

# Oxidative intermediates captured during demethylation of DNA and RNA

Ying Wang<sup>1</sup>, and Jianyu Zhang<sup>1,\*</sup>

<sup>1</sup>School of Pharmaceutical Science and Technology, Tianjin University, No. 92 Weijin Road, Nankai District, Tianjin 300072, P.R. China

**Abstract.** DNA and RNA have various methylation modifications or damage that are directly related to some human diseases and physiological regulation. Most of these methylation modifications are reversible and can be dynamically repaired by RNA or DNA demethylases. Over the past few decades, enzymes from the ALKB and TET families have been shown to have the ability to demethylate nucleic acids, which involves intermediates in the oxidative repair process. These intermediates can be accurately captured by advanced methods such as HPLC, LC-MS, TLC, and crystallization, which can significantly promote our understanding of the dynamic mechanism of demethylation. In this review, we discuss recent research advances in this area and raise open questions and constructive opinions about the capture of nucleic acid demethylation intermediates.

## 1 Introduction

The presence of radiation, endogenous, or exogenous substances in the environment can cause nucleic acid alkylation. Among them, methylation is a common and widely existing modification in eukaryotic cells, and it is also the main epigenetic form of mammalian gene expression regulation. Methylation accounts for two-thirds of more than 100 chemical modifications of RNA[1]. Nucleic acid methylation comes in both modified and damaged forms[2]. Methylation damage, such as N<sup>1</sup>-methyladenine (m<sup>1</sup>A) and N<sup>3</sup>-methylcytosine (m<sup>3</sup>C), can interfere with Watson-Crick base pairing and be cytotoxic and mutagenic. In contrast, methylation modifications do not interfere with base pairing but exhibit important biological regulatory functions[3, 4]. For example, N<sup>6</sup>-methyladenine (m<sup>6</sup>A) is widely present in both DNA and RNA, and its formation involves the methyl group transfer of S-adenosylmethionine (SAM) under the catalysis of the methyltransferase complexes METTL3, METTL4, and WTAP[5-7]. In 2013, Fustin et al. showed that m<sup>6</sup>A methylation modification on mRNA could regulate the circadian cycle. Methylation-modified disorders are closely related to some human diseases, including cancer[8].

As shown in Fig. 1, DNA and RNA methylation has diversity and site preference, most commonly occurring on amine groups outside the loop. Among them, N<sup>6</sup>-methyladenosine (m<sup>6</sup>A) was discovered as early as 1970[9]. Subsequent research proved that the level of m<sup>6</sup>A in DNA regulates the chromosome replication and plays an indispensable role in the genetic process. Besides, as the most abundant methylation modification in mRNA, m<sup>6</sup>A is also the first reversible methylation modification

to be found, which affects the translation and transport of mRNA. 5-methylcytosine (m<sup>5</sup>C), as an epigenetic marker in DNA, is another typical methylation modification.

Abnormal methylation can lead to the accumulation of cytotoxicity and the generation of carcinogenic mutations. Methylation of DNA and RNA is a dynamic process that can be reversed under the catalysis of nucleic acid demethylases. *E. coli* AlkB, which relies on Fe (II) and  $\alpha$ -ketoglutarate ( $\alpha$ -KG), is a dioxygenase and catalyzes the demethylation of nucleic acid bases by oxidation. ALKB can repair the methylation of single-stranded DNA and double-stranded DNA, such as m<sup>1</sup>A and m<sup>4</sup>C, with a significant tendency to single-stranded DNA substrates[10]. ALKB has nine homologous protein members in humans and mammals: ALKBH 1 ~ 8 and FTO. Although the human ALKB family proteins are not similar to *E. coli* ALKB in sequence, their active catalytic centers are highly conserved[11-14]. In 2002, the research found that ALKB and its human homologs ALKBH2 and ALKBH3 can repair alkylation damage on DNA single strands by oxidizing methyl groups to aldehyde groups, revealing a new way to fix DNA methylation. FTO is one of the critical factors leading to human obesity by regulating energy metabolism balance, which is also known as ALKBH9. Jia et al. found that FTO can reverse m<sup>6</sup>A both in DNA and RNA successfully.

TET (ten-eleven translocation) protein family is also an  $\alpha$ -KG and Fe (II) dependent dioxygenase, which contains TET 1, TET 2, and TET 3. All of them can catalyze the conversion of m<sup>5</sup>C to 5-hydroxymethylcytosine (hm<sup>5</sup>C), which is an essential intermediate in the DNA demethylation process. After thymine-DNA glycosylase (TDG) recognition and excision, hm<sup>5</sup>C can be converted into C by base excision repair (BER) finally[15, 16]. The discovery of this

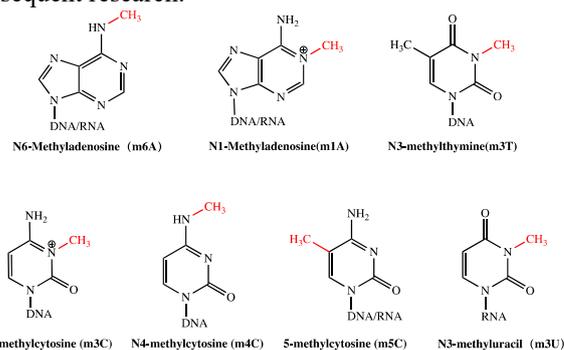
\*Corresponding author: Jianyu.Zhang@tju.edu.cn

intermediate has demonstrated a deeper understanding of the DNA demethylation mechanism. Similarly, the ALKB family forms intermediates while catalyzing the demethylation of DNA and RNA by oxidation, mostly in the form of methylation, such as hm<sup>6</sup>A. In addition, sometimes there are more than one intermediate during demethylation. When FTO repairs the m<sup>6</sup>A in RNA, a tandem oxidation mechanism is observed by capturing hm<sup>6</sup>A and f<sup>6</sup>A.

