Mammalian ALKBH1 serves as an $N^6$-mA demethylase of unpairing DNA

Min Zhang1, Shumin Yang1, Raman Nelakanti2, Wentao Zhao1, Gaochao Liu1, Zheng Li2, Xiaohui Liu3, Tao Wu2,4, Andrew Xiao2 and Haitao Li5

$N^6$-methyladenine ($N^6$-mA) of DNA is an emerging epigenetic mark in mammalian genome. Levels of $N^6$-mA undergo drastic fluctuation during early embryogenesis, indicative of active regulation. Here we show that the 2-oxoglutarate-dependent oxygenase ALKBH1 functions as a nuclear eraser of $N^6$-m in unpairing regions (e.g., SIDD, Stress-Induced DNA Double Helix Destabilization regions) of mammalian genomes. Enzymatic profiling studies revealed that ALKBH1 prefers bubbled or bulged DNAs as substrate, instead of single-stranded (ss-) or double-stranded (ds-) DNAs. Structural studies of ALKBH1 revealed an unexpected “stretch-out” conformation of its “Flip1” motif, a conserved element that usually bends over catalytic center to facilitate substrate base flipping in other DNA demethylases. Thus, lack of a bending “Flip1” explains the observed preference of ALKBH1 for unpairing substrates, in which the flipped $N^6$-mA is primed for catalysis. Co-crystal structural studies of ALKBH1 bound to a 21-mer bulged DNA explained the need of both flanking duplexes and a flipped base for recognition and catalysis. Key elements (e.g., an ALKBH1-specific α1 helix) as well as residues contributing to structural integrity and catalytic activity were validated by structure-based mutagenesis studies. Furthermore, ssDNA-seq and DIP-seq analyses revealed significant co-occurrence of base unpairing regions with $N^6$-mA in mouse genome. Collectively, our biochemical, structural and genomic studies suggest that ALKBH1 is an important DNA demethylase that regulates genome $N^6$-mA turnover of unpairing regions associated with dynamic chromosome regulation.

Cell Research (2020) 30:197-210; https://doi.org/10.1038/s41422-019-0237-5

INTRODUCTION

AlkB family of 2-oxoglutarate (2OG or α-KG)-dependent oxygenases are evolutionarily conserved and implicated in nucleotide demethylation.1-3 As a founding member of its family, AlkB homolog 1 (ALKBH1) was firstly identified in 1996;4 however, its enzymatic activity remains controversial to date. Among the reported substrates of ALKBH1 are different types of methylated nucleotides of DNA5,6 or RNA,7,8 methylated lysine of histone H2A,9 as well as abasic sites of DNA.10 Meanwhile, the functional importance of ALKBH1 in embryogenesis has been demonstrated in Alkbh1 knockout mice from several independent studies, suggesting a distinct role of ALKBH1 in gene regulation and early development.11-13

DNA $N^6$-deoxymethyladenosine ($N^6$-dA) was recently identified in eukaryotic genomes,6,14-21 expanding the repertoire of epigenetic marks. In mouse embryonic stem cells (mESCs), $N^6$-dA was shown to enrich at young LINE-1 transposon and its deposition correlates with epigenetic silencing.6 In vertebrates, $N^6$-dA was further shown to accumulate to high abundance (~0.1%-0.2% $N^6$-dA in DNA) after fertilization and diminish to the background level (<0.001%) with the progression of the embryo development.15 Besides, higher level of $N^6$-dA was also detected in mouse brain upon environmental stress,22,23 as well as the highly malignant brain cancer glioblastoma,24 all of which suggest a process of active regulation of $N^6$-dA. Several studies including ours demonstrated that ALKBH1 is the demethylase of $N^6$-dA on ssDNA;5,16-18,24 however, the exact biochemical activity of ALKBH1 and the underlying molecular basis remains unclear so far.

Here, we report the biochemical, structural and genomic studies of DNA $N^6$-mA demethylation by ALKBH1. By establishing a reliable in vitro enzymatic assay, we demonstrated that ALKBH1 prefers bubbled and bulged DNAs, instead of ssDNA or dsDNA substrates. We further generalized the favorable substrates of ALKBH1 as nucleic acids that share a locally unpairing feature with flanking duplex(es), such as D-loop, R-loop, and DNA or RNA stem loop. Structural studies of ALKBH1 in substrate-free and a 21-mer bulge DNA-bound state revealed the molecular determinants underlying the observed substrate preference, including unique features of a stretch-out Flip1 motif and a functionally indispensable N-terminal “α1” helix. Furthermore, ssDNA-seq and DIP-seq analyses revealed significant co-localization of $N^6$-mA with base unpairing regions in mouse early developmental cell line, highlighting a role of mammalian ALKBH1 in dynamic genomic regulation.

1MOE Key Laboratory of Protein Sciences, Beijing Advanced Innovation Center for Structural Biology, Beijing Frontier Research Center for Biological Structure, Tsinghua-Peking Joint Center for Life Sciences, Department of Basic Medical Sciences, School of Medicine, Tsinghua University, 100084 Beijing, China; 2Department of Genetics and Yale Stem Cell Center, Yale School of Medicine, New Haven, CT 06520, USA and 3National Protein Science Technology Center, School of Life Sciences, Tsinghua University, 100084 Beijing, China
Correspondence: Andrew Xiao (andrew.xiao@yale.edu) or Haitao Li (lht@tsinghua.edu.cn)
4Present address: Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA
These authors contributed equally: Min Zhang, Shumin Yang, Raman Nelakanti

Received: 13 May 2019 Accepted: 30 August 2019
Published online: 12 February 2020
RESULTS

ALKB1 prefers bubbled DNAs as substrate instead of ss-/ds-DNAs because the catalytic activity of ALKB1 was able to demethylate the DNA, which is strictly dependent on Fe(II) and 2-oxoglutarate (Supplementary information, Table S1). Consistent with the previous study, full-length ALKB1 was able to catalyze the demethylation of N6-mA of ssDNA but not dsDNA (Fig. 1a). However, the demethylation fraction of ssDNA can only reach about 50% despite extensive efforts in reaction condition optimization (data not shown), suggesting that ssDNA may not be the optimal substrate. The preference of ALKB1 to ssDNA prompted us to inspect its activity towards bubbled DNA, a form of non-B DNA structure with regulatory potential in genome.25 Remarkably, ALKB1 displayed pronounced N6-mA demethylation activity over a 6-nt (nucleotide) DNA bubble compared to ssDNA and dsDNA (Fig. 1a). Optimal activities (demethylation fraction > 90%) were observed for 6- and 9-nt bubbles on the 41-bp dsDNA templates, and this activity dropped dramatically when the bubble size was shrunk to 3 nt or less (more dsDNA-like) and reduced gradually following bubble extensions to 16 nt (more ssDNA-like) (Fig. 1b). To explore the sequence dependence, three more oligos mimicking the natural sequences identified by SMRT-ChIP-seq analyses18,19 were screened for demethylation activity; N6-mA in all tested 6-nt bubbled DNAs were efficiently demethylated regardless of the sequence motif (Supplementary information, Fig. S2a).

