

Molecular basis for the substrate specificity and catalytic mechanism of thymine-7-hydroxylase in fungi

Wenjing Li¹, Tianlong Zhang¹ and Jianping Ding^{1,2,*}

¹National Center for Protein Science Shanghai, State Key Laboratory of Molecular Biology, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 320 Yue-Yang Road, Shanghai 200031, China and ²Collaborative Innovation Center for Genetics and Development, Fudan University, 2005 Song-Hu Road, Shanghai 200438, China

Received June 29, 2015; Revised September 16, 2015; Accepted September 18, 2015

ABSTRACT

TET proteins play a vital role in active DNA demethylation in mammals and thus have important functions in many essential cellular processes. The chemistry for the conversion of 5mC to 5hmC, 5fC and 5caC catalysed by TET proteins is similar to that of T to 5hmU, 5fU and 5caU catalysed by thymine-7-hydroxylase (T7H) in the nucleotide anabolism in fungi. Here, we report the crystal structures and biochemical properties of *Neurospora crassa* T7H. T7H can bind the substrates only in the presence of co-substrate, and binding of different substrates does not induce notable conformational changes. T7H exhibits comparable binding affinity for T and 5hmU, but 3-fold lower affinity for 5fU. Residues Phe292, Tyr217 and Arg190 play critical roles in substrate binding and catalysis, and the interactions of the C5 modification group of substrates with the cosubstrate and enzyme contribute to the slightly varied binding affinity and activity towards different substrates. After the catalysis, the products are released and new cosubstrate and substrate are reloaded to conduct the next oxidation reaction. Our data reveal the molecular basis for substrate specificity and catalytic mechanism of T7H and provide new insights into the molecular mechanism of substrate recognition and catalysis of TET proteins.

INTRODUCTION

Methylation at the C5 position of cytosine in DNA, which mainly exists in the CpG dinucleotide context in mammals, is an important epigenetic modification. Proper establishment and maintenance of the DNA methylation patterns are crucial for many important cellular processes, including embryonic development, genomic imprinting, X-chromosome inactivation and chromosome sta-

bility (1–4). The ten-eleven translocation (TET) proteins are a group of Fe²⁺ and α -ketoglutarate (α -KG) dependent dioxygenases which can catalyse the conversion of 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) in DNA (5–9). The final product 5caC can be recognized and excised by thymine-DNA glycosylase (TDG) and subsequently substituted by unmodified cytosine via base excision repair (BER) pathway (9,10). This process is currently believed to be an active DNA demethylation pathway. Thus, the TET proteins are proposed to play critical roles in epigenetic regulation (11–14), and aberrant expression, abnormal regulation and mutations of TET proteins are associated with many serious human diseases including cancers (13,15).

The crystal structures of the catalytic domain of *Homo sapiens* TET2 (HsTET2) and the full-length *Naegleria gruberi* TET-like dioxygenase 1 (NgTET1) in complexes with 5mC-containing DNA have been reported recently (16,17). In these structures, the base to be oxidized is flipped out of the double-stranded DNA (dsDNA) and inserted into the active site of HsTET2 or NgTET1. In the HsTET2-DNA and NgTET1-DNA complexes, the C5-methyl group of 5mC is not specifically recognized by any residues but has van der Waals contacts with the C1-carboxylate of NOG and the side chains of two surrounding residues (16,17). Nevertheless, how the TET proteins recognize or distinguish different modification groups on the C5 position of cytosine and catalyse the consecutive oxidations of 5mC to 5hmC, 5fC and 5caC are not well understood.

Intriguingly, the chemical characteristics of the successive oxidations of 5mC to 5hmC, 5fC and 5caC in mammals share great similarities with that of the successive oxidations of thymine (T or 5-methyluracil) to 5-hydroxymethyluracil (5hmU), 5-formyluracil (5fU) and 5-carboxyluracil (5caU), an essential process in the thymidine salvage pathway of nucleotide anabolism in some fungi. The thymidine salvage pathway, which was first discovered in *Neurospora crassa*, consists of a series of catalytic reactions (18). After the hydrolysis of thymine ribonucleoside to ribose and thymine

*To whom correspondence should be addressed. Tel: +86 21 5492 1619; Fax: +86 21 5492 1116; Email: jpding@sibcb.ac.cn

(19), the consecutive oxidations of thymine to 5hmU, 5fU and 5caU are catalysed by thymine-7-hydroxylase (T7H) (20–23), and the subsequent decarboxylation of 5caU to U is catalysed by isoorotate decarboxylase (IDCase) (24–26). Like the TET proteins, T7H is also a Fe^{2+} and α -KG dependent dioxygenase (23). Previous biochemical studies have characterized the enzymatic properties and revealed some insights into the catalytic mechanism of T7H (27–29). However, the structural and molecular basis for the substrate recognition and binding and the catalytic mechanism of T7H remains elusive.

Here we report the crystal structures of *N. crassa* T7H (NcT7H) in apo form, in α -KG-bound form and in complexes with α -KG and T, 5hmU or 5fU. Our structural and biochemical data together reveal the molecular basis for how T7H recognizes different substrates and catalyses the consecutive oxidations and provide new insights into the molecular mechanism of the substrate recognition and catalysis of the TET proteins.

MATERIALS AND METHODS

Cloning, expression and purification

The genes encoding the full-length NcT7H (residues 1–333) and the C-terminal truncated NcT7H (residues 1–299, NcT7H Δ C) were amplified by polymerase chain reaction from the genomic DNA of *N. crassa*. Each gene was cloned into the pET-28a vector (Novagen) with a His₆-tag at the C-terminus. The constructed plasmids were transformed into *Escherichia coli* BL21(DE3) Condon Plus strain (Novagen). The bacterial cells were grown in LB medium supplemented with 0.05 mg/ml kanamycin at 37°C to OD₆₀₀ of 0.6, and then induced with 0.2 mM IPTG at 16°C overnight. The cells were collected, resuspended and lysed on ice by sonication in the lysis buffer (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 5% glycerol, 2 mM β -mercaptoethanol, and 1 mM phenylmethylsulfonyl fluoride). The cell debris was precipitated by centrifugation at 18 000 g, and the supernatant was collected for protein purification.

The proteins were purified by affinity chromatography using a Ni-NTA column (Qiagen) with the storage buffer (20 mM Tris-HCl, pH 8.0, and 150 mM NaCl) supplemented with 20 mM imidazole and 250 mM imidazole serving as the washing buffer and the elution buffer, respectively. The proteins were further purified with gel filtration using a Superdex G75 16/60 column (GE Healthcare). Expression and purification of the Se-Met substituted NcT7H Δ C followed the same procedures as the native protein except that the bacterial cells were grown in M9 medium containing amino acids Lys, Thr, Phe, Leu, Ile, Val, Se-Met, and 1% lactose. Constructs of the NcT7H mutants containing point mutations were generated using the QuikChange Site-Directed Mutagenesis kit (Stratagene) and verified by sequencing. Expression and purification of the mutants were the same as the wild-type protein. The purified proteins were of high purity (above 95%) as analysed by SDS-PAGE.

Crystallization, data collection and structure determination

Crystallization was performed using the sitting drop vapour diffusion method at 16°C. Crystals of Se-Met NcT7H Δ C

in apo form were grown in drops consisting of an equal volume of the protein solution (30 mg/ml) and the reservoir solution [0.2 M $(\text{NH}_4)_2\text{SO}_4$, 0.1 M MES, pH 6.5, and 30% (w/v) PEGMME 5000]. In order to obtain catalytically inactive structures of NcT7H in complexes with different substrates, the protein solution was supplemented with 2 mM NiCl_2 to mimic Fe^{2+} . Crystals of NcT7H in complex with α -KG were grown in drops consisting of the protein solution supplemented with α -KG (1:4 molar ratio) and the reservoir solution (0.1 M BIS-TRIS, pH 5.5, and 25% (w/v) PEG 3,350). Crystals of NcT7H in complexes with α -KG and T, 5hmU or 5fU (Sigma) were grown in drops containing the protein solution supplemented with α -KG and the substrate (1:4:8 molar ratio) and the same reservoir solution. The crystals were cryoprotected using the reservoir solution supplemented with 25% ethylene glycol and then flash-cooled into liquid N_2 . Diffraction data were collected at 100 K at BL17U of Shanghai Synchrotron Radiation Facility and BL19U1 of National Facility for Protein Science Shanghai, and processed with HKL2000 (30). The statistics of the diffraction data are summarized in Table 1.

The apo NcT7H Δ C structure was solved at 2.3 Å resolution using the single-wavelength anomalous dispersion (SAD) method as implemented in Phenix (31), which identified 4 Se atoms with a figure-of-merit of 0.40 and produced an interpretable electron density map. All of the other NcT7H structures were solved by the molecular replacement (MR) method using the apo NcT7H Δ C structure as the search model. Model building was performed with Coot (32) and structure refinement was carried out using Phenix (31) and Refmac5 (33). Structural analysis was carried out using programs in CCP4 (34). The stereochemical geometry and quality of the structure models were analysed using Procheck (35). All graphics were generated using Pymol (www.pymol.org). The statistics of the refinement and the final structure models are summarized in Table 1.