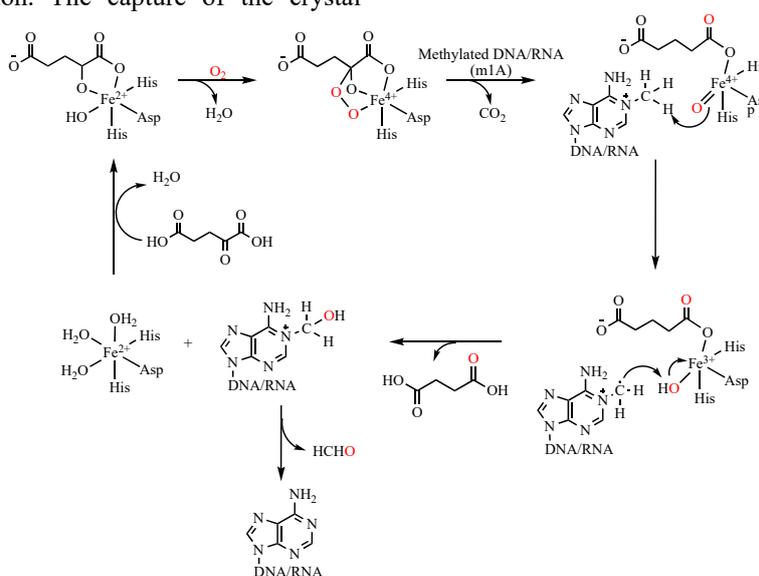
In addition to revealing the specific mechanism of demethylation, intermediates also have crucial physiological significance. The decrease of TET protein content and the decrease of hm<sup>5</sup>C is an essential manifestation of cell differentiation. The difference of hm<sup>5</sup>C content may have a vital regulatory role in determining the function of nerve cells. With the development of technology and the increasing attention of scholars, capturing intermediates has gradually become an area of concern when studying demethylases. Elementary thin-layer chromatography (TLC) can be used to identify the presence of intermediates. High performance liquid chromatography (HPLC) can isolate and identify substrates, products, and intermediates if we quench the reaction in time. Liquid chromatography-mass spectrometry (LC-MS) can accurately analyze the ionic structure of intermediates and provide the most direct identification information. The capture of the crystal

structure visually shows the three-dimensional structure information of the intermediate.

In the past few decades, a large number of researches have focused on revealing the dynamic process and specific mechanism of nucleic acid demethylation by capturing intermediates through different methods. From initial identification to accurate analysis, significant progress has been made. Summarizing the capture of intermediates related to the two major demethylase families, ALKB and TET, can help us grasp the oxidation mechanism at an overall level and provide guidance for subsequent research.



**Fig. 1.** Chemical structures and abbreviations of common DNA and RNA methylation lesions.



**Fig. 2.** The reaction mechanism of the ALKB family in the demethylation of nucleic acids, which involves two kinds of intermediates, Fe (IV) = O and hydroxymethylated bases.

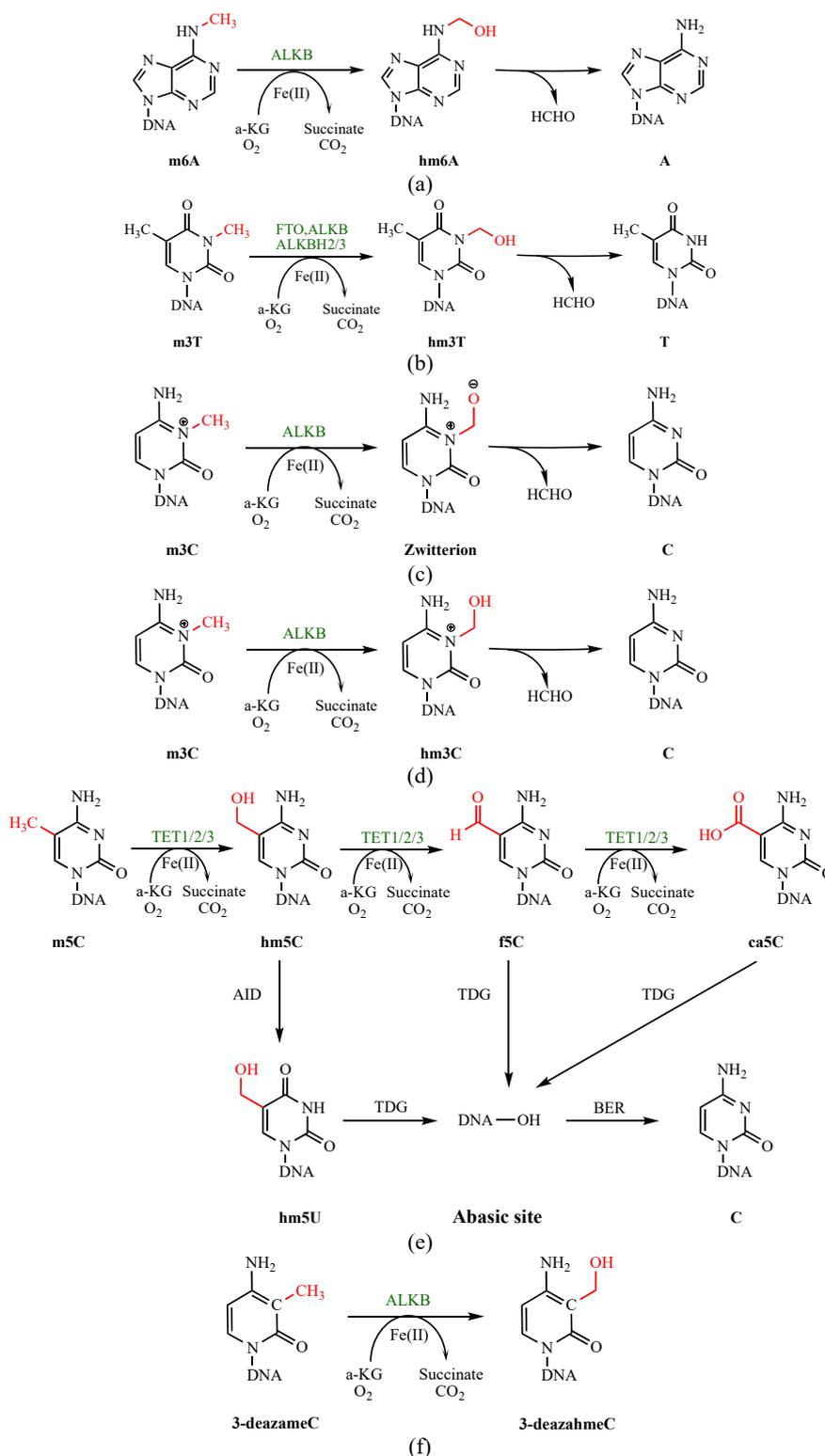
**Table 1.** DNA methylation lesions and intermediates during demethylation by ALKB family or TET family.

