Cell Research (2010) 20:886-898.
© 2010 IBCB, SIBS, CAS All rights reserved 1001-0602/10 $ 32.00
www.nature.com/cr

ORIGINAL ARTICLE

Structural insights into a dual-specificity histone demethylase ceKDM7A from *Caenorhabditis elegans*

Ying Yang[1, 2, 3, *], Lulu Hu[2, 3, *], Ping Wang[2, 3, *], Haifeng Hou[4], Yan Lin[2, 3], Yi Liu[2, 3], Ze Li[2, 3], Rui Gong[2, 3], Xiang Feng[2], Lu Zhou[5], Wen Zhang[2, 3], Yuhui Dong[4], Huirong Yang[2, 3], Hanqing Lin[6], Yiqin Wang[6], Charlie Degui Chen[6], Yanhui Xu[1, 2, 3]

[1]Cancer Institute, Shanghai Cancer Center, Fudan University; Department of Oncology, Shanghai Medical College, Fudan University, shanghai 200032, China; [2]School of Life Sciences, Fudan University, 220 Han-Dan Road, Shanghai 200433, China; [3]Institutes of Biomedical Sciences, Fudan University, 130 Dong-An Road, Shanghai 200032, China; [4]Beijing Synchrotron Radiation Facility, Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049, China; [5]School of Pharmacy, Fudan University, South Chemistry Building, Room 315, 826 Zhangheng Road, Shanghai 201203, China; [6]State Key Laboratory of Molecular Biology, Shanghai Key laboratory of Molecular Andrology, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 320 Yueyang Road, Shanghai 200031, China

Histone lysine methylation can be removed by JmjC domain-containing proteins in a sequence- and methylation-state-specific manner. However, how substrate specificity is determined and how the enzymes are regulated were largely unknown. We recently found that ceKDM7A, a PHD- and JmjC domain-containing protein, is a histone demethylase specific for H3K9me2 and H3K27me2, and the PHD finger binding to H3K4me3 guides the demethylation activity *in vivo.* To provide structural insight into the molecular mechanisms for the enzymatic activity and the function of the PHD finger, we solved six crystal structures of the enzyme in apo form and in complex with single or two peptides containing various combinations of H3K4me3, H3K9me2, and H3K27me2 modifications. The structures indicate that H3K9me2 and H3K27me2 interact with ceKDM7A in a similar fashion, and that the peptide-binding specificity is determined by a network of specific interactions. The geometrical measurement of the structures also revealed that H3K4me3 associated with the PHD finger and H3K9me2 bound to the JmjC domain are from two separate molecules, suggesting a *trans*-histone peptide-binding mechanism. Thus, our systemic structural studies reveal not only the substrate recognition by the catalytic domain but also more importantly, the molecular mechanism of dual specificity of ceDKM7A for both H3K9me2 and H3K27me2.

***Keywords*:** histone; methylation; demethylase; structure; PHD; JmjC

Cell Research (2010) **20**:886-898. doi:10.1038/cr.2010.86; published online 22 June 2010

Introduction

Histone lysine methylation is a reversible modification involved in many physiological and pathological processes [1-4]. Many lysine residues in histones can be methylated, and each lysine can have three states of methylation. The addition of methyl groups to histone lysines is catalyzed by histone methyltransferases, which includes the set-domain and non-set-domain proteins [5]. Histone lysine methylation can be removed by histone demethylases. LSD1 is the first histone demethylase identified, which can remove di- and mono-methylation from H3K4 using an amine oxidase reaction [6]. Subsequently, a JmjC domain-containing protein was identified to possess histone demethylase activity, and the JmjC domain was shown as a demethylase signature motif [7]. This class of enzymes catalyzes the demethylation by a hydroxylation reaction and requires both iron and α-ketoglutarate as cofactors.

*These three authors contributed equally to this work.
Correspondence: Charlie Degui Chen[a], Yanhui Xu[b]
[a]E-mail: cdchen@sibs.ac.cn
[b]E-mail: xuyh@fudan.edu.cn
Received 23 March 2010; revised 4 May 2010; accepted 10 May 2010; published online 22 June 2010

Previously, we and others identified 30 JmjC domain-containing proteins in humans through analysis of public protein-domain databases [8-10] These proteins can be classified into seven groups based on domain structure of the full-length proteins. Proteins in all of the seven groups have been demonstrated to be histone lysine demethylases, which remove methyl groups from histone H3 in a sequence- and methylation-state-specific manner [11-23]. The structures of JMJD2A, one of the JmjC domain-containing histone lysine demethylase specific for H3K9me3 and H3K36me3, have been solved in apo form or in complex with its substrates [24-27]. Although those structural studies had provided insight into the molecular mechanism of catalysis and substrate recognition, how substrate specificity is determined for most of the other JmjC domain-containing histone lysine demethylases was still unknown.

In addition to the JmjC domain, most of the JmjC domain-containing histone lysine demethylases contain domains predicted to be involved in protein or DNA interactions. These domains include PHD, Tudor, FBOX, Bright/Arid, Zinc Finger, and TPR [8]. The PHD finger was shown be a binding motif for both methylated and unmethylated histones [17, 28-29]. Similarly, the Tudor domain is also an H3K4me3-binding module [30]. Bright/Arid domain was shown to be a DNA-binding domain and the Bright/Arid domain of JARID1B bound DNA with little or no sequence specificity [31]. Biochemical studies indicate that these protein or DNA-binding domains are important for the demethylase activity *in vivo* [10, 19]. However, how these domains cooperate with the JmjC domain to regulate histone modifications and gene transcription remains unclear.

Previously, we identified KIAA1718 (KDM7A) as a dual-specificity histone demethylase for H3K9me2 and H3K27me2 that regulates neural differentiation through controlling the expression of FGF4 [32]32. We also found that its *C. elegans* ortholog ceKDM7A, a PHD- and JmjC domain-containing protein, is a histone demethylase specific for H3K9me2 and H3K27me2, and the PHD finger binding to H3K4me3 guides the demethylation activity *in vivo* (Lin *et al*., accompanying paper in this issue [33]). However, the molecular mechanisms for both the enzymatic activity and the function of the PHD finger remain unknown. In this report, we present six crystal structures of the enzyme in apo form and in complex with single or two peptides containing various combinations of H3K4me3, H3K9me2, and H3K27me2 modifications. Structural and biochemical analyses reveal that ceKDM7A binds both H3K9me2- and H3K27me2-containing peptides in a similar fashion, and the substrate recognition and specificity are determined by extensive hydrogen-bonding and hydrophobic interactions. Based on comparisons to the structures of JMJD2A and LSD1, we propose that a shallow binding pocket determines the specificity of ceKDM7A for di- but not trimethylated lysine. Crystal structure of ceKDM7A complexed with H3K4me3K9me2 also reveals that the enzyme cannot bind H3K4me3 and demethylate H3K9me2 in the same histone peptide, suggesting a *trans*-histone regulation mechanism for histone binding and erasing of histone methylation.

Results

Overall structure of the enzyme

We have recently demonstrated that ceKDM7A, a PHD- and JmjC domain-containing protein, is a histone demethylase specific for H3K9me2 and H3K27me2. The PHD finger binds H3K4me3, and the binding directs the demethylation by the JmjC domain *in vivo* (Lin *et al*., accompanying paper in this issue [33]). To gain insights into the substrate specificity of the enzyme and the molecular mechanism of the cross-regulation between the PHD finger and the JmjC domain, we solved cocrystal structures of the enzyme in complex with a histone H3 peptide with dual methylation marks H3K4me3K9me2 and H3K4me3K27me2 (Figure 1 and Supplementary information, Figure S1). The protein fragment used for crystallization contains residues 188-711, which includes both the PHD finger and the JmjC domain (Figure 1A). Because H3K4me3 is an active mark and both H3K9me2 and H3K27me2 are repressive marks, and the active and repressive marks usually distribute in a mutually exclusive manner, we also obtained cocrystal structures of the enzyme in complex with two separate peptides, one containing H3K4me3 and the other containing either H3K9me2 or H3K27me2. The statistics of data collection have been listed in Supplementary information, Table S1.