Enzymatic activity assay

The apparent enzymatic activities of the wild-type and mutant NcT7H and the wild-type NcT7H Δ C to catalyse the consecutive oxidations of T to 5hmU, 5fU and 5caU were analysed using the HPLC method, which was modified based on the method used in the enzymatic activity assay of IDCase (26). The reaction solution (100 μ l) consisted of 5 μ M (for T and 5hmU) or 2 μ M (for 5fU) enzyme, 2 mM α -KG, and 2 mM substrate in the reaction buffer [20 mM K_3PO_4 , 1 mM $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$, and 2 mM ascorbic acid]. The reaction was carried out at 37°C for 1 h and then stopped by heating the solution to 100°C for 3 min. The reaction mixture was analysed using an Agilent 1200 HPLC instrument (Agilent Technologies) with an AQ-C18 column (5 μ m particle size, 25 cm \times 4.6 mm). The mobile phase was 20 mM NH_4OAc (pH 3.5) running at the rate of 0.5 ml/min, and the detector was set at 260 nm. Standard T, 5hmU, 5fU, and 5caU were used as references. T, 5hmU, 5fU, and 5caU have distinct retention times and their contents in the reaction mixture were semi-quantified based on integrations of the respective retention peaks. The apparent enzymatic activity was represented by the fraction of the substrate oxidized.

Table 1. Summary of diffraction data and refinement statistics

	NcT7H Δ C	NcT7H-AKG	NcT7H-T	NcT7H-5hmU	NcT7H-5fU
Data collection					
Wavelength (Å)	0.9792	0.9792	0.9792	0.9792	0.9199
Space group	$P6_3$	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Cell dimensions					
<i>a</i> (Å)	119.1	57.2	56.9	56.8	56.8
<i>b</i> (Å)	119.1	154.7	156.3	155.8	155.6
<i>c</i> (Å)	56.6	75.4	76.0	76.0	76.0
α (°)	90	90	90	90	90
β (°)	90	90.6	91.8	91.7	91.7
γ (°)	120	90	90	90	90
Resolution (Å)	50.0–2.30 (2.38–2.30) ^a	50.0–2.10 (2.18–2.10)	50.0–2.05 (2.12–2.05)	50.0–2.35 (2.43–2.35)	50.0–2.15 (2.23–2.15)
Observed reflections	217 111	249 555	314 882	166 072	290 491
Unique reflections ($I/\sigma(I) > 0$)	20 591 (2062)	75 367 (7543)	82 717 (8246)	53 781 (5427)	69 438 (6848)
$I/\sigma(I)$	29.2 (5.0)	17.4 (3.2)	16.6 (3.0)	19.3 (10.2)	16.1 (3.7)
Completeness (%)	100.0 (100.0)	99.3 (99.9)	99.9 (100.0)	98.3 (99.4)	98.5 (97.1)
Redundancy	10.5 (10.6)	3.3 (3.3)	3.8 (3.8)	3.1 (3.1)	4.2 (4.1)
R_{merge} (%) ^b	10.2 (54.0)	7.9 (40.6)	8.5 (46.8)	6.9 (12.6)	12.3 (59.4)
Refinement					
Resolution (Å)	50.0–2.30	50.0–2.10	50.0–2.05	50.0–2.35	50.0–2.15
No. reflections					
Working set	19 527	71 601	78 516	51 062	66 062
Test set	1053	3731	4130	2689	3343
$R_{\text{work}}/R_{\text{free}}$ (%) ^c	19.6 / 22.4	18.5 / 22.2	18.4 / 21.5	18.1 / 22.5	18.4 / 22.8
No. atoms					
Protein	2049	10 138	10 183	10 140	10 145
Cosubstrate	-	40	40	40	40
Substrate	-	-	36	40	40
Water	109	495	593	557	569
B-factors					
Protein	50.4	46.5	35.8	30.7	28.0
Cosubstrate	-	49.2	30.0	26.2	25.0
Substrate	-	-	37.1	39.7	34.9
Ion	32.7	43.4	27.7	22.3	22.1
Water	54.0	54.6	40.1	32.7	32.0
R.m.s. deviations					
Bond lengths (Å)	0.005	0.006	0.005	0.005	0.006
Bond angles (°)	1.1	1.1	1.0	1.0	1.1
Ramachandran plot (%)					
Most favoured	92.3	92.0	91.8	91.4	91.4
Allowed	7.2	7.7	7.9	8.2	8.3
Generously allowed	0.5	0.3	0.4	0.4	0.4

^aNumbers in parentheses represent the highest resolution shell.

^b $R_{\text{merge}} = \sum_{\text{hkl}} \sum_i |I_i(\text{hkl}) - \langle I(\text{hkl}) \rangle| / \sum_{\text{hkl}} \sum_i I_i(\text{hkl})$.

^c $R = \sum_{\text{hkl}} \|F_o - F_c\| / \sum_{\text{hkl}} F_o$.

To determine the kinetic parameters of T, 5hmU or 5fU oxidation by NcT7H, we used the SCS/PK/LDH-coupled system to monitor the production of succinate (36). Reaction was carried out at 25°C in a buffer containing 100 mM Hepes (pH 7.5), 1 mM α -KG, 0.2 mM ascorbic acid, 0.2 mM $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$, 0.1 mM MgCl_2 , 0.4 mM ATP, 0.4 mM coenzyme A (CoA), 1 mM phosphoenolpyruvate, 0.4 mM NADH and varied concentrations of substrate (T, 5hmU or 5fU). The coupling enzymes succinyl-CoA synthetase (SCS, 5 μM , expressed and purified as His₆-tag fused protein) and pyruvate kinase (PK)/lactate dehydrogenase (LDH) (Sigma, 6–10 units of PK and 9–14 units of LDH), along with the wild-type or mutant NcT7H (1–20 μM), were added to the reaction mixture to a final volume of 500 μl . The reaction was initiated by addition of the wild-type or mutant NcT7H, and the enzymatic activity was measured by monitoring the rate of NADH oxidation, which has an extinction coefficient of 6220 $\text{M}^{-1}\text{cm}^{-1}$ at 340

nm, using a Beckman DU800 spectrophotometer (Beckman Coulter). All kinetic data were fitted to the Michaelis–Menten equation using Prism 5.0 (Graphpad Software). All experiments were performed in triplicates.

Isothermal titration calorimetry analysis

The thermodynamic parameters of the wild-type and mutant NcT7H with different ligands were measured using the isothermal titration calorimetry (ITC) method with an ITC200 Micro-calorimeter (MicroCal) at 25°C. Prior to titrations with the substrates, NcT7H was incubated with NOG or succinate at 1:2 molar ratio. In all experiments, the initial injection of 0.4 μl of the ligand solution was discarded to eliminate the effect of titrant diffusion across the syringe tip during the equilibration process, and each dataset consisted of 20 injections of 2 μl each of 2 mM α -KG, NOG, succinate, U, T, 5hmU, 5fU or 5caU into the sample cell containing 250 μl of 0.2 mM NcT7H. The

heats of dilution were negligible in all cases. Binding constants and other thermodynamic parameters were determined by fitting the integrated titration data using a single binding site model by a nonlinear least-squares method implemented in MicroCal Origin software version 7.0.

RESULTS

Overall structure of NcT7H

Crystallization of the full-length NcT7H (residues 1–333) alone yielded no crystal. Crystallization of a C-terminal truncated NcT7H (residues 1–299, NcT7H Δ C) alone yielded the apo-form crystals. Crystallization of the full-length NcT7H in the presence of α -KG or α -KG and the substrate (T, 5hmU or 5fU) yielded crystals of the α -KG-bound or the substrate-bound complexes, respectively. The Crystal structure of the apo NcT7H Δ C was solved by the SAD method at 2.3 Å resolution (Table 1). The asymmetric unit contains one NcT7H Δ C molecule, which is well defined except four surface exposed regions (residues 95–116, residues 198–207, residues 278–282, and residues 296–299). There is a metal ion bound at the active site, which is provisionally interpreted as calcium as the ICP-AES (inductively coupled plasma atomic emission spectroscopy) analysis revealed that calcium is the most abundant metal in the protein solution (Supplementary Table S1) and the structure refinement yielded a reasonable B-factor (Table 1). Since no divalent metal ion was added in the purification and the crystallization solution, the bound Ca²⁺ is presumably co-purified with the enzyme.

The crystal structures of the full-length NcT7H in complex with α -KG (NcT7H-AKG), and in complexes with α -KG and T (NcT7H-T), with α -KG and 5hmU (NcT7H-5hmU), and with α -KG and 5fU (NcT7H-5fU) were solved by the MR method at 2.1 Å, 2.05 Å, 2.35 Å, and 2.15 Å resolution, respectively (Table 1). In all these structures, the asymmetric unit contains four NcT7H molecules, of which two molecules consist of the whole polypeptide chain and the other two have a disordered region (approximately residues 94–114) due to differed crystal packing environments. Comparison of the four NcT7H molecules shows no significant conformational differences (RMSD of <0.6 Å for all C α atoms). At the active site, α -KG and the substrate are well defined with evident electron density (Supplementary Figure S1). In addition, similar to the apo NcT7H Δ C structure, a metal ion is bound at the active site, which is assigned as Ni²⁺ owing to the presence of NiCl₂ in the crystallization solution and the verification of the ICP-AES analysis (Supplementary Table S1).