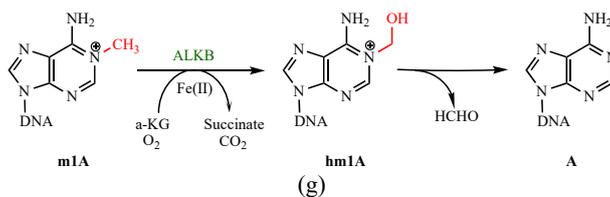
DNA Methylation	Demethylases	Intermediates	Capture
m <sup>1</sup> A	ALKB, ALKBH2, ALKBH3	hm <sup>1</sup> A	×
m <sup>6</sup> A	ALKB, ALKBH1, FTO	hm <sup>6</sup> A, f <sup>6</sup> A	√
m <sup>3</sup> T	FTO, ALKB, ALKBH2, ALKBH3	hm <sup>3</sup> T	√
m <sup>3</sup> C	ALKB	hm <sup>3</sup> C	√
m <sup>4</sup> C	ALKB	hm <sup>4</sup> C	×
m <sup>5</sup> C	TET1/2/3, ALKB, ALKBH2, ALKBH3	hm <sup>5</sup> C, f <sup>5</sup> C, ca <sup>5</sup> C	√

**Table 2.** RNA methylation lesions and intermediates during demethylation by the ALKB family.

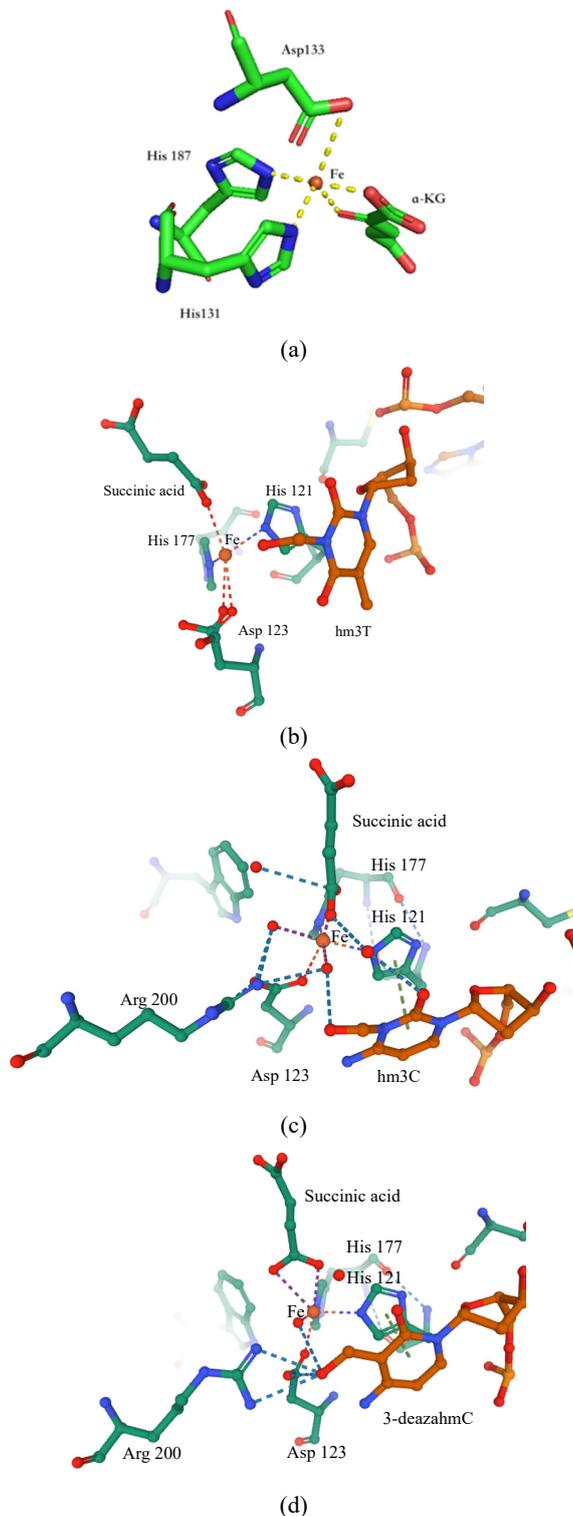
RNA Methylation	Demethylases	Intermediates	Capture
m <sup>1</sup> A	ALKB, ALKBH2, ALKBH3	hm <sup>1</sup> A	×
m <sup>6</sup> A	ALKBH5, FTO	hm <sup>6</sup> A, f <sup>6</sup> A	√
m <sup>3</sup> U	FTO	hm <sup>3</sup> U	×
m <sup>5</sup> C	ALKBH1	hm <sup>5</sup> C, f <sup>5</sup> C	√

(√ means the intermediate was captured, × means it has not been captured yet.)





**Fig. 3.** The reaction mechanism when various DNA methylation lesions are repaired by ALKB family or TET family.



**Fig. 4.** Crystal structure of ALKB family activity center and hydroxylated substrate intermediate. (a)The enzyme activity pocket of ALKB is composed of the metal-ligand ion iron, a common substrate  $\alpha$ -KG, and several key residues (PDB 3O1S) . (b)The complex of ALKB and intermediate hm3T produced when m<sup>3</sup>T on DNA is repaired (PDB 3O1T) . (c)The complex of ALKB and intermediate

hm<sup>3</sup>C proposed when m<sup>3</sup>C on DNA is repaired (PDB 3O1S). (d)The complex of ALKB and intermediate 3-deazameC produced when 3-deazameC on DNA is repaired (PDB 3O1V).