The apo form and the complex structures of ceKDM7A in the presence of various methylated peptides are superimposed well with the highest root mean square deviation (RMSD) value of 0.906 Å for 425 Cα atoms (Supplementary information, Figure S2), indicating that the enzyme does not undergo significant conformational change on peptide binding. These data also suggest that active and repressive methylations in one peptide or in two separate peptides did not affect the conformation of the enzyme. In these structures, ceKDM7A displays an overall dimension of ~90 × 50 × 50 $Å^3$ and is composed of the N-terminal PHD finger (188-283), a JmjC domain-containing catalytic core (284-615), and a C-terminal extended coiled-coil region (616-704) (Figure 1 and

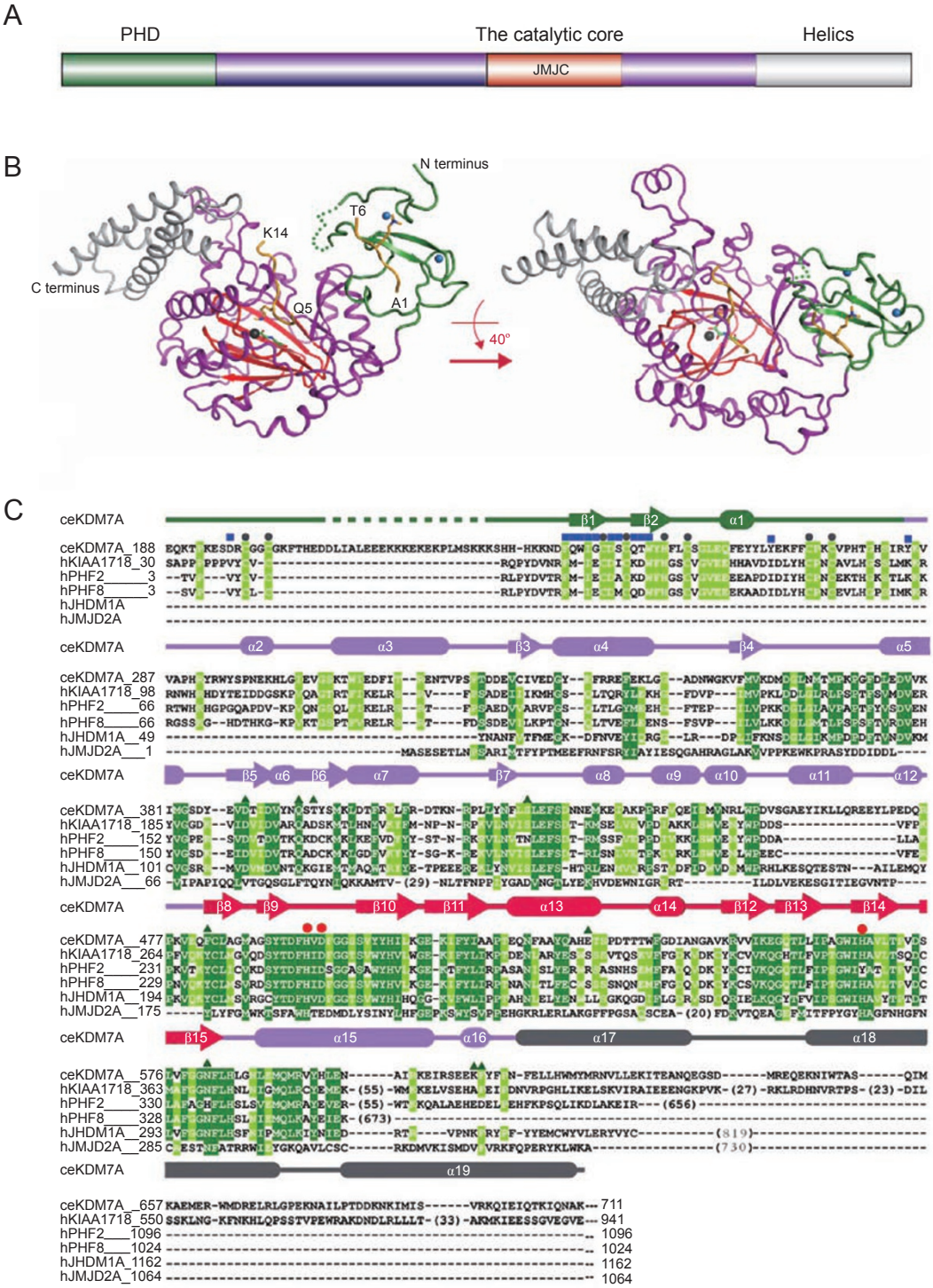


Figure 1 Overall structure of the enzyme complexes. **(A)** The domain structure of the ceKDM7A fragment in crystals. The color scheme is used to illustrate all structure figures. **(B)** Two different views of ceKDM7A complex with a peptide containing both H3K4me3 and H3K9me2. The peptides are colored in yellow, NOG in green, zinc in blue, and Fe (II) in black. N and C terminus of ceKDM7A are indicated. **(C)** Sequence alignment of ceKDM7A (residues 188-711) with human KIAA1718, PHF2, PHF8, JHDM1A, and JMJD2A. Identical and highly conserved residues are highlighted with dark green background and conserved residues with light green background. Secondary structural elements of ceKDM7A (according to complex structure of ceKDM7A with a peptide containing both H3K4me3 and H3K9me2) are colored as in **A** and indicated above the sequences. Residues that are involved in H3K4me3 and H3K9me2 recognition are indicated by blue squares and green triangles, respectively. Residues that are involved in zinc and iron coordination are indicated by bluish gray and red dots, respectively.

Supplementary information, Figure S1). The N-terminal PHD finger is formed by a conserved $Cys_4HisCys_3$ motif coordinated to two Zn^{2+} ions (Supplementary information, Figure S3). The catalytic core is comprised of an interior JmjC domain surrounded by 15 α helices and 5 β strands forming a compact globular structure (Figure 1B and 1C). Eight β strands (β8 to α15) form a conserved jellyroll motif, which is similar to that in other JmjC domain-containing histone demethylase [24, 27]. The PHD finger and the coiled-coil region interact with the catalytic core on both sides.

The PHD finger and the complex structure with H3K4me3

The N-terminus of ceKDM7A is folded as a classic PHD finger [34, 35] in free state and in complex with H3K4me3-containing peptides (Supplementary information, Figure S3A-S3C). The free and bound PHD fingers have similar overall structures with RMSD value of 0.736 Å for 53 Cα atoms. The structure consists of a small α helix (α1), a double-stranded anti-parallel β-sheet (β1 and β2), and four loops (Figure 1C and Supplementary information, Figure S3C). The β/β/α motif forms the core in a compact globular fold, which is stabilized by two zinc atoms in an interleaved manner.

In complexes with H3K4me3-containing peptides, residues 1-6 of the histone peptide adopts a β strand conformation and binds to the PHD finger through β-augmentation (Supplementary information, Figure S3B and Figure 2A). The trimethyl-lysine side chain inserts into an open pocket formed by residues D196, W241, and W250. The R2 side chain forms three intermolecular hydrogen bonds with side chains of residues D245 and Q248, and the carbonyl group of residue C244. The K4- and R2-binding pockets are separated by the aromatic side chain of W250. In general, the binding mode is similar to the previously defined H3K4me3 recognition mode in other PHD-peptide complexes [36, 37]. To determine the contribution of the residues involved in binding, we mutated D196, W241, W250, G243, D245, and Q248 individually or in combination. Circular dichroism spectrum indicates that the mutants maintain the secondary structure of the enzyme (data not shown). Although the wild-type protein bound an H3K4me3-containing peptide, mutation of each residue abolished the binding (Figure 2B), suggesting that these residues contribute to the PHD finger binding to H3K4me3-containing peptides.

The active site

In complex structures, the dimethylated lysine is deeply inserted into the interior of the JmjC domain where Fe (II) and NOG bind (Figure 3A). The Fe(II) is coordinated by side chains of the highly conserved residues H495, D497, and H567, and NOG is stabilized by residues N421, T492, and Y505 (Figure 3B). The methyl side chain of K9/27me2 is checked by F482 and F498 through hydrophobic interactions. One of the methyl groups of dimethylated lysine interacts with side chains of D497 and N581 through two nonconventional CH-O hydrogen bonds with distances of 3.3 and 3.12 Å, respectively. All residues involved in formation of the catalytic complexes are located in the JmjC domain with side chains facing inward. In vitro demethylase assay showed that mutations of Y505A, D497A, F482A, and N581A significantly reduced the activity (Figure 3C and Supplementary information, Table S2), indicating the essential roles of these residues for the demethylase activity.

