npg

ORIGINAL ARTICLE

# Crystal structures of isoorotate decarboxylases reveal a novel catalytic mechanism of 5-carboxyl-uracil decarboxylation and shed light on the search for DNA decarboxylase

Shutong Xu[1, 2, *], Wenjing Li[1, 2, *], Junjun Zhu[1, 2], Rong Wang[1, 2], Zheng Li[1, 2], Guo-Liang Xu[1], Jianping Ding[1]

[1]State Key Laboratory of Molecular Biology, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, [2]Graduate School of Chinese Academy of Sciences, 320 Yue-Yang Road, Shanghai 200031, China

DNA methylation and demethylation regulate many crucial biological processes in mammals and are linked to many diseases. Active DNA demethylation is believed to be catalyzed by TET proteins and a putative DNA decarboxylase that may share some similarities in sequence, structure and catalytic mechanism with isoorotate decarboxylase (IDCase) that catalyzes decarboxylation of 5caU to U in fungi. We report here the structures of wild-type and mutant IDCases from *Cordyceps militaris* and *Metarhizium anisopliae* in apo form or in complexes with 5caU, U, and an inhibitor 5-nitro-uracil. IDCases adopt a typical $(\beta/\alpha)_8$ barrel fold of the amidohydrolase superfamily and function as dimers. A $Zn^{2+}$ is bound at the active site and coordinated by four strictly conserved residues, one Asp and three His. The substrate is recognized by several strictly conserved residues. The functional roles of the key residues at the active site are validated by mutagenesis and biochemical studies. Based on the structural and biochemical data, we present for the first time a novel catalytic mechanism of decarboxylation for IDCases, which might also apply to other members of the amidohydrolase superfamily. In addition, our biochemical data show that IDCases can catalyze decarboxylation of 5caC to C albeit with weak activity, which is the first *in vitro* evidence for direct decarboxylation of 5caC to C by an enzyme. These findings are valuable in the identification of potential DNA decarboxylase in mammals.

*Keywords*: catalytic mechanism; crystal structure; DNA demethylation; DNA decarboxylation; decarboxylase; pyrimidine metabolism; thymidine salvage pathway

## Introduction

Methylation at the C5 position of cytosine (5-methyl-cytosine, 5mC) in DNA is an important form of epigenetic modifications. It plays critical roles in many key cellular processes, including embryonic development, transcription, chromatin remodeling, X-chromosome inactivation, genomic imprinting, and chromosome stability [1-4]. Aberrant DNA methylation patterns have been associated with many human diseases such as imprinting disorders and cancers [5, 6]. It is well established that DNA methylation is catalyzed by DNA methyltransferases. In contrast, whether and how active DNA demethylation occurs has been controversial for a long time [7, 8]. Several mechanisms have been proposed for the active DNA demethylation, including direct breakage

of the carbon-carbon bond of 5mC, DNA base-excision repair (BER) and nucleotide excision repair (NER) pathways, deamination of 5mC to thymine followed by replacement of the T:G mismatch via BER, and oxidative demethylation [7, 8]. Recently, it was discovered that TET proteins can catalyze the conversion of 5mC to 5-hydroxymethyl-cytosine (5hmC) using $Fe^{2+}$ and α-ketoglutarate (α-KG) as cofactors [9, 10], and further to 5-formyl-cytosine (5fC) and 5-carboxyl-cytosine (5caC) *in vitro* and in cultured cells [11-13]. Currently, it is believed that the TET proteins play an important role in the active DNA demethylation through this series of oxidation reactions followed by the conversion of 5caC to cytosine through two possible mechanisms. One possibility is that 5fC and 5caC in DNA can be excised by a DNA glycosylase and subsequently repaired via BER; it has been shown that thymine-DNA glycosylase (TDG) possesses an excision activity towards 5fC- and 5caC-containing DNA [11, 14]. Another possibility is that 5caC can be directly decarboxylated by a putative DNA decarboxylase; it has been shown that mouse embryonic stem cell nuclear extract can convert 5caC to C in DNA, suggesting the existence of a DNA decarboxylase [15].

Intriguingly, the chemistry for the conversion of 5mC to 5caC and possibly to C in the active DNA demethylation in mammals is very similar to that for the conversion of T (or 5mU) to 5-carboxyl-uracil (5caU) and further to uracil in the nucleotide anabolism in some fungi (Supplementary information, Figure S1). It is well established that nucleotides can be biosynthesized through two different pathways *in vivo*, namely *de novo* and salvage pathways. In most organisms, the *de novo* pyrimidine nucleotide synthesis pathway is conserved, starting with *de novo* formation of uridine-5′-monophosphate (UMP) that can be converted to various nucleotides for RNA and DNA syntheses [16-18]. In the vast majority of organisms, the conversion of UMP to thymine deoxyribonucleotide is considered to be unidirectional due to the lack of a metabolic pathway to turn the latter back to UMP [19, 20]. However, *Neurospora crassa* and several other fungi can demethylate T to U via a unique thymidine salvage pathway [21]. This pathway consists of a series of catalytic reactions, including the oxidation of thymidine to thymine ribonucleoside by pyrimidine deoxyribonucleoside 2′-hydroxylase, the hydrolysis of thymine ribonucleoside to thymine and ribose by pyrimidine-nucleoside phosphorylase and hydrolase, the sequential oxidation of thymine to 5-hydroxymethyl-uracil (5hmU), 5-formyl-uracil (5fU) and 5-carboxyl-uracil (5caU) by thymine-7-hydroxylase (T7H) [22-26], and finally the non-oxidative decarboxylation of 5caU to U by isoorotate decarboxylase (IDCase) [27, 28].

The decarboxylation activity of *N. crassa* IDCase was demonstrated using two different biochemical assays [27, 28]. The IDCase gene is adjacent to the T7H gene in the *N. crassa* genome with its protein sequence sharing a low similarity to 2-amino-3-carboxymuconate-6-semialdehyde decarboxylase (ACMSD) [29]. Despite its important biological function in the regulation of the cellular pyrimidine pool, the molecular bases of the substrate recognition and binding and the catalytic mechanism of IDCase remain unknown.

The similarity in the chemistry between the 5mC-to-C conversion and the 5mU-to-U conversion and the lack of biochemical and mechanistic knowledge of IDCases inspired us to carry out this work. Here we report the crystal structures of the wild-type and mutant *Cordyceps militaris* IDCase (CmIDCase) in apo form and in complexes with the substrate 5caU, a substrate analog 5-nitro-uracil (5niU), and the product U, and the wild-type *Metarhizium anisopliae* IDCase (MaIDCase) in apo form. Our structural and biochemical data together reveal the molecular bases of the substrate recognition and binding and the catalytic mechanism for IDCases and shed light on the search for potential DNA decarboxylase in mammals.

## Results and Discussion

### *Overall structures of CmIDCase and MaIDCase*

The apo CmIDCase structure was solved at 1.9 Å resolution (Table 1). CmIDCase assumes a closed barrel structure: a distorted $(\beta/\alpha)_8$ barrel domain (residues 1-17 and 77-355) constitutes the main body; an insertion domain between strand β1 and helix α1 (residues 18-76) forms the cap; and the C-terminal region composing the last three helices (α9-α11) with an "L" shape forms the bottom (Figure 1A and Supplementary information, Figure S2). The insertion domain has a mixed structure consisting of α-helices (α1′ and α2′), β-strands (β1′-β3′), and a $3_{10}$ helix (η1′). The active site is located near the C-terminus of the eight β-strands of the barrel domain and is partially covered by the insertion domain. There is a metal ion bound at the active site with well-defined electron density (Figure 1B). The apo MaIDCase structure was solved at 2.6 Å resolution (Table 1). CmIDCase and MaIDCase share a high sequence similarity (72% identity) (Supplementary information, Figure S3A) and a very similar overall structure (a RMSD of 0.54 Å for all Cα atoms) (Supplementary information, Figure S3B). Similarly, a metal ion is identified at the active site of the apo MaIDCase.