We also demonstrated that the removal of N6-mA from bubbled DNA is strictly dependent on Fe(II) and 2-oxoglutarate (Supplementary information, Fig. S2b). Alanine mutation of two conserved ion-coordinating residues, H231 and D233, completely abolished the activity (Supplementary information, Figs. S1d, e, S2c).

Substrates of ALKB1 share a locally unpairing feature We further interrogated the substrate preference of ALKB1 by changing the position of N6-mA in the 6-nt bubble (Fig. 1c). We observed that N6-mA was more efficiently removed when placed at the 2nd and 3rd positions from the 5′-end. We hypothesized that N6-mA at these positions is more flexible to flip out for demethylation. In this respect, we designed a 3-nt bulged DNA, which forces nucleotides in the bulge to flip out. Notably, the middle N6-mA in the 3-nt bulge was efficiently demethylated compared to that in the 3-nt bubble (97% vs. 18% demethylation fraction, Fig. 1d), and the 1st and 3rd N6-mAs, which are less flexible than the middle one, were shown to be more resistant to demethylation. These results stressed that the flipping-out of N6-mA is crucial to promote its demethylation by ALKB1.

In parallel, we also examined N6-mA demethylation of ALKB1 towards other reported secondary structures of our genome.25,26 N6-mAs deposited in the flexible region of hairpin (DNA stem loop), D-loop and R-loop were efficiently demethylated as in bubble and bulge, while those in cruciform and replication fork, which are locally more restricted, were less demethylated (Fig. 1e). In support, we generated three additional hairpins by placing the preparation of our manuscript, the existence of various degrees; however, m1rA in bubbled RNA (RNA-RNA and RNA-DNA hybrid) as well as 1-methyldeoxynucleosine (1 mA) in ss-, ds-, and bubbled DNA were more resistant to demethylation (demethylation fraction < 20%) (Supplementary information, Fig. S2d). It is noteworthy that all potential substrates of ALKB1, as we have shown above, share a locally unpairing feature in which the modified nucleotide is prone to flipping out for demethylation.

The subcellular localization of ALKB1 is also under debate. Although in certain cell lines, such as HeLa cells27 and HEK293 cells,28 ALKB1 was suggested to localize and function in mitochondria. Studies in mouse trophoblasts, human and mouse ESCs,29 human mesenchymal stem cells and human cancers, all demonstrated its localization in nucleus and functions. Consistent with our previous works,29 we showed that ALKB1 is located in nucleus in mESC, and its localization was not influenced with fusion tag switched from N- to C-terminal (Supplementary information, Fig. S2e). Thus, we focus on ALKB1 function as genomic DNA N6-mA demethylase in the current study.

Detection of N6-hydroxymethyladenosine (N6-hmA) during N6-mA demethylation by ALKB1 To investigate the demethylation process of ALKB1, we monitored the reaction with online LC-MS/MS to detect N6-hmA, an expected intermediate of N6-mA oxidized by Fe(II)/α-KG-dependent dioxygenase. First, we analyzed the demethylation products in the form of digested nucleosides as we did above (Methods). A peak corresponding to N6-hmA (m/z = 282.12023, with mass error 0.3 ppm) was detected 10 min after reaction started, which was not found when the N6-mA substrate was incubated with the heat-inactivated ALKB1 (Control). The quantity of N6-hmA substantially accumulated as reaction proceeded (30 min), while the peak corresponding to N6-mA (m/z = 266.12477) gradually decreased (Fig. 2a). In support, the MS/MS spectrum of potential N6-hmA peak (30 min) was acquired, and the putative fragments were assigned, confirming the presence of N6-hmA (Fig. 2b). Unlike the FTO-mediated m6A oxidation,29 no confident peak corresponding to N6-formyladenosine (N6-FA) was observed in our study of ALKB1. The intermediate was further confirmed based on the intact oligo (undigested) analysis using UHPLC-Orbitrap MS (Methods). Consistently, the peak corresponding to N6-hmA-oligo, the expected oxidation product of N6-mA-oligo, was observed at the endpoint of the reaction, while no comparable N6-FA-oligo signal was identified (Fig. 2c). Remarkably, the N6-hmA-oligo signal dropped to background level after overnight incubation at 10°C following the stop of enzymatic reaction at 30 min, suggesting spontaneous decomposition of N6-hmA to A independent of ALKB1 catalysis under the reaction condition (pH 8.0) (Supplementary information, Fig. S3). During the preparation of our manuscript, the existence of N6-hmA in mammalian genomes as well as its accumulation in lung carcinoma tissues was reported and ALKB1 was shown to convert N6-mA to N6-hmA in vivo.30 With these intriguing results, further studies are needed to delineate the biological function and regulatory roles of N6-hmA generated by ALKB1.

Overall structure of ALKB1 Mouse ALKB1 protein has 389 residues and consists of an N-terminal extension (NTE), a nucleotide-recogniton lid (NRL), and a C-terminal double-stranded β helix (DSBH) domain (Fig. 3a). To elucidate the molecular mechanisms underlying ALKB1 function, we determined the 2.5 Å crystal structure of mouse ALKB1 (residues 1–359, ALKB1_1–359) bound to Mn(II) and N-oxalylglycine (NOG), the substitutes for the reactive Fe(II) and α-KG (Fig. 3b, c, data collection and refinement statistics summarized in Supplementary information, Table S2). Residues 20–358 were modeled based on the electron densities. The

SPRINGER NATURE

Cell Research (2020) 30:197 – 210
DSBH catalytic core of ALKBH1 adopts a typical eight-stranded (I–VIII) jelly-roll fold that contains a major sheet (βI, βVII, βIII, βVI) and a minor sheet (βVII, βV, βV and βII) to sandwich Mn(II) and NOG at the active center. The central jelly-roll fold is notably wrapped around and stabilized by additional elements from NTE, NRL and DSBH (Fig. 3d, e). In particular, an N-terminal α1 helix of NTE buttresses the minor sheet from the bottom; an extended "coil-sheet" structure of NTE extends the β(V-VI) edge from the back; α3 of NTE and α7 of DSBH organize a helical cluster involving elements α2 and α8, and stack against the major sheet from the top. Moreover, the NRL subdomain composed of elements Flip1 and Flip2 attaches to the β(I) edge of the major sheet, and functionally constitutes the substrate recognition surface along with α1 helix (Supplementary information, Fig. S4).

An extra "α1 helix" stabilizes the active center of ALKBH1. Although ALKBH1 is much longer than its bacterial homolog AlkB (389 vs. 216 residues, Supplementary information, Fig. S5), the catalytic core of ALKBH1 superimposed well with bacterial AlkB,31,32 with a root mean square deviation (r.m.s.d) of 0.759 Å over 105 Ca atoms. In contrast, three notably unique structural features to ALKBH1 arise outside the core structure: a far N-terminal α1 helix packed aside the catalytic center, the stretch-out of an extended Flip1, and a long insertion between β(VII) and β(VIII) (INS) that stabilizes Flip1 through helical cluster formation (Fig. 4a). Such features hold unique for ALKBH1 when comparing with all structures of its human paralogues determined so far, including ALKBH2, 3, 5, 7, 8 and FTO 32–35,37,38,39 (Supplementary information, Fig. S6).