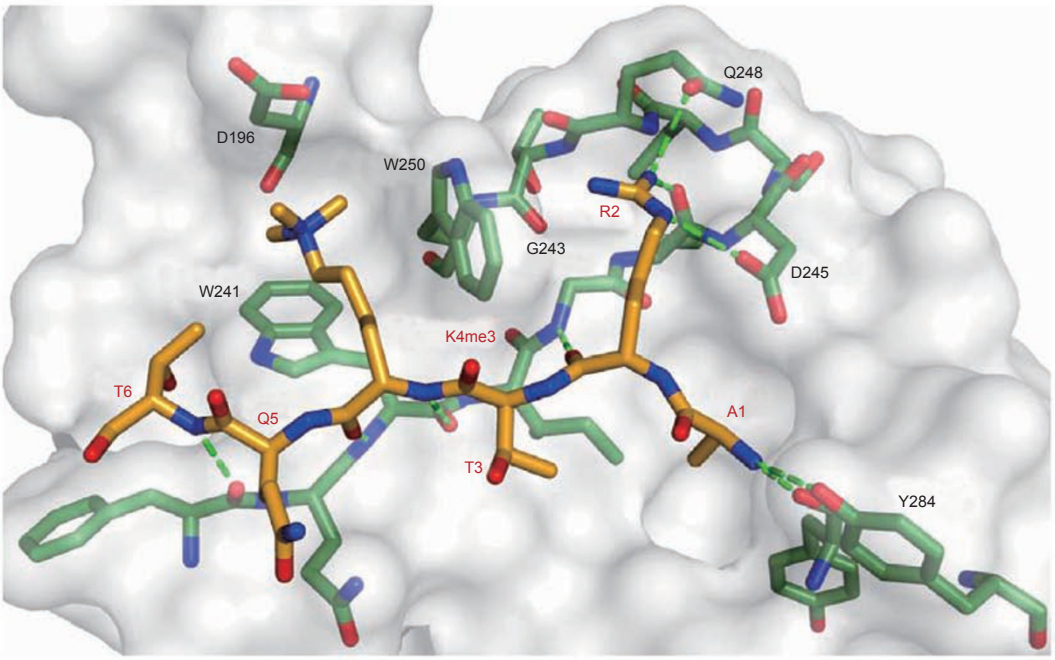
Substrate recognition

H3K9me2 and H3K27me2 peptides bind to the catalytic core in a similar fashion (Figure 4A-4B and Supplementary information, Figure S4A-S4B). The peptides adopt an extended conformation along a groove on the surface of the catalytic core. Substrate recognition is mainly achieved through a network of hydrogen bonds. Most atoms involved in hydrogen bonds in peptide are from the main chains of the peptides. In complexes with H3K9me2-containing peptides, the side chains from R8, S10, and T11 also form hydrogen bonds with E531, D389, and E609, which may be involved in substrate specificity determination. *In vitro* demethylase assay showed that mutations of D389A, Q396A, T398A, and E531I decreased the enzymatic activity (Figure 4C), indicating that the interactions are important for substrate recognition and the enzymatic activity. In addition, four residues D389, T398, K610, and F611 form a lid above the Cα of K9/27, which forces the peptide into a V shape conformation and pushes K9/27 deeply into the catalytic site (Figure 4A). This conformation is further stabilized by the hydrogen bonds between S424 of the enzyme and S10 from the peptide (Figure 4A-4B). Mutations of S424A and E609K610F611AAA abolished the enzyme activity, suggesting the essential role of the V-shaped conformation for demethylase activity. Detailed interactions are summarized in Supplementary information, Table S2.

Substrate specificity

ceKDM7A specifically demethylates H3K9me2 and H3K27me2, two important epigenetic marks associated with transcription repression. Both lysine residues are flanked by identical amino acids ($\underline{ARK_{9/27}S}$), suggesting that the substrate specificity of ceKDM7A may be defined by the three flanking residues (Figure 5A). Indeed, R8/R26 and S10/S28 interact with the enzyme through

A



B

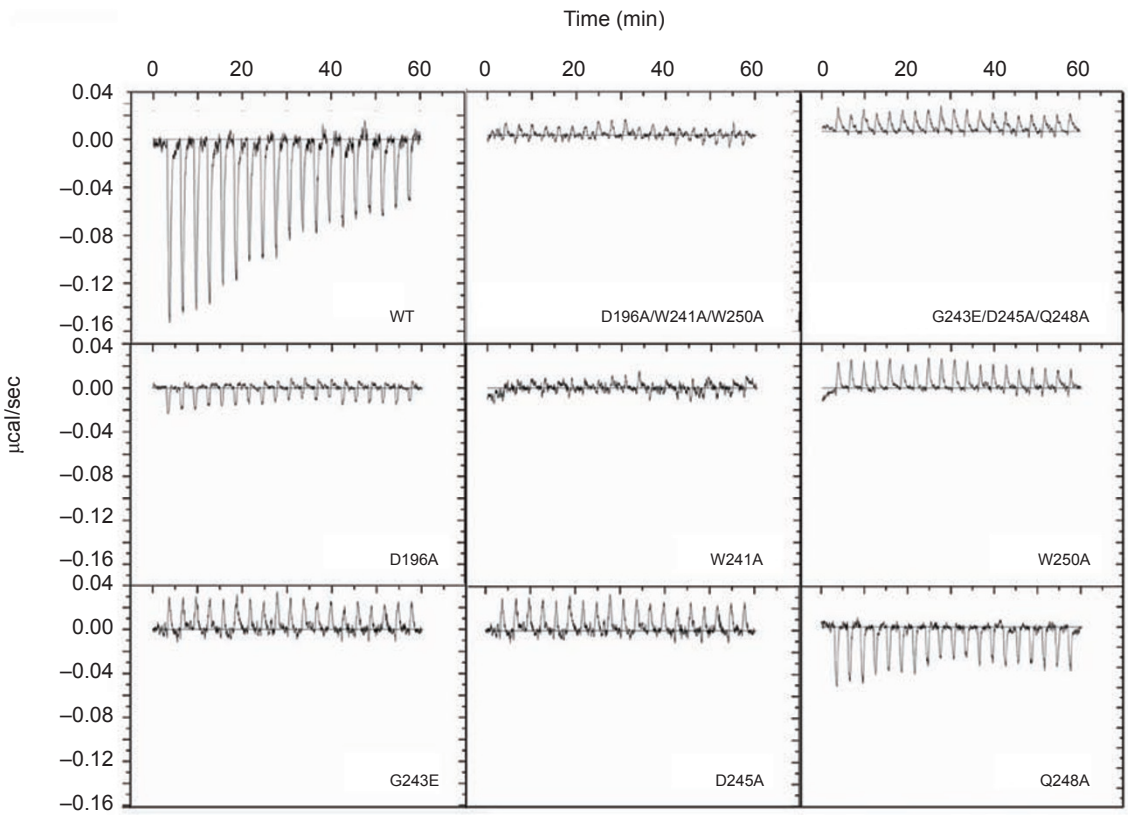


Figure 2 Binding of H3K4me3 by the PHD finger. **(A)** Interaction between the PHD finger and H3K4me3-containing peptide. Hydrogen bonds are indicated as green dotted line. **(B)** Isothermal titration calorimetry (ITC) assays of H3K4me3 peptide interacting with the wild-type and various mutants of ceKDM7A.