The structures of the wild-type CmIDCase in complexes with U and 5niU were determined at 2.2 Å and 2.3 Å resolution, respectively; the H195A and D323N mutants

**Table 1** Summary of diffraction data and refinement statistics

| | CmIDCase | CmIDCase-U | CmIDCase-5niU | MaIDCase | D323N-5caU | D323A-5caU | D323N | H195A | H195A+Zn |
|---|---|---|---|---|---|---|---|---|---|
| **Data collection** | | | | | | | | | |
| Space group | $P2_1$ | $P2_1$ | $P2_1$ | $P3_112$ | $P4_1$ | $C2$ | $P4_1$ | $C2$ | $C2$ |
| Cell dimensions | | | | | | | | | |
| a (Å) | 77.7 | 77.3 | 77.3 | 133.0 | 77.4 | 192.3 | 77.6 | 144.9 | 144.4 |
| b (Å) | 146.2 | 144.9 | 145.3 | 133.0 | 77.4 | 54.3 | 77.6 | 56.4 | 57.5 |
| c (Å) | 77.8 | 77.3 | 77.5 | 175.4 | 134.4 | 143.8 | 140.1 | 105.7 | 105.6 |
| α (°) | 90.0 | 90.0 | 90.0 | 90.0 | 90.0 | 90.0 | 90.0 | 90.0 | 90.0 |
| β (°) | 96.5 | 97.1 | 95.8 | 90.0 | 90.0 | 121.0 | 90.0 | 105.4 | 105.3 |
| γ (°) | 90.0 | 90.0 | 90.0 | 120.0 | 90.0 | 90.0 | 90.0 | 90.0 | 90.0 |
| Resolution (Å) | 50.0-1.90 | 50.0-2.20 | 50.0-2.30 | 50.0-2.60 | 50.0-2.10 | 50.0-2.10 | 50.0-2.40 | 50.0-1.75 | 50.0-2.00 |
| | $(1.97-1.90)^a$ | (2.28-2.20) | (2.38-2.30) | (2.69-2.60) | (2.18-2.10) | (2.18-2.10) | (2.49-2.40) | (1.81-1.75) | (2.07-2.00) |
| Observed reflections | 621,021 | 290,157 | 267,357 | 264,415 | 175,702 | 269,398 | 108,427 | 309,724 | 234,039 |
| Unique reflections | 135,002 | 86,105 | 75,264 | 53,820 | 46,222 | 74,453 | 31,075 | 82,427 | 55,498 |
| $(I/\sigma(I) > 0)$ | | | | | | | | | |
| $R_{merge}$ (%)$^b$ | 8.7 (41.5) | 8.5 (48.1) | 12.3 (48.1) | 10.6 (62.3) | 8.5 (23.2) | 7.7 (26.7) | 10.3 (65.6) | 5.7 (29.7) | 6.2 (35.5) |
| $I/\sigma I$ | 15.1 (4.0) | 14.7 (3.2) | 11.2 (3.3) | 14.1 (2.6) | 18.1 (6.1) | 18.2 (4.6) | 13.4 (2.6) | 21 (4.9) | 24.6 (5.7) |
| Completeness (%) | 99.9 (100) | 99.5 (96.5) | 99.7 (97.6) | 98.3 (99.3) | 99.9 (100) | 99.7 (99.5) | 97.1 (98.4) | 98.8 (98.1) | 97.9 (97.1) |
| Redundancy | 4.6 (4.6) | 3.4 (3.3) | 3.6 (3.4) | 4.9 (4.9) | 3.8 (3.8) | 3.6 (3.5) | 3.5 (3.5) | 3.8 (3.5) | 4.2 (4.2) |
| **Refinement** | | | | | | | | | |
| Resolution (Å) | 50.0-1.90 | 50.0-2.20 | 50.0-2.30 | 50.0-2.60 | 50.0-2.10 | 50.0-2.10 | 50.0-2.40 | 50.0-1.75 | 50.0-2.00 |
| No. reflections | 128,186 | 81,682 | 71,426 | 51,062 | 43,835 | 70,702 | 29,494 | 78,200 | 52,674 |
| Working set | 121,455 | 77,360 | 67,618 | 48,379 | 41,495 | 66,952 | 27,914 | 74,103 | 49,850 |
| Test set | 6,731 | 4,322 | 3,808 | 2,683 | 2,340 | 3,750 | 1,580 | 4,097 | 2824 |
| $R_{work}/R_{free}$ (%)$^c$ | 15.7/19.7 | 15.3/21.0 | 15.3/20.2 | 17.5/22.2 | 15.9/19.0 | 17.5/22.0 | 18.1/24.0 | 15.6/17.9 | 16.9/20.5 |
| No. atoms | | | | | | | | | |
| Protein | 11,273 | 11,583 | 11,311 | 8,690 | 5,590 | 11,241 | 5,339 | 5,643 | 5,633 |
| Ligand | – | 32 | 44 | – | 22 | 44 | – | – | – |
| Water | 931 | 429 | 649 | 283 | 280 | 425 | 99 | 589 | 393 |
| B-factors | | | | | | | | | |
| Protein | 21.5 | 36.6 | 24.7 | 49.7 | 32.4 | 30.4 | 57.2 | 29.8 | 37.7 |
| Ligand | – | 29.1 | 21.3 | – | 27.1 | 37.6 | – | – | – |
| Ion | 19.5 | 36.3 | 22.7 | 52.7 | 27.0 | 38.8 | 49.3 | 46.4 | 32.5 |
| Water | 26.9 | 37.6 | 29.8 | 43.0 | 35.4 | 38.9 | 50.1 | 43.2 | 41.9 |

**Table 1** Summary of diffraction data and refinement statistics

| | CmIDCase | CmIDCase-U | CmIDCase-5niU | MaIDCase | D323N-5caU | D323A-5caU | D323N | H195A | H195A+Zn |
|---|---|---|---|---|---|---|---|---|---|
| R.m.s. deviations | | | | | | | | | |
| Bond lengths (Å) | 0.006 | 0.006 | 0.007 | 0.008 | 0.005 | 0.004 | 0.005 | 0.005 | 0.005 |
| Bond angles (°) | 1.1 | 1.1 | 1.1 | 1.1 | 1.1 | 1.0 | 1.0 | 1.0 | 1.0 |
| Ramachandran plot (%) | | | | | | | | | |
| Most favored | 91.8 | 91.5 | 92.0 | 90.2 | 91.0 | 91.5 | 91.4 | 91.1 | 90.9 |
| Allowed | 7.5 | 7.8 | 7.2 | 8.9 | 8.2 | 7.8 | 8.0 | 8.3 | 8.1 |
| Generously allowed | 0.7 | 0.7 | 0.8 | 0.9 | 0.8 | 0.7 | 0.6 | 0.6 | 1.0 |

[a]Numbers in parentheses represent the highest resolution shell.

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

[c]$R = \sum_{hkl} ||F_o| - |F_c|| / \sum_{hkl} |F_o|$.

in apo form at 1.75 Å and 2.4 Å resolution, respectively; and the D323N and D323A mutants in complexes with 5caU both at 2.1 Å resolution (Table 1). The overall structure of CmIDCase in these structures is very similar to that in the apo form (RMSDs of 0.25-0.59 Å for all Cα atoms), except that the apo H195A mutant exhibits some conformational differences in the insertion domain (a RMSD of 1.2 Å for all Cα atoms). In all these structures, a metal ion is also bound at the active site with well-defined electron density (Figure 1B). In the ligand-bound structures, the ligands are defined unambiguously with evident electron density (Figure 1B).

Size-exclusion chromatography analyses showed that both CmIDCase and MaIDCase exist predominantly as dimers in solution (Supplementary information, Figure S4A). In the apo CmIDCase structure, there are four CmIDCase molecules in an asymmetric unit, forming two dimers. No significant conformational difference exists among the four molecules and between the two dimers. The dimer interface buries about 2 938 Å$^2$ or 18.1% of the solvent-accessible surface area of each sub-unit, and is mediated mainly via two $3_{10}$ helices (η2′ and η3′) and three α-helices (α5, α6 and α4′) (Supplementary information, Figure S4B). In the other CmIDCase structures and the apo MaIDCase structure, the enzyme adopts a similar dimeric architecture. These results indicate that CmIDCase and MaIDCase form dimers in both solution and crystal structure, suggesting that dimerization of IDCases is a conserved property and is required for the enzymatic activity, in agreement with our structural and biochemical data (see results later).

*Metal-binding site*

In all the structures, there is a metal ion bound at the active site. There was 0.2 M MgCl$_2$ in the crystallization solutions of all the crystals except for the H195A mutant and the apo MaIDCase. To identify the type of the metal ion, we performed fluorescence scans of the apo CmID-Case and MaIDCase crystals and the results showed evident signal at the K-edge of zinc (1.28 Å) but not other metals. We collected the zinc anomalous dispersion data for both CmIDCase and MaIDCase crystals, and the anomalous difference Fourier maps showed strong electron density only at the active site (Supplementary information, Figure S5). These results together indicate that the bound metal ion is Zn$^{2+}$.