Similar to other α -KG dependent dioxygenases (37), NcT7H adopts a variation of double-stranded β -helix (DSBH) fold as the core (Figure 1A and Supplementary Figure S2). The major β -sheet of the DSBH fold comprises of eight β -strands (β 3–7, β 10, β 13, and β 16), and the minor β -sheet comprises of four β -strands (β 9, β 11–12, and β 14). In addition, two short β -strands (β 1 and β 2) are located at one end (called the back end) of the DSBH core, and eight α -helices wrap around the DSBH core. These structure elements together form the compact globular catalytic domain. Besides, there are two large insertions (one between β 5 and β 6 and the other between β 7 and β 9) which form

several extra structure elements flanking the catalytic domain (Figure 1A and Supplementary Figure S2).

Structure of the active site

The active site of NcT7H is located at the front end (opposite to the back end) of the DSBH core and partially covered by the α 10 helix at the C-terminus (Figure 1A). It is composed of residues mainly from strands β 6, β 7, β 9, β 10, β 14, and β 16. Structural comparison shows that the α -KG binding does not cause significant conformational changes in the overall structure (RMSD of <1.0 Å for all C α atoms) but induces small conformational changes at the active site (Figure 1B and Supplementary Figure S3A). Compared to the apo NcT7H Δ C structure, in the α -KG-bound NcT7H structure, the residues composing the active site and the associated β -strands move slightly towards α -KG, and the active site assumes a relatively more compact conformation. In addition, the binding of the substrate also does not cause obvious conformational changes at the active site among different substrate-bound NcT7H structures as well as compared to the α -KG-bound NcT7H structure (Supplementary Figure S3B–F). However, the lack of major conformational changes at the active site of NcT7H upon α -KG and substrate binding could be due to the crystalline state, which might not exactly reflect the solution situation of the enzyme. For example, the NMR and crystallographic studies of the dioxygenase AlkB have shown that the active site of AlkB is conformationally flexible in solution and the cosubstrate and substrate binding may induce substantial conformational changes at the active site (38,39).

The active site of NcT7H is consisted of the binding sites for the metal ion, the cosubstrate and the substrate. In the apo NcT7H Δ C structure, the Ca²⁺ is coordinated by three strictly conserved residues His214, Asp216 and His271, which constitute the characteristic HXD/E...H motif of α -KG dependent dioxygenases (37,40), and three water molecules in an octahedral geometry (Figure 2A and Supplementary Figure S4). In the ligand-bound NcT7H structures, the Ni²⁺ retains the coordinations with the three residues and one water molecule (Wat1), but the other two water molecules are replaced by the C1-carboxylate and C2-oxo groups of α -KG (Figure 2B–E). The binding and coordination geometry of the Ca²⁺ and Ni²⁺ may resemble these of the Fe²⁺ to some extent (41).

In the ligand-bound NcT7H structures, α -KG binds to a positively charged pocket with the C5-carboxylate embedded in the pocket and the C1-carboxylate oriented towards the substrate (Figure 2B–E and Supplementary Figure S5). The C5-carboxylate is stabilized by four hydrogen-bonding interactions with the side chains of Tyr194, Arg286 and Ser288. The C1-carboxylate forms a hydrogen bond with the side chain of Arg190. In addition, several hydrophobic residues (including Leu192, Leu223 and Val273) are also involved in interactions with α -KG. Sequence alignment shows that all the residues involved in α -KG binding are strictly conserved in T7Hs from different species (Supplementary Figure S4). The interactions of α -KG with NcT7H are very similar to those observed in many other α -KG dependent dioxygenases (39,40,42,43).

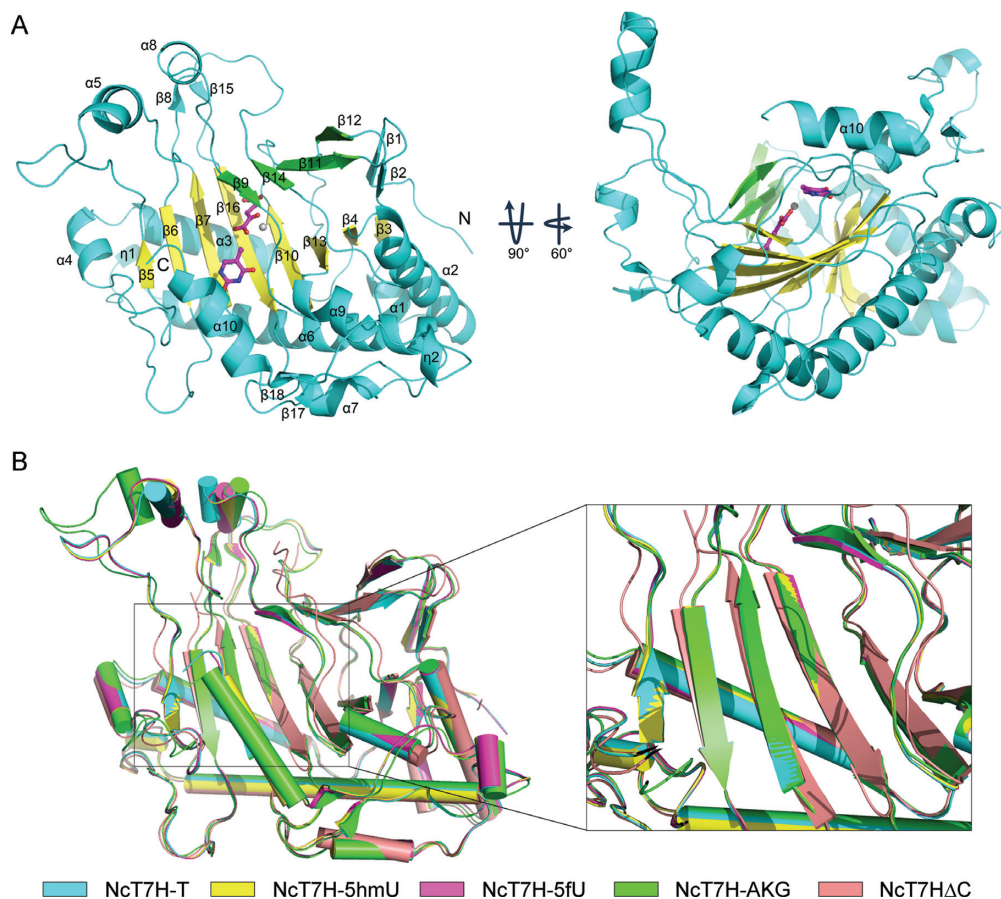


Figure 1. Overall structure of NcT7H. (A) Overall structure of NcT7H in complex with α -KG and T. The minor and major β -sheets that constitute the DSBH fold are coloured in green and yellow, respectively. Other structure elements flanking the DSBH core are shown in cyan. The location of the active site is indicated with a metal ion shown as a grey sphere, the cosubstrate α -KG and the substrate T shown as ball-and-stick models and coloured in magenta. The secondary structure elements are labeled. (B) Comparison of the overall structures of NcT7H Δ C (salmon), NcT7H-AKG (green), NcT7H-T (cyan), NcT7H-5hmU (yellow) and NcT7H-5fU (magenta).

In the substrate-bound NcT7H structures, T, 5hmU or 5fU binds to a pocket adjacent to the metal ion and α -KG, and has largely hydrophobic interactions and a few hydrophilic interactions with the surrounding residues (Figure 2C–E and Supplementary Figure S5B–D). Specifically, in the NcT7H-T complex, the uracil moiety of T makes π – π stacking interaction with the side chain of Phe292 and is covered on top in part by the side chain of Tyr217 via hydrophobic interactions (Figure 2C). In addition, the N1 atom of the uracil moiety makes indirect hydrogen bonds with the side chains of Asn87, Glu122 and Arg190 via a water molecule; the O2 group makes a direct hydrogen bond with the side chain of Asn87 and an indirect hydrogen bond with the side chain of Asn294 via a water molecule; the N3 atom makes an indirect hydrogen bond with the side chain of Asn294 via a water molecule; and the O4 group makes a direct hydrogen bond with the main-chain amine of Tyr217 (Figure 2C). Interestingly, the C5-methyl group (C51 atom) of T is not specifically recognized by any residues but has van der Waals contacts with the C1-carboxylate of α -KG and the side chains of Asp216 and His214 (about 3.6–3.9 Å) (Figure 2C), which is similar to 5mC in the HsTET2 and NgTET1 structures (16,17).

In the NcT7H-5hmU and NcT7H-5fU complexes, the uracil moiety of 5hmU or 5fU maintains almost identical interactions with the surrounding residues as that in the NcT7H-T complex (Figure 2C–E). The C51 atom of 5hmU and 5fU is located about 3.7 Å and 3.9 Å away from the C1-carboxylate of α -KG, respectively. In addition, the 5-hydroxymethyl group of 5hmU or the 5-formyl group of 5fU forms a hydrogen bond with the C1-carboxylate of α -KG and a hydrogen bond with the side chain of Arg190 (Figure 2D and E). In all the substrate-bound NcT7H structures, Arg190 is located at the entrance to the active site and its side chain makes hydrogen bonds with both α -KG and the substrate, suggesting that Arg190 might play a critical role in the recognition and binding of the cosubstrate and substrate and the catalytic reaction. Sequence alignment shows that all the residues involved in the substrate binding are strictly or highly conserved in T7Hs from different species (Supplementary Figure S4).