## 2 The capture of Fe(IV)=O intermediates during demethylation

Nucleic acid demethylation is carried out by two steps of dioxygen activation and substrate oxidation (Fig. 2). During the activation of dioxygen, Fe(IV)=O is formed as an essential intermediate. The Fe(IV) intermediate in the catalytic process of the ALKB family has been supported by calculations and spectroscopy[17]. As another member of the Fe<sup>2+</sup>/α-KG dioxygenase family, TauD can catalyze the hydroxylation of taurine at C1. John C et al. discovered that the complex of ferrous iron ion and substrate forms a new intermediate after adding oxygen through the Mössbauer spectroscopy[18]. In the process, the Fe(IV)=O intermediate hypothesis was also proposed, and the existence of EPR technology and freeze-quench Mössbauer spectroscopy made its existence clear for the first time. The absorption of Fe(IV)=O decreases with increasing time. EPR is used to test whether the intermediate can be reduced to ferric iron, in order to prove its valence. Detailed molecular mechanics (MM) and molecular dynamics (MD) studies determine the flexibility of the intermediate. Dioxygen first binds to the iron(II) center, and then the iron(IV)-oxygen hybrid is rapidly formed, with the double ring structure as the transition state. The clustering model reveals the mechanism of substrate hydroxylation. The energy barrier required for substrate activation is very low and is feasible in terms of energy[19, 20].

## 3 The capture of hydroxylated intermediates during DNA demethylation

The substrate of the ALKB enzyme family is diverse. However, the TET enzyme family seems to exhibit demethylation function only for m<sup>5</sup>C (Table 1). In the presence of Fe(II) and α-KG, DNA methylation modification is reversed by oxidation of the ALKB family or TET family, and intermediates with different biochemical properties and stability are formed in the process. The form of the intermediate mainly includes hydroxymethylation product (hm-), formylation product (f-), and carboxylation product (ca-). Over the past few years, we have proposed hypothetical reaction mechanisms (Fig. 3), in which the intermediates for m<sup>6</sup>A, m<sup>3</sup>T, m<sup>3</sup>C, and m<sup>5</sup>C had been confirmed by considerable research in different methods.

### 3.1 Intermediates captured during m<sup>6</sup>A demethylation

As the homologs of ALKB, ALKBH1 and FTO can also demethylate m<sup>6</sup>A modifications of DNA. ALKB oxidizes the methyl group on the nitrogen atom to hydroxymethyl to generate hm<sup>6</sup>A, which can continue to complete demethylation by removing one molecule of

formaldehyde. Interestingly, FTO exhibits a unique tandem oxidation mechanism. m<sup>6</sup>A is still oxidized to hm<sup>6</sup>A first, the difference is that hm<sup>6</sup>A can be recognized by FTO as a substrate and then oxidized to f<sup>6</sup>A. hm<sup>6</sup>A and f<sup>6</sup>A can be converted into unmodified adenine by removing one molecule of formaldehyde or formic acid, respectively.

The effective combination of HPLC and high-resolution mass spectrometry is the primary method for intermediate capture. In 2012 Deyu Li et al. discovered that ALKB could ultimately convert 1, N<sup>6</sup>-ethanoadenine (EA) in DNA to unmodified adenine. The results of high-resolution HPLC-ESI-TOF experiments revealed the reaction process and intermediates conversion mechanism. This process involved the oxidation of N<sup>1</sup> and N<sup>6</sup>, which inspired them to continue to discover that another N<sup>6</sup>-modified product m<sup>6</sup>A is also a substrate of ALKB[21]. HPLC-ESI-TOF results showed that after incubating ALKB with m<sup>6</sup>A for 1 hour, in addition to the production of adenine, an intermediate hm<sup>6</sup>A was also detected, which was the first time to confirm the oxidative demethylation mechanism of ALKB. In order to maintain the integrity of the products and intermediates, the temperature of the autosampler was kept at 5 °C. As mentioned, FTO demethylates m<sup>6</sup>A in RNA through a tandem oxidation mechanism and generates two intermediates, hm<sup>6</sup>A and f<sup>6</sup>A. Although there have been studies showing that FTO can also recognize m<sup>6</sup>A in DNA as a substrate, there is currently no research to prove whether FTO will generate the same intermediate during DNA methylation.

In recent years, the biological function of ALKBH1 has gradually attracted attention. In 2017, Layla Kawarada et al. found that ALKBH1 can repair m<sup>5</sup>C methylation modifications in RNA and capture intermediates in the oxidation process by mass spectrometry[1]. In 2019, Key Laboratory of Analytical Chemistry for Biology and Medicine at Wuhan University discovered that ALKBH1 could also repair m<sup>6</sup>A modifications in DNA. Similar to the mechanism of ALKB demethylation, hm<sup>6</sup>A was also captured by mass spectrometry in the catalytic process of ALKBH1. Variable control experiments on the two cofactors of the ALKBH1, Fe(II) and α-KG, show that reducing their content will cause a significant reduction or almost disappearance of hm<sup>6</sup>A, which proves that hm<sup>6</sup>A is generated by ALKBH1[22]. Unlike the tandem oxidation mechanism of FTO, hm<sup>6</sup>A is not further oxidized to f<sup>6</sup>A.

The half-life of hm<sup>6</sup>A and f<sup>6</sup>A under neutral solution conditions were about 3.10 and 3.13 hours, respectively. They can be hydrolyzed to unmodified adenine. We do not rule out that other protein factors may play a role in promoting this process. Although hm<sup>6</sup>A and f<sup>6</sup>A are fundamentally not stable enough, they are stable if they bind to potential proteins and are embedded in a hydrophobic environment[2, 9].

### 3.2 Intermediates captured during m<sup>3</sup>T and m<sup>3</sup>C demethylation

The weaker binding of ALKB family proteins to DNA is not conducive to the identification of the crystal structure of ALKB-DNA complexes. Unlike the reaction mechanism of DNA dealkylase ABH2, when the substrate is dsDNA containing methylation damage, ALKB only interacts with the damaged DNA strand[23-26]. Verdine and other groups have developed a disulfide bond cross-linking strategy for stabilizing the structure of ALKB-dsDNA complexes, initially revealing the interaction between proteins and DNA[25]. Through a unique base reversal mechanism, bases with methylation damage are flipped out of the double helix and enter the active center of ALKB. At the same time, the bases far away from the active site of ALKB are modified with disulfide bonds to form a stable covalent bond with the residue Cys of ALKB, while the two adjacent bases of the damaged base that is turned over are stacked in together. Both of these factors stabilize the structure of the ALKB-dsDNA complex. Base stacking and hydrogen bonding between base pairs are important factors for stabilizing the structure of double-stranded DNA. The base flipping process that takes place in dsDNA requires higher energy than the same process in ssDNA. Another complementary DNA strand that does not contain methylation damage is a non-competitive inhibitor of ALKB, so ssDNA is a better substrate of ALKB.