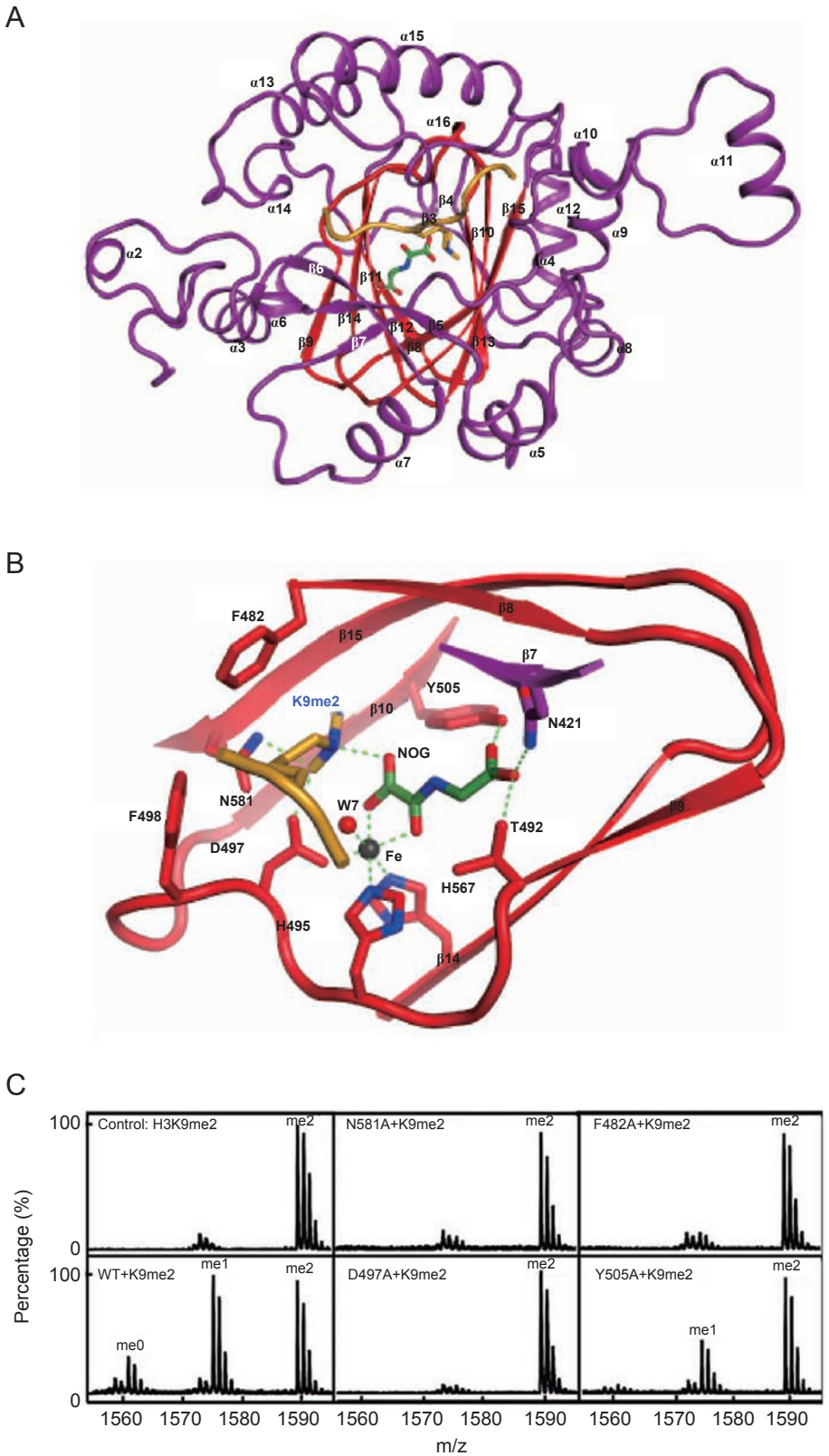


Figure 3 The catalytic core and active site. **(A)** Structure of the catalytic core of ceKDM7A with a H3K9me2-containing peptide. **(B)** Structure of the catalytic intermediate. Hydrogen bonds and Fe (II) coordination are represented by green dotted lines. W7 is a water molecule. **(C)** Demethylase activity of the wild type and various mutants of ceKDM7A. The molar ratios of the enzyme and the substrate were kept at 1:10. H3K9me2 peptide was used as the substrate.

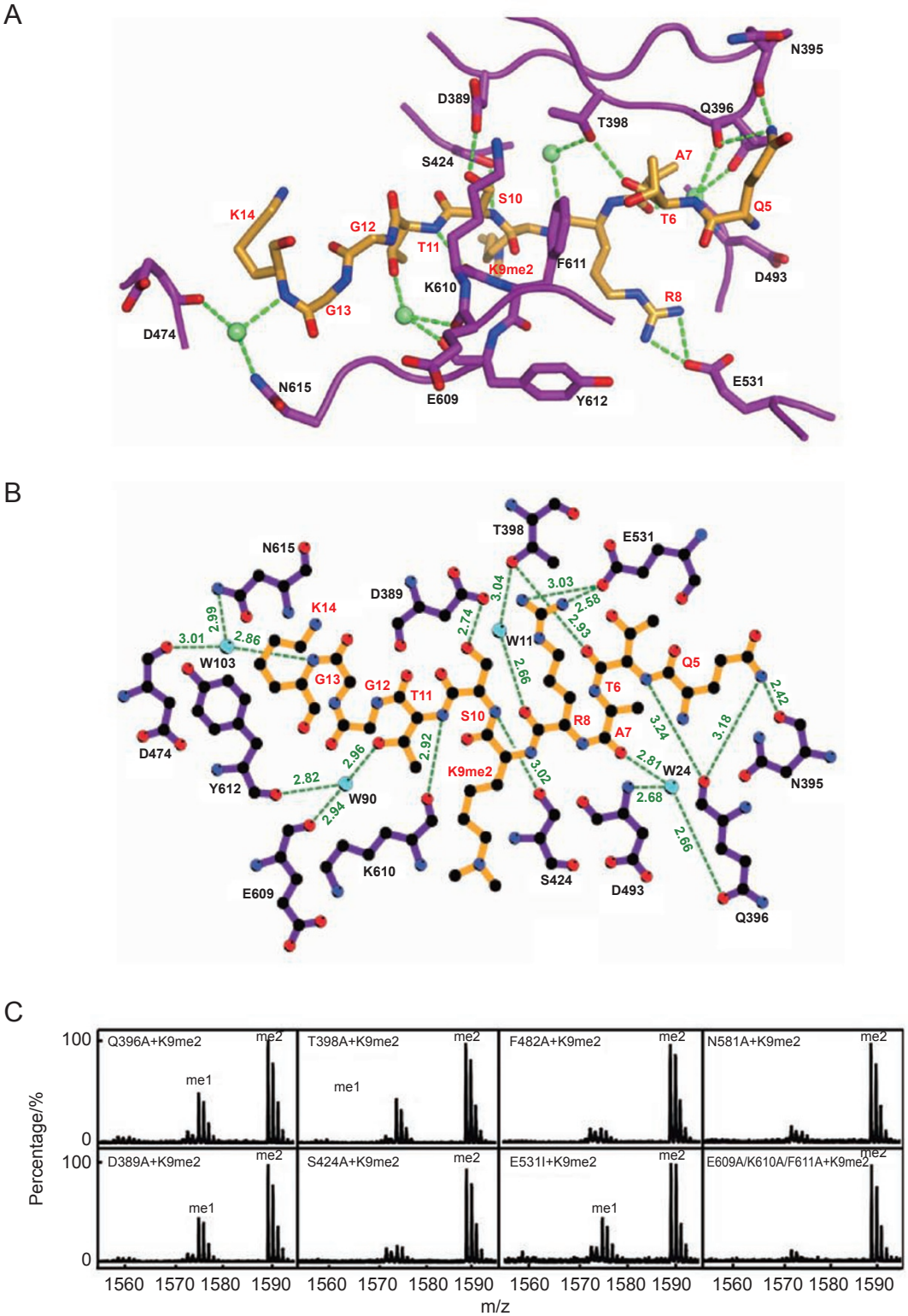


Figure 4 Substrate recognition of ceKDM7A. **(A)** Detailed interactions between ceKDM7A and a H3K9me2 peptide. Water molecules are shown as green balls. Hydrogen bonds are represented by green dotted lines. **(B)** Ligplot representation of the interactions between the enzyme and a H3K9me2 peptide. The carbon, oxygen, and nitrogen are colored in black, red, and blue, respectively. Lengths of hydrogen bond (dashed lines) are given in Å. **(C)** Mass spectrometry of demethylation reaction for various mutants of ceKDM7A. Please compare their activity to the wild type in Figure 3C. H3K9me2 peptide was used as the substrate.