The extra α1 helix forms extensive hydrophobic contacts with the minor sheet involving residues F23, L26, F27, Y30 of α1, I230 of β(II), P290 of β(VII), and I253, L255 of β(IV), and sequence alignment shows that these residues are highly conserved in vertebrates (Fig. 4b, f). The α1 helix is critical to maintain the structural integrity of the active center; in particular, F23 and F27 directly stabilize the minor β sheet through extensive hydrophobic contacts. In support, the structure of an N-truncated ALKBH1 (residue 37–369, ALKBH1Δα1) determined at 2.3 Å by the Se-SAD method (Supplementary information, Table S2) demonstrates that missing of α1 helix (resides 22–32) results in distortion of the cofactor coordination geometry as reflected by flipping away of H287 and disordering of β(II) that comprises H231 and D233 (Fig. 4c). Neither Mn(II) nor NOG was observed in the crystal (Supplementary information, Fig. S7a), and the impaired cofactor coordination of ALKBH1Δα1 in solution was also confirmed by calorimetric titration (Fig. 4d). Furthermore, enzymatic assays showed that ALKBH1Δα1 was totally inactive in N6-mA demethylation (Fig. 4e). Although an α1-truncated human ALKBH1 protein was reported to have demethylation activities towards m1rA in tRNA,8 we performed in vitro enzymatic assay using the α1-truncated human ALKBH1 and hardly detected any activities towards m1rA in tRNA stem loop nor N6-mA in DNA bubble (Supplementary information, Fig. S7b, c), further underlying the importance of α1 helix.

A "stretch-out" Flip1 and "α1" form a unique substrate channel of ALKBH1. The NRL subdomain, further composed of two sections named Flip1 and Flip2,39 is a unique feature of the AlkB family dioxygenases.31,37 Unlike the highly conserved catalytic core, the NRL subdomain is structurally variable among AlkB members and has been shown to play a critical role in substrate interaction.31–33,35,37,39,41 In most ALKBH1 homologs,31–33,37,41,42 the Flip1 motif adopts a bending conformation covering the active center (Supplementary information, Fig. S6), and forms a positively charged groove for nucleic acid binding (Fig. 4c, right). In the case of AlkB and ALKBH2, Flip1 has been shown to facilitate the recognition of modified base as well as

**Fig. 1** ALKBH1 prefers locally unpaired DNAs as substrates, but not ss-/ds-DNAs. a–e LC-MS/MS analyses of nucleoside hydrolytes for in vitro N°-mA demethylation assays using purified mouse ALKBH1 in full length and various DNA substrates. a 41-nt ssDNA (41s), 41-bp dsDNA (41d), and 41-bp dsDNA with a 6-nt mismatched region in the middle to mimic a bubble (41b6); b 41-bp bubbled DNAs, varied in the bubble size from 1 to 16 nt; c 41-bp bubbled DNAs with N°-mA in different position of a 6-nt bubble (pos1-6); d N°-mA in the middle of a 3-nt bubble (bubble_3) and different position of a 3-nt bulge (bulge_3A1, 3A2, and 3A3); e DNAs with different secondary structures. Substrates were shown as cartoon and N°-mA was shown as a green dot. Relative amounts of N°-mA were calculated for each measurement according to standard curves of N°-mA in nucleoside. Demethylation results were reported as residual N°-mA compared to its blank control (sample without ALKBH1). *, **, ***, and ns indicate p < 0.05, 0.01, and 0.001 and ≥ 0.05, respectively, t-test; error bars, ± SD of three biological replicates.
base flipping for demethylation.32,42 In contrast, Flip1 of ALKBH1 assumes a fully extended conformation (Fig. 5a), which is stabilized by extensive hydrophobic and hydrogen-bonding interactions through helical cluster formation among α3 and α5' "INS" motif, a non-conserved long insertion between β(VII) and β(VIII) unique to ALKBH1 (Fig. 5b; Supplementary information, Fig. S6). Conceivably, the stretch-out of Flip1 likely leads to the lack of autonomous base-flipping activity of ALKBH1, thus accounting for the observed incompetence of dsDNA demethylation by ALKBH1. By contrast, the locally unpaired DNA bubble can well serve as substrate for ALKBH1 since base-flipping has been primed for direct recognition and catalysis.

On the other hand, considerable lysine and arginine residues are enriched within Flip1, which constitute one extended and electrostatically positive wing around the catalytic cleft (Fig. 5a, c). In parallel, an additional string of basic residues was found around α1 helix and Flip2, forming the other positive wing of the substrate channel (Fig. 5a, c). The dimension of the substrate channel is ~66 Å in length and ~30 Å in diameter (Fig. 5c, left and middle), which is well-positioned for duplex substrate recognition of ~20 bp in length (contour length = 68 Å for B-form DNA). By contrast, AlkB has a relatively narrow binding groove of ~13 Å in diameter (Fig. 5c, right), consistent with its preference for ssDNA substrates.32,42

We next performed mutagenesis studies to explore the importance of the basic residues around the substrate channel. Alanine mutation of basic residues of Flip1 region (K158A/R159A/R160A, R167A/R169A) dramatically abolished N6-mA demethylation activity of ALKBH1 (Fig. 5d), and electrophoretic mobility shift assays further demonstrated impaired DNA binding of these mutants (Supplementary information, Fig. S7d). As well, the basic residues of the α1 and Flip2 regions were similarly verified to be indispensable for substrate binding and enzymatic activities (Fig. 5d; Supplementary information, Fig. S7d). Hence, the above data underscore a role of those basic residues for nucleic acid substrate recognition. In the meantime, the K/R-rich substrate-
binding surface less likely supports the positively charged histones as substrates for ALKBH1.

Co-crystal structure of ALKBH1 in complex with bulged DNA

To further explain the preference of ALKBH1 for locally unpairing DNAs, we determined the 2.4 Å crystal structure of ALKBH1 bound to a 21-mer bulged DNA (data collection and refinement statistics summarized in Supplementary information, Table S2). To obtain a stable and homogenous complex sample for crystallization, we adopted a disulfide cross-linking strategy by introducing a cysteine residue (S235C) in the catalytic pocket and a disulfide-modified cytosine within the bulge (Fig. 6a and methods). The overall structure of DNA-bound and the free state ALKBH1 superimposed well (r.m.s.d Cα = 0.555 Å), except for conformational adjustments of several flexible loops of Flip2 and Flip1 regions (Fig. 6b). As expected, the structurally stabilized Flip1 element (Fig. 5b) remains in a "stretch-out" conformation even in the presence of DNA substrate, lending further supports to the lack of autonomous base-flipping capability of ALKBH1.

Consistent with our previous analysis, the 21-mer bulged DNA substrate is docked onto a wide and positively charged concave surface in between Flip1, Flip2 and α1 (Fig. 6c, left). The bulged DNA adopts a B-form structure with both strands contributing to enzyme-substrate recognition primarily through "basic-backbone" interactions. Notably, two arginine residues (R24 of α1 and R159 of Flip1) insert into DNA minor grooves from both ends and function as two fingers to hold the DNA substrate (Fig. 6c, right). Such a recognition mode likely helps to orientate the DNA substrate for proper insertion of the flipped base into the active pocket, which is supported by dramatic activity loss of the R24A and R159A mutants (Supplementary information, Fig. S7e). The observed interaction mode between ALKBH1 and bulged DNA underscores the need of both the flanking duplex/stem region and the flipped base for recognition and catalysis. In support, enzymatic kinetics analyses showed that ALKBH1 displayed higher catalytic efficiency towards bulged DNA (kcat/Km = 0.71 × 10³ M⁻¹ s⁻¹) than ssDNA (kcat/Km = 0.46 × 10³ M⁻¹ s⁻¹) (Fig. 6d). Interestingly, the calculated apparent Km values are comparable for both ssDNA and bulged DNA (2.79 vs. 3.18 μM). By contrast, the kcat value is higher for bulged DNA (1.27 × 10⁻³ vs. 2.27 × 10⁻³ s⁻¹), which suggests that the duplex regions of the bulged DNA substrate promote efficient enzymatic turnover likely by facilitating optimal alignment of the flipped N⁶-mA in the catalytic center.

