Supplementary information for

Structural basis of nucleic acid recognition and 6mA demethylation by human ALKBH1

Li-Fei Tian[1,2,†], Yan-Ping Liu[2,†], Lianqi Chen[2,3], Qun Tang[2], Wei Wu[1], Wei Sun[2], Zhongzhou Chen[1,*] and Xiao-Xue Yan[2,*]

[1] State Key Laboratory of Agrobiotechnology, College of Biological Sciences, China Agricultural University, Beijing 100193, China;

[2] National Laboratory of Biomacromolecules, Chinese Academy of Sciences (CAS) Center for Excellence in Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China.

[3] College of Life Sciences, University of Chinese Academy of Sciences, Beijing, 100049, China.

[†] These authors contributed equally: Li-Fei Tian, Yan-Ping Liu

* Correspondence: Zhongzhou Chen (chenzhongzhou@cau.edu.cn) or Xiao-Xue Yan (snow@ibp.ac.cn)

MATERIAL AND METHODS

Protein expression and purification

DNAs encoding wild type and mutants of human ALKBH1 were amplified by PCR and subcloned into pET-28a (Novagen) vector containing protease recognition site of tobacco etch virus (TEV). The final clones were verified by DNA sequencing. All recombinant plasmids were transformed into *E. coli* BL21 (DE3). The cells were grown in LB at 37˚C until A_{600} reached 0.8-1.0, and then the overexpression of protein was induced by addition 0.4 mM IPTG at 16˚C for 12-16 h. The cultures were harvested by centrifuging at 4000 g for 12 minutes and resuspended in buffer A (50 mM Tris pH 8.0, 500 mM NaCl, 1 mM PMSF, 5% glycerol, 5 mM beta-mercaptoethanol). The cells were lysed by sonication and lysates were clarified by centrifuging at 20000 g for 40 minutes. The supernatant was then filtrated through a 0.45 μm filter membrane to remove cell debris and then applied to His affinity column (GE Healthcare). After the sample was loaded, the column was washed with buffer B (buffer A containing 20 mM imidazole) and the target protein was eluted with buffer C (buffer A containing 300 mM imidazole). TEV protease (1:15 weight ratio) was added to the eluent containing protein, and the His-Tag was removed overnight. The digestion production was reloaded onto His affinity beads to remove His Tag and His-tagged TEV protease. Gel-filtration buffer (20 mM Bis-Tris pH 6.5, 150 mM NaCl, 2 mM dithiothreitol, 5% v/v glycerol) was used for further purification by size exclusion chromatography (GE Healthcare). Then 6 ml Resource S column (GE Healthcare) was used for further purification. The purified *h*ALKBH1 proteins were stored in storage buffer (20 mM Bis-Tris pH 6.5, 100 mM NaCl, 2 mM dithiothreitol, 5% v/v glycerol) at concentrations of 20-25 mg/ml. Fractions were analyzed by SDS-PAGE and the target protein was combined and concentrated to 20 mg/ml for crystallization.

In vitro demethylase activity assays by HPLC

The demethylation activity assay was carried out in a 50-μL reaction mixture containing 10 μM 6mA-containing ssDNA or bubble/bugle DNA substrates, 10 μM human ALKBH1 protein, 1 mM L-ascorbic acid, 1 mM $(NH_4)_2Fe(SO_4)_2$, 2 mM α-KG, 50 mM Bis-Tris pH 7.0, 50 mM KCl, 2 mM $MgCl_2$. The sample was incubated at 37˚C for 30 min, then the reaction was quenched by heating at 96 °C for 5 min. We further digested the oligonucleotides into single nucleosides

using nuclease P1 (Sigma N8630) and alkaline phosphatase (Takara 2250A). The treated solution was analyzed in an HPLC system equipped with an Agilent Eclipse XDB-C18 analysis column (150 x 4.6 mm), and was eluted with buffer A (H_2O) and buffer B (100% methanol) with a flow rate of 0.8 ml/min at 20 °C. The detection wavelength was set at 254 nm.

In vitro demethylase activity assays by a methylation-sensitive restriction digest method

*h*ALKBH1 6mA demethylation activity assays of ssDNA were modified from the publication[1]. 1 μM *h*ALKBH1 was incubated with 1 μM ssDNA_6mA1 (5'-AACTTCGTGCAGGCATGGG (6mA) TCTTGTCTACT-3', Invitrogen) in reaction buffer containing 1 mM L-ascorbic acid, 1 mM $(NH_4)_2Fe(SO_4)_2$, 2 mM α-KG, 50 mM Bis-Tris pH 7.0, 50 mM KCl, 2 mM $MgCl_2$ at 37˚C for various time. The reactions were stopped by heating to 95˚C after adding 1.1 equivalent of FAM-ssDNA2 (5'-FAM-AGTAGACAAGATCCCATGCCTGCACGAAGTT-3', Invitrogen) and DpnII buffer. The samples were slowly cooled to room temperature for DNA annealing. DpnII (0.5 U/ml) was added to the sample and kept at 37˚C for 1 hour. All samples were analyzed by Tris-Tricine electrophoresis[31] and DNA was shown with ChemiDoc™ MP Imaging System (Bio-Rad) at 520 nm. Quantitative calculation of gray value in gel image is calculated by using Quantity One software and statistical mapping by Excel software.

Bubble DNA and Bulge DNA 6mA demethylation activity assays. Firstly, ssDNA_6mA1 was annealed with different paired DNA (Supplementary information, Fig. S11, S14a) to form Bubble/Bulge DNA by gradient annealing procedure. The paired primers were mixed in a ratio of 1:1 and put into the PCR machine. The temperature was 95˚C for 5 minutes, then lowered to 80˚C for 2 minutes. After that, the temperature was lowered by 3˚C for each 2-minute until it reached 14˚C and then for 5 minutes. Then the enzymatic activity was determined by the same method of ssDNA, except for the gradient annealing method after adding 1.1 equivalent of FAM-ssDNA2, FAM-ssDNA2 was paired with ssDNA_6mA1 to ensure the elimination of mismatched DNA. The temperature was 95˚C for 5 minutes, then lowered to 85˚C for 2 minutes. After that, the temperature was lowered by 3˚C for each 2-minute until it reached 10˚C and then for 5 minutes. Note that due to the excess of FAM-ssDNA2, all the ssDNA_6mA1 and its demethylated product were paired completely with FAM-ssDNA2 except

for Bubble1.

Crystallization and data collection

To obtain the *h*ALKBH1 complex crystals, 0.5 mM protein was mixed with 2 mM $MnCl_2$, 5 mM α-KG, and incubated overnight on ice. The *h*ALKBH1 and *h*ALKBH1 complexes were subjected to initial crystallization screening at 4˚C and 12˚C by sitting-drop vapor diffusion. A series of truncations were screened for crystallization. The *h*ALKBH1$_{19\text{-}369}$-Mn^{2+} formed diamond-shaped crystals at 12˚C in the reservoir solution of 5% ethanol, 5% MPD, 100 mM NaF, 100 mM Hepes pH7.5. The high-quality crystals (Supplementary information, Fig. S1a) were obtained by hanging-drop vapor diffusion in the solution of 2% sucrose, 3% ethanol, 5% MPD, 80 mM NaF, 100 mM Hepes pH7.0 for 7 days. SeMet labeled *h*ALKBH1$_{19\text{-}369}$-Mn^{2+} crystals were obtained under the same condition. The *h*ALKBH1$_{19\text{-}369}$-Mn^{2+}-α-KG formed flake-shaped crystals (Supplementary information, Fig. S1c) at 12˚C in the reservoir solution of 5% PEG3350, 5% MPD, 100 mM $CaCl_2$, 200 mM NaF, 100 mM Bis-Tris pH 6.5. To collect the data, the crystals were gradually transferred to a cryo-buffer (10% ethylene glycol was added to reservoir buffer) and flash frozen in liquid N_2. The data were collected on the beamline BL19U1 of the Shanghai Synchrotron Radiation Facility (Supplementary information, Fig. S1b, d). All data were integrated and scaled with the HKL2000 suite of programs[2]. Data collection and processing statistics were shown in Supplementary information Table S1.

Structure determination and refinement

The diffracted resolution of native *h*ALKBH1 crystals reached 1.97 Å (Supplementary information, Table S1). Due to the low sequence identity (the highest is 19%) with other structure-solved proteins, the correct phase of *h*ALKBH1 was not obtained by molecular replacement. Therefore, the SeMet labeled crystal data were collected and the structure was solved by single-wavelength anomalous scattering. The diffraction data of SeMet-*h*ALKBH1 crystals were collected at the wavelength 0.9793 Å with an oscillation of 0.5˚ per image. The initial phase is calculated and small number of β-strand bundles can be automatically modeled into electron density map by PHENIX[3]. Then we use COOT[4] to build polypeptide model and use PHENIX[3] to finish the refinement in the iterative cycle. Finally, the best solution is

improved to R_{work}/R_{free} values of 0.35 / 0.43 and the figure of merit was 0.46. Using this model, after several rounds of model building and refinement by PHENIX[3], COOT[4] and REFMAC5[5], the high-resolution *h*ALKBH1-Mn^{2+} structure was refined to an R_{work} of 20.3% and R_{free} of 23.3%. The final structure model agreed well with the electron density (Supplementary information, Fig. S18a,b). *h*ALKBH1-Mn^{2+}-α-KG was determined by molecular replacement using *h*ALKBH1-Mn^{2+} structure as a model. The final structure was refined to an R_{work} of 21.6% and an R_{free} of 25.4%. The space group of *h*ALKBH1-Mn^{2+}-α-KG crystal was P1. Each cell contains eight protein molecules, and the structures of eight molecules are highly similar (Supplementary information, Fig. S18c). All the figures in this article showing molecular structure were made by PYMOL[6].

Isothermal titration calorimetry (ITC) assays

Size exclusion chromatography of *h*ALKBH1 protein was performed with buffer D (1 mM $MnCl_2$, 20 mM Bis-Tris pH 6.5, 150 mM NaCl, 2 mM dithiothreitol, 5% v/v glycerol). The α-KG was dissolved in buffer D and the final concentration was 200 mM, and the pH was adjusted to 6.5 by NaOH. Then the solution was diluted with buffer D until the final concentration was 4 mM. α-KG in the syringe was titrated into 0.22 and 0.38 mM *h*ALKBH1 protein in the cell. ITC_{200} calorimeter (GE Life Science and MicroCal) was used to carry out the experiment at 20˚C, and the ITC data were processed by Origin software (version 7.0).

DNA binding assay

DNA binding characteristics of wild and mutant *h*ALKBH1 were examined by non-radioactive electrophoresis mobility shift assay (EMSA). 21 mers ssDNA and dsDNA were used to detect binding ability. The 0.5 μM substrate DNA and 4 μM protein (wild type and mutant *h*ALKBH1) were incubated in 20 μl reaction buffer containing 30 mM Bis-Tris pH 6.5 and 100 mM NaCl, 5% v/v glycerol at 25˚C for 30 minutes. Protein-DNA complexes were loaded onto 6% non-denatured polyacrylamide gel with TBE buffer solution. DNA and protein-DNA complexes were visualized by SYBR Green (Molecular Probes EMSA kit, Invitrogen). The same experiments were performed with Superdex-200 HR 10/300 column on fast protein liquid chromatography system (GE Healthcare) at a flow rate of 0.5 ml/min. *h*ALKBH1 (including wild

type and mutant) and *h*ALKBH1-DNA complex were loaded at a concentration of 200 μg/ml onto a column balanced with buffer (30 mM Bis-Tris pH 6.5, 200 mM NaCl, 2 mM dithiothreitol, 5% v/v glycerol) and eluted with the same buffer. The elution peaks of *h*ALKBH1 and *h*ALKBH1-DNA complex were monitored by measuring absorbance at 280/254 nm.

DNA binding assay by Octet

The instrument was Octet Red 96. Biotin-labeled ssDNA, dsDNA, Bubble6 DNA and Bulge6 DNA were immobilized on SA sensors. Eight different concentrations of *h*ALKBH1 proteins (0, 0.0625, 0.125, 0.25, 0.5, 1, 2, 4μM) reacted with fixed DNA SA sensors respectively, and then they were subjected to baseline (60s), associations (60s), dissociations (120s). Octet monitors the binding of *h*ALKBH1 protein on SA sensor surface in real time. At last, the dissociation constant K_D is calculated by Octet software. Binding buffers, 30 mM Bis-Tris 6.5,100 mM NaCl, 1 mM $MnCl_2$, 1 mM α-KG,5% glycerol. The DNA sequence is used in demethylation experiment.

Supplementary References

1 Müller, T. A., Tobar, M. A., Perian, M. N. & Hausinger, R. P. Biochemical Characterization of AP Lyase and m(6)A Demethylase Activities of Human AlkB Homologue 1 (ALKBH1). *Biochemistry-us.* **56**, 1899-1910 (2017).

2 Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Macromolecular Crystallography, Pt A* **276**, 307-326 (1997).

3 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213-221 (2010).

4 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486-501 (2010).

5 Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67**, 355-367 (2011).