Binding affinity of NcT7H for different substrates

NcT7H can catalyse the consecutive oxidations of T to 5hmU, 5fU and 5caU, which differ only at the C5 position of the uracil moiety. To investigate whether NcT7H has varied

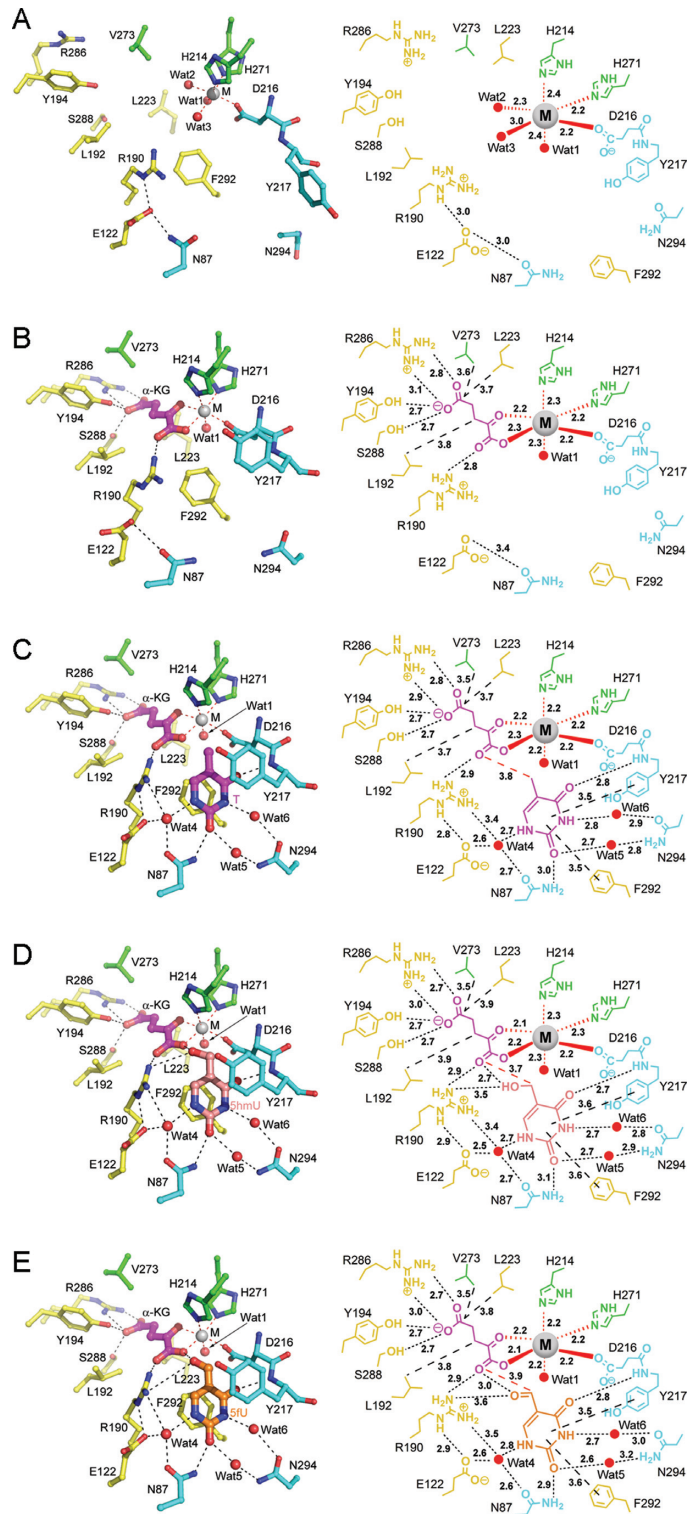


Figure 2. Structure (left panel) and schematic diagram (right panel) of the active site showing the interactions of the metal ion, α -KG and the substrate with the surrounding residues. (A) The apo NcT7HAD. (B) The NcT7H-AKG complex. (C) The NcT7H-T complex. (D) The NcT7H-5hmU complex. (E) The NcT7H-5fU complex. The substrate, α -KG and the residues are shown with ball-and-stick models, the metal ion with a grey sphere, and the water molecules with red spheres. The colour coding for the residues is the same as in Figure 1A. The substrates T, 5hmU and 5fU are coloured in magenta, salmon and orange, respectively. The coordination bonds of the metal ion are shown with red dotted lines, the hydrogen bonds with black dotted lines, and the key hydrophobic interactions with black dashed lines. The bond lengths (\AA) are indicated.

binding affinities for different substrates, we performed ITC experiments to analyse the binding affinities of NcT7H for α -KG and the substrates. NcT7H has a moderate binding affinity for α -KG (K_d of $34.4 \pm 6.63 \mu\text{M}$) and a 6-fold higher affinity for the α -KG analogue NOG (K_d of $5.35 \pm 1.26 \mu\text{M}$) (Figure 3 and Table 2), consistent with the structural data that α -KG can bind tightly to the active site of NcT7H. In the α -KG-bound NcT7H structure, the C3 atom of α -KG is about 3.6 Å away from the hydroxyl group of Tyr194. When NOG is docked into the active site, the N3 atom of NOG could form a hydrogen bond with the side chain of Tyr194 to strength its binding, explaining the higher affinity for NOG.

In the absence of α -KG or NOG, NcT7H has no detectable binding with all the substrates (data not shown). In the presence of NOG (substituting α -KG to avoid catalytic reaction), NcT7H has moderate binding affinities for T and 5hmU (K_d of $32.0 \pm 5.24 \mu\text{M}$ and $30.5 \pm 5.59 \mu\text{M}$, respectively) and a 3-fold lower affinity for 5fU (K_d of $108 \pm 13.0 \mu\text{M}$), but has no detectable binding with 5caU and U (Figure 3 and Table 2). These results indicate that NcT7H binds the substrates only in the presence of α -KG, which is consistent with the previous biochemical and structural studies of other α -KG dependent dioxygenases (44). Moreover, these results suggest that the C5 modification groups of the substrates play some roles in determining the substrate specificity and in the substrate binding.

Furthermore, our ITC results show that NcT7H has no detectable binding with the α -KG product succinate (Figure 3 and Table 2) and in the presence of succinate, NcT7H also has no detectable binding with the substrates (data not shown), consistent with our crystallization experiment results that NcT7H yielded no crystals in the presence of succinate alone or succinate and 5hmU, 5fU or 5caU. These data suggest that after each oxidation reaction, the products (succinate and 5hmU, 5fU or 5caU) are released from the enzyme, and then new α -KG and substrate are reloaded for next oxidation reaction.

Biochemical and mutagenesis analyses

We first characterized the apparent enzymatic activities of NcT7H towards different substrates using the HPLC method (26) (Figure 4A). At the standard condition, about 80% of T was oxidized to 5hmU, 5fU and 5caU with gradually decreasing amounts; and about 80% of 5hmU was converted to 5fU and 5caU with approximately equal amounts. As the catalytic reaction towards 5fU is faster than that towards T and 5hmU, the amount of enzyme used in the assay was 2.5 fold less. At that condition, about 60% of 5fU was converted to 5caU.

We also performed kinetic studies to measure the K_m and k_{cat} values of NcT7H towards different substrates using the SCS/PK/LDH-coupled system (36). NcT7H exhibits comparable K_m and k_{cat} values towards T and 5hmU but displays about 1.8-fold higher K_m value and 2-fold higher k_{cat} value towards 5fU, and the catalytic efficiency of NcT7H towards different substrates is in the order of $5fU \geq T > 5hmU$ (Table 3 and Supplementary Figure S6), which are consistent with our binding affinity and activity assay results (Figures 3 and 4, and Table 2). However, it should be noted that

at the assay condition, the catalytic reaction for 5fU is faster than that for T and 5hmU and 5fU is also less stable than T and 5hmU, and thus the variation of the velocity for 5fU is much higher than that for T and 5hmU. This might explain that the kinetic parameters reported in this study and the previous studies of *Rhodotorula glutinis* T7H (27,28) are comparable for T and 5hmU but vary greatly for 5fU.

In the full-length NcT7H structures, the C-terminal region forms a long helix ($\alpha 10$) that covers on top of the substrate-binding pocket (Figure 4B). Although helix $\alpha 10$ does not directly interact with α -KG and substrate, it contributes to the formation of the substrate-binding pocket. Specifically, several residues of helix $\alpha 10$ (particularly Arg328) stabilize the side chain of Tyr217 which makes hydrophobic interactions with the substrate. In addition, the strictly conserved residues Tyr324 and Arg328 of helix $\alpha 10$ are involved in a hydrogen-bonding network with three highly conserved residues (Thr215, Asp256 and Ser268) located adjacent to the HXD/E...H motif. These interactions would stabilize the conformations of the structure elements at the active site and in turn constrain the conformation of helix $\alpha 10$ to form a closed substrate-binding pocket (Figure 4B). Indeed, our crystallization experiments show that co-crystallization of NcT7H Δ C with α -KG and the substrates yielded only the α -KG-bound form (data not shown). In addition, our biochemical data show that NcT7H Δ C could not catalyse the oxidation reaction of T, 5hmU or 5fU (Figure 4A). The structural and biochemical data together indicate that the C-terminal region of NcT7H is essential for the substrate binding and catalysis.