Unlike ALKBH1, both AlkB and ALKBH2 showed comparable activities on ss-, ds-, and bubbled substrates owing to the existence of a bending Flip1 for base flipping and a relatively narrow basic groove for nucleic acid strand recognition (Supplementary information, Fig. S8). Collectively, our complex structural study demonstrated a distinct substrate recognition mode of ALKBH1 as compared to other AlkB family members, which ultimately determines the unique preference of ALKBH1 for base unpairing DNA.
Substrate-binding pocket of ALKBH1

Substrate-binding pocket of ALKBH1 family members is formed by residues from NRL subdomain and DSBH fold, in which a target base is inserted, with its alkyl group pointing to the active site. The substrate-binding pocket of ALKBH1, compared to those of its homologs, was shown to be more similar to that of AlkB: comprising residues from DSBH fold (βI-III, βVIII and a loop between βII-III) assume nearly identical conformation in both proteins. Major differences come from the NRL subdomain, with a shorter Flip2 from ALKBH1 showing a "back-drawing" conformation, and its "stretch-out" Flip1 resulting in a more exposed pocket (Figs. 4a and 7a).

At this stage, covalent crosslinking of the flipped base with ALKBH1 within the active pocket prevents us from in-depth analysis of methyl base recognition in a native state. We then turn to structure-based comparison and mutagenesis studies to further explore the methyl base-binding pocket of ALKBH1. Structural alignment of ALKBH1-359 with AlkB bound to its 1 mA substrate revealed potential residues involved in methyl N6-base interaction of ALKBH1 (Fig. 7a). In particular, Y177 and H231 of ALKBH1 overlap with W69 and H131 of AlkB, and likely play a similar role in sandwiching N6-mA for demethylation; S235 of ALKBH1 overlaps with D135 of AlkB, and is located next to the hydrogen-bonding edge of the modified base. In AlkB, D135 has been shown to modulate base selectivity, and D135S or D135A mutations disrupted enzyme activity towards 1 mA due to loss of favorable interactions. Sequence alignment of ALKBH1 orthologues revealed that S235 is conserved among vertebrates (Fig. 4f). Considering that the dimension of N6-mA is bulkier than 1 mA (Fig. 7b), a shorter side chain of S235 may endow additional interactions with the modified base. This suggests that the methyl group on N6 of mA may act as a triphosphate moiety, stabilizing the transition state of the demethylation reaction. Further studies are needed to confirm this hypothesis and to understand the role of S235 in methyl demethylation specificity.
space for substrate accommodation. In support, S235A mutation did not affect ALKBH1 activity; while S235D and S235E resulted in a dramatic activity loss likely due to introduced steric clashes, especially in the case of S235E (Fig. 7c). We also tested the enzymatic activities of Y177A, S181A, and K182A, and revealed that Y177A, K182A but not S181A displayed clear activity decrease, consistent with a role of Y177 and K182 but not S181 in direct substrate pocket formation (Fig. 7c).

Fig. 5 Substrate-binding surface analysis and mutagenesis studies. a Distribution of basic residues along α1, Flip1 and Flip2 regions of ALKBH1 (9-359). Lysine and arginine are shown in stick. Residues critical to enzyme activity, revealed by mutagenesis experiments, are colored in deep salmon; residues shown to be less indispensable are colored cyan. The catalytic center is indicated by a red star. Right panel, structural comparison of ALKBH1 (9-359) and AlkB-DNA (PDB: 3BI3) complex. α1 and Flip1 of ALKBH1 are colored cyan; DSBH of ALKBH1 is colored yellow. Flip 1 of AlkB is colored magenta; DNA of AlkB and other parts of both proteins are colored gray. b Hydrophobic and hydrogen-bonding interactions among Flip1, α3 (from NTE), and ins. Flip1, α3 and ins are colored green, cyan, and yellow, respectively. Residues involved in hydrophobic and hydrogen-bonding interaction are shown as stick. c Comparison of electrostatic potential surface of ALKBH1 and AlkB. Both surfaces are colored as a spectrum of the surface electrostatic potential ranging from blue (10 kT/e) to red (−10 kT/e). d In vitro demethylation assays of ALKBH1 mutants towards the 41b6 DNA substrate.

N6-mA co-localizes with genomic unpairing regions during TSC development

A recently developed single-stranded DNA sequencing approach based on permanganate and S1 nuclease mapping (ssDNA-seq) successfully detected ssDNA regions (not single-stranded DNA) as a common feature across genome of activated B cells.25 We validated the ssDNA-seq approach in mouse ESCs using a wild-type line (TT2) which has low N6-mA levels in the self-renewing state (Fig. 8a and Methods). Totally, 8134 ssDNA peaks (false discovery rate cutoff = 0.01) were identified in mouse ESCs (Supplementary information, Fig. S9a). Owing to the more stringent peak calling with an input control utilized in this study, this number is expectedly less than that identified in activated B cells. Still, the genomic distribution of the peaks was consistent with the previously published findings, with a majority of peaks within genes (56.3%, \( P = 4.9 \times 10^{-24} \)) and enrichment within 1 kb of transcription start sites (TSS; 1.7%, \( P = 2.4 \times 10^{-11} \)) (Supplementary information, Fig. S9a, b). As expected, ssDNA was present at actively transcribed pluripotency genes (e.g., Nanog and Scl2a3, Supplementary information, Fig. S9g).

We further interrogated whether N6-mA was present on the unpairing regions in mouse genome. As the genome-wide N6-mA level in mouse ESCs is too low (6–7 p.p.m),6 we turned to another early development system with upregulated N6-mA for co-localization analyses of N6-mA and ssDNA regions. Our most recent study observed a substantial upregulation of N6-mA level during the transition of mouse ESCs to trophoblast stem cells (TSCs) (in press). By applying ssDNA-seq to ESCs undergoing cell fate transition (named as “LV1” hereafter), we identified more than 67,000 ssDNA-enriched regions, a majority of which were within 1 kb of predicted SIDD regions (65.4%). Notably, ssDNA peaks in LV1 were detected primarily within distal intergenic regions (55.9%) while significantly depleted within genes (35.1%, \( P = 2.9 \times 10^{-320} \)) and near TSS (0.7%, \( P = 6.2 \times 10^{-74} \)) (Supplementary information, Fig. S9d, e), which is in contrast to the predominant gene-associated ssDNA regions in self-renewing ESCs (Supplementary information, Fig. S9a, b) and activated B cells.25 The genome-wide distribution of N6-mA in LV1 was interrogated in parallel using DNA immunoprecipitation by anti-N6-mA antibody followed by next generation sequencing (DIP-seq). Reads of DIP-seq (95.5%) were confidently mapped to the mouse genome rather than mitochondrial or genomic DNA from other species.
suggest the presence of N6-attachment region-S/MAR previously was known to play an important role in gene expression and chromatin structure regulation at the H19 imprinting locus. At the H19 locus, in particular, our identification of N6-attachment regions (Fig. 8b) showed ssDNA enrichment within N6-attachment peaks but not in adjacent or randomly shuffled regions (Fig. 8b). ssDNA region (referred to as scaffold/matrix attachment region-S/MAR previously) was known to play an important role in gene expression and chromatin structure regulation at the H19 imprinting locus.66 At the H19 locus, in particular, our identified ssDNA peaks were shown to overlap with the well-known S/MAR regions identified by previous studies.25,45 with N6-attachment also enriched in the region (Fig. 8c). Our results suggest the presence of N6-attachment at locally unpairing regions in physiological conditions, which is consistent with the substrate preference of its demethylase.