6 Alexander, N., Woetzel, N. & Meiler, J. bcl::Cluster : A method for clustering biological molecules coupled with visualization in the Pymol Molecular Graphics System. *IEEE International Conference on Computational Advances in Bio and Medical Sciences* **2011**,

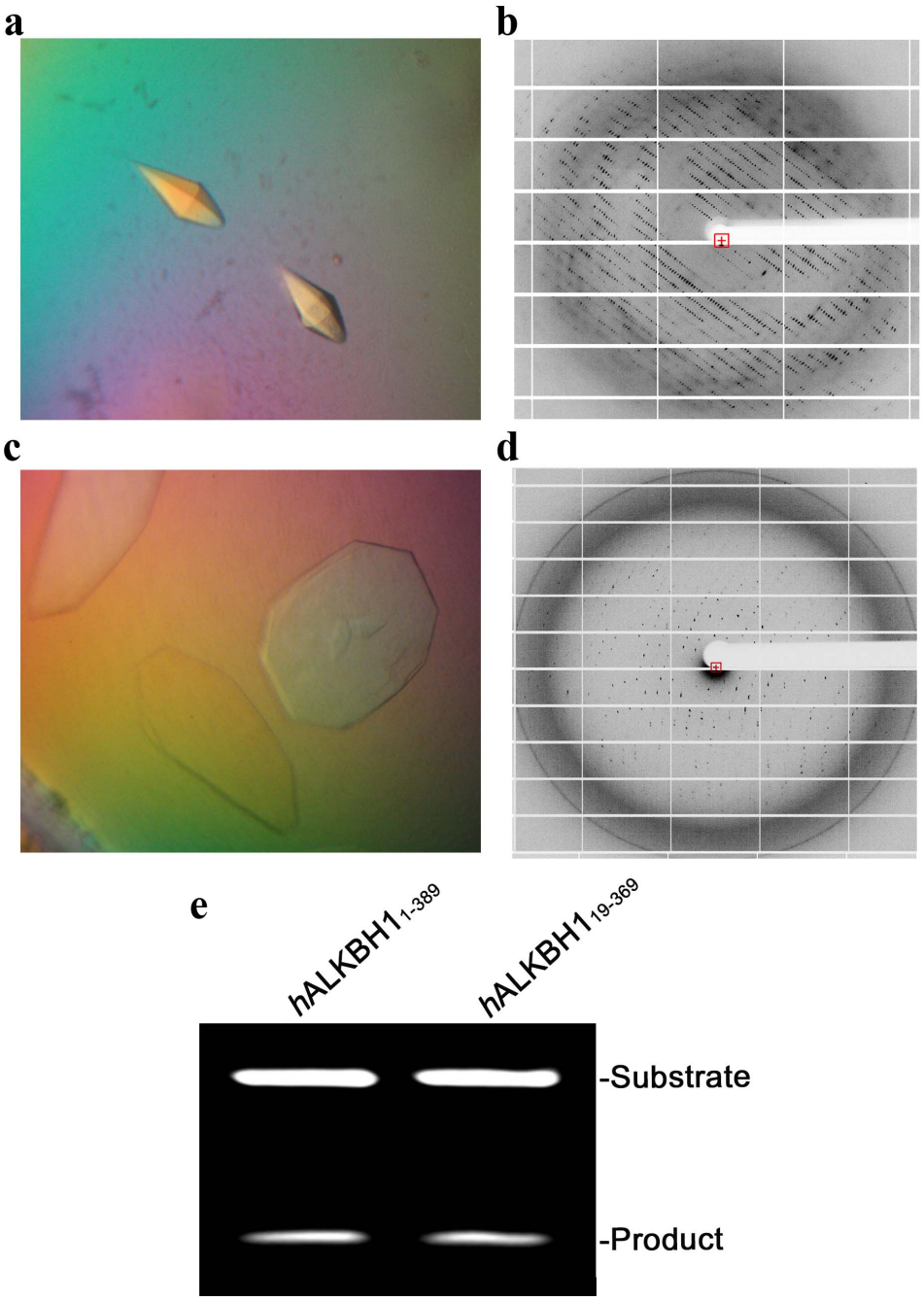
13-18 (2011).

7 Poirot, O., O'Toole, E. & Notredame, C. Tcoffee@igs: a web server for computing, evaluating and combining multiple sequence alignments. *Nucleic Acids Res.* **31**, 3503-3506 (2003).

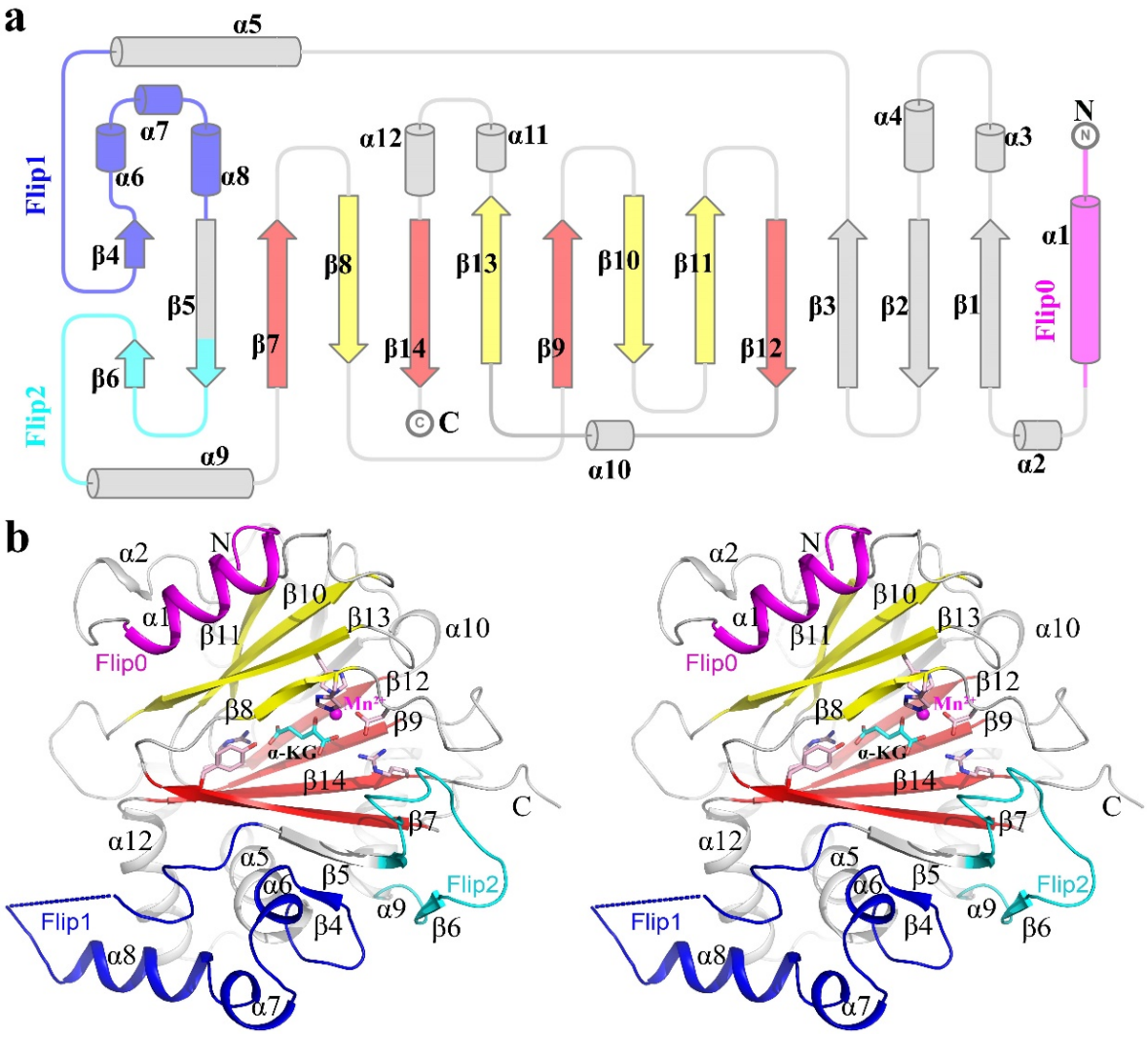
8 Bond, C. S. & Schuttelkopf, A. W. ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments. *Acta Crystallogr. D* **65**, 510-512 (2009).

9 Zhu, C. X. & Yi, C. Q. Switching Demethylation Activities between AlkB Family RNA/DNA Demethylases through Exchange of Active-Site Residues. *Angew. Chem. Int. Edit.* **53**, 3659-3662 (2014).

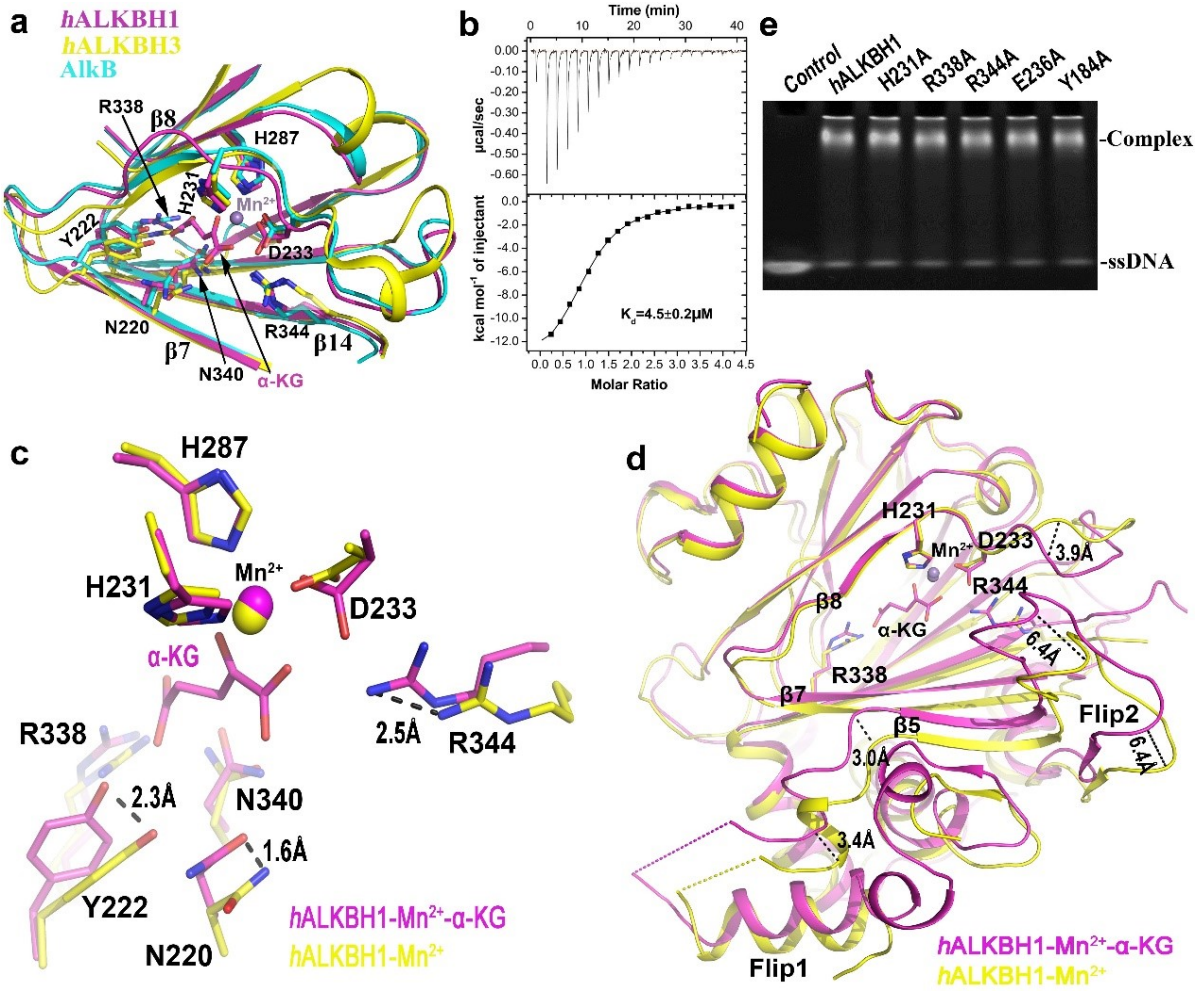
10 Zhou, C. *et al.* Identification and analysis of adenine N6-methylation sites in the rice genome. *Nature Plants* **4**, 554-563 (2018).



Supplementary information, Fig. S1. The crystals of *h*ALKBH1. **a,** Diamond-shaped crystals of *h*ALKBH1$_{19\text{-}369}$-Mn^{2+} after seeding. **b,** Representative X-ray diffraction image of native *h*ALKBH1$_{19\text{-}369}$-Mn^{2+}. **c,** Flake-shaped crystals of *h*ALKBH1$_{19\text{-}369}$-Mn^{2+}-α-KG after seeding. **d,** Representative X-ray diffraction image of native *h*ALKBH1$_{19\text{-}369}$-Mn^{2+}-α-KG. **e,** Detection of *h*ALKBH1 demethylation activity by electrophoresis. The samples were detected at 1 h after demethylation reaction, the demethylated *h*ALKBH1 product was digested by nuclease DpnII. The results of detection showed that the *h*ALKBH1$_{1\text{-}389}$ and *h*ALKBH1$_{19\text{-}369}$ exhibited similar activity toward ssDNA containing 6mA *in vitro*.