In the apo CmIDCase, the Zn$^{2+}$ is coordinated by six ligands in an octahedral geometry: the side-chain Nε2 of His12, His14 and His195, the side-chain carboxyl of Asp323, and two water molecules (Figure 2A). In the apo MaIDCase, the Zn$^{2+}$ is also coordinated by the same six ligands in an octahedral geometry (Supplementary in-
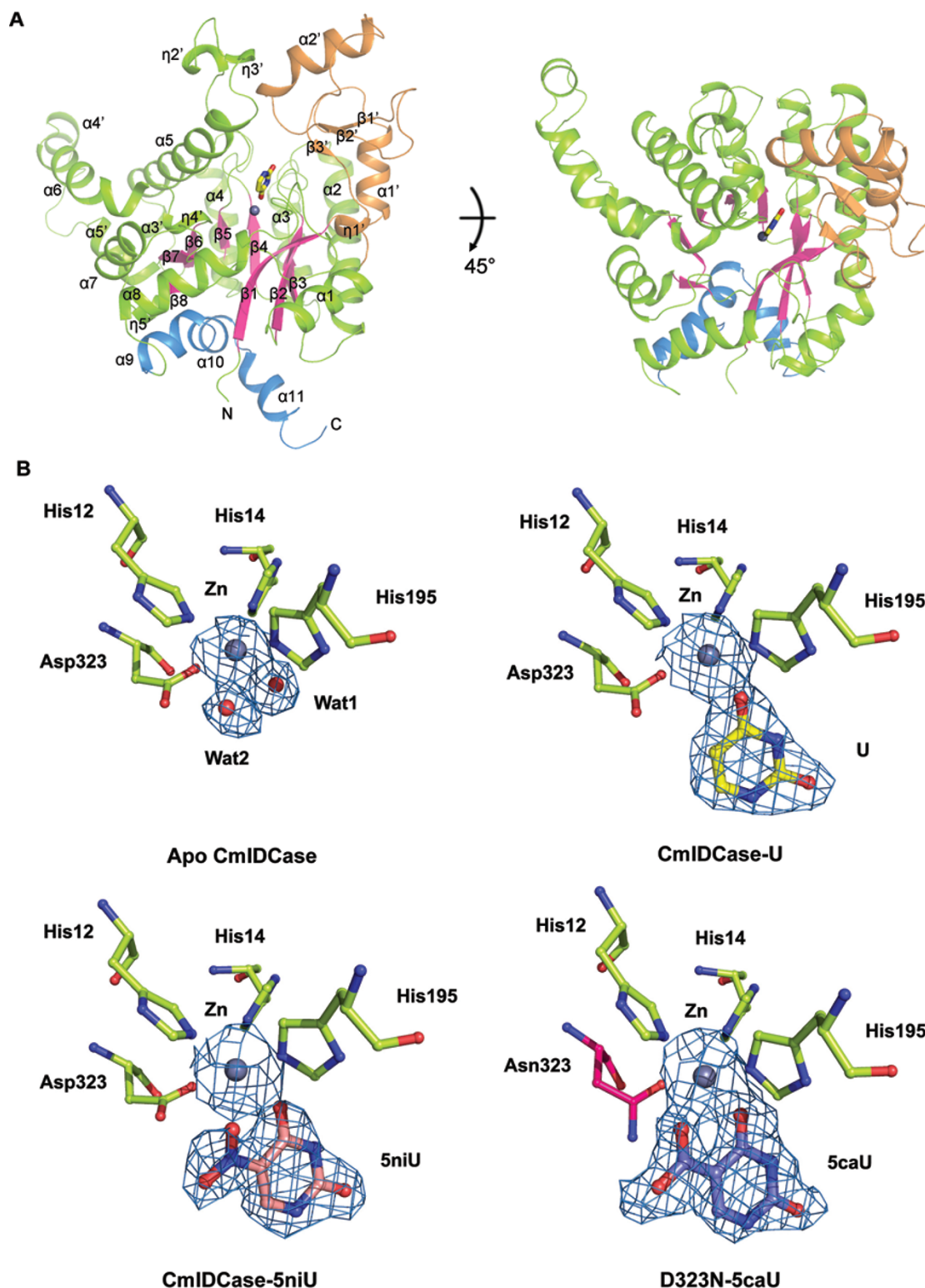
**Figure 1** Overall structures of CmIDCase. **(A)** Overall structure of CmIDCase in ribbon diagram. The $(\beta/\alpha)_8$ barrel domain is shown with $\alpha$-helices in green and $\beta$-strands in magenta. The insertion domain is colored in orange, and the C-terminal three $\alpha$-helices in blue. The location of the active site is indicated with a $Zn^{2+}$ shown as a gray sphere and the product U shown as a yellow ball-and-stick model. **(B)** Simulated annealing composite omit 2Fo-Fc maps (contoured at 1.0 $\sigma$ level) for the $Zn^{2+}$ and/ or the ligand at the active site in the structures of the apo CmIDCase, the CmIDCase-U complex, the CmIDCase-5niU complex, and the D323N-5caU complex as representatives.

formation, Figure S6A). In the ligand-bound CmIDCase structures, the $Zn^{2+}$ maintains the coordinations with the four residues. However, in the CmIDCase-U complex, one water molecule position is occupied by the O4 of U and the other is empty, forming a distorted square pyramidal geometry (Figure 2B); in the CmIDCase-5niU complex, both water molecule positions are occupied by the O4 and O52 of 5niU, retaining an octahedral geometry (Figure 2C). Sequence alignment shows that the four residues involved in the coordinations of $Zn^{2+}$ are strictly conserved in IDCases from different species (Supplementary information, Figure S3A), suggesting that the $Zn^{2+}$ binding is conserved and required for the activity.
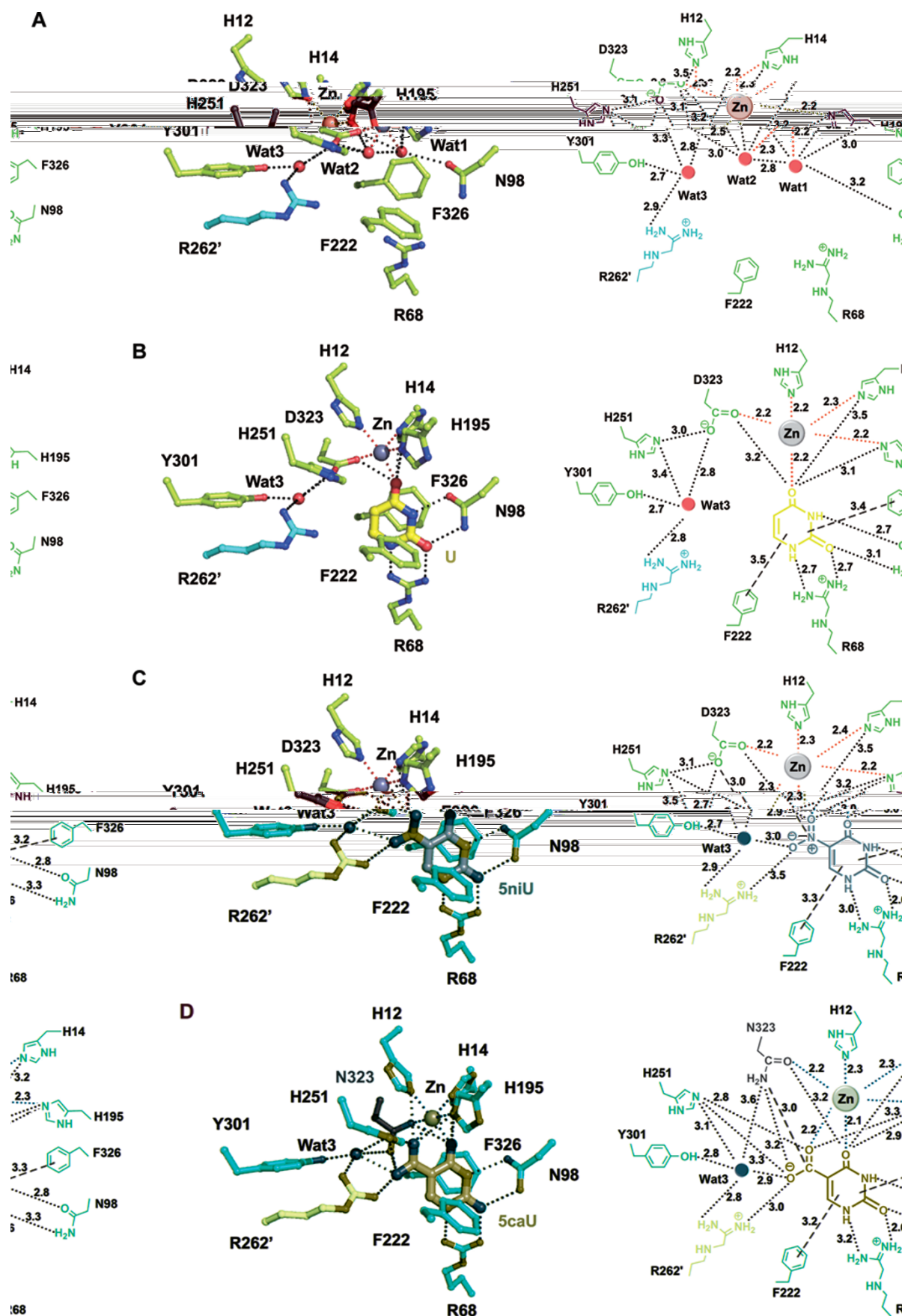
*Substrate-binding site*

In the structures of CmIDCase in complexes with U and 5niU, the ligands are bound in a small pocket adjacent to the metal-binding site (Figure 1A). Structural comparisons show that the active site structure in these complexes is very similar to that in the apo form; the residues forming both the metal-binding site and the substrate-binding site assume almost identical conformations (Figure 2 and Supplementary information, Figure S6), indicating that the ligand binding does not induce notable conformational changes in the overall structure and at the active site.