**DISCUSSION**

Early studies on AlkB family members focused on its function in damage repair due to lack of knowledge in the occurrence and regulation of physiological DNA/RNA methylations and their derivatives.5,47–50 Despite being the founding member of its family, the cellular function of ALKBH1 has long been a debate in the field given its broad substrate spectrum beyond damage repair reported in different studies.5,8 Recently, ALKBH1 was reported as a N6-attachment eraser of ssDNA substrate.6 Here, our biochemical profiling and structural analyses provide key evidence on novel features of ALKBH1 substrates, characterized by a locally unpairing structure that contains flipped N6-attachment with flanking duplex stems. The lack of autonomous base-flipping activity distinguishes ALKBH1 from other homologs such as ALKBH232 and TET251 that can directly demethylate dsDNA substrates. Our co-crystal structural studies demonstrated that the stretch-out conformation of a K/R-rich Flip1 creates a wide and basic concave surface best for duplex DNA engagement with the flipped methyl base inserted into the active center. The dimension of the substrate-binding surface of ALKBH1 well matches a B-form DNA of ~20 bp in length, which might serve as a physiological recognition unit centered on N6-attachment-mediated epigenetic regulation.

The catalytic pocket of ALKBH1 is lined with a small side chain residue of S235 that is orthologously conserved in vertebrates, which allows an accommodation of the bulkier N6-attachment base for catalysis, and thus plays a key role in modification base selectivity. In support, mutation of S235 to bulkier residues (Asp/Glu) disrupted N6-attachment demethylation activity. Complete demethylation of N6-attachment by Fe(II)/2-OG-dependent dioxygenases contains several catalytic steps.55 We were able to detect the existence of N6-hmA as an oxidized intermediate during N6-attachment demethylation by ALKBH1. Our detection of N6-hmA, together with the published findings,30 suggesting that N6-hmA is a potential epigenetic mark reminiscent of 5hmC, and ALKBH1 may function as a key modulator of N6-attachment and N6-hmA levels.

**Fig. 6** Structure of ALKBH1 in complex with a 21-mer bulged DNA. A diagram of the cross-linking strategy. A disulfide bond can be formed between Cys in the bulge and Cys235 in the active site of ALKBH1. B Structural alignment of ALKBH120–355 in complex with bulged DNA and ALKBH11–359. C Left, ALKBH1 in complex is shown as electrostatic potential surface in the front view, colored as a spectrum of its surface electrostatic potential ranging from blue (10 kT/e) to red (−10 kT/e). The bulged DNA is shown as sticks with the 2Fo−Fc omit map contoured at 2.0σ. Right, cartoon of ALKBH1 in complex with bulged DNA in bottom view. NTE, NRL, and DSBH are colored cyan, green, and yellow, respectively. The bulged DNA is colored wheat. R24 and R159 are shown as sticks and space-filling dots. D Michaelis–Menten plots of the steady-state kinetics of mouse ALKBH1-catalyzed N6-attachment demethylation in bulged and ssDNA. N6-attachment-containing bulged/ss DNAs at various concentration were incubated with 1.5 μM of mouse ALKBH1. Demethylation products were analyzed as undigested oligos (see methods) by UHPLC-Orbitrap MS. Relative product amounts were calculated according to standard curve of dA-DNA oligo. The initial velocity (V0) was calculated based on the da-DNA generated in the assumed linear interval of the first 2.5 min. Less than 20% of the substrate was consumed in all reactions. The kinetic parameters, Km and kcat were generated by GraphPad software with the Michaelis–Menten equation. Error bars, SD of biological triplicates.
It has not escaped our attention that, in addition to N6-mA of unpairing DNA, ALKBH1 displayed in vitro activities towards m1rA, m6rA, m5rC to various degrees as long as a locally unpairing DNA, ALKBH1 displayed in vitro activities towards m1rA, alkylation position, and alkyl group.31,43,44 Comparing to its ability to repair chemically diverse substrates different in base, observed for other AlkB family members. For example, structural and biochemical studies have established that bacterial AlkB is able to repair chemically diverse substrates different in base, alkylation position, and alkyl group.31,43,44 Comparing to its relatively broad base selectivity, ALKBH1 displays a high dependence on the secondary structure of nucleic acid substrates. Therefore, the biological function of ALKBH1 is likely regulated by its cellular localization and genomic targeting in cells. The reports of such additional activities to date are limited to certain cell lines, which have not been observed in developments or human diseases. Future studies will determine the physiological significance of those observations. Nonetheless, we observed dominant nuclear localization of ALKBH1 in mESCs, which, in concert with the reported cellular N6-mA demethylation activity,6,16,17 supports an important function of ALKBH1 in DNA N6-mA biology.

In living cells, the prevailing DNA conformation is the right-handed, double-stranded B-DNA. However, studies in the past decades have shown that our genome DNA is a dynamic molecule and is able to adopt various non-B structures, including melted bubbles in SIDD regions and DNA hairpins that share a locally unpairing feature.5,35,36,52-54 Formation of non-B DNA is influenced by either base sequence per se or biological processes that introduce local supercoiling. In turn, non-B DNAs usually play active roles in chromosome structure organization or gene regulation. For example, DNA bubbling is often coupled to topological stress occurring at locally AT-rich regions including S/MARs that are enriched with SIDD.45,54,56 In addition, nucleosome formation usually stabilizes negative supercoil with a superhelical density (σ) of about −0.07.57 Conceivably, in the processes like transcription or replication, nucleosome disassembly might facilitate ALKBH1 function by trigging local base unpairing. Here, we revealed co-localization of N6-mA and putative non-B DNA sites (ssDNA regions) in mouse genome. Our work provides an intriguing model for studying N6-mA demethylation on unpairing substrates in physiological environments, and further suggests that ALKBH1 is an important DNA demethylase associated with dynamic chromosome regulation.