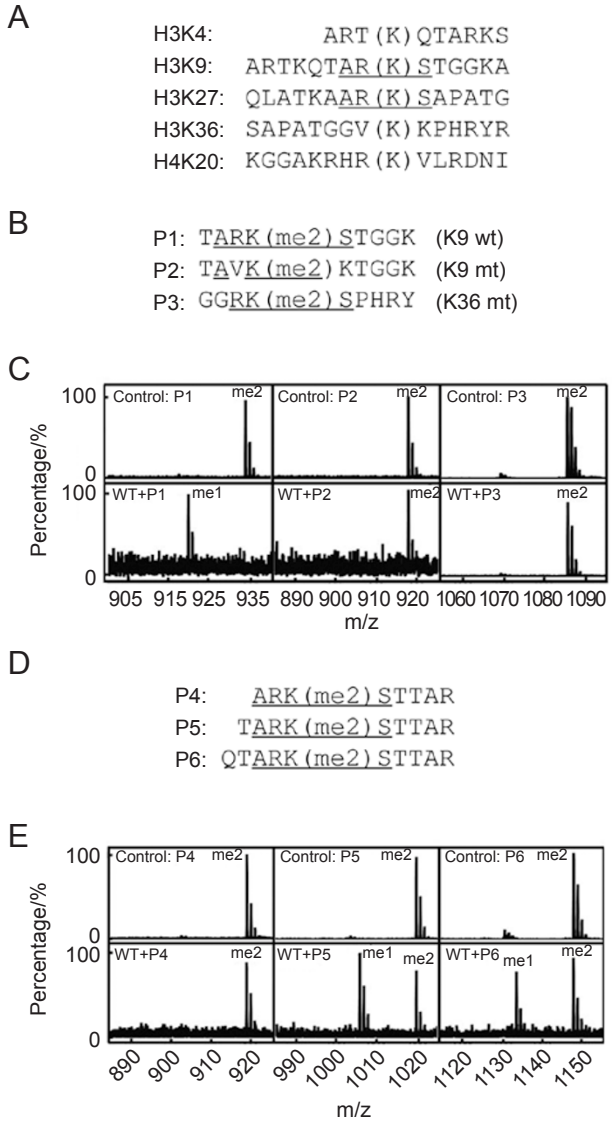


Figure 5 Substrate specificity determination. **(A)** Sequence alignment of methylated histone. Conserved residues between H3K9 and H3K27 are underlined. **(B)** Sequences of the swapped peptides used for demethylation assays. P1 is the wild-type H3K9me2 peptide, P2 is a mutated H3K9me2 peptide in which R8 and S10 were replaced by V35 and K37, as in H3K36me2 peptide, and P3 is a mutated H3K36me2 peptide in which V35 and K37 were replaced by R8 and S10 as in H3K9me2 peptide. **(C)** Mass spectrometry of demethylation reaction for the native and swapped peptides. **(D)** Sequences of the hybrid peptides used for demethylation assays. Underlined residues are from H3K9, and TTAR is from H3K4. **(E)** Mass spectrometry of demethylation reaction for the hybrid peptides.

hydrogen bonds (Figure 4A-4B and Supplementary information, Figure S4A-S4B). To test the hypothesis, we performed enzymatic activity assay using the following sequence-swapped peptides P1 (TARK$_{(me2)}$STGGK, the wild-type H3K9me2 peptide), P2 (TAVK$_{(me2)}$KTGGK, a mutated H3K9me2 peptide in which R8 and S10 were replaced by V35 and K37 as in H3K36me2 peptide), and P3 (GGRK$_{(me2)}$SPHRY, a mutated H3K36me2 peptide in which V35 and K37 were replaced by R8 and S10 as in H3K9me2 peptide) (Figure 5A). Mass spectrometric analysis showed that ceKDM7A had demethylase activity for P1, but not for P2 and P3 (Figure 5C). These results suggest that R8 and S10 are necessary but not sufficient for specificity determination.

To further study the substrate specificity, we performed enzymatic activity assay using hybrids of H3K4- and H3K9-flanking sequences as substrates. These peptides are P4 (ARK$_{(me2)}$STTAR, the first five residues flanking H3K4 were replaced by ARK$_{(me2)}$ST as in H3K9), P5 (TARK$_{(me2)}$STTAR, addition of a T to P4 peptide as in H3K9), and P6 (QTARK$_{(me2)}$STTAR, addition of a QT to P4 peptide as in H3K9) (Figure 5A and 5D). Mass spectrometric analysis showed that ceKDM7A had no demethylase activity for P4, but could remove the methyl groups from P5 and P6 (Figure 5E). These results suggest that, in addition to ARK$_{(me2)}$S, other surrounding residues involved in substrate binding are also important for substrate selection. These results are consistent with our structure data that residues G12 and G13 are not involved in enzyme-substrate interaction (Figure 4A-4B), and that the substrate sequence preference of ceKDM7A is different from JMJD2A [24, 27].

To study the molecular determinants responsible for the specificity for dimethylation, we measured distances from Cα and Nε of methylated lysine to C1 of NOG (Figure 6). The distance between Nε of methylated lysine and C1 of NOG in ceKDM7A complexes ranges from 3.05 to 4.29 Å and two methyl group positions away from the active site (Figure 6A). There is no space for one more methyl group to sit between. This distance cannot be extended because a lid composed of four residues D389, T398, K610, and F611 pushes K9/K27 deeply into the catalytic site, leaving only a short distance of 7.32 Å from Cα of methylated lysine to C1 of NOG (Figure 6A). These results suggest that the shallow binding pocket may determine the substrate specificity for dimethylated lysine. Consistently, the H3K4me2 demethylase LSD1 also has a shallow active site, even though it uses a different enzymatic mechanism, and its active site shares no sequence similarity to ceKDM7A and other JmjC domain-containing proteins. In the LSD1 complex with H3 peptide analog (K4 is replaced by Met) [38], the distance between Cα and the reactive N-5 of flavin is 7.69 Å (Figure 6B and Supplementary information, Figure S7). On the contrary, H3K9me3 and H3K36me3 demethylase

JMJD2A has a large active site. In JMJD2A complexes [24, 26-27], the trimethylated lysine-containing peptide resides in an open surface with side chain pointing to a deep pocket (10.54-11.08 Å) in a relaxed conformation (Figure 6B). The distance between Nε of methylated lysine and C1 of NOG ranges from 4.83 to 5.59 Å, which provides enough space for one (additional) methyl group. Detailed analyses for the active sites are summarized in Supplementary information, Table S3.

Peptide-binding mode by ceKDM7A

The cocrystal structure of ceKDM7A with a single peptide containing both H3K4me3 and H3K9me2 is highly similar to that with two separate peptides containing either H3K4me3 or H3K9me2, with a RMSD value of 0.225 Å for 448 Cα atoms. In the cocrystal structure, a stretch of peptide bound to the PHD finger contains residues 1 to 6 of histone H3, and another stretch of peptide bound to the JmjC domain contains residues 5 to 14 (Figure 1B). Because both stretches of peptides contain residues 5 and 6 of histone H3, these results indicate that the stretches of peptides should be from two separate histone H3 molecules. In addition, the Cα of H3K4 in the peptide associated with the PHD finger is 28.6 Å from the Cα of H3K9 in the peptide associated with the JmjC domain. Because the distance between the two Cα of K4 and K9 in one histone is 19 Å (3.8 Å × 5 residues) at the most extended conformation, these results further support the conclusion that the two stretches of peptides are not from the same molecule. The cocrystal structure of ceKDM7A with a single peptide containing both H3K4me3 and H3K27me2 is also highly similar to that with two separate peptides containing either H3K4me3 or H3K27me2, with a RMSD value of 0.366 Å for 407 Cα atoms. In these cocrystal structures, a stretch of peptide containing residues 1 to 6 of histone H3 binds to the PHD finger, and another stretch of peptide containing residues 23 to 32 of histone H3 binds to the JmjC domain (Supplementary information, Figures S1 and S4). However, the cocrystal structures did not allow us to determine whether the bound H3K4 peptide and H3K27 peptide are from one or two different histone H3 molecules, since the distance between the two Cα of K4 and K27 in one H3 histone (3.8 Å × 23 residues = 87.4 Å at the most extended conformation) is longer than 28.6 Å and the two bound peptides do not share common residues.