In the CmIDCase-U complex, the pyrimidine ring of U is sandwiched between the side chains of Phe222 and Phe326 (Figure 2B). Phe222 makes an edge-to-face type aromatic interaction with the pyrimidine ring and Phe326 makes a parallel ring stacking π-π interaction. The N1 and O2 of the pyrimidine ring form two hydrogen bonds with the side chain of Arg68; the O2 and N3 form two hydrogen bonds with the side chain of Asn98; the O4 is coordinated to $Zn^{2+}$ and forms three hydrogen bonds with the side chains of His14, His195 and Asp323. In the CmIDCase-5niU complex, the uracil moiety of 5niU forms almost identical interactions with the surrounding residues and the $Zn^{2+}$ (Figure 2C). Additionally, the O52 of the 5-nitro group forms one coordination bond with the $Zn^{2+}$ and a hydrogen bond with the side chain of His251, and the O53 forms a hydrogen bond with the side chain of Arg262 from the other subunit (designated as Arg262′) and interacts with the side chain of Tyr301

via a water molecule. These extra interactions between the 5-nitro group of 5niU and the protein indicate a tighter binding of 5niU (or 5caU, see below) than U, which is advantageous for the release of U and the reloading of 5caU.

To investigate the binding mode of 5caU at the active site and the functional roles of Asp323 and His195 in the catalysis, we determined the structures of several CmIDCase mutants. The H195A mutant could only be crystallized in apo form despite of the presence of 5caU in the crystallization. Interestingly, in this mutant structure, the $Zn^{2+}$ could only be refined with a low occupancy of 0.3 to yield a reasonable B factor; however, addition of 0.2 M $ZnCl_2$ in the crystallization could elevate the occupancy of $Zn^{2+}$ to 1.0 with a reasonable B factor (Table 1), indicating that the H195A mutation significantly impairs the $Zn^{2+}$ binding. In the apo H195A mutant cocrystallized with Zn, the active site structure is similar to that in the apo CmIDCase; however, due to the H195A mutation, the $Zn^{2+}$ makes only four coordinations with His12, His14, Asp323, and one water molecule in a distorted tetrahedral geometry (Supplementary information, Figure S6B). On the other hand, the D323N and D323A mutants could be crystallized in either apo form or in complex with 5caU. In the apo D323N mutant, the active site structure is very similar to that in the apo CmIDCase, and the $Zn^{2+}$ maintains six coordinations with the four residues and two water molecules in an octahedral geometry (Supplementary information, Figure S6C). In the D323N-5caU complex, the active site structure is very similar to that in the CmIDCase-5niU complex, and the $Zn^{2+}$ and 5caU make almost identical interactions between them and with the surrounding residues (Figure 2D). In the D323A-5caU complex, the active site structure is also similar to that in the CmIDCase-5niU complex; however, due to the D323A mutation, the $Zn^{2+}$ maintains five coordinations with His12, His14, His195, and the O4 and O52 of 5caU in a distorted square pyramidal geometry (Supplementary information, Figure S6D). These results together demonstrate that 5caU binds to the active site with almost identical interactions with the $Zn^{2+}$ and the surrounding residues to 5niU. The H195A mutation abolishes the substrate binding and impairs the metal binding, whereas the D323N and

**Figure 2** Structure of the active site. Structure of the active site (left panel) and a schematic diagram showing the interactions of the $Zn^{2+}$ and/or the ligand with the surrounding residues (right panel) in **(A)** the apo CmIDCase, **(B)** the CmIDCase-U complex, **(C)** the CmIDCase-5niU complex, and **(D)** the D323N-5caU complex. The ligands and the key residues involved in interactions with the $Zn^{2+}$ and the ligand are shown with ball-and-stick models, the $Zn^{2+}$ with a gray sphere, and the water molecules with red spheres. Arg262 is contributed from an adjacent subunit. The coordination bonds of the $Zn^{2+}$ are indicated with red dotted lines, the hydrogen bonds with black dotted lines, and the key hydrophobic interactions with black dashed lines. All bond lengths (Å) are indicated.

D323A mutations do not significantly affect the binding of either the metal ion or the substrate. As the residues involved in the recognition and binding of the substrate are strictly conserved (Supplementary information, Figure S3A), this substrate-binding mode should be conserved in all IDCases.

*Biochemical and mutagenesis analyses*

The decarboxylation activities of CmIDCase and MaIDCase were first characterized using a sensitive high-pressure liquid chromatography (HPLC) assay. When 5caU was incubated with the enzyme for 15 min, a new HPLC peak was detected with a retention time (~15 min), the same as the U standard, indicating that both enzymes can catalyze the decarboxylation of 5caU to U (Figure 3A). The kinetic parameters of the enzymes were then measured using a spectrophotometry assay (Supplementary information, Figure S7). The $K_m$ values of CmIDCase and MaIDCase are determined to be 22.4 ± 1.3 μM and 18.6 ± 1.9 μM, respectively, which are comparable to that of NcIDCase [28, 29]. The $k_{cat}$ values of CmIDCase and MaIDCase are determined to be 4.17 ± 0.09 min$^{-1}$ and 2.02 ± 0.08 min$^{-1}$, respectively. To test the possible decarboxylation activity for 5caC, we incubated 5caC with the enzyme for 36 h and analyzed the reaction mixture using the HPLC assay. Interestingly, a new HPLC peak was detected with a retention time (~4.2 min), the same as the C standard, indicating that both enzymes can also catalyze the 5caC-to-C conversion albeit with much weaker activities (Figure 3B). These results suggest that the 5caC-to-C conversion appears to utilize a similar catalytic reaction as the 5caU-to-U conversion. This is the first *in vitro* evidence for the direct decarboxylation of 5caC to C by an enzyme.

To analyze the functional roles of the key residues at the active site, we performed mutagenesis studies for CmIDCase (Figure 3C). Mutation of any of the four residues involved in the $Zn^{2+}$ binding (His12, His14, His195, and Asp323) abolished the activity; these mutations might affect the binding and/or the coordination geometry of $Zn^{2+}$ that are critical for the catalysis. Mutations of most of the residues that have direct interactions with the substrate, including Arg68, Asn98, Phe222, Arg262 and Phe326 in addition to His195 and Asp323, also resulted in undetectable activity; these mutations might affect the binding and/or the precise positioning of 5caU and thus impair the catalysis.

The H195A mutant possessed a significantly decreased binding ability for $Zn^{2+}$, and lost the binding ability for 5caU and thus the activity, indicating that His195 is critical for the binding of both the metal ion and the substrate. Intriguingly, the D323A and D323N mutants can bind both $Zn^{2+}$ and 5caU but lose the catalytic activity, indicating that the side-chain carboxyl of Asp323 is essential for the catalytic reaction. The R262A mutant also completely lost the activity, indicating that Arg262 plays a critical role in catalysis as well. Although His251 forms a hydrogen bond with the O52 of 5caU, the H251A mutant retained ~20% of the activity, suggesting that His251 plays a less critical role in the substrate binding and the catalysis. The Y301F mutant retained ~30% of the activity and the Y310A mutant ~20% of the activity, in agreement with our structural data that Tyr301 interacts with 5caU via a water molecule, and thus its mutation has a less severe impact on the substrate binding and the catalysis. As Arg68 is located in the insertion domain, our data indicate that the insertion domain is involved in the substrate binding and the catalytic reaction. Furthermore, as Arg262 is contributed by the adjacent subunit, our data also indicate that the dimeric state of CmIDCase is essential for its function.

*IDCases belong to the amidohydrolase superfamily*

Structural comparison of CmIDCase with those in the Protein Data Bank using the Dali server (http://ekhidna.biocenter.helsinki.fi/dali_server) reveals that the overall structure of CmIDCase resembles that of ACMSD of the amidohydrolase superfamily [30] (PDB code 2WM1, RMSD of 2.5 Å for the overall structure and 2.1 Å for the (β/α)$_8$ barrel domain) (Supplementary information, Figure S8). The amidohydrolase superfamily proteins are responsible for hydrolysis of a variety of substrates with functional groups at carbon or phosphorus centers [30, 31]. Although CmIDCase shares only low sequence similarities (< 30% identity) with ACMSD and other members of the superfamily, they have a similar active site structure (Supplementary information, Figure S8). The metal-binding site of CmIDCase resembles mostly that of the Zn-dependent ACMSD [30, 31], the Zn-dependent adenosine deaminase [32, 33], and the Fe-dependent cytosine deaminase [34, 35], and the four residues involved in the metal binding are strictly conserved. Based on the structural and sequence similarities, we conclude that IDCases belong to the amidohydrolase superfamily.