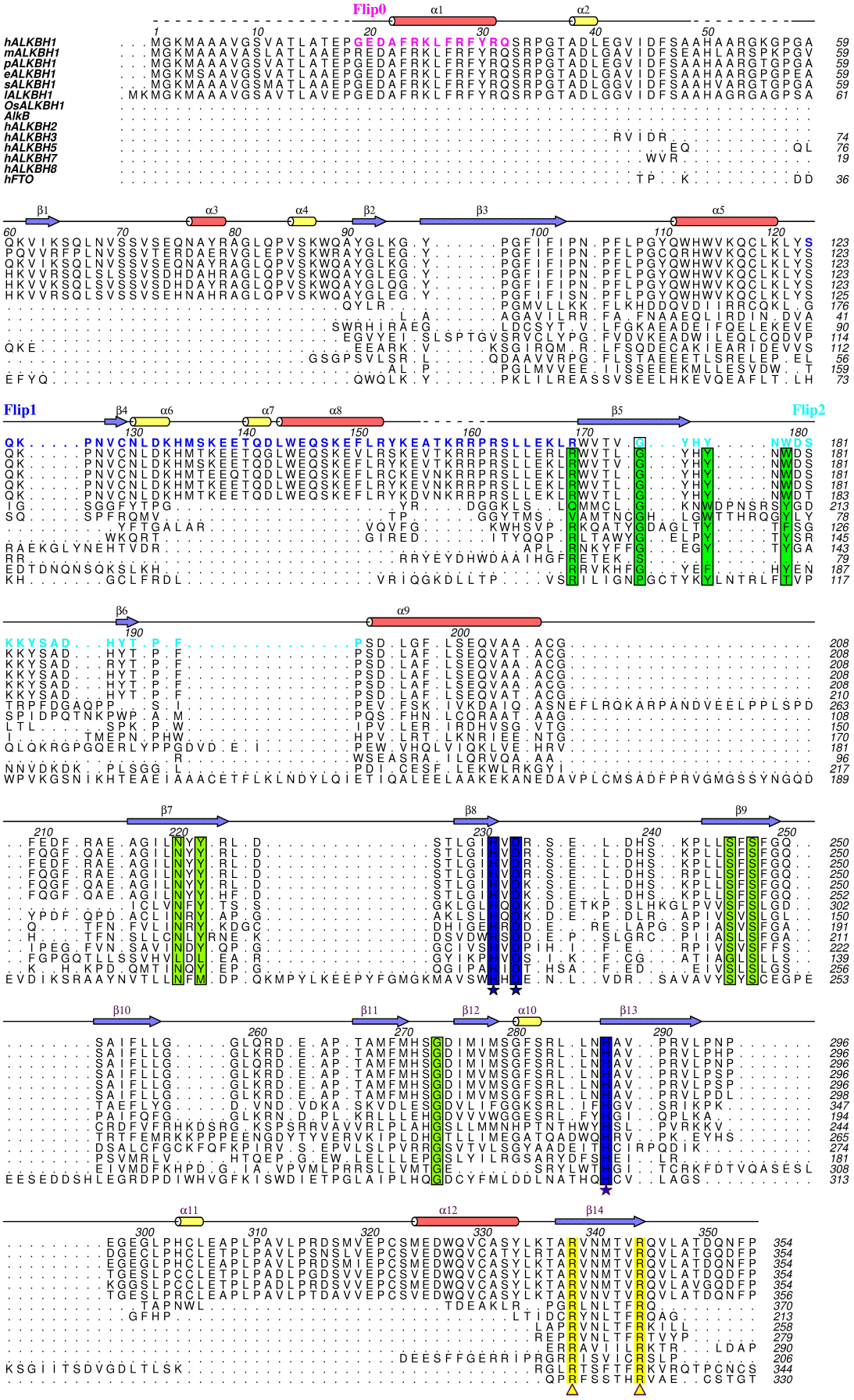


Supplementary information, Fig. S2. Overall structure of *h*ALKBH1-Mn^{2+}-α-KG. a, Topology structure of *h*ALKBH1. β-strands are shown as arrows, α-helices as cylinder. **b,** The stereo cartoon representation of *h*ALKBH1-Mn^{2+}-α-KG structure. Mn^{2+} is shown in magenta ball. The residues that have no visible electron density are shown in dashed lines. Key residues as well as α-KG are shown as sticks, and secondary structures are labeled. Flip0, Flip1, and Flip2 are shown in magenta, blue, and cyan, respectively.

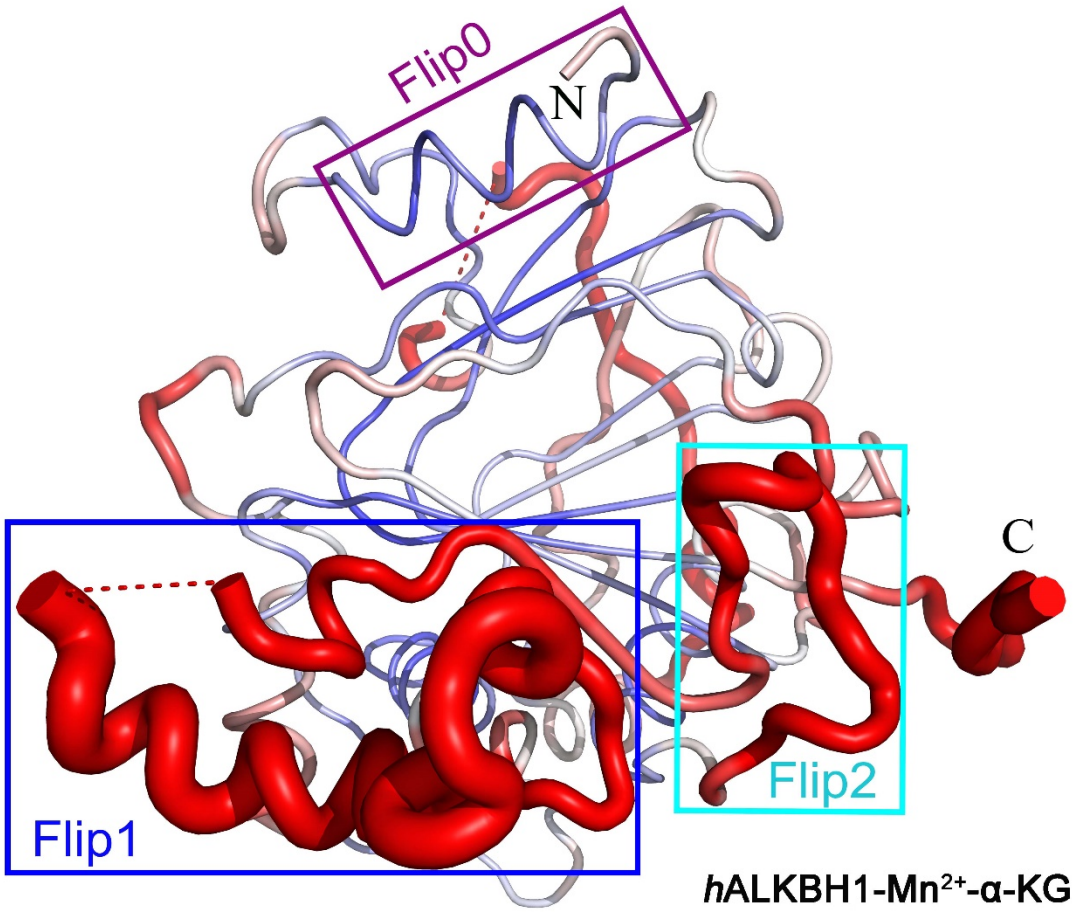


Supplementary information, Fig. S3. The detailed interaction of the active center of *h*ALKBH1. a, The superposition of *h*ALKBH1 (purple) with AlkB (cyan; PDB ID: 3T4V), *h*ALKBH3 (yellow; PDB 2IUW) at the active site. All the critical residues are shown as sticks. The two residues following the conserved HX(D/E) motif from the three proteins, α-KG, and Mn^{2+} from *h*ALKBH1 are labeled. **b,** Equilibrium binding analysis of *h*ALKBH1 with α-KG in the presence of the Mn^{2+} ion using ITC. **c,** Superposition of *h*ALKBH1-Mn^{2+}-α-KG (purple) with *h*ALKBH1-Mn^{2+} (yellow) at the active site. All the critical residues are shown as sticks. The two residues following the conserved HX(D/E) motif from the two proteins, α-KG, and Mn^{2+} are labeled. The distance between side chains of residues is represented by black dashed lines. **d,** Comparison diagram of Flip2 in *h*ALKBH1-Mn^{2+}-α-KG with *h*ALKBH1-Mn^{2+}. Although the total volume of ALKBH1 did not change during the binding of α-KG, the surface area obviously decreased from 14,445 A^2 to 13,893 A^2 if the inconsistent residues in the solved structures

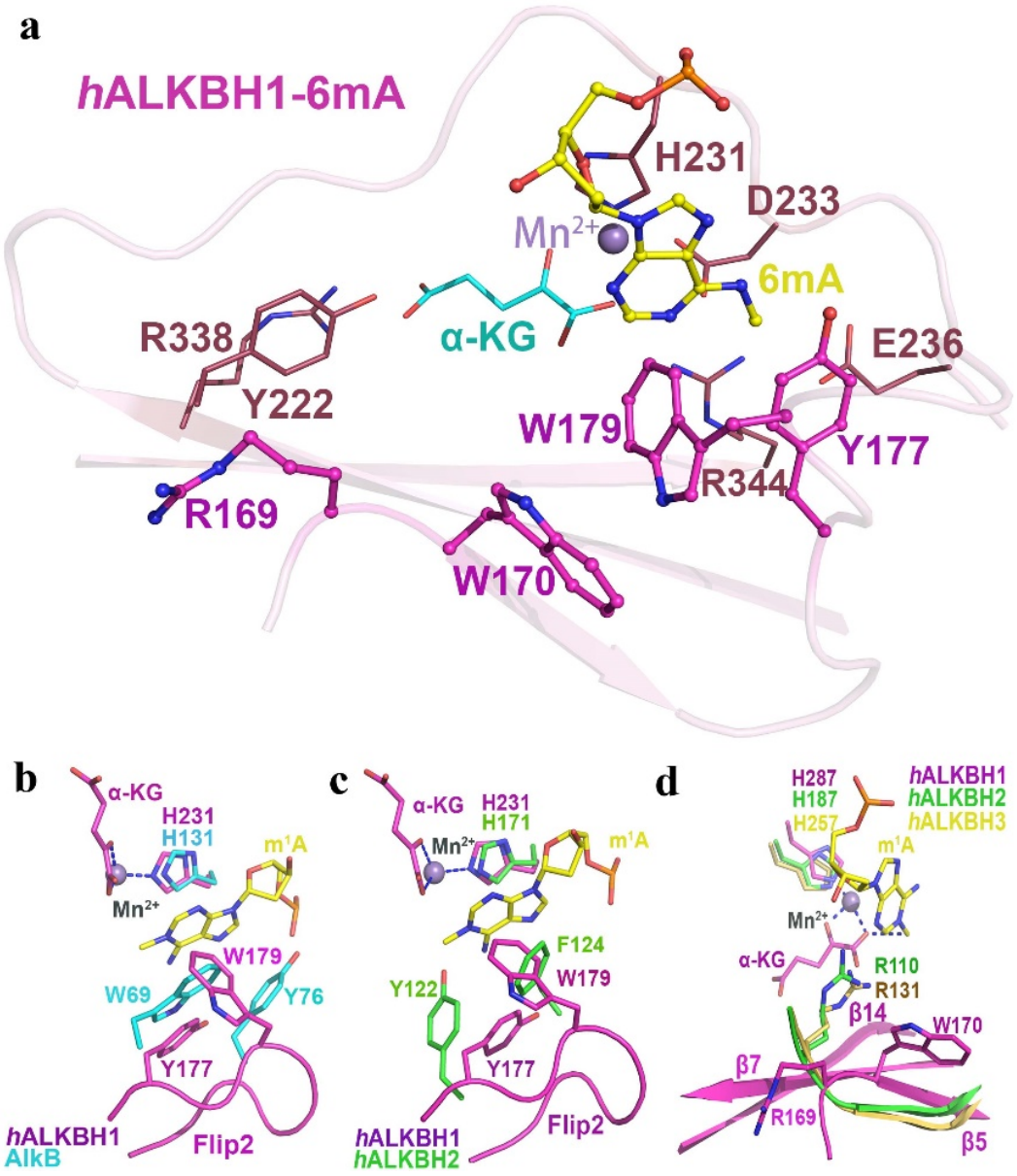
were not calculated. The data indicate that ALKBH1 whole structure shrink with the association of α-KG. Compare with the structure of *h*ALKBH1-Mn^{2+}, α-KG binding make the Flip1 shifts 3.4 Å, the Flip2 shifts 6.4 Å to the core area, the loop after β8 reduces shifts 3.9 Å , β5 shifts 3.0 Å to the core area. **e,** EMSA binding assays were performed for ssDNA with *h*ALKBH1 mutants and WT. Assays were repeated three times.