**MATERIALS AND METHODS**

**Protein expression and purification**

The gene fragment encoding residues 1–389 (full length), 1–359 (ALKBH11–359), 20–355 (ALKBH120–355), and 37–369 (ALKBH137–369) of mouse ALKBH1 were inserted into a modified pRSFduet-1 vector, respectively. Point mutations were generated by a site-directed mutagenesis kit (Stratagene). Proteins with amino (N)-terminal His-tag were expressed in *Escherichia coli* strain BL21 (DE3) (Novagen) in the presence of 0.2 mM isopropyl-β-D-thiogalactopyranoside. Cells after overnight induction were collected by centrifugation and re-suspended in the lysis buffer: 20 mM Tris-HCl, pH 8.0, 100 mM NaCl. After cell lysis by an EmulsiFlex C3 (Avestin) homogenizer, the centrifugation-cleared supernatant was applied to a HisTrap nickel column (GE Healthcare), and the bound protein was eluted with a linear imidazole gradient from 20 mM to 500 mM. The eluents were subjected to His-tag removal cleaved by PreScission protease, followed by ion-exchange chromatography on HiTrap QHP and HiTrap Heparin HP columns (GE Healthcare). Finally, the protein was purified to homogeneity through size exclusion chromatography on a Superdex 200 increase 10/300 column (GE Healthcare), and concentrated to 15 mg/mL in a buffer containing 20 mM Tris-HCl, pH 8.0, 100 mM NaCl for future use. Proteins used for enzymatic assays were purified in essentially the same procedures as described above. Of note, the full-length wild-type and mutated constructs of ALKBH1 do not contain PreScission protease cleavage site and the N-terminal His-tag was retained. Selenomethionine (SeMet)-labeled protein (ALKBH1137–369) was
expressed in the methionine-auxotrophic strain B834(DE3) and purified using the same strategy as the native protein.

Cross-linked ALKBH1-DNA sample preparation
To solve the crystal structure of ALKBH1-DNA complex, a slightly truncated frame of ALKBH1–ALKBH120–355 was chosen for cross-linking and crystallization trials based on activity profiling and structural analysis. To generate the cross-linked ALKBH120–355-DNA complex, the S235C mutation was introduced to the catalytic center for disulfide cross-linking, and other four C to S mutations (C300S, C304S, C129S, and C322S) were introduced to avoid non-specific cross-linking. The mutant protein was incubated with synthetic oligonucleotides (Supplementary information, Table S1) at 4 °C for 10 h, and further purified by Mono-Q anion exchange chromatography and size exclusion chromatography on a Superdex 200 increase 10/300 column (GE Healthcare).

Crystallization, data collection and structure determination
Crystallization was performed via the sitting or hanging-drop vapor diffusion method under 4 °C by mixing equal volumes (0.2–1.0 µL) of the protein sample and the reservoir solution. Prior to crystallization, proteins were mixed with 2.3 mM MnCl2 and 4.3 mM NOG/succinic acid at a concentration of 10 mg/mL. SeMet-labeled ALKBH137–369 crystals were grown in the solution containing 16% (w/v) PEG3350, 0.13 M ammonium citrate dibasic; native ALKBH11–359 crystals were grown in the solution containing 10% (w/v) PEG6000, 0.1 M BICINE-HCl, pH 7.8; ALKBH120–355-DNA complex crystals were grown in the solution containing 0.2 M NaCl, 0.1 M phosphate citrate, pH 4.2, 10% PEG3000. Crystals were briefly soaked in a cryo-protectant composed of reservoir solution supplemented with 20% glycerol, and then flash frozen in liquid nitrogen for data collection at 100 K. The data of SeMet-labeled ALKBH137–369 was collected at wavelength of 0.9785 Å at Shanghai Synchrotron Radiation Facility beamline BL19U, and the data of native ALKBH11–359 and ALKBH120–355-DNA complex were collected at wavelength of 0.9792/0.9791 Å at Shanghai Synchrotron Radiation Facility beamline BL17U. Data were indexed, integrated and merged using the HKL2000 software package. Data collection statistics are shown in Supplementary information, Table S2.
The phase of ALKBH1, a single-wavelength anomalous dispersion method using PHENIX. The ALKBH1 and the ALKBH1-DNA complex structures were solved by molecular replacement using MOLREP. All structures were refined using PHENIX with iterative manual model building using COOT. Model quality was analyzed with PROCHECK. Detailed structural refinement statistics are summarized in Supplementary information, Table S2.

Preparation of nucleic acid substrates
All modified DNA/RNA oligos were purchased from Gene Link and the unmodified DNA/RNA oligos were purchased from GenScript. The ds-, ss-, bubbled DNA/RNA and stem-loop DNA/RNA were annealed at 25 μM in the annealing buffer (50 mM KCl, 4 mM MgCl₂, 50 mM Tris-HCl, pH 8.0) using a thermocycler, during which the oligos were heated up to 95 °C for 2 min and gradually cooled to 25 °C over an hour. R-loop, D-loop, and replication fork were annealed according to the published method with minor modifications. Cruciform was generated by separately annealing the stem loop-containing forward and reversed strands as stated above, and then annealing both strands in equal molar ratio from 60 °C to 30 °C. All DNA and RNA substrates used in this study are summarized in Supplementary information, Table S1.

In vitro demethylation assay
Demethylation assays were performed in 20 μL volume, which contains 30 pmol of annealed oligos and 4.08 μg of recombinant mouse ALKBH1 protein or mutant. The reaction buffer consisted of 50 mM KCl, 1 mM MgCl₂, 50 mM Tris-HCl, pH 8.0, 2 mM ascorbic acid, 1 mM α-ketoglutarate and 100 μM (NH₄)₂Fe(SO₄)₂·6H₂O. For demethylation assay of human ALKBH1, the same condition was applied except that MOPS buffer at pH 6.6 was used, which is adapted from previous report. Reactions were performed at 37 °C for 90 min, and stopped by adding EDTA to a final concentration of 5 mM (2.5 mM for reactions containing RNA). Then the products were subjected to dot blotting or digested for LC-MS/MS detection.

Dot blotting
DNA samples were denatured at 95 °C for 10 min, snap-cooled on ice and neutralized with 10% volume of 6.6 M ammonium acetate. Samples (2 μL) were spotted on the membrane (Amersham Hybond-N+, GE), air-dried and then UV-cross-linked (2× auto-crosslink, 1800 UV Stratalinker, STRATAGENE). Membranes were blocked in blocking buffer (5% milk, PBS-T) for 2 h at room temperature, and then incubated with N⁺-m mA antibodies (200–003, Synaptic Systems, 1:1000) overnight at 4 °C. After 5-time washes, membranes were incubated with HRP-linked secondary anti-rabbit IgG antibody (ZB-2301, ZSG-BIO, 1:5000) for 1 h at room temperature. Signals were detected with Immun-Star™ WesternC Chemiluminescence (Bio-Rad).

LC-MS/MS detection of demethylation products as undigested DNA oligos
QExactive mass spectrometer (Thermo Fisher, CA), equipped with a heated electrospray ionization (HESI) probe was used in positive ion mode. Nucleosides were separated by a Luna omega P5 C18 column (2.1 × 100 mm, 1.6 μm, Waters). A binary solvent system was used, in which mobile phase A consisted of 2 mM ammonium acetate and 100% aqueous, and mobile phase B of 100% acetonitrile. A 10-min gradient with flow rate of 200 μL/min was used as follows: 0–2 min at 1% B; 2–4 min, 1%–30% B; 4–6 min, 30%–98% B; 6–7.1 min, 98% B; 7.1–10 min, 1% B. Column chamber and sample tray were held at 35 °C and 10 °C, respectively. Data acquired in data-dependent MSMS acquisition mode. The MS and MSMS scans were collected with resolution of 70,000 and 17,500, respectively. The Source parameters are as follows: spray voltage: 3000 V; ion transfer tube temperature: 320 °C; vaporizer temperature: 300 °C; sheath gas flow rate: 35 Arb; auxiliary gas flow rate: 10 Arb. Data analysis and quantitation were performed by the software Xcalibur 3.0.63 (Thermo Fisher, CA).