Nevertheless, a detailed comparison of the CmID-Case-5niU and D323N-5caU complexes with the HsAC-MSD-DHAP complex [30] reveals notable differences at the active site. In the HsACMSD-DHAP complex, the substrate analog DHAP interacts indirectly with the $Zn^{2+}$ via a water molecule that is proposed to act as the nucleophile in the catalysis (Supplementary information, Figure S8C). In addition, the residues involved in the substrate binding exhibit substantial variations in
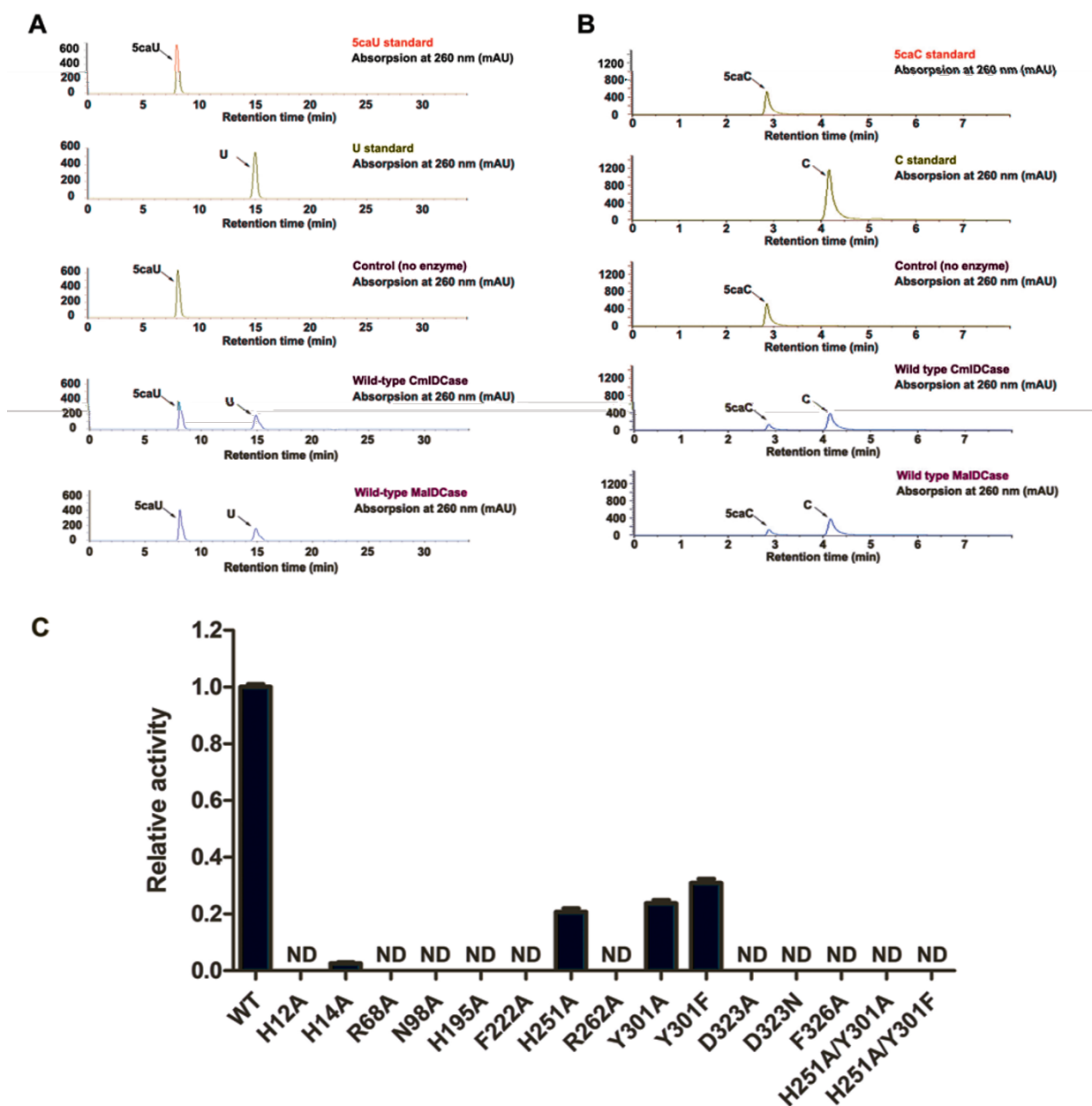
**Figure 3** Decarboxylation activities of wild-type and mutant CmIDCase and MaIDCase. **(A)** Decarboxylation activities of wild-type CmIDCase and MaIDCase for the conversion of 5caU to U analyzed using the HPLC method. The standard 5caU and U were used as references. The reaction took place at 30 °C for 15 min. **(B)** Decarboxylation activities of wild-type CmIDCase and MaIDCase for the conversion of 5caC to C analyzed using the HPLC method. The standard 5caC and C were used as references. The reaction took place at 30 °C for 36 h. **(C)** Relative activities of wild-type and mutant CmIDCase containing point mutations at the active site for the conversion of 5caU to U measured using the HPLC method. ND means that the activity was undetectable.

both sequence and structure. In particular, Arg235 of HsACMSD (equivalent to Arg262 of CmIDCase) is not involved in interaction with DHAP. These differences might indicate their differed substrate specificities and

catalytic mechanisms.

*Catalytic mechanism of decarboxylation for IDCases*
 As discussed above, CmIDCase belongs to the ami-

Shutong Xu *et al*. npg

1305

dohydrolase superfamily. A key feature of the catalytic mechanism proposed for the amidohydrolase superfamily members is the nucleophilic attack of a metal-bound hydroxide at a carbon atom of the substrate to form a tetrahedral intermediate [31, 36]. We determined a series of structures of CmIDCase or its mutants in apo form and in complexes with 5caU, 5niU, and U, representing different enzymatic states in the catalytic reaction. The structures of the apo CmIDCase and D323N mutant represent the initial state. In this state, the $Zn^{2+}$ is coordinated by six ligands including His12, His14, His195, Asp323, and two water molecules in an octahedral geometry; the substrate-binding pocket is empty or occupied by a few water molecules (Figure 2A and Supplementary information, Figure S6C). The structures of the D323N-5caU and CmIDCase-5niU complexes represent the substrate-bound state. In this state, the $Zn^{2+}$ is coordinated by six ligands including the four conserved residues and the O4 and O52 of 5caU in an octahedral geometry; 5caU makes direct interactions with $Zn^{2+}$ and several strictly conserved residues, including Arg68, Asn98, Phe222, His251, Arg262, and Phe326 (Figure 2C and 2D). The positions of the two water molecules to coordinate the $Zn^{2+}$ in the initial state are occupied by the O4 and O52 of 5caU and no water molecule is found within 5.0 Å of $Zn^{2+}$. The structure of the CmIDCase-U complex represents the product state. In this state, the $Zn^{2+}$ is coordinated by five ligands including the four conserved residues and the O4 of U in a distorted pyramidal geometry; U maintains the interactions with $Zn^{2+}$ and the surrounding residues (Figure 2B). The position of one water molecule in the initial state is occupied by the O4 of U and the other is unoccupied.

A detailed analysis of the D323N-5caU and CmID-Case-5niU complexes reveals that in the substrate-bound state, there is no room for a water molecule to bind directly to $Zn^{2+}$ and then to act as the nucleophile in catalysis, suggesting that CmIDCase is unlikely to utilize a catalytic mechanism similar to that proposed for other members of the amidohydrolase superfamily. On the other hand, the side-chain carboxyl of Asp323 is not only coordinated to $Zn^{2+}$, but also is oriented towards the 5-carboxyl of 5caU from the side way with a distance of ~3.0 Å to the C51 and an angle of 103° to the C51-C5 bond, suggesting that Asp323 is a potential candidate that functions as the nucleophile to attack the C51 of 5caU. This is in good agreement with our biochemical data that the side-chain carboxyl of Asp323 is essential for the catalytic reaction. p*K*a analysis of the ionizable residues at the active site of CmIDCase using the PROPKA 3.1 server (http://propka.ki.ku.dk/) [37] shows that Asp323 has a negative p*K*a of -2.6 that is much lower than any

other residues nearby, supporting its possibility to function as a nucleophile. Previous biochemical studies of L-2-haloacid dehalogenase from *Pseudomonas sp.* YL [38] and oxalate decarboxylase from *Bacillus subtilis* [39] using an [18]O isotope labeling technique have shown that a conserved Asp can serve as a nucleophile to attack the substrate to cleave a carbon-halogen or carbon-carbon bond. The suggestion of Asp323 functioning as the nucleophile to attack the C51 of 5caU can also explain why 5niU is an inhibitor of CmIDCase: as the N51 of 5niU is a poorer electrophile than the C51 of 5caU, Asp323 cannot perform a nucleophilic attack on the N51 of 5niU to cleave the N51-C5 bond. In addition, the fact that 5niU is an inhibitor also implies that the nucleophilic attack is unlikely to occur at other position, for example the C6, of the pyrimidine ring, leading to the weakening and cleavage of the C5-C51 bond as otherwise both 5caU and 5niU would become the substrate.