Flip0
Flip1
Flip2
hALKBH1
mALKBH1
pALKBH1
eALKBH1
sALKBH1
lALKBH1
OsALKBH1
AlkB
hALKBH2
hALKBH3
hALKBH5
hALKBH7
hALKBH8
hFTO

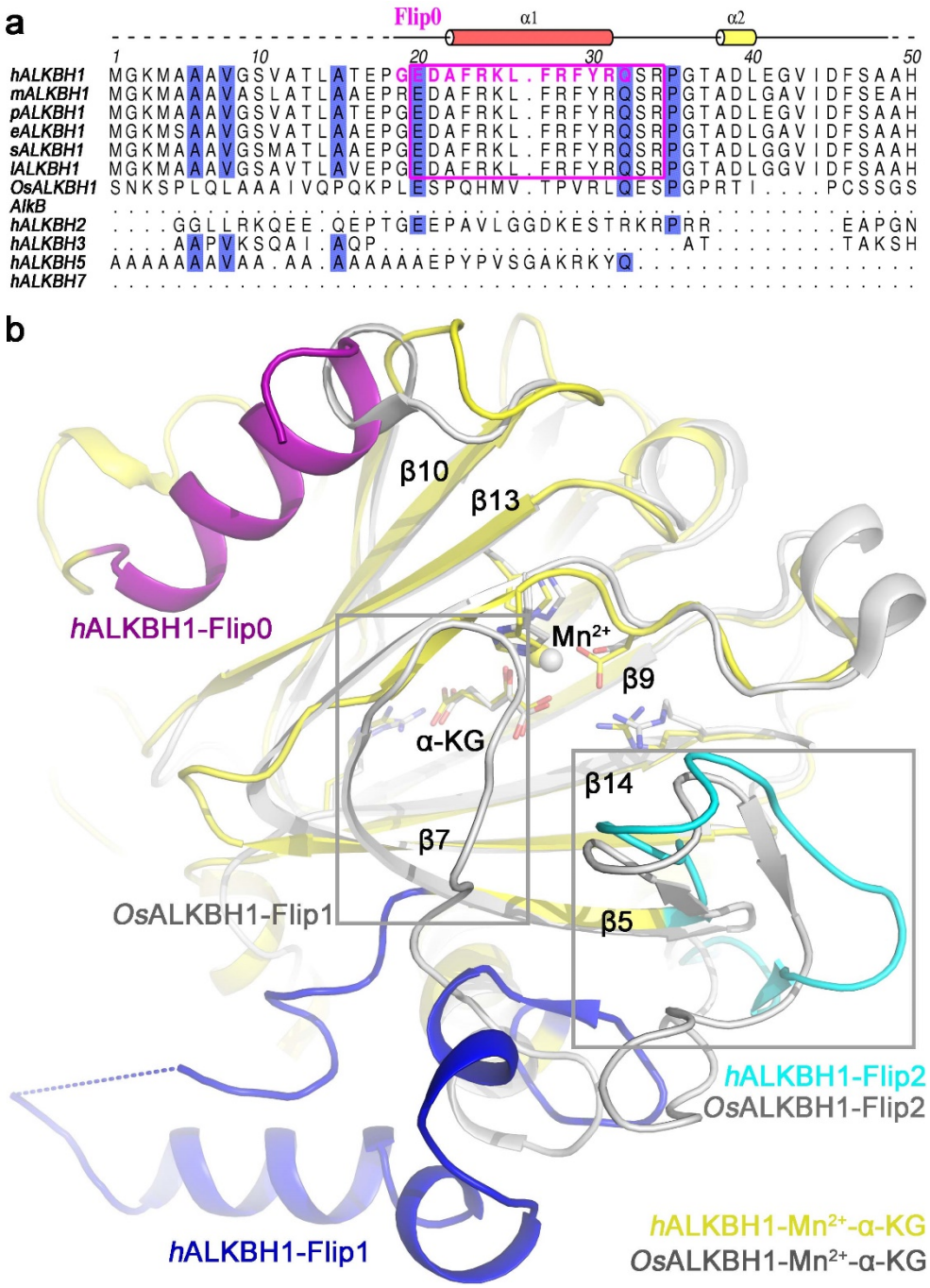
Supplementary information, Fig. S4. Structure-based sequence alignment of ALKBH1 from different species and its family members. The sequences were aligned using the server T-Coffee[7]. The ALINE program[8] was used to prepare the figure. Secondary structure elements and sequence numbering of *h*ALKBH1 are shown above the sequence. Flip1, Flip2, and Flip0 of *h*ALKBH1 are expressed in blue, cyan, and magenta, respectively. The residues involved in the binding of α-KG are labeled with yellow triangle, and the key residues in the nucleotide recognition lid that are involved in 6mA recognition are colored blue and highlighted by blue asterisks. h, *Homo sapiens*; m, *Mus musculus*; p, *Pongo abelii*; e, *Equus caballus*; s, *Sus scrofa*; l, *Loxodonta africana*; Os, *Oryza sativa*. The sequences are used as follows: *h*ALKBH1: NP_006011; *e*ALKBH1: XP_001493983; *p*ALKBH1: XP_002825034; *l*ALKBH1: XP_003408846; *s*ALKBH1: XP_020955215; *m*ALKBH1: NP_001096035; *Os*ALKBH1, XP_015628051; AlkB: NP_416716; *h*ALKBH2: NP_001001655; *h*ALBKH3: NP_631917; *h*ALKBH5: NP_060228; *h*ALKBH7: NP_115682; *h*ALKBH8: XP_016874046; *h*FTO: NP_001350823.



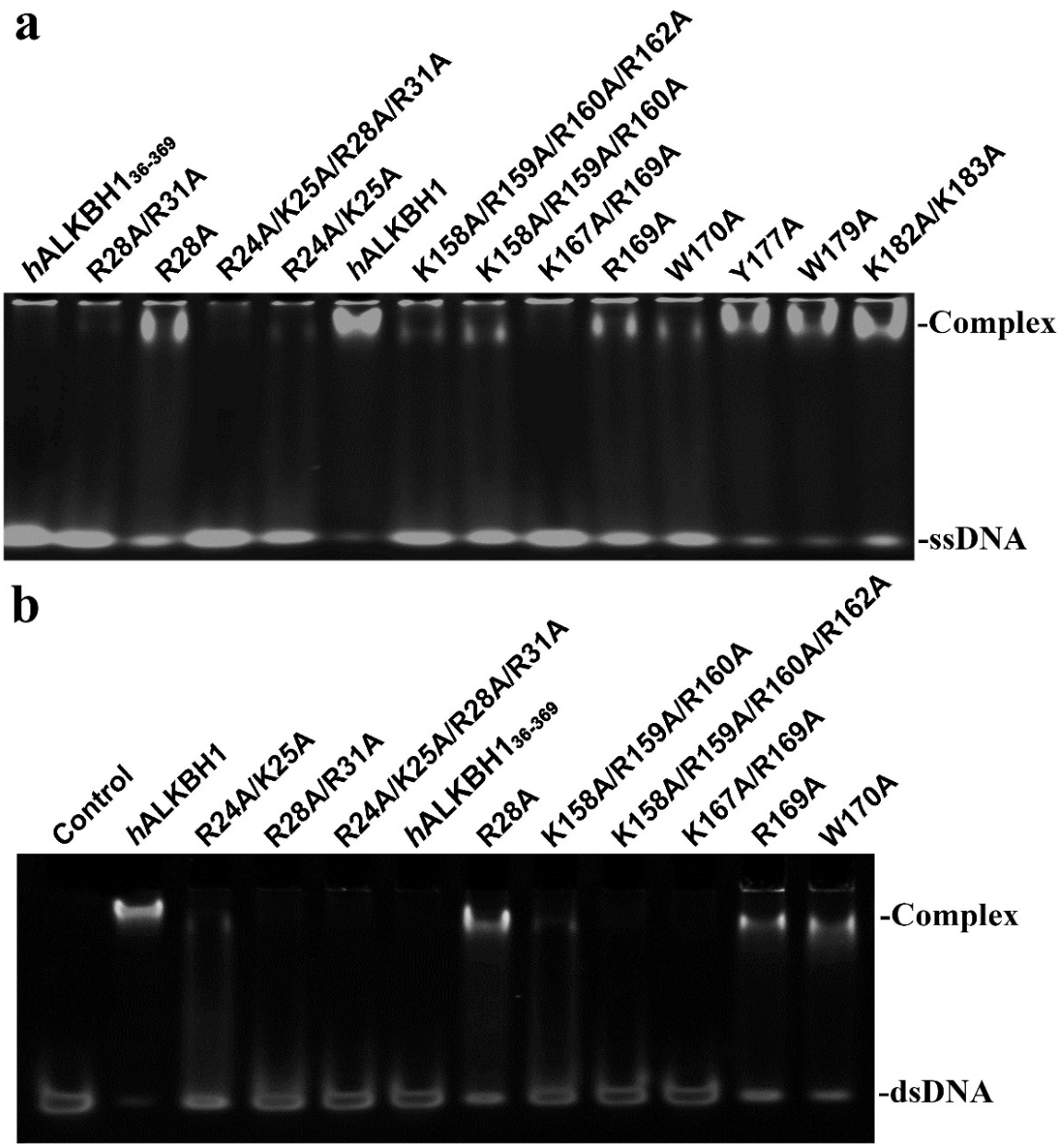
Supplementary information, Fig. S5. Overall B-factors of *h*ALKBH1. Color-coded on the basis of the calculated B-factors. The colors range from blue to red corresponding to increasing fluctuations. Flip0, Flip1, and Flip2 are highlighted in purple, blue, and cyan boxes, respectively.



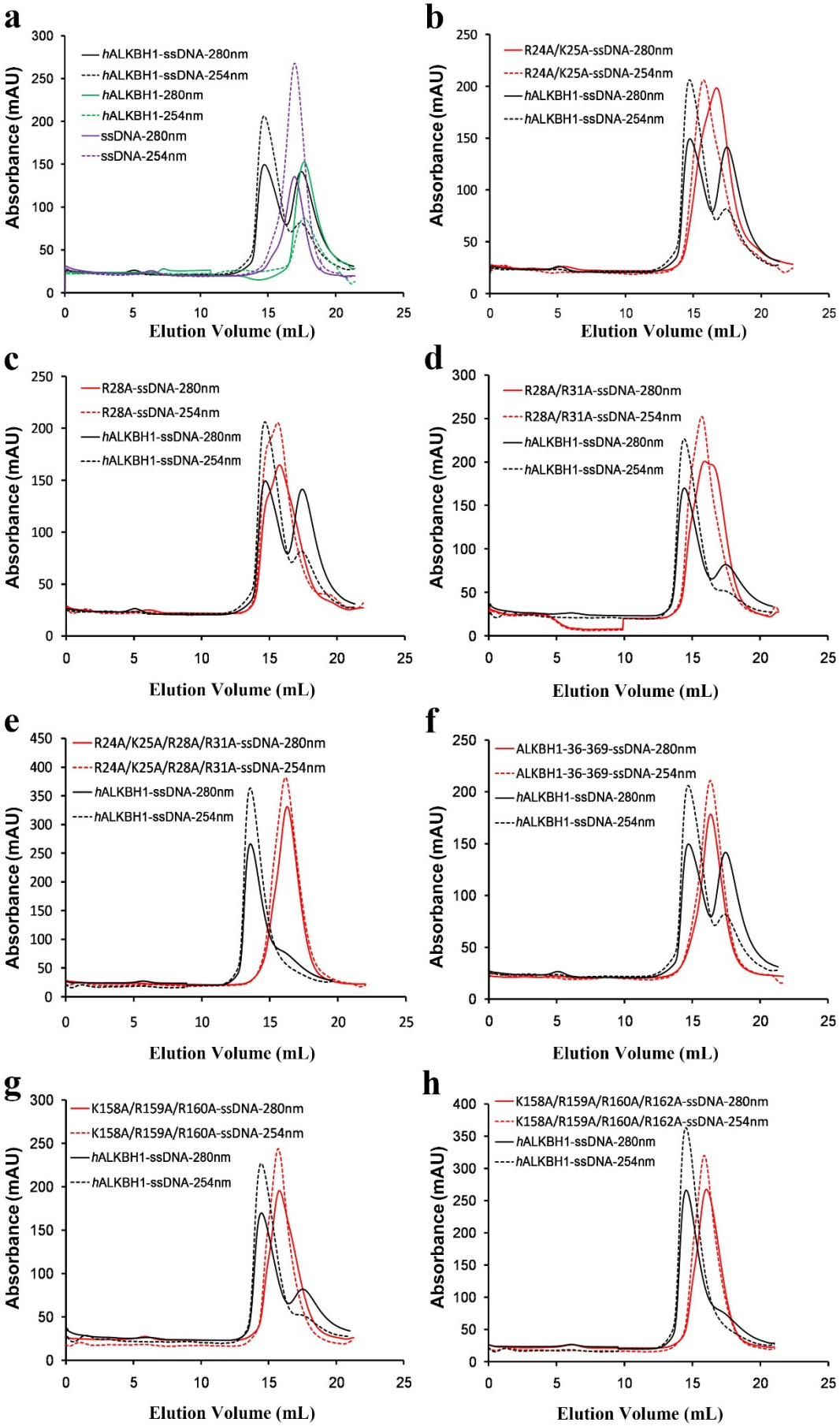
Supplementary information, Fig. S6. The distinct lid region of *h*ALKBH1 is vital for its substrate recognition and catalysis. **a,** *h*ALKBH1 conservative residues Arg169, Trp170, Tyr177, and Trp179 on Flip2 form a recognition pocket to bind substrates. 6mA is superimposed from AlkB-6mA complex (PDB 4NID)[9] and colored yellow. **b-d,** Structural comparison of *h*ALKBH1 (cyan) with other AlkB proteins. AlkB(magenta; PDB 4NIH), *h*ALKBH2 (green; PDB 3BTZ) and *h*ALKBH3 (yellow; PDB 2IVW)**.** Superposition of the partial key residues involved in nucleoside recognition. The nucleoside is colored yellow. α-KG and Mn^{2+} from Alkbh1 are labeled.

