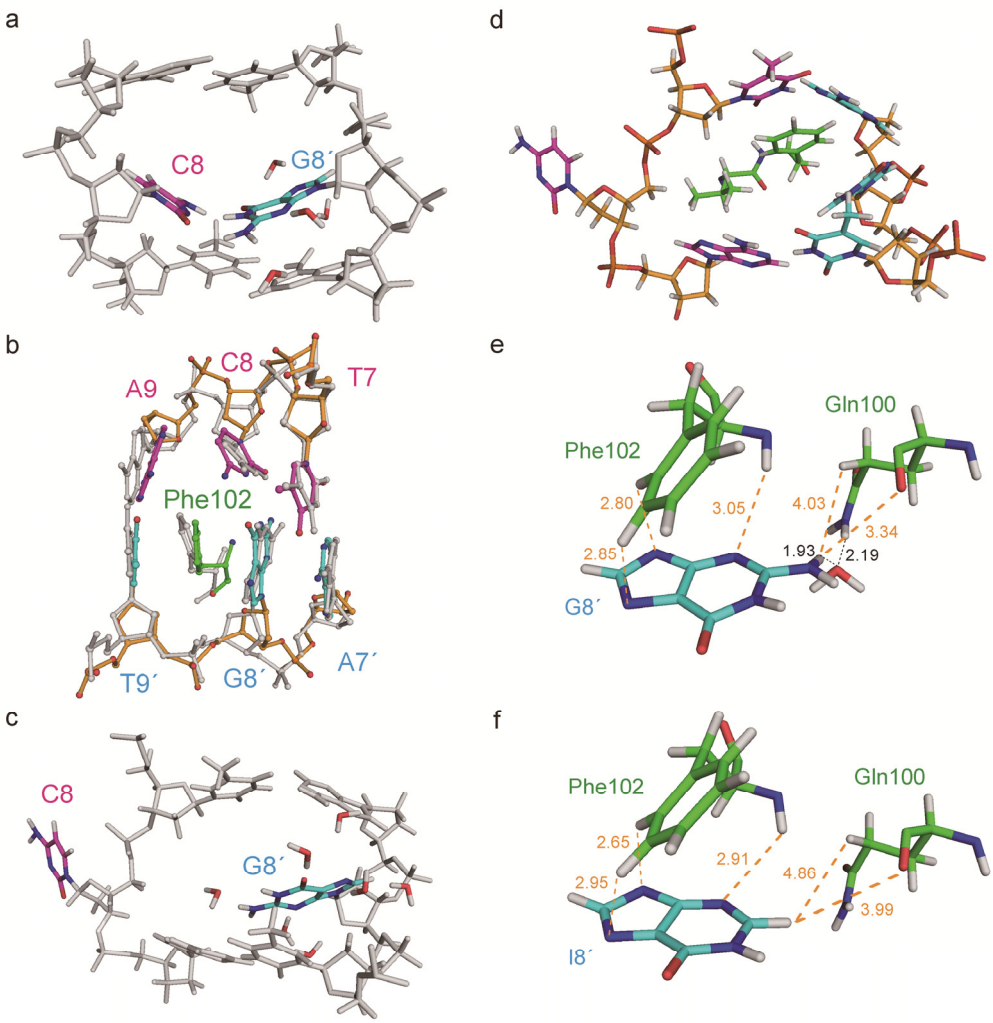

| | Fold of Fluorescent Increase | | |
|---|---|---|---|
| | Wt ALKBH2 | F102A | V101G F102A |
| DNA12:<br>(1-meA)(2AP)A<br>T T T | 2.7 ± 0.3 | 1.7 ± 0.2 | 0.9 ± 0.1 |
| A(2AP)(1-meA)<br>T T T | 1.1 ± 0.1 | N.D. | N.D. |
| A(2AP)A<br>T T T | 1.0 ± 0.1 | 1.0 ± 0.1 | 1.0 ± 0.1 |
| T (2AP)T<br>(1-meA) T A | 1.3 ± 0.2 | 1.1 + 0.2 | 1.0 ± 0.1 |
| T(2AP) T<br>A T (1-meA) | 1.1 ± 0.1 | N.D. | N.D. |
| T(2AP)T<br>A T A | 1.0 ± 0.1 | 1.0 ± 0.2 | 1.0 ± 0.1 |
| DNA11:<br>A(2AP)A<br>T T C | 1.7 ± 0.1 | 1.3 ± 0.1 | 1.1 ± 0.1 |
| A(2AP)A<br>C T T | 1.1 ± 0.1 | N.D. | N.D. |