Meanwhile, in the structure of the D323N-5caU complex, there is a water molecule (Wat3) at the active site that forms hydrogen bonds with the 5-carboxyl of 5caU and the side chains of Asp323, Arg262, His251, and Tyr301. This water molecule is also conserved in the structures of the CmIDCase-5niU and D323A-5caU complexes and the apo CmIDCase, in which it maintains similar interactions with the substrate and/or the surrounding residues, suggesting that Wat3 might play some functional role(s) in the catalysis. A detailed analysis shows that Wat3 is positioned ~5.3 Å away from $Zn^{2+}$ and thus cannot be an equivalent of the metal-bound water molecule in the catalytic mechanism proposed for the amidohydrolase superfamily members. Wat3 is positioned in line with the C51-C5 bond of 5caU with an angle of 153° and thus is not in an ideal geometry to perform a direct nucleophilic attack on the C51 of 5caU because of potential negative charge repulsion. Nonetheless, we cannot completely rule out the possibility that Wat3 functions as the nucleophile due to the following reasons: (1) Wat3 is directly hydrogen-bonded to the 5-carboxyl of 5caU and is conserved in both the apo form and the substrate-bound CmIDCase structures; (2) Wat3 is directly hydrogen-bonded to Asp323 and Arg262, both of which are required for the activity; and (3) subtle conformational changes might occur at the active site during the catalytic reaction, which could place Wat3 in a better position to perform a direct nucleophilic attack on the C51 of 5caU.

Based on the structural and biochemical data together, we can propose the catalytic mechanism of CmIDCase decarboxylation with two possible models depending on whether Asp323 or Wat3 functions as the nucleophile (Figure 4). In both cases, the substrate 5caU binds to the
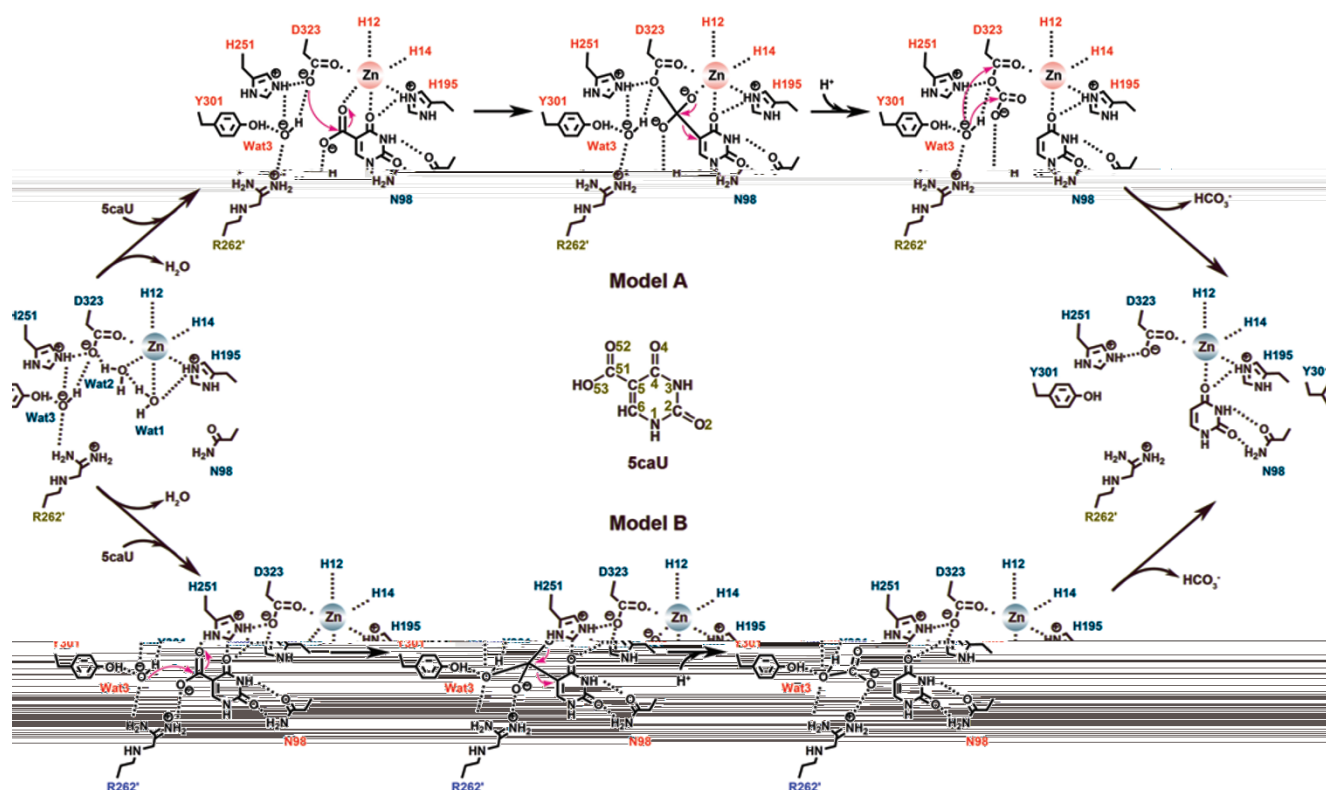
**Figure 4** A schematic diagram showing the catalytic mechanism of decarboxylation for CmIDCase. The catalytic mechanism with two possible models: in model A, Asp323 functions as the nucleophile to initiate the attack on the carboxyl C51 atom of 5caU; and in model B, Asp323 functions as the catalytic base to deprotonate Wat3 to form a hydroxide ion that acts as the nucleophile to initiate the attack on the carboxyl C51 atom of 5caU. In both models, Arg262 plays key roles in the stabilization of the intermediates and Wat3. The covalent bonds are indicated with black lines, and the hydrogen bonds and the coordination bonds with dash lines.

active site with its O4 and O52 replacing both Wat1 and Wat2 to interact directly with $Zn^{2+}$, and the bound 5caU is stabilized by the hydrogen bonding and hydrophobic interactions with the conserved residues, including Arg68, Asn98, His195, His251, Phe222, Arg262, Tyr301, Asp323, and Phe326. In model A, the side-chain carboxyl of Asp323 is activated by the $Zn^{2+}$ and functions as the nucleophile to attack the C51 of 5caU, leading to the formation of an unstable tetrahedral intermediate. Concomitantly, the C51-C5 bond of 5caU is polarized and then cleaved immediately through the protonation of C5 by a proton abstracted from solution to produce an unstable aspartate-carboxyl intermediate and the product U. Further, a water molecule (possibly Wat3) carries out a nucleophilic attack at either the leaving carboxyl carbon or the carboxyl carbon of Asp323 of the aspartate-carboxyl intermediate to release a $HCO_3^-$ ion, leading to the formation of the product state. In this process, Arg262 plays key roles in the stabilization of the intermediates

and Wat3. Finally, the product U is dissociated and the substrate 5caU can be reloaded to enter the next round of catalysis. In model B, Asp323 serves as a general base to deprotonate Wat3 into a hydroxide ion, which is stabilized by the positively charged side chain of Arg262. The activated Wat3 acts as the nucleophile to attack the C51 of 5caU to form an unstable tetrahedral intermediate. Concurrently, the C51-C5 bond of 5caU is polarized and then cleaved immediately through the protonation of C5 by a proton from solution to produce a $HCO_3^-$ ion and the product U. As both models can explain the structural and biochemical data well, additional data are required to discern the two possibilities. As all the residues at the active site are strictly conserved in IDCases from different species, the proposed catalytic mechanism for CmIDCase should apply to all IDCases. As the key residues involved in the metal binding including Asp323 are strictly conserved in the other members of the amidohydrolase superfamily, this catalytic mechanism might

also apply to these enzymes.