To examine the functional roles of key residues at the active site in the catalytic reaction, we performed mutagenesis and biochemical studies (Figure 4C–E). Single mutation of any of the three residues involved in the metal ion binding (His214, Asp216 and His271) results in almost complete loss of the activity towards all substrates as these mutations would affect the Fe^{2+} binding which is essential for catalysis. Among the residues involved in the α -KG binding, Arg190 has hydrogen-bonding interactions with the C1-carboxylate of α -KG and the C5-hydroxymethyl group of 5hmU (or the C5-formyl group of 5fU), and thus its mutation to Lys or Ala leads to drastic loss of the activity towards all substrates (Figure 4C–E). Mutations of the other residues (Leu192, Tyr194, Leu223, Val273, Arg286 and Ser288) moderately decrease the activity towards T, 5hmU and 5fU with gradually increasing effects in the order of $T < 5hmU < 5fU$. These results indicate that the activity of NcT7H towards T, 5hmU and 5fU has an increasing dependency on the proper binding of α -KG.

Among the residues involved in the substrate binding, the side chain of Phe292 has π – π interaction with the substrates and mutation F292A abolishes the activity towards all substrates. The side chain of Tyr217 makes hydrophobic interactions with the substrates; mutation Y217F has insignificant effects on the activity towards T and 5hmU but a moderate effect on the activity towards 5fU, and mutation Y217A has moderate effects on the activity towards T and 5hmU but abolishes the activity towards 5fU. Sequence alignment shows that Phe292 is strictly conserved, and Tyr217 is highly conserved but can be substituted by Phe or Trp in some species (Supplementary Figure S4). As

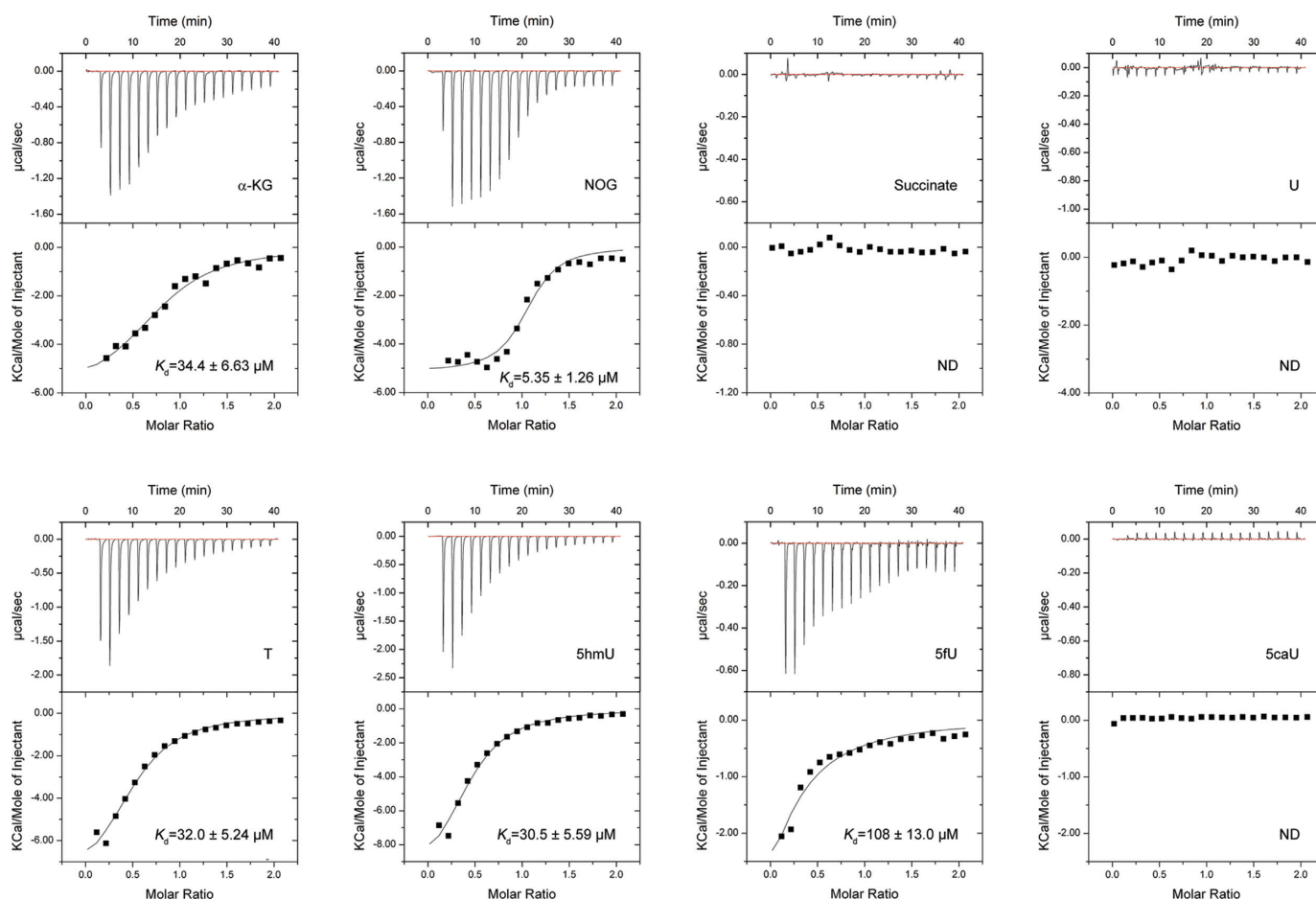


Figure 3. ITC analyses of the binding affinities of NcT7H towards different ligands. The binding affinities of NcT7H for α -KG, NOG and succinate and the binding affinities of NcT7H for U, T, 5hmU, 5fU and 5caU in the presence of NOG were analysed. The dissociation constant (K_d) is indicated. ND, not detected.

Table 2. ITC measured thermodynamic parameters of NcT7H with different ligands

Ligand	K_d (μ M)	ΔH (kcal/mol)	$T\Delta S$ (kcal/mol)	n Value
α -KG	34.4 ± 6.63	-6.03 ± 0.56	0.06	0.80 ± 0.05
NOG	5.35 ± 1.26	-5.14 ± 0.20	2.05	1.04 ± 0.03
Succinate	ND	ND	ND	ND
U	ND	ND	ND	ND
T	32.0 ± 5.24	-8.45 ± 0.74	-2.32	0.52 ± 0.03
5hmU	30.5 ± 5.59	-10.8 ± 1.19	-4.59	0.45 ± 0.04
5fU	108 ± 13.0	-8.73 ± 0.60	-3.31	0.20 ± 0.00
5caU	ND	ND	ND	ND

The protein used for analysing the binding affinity of NcT7H for U, T, 5hmU, 5fU and 5caU was pre-incubated with the cosubstrate analogue NOG. ND, not detected.

residues Asn87, Glu122 and Asn294 are mainly involved in water molecule mediated hydrogen-bonding interactions with the substrate, mutations N87A and E122A do not severely impair the activity towards T but can dramatically reduce the activity towards 5hmU and 5fU, and mutation N294A does not affect the activity towards all substrates. These results indicate that mutations of the residues participating in the substrate binding affect the activity to varied extents with gradually increasing effects in the order of $T < 5hmU < 5fU$ as well.

Furthermore, we carried out mutagenesis and kinetic studies to verify the functional roles of several key residues at the active site (Table 3 and Supplementary Figure S6). Mutations N87A, R190K and R286A have moderate effects on the K_m value (increased by 4–6 folds) and varied effects on the k_{cat} value (decreased by 1–9 folds) towards T and 5hmU, and the catalytic efficiencies of these mutants are substantially decreased by 5–55 folds. Mutation Y217A substantially increases the K_m value (11–12 folds) towards T and 5hmU but has minor effect on the k_{cat} value, and thus the catalytic efficiency of this mutant is decreased by 18