Isothermal titration calorimetry
For ITC measurement, MnCl₂, NOG and the proteins were prepared in the buffer containing 20 mM Tris-HCl, pH 7.5, and 100 mM NaCl. The titration was performed using a MicroCal iTC200 system (GE Healthcare) at 25 °C. Each ITC titration consisted of 17 successive injections with 0.4 μL for the first and 2.4 μL for the rest. MnCl₂ at 1.0 mM was titrated into proteins at 0.1 mM; and NOG at 1.0 mM was titrated into proteins titrated with MnCl₂ successively. The blocked ITC curves were processed using Origin 7.0 software (OriginLab) according to the “One Set of Sites” fitting model.

Circular dichroism spectroscopy
Purified proteins were diluted to 0.1 mg/mL in 1× PBS buffer. Circular dichroism spectra were recorded under room temperature using an Applied Photo-physics Chirascan plus spectropolarimeter with a 1 nm path-length cell and a bandwidth of 1 nm. Spectra were scanned from 190–260 nm with a step size of 1 nm and were repeated for three times. Each reported circular dichroism curve was averaged with three scans after the subtraction of buffer control and smoothed.

Electrophoretic mobility shift assay
The DNA substrate (41b6, 1 μM) was incubated with increasing amount of ALKBH1 proteins (0, 1, 5, 10, 15 μM) in buffer containing 20 mM Tris-HCl, pH 7.5, 100 mM NaCl, 0.1 mM MnCl₂, and 1 mM NOG at 4 °C for 30 min. The samples were subjected to 5% native polyacrylamide gel and run at 100 V in 0.5× TBE buffer.

Immunofluorescence
The stable ES cells carrying pLV-Flag-HA-Alkbh1 and pLV-Alkbh1-HA constructs were grown on gelatin-treated slides (Thermo, PEZGS0816) for 24 h. Cells were fixed with 1% PFA and
permeabilized with 0.2% triton X-100 in 1 × phosphate buffered saline (PBS). Then the cells were incubated in the blocking buffer (2% BSA, 0.2% triton X-100 in PBS) for 1 h, incubated with 1:500 HA antibody (Cell signaling, 3724 S) in blocking buffer at 4 °C overnight, and then incubated with the secondary antibody (Thermo, A11004). DNA was stained with DAPI for 5 min. Slides were mounted with mounting gel (Electron Microscopy Sciences). Images were acquired with Leica SP5 confocal laser microscope.

Single-stranded DNA-sequencing (ssDNA-seq) ssDNA-seq was performed as the previously described protocol with minor modifications.22 ES cells (8 × 10³) cultured in feeder-free conditions on gelatin-coated tissue culture plates were washed with 1× PBS buffer at 37 °C. Low salt buffer (15 mM Tris-HCl, pH 7.5, 60 mM KCl, 15 mM NaCl, 5 mM MgCl₂, 0.5 mM EGTA, and 300 mM sucrose) at 37 °C was added for 5 min. Cells were then treated with 100 mM KMnO₄ for 80 s at 37 °C (and 300 mM sucrose) at 37 °C was added for 5 min. Cells were then treated with 100 mM KMnO₄ for 80 s at 37 °C ("Treated" sample), and the reaction was quenched by the addition of 50 mM EDTA, 700 mM β-mercaptoethanol, and 1% (w/v) SDS. In parallel, the same number of cells were treated with water ("Blank" sample) and processed similarly. Lysates were incubated with 200 μg/mL of proteinase K (Qiagen, 19133) overnight at 37 °C. DNA was extracted twice with phenol. DNA was extracted a final time with phenol:chloroform:isoamyl alcohol (25:24:1, v/v/v; PCI), precipitated with 2 M ammonium acetate in ethanol, and washed with 70% ethanol (hereafter called PCI extraction with ethanol precipitation) and resuspended in 1 mL of 10 mM Tris-HCl, pH 8.0, and 1 mM EDTA buffer (TE buffer).

Genomic DNA was treated with an RNase A/T1 mix (Thermo EN0551; 1:50 final concentration, 40 μg/mL and 100 U/mL, respectively) for 1 h at 37 °C, then PCI extracted with ethanol precipitation and resuspended in 1 mL of TE buffer. Free 3′ ends formed due to DNA breakage during sample preparation were blocked by treatment with 100 μM corydcepin-5'-triphosphate sodium salt (Sigma Aldrich, C9137) and 400 U of Terminal transferase (NEB, M0315; TdT) in 1× TdT buffer in a reaction volume of 3 mL for 2 h at 37 °C. DNA was PCI extracted with ethanol precipitation and resuspended in 1 mL TE buffer.

Digestion of ssDNA was carried out by dividing each of the Treated and Blank samples equally into four micro-tubes with 0, 50, 100, or 200 U of S1 nuclease (ThermoFisher, EN0321) in 500 μL of the supplied reaction buffer and incubating for 30 min at 37 °C. DNA was PCI extracted with ethanol precipitation and resuspended in 1 mL of TE buffer. Free 3′ ends formed due to DNA breakage during sample preparation were blocked by treatment with 100 μM corydcepin-5'-triphosphate sodium salt (Sigma Aldrich, C9137) and 400 U of Terminal transferase (NEB, M0315; TdT) in 1× TdT buffer in a reaction volume of 3 mL for 2 h at 37 °C. DNA was PCI extracted with ethanol precipitation and resuspended in 1 mL TE buffer.

Quantification of ssDNA enrichment in transposable elements Raw reads in fastq format were analyzed by SalmonTE (version 0.8.2) with default parameters and fold enrichment of samples over input was calculated for each family of transposable elements in the mouse index.65 LINE1 subfamilies were assigned to bins of young, middle, and old age as performed previously.65

Mouse ES cell culture Mouse tetracycline-off Cdx2 ES cells (C-ES) and TT2 ES cells were cultured on gelatin-coated tissue culture plates with recombinant LIF. ES cells were grown in DMEM supplemented with 15% fetal bovine serum, 1% non-essential amino acids, 1 mM sodium pyruvate, 2 mM l-glutamine, 1,000 units of mLIF (EMD Millipore), 0.1 mM β-mercaptoethanol (Sigma Aldrich) and antibiotics. Additionally, media for C-ES cells was supplemented with 0.5 μg/mL doxycycline to turn off Cdx2 expression. Cells were passaged at 70%–80% confluence using 0.05% trypsin-EDTA.

Antibodies and reagents for western blot and immunofluorescence HA-Tag Rabbit mAb; CST; 3724; Clone C29F4; Lot 8; IF/WB; 1:1500.