### Implication for active DNA decarboxylation

DNA methylation and demethylation regulate many crucial biological processes in mammals and are linked to many diseases. The underlying molecular mechanisms of the active DNA demethylation are still elusive and research in this area has been making rapid progresses in the past few years. Recently, the fungal thymidine salvage pathway has attracted great attention due to the similarity in the chemistries of the T-to-U and the 5mC-to-C conversions [8, 12, 15] (Supplementary information, Figure S1). In fungi, the demethylation of T to U is a multi-step process including the conversion of T to 5hmU, 5fU and 5caU through three consecutive oxidation reactions catalyzed by T7H and the non-oxidative decarboxylation of 5caU to U catalyzed by IDCase [22-26]. In mammals, the TET proteins can sequentially oxidize 5mC to 5hmC, 5fC and 5caC [9-12]. By analogy, it is plausible that the active demethylation of 5mC could end up with a direct decarboxylation of 5caC to C catalyzed by a putative DNA decarboxylase. The TET proteins in mammals were identified based on their low sequence similarities to the JBP1 and JBP2 proteins in trypanosome, which have been proposed to oxidize the 5-methyl group of thymine [9]. The previous biochemical data show that mouse embryonic stem cell nuclear extract can convert 5caC to C in DNA [15], and our biochemical data show that IDCase can catalyze the direct decarboxylation of 5caC to C albeit with weak activity, suggesting the existence of a putative DNA decarboxylase that may share some similarities in sequence, structure and catalytic mechanism with IDCases. Our structural and biochemical data of IDCases provide useful hints for the search of potential DNA decarboxylase.

The potential DNA decarboxylase may belong to the amidohydrolase superfamily with its catalytic domain adopting a $(\beta/\alpha)_8$ barrel fold. The architecture of the active site in the DNA decarboxylase should be similar to that in IDCase, consisting of a divalent metal ion (most likely $Zn^{2+}$) and four conserved residues (most likely one Asp and three His residues) that are critical to the catalysis. Owing to the structure similarity of C and U, it is possible that the key residues important for the binding of 5caU in CmIDCase, such as Arg68, Asn98, Phe222, His251, Arg262, Tyr301, and Phe326, may also be conserved in the DNA decarboxylase. Nevertheless, as the 4-amine of 5caC is a weaker electron donor than the 4-carbonyl of 5caU, it would form less tight interactions with the $Zn^{2+}$ and/or the surrounding residues, which is consistent with our data that CmIDCase cannot cocrystallize with 5caC and has a much weaker activity

towards 5caC. This suggests that 5caC might interact with the metal ion and/or the surrounding residues in a slightly different way. In addition, as the substrate is a 5caC-containing DNA, the DNA decarboxylase should contain a DNA-binding domain, which does not exist in IDCase, and hence its size should be larger than that of IDCase. It is expected that the 5caC base in DNA needs to be flipped out and inserts into the active site of the DNA decarboxylase for the catalytic reaction.

## Materials and Methods

### Cloning, expression and purification of CmIDCase and MaIDCase

The full-length *C. militaris* and *M. anisopliae* IDCase genes were amplified by PCR from the genomic DNAs of *C. militaris* and *M. anisopliae*, respectively (kind gifts from Dr Chengshu Wang, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences). Each gene was inserted into the pET-28b vector (Novagen) that attaches a 6× His tag at the C-terminus and the pET-28Sumo vector (Novagen) that attaches a 6× His tag plus a SUMO tag at the N-terminus. The plasmids were transformed into *E. coli* BL21 (DE3) Codon Plus strain (Novagen). The transformed bacterial cells were grown in LB medium supplemented with 0.05 mg/ml kanamycin at 37 °C until $OD_{600}$ reached 0.6, and then the protein expression was induced with 0.2 mM IPTG at 16 °C overnight. The cells were collected, resuspended in a lysis buffer (50 mM Tris-HCl, pH 8.0, 300 mM NaCl, 5 mM β-mercaptoethanol, 10% glycerol, and 1 mM PMSF), and lysed on ice by sonication. The cell lysate was precipitated by centrifugation at 18 000× *g* and the supernatant was used for protein purification.

The 6× His-tagged proteins were purified by affinity chromatography using a Ni-NTA Superflow column (Qiagen). The 6× His-Sumo-tagged proteins were first purified by affinity chromatography using a Ni-NTA column, and the tag was cleaved by the Ulp1 protease at 4 °C for 6 h and then removed by a second step Ni-NTA column purification. The target proteins were further purified by gel filtration using a Superdex G200 16/60 column (GE Healthcare). The purified proteins were of high purity (above 95%) as analyzed by SDS-PAGE, and were concentrated to about 10 mg/ml in a storage buffer (20 mM Tris-HCl, pH 8.0, 300 mM NaCl, and 1 mM DTT) for structural and biochemical studies.

Constructs of the CmIDCase mutants containing point mutations were generated using the QuikChange® Site-Directed Mutagenesis kit (Strategene) and verified by DNA sequencing. Expression and purification of the mutants were the same as the wild-type enzyme.

### Crystallization, diffraction data collection, and structure determination

Crystallization was performed using the sitting drop vapor diffusion method at 16 °C by mixing equal volumes (1 μl) of the protein solution and the reservoir solution and equilibrated against 0.5 ml of the reservoir solution. For obtaining crystals of CmIDCase and its mutants in complexes with 5caU (Alfa Aesar), 5caC (Alfa Aesar), 5niU (Sigma), and U (Sigma), the protein was incubated with the ligand at a molar ratio of 1:8 prior to crystallization.

Crystals of CmIDCase in apo form and in complexes with U and 5niU were grown in drops consisting of the 6× His-tagged protein solution and the reservoir solution (0.1 M Tris-HCl, pH 8.5, 0.2 M MgCl₂, and 25% PEG3350). Co-crystallization of CmIDCase with 5caC yielded crystals of CmIDCase in apo form. Crystals of the D323N mutant in apo form were grown in drops consisting of the tag-removed protein solution and the same reservoir solution in the presence of 5caC; and crystals of the D323N and D323A mutants in complexes with 5caU were grown in drops consisting of the tag-removed protein solution and the same reservoir solution. Crystals of the H195A mutant in apo form were grown in drops consisting of the tag-removed protein solution and the reservoir solution (0.1 M sodium citrate, pH 5.6, 0.2 M NH₄Ac, and 30% PEG4000) in the presence of 5caU. Crystals of MaIDCase in apo form were grown in drops consisting of the 6× His-tagged protein solution and the reservoir solution (0.1 M BIS-TRIS propane, pH 7.0, and 60% (v/v) tacsimate™). Diffraction data were collected at 100 K at beamline 17U of Shanghai Synchrotron Radiation Facility, and processed using HKL2000 [40]. The statistics of the diffraction data are summarized in Table 1.

The structure of the apo CmIDCase was solved by the molecular replacement (MR) method using Phenix [41] and the structure of human ACMSD in complex with DHAP (PDB code 2WM1) [30] as the search model. All of the other structures of the wild-type or mutant CmIDCase and MaIDCase were solved by the MR method using the apo CmIDCase structure as the search model. Model building was performed using Coot [42] and structure refinement was carried out using Phenix [41] and Refmac5 [43]. Stereochemistry of the structure models was analyzed using Procheck [44]. Structural analysis was carried out using programs in CCP4 [45] and the PISA server [46]. All graphics were generated using Pymol (http://www.pymol.org). Statistics of the structure refinement and the structure models are summarized in Table 1.

*Decarboxylation activity assay using HPLC*

Decarboxylation activities of the wild-type and mutant CmIDCase and MaIDCase to convert 5caU to U or 5caC to C were analyzed using a sensitive HPLC method. Specifically, the reaction mixture consisted of 75 nM of the tag-removed enzyme, 0.3 mM 5caU (or 5caC), and 50 mM Tris-HCl (pH 7.4) in a total volume of 50 μl. The reaction took place at 30 °C for 15 min for the 5caU to U conversion or 36 h for the 5caC to C conversion, and then was stopped by addition of 0.3 mM 5niU, a potent inhibitor of IDCases [28]. The reaction mixture was analyzed using an Agilent 1200 HPLC instrument (Agilent Technologies) with an AQ-C18 column (5-μm particle size, 25 cm × 4.6 mm). The mobile phase was 20 mM NH₄Ac (pH 5.2) running at the rate of 0.6 ml/min, and the detectors were set at 260/280/300 nm. The standard 5caU, U, 5caC, and C were used as references.

*Decarboxylation activity assay using spectrophotometry*

Decarboxylation activities of the wild-type and mutant CmIDCase and MaIDCase to convert 5caU to U were also assayed using a spectrophotometric method described previously [28]. Briefly, the reaction mixture consisted of 50 mM Tris-HCl (pH 7.4), 0.15 μM of the tag-removed enzyme, and a varied concentration (10 to 80 μM) of the substrate mimic 2-thioIOA (Sigma) in a total volume of 1 ml. The conversion of 2-thioIOA to 2-thio-uracil was monitored by measuring the decrease of absorption at 334 nm using a Beckman DU800 spectrophotometer (Beckman Coulter).