The active site of demethylase is a nanoscale reaction pocket. The enzymatic reaction carried out in the single crystal of the enzyme is very conducive to the study of the reaction mechanism and the capture of intermediates. In 2010, Chengqi Yi et al. further optimized the chemical cross-linking strategy, and pioneered the capture of intermediates in the m<sup>3</sup>T demethylation process through the crystal oxidation study of the ALKB-dsDNA complex, and proposed the possibility that there is a zwitterionic intermediate during m<sup>3</sup>C demethylation process[4]. Although the methylation modification of m<sup>3</sup>T and m<sup>3</sup>C occurs on the nitrogen atom at the same position, their chemical properties are different: m<sup>3</sup>T is a neutral structure, and m<sup>3</sup>C carries a positive charge. Different charge properties lead them to have different stability and group leaving ability under the same physiological PH environment. Studies have shown that ALKB repairs m<sup>3</sup>C methylation faster than m<sup>3</sup>T. ALKB repairs base damage by activating hydrogen peroxide molecules and substrate oxidation. The ALKB-dsDNA complex containing m<sup>3</sup>T and m<sup>3</sup>C was grown under anaerobic conditions together with Fe<sup>2+</sup> and a-KG. The single crystal was exposed to hydrogen peroxide to initiate the oxidation of the crystal. The single plate-like structure of the single crystal is conducive to full oxidation, and the disulfide bond cross-linked with ALKB ensures that the inverted damaged base is within the active site pocket of ALKB (Fig. 4). The activation process of dioxygen molecules in the crystal seems to be fast, and no intermediates of the process are captured. In contrast, the oxidized intermediates of the substrate can be captured and stable for tens of hours.

When the compound crystals containing m<sup>3</sup>T are exposed to hydrogen peroxide, a-KG is oxidized to

succinate, and the methylated base is oxidized to a hydroxymethyl group directed to iron ions, forming the intermediate hm<sup>3</sup>T. The new hydroxyl group formed by oxidation forms hydrogen bonds with Arg210, Asn120, and succinate. The crystal resolution of the compound crystal containing m<sup>3</sup>C after being exposed to dioxygen is not enough to confirm its structure. However, the same study was carried out with the analog of m<sup>3</sup>C, 3-deaza-3-methylcytosine (3-deazamC). The definite intermediate structure 3-deazahmC was obtained. If the carbon atom at position three is replaced with a nitrogen atom again, there will be a proton transfer and form a zwitterionic structure. Therefore, hm<sup>3</sup>C and zwitterionic structures are likely to exist in the process of m<sup>3</sup>C oxidation. Due to different charging properties, hm<sup>3</sup>C and hm<sup>3</sup>T have different transformation mechanisms. Neutral hm<sup>3</sup>T is relatively stable and tends to protonate, releasing formaldehyde, generating unmodified thymine. The positively charged hm<sup>3</sup>C tends to deprotonate, forming a relatively stable zwitterionic intermediate, which is further decomposed into unmodified cytosine and formaldehyde. In addition to ALKB, FTO, ALKBH2, and ALKBH3 can also demethylate m<sup>3</sup>T[27-30]. Due to the similar charge properties of the substrates, we suppose that the same intermediate structure exists during their oxidation process. Crystal oxidation research dramatically expands the stability and visibility of intermediate capture and provides key structural information for revealing the specific oxidative demethylation mechanism.

### 3.3 Intermediates captured during m<sup>5</sup>C demethylation

5-Methylcytosine (m<sup>5</sup>C) is an epigenetic marker widely present in DNA and the main form of DNA methylation in eukaryotes such as plants and animals[31]. m<sup>5</sup>C is catalyzed by DNMT and plays a key role in cell processes such as gene regulation and differentiation, so it is also called the fifth base. The TET family also belongs to  $\alpha$ -KG and Fe<sup>2+</sup>-dependent dioxygenases, which mediate the demethylation process through oxidation. TET1 was first identified as having catalytic activity for converting m<sup>5</sup>C to 5-hydroxycytosine (hm<sup>5</sup>C)[32]. Subsequent research proves that the other two members of the TET family, TET2 and TET3, also have the same function[33]. At first, hm<sup>5</sup>C was mistaken for the product of automatic oxidation during sample processing. With the widespread use of 2D-TLC, HPLC-MS, and other technologies, the intermediates in the m<sup>5</sup>C demethylation process were gradually identified. Wyatt first found that hm<sup>5</sup>C is a hydroxylated form of m<sup>5</sup>C in phage DNA[34]. Penn et al. also found hydroxylated modified cytosine in DNA extracted from the brain tissue of mammalian rats, mice, and oviparous bullfrogs, accounting for about 15% of the total cytosine of DNA[35]. In 2009, TLC technology detected hm<sup>5</sup>C in mouse brain cells and embryonic stem cells, which triggered a follow-up research boom[16, 32, 36].

Hm<sup>5</sup>C is formed by the oxidation of methyl groups, often referred to as the sixth base, and plays an important role in the entire TET reaction mechanism (Fig. 3e). On the