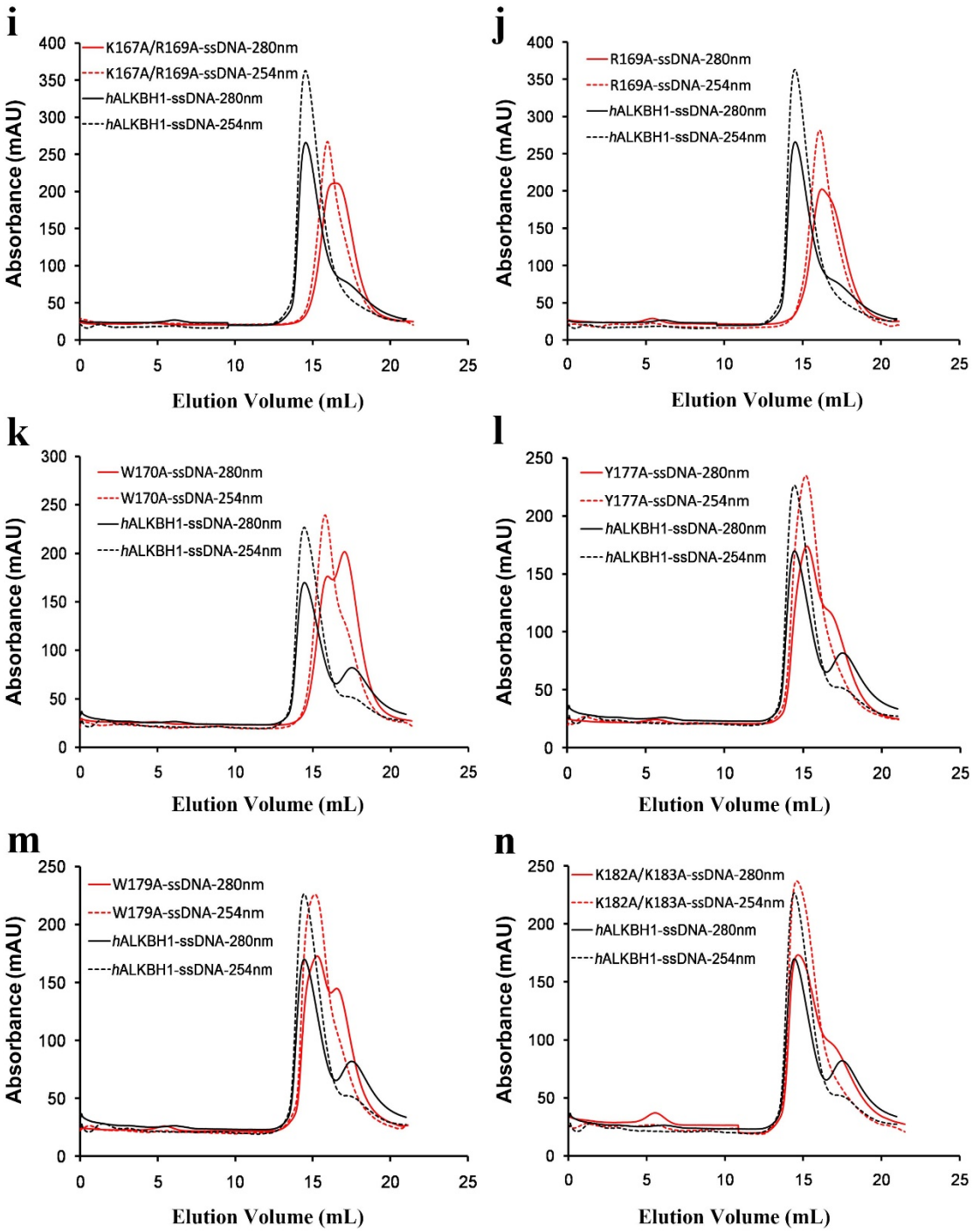


Supplementary information, Fig. S7. Relative structural information of *h*ALKBH1. a, Structure-based sequence alignment of *h*ALKBH1 of different species and its family members within Flip0 region (purple). h, *Homo sapiens*; m, *Mus musculus*; p, *Pongo abelii*; e, *Equus caballus*; s, *Sus scrofa*; l, *Loxodonta africana;* Os, *Oryza Sativa*. **b,** Structure comparison of *h*ALKBH1 (yellow) and *Os*ALKBH1 (gray, PDB ID: 5XOI) [10]. α-KG and key residues are shown as sticks, Mn^{2+} is shown in small balls. Flip0, Flip1, and Flip2 of *h*ALKBH1 are shown in purple, blue, and cyan, respectively. Flip1 and Flip2 of *Os*ALKBH1 are highlighted within the gray frame.



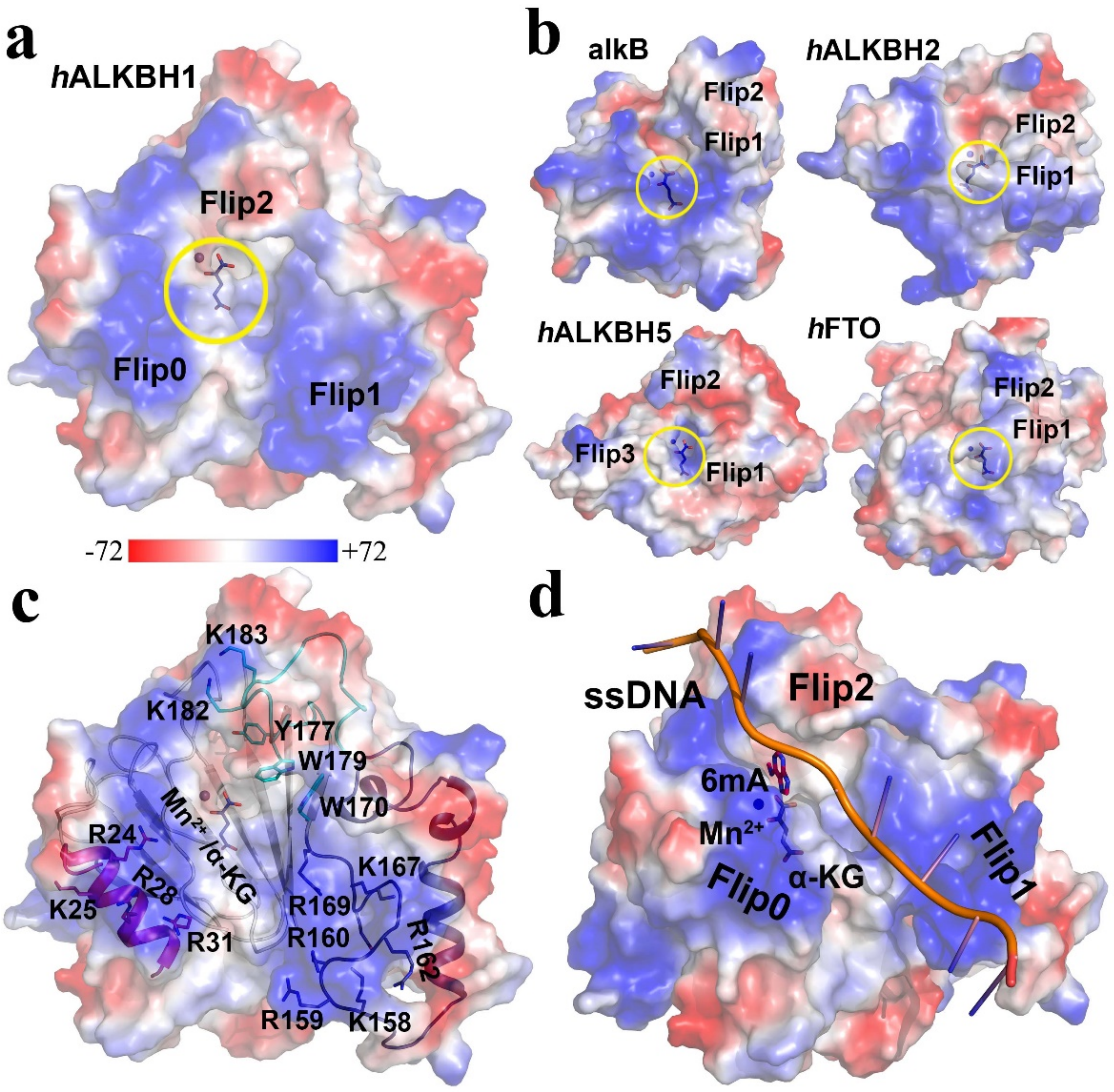
Supplementary information, Fig. S8. EMSA experiments of *h*ALKBH1 mutants with ssDNA and dsDNA. a, EMSA binding assays were performed for ssDNA with *h*ALKBH1 mutants or WT. The sequence of the ssDNA was 5'-CGGACTGGCGGCAGCACTGC-3'. substrate used was 1 μM. **b,** EMSA binding assays were performed for dsDNA with *h*ALKBH1 mutants or WT. The sequence of one chain in the dsDNA was 5'-CGGACTGGCGGCAGCACTGC-3'. The complementary chain sequence: 5'-GCAGTGCTGCCGCCAGTCCG-3'. substrate used was 1 μM.


a
hALKBH1-ssDNA-280nm
hALKBH1-ssDNA-254nm
hALKBH1-280nm
hALKBH1-254nm
ssDNA-280nm
ssDNA-254nm
Absorbance (mAU)
Elution Volume (mL)
b
R24A/K25A-ssDNA-280nm
R24A/K25A-ssDNA-254nm
hALKBH1-ssDNA-280nm
hALKBH1-ssDNA-254nm
Absorbance (mAU)
Elution Volume (mL)
c
R28A-ssDNA-280nm
R28A-ssDNA-254nm
hALKBH1-ssDNA-280nm
hALKBH1-ssDNA-254nm
Absorbance (mAU)
Elution Volume (mL)
d
R28A/R31A-ssDNA-280nm
R28A/R31A-ssDNA-254nm
hALKBH1-ssDNA-280nm
hALKBH1-ssDNA-254nm
Absorbance (mAU)
Elution Volume (mL)
e
R24A/K25A/R28A/R31A-ssDNA-280nm
R24A/K25A/R28A/R31A-ssDNA-254nm
hALKBH1-ssDNA-280nm
hALKBH1-ssDNA-254nm
Absorbance (mAU)
Elution Volume (mL)
f
ALKBH1-36-369-ssDNA-280nm
ALKBH1-36-369-ssDNA-254nm
hALKBH1-ssDNA-280nm
hALKBH1-ssDNA-254nm
Absorbance (mAU)
Elution Volume (mL)
g
K158A/R159A/R160A-ssDNA-280nm
K158A/R159A/R160A-ssDNA-254nm
hALKBH1-ssDNA-280nm
hALKBH1-ssDNA-254nm
Absorbance (mAU)
Elution Volume (mL)
h
K158A/R159A/R160A/R162A-ssDNA-280nm
K158A/R159A/R160A/R162A-ssDNA-254nm
hALKBH1-ssDNA-280nm
hALKBH1-ssDNA-254nm
Absorbance (mAU)
Elution Volume (mL)

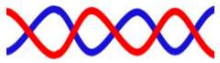
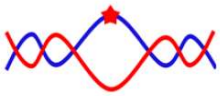
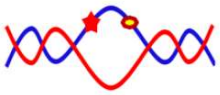
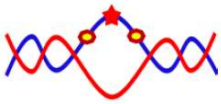
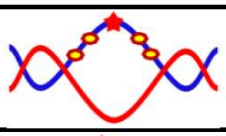
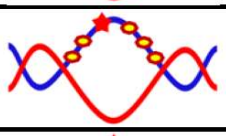
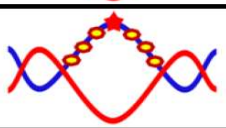
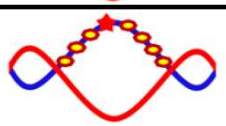
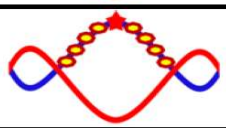
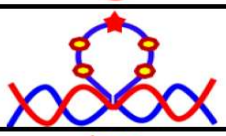
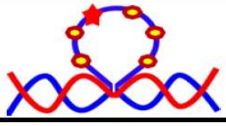
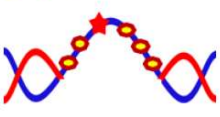