The apparent kinetic parameters $K_m$ and $k_{cat}$ were determined by fitting the kinetic data to the Michaelis-Menten equation using a nonlinear regression analysis implemented in GraphPad Prism 5. All the experiments were performed at 25 °C in triplicates.

*Accession codes*

The crystal structures of the apo CmIDCase, the CmIDCase-5niU complex, the CmIDCase-U complex, the apo H195A CmIDcase, the apo H195A CmIDcase plus Zn, the apo D323N CmIDcase, the D323N-5caU complex, the D323A-5caU complex, and the apo MaIDCase have been deposited with the Protein Data Bank under accession codes 4HK5, 4HK6, 4HK7, 4LAN, 4LAO, 4LAK, 4LAM, 4LAL and 4HJW, respectively.

# References

1   Chow JC, Yen Z, Ziesche SM, Brown CJ. Silencing of the mammalian X chromosome. *Annu Rev Genom Hum Genet* 2005; **6**:69-92.
2   Morison IM, Ramsay JP, Spencer HG. A census of mammalian imprinting. *Trends Genet* 2005; **21**:457-465.
3   Miranda TB, Jones PA. DNA methylation: the nuts and bolts of repression. *J Cell Physiol* 2007; **213**:384-390.
4   Suzuki MM, Bird A. DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet* 2008; **9**:465-476.
5   Robertson KD. DNA methylation and human disease. *Nat Rev Genet* 2005; **6**:597-610.
6   Gal-Yam EN, Saito Y, Egger G, Jones PA. Cancer epigenetics: modifications, screening, and therapy. *Annu Rev Med* 2008; **59**:267-280.
7   Ooi SK, Bestor TH. The colorful history of active DNA demethylation. *Cell* 2008; **133**:1145-1148.
8   Wu SC, Zhang Y. Active DNA demethylation: many roads lead to Rome. *Nat Rev Mol Cell Biol* 2010; **11**:607-620.
9   Tahiliani M, Koh KP, Shen Y, *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 2009; **324**:930-935.
10  Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, Zhang Y. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* 2010; **466**:1129-1133.
11  He YF, Li BZ, Li Z, *et al.* Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* 2011; **333**:1303-1307.
12  Ito S, Shen L, Dai Q, *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* 2011; **333**:1300-1303.
13  Pfaffeneder T, Hackner B, Truss M, *et al.* The discovery of

5-formylcytosine in embryonic stem cell DNA. *Angew Chem Int Ed Engl* 2011; **50**:7008-7012.

14 Zhang L, Lu X, Lu J, *et al.* Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. *Nat Chem Biol* 2012; **8**:328-330.

15 Schiesser S, Hackner B, Pfaffeneder T, *et al.* Mechanism and stem-cell activity of 5-carboxycytosine decarboxylation determined by isotope tracing. *Angew Chem Int Ed Engl* 2012; **51**:6516-6520.

16 Jones ME. Pyrimidine nucleotide biosynthesis in animals: genes, enzymes, and regulation of UMP biosynthesis. *Annu Rev Biochem* 1980; **49**:253-279.

17 Zrenner R, Stitt M, Sonnewald U, Boldt R. Pyrimidine and purine biosynthesis and degradation in plants. *Annu Rev Plant Biol* 2006; **57**:805-836.

18 Carter NS, Yates P, Arendt CS, Boitz JM, Ullman B. Purine and pyrimidine metabolism in *Leishmania*. *Adv Exp Med Biol* 2008; **625**:141-154.

19 Carreras CW, Santi DV. The catalytic mechanism and structure of thymidylate synthase. *Annu Rev Biochem* 1995; **64**:721-762.

20 Costi MP, Ferrari S, Venturelli A, Calo S, Tondi D, Barlocco D. Thymidylate synthase structure, function and implication in drug discovery. *Curr Med Chem* 2005; **12**:2241-2258.

21 Fink RM, Fink K. Utilization of radiocarbon from thymidine and other precursors of ribonucleic acid in *Neurospora crassa*. *J Biol Chem* 1962; **237**:2289-2290.

22 Abbott MT, Schandl EK, Lee RF, Parker TS, Midgett RJ. Cofactor requirements of thymine 7-hydroxylase. *Biochim Biophys Acta* 1967; **132**:525-528.

23 Watanabe MS, McCroskey RP, Abbott MT. The enzymatic conversion of 5-formyluracil to uracil 5-carboxylic acid. *J Biol Chem* 1970; **245**:2023-2026.

24 Holme E, Lindstedt G, Lindstedt S, Tofft M. 18-O studies of the 2-ketoglutarate-dependent sequential oxygenation of thymine to 5-carboxyuracil. *J Biol Chem* 1971; **246**:3314-3319.

25 Liu CK, Shaffer PM, Slaughter RS, McCroskey RP, Abbott MT. Stoichiometry of the pyrimidine deoxyribonucleoside 2′-hydroxylase reaction and of the conversions of 5-hydroxymethyluracil to 5-formyluracil and of the latter to uracil-5-carboxylic acid. *Biochemistry* 1972; **11**:2172-2176.

26 Liu CK, Hsu CA, Abbott MT. Catalysis of three sequential dioxygenase reactions by thymine 7-hydroxylase. *Arch Biochem Biophys* 1973; **159**:180-187.

27 Palmatier RD, McCroskey RP, Abbott MT. The enzymatic conversion of uracil 5-carboxylic acid to uracil and carbon dioxide. *J Biol Chem* 1970; **245**:6706-6710.

28 Smiley JA, Angelot JM, Cannon RC, Marshall EM, Asch DK. Radioactivity-based and spectrophotometric assays for isoorotate decarboxylase: identification of the thymidine salvage pathway in lower eukaryotes. *Anal Biochem* 1999; **266**:85-92.

29 Smiley JA, Kundracik M, Landfried DA, Barnes Sr VR, Axhemi AA. Genes of the thymidine salvage pathway: thymine-7-hydroxylase from a *Rhodotorula glutinis* cDNA library and iso-orotate decarboxylase from *Neurospora crassa*. *Biochim Biophys Acta* 2005; **1723**:256-264.

30 Garavaglia S, Perozzi S, Galeazzi L, Raffaelli N, Rizzi M. The crystal structure of human alpha-amino-beta-carboxy-muconate-epsilon-semialdehyde decarboxylase in complex with 1,3-dihydroxyacetonephosphate suggests a regulatory

link between NAD synthesis and glycolysis. *FEBS J* 2009; **276**:6615-6623.

31 Martynowski D, Eyobo Y, Li T, Yang K, Liu A, Zhang H. Crystal structure of alpha-amino-beta-carboxymuconate-epsilon-semialdehyde decarboxylase: insight into the active site and catalytic mechanism of a novel decarboxylation reaction. *Biochemistry* 2006; **45**:10412-10421.

32 Wilson DK, Rudolph FB, Quiocho FA. Atomic structure of adenosine deaminase complexed with a transition-state analog: understanding catalysis and immunodeficiency mutations. *Science* 1991; **252**:1278-1284.

33 Kinoshita T, Nakanishi I, Terasaka T, *et al.* Structural basis of compound recognition by adenosine deaminase. *Biochemistry* 2005; **44**:10562-10569.

34 Porter DJ, Austin EA. Cytosine deaminase. The roles of divalent metal ions in catalysis. *J Biol Chem* 1993; **268**:24005-24011.

35 Ireton GC, McDermott G, Black ME, Stoddard BL. The structure of *Escherichia coli* cytosine deaminase. *J Mol Biol* 2002; **315**:687-697.

36 Seibert CM, Raushel FM. Structural and catalytic diversity within the amidohydrolase superfamily. *Biochemistry* 2005; **44**:6383-6391.

37 Olsson MHM, Søndergaard CR, Rostkowski M, Jensen JH. PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. *J Chem Theory Comput* 2011; **7**:525-537.

38 Liu JQ, Kurihara T, Miyagi M, Esaki N, Soda K. Reaction mechanism of L-2-haloacid dehalogenase of *Pseudomonas sp*. YL. Identification of Asp10 as the active site nucleophile by 18O incorporation experiments. *J Biol Chem* 1995; **270**:18309-18312.

39 Reinhardt LA, Svedruzic D, Chang CH, Cleland WW, Richards NGJ. Heavy atom isotope effects on the reaction catalyzed by the oxalate decarboxylase from *Bacillus subtilis*. *J Am Chem Soc* 2003; **125**:1244-1252.

40 Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 1997; **276**:307-326.

41 Adams PD, Afonine PV, Bunkoczi G, *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 2010; **66**:213-221.

42 Emsley P, Cowtan K. COOT: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 2004; **60**:2126-2132.

43 Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 1997; **53**:240-255.

44 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.

45 Winn MD, Ballard CC, Cowtan KD, *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 2011; **67**:235-242.

46 Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 2007; **372**:774-797.

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