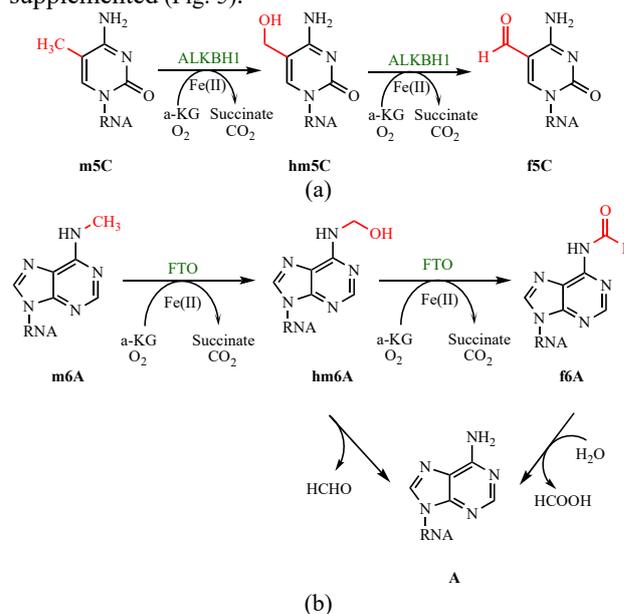
one hand, hm<sup>5</sup>C can be catalyzed by deaminase (AID) to produce hm<sup>5</sup>U, which is recognized and excised by thymine-DNA glycosylase (TDG), and then the base excision repair (BER) pathway can convert the site to cytosine to complete the demethylation[37]. On the other hand, hm<sup>5</sup>C can continue to be oxidized and form two other intermediates: 5-formylcytosine (f<sup>5</sup>C) and 5-carboxycytosine (ca<sup>5</sup>C). f<sup>5</sup>C and ca<sup>5</sup>C can also be recognized and excised by TDG, and the site is restored to cytosine through the BER pathway. In conclusion, with the participation of the BER pathway, TET family enzymes and TDG mediate the dynamic balance of the demethylation process. Although m<sup>5</sup>C, hm<sup>5</sup>C, and f<sup>5</sup>C can all be oxidized by TET, TET2 shows obvious substrate preference for m<sup>5</sup>C[38-40]. When three substrates are present at the same time, the reaction rate of m<sup>5</sup>C is about 5-8 times that of the other two. Because hm<sup>5</sup>C was oxidized to ca<sup>5</sup>C and then quickly identified and excised, the whole gene map showed that the expression level of hm<sup>5</sup>C was relatively insufficient in the TET binding sites of most ES cells[16].

Unlike the well-known ALKB family substrates such as m<sup>1</sup>A and m<sup>6</sup>A, m<sup>5</sup>C methylation modification occurs on the carbon atom rather than the nitrogen atom. Surprisingly, in 2019 Ke Bian et al. Found that ALKB, ALKBH2, and ALKBH3 can oxidize m<sup>5</sup>C in vitro and produce the same intermediates: hm<sup>5</sup>C, f<sup>5</sup>C, and ca<sup>5</sup>C. The experiment used enzymatic reaction conditions similar to TET, and the presence of hm<sup>5</sup>C, f<sup>5</sup>C, and ca<sup>5</sup>C were detected in the oxidation products of the three enzymes[41]. ALKB and ALKBH3 have a clear preference for single-stranded DNA substrates, while ALKBH2 seems to have a slight preference for double-stranded DNA substrates. High-resolution mass spectrometry indicates that the three intermediates produced by different enzymes have different specific gravities. Hm<sup>5</sup>C occupies the majority of the products of the oxidation reaction catalyzed by ALKBH2 and ALKBH3, while f<sup>5</sup>C and ca<sup>5</sup>C have less accumulation than it. In contrast, the most accumulated intermediate after ALKB oxidation of the substrate is f<sup>5</sup>C. Molecular dynamics simulation is an effective mean to study the microstructure and explore the underlying mechanism. In different molecular models, when m<sup>5</sup>C binds to ALKB or ALKBH2 in a *syn*-conformation, the distance between the methyl group at the C5 site and the active Fe center of the enzyme is only 3.8 Å, which is much shorter than that in *anti*-conformation binding. The torsion angle predicted by the *syn*-conformation binding model is consistent with the data of the crystal structure of the TET-m<sup>5</sup>C complex. Therefore, we speculate that the ALKB family is only more conducive to the generation of several oxidized intermediates when the *syn*-conformation of m<sup>5</sup>C is combined. This study provides the first evidence that the ALKB family can also repair the methylation of carbon sites, and expands the new field of ALKB family regulation of epigenetic modification. ALKB, ALKBH2, and ALKBH3 can produce the same oxidation products as the TET enzyme, but whether they can oxidize m<sup>5</sup>C modification of DNA in vivo and whether other similar enzymes (such as FTO) also have the same function remains to be confirmed by subsequent studies. Hm<sup>5</sup>C is

involved in gene expression and regulation and is related to stem cells and cancer, which means that ALKB family enzymes may play a role in epigenetics in a new way.

## 4 The capture of hydroxylated intermediates during RNA demethylation

RNA methylation proved to be a reversible dynamic process for the first time in 2011, mainly including m<sup>1</sup>A and m<sup>6</sup>A[27]. In addition to DNA methylation, the ALKB family also has enzymatic activity for RNA methylation. For example, ALKBH5 and FTO can convert m<sup>6</sup>A in mRNA to adenine, which are two widely studied demethylases[42]. As the most extensive modification of mRNA methylation in mammalian cells, the demethylation process of m<sup>6</sup>A-RNA produces intermediates similar to the DNA demethylation process. Recent studies have found that ALKBH1 can reverse the m<sup>5</sup>C modification in tRNA to cytosine through oxidation, generating two intermediates, hm<sup>5</sup>C and f<sup>5</sup>C[1]. At present, the research on RNA demethylase is relatively insufficient (Table 2), so the mechanism involved in the process and the capture of intermediates still need to be supplemented (Fig. 5).



**Fig. 5.** The reaction mechanism when various RNA methylation lesions are repaired by ALKB family.

### 4.1 Intermediates captured during m<sup>6</sup>A demethylation

The m<sup>6</sup>A modification in RNA occupies 1% to 2% of all adenine and plays an important role in gene regulation. In 2013, Ye Fu et al. discovered the unique tandem oxidation mechanism of FTO through HPLC separation and MALDI-TOF analysis. During the demethylation of m<sup>6</sup>A-RNA, two oxidized intermediates were captured, hm<sup>6</sup>A and f<sup>6</sup>A. Under the ionization environment, the mass spectrum peak of m<sup>6</sup>A-2 Da can be generated after the dehydration of hm<sup>6</sup>A. In addition, the observed mass