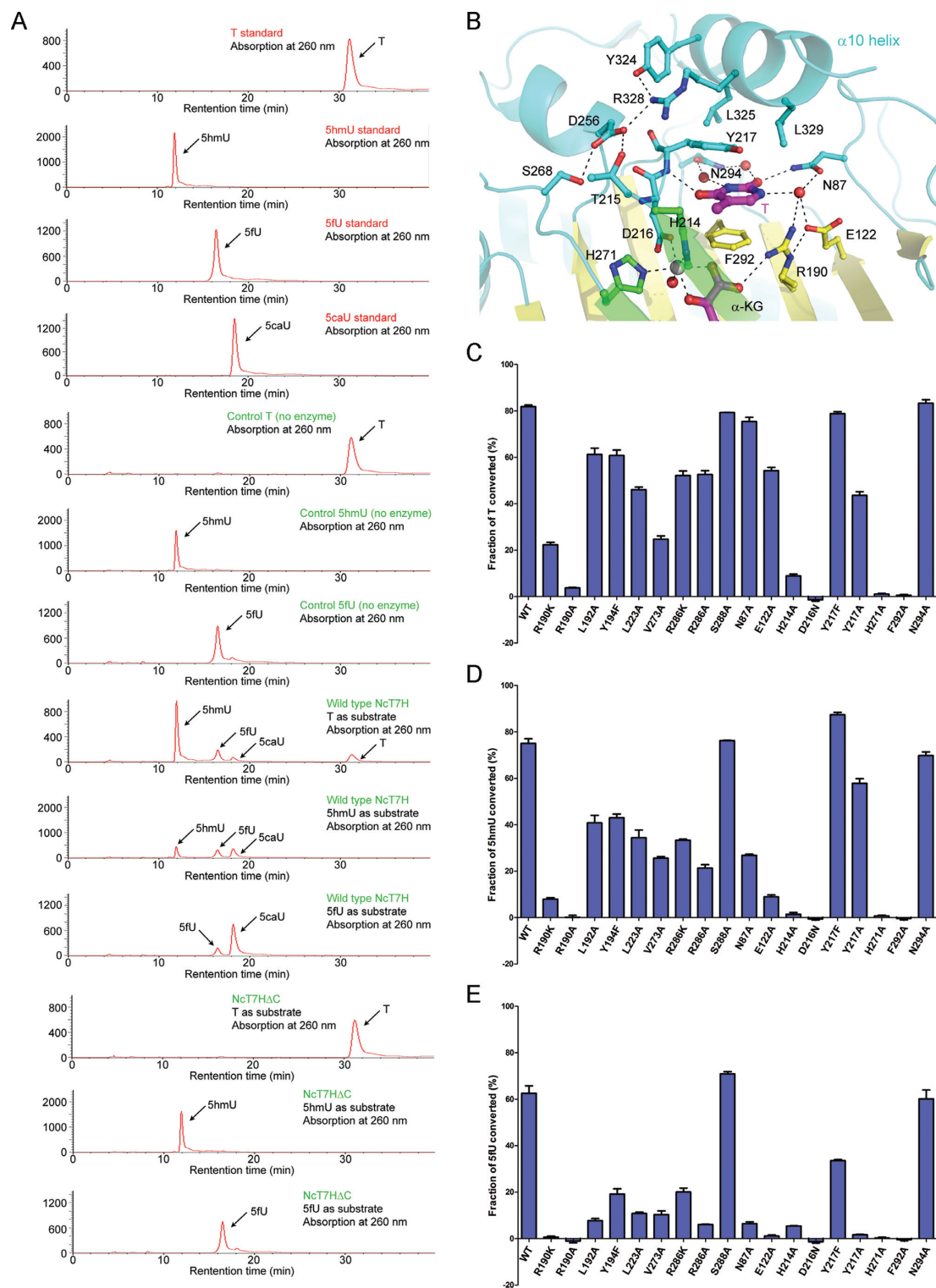


Figure 4. Apparent enzymatic activities of wild-type and mutant NcT7H. **(A)** Activities of the full-length of NcT7H and the C-terminal truncated NcT7H Δ C for the oxidation of T, 5hmU and 5fU analysed with the HPLC method. The standard T, 5hmU, 5fU and 5caU were used as references. **(B)** Interactions of the α 10 helix with the residues composing the active site in the structure of NcT7H-T complex. The key residues, the substrate and α -KG are shown with ball-and-stick models. The hydrogen bonds are indicated with black dotted lines. **(C–E)** The effects of mutations of the key residues at the active site on the apparent enzymatic activity for the oxidation of **(C)** T, **(D)** 5hmU and **(E)** 5fU, respectively.

Table 3. Kinetic parameters of wild-type and mutant NcT7H towards different substrates

	Substrate	K_m (μM)	k_{cat} (min^{-1})	k_{cat}/K_m ($\text{min}^{-1}\cdot\mu\text{M}^{-1}$)
WT	T	141 \pm 13.4	25.0 \pm 0.74	0.177
	5hmU	178 \pm 16.4	19.5 \pm 0.58	0.110
	5fU	253 \pm 32.2	45.9 \pm 2.14	0.181
R190K	T	542 \pm 64.7	3.30 \pm 0.18	0.006
	5hmU	1016 \pm 200	2.12 \pm 0.23	0.002
	5fU	ND	ND	ND
R286A	T	649 \pm 74.1	24.3 \pm 1.29	0.037
	5hmU	950 \pm 114	8.46 \pm 0.55	0.009
	5fU	ND	ND	ND
N87A	T	638 \pm 128	13.8 \pm 1.32	0.022
	5hmU	1065 \pm 156	5.77 \pm 0.48	0.005
	5fU	ND	ND	ND
Y217A	T	1500 \pm 332	15.4 \pm 2.13	0.010
	5hmU	2062 \pm 360	25.6 \pm 3.02	0.012
	5fU	ND	ND	ND
Y217F	T	128 \pm 18.5	32.5 \pm 1.37	0.254
	5hmU	247 \pm 35.2	33.2 \pm 1.68	0.134
	5fU	718 \pm 213	14.2 \pm 2.11	0.020

ND, not detected.

folds and 9 folds, respectively. Additionally, these mutations abolish the activity for 5fU, leading to undetectable K_m and k_{cat} values. Interestingly, mutation Y217F has minor effects on the K_m value but causes a slightly increased k_{cat} value for T and 5hmU, leading to a slightly increased catalytic efficiency. However, this mutation has moderate effects on both the K_m (increased by 3 folds) and k_{cat} values (decreased by 3 folds) for 5fU, and thus the Y217F mutant has a substantially decreased catalytic efficiency towards 5fU. These results suggest that the hydroxyl group of Tyr217 has differed effects on the catalytic reaction for different substrates. Overall, the kinetic results are in good agreement with the activity assay results (Figure 4C–E).

Taken together, our structural and biochemical data show that His214, Asp216 and His271 play essential roles in the metal ion binding; Arg286 plays an important role in the binding of α -KG; Arg190 plays a vital role in the binding of both α -KG and the substrate, and Phe292 and Tyr217 play critical roles in the substrate binding. The other residues at the active site have varied effects on the binding of α -KG and the substrate. The proper binding of α -KG and the substrate has gradually increasing effects on the activity towards the substrates in the order of T < 5hmU < 5fU.

Substrate specificity and catalytic mechanism of NcT7H

Our structural and biochemical data demonstrate that NcT7H can bind T, 5hmU and 5fU with slightly differed binding affinity only in the presence of α -KG but can not bind U and 5caU (Figure 3 and Table 2). In the substrate-bound NcT7H structures, the substrates maintain almost identical interactions with the enzyme, including the hydrophobic interactions with the side chains of Phe292 and Tyr217 and the hydrogen-bonding interactions with several conserved residues (Figure 2C–E). These interactions play important roles in the substrate binding as mutations of these residues have severe to moderate effects on the substrate binding and the activity (Figure 4C–E, Table 3 and Supplementary Figure S6). As the C5 modification groups

of T, 5hmU and 5fU make slightly differed interactions with α -KG and the enzyme, they might play important roles in the substrate binding and differentiation (Figure 3 and Table 2). Presumably, the differed C5 modification groups would have varied effects on the chemical property of the uracil moiety and thus could affect its interactions with the enzyme. Furthermore, the strictly conserved Arg190 makes interactions with both α -KG and the substrate and thus also plays a critical role in the substrate recognition and the catalytic reaction (Figure 2D and E, Figure 4C–E, Table 3 and Supplementary Figure S6). These factors together allow NcT7H to distinguish the substrates T, 5hmU and 5fU from U and 5caU, and to determine the slightly varied binding affinity and/or enzymatic activity towards different substrates.

Previous structural and biochemical studies of other α -KG dependent dioxygenases have shown that α -KG binds to the enzyme with either ‘in-line’ or ‘off-line’ mode and consequently there are two slightly differed catalytic mechanisms (37,43,44). In the substrate-bound NcT7H structures, the C1-carboxylate and C2-oxo groups of α -KG are located in opposite to His271 and Asp216 of the HXD/E...H motif, respectively, and Wat1 (which occupies presumably the oxygen in catalysis) is located below the plane formed by the C1-carboxylate of α -KG, Asp216 and His271 and in opposite to His214 (Figure 2B–E), indicating that the α -KG binding assumes the ‘off-line’ mode in NcT7H. As our structural and biochemical data have identified the key residues involved in the binding of the metal ion, α -KG and the substrate and their functional roles in the catalysis, we can propose a detailed catalytic mechanism for NcT7H, which is very similar to that for other α -KG dependent dioxygenases with the ‘off-line’ α -KG binding mode (44) (Supplementary Figure S7).

DISCUSSION

The TET proteins can catalyse the consecutive oxidations of 5mC to 5hmC, 5fC and 5caC in active DNA demethylation

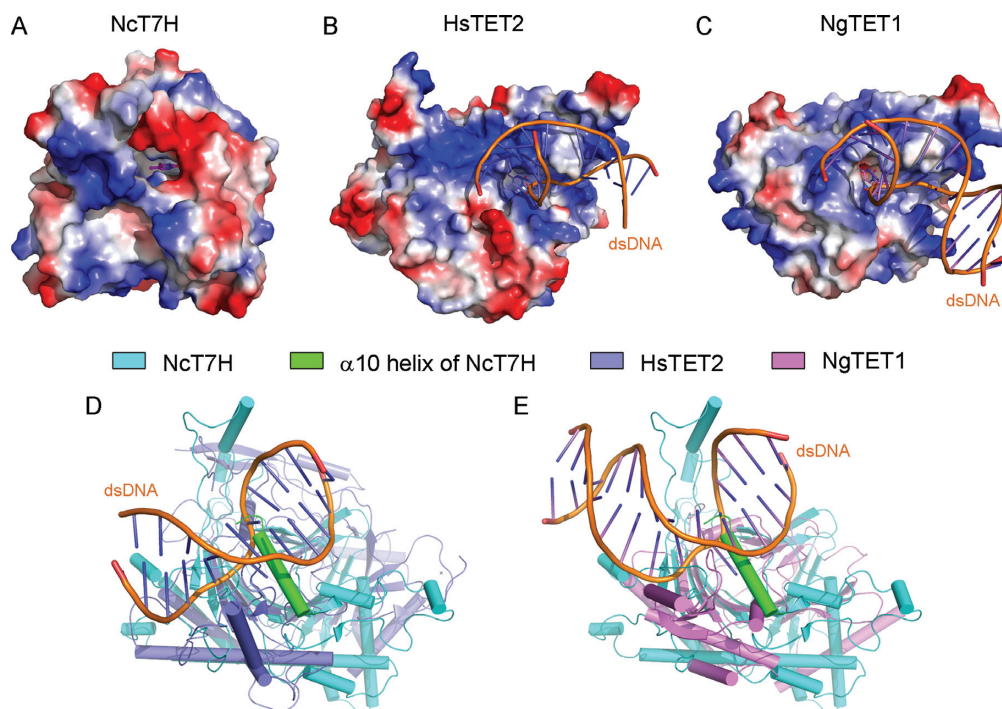


Figure 5. Structural comparison of NcT7H with the TET proteins. (A–C) Electrostatic potential surfaces of (A) NcT7H, (B) HsTET2 and (C) NgTET1. The surface charge distribution is displayed as blue for positive, red for negative and white for neutral. (D) Structural comparison of NcT7H with HsTET2-DNA. (E) Structural comparison of NcT7H with NgTET1-DNA. The bound dsDNA are shown with yellow ribbons.