**Supplementary Figure 1** Activity assays and fluorescence analysis of the roles of Val101 and Phe102 residues in base-flipping and repair activity. (**a**) A typical HPLC spectrum of a reaction mixture with ALKBH2 Phe102Ala mutant and a 1-meA containing dsDNA. The substrate, which bears a positive charge, elutes faster than the repair product. The red line represents the gradient of the HPLC program. (**b**) A typical HPLC spectrum of a reaction mixture with ALKBH2 Phe102Ala mutant and a

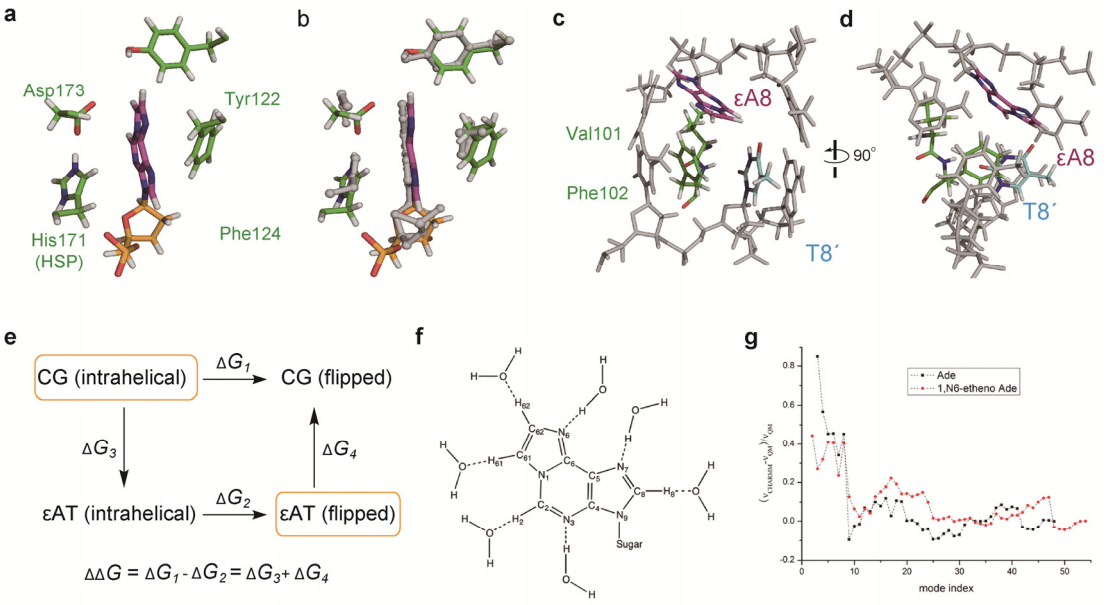
ssDNA substrate. (**c**) Summary of repair activity. Each experiment was performed in triplicate.(**d**) A typical fluorescent trace when increasing amount of concentrated ALKBH2 protein was added to dsDNA of constant concentration. Fluorescence from ALKBH2 has been subtracted. (**e**) When bound to ALKBH2, a mismatched A:C pair gave enhance fluorescence of the adjacent 2AP similar to a 1-meA:T pair. Multiple duplex DNAs, including control sequences, were tested in our fluorescence study. A 1-meA:T pair has four possibilities to flank a 2AP:T pair (to its 5′ or 3′, and 1-meA can be in the 2AP strand or the complementary strand); all of them showed increased fluorescence upon ALKBH2 binding. However, the level of signal increase varies from one sequence to another; and some free duplexes are already very fluorescent when excited, causing the net increase of fluorescence by protein binding to be relatively small. Thus, only results for DNA11 and DNA12 are shown in **b** while some other results (including those of control sequences) are listed in **c**. Fold of fluorescence change (the fluorescence when 1 equivalent of ALKBH2 was added to DNA [fluorescence from ALKBH2 has been subtracted] divided by that of free DNA) is plotted. All experiments were performed in triplicate. (**f**) Fold of fluorescence change when ALKBH2 was added to various DNA containing 2AP, which differ from each other by the central DNA sequence (indicated as a cyan box, and written in the table). For sequences which show only very modest fluorescence increase upon ALKBH2 addition, experiments were not performed for the Phe102Ala and Val101Gly Phe102Ala mutants (N.D.: not determined). All experiments were performed in triplicate.