spectrum peak of  $m^6A + 14$  Da confirmed the existence of  $f^6A$  and verified that FTO has a tandem oxidation mechanism similar to TET (Fig. 5b). The specific reaction mechanism involves that  $m^6A$  is first oxidized to  $hm^6A$ , and  $hm^6A$  is further oxidized to  $f^6A$ . The kinetic data indicates that the reaction rate of the first-step oxidation is 12.3 times that of the second-step oxidation. As a decisive step, the rate of the second oxidation is positively correlated with the local concentration of FTO.  $hm^6A$  and  $f^6A$  are relatively stable under physiological conditions, with a half-life of approximately 3 hours. In a neutral solution, both  $hm^6A$  and  $f^6A$  can spontaneously hydrolyze to produce unmodified adenine. Both acidic and alkaline conditions can accelerate the process. In addition,  $hm^6A$  can also be recognized as a substrate by FTO. According to the predicted free energy data,  $hm^6A$  has a more potent binding force. Molecular dynamics simulations were used to construct  $m^6A$ -RNA and  $hm^6A$ -RNA models. The results showed that the  $N^6$  methyl groups of both substrates were at the iron center of the active site of FTO. YTHDF2 protein is considered to be a potential  $m^6A$ -RNA binding protein, but its binding affinity to  $hm^6A$  and  $f^6A$  is significantly lower than that to  $m^6A$ -RNA, which means that there may be other binding proteins that regulate  $hm^6A$  and  $f^6A$ , while the formation of intermediates inhibits the protein interaction between  $m^6A$ -RNA and YTHDF2[8].

ALKBH5 is the second discovered RNA demethylase. Further research revealed that both ssDNA and ssRNA with methylation modification could be recognized as substrates by ALKBH5. ALKBH5 has a particular preference for  $m^6A$  in ssRNA. Knocking out the ALKBH5 gene will affect the sperm count of male mice. Interestingly, ALKBH5 only has catalytic activity on  $m^6A$  of mRNA, while there is no enzyme activity on  $m^6A$  in tRNA. Unlike FTO, MALDI-TOF / TOF mass spectrometry analysis showed that when ALKBH5 demethylated  $m^6A$ -ssRNA,  $hm^6A$  and  $f^6A$  were not produced, only the  $m^6A$ -14 Da peak representing adenine was observed. ALKBH5 seems to have a different demethylation mechanism with FTO, which can directly convert  $m^6A$  to unmodified adenine. ALKBH5 also demethylates the dimethylated adenosine  $m^{(6)}2A$ , which is a non-canonical base present in ribosomal RNA[43]. One hypothesis is that when  $hm^6A$  is produced, it will exist in the enzyme active site very shortly, and then it will be catalyzed and hydrolyzed to adenine very quickly. This hypothesis needs to be proved by further research. Arg-130 and Lys-132 are two key residues in the conserved  $\alpha$  chain of the ALKB family, and ALKBH5 has both two key residues, which makes it unique from other protein family members. In addition, these two residues may have a strong interaction with  $m^6A$ , making ALKBH5 have a stronger affinity to  $m^6A$  compared with FTO. The active site cavity of ALKBH5 is relatively small, and the calculation shows its volume is  $490.2 \text{ \AA}^3$ , while the active site volume of FTO is  $817.5 \text{ \AA}^3$ . The larger active site volume may explain the apparent substrate diversity of FTO compared to ALKBH5[44]. The difference in the structure of these proteins may be the internal reason that ALKBH5 and FTO have different catalytic mechanisms[45].

## 4.2 Intermediates captured during $m^5C$ demethylation

There are more than 100 modifications in RNA, 80% of which are located in tRNA. There are multiple modifications at the first position (position 34) of the anticodon of tRNA. By regulating the interaction of codons and anticodons, these modifications ensure accurate translation. NSUN2 / 3 is a methyltransferase directed to mitochondrial tRNA (mt-tRNA)[46]. In the presence of the coenzyme Adomet, it can act on the cytosine of the tRNA moving position and methylate it to  $m^5C$ . ALKBH1 is an oxidized mt-tRNA methyl-modified dioxygenase. In 2016, ESI-MS research proved that ALKBH1 could oxidize  $m^5C$  in mt-tRNA to  $f^5C$  in vitro (Fig. 5a). The RNA can be labeled with 1-styrene-2,3,3-trimethylindole instead of 5-sulfonate (TMI)[46]. In 2017, Layla Kawarada et al. Proved that there is another oxidation intermediate  $hm^5C$ , which improves the mechanism of ALKBH1 oxidation of  $m^5C$ : in the presence of oxygen and  $\alpha$ -KG,  $m^5C$  is first oxidized by ALKBH1 to  $hm^5C$ , and  $hm^5C$  continues to be oxidized  $f^5C$ [1]. The intermediate  $f^5C$  is necessary for AUA (non-universal codon in mitochondria) translation, so knocking out ALKBH1 has a high impact on mitochondrial translation function. The successful capture of  $hm^5C$  and  $f^5C$  clearly clarified the mechanism of  $m^5C$  oxidation and laid the indispensable position of ALKBH1 in the regulation of mitochondrial function[47-49].

## 5 Summary and outlook

In the past few decades, the study of intermediates in the demethylation process of nucleic acids has made great leaps, which not only makes us clearly understand the specific enzyme catalytic mechanism, but also brings us many new topics. Demethylation is accomplished by dioxygen activation and substrate oxidation, and the corresponding intermediates are captured in both steps. In the hydrogen peroxide activation stage, Fe (II) and  $\alpha$ KG activate hydrogen peroxide to form highly active Fe (IV) = O. In the substrate oxidation stage, Fe (IV) = oxo inserts oxygen atoms into the C-H bond of the substrate, and Fe (IV) is reduced back to Fe (II) to complete the catalytic cycle. Spectroscopic technology and TLC are the earliest technique used to identify oxidized intermediates, but the accuracy of the data is insufficient. In recent years, molecular dynamics simulation, oxidation crystallization, and high-resolution liquid chromatography/mass spectrometry have been widely used in the research of intermediates, especially LC-MS is the most common and intuitive method to identify the presence of an intermediate. Due to the rapid formation of intermediates and the short half-life, most methods require low temperatures to ensure the stability of substrates and intermediates. EPR technology can capture the reaction changes within tens of milliseconds by virtue of its high time resolution, which is helpful for the study of transient intermediates. However, we still have many problems to be solved. For example, the physiological role of each intermediate is still unclear. In addition, there are many