Library preparation for high-throughput sequencing “Treated” samples and their input controls were further processed for high-throughput sequencing. Sequencing libraries were made using the NEBNext Ultra II DNA Library Prep kit with 10–100 ng of each sample and input. Libraries were prepared according to the manufacturer’s instructions without size selection. Library concentrations were measured with Qubit and quality control performed with Agilent 2100 Bioanalyzer. Libraries were pooled and sequenced by paired-end 2 × 100 bp in one lane of an Illumina HiSeq4000.

ACKNOWLEDGEMENTS

We thank Prof. Z.X. Wang for his valuable advices on enzyme kinetic studies. We thank the staff members at beamline BL17U of the Shanghai Synchrotron Radiation
Facility and Dr S. Fan at Tsinghua Center for Structural Biology for their assistance in data collection, and Metabolomics Center in the China National Center for Protein Sciences Beijing for providing facility support. This work was supported by grants from the National Natural Science Foundation of China (91753203 and 31725014 to H.L.) and the National Key R&D Program (2016YFA0500700 to H.L.) and the NIH Medical Scientist Training Program Training Grant (T32GM007205 to R.N.).

AUTHOR CONTRIBUTIONS
H.L. and A.X. conceived and supervised the study; M.Z., S.Y. designed and performed biochemical and structural studies with assistance from W.Z., G.L.; R.N. performed cellular and genomic profiling studies with help from Z.L., T.W.; K.L. helped with mass spectrometry support; H.L. and A.X. wrote the manuscript.

ADDITIONAL INFORMATION
Supplementary information accompanies this paper at https://doi.org/10.1038/s41422-019-0237-5.

Competing interests: The authors declare no competing interests.

REFERENCES
1. Koonin, L. A. A. E. V. The DNA-repair protein AlkB, EGL-9, and leperecan define new families of 2-oxogluurate- and iron-dependent glycosylases. Genome Biol. 2, research0007.1–research007.8 (2001).
2. Kurowiski, M. A., Bhagwat, A. S., Papaj, G. & Bujnicki, J. M. Phylogenetic identification of five new human homologs of the DNA repair enzyme AlkB. BMC Genom. 4, 48 (2003).
3. Johansson, C. et al. The roles of Jumonji-type oxygenases in human disease. Epigenomics 6, 89–120 (2014).
4. Wei, Y. F., Carter, K. C., Wang, R.-P. & Shell, B. K. Molecular cloning and functional analysis of a human cDNA encoding an Escherichia coli AlkB homolog, a protein involved in DNA alkylation damage repair. Nucleic Acids Res. 24, 931–937 (1996).
5. Westby, M. P. et al. Human AlkB homolog 1 is a mitochondrial protein that demethylates 3-methylcytosine in DNA and RNA. J. Biol. Chem. 283, 25046–25056 (2008).
6. Wu, T. P. et al. DNA methylation on N(6)-adenine in mammalian embryonic stem cells. Nature 532, 329–333 (2016).
7. Haag, S. et al. NUSN3 and ABH1 modify the wobble position of mt-tRNAMet to expand codon recognition in mitochondrial translation. EMBO J. 35, 2104–2119 (2016).
8. Liu, F. et al. ALKBH1-mediated RNA demethylation regulates translation. Cell 167, 816–828 e816 (2016).
9. Ougland, R. et al. ALKBH1 is a histone H2A dioxygenase involved in neural differentiation. Stem Cells 30, 2672–2682 (2012).
10. Muller, T. A., Meek, K. & Hausinger, R. P. Human AlkB homolog 1 (ABH1) exhibits DNA lyase activity at abasic sites. DNA Repair 9, 58–65 (2010).
11. Pan, Z. et al. Impaired placental trophoblast lineage differentiation in Alkbh1(-/-) mice. Dev. Dyn. 237, 316–327 (2008).
12. Nordstrand, L. M. et al. Mice lacking Alkbh1 display sex-ratio distortion and unilateral eye defects. PLoS ONE 5, e13827 (2010).
13. Muller, T. A., Yu, K., Hausinger, R. P. & Meek, K. ALKBH1 is dispensable for abasic site cleavage during base excision repair and class switch recombination. PLoS ONE 8, e67403 (2013).
14. Kozioł, M. J. et al. Identification of methylated deoxyadenosines in vertebrates reveals diversity in DNA modifications. Nat. Struct. Mol. Biol. 23, 24–30 (2016).
15. Liu, J. et al. Abundant DNA 6mA methylation during early embryogenesis of zebrafish and pig. Nat. Commun. 7, 13052 (2016).
16. Xiao, C.-L. et al. N6-methyladenine DNA modification in the human genome. Mol. Cell 71, 306–318 e307 (2018).
17. Xie, Q. et al. N6-methyladenine DNA modification in globostoma. Cell 175, 1228–1243 (2018).
18. Zhu, S. et al. Mapping and characterizing N6-methyladenine in eukaryotic genomes using single-molecule real-time sequencing. Genome Res. 28, 1067–1078 (2018).
19. Fu, Y. et al. N6-methyldeoxyadenosine marks active transcription start sites in Chlamydomonas. Cell 161, 879–892 (2015).
20. Greer, E. L. et al. DNA methylation on N6-adenine in c. elegans. Cell 161, 868–878 (2015).
21. Zhang, G. et al. N6-methyladenine DNA modification in Oroschophila. Cell 161, 893–906 (2015).
22. Kigar, S. L. et al. N(6)-methyladenine is an epigenetic marker of mammalian early life stress. Sci. Rep. 7, 18078 (2017).
53. Bacolla, A. & Wells, R. D. Non-B DNA conformations, genomic rearrangements, and human disease. *J. Biol. Chem.* **279**, 47411–47414 (2004).

54. Ghosh, R. P. et al. Satb1 integrates DNA binding site geometry and torsional stress to differentially target nucleosome-dense regions. *Nat. Commun.* **10**, 3221 (2019).

55. Terumi, K.-S. & Kohwi Y. Torsional stress stabilizes extended base unpairing in suppressor sites flanking immunoglobulin heavy chain enhancer. *Biochemistry* **29**, 9551–9560 (1990).

56. J Bode, Y. K. et al. Biological significance of unwinding capability of nuclear matrix-associating DNAs. *Science* **255**, 195–197 (1992).

57. Corless, S. & Gilbert, N. Effects of DNA supercoiling on chromatin architecture. *Biophys. Rev.* **8**, 245–258 (2016).

58. MINOR, Z. O. A. W. Processing of X-ray diffraction data collected in oscillation mode. * Methods Enzymol.* **276**, 307–326 (1997).

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

60. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr D Biol. Crystallogr.* **66**, 22–25 (2010).

61. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol. Crystallogr.* **60**, 2126–2132 (2004).

62. Laskowski, R. A. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283–291 (1993).

63. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

64. Zang, C. et al. A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics* **25**, 1952–1958 (2009).

65. Jeong, H.-H., Yalamanchili, H. K., Guo, C., Shulman, J. M. & Liu, Z. An ultra-fast and scalable quantification pipeline for transposable elements from next generation sequencing data. *Pac. Symp. Biocomput.* **2018**, 168–179 (2018).

66. Nishiyama, A. et al. Uncovering early response of gene regulatory networks in ESCs by systematic induction of transcription factors. *Cell Stem Cell* **5**, 420–433 (2009).