a
Tyr122
β6
β6
β7
β7
Phe124
Ser125
b
c
Asp173
Tyr122
Phe124
His171
Tyr122
Glu175
His171
Phe124
d
εA
1-meA
3-meC
e
Distance
to metal
Mn(II)

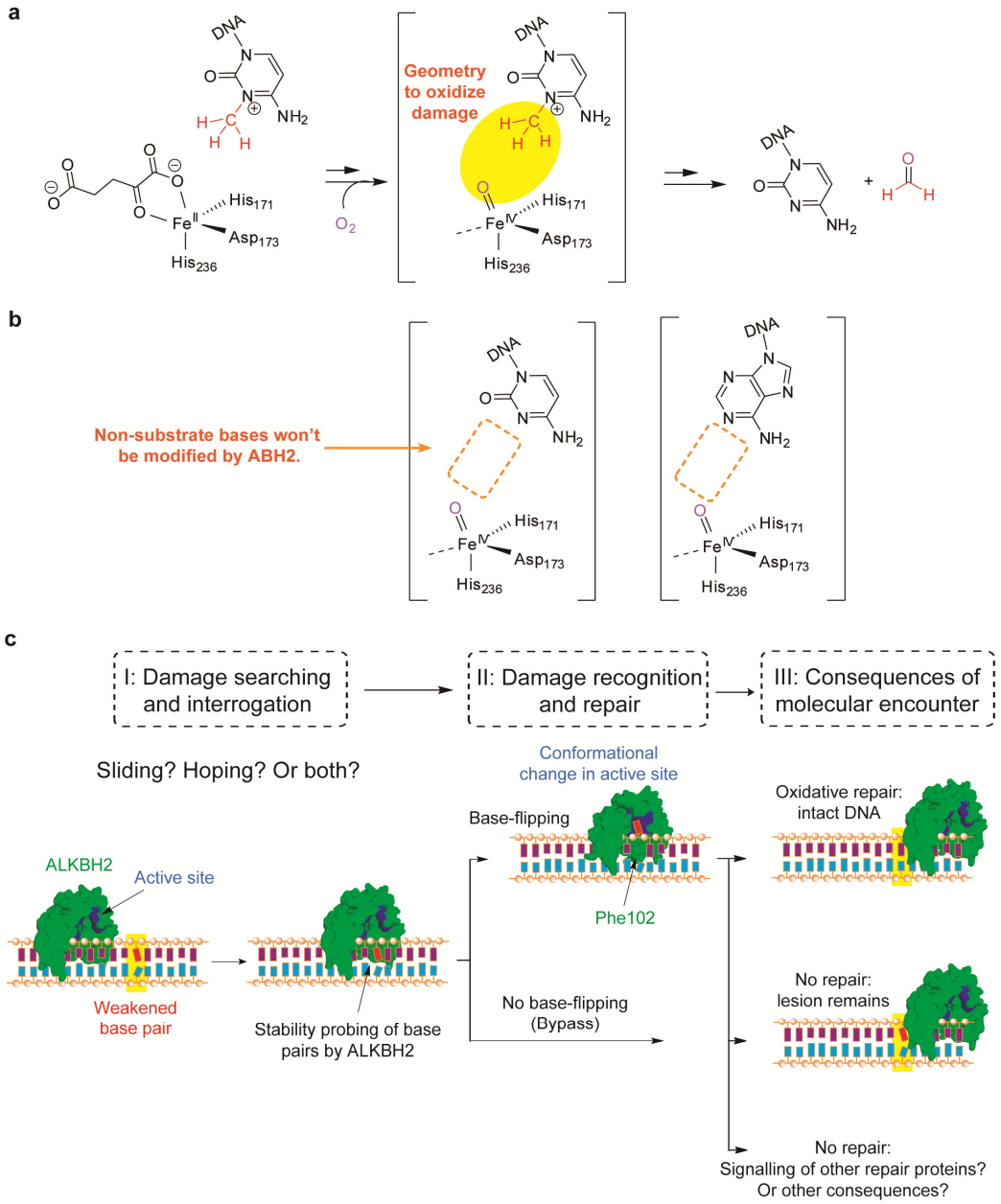
**Supplementary Figure 2** Structures and conformational changes of ALKBH2 when a substrate base is absent or present in its active site. (**a**) Conformational change of the substrate recognition lid in the absence of a bound substrate. The CG structure (green) is superimposed onto ALKBH2 structure with a bound 1-meA (3BTY) (protein in cyan and DNA in bright orange). DNA from the CG structure is omitted for clarity purpose. The hairpin loop motif moves away from the protein body in the absence of a flipped nucleotide. (**b**) Crystal structures of the ALKBH2/dsDNA complexed containing εA and 3-meC, respectively (overall view). (**c**) Active site interactions of εA and 3-meC with ALKBH2. Same color coding as in Fig. 1 is used. Dotted lines indicate hydrogen bonds while curved lines represent van der waals interaction. (**d**) Final positions of 1-meA, εA, and 3-meC in the active site. The protein part of these structures was used for superposition but is omitted for the sake of clarity. The aberrant methyl portions of the damages are colored in red. (**e**) Side view of the same superposition, showing that the aberrant methyl portions are in optimal distance and geometry to the catalytic metal center.