in mammals, and play critical roles in epigenetic regulation (11–14). Although the crystal structures of the HsTET2-DNA and NgTET1-DNA complexes have been reported (16,17), the molecular basis for how the TET proteins recognize different C5 modification groups on cytosine and catalyse the consecutive oxidations are still elusive. As the conversion of T to 5hmU, 5fU and 5caU catalysed by T7H in fungi is chemically similar to that of 5mC to 5hmC, 5fC and 5caC catalysed by the TET proteins in mammals, we carried out the structural and biochemical studies of NcT7H to investigate the molecular basis for the substrate specificity and catalytic mechanism of T7H, hoping that this knowledge could advance our understanding of the structure and function of the TET proteins. Indeed, structural comparison of NcT7H with other α -KG dependent dioxygenases and in particular the TET proteins provides new insights into the molecular mechanism of the substrate recognition and catalytic reaction of the TET proteins.

As expected, structural similarity search using the Dali server (45) shows that the structure of NcT7H is similar to several α -KG dependent dioxygenases, including HsTET2, NgTET1, *Aspergillus nidulans* isopenicillin N synthase (AnIPNS), *H. sapiens* AlkB homology 5 (HsAlkBH5), *E. coli* AlkB (EcAlkB) and *H. sapiens* FTO (HsFTO). The DSBH core of these enzymes can be superimposed well albeit the flanking structure elements are much different (Supplementary Table S2 and Figure S8). Particularly, despite of the low sequence identity, the HXD/E...H motif which is involved in the metal ion binding and the Arg residue which is involved in stabilization of the C5-carboxylate of α -KG (corresponding to Arg286 in NcT7H) in these enzymes,

the common features of the α -KG dependent dioxygenases (37,40), could be structurally aligned very well (Supplementary Figures S9 and S10).

Nevertheless, a detailed structural comparison of NcT7H, HsTET2 and NgTET1 shows that there are substantial structural differences at the active site. The active sites of HsTET2 and NgTET1 are relatively open, whereas the active site of NcT7H is relatively closed because the C-terminal α 10 helix covers on top of the active site and it would cause steric conflict with the bound DNA in the HsTET2-DNA and NgTET1-DNA complexes (Figure 5). Besides, the surface surrounding the active site in HsTET2 and NgTET1 is largely positively charged (Figure 5B and C), while that in NcT7H is largely negatively charged which is unfavourable for binding DNA (Figure 5A). These differences may determine that the active site of NcT7H binds only a free base but not a modified nucleotide in DNA.

Furthermore, the structural comparison also reveals that the recognition and binding manners of the bases have some commonalities and differences between NcT7H and the TET proteins. As the substrate of NcT7H is a free base, the substrate is sandwiched by two aromatic residues (Phe292 and Tyr217) to stabilize its binding (Figure 2C–E). However, as the substrate of the TET proteins is a modified nucleotide in DNA and its conformation is constrained by the DNA, only one side of 5mC is involved in hydrophobic interactions with an aromatic residue (Tyr1902 of HsTET2 and Phe295 of NgTET1) (Supplementary Figure S9A and C). In addition, the substrates of both NcT7H and the TET proteins make hydrogen-bonding interactions with several

residues either directly or indirectly via water molecules; however, these involved residues are not very conserved (Supplementary Figures S9B and D and S10). On the other hand, the interactions of α -KG with the proteins are conserved in all these structures (Figure 2B and Supplementary Figure S9A–D). Specifically, the C1-carboxylate and C5-carboxylate of α -KG are each stabilized by an Arg residue (Arg190 and Arg286 of NcT7H, Arg1261 and Arg1896 of HsTET2, and Arg224 and Arg289 of NgTET1, respectively) via hydrogen-bonding interactions albeit the Arg residue interacting with the C1-carboxylate is provided by different structure elements (Supplementary Figure S10).

As Phe292 and Tyr217 of NcT7H play important roles in the substrate binding, we also compared the active site of NcT7H with that of EcAlkB and HsFTO, two other dioxygenases with nucleic acid substrates. Similar to that in NcT7H, the substrate base in EcAlkB and HsFTO is also sandwiched by two large side-chain residues (Trp69 and His131 in EcAlkB and Tyr108 and His231 in HsFTO) (Supplementary Figure S9E–H). However, these residues are contributed by different structural elements and thus the binding orientation of the substrate base in EcAlkB and HsFTO is different from that in NcT7H and the TET proteins. These results indicate that stabilization of the substrate base by large side-chain residue(s) is a common feature of these dioxygenases.

Our structural and biochemical data have demonstrated that the conserved Arg190 of NcT7H plays a critical role in the binding of α -KG and the substrate and the catalytic reaction. In the substrate-bound NcT7H structures, the side chain of Arg190 makes hydrogen-bonding interaction and/or van der Waals contacts with both the C1-carboxylate of α -KG and the C5 modification group of the substrate (Figure 2C–E). In the HsTET2-DNA and NgTET1-DNA structures, there is a conserved Arg residue (Arg1261 of HsTET2 and Arg224 of NgTET1) from a different structure element, which occupies a similar spatial position as Arg190 of NcT7H and has hydrogen-bonding interaction with the C1-carboxylate of α -KG. However, secondary structural alignment shows that Arg190 of NcT7H is equivalent to Thr1372 of HsTET2 or Ala212 of NgTET1 (Supplementary Figure S10). These residues are located on a β -strand (corresponding to β 7 of NcT7H) which is proposed to be involved in substrate binding (37). In the HsTET2-DNA and NgTET1-DNA structures, Thr1372 of HsTET2 and Ala212 of NgTET1 make van der Waals contacts with the 5-methyl group of 5mC and thus might play some roles in the recognition of the 5-methyl group of 5mC (16,17). These results suggest that HsTET2 and NgTET1 might use two residues to recognize and bind the C1-carboxylate of α -KG and the C5 modification group of the substrate, respectively, whereas NcT7H uses a single Arg190 to exert the dual functions.

Our structural and biochemical data of NcT7H show that the products are released after each oxidation reaction and new cosubstrate and substrate are reloaded to conduct the next oxidation reaction. In the HsTET2 and NgTET1 structures, the 5mC is flipped out of the DNA and inserted into the active site, and α -KG is deeply buried at the active site (16,17). Although the DNA blocks the entrance to the active site, it has extensive interactions with the protein and

thus its dissociation from the protein might be difficult. Hence, there might be two possibilities for the TET proteins to reload a new α -KG and carry out the next oxidation reaction. One possibility is that after each oxidation reaction, the DNA is released from the protein to unblock the entrance to the active site, allowing release of succinate and reloading of a new α -KG. In this case, the consecutive oxidations of 5mC by the TET proteins is also discontinuous, and thus the oxidized intermediates of 5mC might have specific functions and should be regulated precisely, which is in agreement with the previous studies showing that the oxidation products of 5mC could be positioned at different regulatory regions and contribute to different transcriptional states (46,47). It is also possible that after each oxidation reaction, the DNA is not dissociated from the protein; instead, the active site undergoes conformational changes to release succinate and then reload a new α -KG. In this case, the TET proteins could catalyse the consecutive oxidations of 5mC to 5caC continuously and thus the oxidized intermediates of 5mC might have less functional roles. Further structural and functional studies of the TET proteins are needed to resolve this issue.

ACCESSION NUMBERS

The crystal structures of the apo NcT7H Δ C and the full-length NcT7H in complexes with α -KG, with α -KG and T, with α -KG and 5hmU, and with α -KG and 5fU have been deposited in the Protein Data Bank under accession codes 5C3O, 5C3P, 5C3Q, 5C3R, and 5C3S, respectively.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank the staff members at BL17U of Shanghai Synchrotron Radiation Facility of China and BL19U1 of National Facility for Protein Science Shanghai for technical supports in diffraction data collection, and other members of our group for discussion.

Authorship Contributions: W.L. carried out the biochemical and structural studies, and drafted the manuscript. T.Z. participated in the structure determination. J.D. conceived the study, participated in the experimental design, data analyses and discussion, and wrote the manuscript.

FUNDING

The Ministry of Science and Technology of China [2011CB966301, 2013CB910404]; National Natural Science Foundation of China [31221001]; Chinese Academy of Sciences [XDB08010302]. Funding for open access charge: Ministry of Science and Technology of China [2011CB966301, 2013CB910404].