Supplementary information, Fig. S9. Chromatography of *h*ALKBH1 and its ssDNA complex. The elution peaks of *h*ALKBH1 and *h*ALKBH1-ssDNA complex were monitored by measuring absorbance at 280 and 254 nm. The sequence of the ssDNA was 5'-CGGACTGGCGGCAGCACTGC-3'. **a,** *h*ALKBH1-ssDNA. **b,** R24A/K25A. **c,** R28A. **d,** R28A/R31A. **e,** R24A/K25A/R28A/R31A. **f,** *h*ALKBH1$_{36\text{-}369}$. **G,** K158A/R159A/R160A. **h,** K158A/R159A/R160A/R162A. **I,** K167A/R169A. **j,** R169A. **k,** W170A. **l,** Y177A. **m,** W179A. **n,** K182A/K183A. The experiments were performed with Superdex-200 HR 10/300 column at a

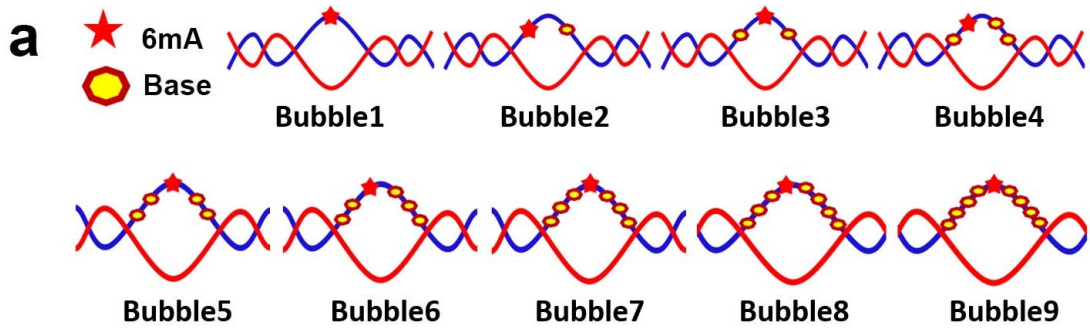
flow rate of 0.5 ml/min. *h*ALKBH1 and *h*ALKBH1-DNA complex were loaded onto the column balanced with buffer (30 mM Bis-Tris pH 6.5, 200 mM NaCl, 2 mM dithiothreitol, 5% v/v glycerol) and eluted with the same buffer.

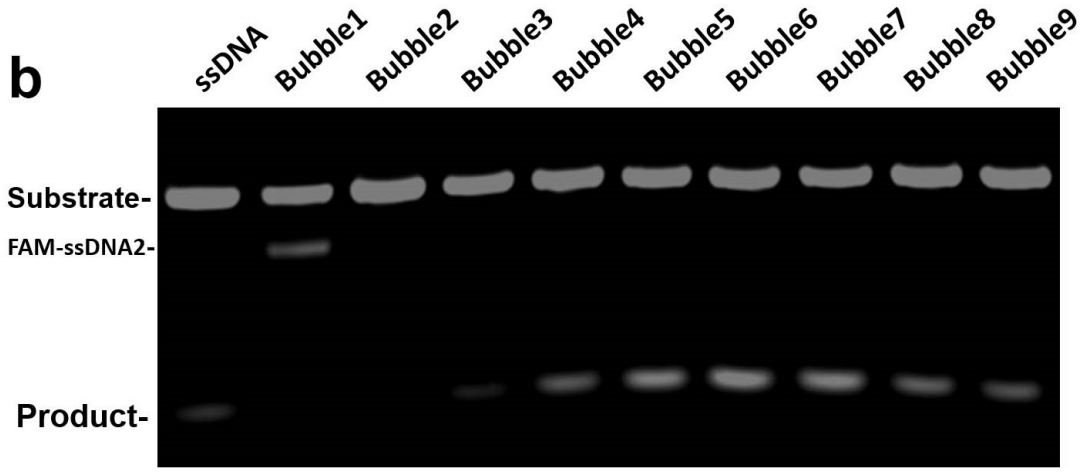


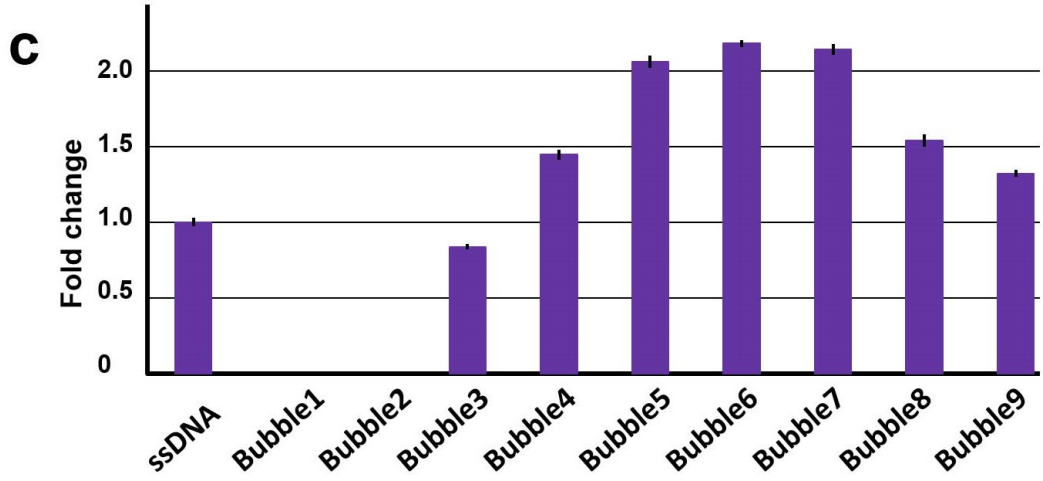
Supplementary information, Fig. S10. The substrate recognition and potential binding model of *h*ALKBH1. a-b, The surface potential (± 72 kBT/e) of *h*ALKBH5 (PDB 4NRM), AlkB (PDB 3BIE), hFTO (PDB 3LFM), *h*ALKBH2 (PDB 3BTZ) and *h*ALKBH1. The surface coloured as a gradient ranging from red (negative) to blue (positive). Flips, active center (yellow cycles), α-KG and Mn^{2+} are labeled. **c,** The critical residues located at the positively charged surface of *h*ALKBH1. **d,** The proposed model of *h*ALKBH1 binding to ssDNA (orange). The 6mA nucleobase inserted into the catalytic pocket is indicated.

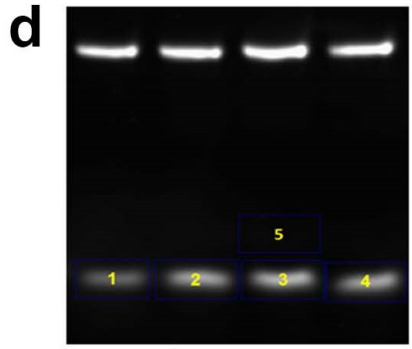
Name	Sequence	6mA: Base:
dsDNA	5'-AACTTCGTGCAGGCATGGG(6mA)TCTTGTCTACT-3' 3'-TTGAAGCACGTCCGTACCC T AGAACAGATGA-5'	
Bubble1	5'-AACTTCGTGCAGGCATGGG(6mA)TCTTGTCTACT-3' 3'-TTGAAGCACGTCCGTACCC G AGAACAGATGA-5'	
Bubble2	5'-AACTTCGTGCAGGCATGGG(6mA)TCTTGTCTACT-3' 3'-TTGAAGCACGTCCGTACCC G CGAACAGATGA-5'	
Bubble3	5'-AACTTCGTGCAGGCATGGG(6mA)TCTTGTCTACT-3' 3'-TTGAAGCACGTCCGTACCA G CGAACAGATGA-5'	
Bubble4	5'-AACTTCGTGCAGGCATGGG(6mA)TCTTGTCTACT-3' 3'-TTGAAGCACGTCCGTACCA G CTAACAGATGA-5'	
Bubble5	5'-AACTTCGTGCAGGCATGGG(6mA)TCTTGTCTACT-3' 3'-TTGAAGCACGTCCGTACTA G CTAACAGATGA-5'	
Bubble6	5'-AACTTCGTGCAGGCATGGG(6mA)TCTTGTCTACT-3' 3'-TTGAAGCACGTCCGTACTA G CTGACAGATGA-5'	
Bubble7	5'-AACTTCGTGCAGGCATGGG(6mA)TCTTGTCTACT-3' 3'-TTGAAGCACGTCCGTATTA G CTGACAGATGA-5'	
Bubble8	5'-AACTTCGTGCAGGCATGGG(6mA)TCTTGTCTACT-3' 3'-TTGAAGCACGTCCGTATTA G CTGCCAGATGA-5'	
Bubble9	5'-AACTTCGTGCAGGCATGGG(6mA)TCTTGTCTACT-3' 3'-TTGAAGCACGTCCGTCTTA G CTGCCAGATGA-5'	
Bulge5	5'-AACTTCGTGCAGGCATGGG(6mA)TCTTGTCTACT-3' 3'-TTGAAGCACGTCCGTAC---------AACAGATGA-5'	
Bulge6	5'-AACTTCGTGCAGGCATGGG(6mA)TCTTGTCTACT-3' 3'-TTGAAGCACGTCCGTAC----------ACAGATGA-5'	
Bulge7	5'-AACTTCGTGCAGGCATGGG(6mA)TCTTGTCTACT-3' 3'-TTGAAGCACGTCCGTA----------ACAGATGA-5'	
Bulge8	5'-AACTTCGTGCAGGCATGGG(6mA)TCTTGTCTACT-3' 3'-TTGAAGCACGTCCGTAC-5' 3'-ACAGATGA-5'	

Supplementary information, Fig. S11. Schematic diagram of Bubble and Bulge DNA.







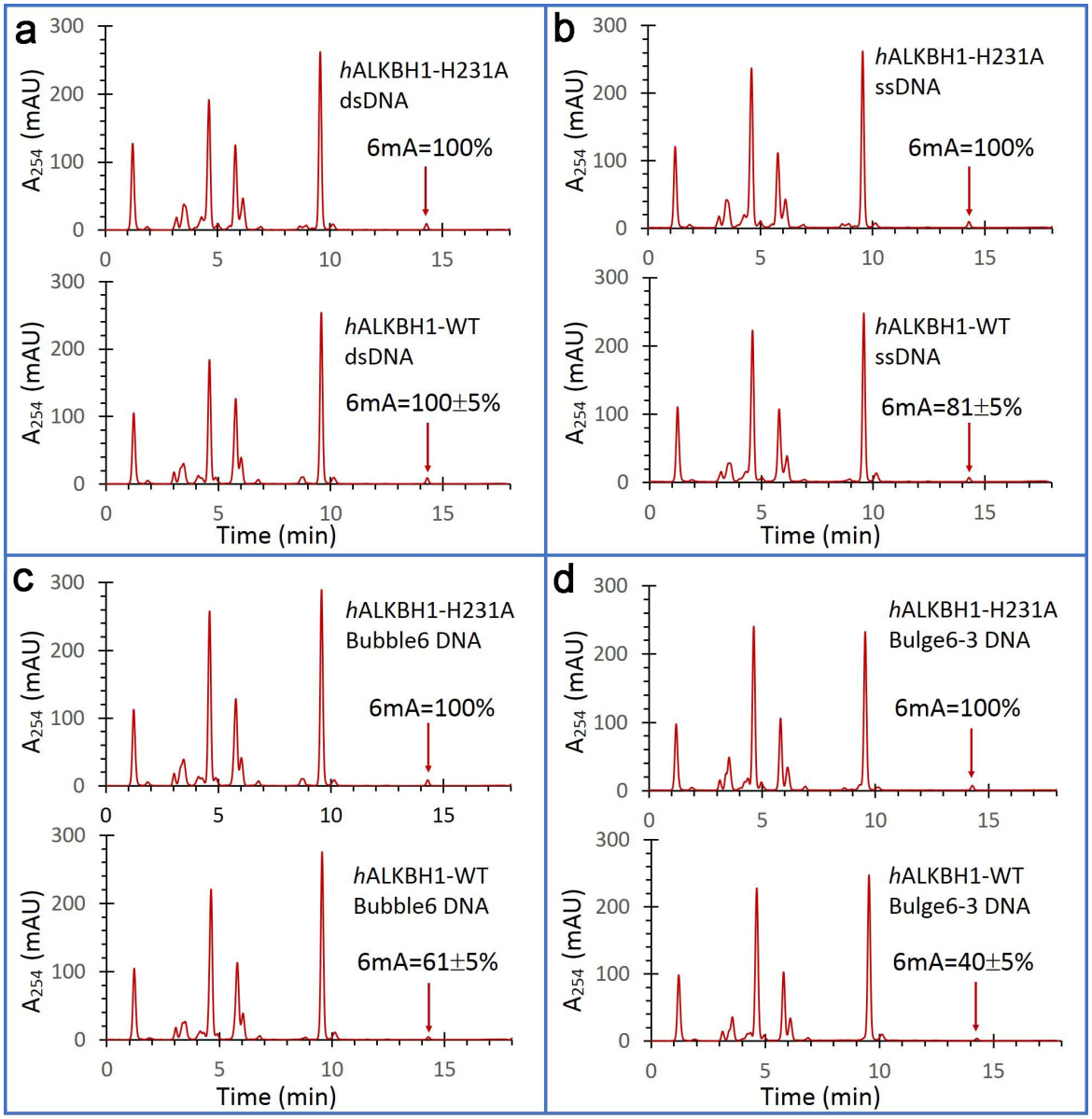


Index	Name	Adj.Vol. INT*mm²	%Adj. Vol.
1	ssDNA	466275.34	14.77
2	Bubble5	869243.80	27.53
3	Bubble6	880707.15	27.89
4	Bubble7	816069.07	25.84
5	Background	125256.79	3.97