**Supplementary Figure 3** "Alchemical" molecular dynamics simulations of the computational cycle shown in Fig. 2b and related results. (**a**) The computationally equilibrated CG structure, using the CG crystal structure. Water molecules immediately surrounding the guanine base are shown. (**b**) Comparison of the computationally equilibrated CG structure (in grey) and the CG crystal structure (colored as in Fig. 1b) shows that the two structures overlay well with each other. (**c**) The computationally constructed, hypothetical extra-helical CG state. (**d**) Computationally optimized extra-helical CI structure. During the equilibration phase of the calculation for the CI crystal structure, Phe 102 and the I base shift such as to increase their interactions, but these do not change significantly in going from I to G (G base of the hypothetical extra-helical CG state). (**e**) Contributions from ALKBH2 residues Phe102 and Gln100 during stability probing of DNA base pairs. Interactions between guanine base and Phe102 & Gln100 in the optimized CG structure. (**f**) Interactions between the inosine base and Phe102 & Gln100 in the computationally constructed intra-helical CI state. Van der Waals interactions and hydrogen bonds are represented as orange or black dotted lines, with distances (in Å) marked.




**Supplementary Figure 4** "Alchemical" molecular dynamics simulations of the εA crystal structure and related results. (**a**) The computationally equilibrated εA structure, using the εA crystal structure as the starting model. His 171, which is a metal-chelating ligand, is of HSP state. (**b**) Overlay of the equilibrated εA structure to the εA crystal structure (grey). The same orientation is used as that in Supplementary Fig. 2c. The εA base is colored in magenta and the opposite thymine in cyan. ALKBH2 residues Phe102 and Val101 are shown in green. View in **c** is obtained by a ~90°rotation of **d**. (**e**)Thermodynamic cycles used for the free energy calculations of CG/εAT case. (**f**) Parameterization of εA, with its interactions with water considered. (**g**) Comparison of normal mode difference spectra (empirical frequencies minus quantum mechanical ones) for εA and adenine (Ade).




**Supplementary Figure 5** Oxidation chemistry of ALKBH2 and a proposed working model for ALKBH2-mediated damage-searching process. (**a**) The putative iron(IV)-oxo species generated by ALKBH2 is in good geometry to oxidize the aberrant methyl group. (**b**) The reactive iron(IV)-oxo species, even if generated in the presence of a non-substrate base, won't cause unwanted oxidation modification to the flipped base. (**c**) A proposed working model for the damage-searching and repair process of ALKBH2. Such a bi-molecular encounter event is divided into three steps: (I) ALKBH2 senses the weakened base pairs during the damage detection process. Sliding, hopping or a combination of both mechanisms could be used; (II) upon base-binding to the active site, an induced conformational change is observed (blue color in the active site). A cognate lesion should be oxidatively repaired, while non-cognate base that is flipped by ALKBH2 should not be modified; (III) three