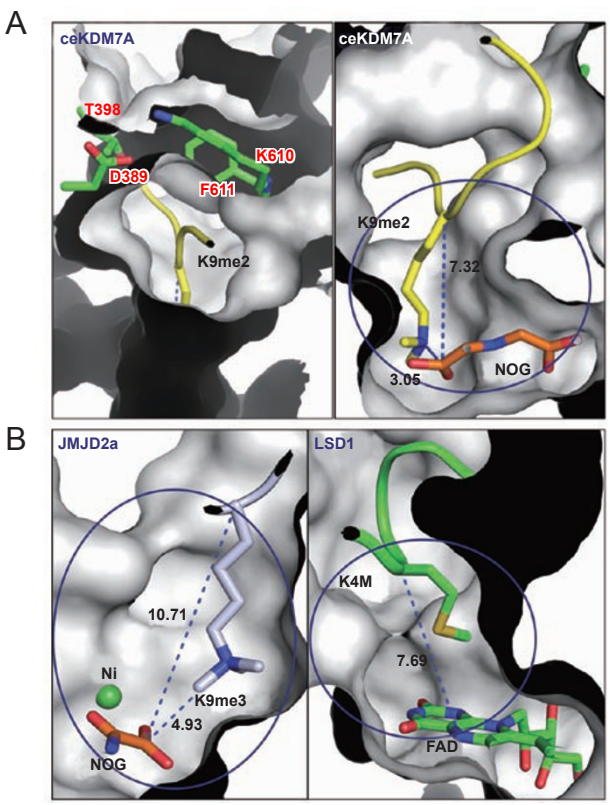


Figure 6 Comparison of the active sites among ceKDM7A (with H3K9me2), JMJD2A (with H3K9me3, 2OQ6.pdb), and LSD1 (with pK4M peptide, in which lysine is replaced by methionine, 1V1D.pdb). Distances are given in angstroms. Side chains of the residues for the lid formation in ceKDM7A structure are indicated in **A**.

Discussion

Up to now, more than 20 histone lysine demethylases have been identified that can remove methyl groups from histone H3 in a sequence- and methylation-state-specific manner [6, 11-23, 39]. These histone lysine demethylases can be divided into two groups with different catalytic mechanisms. One group has only two members LSD1 and LSD2 and can only remove di- and mono-methylation from H3K4 using an amine oxidase reaction [6, 39]. All others are JmjC domain-containing proteins that catalyze the demethylation by a hydroxylation reaction and require both iron and α-ketoglutarate as cofactors [7]. This group of proteins can remove methyl groups from many histone lysine residues such as H3K4, H3K9, H3K27, and H3K36 with mono-, di-, and trimethylation. Some of them can remove methyl groups from two different lysine residues and, therefore, are dual-specificity histone demethylases. The crystal structure of LSD1 has been solved [38, 40-42] and the molecular determination for the sequence and methylation-state specificity has been uncovered. Although the structure of JMJD2A has been solved in apo form or in complex with its substrates [24-27], which provides insights into the molecular

mechanism of catalysis and substrate recognition, how substrate specificity is determined for all other members of the JmjC domain-containing histone lysine demethylases remains to be elucidated.

Substrate specificity determined by primary sequence

Our structural and biochemical studies indicate that ceKDM7A binds both H3K9me2- and H3K27me2-containing peptide substrates in a similar fashion, and both main chains and side chains from the peptides are involved in substrate recognition (Figure 4A-4B and Supplementary Information, Figure S4A-S4B). Because both H3K9me2 and H3K27me2 are substrates for ceKDM7A, we hypothesized that the core substrate sequence for ceKDM7A is ARK$_{(me2)}$S, four residues shared by both substrates. Indeed, mutation of ARK$_{(me2)}$S to AVK$_{(me2)}$K abolished the activity, suggesting that R8 and S10 are necessary for specificity determination. However, ARK$_{(me2)}$STTAR, a hybrid of H3K4- and H3K9-flanking sequences, is not a substrate for ceKDM7A. These results suggest that other surrounding residues are also important for substrate selection. Indeed, addition of a T makes ARK$_{(me2)}$STTAR a substrate. These studies demonstrate that the primary sequence flanking the methylated lysine determines the substrate specificity, consistent with the proposed mechanism for JMJD2A [24, 26-27]. However, there are major differences for substrate recognition and specificity determination between ceKDM7A and JMJD2A. First, JMJD2A binds H3K9me3- and H3K36me3-containing peptide substrates in an opposite orientation, whereas ceKDM7A binds both H3K9me2- and H3K27me2-containing peptide substrates in a similar fashion with the same orientation. Second, although both proteins can remove methyl groups from H3K9, the secondary structural elements essential for substrate binding are highly variable (Supplementary information, Figure S5). Unlike ceKDM7A, JMJD2A achieves the substrate specificity through few polar side chain interactions and requires a bent peptide conformation. Third, two glycine residues G12 and G13 are required for substrate selection, which renders H3K9 the preferred substrate for JMJD2A [24, 27]. However, both glycine residues are dispensable for ceKDM7A substrates.

The structural features revealed by our studies of ceKDM7A complexes may be applicable to other JmjC domain-containing proteins. JHDM1A is a histone demethylase specific for H3K36me2 [7], and the crystal structure of the protein in apo form has been solved [43]. The JmjC domains of JHDM1A and ceKDM7A share high sequence homology and many critical residues for demethylase activity are conserved between these two proteins (Figure 1C). For example, residues D389, K610, and F611, which form the lid of the substrate-binding pocket and force the peptide into a V-shaped conformation, are identical between the two proteins. The lid pushes the substrates deeply into the catalytic site and is required for the demethylase activity (Supplementary information, Figure S6). These comparisons suggest that JHDM1A may use a similar mechanism for substrate recognition. However, there are some differences between the two structures. For example, T398 and E531, two important residues for substrate recognition in ceKDM7A, are replaced by G118 and L248 in JHDM1A, respectively. A β-strand (β6), an important secondary structure for peptide binding in ceKDM7A, is also different in the corresponding region in JHDM1A (Figure 1C). These differences may explain the different substrate specificity for these two proteins, but this remains to be determined in further studies.

Cross-regulation of histone demethylases

In this study, we solved the cocrystal structures of ceKDM7A using a fragment containing both PHD and JmjC domains. Structure of its apo form shows that the PHD and JmjC domains are two separate modules (Supplementary information, Figure S3A). Cocrystal structures of ceKDM7A with H3K4me3-containing peptides bound to the PHD finger and H3K9me2 or H3K27me2 peptides bound to the JmjC domain indicate that binding of the peptides does not cause significant conformational changes (Supplementary information, Figure S2A-S2C), suggesting that each domain can perform their biochemical functions independently. Consistently, a fragment containing only the PHD finger can bind to H3K4me3-containing peptide, and disrupting the interaction between the PHD finger and H3K4me3 did not affect the demethylase activity of the JmjC domain *in vitro* (Lin et al., accompanying paper in this issue [33]). However, the PHD finger binding to H3K4me3 is required for demethylation by the JmjC domain *in vivo*. Further analyses indicate that the PHD finger binding to H3K4me3 guides the demethylation by the JmjC domain in physiological conditions, indicating a cross-regulation between the two domains (Lin *et al*., accompanying paper in this issue [33]). Geometrical measurement of the structures revealed that H3K4me3 associated with the PHD finger and H3K9me2 bound to the JmjC domain are from two separate molecules, suggesting that the cross-regulation could be a *trans*-histone event. We speculate that this mode of cross-regulation may be applicable to other JmjC domain-containing histone lysine demethylases, since most of them contain domains predicted to be involved in protein or DNA interactions, in addition to the JmjC domain.

Structural comparison with PHF8 and KIAA1718

The crystal structures of PHF8 complexed with a peptide containing both H3K4me3 and H3K9me2, and the apo form of KIAA1718 were recently reported [44]. Through enzymatic and structural analysis, Horton *et al*., demonstrated that the presence of H3K4me3 in *cis*-form made H3K9me2 a better substrate for PHF8 and H3K9me2 a worse substrate for KIAA1718 *in vitro*, which was explained by the fact that PHF8 adopts a bent conformation between the PHD finger and the JmjC domain, while KIAA1718 uses an extended conformation. The apo form and five complex structures of ceKDM7A in our study demonstrate that ceKDM7A used an extended conformation similar to KIAA1718 and that the peptide binding did not cause conformation changes of the enzyme. It is not yet known whether the peptide binding causes the conformational change of PHF8, because the structure of apo form of PHF8 has not been solved. Structural comparison indicates that ceKDM7A is similar to PHF8 and KIAA1718, especially at the active site and in the substrate-binding mode. In general, ceKDM7A is more similar to KIAA1718, with a RMSD value of 2.064 Å for 340 Cα atoms, which is consistent with the sequence alignment (Figure 1C). This suggests that KIAA1718 probably will not make significant conformational changes upon histone peptide binding. Cocrystal structures of KIAA1718 with its substrate peptides will answer this question. Since ceKDM7A shares highly conserved catalytic core with PHF8 and KIAA1718, we speculate that PHF8 and KIAA1718 may have similar sequence and methylation-state specificity for H3K9me2 and H3K27me2 as ceKDM7A (Lin *et al*., accompanying paper in this issue [33]). Indeed, KIAA1718 could demethylate both H3K9me2 and H3K27me2, but PHF8 was nearly inactive on H3K27me2 [44], indicating that PHF8 has different substrate specificity than KIAA1718 and ceKDM7A.