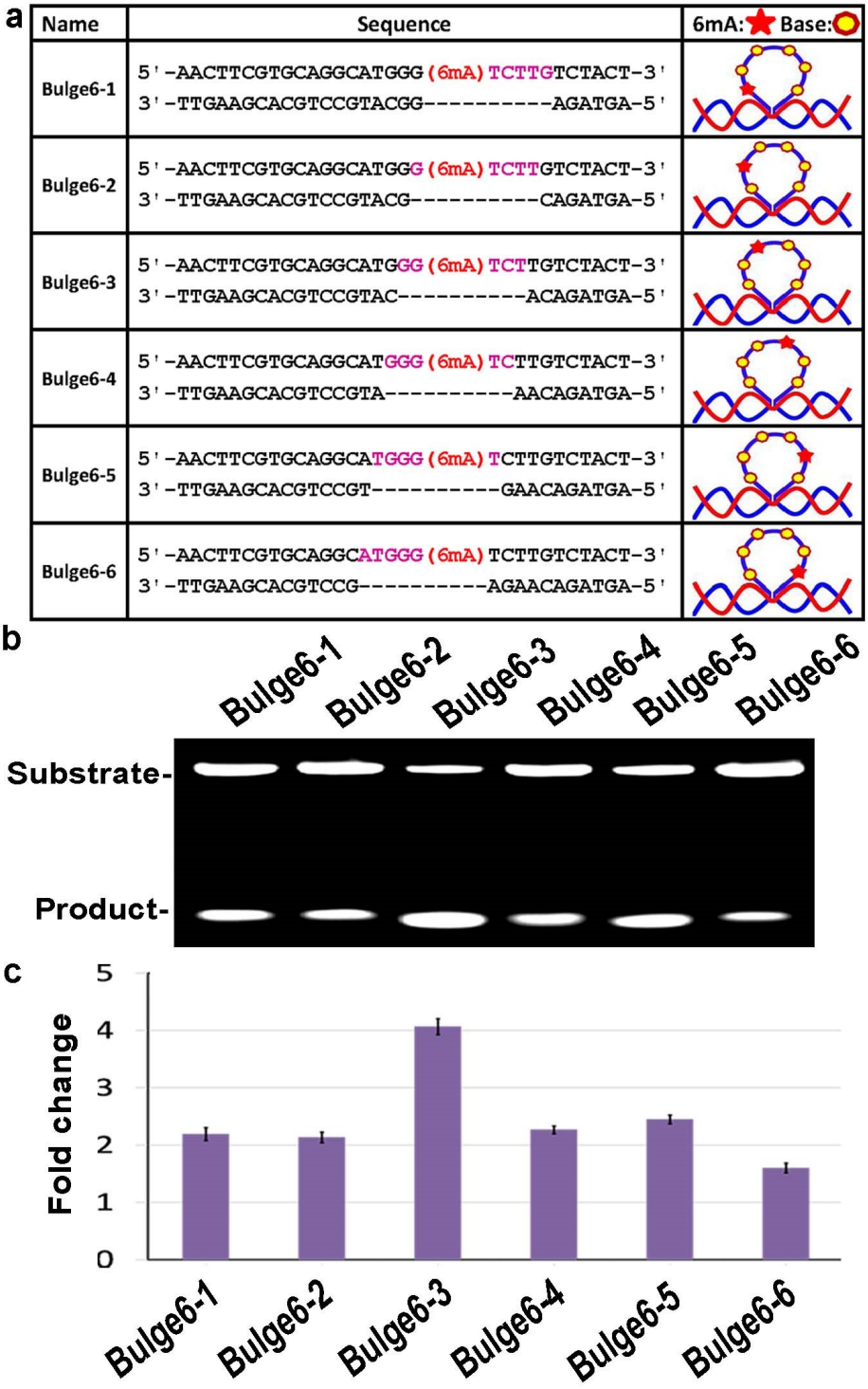
Background Subtraction Method: Global
Data units: Intensity (INT)

Supplementary information, Fig. S12. 6mA demethylation of *h*ALKBH1 on Bubble DNA.

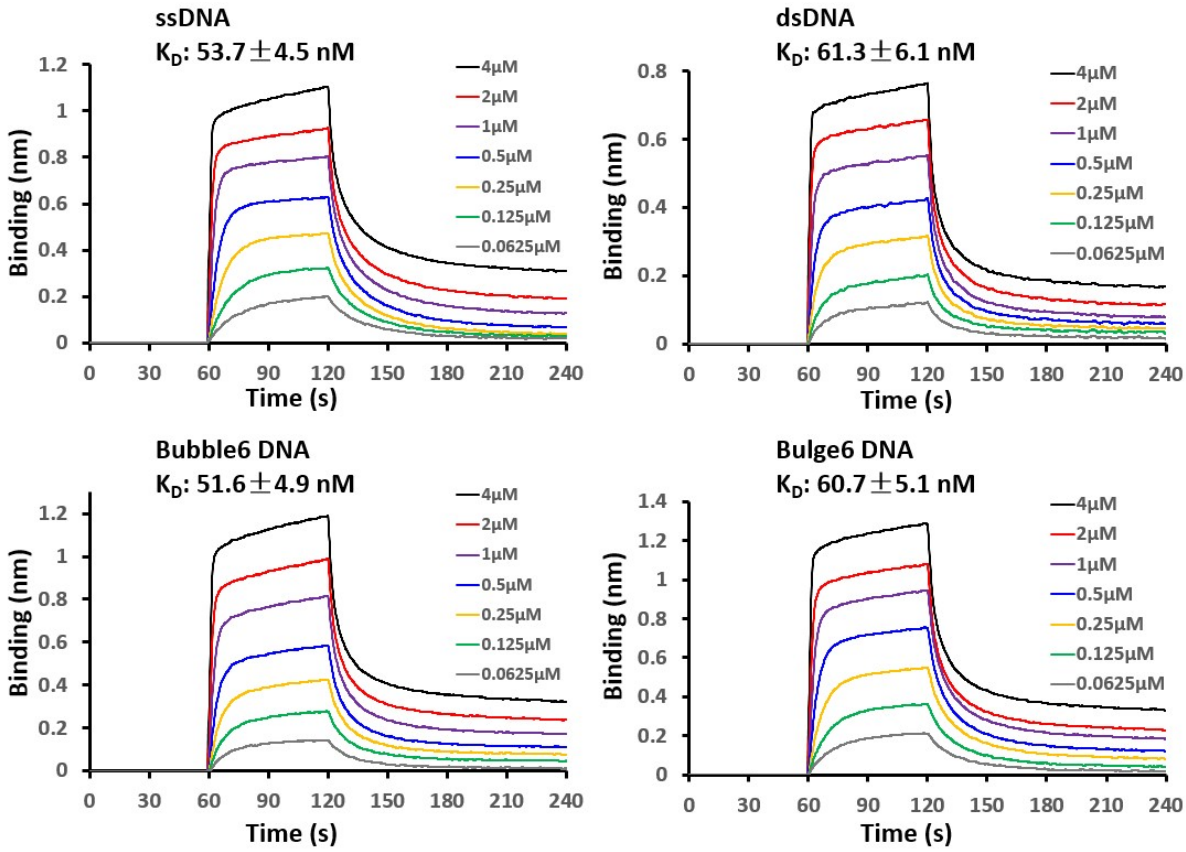
a, Schematic diagram of bubble DNA. **b,** 6mA demethylation of *h*ALKBH1 with different bubble DNA by electrophoresis. The samples were detected at 30 min after demethylation reaction. Before detection, the demethylated *h*ALKBH1 product was enzymatically digested using nuclease DpnII. Among them, *h*ALKBH1 had the strongest demethylation activity to Bubble6. There was an obvious band of FAM-ssDNA2 for Bubble1, due to the less mismatch (only one mismatch) between Bubble1-ssDNA and ssDNA_6mA1. During the formation of paired dsDNA annealing, Bubble1-ssDNA strongly competed with FAM-ssDNA2 to bind ssDNA_6mA1. Thus, much FAM-ssDNA2 existed as free single strand and was not cut by DpnII. This observation also proved that for other Bubble DNAs, ssDNA_6mA1 and its product did not form mismatched DNA during annealing after adding FAM-ssDNA2. **c,** Statistical results of *h*ALKBH1 activity. The activity of *h*ALKBH1 to bubble6 was two fold higher than to ssDNA. Assays were repeated three times. **d,** Quantity One software is used to calculate the gray value of gel image.



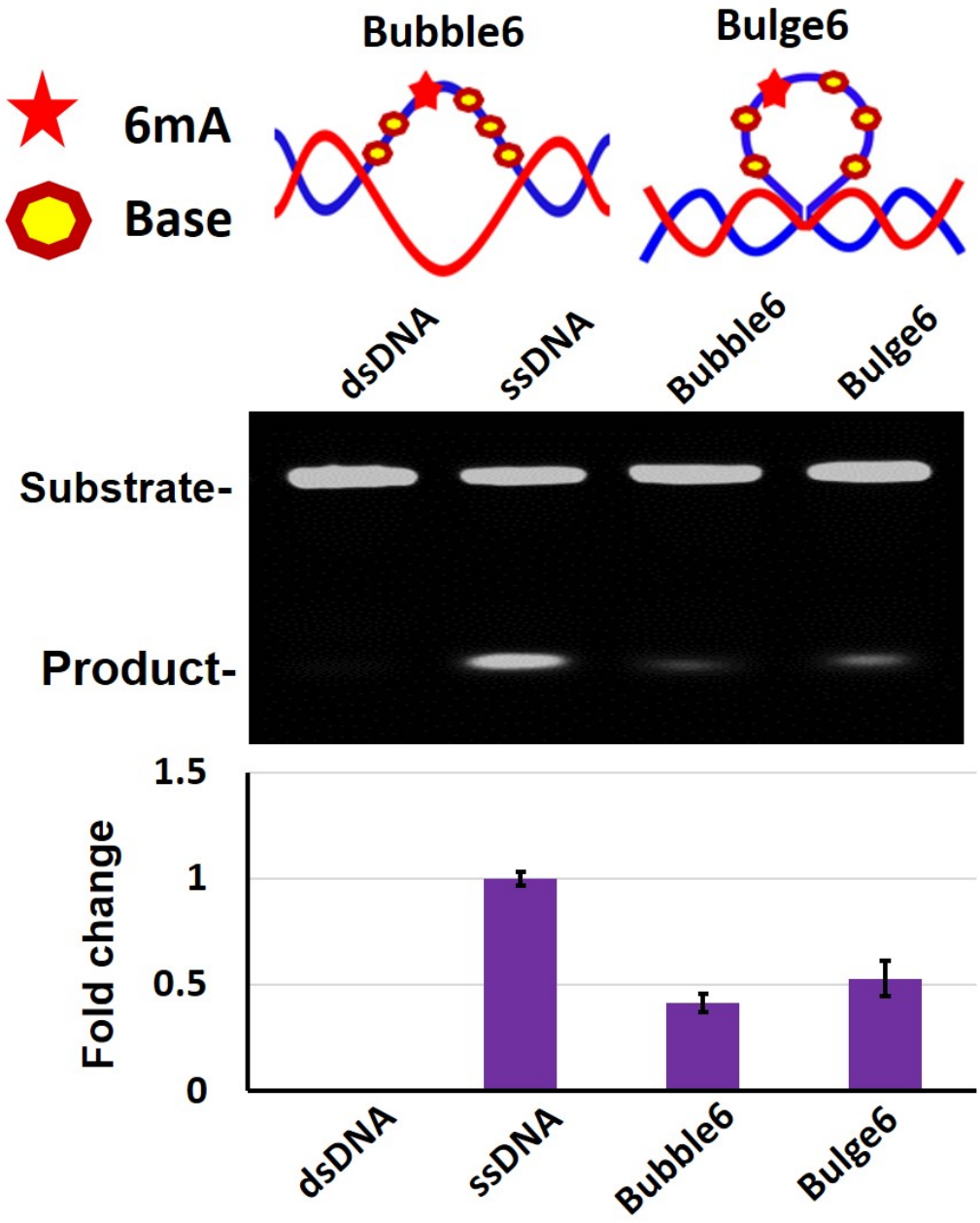
Supplementary information, Fig. S13. ***In vitro*** **6mA demethylation of *h*ALKBH1 to DNA by HPLC**. **a,** Compared with the mutant H231A, *h*ALKBH1 showed no demethylation activity to dsDNA. **b,** *h*ALKBH1 showed weak demethylation activity to ssDNA, and could remove ~20% of 6mA modification from the substrates. **c,** *h*ALKBH1 had better demethylation activity to Bubble6 DNA and could remove ~40% of 6mA modification. **d,** *h*ALKBH1 had the best demethylation activity for Bulge6-3 DNA, and could remove ~60% of 6mA modification from the substrate. The relative content of 6mA is same when using T, C or G as internal standard, respectively. In each panel, the relative content of 6mA in the reaction of the inactive mutant H231A is considered as 100%.