different consequences of the molecular encounter event is proposed: firstly, a cognate lesion is repaired and the original DNA base is restored; secondly, a non-cognate base, which could be flipped by ALKBH2, is expected to be expelled by ALKBH2 and the lesion remains in the genome; thirdly, when a non-cognate lesion is bound by ALKBH2, although oxidative repair would not occur, it remains to be seen whether or not other events, for instance repair signaling, will be triggered. Same color coding as in Supplementary Fig. 1 is used.

## Supplementary Note

**Additional crystal structures containing an intra-helical C:G base pair.** Two additional structures were solved, which are Mn(II) and 2KG bound. Both structures have a central C:G pair that is interrogated by ALKBH2 (very similar to Fig. 1b): (Mn/2KG) CG structure (PDB code: 3S57) having the same sequence as the CG structure but with co-factors bound, and (Mn/2KG) CG-DNA2 structure (PDB code: 3S5A) having a slightly different DNA sequence still with an intra-helical C:G pair. These additional structures indicate that the intra-helical C:G pair observed in the CG structure are not affected by binding of the co-factors and by altering DNA sequences.

### Computational details

**Methods.** Free energy calculations can provide useful insights into the thermodynamic basis for the observed structures; here, we focus on the free energy difference for ALKBH2 flipping C:G and C:I, and C:G and εA:T pairs, denoted $\Delta\Delta G = \Delta G_1 - \Delta G_2$ in Supplementary Fig. 4e. Because $\Delta\Delta G$ is a state function, we can obtain it by following the hypothetical paths for mutating an intra-helical (or "unflipped") C:G pair to a C:I (εA:T) pair ($\Delta G_3$) and a flipped C:I (εA:T) pair to a C:G pair ($\Delta G_4$). The free energy difference can then be calculated by $\Delta\Delta G = \Delta G_1 - \Delta G_2 = \Delta G_3 + \Delta G_4$. This route is computationally preferable because the conformational changes associated with the mutations are small compared with the flipping process, accelerating convergence.

To calculate the free energy changes for mutating between C:G and C:I (εA:T) pairs, we employ thermodynamic integration[1,2]. A hybrid structure is constructed that contains both the G (CG) and I (εAT) bases, and a coupling parameter $\lambda$ ($0 \le \lambda \le 1$) is varied to tune from one to the other. The potential thus has the form

$$U(\lambda) = (1-\lambda)U_A + \lambda U_B + U_C, \qquad (1)$$

where A and B denote the bases before and after the mutation, respectively, and C is the remainder of the system. A series of simulations is run in which $\lambda$ is increased from 0 to 1 in steps (detailed below). The free energy difference can then be calculated as

$$\Delta G = G_B - G_A = \int_0^1 d\lambda \frac{\partial G}{\partial \lambda} = \int_0^1 d\lambda \left\langle U_B - U_A \right\rangle_\lambda , \quad (2)$$

where $\left\langle ... \right\rangle_\lambda$ denotes an ensemble average with the given $\lambda$ value. To obtain the integral in Eq. 2, we divide the range of $\lambda$ into 19 windows with $\Delta\lambda = 0.01$ when $0.01 \le \lambda \le 0.05$ and $0.95 \le \lambda \le 0.99$ and with $\Delta\lambda = 0.1$ when $0.1 \le \lambda \le 0.9$. The smaller interval near the endpoints is chosen because of the singularity of the free energy derivatives at $\lambda = 0$ and $\lambda = 1$ ( $\partial G / \partial \lambda \sim \lambda^{-3/4}$ and $(1-\lambda)^{-3/4}$, respectively)[3]. For each $\lambda$, it is necessary to determine the equilibration and sampling periods that are required to obtain a converged ensemble average. To this end, we monitor the convergence of the reverse cumulative averages of the potential energy derivative[4]. We assume that convergence is achieved when a plateau lasting 20 ps or longer is present in the reverse cumulative average. We used a standard trapezoidal method to numerically integrate between $\lambda = 0.1$ and $\lambda = 0.9$, and we fit the remaining $\left\langle U_B - U_A \right\rangle_\lambda$ values to the known functional forms $\lambda^{-3/4}$ and $(1-\lambda)^{-3/4}$ to analytically integrate the endpoint contributions. Summing up the numerical and analytical parts we obtain the free energy change of the mutation from A to B.