Materials and Methods

Protein purification

All constructs were generated using PCR-based cloning strategy and all mutants were generated by the Quick-Change mutagenesis protocol (Stratagene) and verified by DNA sequencing. ceKDM7A (residues 188-711) and ceKDM7A mutants (residues 188-711) were subcloned into a pET-15b derivative encoding a 3C protease cleavage site. All constructs were transformed into and overexpressed in LB culture medium at 15 °C in *Escherichia coli* strain BL21 (DE3). His6-tagged proteins were purified by nickel nitrilotriacetic acid affinity chromatography followed by 3C protease cleavage. The proteins were purified to homogeneity using anion exchange and gel filtration chromatography. The purified proteins were concentrated to 15 mg/ml and used for crystallization and enzymatic activity assay.

Crystallization and data collection

Five peptides used for cocrystallization were H3K4me3 (ARTK$_{(me3)}$QTARKSTGGKA), H3K9me2 (ARTKQTARK$_{(me2)}$STGGKAPR), H3K27me2(QLATKAARK$_{(me2)}$SAPASGGV), H3K4me3K9me2(ARTK$_{(me3)}$QTARK$_{(me2)}$STGGKAPRKQLATKAARKSAPAS), and H3K4me3K27me2(ARTK$_{(me3)}$QTARKSTGGKAPRKQLATKAARK$_{(me2)}$SAPAS).

The six distinct complexes used for crystallization are listed below.

(1) ceKDM7A alone; (2) ceKDM7A complex with H3K4me3 peptide and NOG; (3) ceKDM7A complex with H3K4me3 peptide, H3K9me2 peptide, and NOG; (4) ceKDM7A complex with H3K4me3 peptide, H3K27me2 peptide, and NOG; (5) ceKDM7A complex with H3K4me3K9me2 peptide and NOG; (6) ceKDM7A complex with H3K4me3K27me2 peptide and NOG.

For crystallization with peptides, 2 mM peptides were added to the protein solution before crystallization. NOG was added to a final concentration of 3 mM before setting up drops. Crystals with good diffracting quality for ceKDM7A and ceKDM7A with peptides were obtained after incubating for 5 to 10 days at 4 °C. Diffracting crystals were grown by hanging-drop vapor-diffusion method by mixing the protein (~15 mg/ml) with an equal volume of reservoir solution containing 18% PEG10000, 0.1 M Bis-Tris, pH 6.6, 0.2 M Sodium Formate for crystals without peptides, and 16% PEG3350, 0.2 M sodium fluoride, 0.1 M Bis-Tris, or HEPES with pH ranging from 6.6 to 7.4 for crystals with various peptides. Small crystals appeared within 2 days. Microseeding and streak seeding were used to generate single large crystals. The crystals of ceKDM7A without peptides belong to the space group $P2_1$, with a= 66.0 Å, b = 144.5 Å, c=78.4 Å, and β = 106.7 °. The crystals of ceKDM7A with H3k4me3K27me2 peptide belong to the space group $P2_1$, with a= 60.1 Å, b = 86.4 Å, c=62.7 Å, and β = 113.2 °. The crystals of ceKDM7A with other peptides belong to the space group $P2_12_12_1$, with a= 68.4 Å, b = 85.5 Å, and c= 102.6 Å. Crystals were slowly equilibrated with a cryoprotectant buffer containing reservoir buffer plus 20% glycerol (v/v) and were flash frozen in a cold nitrogen stream at –173 °C. Crystals were examined on X8 Proteum system (Bruker AXS) and data sets were collected on beamline BL17U at SSRF (Shanghai, China). All data were processed using the program HKL2000 [45].

Structure determination

The structure of ceKDM7A was determined by molecular replacement using the JHDM1 (2YU1.pdb) as a searching model [43], against an initial 2.5 Å native data set in $P2_1$ form. The crystals contain two molecules in one asymmetric unit. Rotation and translation function searches were performed with the program PHASER [46]. The structure of ceKDM7A with peptides were determined by difference Fourier method and the models were manually built with COOT [47]. All refinements were performed using the refinement module phenix.refine of PHENIX package [48]. The model quality was checked with the PROCHECK program [49], which showed good stereochemistry according to the Ramachandran plot for all structures. The structure similarity search was performed with DALI Server [50] and structure superimposition was performed with COOT [47]. All structure figures

were generated by PyMol [51].

MALDI-TOF analysis of enzymatic activity

In demethylase activity assays, purified protein or mutants (3 μM) were incubated with peptide (30 μM) in buffer containing 30 mM Tris-HCl (pH 7.4), 0.15 M NaCl, 50 μM $(NH_4)_2Fe(SO_4)_2$, 1 mM 2-oxoglutarate, and 2 mM ascorbate. After 60 min at 37 °C, 10 μl of the demethylation reaction was desalted through a C18 Ziptip (Millipore) according to the instructions from manufacturer. The bound peptides were eluted using 2 μl buffer containing 70% acetonitrile and 0.1% TFA and then spotted directly onto the target. Samples were analyzed on a MALDI-TOF micro MX mass spectrometer (ABI 4700). In demethylase activity assays for swapped peptides and hybrid peptides, wild-type ceKDM7A protein with 10 μM and peptides with 50 μM were used. The six peptides are P1 ($TARK_{(me2)}STGGK$), P2 ($TAVK_{(me2)}KTGGK$), P3 ($GGRK_{(me2)}SPHRY$), P4 ($ARK_{(me2)}STTAR$), P5 ($TARK_{(me2)}STTAR$), and P6 ($QTARK_{(me2)}STTAR$).

Isothermal titration calorimetry

To obtain binding affinity between ceKDM7A and histone tail modifications, purified ceKDM7A or mutants with 100 μM were titrated against various peptides (H3K4me3, H3K4me2, H3K9me3, H3K9me2, H3K27me3, and H3K27me2) with 1 mM concentration using VP-isothermal titration calorimetry microcalorimeter (MicroCal) at 10 °C. All proteins and peptides were prepared in a buffer containing 10 mM HEPES (pH 8.0) and 0.1 M NaCl. The data were fitted by software Origin 7.0.

Accession codes

The coordinates and structure factors have been deposited with accession numbers 3N9L (for ceKDM7A complex with H3K4me3 peptide and NOG), 3N9M (for ceKDM7A alone), 3N9N (for ceKDM7A complex with H3K4me3K9me2 peptide and NOG), 3N9O (for ceKDM7A complex with H3K4me3 peptide, H3K9me2 peptide and NOG), 3N9P (for ceKDM7A complex with H3K4me3K27me2 peptide and NOG), 3N9Q (for ceKDM7A complex with H3K4me3 peptide, H3K27me2 peptide and NOG).

Acknowledgments

We thank Yang Shi (Harvard Medical School, USA) and Ruiming Xu (Institute of Biophysics, CAS, China) for critical reading of the manuscript, staff members of beamline BL17U at SSRF (China) and BL17A, NW12 at Photon Factory (Japan) for data collection, the cell biology and molecular biology core facilities for MS and biacore experiments, and other members in Chen lab and Xu lab for technical help. This work was supported by grants from National Basic Research Program of China (2010CB529700, 2009CB918600, and 2007CB947900), the National Natural Science Foundation of China (30870493, 90919026), Chinese Academy of Sciences (KSCX2-YW-R-04), Shanghai Pujiang Program (08PJ14010, 0757S11361), Shanghai Leading Academic Discipline Project (B111), and the Council of Shanghai Municipal Government for Science and Technology.