substrate intermediates that are only assumed to exist while lack direct experimental data to support. Both FTO and ALKBH5 are similar RNA demethylases. FTO can produce two intermediates when catalyzing m<sup>6</sup>A demethylation, hm<sup>6</sup>A, and f<sup>6</sup>A, but whether ALKBH5 has intermediates during m<sup>6</sup>A demethylation still has no precise answer. MIA and m<sup>3</sup>C have similar charge properties. Many views assume that they have hydroxylated intermediates, but there has been no captured study to prove their existence. A recent study identified N<sup>6</sup>-methyladenine (N<sup>6</sup>-meA) in mammalian transfer RNA (tRNA) as a novel ALKBH3 substrate. An in vitro translation assay showed that ALKBH3-demethylated tRNA significantly enhanced protein translation efficiency [50]. Whether hm<sup>6</sup>A will exist during ALKBH3 catalyzation still need to explore. How the structural differences between enzymes of the same enzyme family lead to different catalytic mechanisms is the direction of our subsequent research. Intermediate capture is a crucial requirement to reveal the answer.

## References

1. Kawarada L., et al., *Nucleic Acids Res*, **45**(12): p. 7401-7415(2017)
2. Shen L., et al., *Annu Rev Biochem*, **83**: p. 585-614(2014)
3. XU Ping C.S., *Chin J Gastroenterol*, **15**, No.4(2010)
4. Yi C., et al., *Nature*, **468**(7321): p. 330-3(2010)
5. Wu H. and Zhang Y., *Cell*, **156**(1-2): p. 45-68(2014)
6. Xiong J., et al., *Nucleic Acids Res*, **47**(3): p. 1268-1277(2019)
7. Yi C. and He C., *Cold Spring Harb Perspect Biol*, **5**(1): p. a012575(2013)
8. Fu Y., et al., *Nat Commun*, **4**: p. 1798(2013)
9. Liu J. and Jia G., *J Genet Genomics*, **41**(1): p. 21-33(2014)
10. Li D., et al., *Chem Res Toxicol*, **26**(8): p. 1182-7(2013)
11. Sundheim O., Vågbø, C. B., Bjørås, M., Sousa, M. M., Talstad, V., Aas, P. A., Drabløs, F., Krokan, H. E., Tainer, J. A., and Slupphaug, G., (2006)
12. Lee D.H., Jin, S. G., Cai, S., Chen, Y., Pfeifer, G. P., and O'Connor, T. R., *J Biol Chem*, **280**(47): p. 39448-39459(2005)
13. Yang C.G., et al., *Nature*, **452**(7190): p. 961-965(2008)
14. Pastore C., et al., *J Biol Chem*, **287**(2012)
15. GUO Xiao Qiang W.Y.J., GUO Zhen Qing, *Chinese Journal of Biochemistry and Molecular Biology*, **2011**,**27**( 12) p. 1101-1106(2011)
16. He Y.F., et al., *Science*, **333**(6047): p. 1303-7(2011)
17. Price J.C., et al., *Biochemistry*, **44**(22): p. 8138-47(2005)
18. Price J.C., et al., *Biochemistry*, **42**(24): p. 7497-508(2003)
19. Quesne M.G., et al., *Chemistry*, **20**(2): p. 435-46(2014)
20. Latifi R., et al., *Dalton Trans*, **49**(14): p. 4266-4276(2020)
21. Li D., et al., *J Am Chem Soc*, **134**(21): p. 8896-901(2012)
22. Zhang M., et al., *Cell Res*, **30**(3): p. 197-210(2020)
23. Huang H., et al., *Science*, **282**(5394): p. P.1669-1675(1998)
24. Dinglay S., et al., *Genes & Development*, **14**(16): p. 2097-2105(2000)
25. Verdine G.L. and Norman D.P.G., *Annual Review of Biochemistry*, **72**(1): p. 337-366(2003)
26. Mishina Y., Chen L.X., and He C., *Journal of the American Chemical Society*, **126**(51): p. 16930-16936(2005)
27. Jia G., et al., *Nat Chem Biol*, **7**(12): p. 885-7(2011)
28. Zou S., et al., *Sci Rep*, **6**: p. 25677(2016)
29. Yang C.G., et al., *Nature*, **452**(7190): p. 961-5(2008)
30. Yang T., et al., *Chem Commun (Camb)*, **52**(36): p. 6181-4(2016)
31. Kriaucionis S. and Heintz N., *Science*, **324**(5929): p. 929(2009)
32. Tahiliani M., et al., *Science*, **324**(5929): p. 930-5(2009)
33. Klug M., et al., *Genome Biol*, **14**(5): p. R46(2013)
34. Wyatt G.R. and Cohen S.S., *Biochemical Journal*, **55**(5): p. 774(1953)
35. *Biochemical Journal*, **126**(4): p. 781-790(1972)
36. Gerken T., et al., (2007)
37. Zhang L., et al., *Nat Chem Biol*, **8**(4): p. 328-30(2012)
38. Ito S., et al., *Science*, **333**(6047): p. 1300-3(2011)
39. Kramer G., et al., *Nature*, **419**(6903): p. 171-4(2002)
40. Kriukiene E., Liutkeviciute Z., and Klimasauskas S., *Chem Soc Rev*, **41**(21): p. 6916-30(2012)
41. Bian K., et al., *Nucleic Acids Res*, **47**(11): p. 5522-5529(2019)
42. Zhang X., et al., *Proc Natl Acad Sci U S A*, **116**(8): p. 2919-2924(2019)
43. Ensfielder T.T., et al., *Chem Commun (Camb)*, **54**(62): p. 8591-8593(2018)
44. Feng C., et al., *J Biol Chem*, **289**(17): p. 11571-83(2014)
45. Toh J.D.W., et al., *Chem Sci*, **6**(1): p. 112-122(2015)
46. Haag S., et al., *EMBO J*, **35**(19): p. 2104-2119(2016)
47. Martinez S. and Hausinger R.P., *J Biol Chem*, **290**(34): p. 20702-11(2015)
48. Neri F., et al., *Cell Rep*, **10**(5): p. 674-683(2015)
49. Thapar R., et al., *Biochemistry*, **58**(5): p. 312-329(2019)
50. Ueda Y., et al., *Sci Rep*, **7**: p. 42271(2017)