Decomposition of the total free energy difference $\Delta\Delta G$ into group-by-group contributions can further provide valuable insight into the physical basis of the free energy values, even though these contributions are path-dependent (see Refs. [5-6] for further discussion). For a group of atoms denoted as *i* in C in Eq. 1,

$$\frac{\partial G_i}{\partial \lambda} = \sum_{j \in i} \frac{\partial G_j}{\partial \lambda} = \sum_{j \in i} \left\langle U_{jB} - U_{jA} \right\rangle_\lambda , \quad (3)$$

where $\Delta G_i$ is the contribution of group *i* to a calculated free energy difference $\Delta G$,

$U_{jX}$ is the interaction between atom j and block X (X = A or B), and the sum runs over the atoms in *i*[7]. The free energy contribution of group *i* can be calculated as

$$\Delta G_i = \int_0^1 d\lambda \frac{\partial G_i}{\partial \lambda}. \quad (4)$$

In our simulation, we divide the contributions to the total free energy difference between flipping a C:G pair and a C:I (εA:T) pair into three parts: the aqueous solvent, the protein ALKBH2, and the DNA. Specifically, we calculate 1) the interactions between all the water molecules plus counter ions and the G/I (CG/εAT) bases, 2) the interactions between the protein and the G/I (CG/εAT) bases, and 3) the interactions between the two closest base pairs to the CG (CI/ εAT) pair and the G/I (CG/εAT) bases.

**System and molecular dynamics.** The crystallographic structures of the DNA-protein complexes with either a flipped C:I or εA:T pair or an unflipped C:G pair were used as the initial structures for the $\Delta G_3$ and $\Delta G_4$ calculations. These structures were represented by the CHARMM all-hydrogen topology and parameter sets (c22 for the protein [8] and c27 for the nucleic acid [9-10]) plus the εA model described below; the calculations were performed with CHARMM version c36a1[11-12] . Hydrogen atoms were added using the HBUILD command in CHARMM.

The PBEQ module [13] was used to determine the charge states of all ten histidine residues in ALKBH2. Six of them are within 13.5 Å of the DNA; we first determined their charge states by selecting the states with the lowest electrostatic solvation free energies, assuming the remaining histidines to be neutral (HSD). We then fixed those six histidines in their charge states, and determined the charge states of the remaining four histidines by an analogous procedure. For the C:I/G calculation, this procedure was sufficient to obtain reasonable results. For the εA:T/C:G calculation, preliminary results indicated that the protonation states are more subtle owing to the conformations accessible to the flipped εA base. Thus we

performed alchemical molecular dynamics simulations on three different systems that were close in free energy in the electrostatic estimates: one with His136 and His237 in the neutral HSD state, one with HSP 136 and HSD 237, and one with both in the HSP doubly protonated state. These three cases were explicitly considered only for $\Delta G_4$ because the CG base pair in the $\Delta G_3$ step is in the unflipped state and is far from the histidine residues in question. Consistent results were obtained from the simulations with one and two HSP residues as shown in the middle section of Supplementary Table 1; we favor the result from the simulation with one HSD and one HSP for two reasons. The equilibrated structure resembles the experimental one most closely (Supplementary Fig. 4a-b) and the contributions from selected residues, in particular, Asp173, which makes hydrogen bonds to the histidines in question and is close to the changing bases, are smaller in magnitude (data not shown).