References

1 Agger K, Christensen J, Cloos PA, Helin K. The emerging functions of histone demethylases. *Curr Opin Genet Dev* 2008; **22**:1115-1140.
2 Bhaumik SR, Smith E, Shilatifard A. Covalent modifications of histones during development and disease pathogenesis. *Nat Struct Mol Biol* 2007; **14**:1008-1016.
3 Kouzarides T. Chromatin modifications and their function. *Cell* 2007; **128**:693-705.
4 Shi Y. Histone lysine demethylases: emerging roles in development, physiology and disease. *Nat Rev Genet* 2007; **8**:829-833.
5 Martin C, Zhang Y. The diverse functions of histone lysine methylation. *Nat Rev Mol Cell Biol* 2005; **6**:838-849.
6 Shi Y, Lan F, Matson C, *et al.* Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* 2004; **119**:941-953.
7 Tsukada Y, Fang J, Erdjument-Bromage H, *et al*. Histone demethylation by a family of JmjC domain-containing proteins. *Nature* 2006; **439**:811-816.
8 Klose RJ, Kallin EM, Zhang Y. JmjC-domain-containing proteins and histone demethylation. *Nat Rev* 2006; **7**:715-727.
9 Shi Y, Whetstine JR. Dynamic regulation of histone lysine methylation by demethylases. *Mol Cell* 2007; **25**:1-14.
10 Xiang Y, Zhu Z, Han G, *et al.* JARID1B is a histone H3 lysine 4 demethylase up-regulated in prostate cancer. *Proc Natl Acad Sci USA* 2007; **104**:19226-19231.
11 Lee MG, Villa R, Trojer P, *et al.* Demethylation of H3K27 regulates polycomb recruitment and H2A ubiquitination. *Science* 2007; **318**:447-450.
12 Lee MG, Norman J, Shilatifard A, Shiekhattar R. Physical and functional association of a trimethyl H3K4 demethylase and Ring6a/MBLR, a polycomb-like protein. *Cell* 2007; **128**:877-887.
13 Xiang Y, Zhu Z, Han G, Lin H, Xu L, Chen CD. JMJD3 is a histone H3K27 demethylase. *Cell Res* 2007; **17**:850-857.
14 Lan F, Bayliss PE, Rinn JL, *et al*. A histone H3 lysine 27 demethylase regulates animal posterior development. *Nature* 2007; **449**:689-694.
15 Christensen J, Agger K, Cloos PA, *et al.* RBP2 belongs to a family of demethylases, specific for tri- and dimethylated lysine 4 on histone 3. *Cell* 2007; **128**:1063-1076.
16 Cloos PA, Christensen J, Agger K, *et al.* The putative oncogene GASC1 demethylates tri- and dimethylated lysine 9 on histone H3. *Nature* 2006; **442**:307-311.
17 Iwase S, Lan F, Bayliss P, *et al.* The X-linked mental retardation gene SMCX/JARID1C defines a family of histone H3 lysine 4 demethylases. *Cell* 2007; **128**:1077-1088.
18 Klose RJ, Yamane K, Bae Y, *et al.* The transcriptional repressor JHDM3A demethylates trimethyl histone H3 lysine 9 and lysine 36. *Nature* 2006; **442**:312-316.
19 Klose RJ, Yan Q, Tothova Z, *et al.* The retinoblastoma binding protein RBP2 is an H3K4 demethylase. *Cell* 2007; **128**:889-900.
20 Whetstine JR, Nottke A, Lan F, *et al.* Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases. *Cell* 2006; **125**:467-481.
21 Yamane K, Tateishi K, Klose RJ, *et al.* PLU-1 is an H3K4 demethylase involved in transcriptional repression and breast cancer cell proliferation. *Mol Cell* 2007; **25**:801-812.
22 Yamane K, Toumazou C, Tsukada Y, *et al*. JHDM2A, a

JmjC-containing H3K9 demethylase, facilitates transcription activation by androgen receptor. *Cell* 2006; **125**:483-495.
23 Agger K, Cloos PA, Christensen J, *et al.* UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. *Nature* 2007; **449**:731-734.
24 Chen Z, Zang J, Kappler J, *et al.* Structural basis of the recognition of a methylated histone tail by JMJD2A. *Proc Natl Acad Sci USA* 2007; **104**:10818-10823.
25 Chen Z, Zang J, Whetstine J, *et al.* Structural insights into histone demethylation by JMJD2 family members. *Cell* 2006; **125**:691-702.
26 Couture JF, Collazo E, Ortiz-Tello PA, Brunzelle JS, Trievel RC. Specificity and mechanism of JMJD2A, a trimethyllysine-specific histone demethylase. *Nat Struct Mol Biol* 2007; **14**:689-695.
27 Ng SS, Kavanagh KL, McDonough MA, *et al*. Crystal structures of histone demethylase JMJD2A reveal basis for substrate specificity. *Nature* 2007; **448**:87-91.
28 Shi X, Hong T, Walter KL, *et al.* ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. *Nature* 2006; **442**:96-99.
29 Wysocka J, Swigut T, Xiao H, *et al.* A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* 2006; **442**:86-90.
30 Huang Y, Fang J, Bedford MT, Zhang Y, Xu RM. Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A. *Science* 2006; **312**:748-751.
31 Patsialou A, Wilsker D, Moran E. DNA-binding properties of ARID family proteins. *Nucleic Acids Res* 2005; **33**:66-80.
32 Huang C, Xiang Y, Wang Y, *et al.* Dual-specificity histone demethylase KIAA1718 (KDM7A) regulates neural differentiation through FGF4. *Cell Res* 2010; **20**:154-165.
33 Lin H, Wang Y, Wang Y, *et al*. Coordinated regulation of active and repressive histone methylations by a dual specificity histone demethylase ceKDM7A from *C. elegans*. *Cell Res* 2010; **20**:899-907.
34 Bienz M. The PHD finger, a nuclear protein-interaction domain. *Trends Biochem Sci* 2006; **31**:35-40.
35 Pascual J, Martinez-Yamout M, Dyson HJ, Wright PE. Structure of the PHD zinc finger from human Williams-Beuren syndrome transcription factor. *J Mol Biol* 2000; **304**:723-729.
36 Li H, Ilin S, Wang W, *et al.* Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature* 2006; **442**:91-95.
37 Pena PV, Davrazou F, Shi X, *et al.* Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. *Nature* 2006; **442**:100-103.
38 Forneris F, Binda C, Adamo A, Battaglioli E, Mattevi A. Structural basis of LSD1-CoREST selectivity in histone H3 recognition. *J Biol Chem* 2007; **282**:20070-20074.
39 Karytinos A, Forneris F, Profumo A, *et al*. A novel mammalian flavin-dependent histone demethylase. *J Biol Chem* 2009; **284**:17775-17782.
40 Chen Y, Yang Y, Wang F, *et al.* Crystal structure of human histone lysine-specific demethylase 1 (LSD1). *Proc Natl Acad Sci USA* 2006; **103**:13956-13961.
41 Stavropoulos P, Blobel G, Hoelz A. Crystal structure and mechanism of human lysine-specific demethylase-1. *Nat Struct Mol Biol* 2006; **13**:626-632.
42 Yang M, Gocke CB, Luo X, *et al*. Structural basis for CoREST-dependent demethylation of nucleosomes by the human LSD1 histone demethylase. *Mol Cell* 2006; **23**:377-387.
43 Han Z, Liu P, Gu L, *et al*. Structural basis for histone demethylation by JHDM1. *Front Science* 2007; **1**:52-61.
44 Horton JR, Upadhyay AK, Qi HH, Zhang X, Shi Y, Cheng X. Enzymatic and structural insights for substrate specificity of a family of jumonji histone lysine demethylases. *Nat Struct Mol Biol* 2010; **17**:38-43.
45 Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 1997; **276**:307-326.
46 McCoy AJ, Grosse-Kunstleve RW, Storoni LC, Read RJ. Likelihood-enhanced fast translation functions. *Acta crystallogr* 2005; **61**:458-464.
47 Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta crystallogr* 2004; **60**:2126-2132.
48 Adams PD, Grosse-Kunstleve RW, Hung LW, *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta crystallogr* 2002; **58**:1948-1954.
49 Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 1993; **26**:283-291.
50 Holm L, Kaariainen S, Rosenstrom P, Schenkel A. Searching protein structure databases with DaliLite v.3. *Bioinformatics* 2008; **24**:2780-2781.
51 DeLano WL. The PyMOL Molecular Graphics System (2002). *on World Wide Web* http://www.pymol.org.

(**Supplementary information** is linked to the online version of the paper on the *Cell Research* website.)