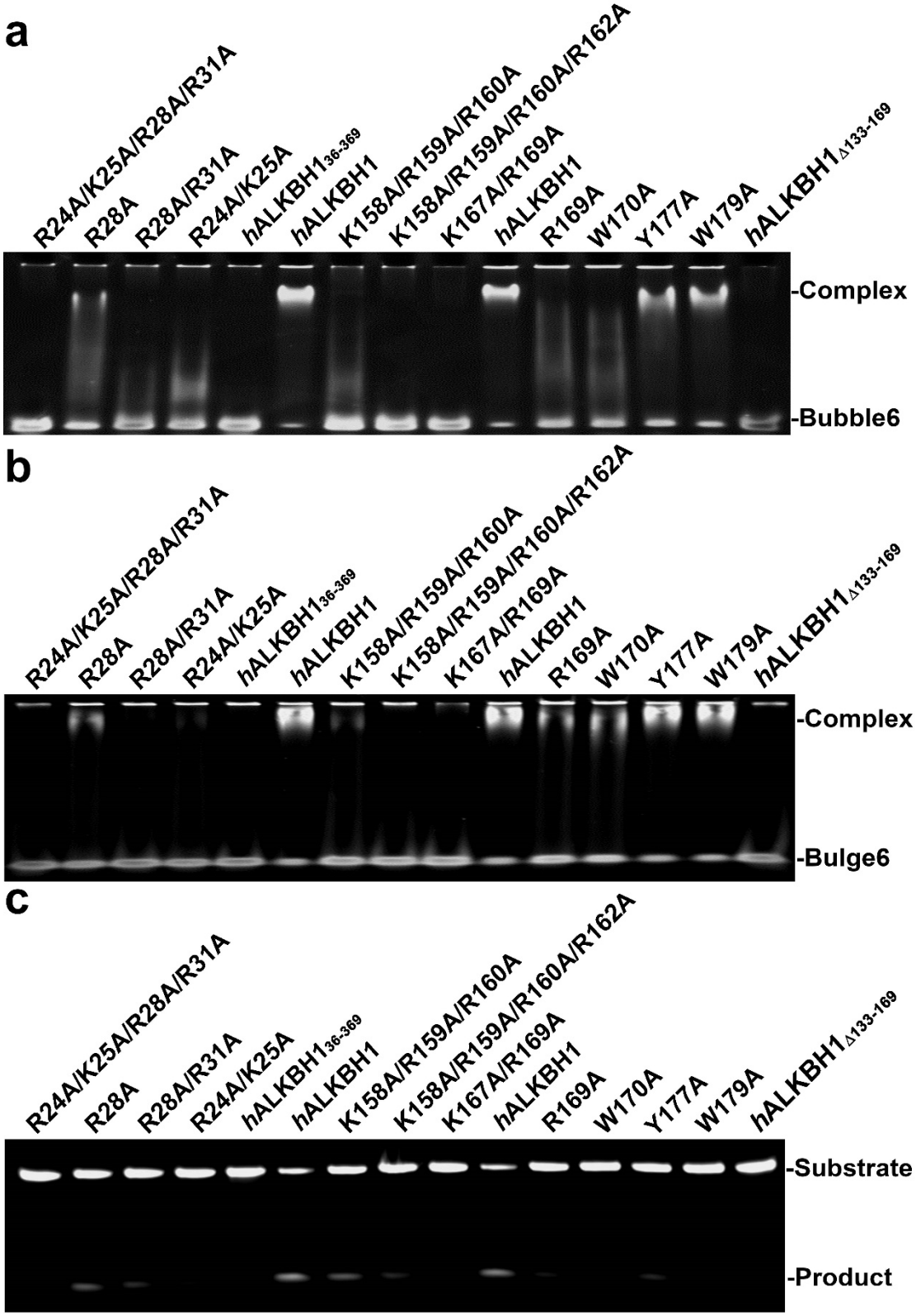
Supplementary information, Fig. S14. *In vitro* **6mA demethylation of *h*ALKBH1 to Bulge DNA. a,** Schematic diagram of Bulge6 DNA. Bulge6-3 is the Bulge6 used in the previous experiments. **b-c,** Demethylation of *h*ALKBH1 with different 6mA sites on the Bulge6 DNA by electrophoresis, the samples were detected at 30 min after demethylation reaction. Before detection, the demethylated *h*ALKBH1 product was enzymatically digested using nuclease DpnII. Assays were repeated three times. Error bars, s.d. of three replicates.



Supplementary information, Fig. S15. The interaction between *h*ALKBH1 and four kinds of DNA (ssDNA, dsDNA, Bubble6 and Bulge6) were detected by Octet. The results showed that there were strong interactions between *h*ALKBH1 and ssDNA, dsDNA, Bubble6 and Bulge6. Their K_D values were ssDNA: 53.7±4.5nM, dsDNA: 61.3±6.1nM, Bubble6: 51.6±4.9nM, Bulge6: 60.7±5.1nM. The K_D values of four kinds of DNA and *h*ALKBH1 were close to each other, which indicated that there was strong interaction between four kinds of DNA and *h*ALKBH1, and there were no significant difference. Assays were repeated three times.

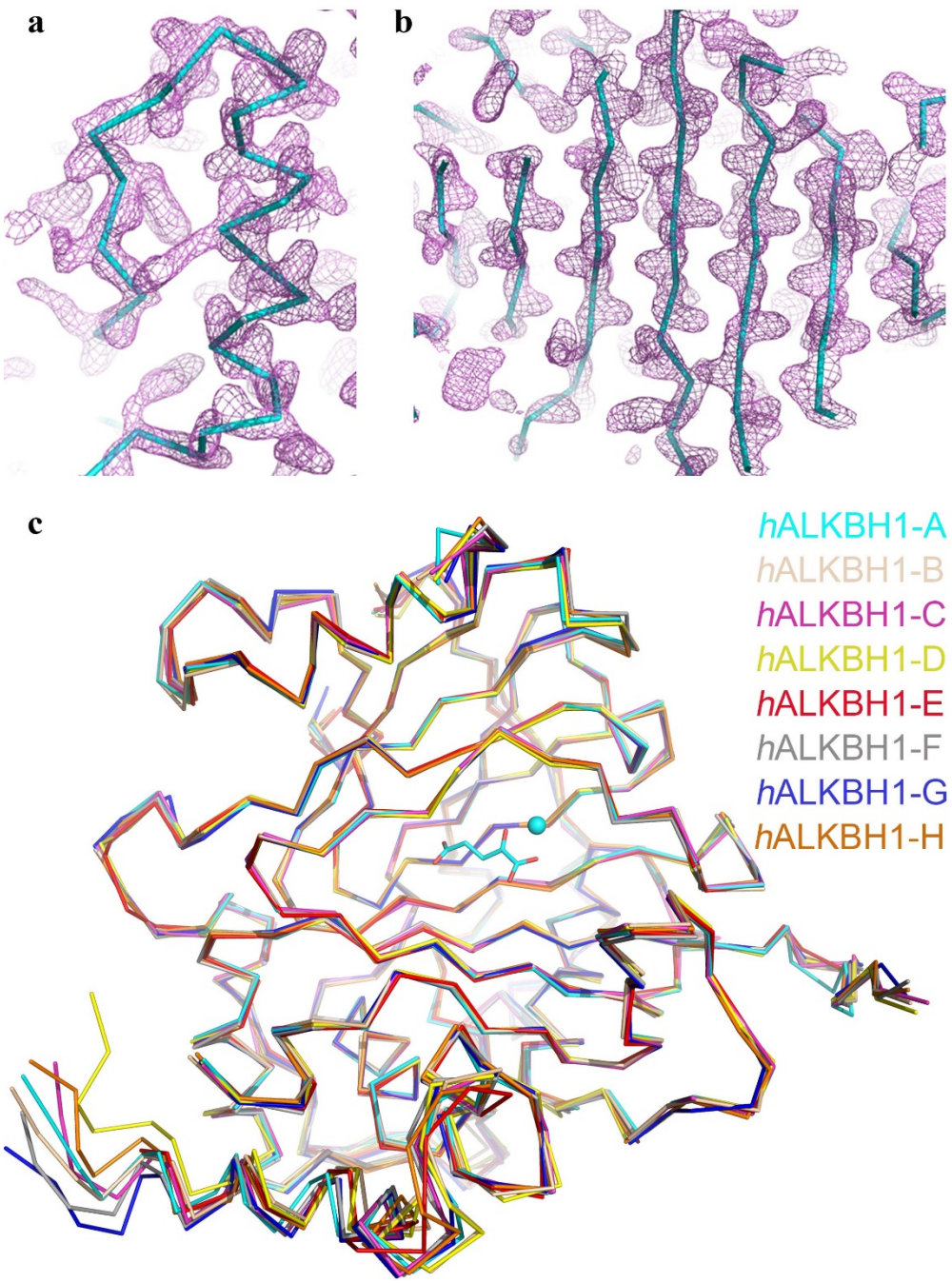


Supplementary information, Fig. S16. The demethylation of 6mA on ssDNA, dsDNA, Bubble6 and Bulge6 of *h*ALKBH5 was studied. The samples were detected at 30 min after demethylation reaction. The results showed that when the substrate was Bubble6 or Bulge6, the demethylation activity of 6mA in *h*ALKBH5 was 41% or 52.5% respectively of ssDNA. The final results showed that ssDNA was the best substrate for demethylation of 6mA in *h*ALKBH5.



Supplementary information, Fig. S17. Biochemical data of *h*ALKBH1 mutants on Bubble6 and Bulge6. **a,** EMSA binding assays were performed for Bubble6 with key mutants

of *h*ALKBH1 Flip0, Flip1, and Flip2. **b,** EMSA binding assays were performed for Bulge6 with key mutants of *h*ALKBH1 Flip0, Flip1, and Flip2. **c,** Detection of the demethylation activity by electrophoresis, The samples were detected at 30 min after demethylation reaction. Before detection, the demethylated *h*ALKBH1 product was enzymatically digested using nuclease DpnII. Mutations of the key residues greatly impaired *h*ALKBH1 demethylation to 6mA for bubble6 DNA. The assays were reproduced three times with similar results.



Supplementary information, Fig. S18. Representative electron density maps of *h*ALKBH1. Representative experimental electron density map of **a,** α-helices and **b,** β-sheets of *h*ALKBH1$_{19\text{-}369}$-Mn^{2+} molecules. The electron density *2mFo-DFc* is contoured at 2.0 σ.**c**, Molecular alignment of the eight *h*ALKBH1-Mn^{2+}-α-KG complex in the asymmetric unit. Eight *h*ALKBH1 molecules are shown as ribbon. Protomer A is shown in cyan, B is shown in wheat, C is shown in magenta, D is shown in yellow, E is shown in red, F is shown in gray, G is shown in blue, H is shown in orange. α-KG is shown as sticks, Mn^{2+} is shown in ball. Note that: The diffraction data cannot be processed using other space groups except P1.

Supplementary information, Table S1.

Data collection and refinement statistics of *h*ALKBH1 complexes

	*h*ALKBH1-Se	*h*ALKBH1-Mn^{2+}	*h*ALKBH1-Mn^{2+}-α-KG
Data collection			
Wavelength (nm)	0.9793	0.9793	0.9793
Space group	$P6_5$	$P6_5$	$P1$
Cell dimensions			
a,b,c (Å)	57.1, 57.1, 192.2	56.7, 56.7, 191.7	89.0, 99.2, 126.8
α,β,γ (˚)	90, 90, 120	90, 90, 120	106.6, 90.4, 93.1
Resolution (Å) †	50-2.7 (2.75-2.70)	50-1.97 (2.0-1.97)	50-2.8 (2.85-2.80)
Rmerge (%)	8.1 (69.3)	9.5 (74.6)	6.5 (72.2)
I/σ	40.9 (2.42)	23.1 (2.52)	11.2 (1.2)
Completeness (%)	99.8 (99.0)	99.4 (98.4)	93 (73.8)
Total No. of reflections	247,402	375,159	353,308
Unique reflections	9,725	24,558	72,797
Redundancy	25.4 (15.6)	15.3 (9.9)	3.7 (3.1)
Refinement			
Resolution (Å)		50-1.97 (2.02-1.97)	50-2.80 (2.87-2.80)
No. of reflections		20,188 (621)	69,068 (1351)
Rwork/Rfree (%)		20.3/23.3	21.6/25.4
No. of atoms			
Protein		2,430	20,086
ligand/ion		4	88
Water		283	486
B-factors (Å2)			
Protein		33.3	55.2
ligand/ion		40.2	44.5
Water		41.9	38.5
r.m.s.deviations			
Bond lengths (Å)		0.007	0.008
Bond angles (º)		1.181	1.24
Ramachandran plot (%)[1]		94.7/5.3/0	90.3/9.7/0

†Statistics for highest resolution shell.

[1]Residues in most favored, generously allowed and disallowed regions of the Ramachandran plot.

Supplementary information, Movie S1

Movie S1

Great conformational changes induced by α-KG were observed in the structure of *h*ALKBH1-Mn^{2+}-α-KG complex.