Sodium ions were added to neutralize the overall system. Each was initially placed 6 Å from a P atom along the bisector of the angle formed by the O1P, P, and O2P atoms, except two in the εA:T/C:G simulation, which were placed at 12 Å to avoid clashes with the base in the flipped state. Then we solvated the complexes with a pre-equilibrated 91×85×73 Å box of TIP3 waters [14-15] subject to periodic boundary conditions. We applied a switch function to the van der Waals interactions over 8-12 Å, and a force shift function [11-12,16] to the electrostatic interactions to truncate them at 12 Å. The positions of hydrogen atoms were first relaxed by energy minimization with fixed heavy atom positions (1000 steps of the steepest descent algorithm (SD) followed by 1000 steps of the adopted basis Newton-Raphson algorithm (ABNR) [11-12]). With the protein and DNA fixed, the same minimization protocol was used to relax the water molecules and the leap frog algorithm was used to heat them from 0 to 300 K over 32 ps and to simulate their dynamics for an additional 20 ps at 300 K. Then, with all atoms free to move the entire system was relaxed using the same minimization and dynamics protocols. The CPT module [17-20] was used in the dynamics for another 40 ps to keep the system under constant normal pressure. The size of the water box was then optimized to 90×81×72 Å. In all molecular dynamics simulations in this study, we used the SHAKE algorithm [21] to

constrain the lengths of the bonds to hydrogen atoms and set the integration time step to be 1 fs.

After preparing the system, we performed free energy simulations with the BLOCK module of CHARMM[22]. In this module, the G (CG) and I (εAT) bases of interest are represented by two separate sets of atoms. In the case of the mutation from a base G to a base I, the initial hybrid structure was constructed by duplicating a G base at the position of interest while replacing the atom N2 by H2 and deleting the atoms H21 and H22. The same operation was performed for the simulations in which an I base was mutated to a G base except that the atom H2 was replaced by N2 and the coordinates of atoms H21 and H22 were generated by the HBUILD command. In the case of the mutation between C:G and εA:T, the same protocol was used except that whole bases were replaced. Non-bonded interactions (van der Waals and electrostatic interactions) between the duplicated bases and the rest of the system were scaled by the coupling parameter $\lambda$. For each window of the simulations, energy minimization was performed followed by leap frog dynamics[11] to heat the system from 0 to 300 K over 32 ps and equilibrate the system at 300 K for additional 760 ps. Configurations were saved every 10 fs. The reverse cumulative averages of the potential energy derivative were well converged after 50-300 ps of equilibration. The configurations of the last 200 ps were used to calculate the ensemble averages of the potential energy derivative.

**Parameterization for 1,N$^6$-etheno adenine (**εA**).** We optimized the parameters for εA following the protocol in refs. 9 and 10. Hartree-Fock (HF) and Møller-Plesset (MP2) calculations with a 6-31G* basis were performed using Gaussian09 [21]to generate the target geometry and estimate hydrogen bonding interactions with water molecules (Supplementary Fig. 4f). Energy values were scaled as in ref. [9] to account for the lack of explicit polarization in the empirical energy function. The partial atomic charges for the base and the van der Waals parameters for $C_{61}$, $C_{62}$, $H_{61}$, and $H_{62}$ were set by analogy with existing CHARMM atom types and optimized manually to reproduce the target data. The root-mean-square error obtained with the resulting

parameters was 0.1 kcalmol$^{-1}$, which is consistent with the existing force field. Reasonable agreement was obtained between quantum mechanical and empirical normal mode spectra for the isolated base (Supplementary Fig. 4g). The optimized partial atomic charges for atoms in the base are shown in the bottom section of Supplementary Table 1.

## Supplementary References


1 Brooks, C. L., Karplus, M. & Pettitt, B. M. *Proteins : a theoretical perspective of dynamics, structure, and thermodynamics*. Vol. v. 71 (J. Wiley, 1988).

2 Frenkel, D. & Smit, B. *Understanding molecular simulation : from algorithms to applications*. 2nd edn, Vol. v. 1 (Academic Press, 2002).

3 Simonson, T. FREE-ENERGY OF PARTICLE INSERTION - AN EXACT ANALYSIS OF THE ORIGIN SINGULARITY FOR SIMPLE LIQUIDS. *Mol. Phys.* **80**, 441-447 (1993).

4 Yang, W., Bitetti-Putzer, R. & Karplus, M. Free energy simulations: Use of reverse cumulative averaging to determine the equilibrated region and the time required for convergence. *Journal of Chemical Physics* **120**, 2618-2628, doi:10.1063/1.1638996 (2004).

5 Boresch, S. & Karplus, M. THE MEANING OF COMPONENT ANALYSIS - DECOMPOSITION OF THE FREE-ENERGY IN TERMS OF SPECIFIC INTERACTIONS. *Journal of Molecular Biology* **254**, 801-807, doi:10.1006/jmbi.1995.0656 (1995).