Conflict of interest statement. None declared.

REFERENCES

1. Chow, J.C., Yen, Z., Ziesche, S.M. and Brown, C.J. (2005) Silencing of the mammalian X chromosome. *Annu. Rev. Genomics Hum. Genet.*, **6**, 69–92.

2. Morison, I.M., Ramsay, J.P. and Spencer, H.G. (2005) A census of mammalian imprinting. *Trends Genet.*, **21**, 457–465.
3. Miranda, T.B. and Jones, P.A. (2007) DNA methylation: the nuts and bolts of repression. *J. Cell. Physiol.*, **213**, 384–390.
4. Suzuki, M.M. and Bird, A. (2008) DNA methylation landscapes: provocative insights from epigenomics. *Nat. Rev. Genet.*, **9**, 465–476.
5. Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L. et al. (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, **324**, 930–935.
6. Iyer, L.M., Tahiliani, M., Rao, A. and Aravind, L. (2009) Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle*, **8**, 1698–1710.
7. Pfaffeneder, T., Hackner, B., Truss, M., Munzel, M., Muller, M., Deiml, C.A., Hagemeyer, C. and Carell, T. (2011) The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew. Chem.-Int. Edit.*, **50**, 7008–7012.
8. Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C. and Zhang, Y. (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science*, **333**, 1300–1303.
9. He, Y.F., Li, B.Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L. et al. (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science*, **333**, 1303–1307.
10. Zhang, L., Lu, X., Lu, J., Liang, H., Dai, Q., Xu, G.L., Luo, C., Jiang, H. and He, C. (2012) Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. *Nat. Chem. Biol.*, **8**, 328–330.
11. Wu, H. and Zhang, Y. (2014) Reversing DNA methylation: mechanisms, genomics, and biological functions. *Cell*, **156**, 45–68.
12. Pastor, W.A., Aravind, L. and Rao, A. (2013) TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nat. Rev. Mol. Cell Biol.*, **14**, 341–356.
13. Cimmino, L., Abdel-Wahab, O., Levine, R.L. and Aifantis, I. (2011) TET family proteins and their role in stem cell differentiation and transformation. *Cell Stem Cell*, **9**, 193–204.
14. Branco, M.R., Ficz, G. and Reik, W. (2012) Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nat. Rev. Genet.*, **13**, 7–13.
15. Huang, Y. and Rao, A. (2014) Connections between TET proteins and aberrant DNA modification in cancer. *Trends Genet.*, **30**, 464–474.
16. Hashimoto, H., Pais, J.E., Zhang, X., Saleh, L., Fu, Z.Q., Dai, N., Correa, I.R. Jr, Zheng, Y. and Cheng, X. (2014) Structure of a Naegleria Tet-like dioxygenase in complex with 5-methylcytosine DNA. *Nature*, **506**, 391–395.
17. Hu, L., Li, Z., Cheng, J., Rao, Q., Gong, W., Liu, M., Shi, Y.G., Zhu, J., Wang, P. and Xu, Y. (2013) Crystal structure of TET2-DNA complex: insight into TET-mediated 5mC oxidation. *Cell*, **155**, 1545–1555.
18. Fink, R.M. and Fink, K. (1962) Utilization of radiocarbon from thymidine and other precursors of ribonucleic acid in *Neurospora crassa*. *J. Biol. Chem.*, **237**, 2289–2290.
19. Shaffer, P.M., McCroskey, R.P. and Abbott, M.T. (1972) Substrate specificity of the hydroxylase reaction in which thymidine is converted to thymine ribonucleoside. *Biochim. Biophys. Acta*, **258**, 387–394.
20. Liu, C.K., Hsu, C.A. and Abbott, M.T. (1973) Catalysis of three sequential dioxygenase reactions by thymine 7-hydroxylase. *Arch. Biochem. Biophys.*, **159**, 180–187.
21. Holme, E., Lindstedt, G., Lindstedt, S. and Tofft, M. (1971) 18-O studies of the 2-ketoglutarate-dependent sequential oxygenation of thymine to 5-carboxyuracil. *J. Biol. Chem.*, **246**, 3314–3319.
22. Watanabe, M.S., McCroskey, R.P. and Abbott, M.T. (1970) The enzymatic conversion of 5-formyluracil to uracil 5-carboxylic acid. *J. Biol. Chem.*, **245**, 2023–2026.
23. Abbott, M.T., Schandl, E.K., Lee, R.F., Parker, T.S. and Midgett, R.J. (1967) Cofactor requirements of thymine 7-hydroxylase. *Biochim. Biophys. Acta*, **132**, 525–528.
24. Palmatier, R.D., McCroskey, R.P. and Abbott, M.T. (1970) The enzymatic conversion of uracil 5-carboxylic acid to uracil and carbon dioxide. *J. Biol. Chem.*, **245**, 6706–6710.
25. Smiley, J.A., Angelot, J.M., Cannon, R.C., Marshall, E.M. and Asch, D.K. (1999) Radioactivity-based and spectrophotometric assays for isoorotate decarboxylase: identification of the thymidine salvage pathway in lower eukaryotes. *Anal. Biochem.*, **266**, 85–92.
26. Xu, S., Li, W., Zhu, J., Wang, R., Li, Z., Xu, G.L. and Ding, J. (2013) Crystal structures of isoorotate decarboxylases reveal a novel catalytic mechanism of 5-carboxyl-uracil decarboxylation and shed light on the search for DNA decarboxylase. *Cell Res.*, **23**, 1296–1309.
27. Neidigh, J.W., Darwanto, A., Williams, A.A., Wall, N.R. and Sowers, L.C. (2009) Cloning and characterization of *Rhodotorula glutinis* thymine hydroxylase. *Chem. Res. Toxicol.*, **22**, 885–893.
28. Thornburg, L.D., Lai, M.T., Wishnok, J.S. and Stubbe, J. (1993) A non-heme iron protein with heme tendencies: an investigation of the substrate specificity of thymine hydroxylase. *Biochemistry*, **32**, 14023–14033.
29. Holme, E. (1975) A kinetic study of thymine 7-hydroxylase from *Neurospora crassa*. *Biochemistry*, **14**, 4999–5003.
30. Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.
31. Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W. et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr.*, **D66**, 213–221.
32. Emsley, P. and Cowtan, K. (2004) COOT: model-building tools for molecular graphics. *Acta Crystallogr.*, **D60**, 2126–2132.
33. Murshudov, G.N., Vagin, A.A. and Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr.*, **D53**, 240–255.
34. Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A. et al. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr.*, **D67**, 235–242.
35. Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283–291.
36. Luo, L., Pappalardi, M.B., Tummino, P.J., Copeland, R.A., Fraser, M.E., Grzyska, P.K. and Hausinger, R.P. (2006) An assay for Fe(II)/2-oxoglutarate-dependent dioxygenases by enzyme-coupled detection of succinate formation. *Anal. Biochem.*, **353**, 69–74.
37. Aik, W., McDonough, M.A., Thalhammer, A., Chowdhury, R. and Schofield, C.J. (2012) Role of the jelly-roll fold in substrate binding by 2-oxoglutarate oxygenases. *Curr. Opin. Struct. Biol.*, **22**, 691–700.
38. Ergel, B., Gill, M.L., Brown, L., Yu, B., Palmer, A.G. 3rd and Hunt, J.F. (2014) Protein dynamics control the progression and efficiency of the catalytic reaction cycle of the *Escherichia coli* DNA-repair enzyme AlkB. *J. Biol. Chem.*, **289**, 29584–29601.
39. Yu, B., Edstrom, W.C., Benach, J., Hamuro, Y., Weber, P.C., Gibney, B.R. and Hunt, J.F. (2006) Crystal structures of catalytic complexes of the oxidative DNA/RNA repair enzyme AlkB. *Nature*, **439**, 879–884.
40. McDonough, M.A., Loenarz, C., Chowdhury, R., Clifton, I.J. and Schofield, C.J. (2010) Structural studies on human 2-oxoglutarate dependent oxygenases. *Curr. Opin. Struct. Biol.*, **20**, 659–672.
41. Harding, M.M. (2001) Geometry of metal-ligand interactions in proteins. *Acta Crystallogr.*, **D57**, 401–411.
42. Valegard, K., van Scheltinga, A.C., Lloyd, M.D., Hara, T., Ramaswamy, S., Perrakis, A., Thompson, A., Lee, H.J., Baldwin, J.E., Schofield, C.J. et al. (1998) Structure of a cephalosporin synthase. *Nature*, **394**, 805–809.
43. Clifton, I.J., McDonough, M.A., Ehrismann, D., Kershaw, N.J., Granatino, N. and Schofield, C.J. (2006) Structural studies on 2-oxoglutarate oxygenases and related double-stranded beta-helix fold proteins. *J. Inorg. Biochem.*, **100**, 644–669.
44. Hausinger, R.P. (2004) FeII/alpha-ketoglutarate-dependent hydroxylases and related enzymes. *Crit. Rev. Biochem. Mol. Biol.*, **39**, 21–68.
45. Holm, L. and Rosenstrom, P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res.*, **38**, W545–W549.
46. Wu, H. and Zhang, Y. (2011) Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. *Genes Dev.*, **25**, 2436–2452.
47. Shen, L., Wu, H., Diep, D., Yamaguchi, S., D'Alessio, A.C., Fung, H.L., Zhang, K. and Zhang, Y. (2013) Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell*, **153**, 692–706.