6 Boresch, S., Archontis, G. & Karplus, M. FREE-ENERGY SIMULATIONS - THE MEANING OF THE INDIVIDUAL CONTRIBUTIONS FROM A COMPONENT ANALYSIS. *Proteins-Structure Function and Genetics* **20**, 25-33, doi:10.1002/prot.340200105 (1994).

7 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, doi:10.1021/bi061061y (2006).

8 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).

9 Foloppe, N. & MacKerell, A. D. All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *Journal of Computational Chemistry* **21**, 86-104, doi:10.1002/(sici)1096-987x(20000130)21:2<86::aid-jcc2>3.0.co;2-g (2000).

10 MacKerell, A. D. & Banavali, N. K. All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution. *Journal of Computational Chemistry* **21**, 105-120, doi:10.1002/(sici)1096-987x(20000130)21:2<105::aid-jcc3>3.0.co;2-p (2000).

11 Brooks, B. R. *et al.* CHARMM: The Biomolecular Simulation Program. *Journal of Computational Chemistry* **30**, 1545-1614, doi:10.1002/jcc.21287 (2009).

12 Brooks, B. R. *et al.* CHARMM - A PROGRAM FOR MACROMOLECULAR ENERGY, MINIMIZATION, AND DYNAMICS CALCULATIONS. *Journal of Computational*

*Chemistry* **4**, 187-217 (1983).

13 Im, W., Beglov, D. & Roux, B. Continuum Solvation Model: computation of electrostatic forces from numerical solutions to the Poisson-Boltzmann equation. *Computer Physics Communications* **111**, 59-75, doi:10.1016/s0010-4655(98)00016-2 (1998).

14 Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. COMPARISON OF SIMPLE POTENTIAL FUNCTIONS FOR SIMULATING LIQUID WATER. *Journal of Chemical Physics* **79**, 926-935, doi:10.1063/1.445869 (1983).

15 Neria, E., Fischer, S. & Karplus, M. Simulation of activation free energies in molecular systems. *Journal of Chemical Physics* **105**, 1902-1921, doi:10.1063/1.472061 (1996).

16 Steinbach, P. J. & Brooks, B. R. NEW SPHERICAL-CUTOFF METHODS FOR LONG-RANGE FORCES IN MACROMOLECULAR SIMULATION. *Journal of Computational Chemistry* **15**, 667-683, doi:10.1002/jcc.540150702 (1994).

17 Andersen, H. C. MOLECULAR-DYNAMICS SIMULATIONS AT CONSTANT PRESSURE AND-OR TEMPERATURE. *Journal of Chemical Physics* **72**, 2384-2393, doi:10.1063/1.439486 (1980).

18 Nose, S. & Klein, M. L. CONSTANT PRESSURE MOLECULAR-DYNAMICS FOR MOLECULAR-SYSTEMS. *Mol. Phys.* **50**, 1055-1076, doi:10.1080/00268978300102851 (1983).

19 Hoover, W. G. CANONICAL DYNAMICS - EQUILIBRIUM PHASE-SPACE DISTRIBUTIONS. *Physical Review A* **31**, 1695-1697, doi:10.1103/PhysRevA.31.1695 (1985).

20 Feller, S. E., Zhang, Y. H., Pastor, R. W. & Brooks, B. R. CONSTANT-PRESSURE MOLECULAR-DYNAMICS SIMULATION - THE LANGEVIN PISTON METHOD. *Journal of Chemical Physics* **103**, 4613-4621, doi:10.1063/1.470648 (1995).

21 Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. C. NUMERICAL-INTEGRATION OF CARTESIAN EQUATIONS OF MOTION OF A SYSTEM WITH CONSTRAINTS - MOLECULAR-DYNAMICS OF N-ALKANES. *Journal of Computational Physics* **23**, 327-341, doi:10.1016/0021-9991(77)90098-5 (1977).

22 Tidor, B. & Karplus, M. SIMULATION ANALYSIS OF THE STABILITY MUTANT R96H OF T4 LYSOZYME. *Biochemistry* **30**, 3217-3228, doi:10.1021/bi00227a009 (1991).