








# Structure of the catalytically active APOBEC3G bound to a DNA oligonucleotide inhibitor reveals tetrahedral geometry of the transition state

Received: 15 August 2022

Accepted: 4 November 2022

Published online: 19 November 2022


 Check for updates

Atanu Maiti <sup>1,5</sup>, Adam K. Hedger <sup>2,3,4,5</sup>, Wazo Myint <sup>1</sup>,  
Vanivilasini Balachandran<sup>1</sup>, Jonathan K. Watts <sup>3,4</sup>, Celia A. Schiffer <sup>2,4</sup> &  
Hiroshi Matsuo <sup>1</sup> 

APOBEC3 proteins (A3s) are enzymes that catalyze the deamination of cytidine to uridine in single-stranded DNA (ssDNA) substrates, thus playing a key role in innate antiviral immunity. However, the APOBEC3 family has also been linked to many mutational signatures in cancer cells, which has led to an intense interest to develop inhibitors of A3's catalytic activity as therapeutics as well as tools to study A3's biochemistry, structure, and cellular function. Recent studies have shown that ssDNA containing 2'-deoxy-zebularine (dZ-ssDNA) is an inhibitor of A3s such as A3A, A3B, and A3G, although the atomic determinants of this activity have remained unknown. To fill this knowledge gap, we determined a 1.5 Å resolution structure of a dZ-ssDNA inhibitor bound to active A3G. The crystal structure revealed that the activated dZ-H<sub>2</sub>O mimics the transition state by coordinating the active site Zn<sup>2+</sup> and engaging in additional stabilizing interactions, such as the one with the catalytic residue E259. Therefore, this structure allowed us to capture a snapshot of the A3's transition state and suggests that developing transition-state mimicking inhibitors may provide a new opportunity to design more targeted molecules for A3s in the future.

The apolipoprotein B mRNA editing enzyme catalytic polypeptide-like (APOBEC) family of proteins are cytosine deaminases which convert cytosine to uracil in DNA and RNA<sup>1,2</sup>. APOBEC1 is the founding member of this family, described to deaminate a cytidine in the apolipoprotein B pre-mRNA in plasma metabolism almost three decades ago<sup>3–5</sup>. Within this larger family, APOBEC3 proteins (A3s) catalyze the deamination of cytidines in single-stranded DNA (ssDNA) providing anti-viral activities in the innate immune response<sup>6–11</sup>. However, while those A3s play a beneficial role in anti-viral immunity, they may lead to drug

resistance<sup>12–15</sup>. Furthermore, A3A and A3B have been found to be a source of cancer-associated mutations in various types of cancer such as breast, bladder, head and neck, cervical, and lung cancer. Recent studies have shown that the dysregulation of A3 proteins is a major endogenous source of DNA mutations in approximately 75% of cancer types and >50% of all cancers analyzed<sup>16,17</sup>. A3 proteins can cause DNA mutations either alone or as a response to cancer therapies which can drive evolution of cancers, and A3 related mutations have been associated with poor prognosis and therapeutic resistance in cancers<sup>18–27</sup>.

<sup>1</sup>Cancer Innovation Laboratory, Frederick National Laboratory for Cancer Research, Frederick, MD, USA. <sup>2</sup>Institute for Drug Resistance, University of Massachusetts Chan Medical School, Worcester, MA, USA. <sup>3</sup>RNA Therapeutics Institute, University of Massachusetts Chan Medical School, Worcester, MA, USA. <sup>4</sup>Department of Biochemistry and Molecular Biotechnology, University of Massachusetts Chan Medical School, Worcester, MA, USA. <sup>5</sup>These authors contributed equally: Atanu Maiti, Adam K. Hedger.  e-mail: [hiroshi.matsuo@nih.gov](mailto:hiroshi.matsuo@nih.gov)

This propensity for driving the evolution of cancers and the involvement in antiviral drug resistance makes specific inhibition of these enzymes highly desirable.

Although interest in targeting APOBEC3 has intensified, screening efforts to find small molecule inhibitors have only yielded weak binding compounds that inhibit A3's at low micromolar concentrations<sup>28–31</sup>. More promising results emerged from efforts to develop sequence selective inhibitors of A3 enzymes using 2'-deoxy zebularine (dZ) incorporated into short oligonucleotides<sup>32,33</sup>. The resulting ssDNA containing 2'-deoxy-zebularine (dZ-ssDNA) inhibited A3 deamination at low micromolar concentrations, and exhibited specificity for A3 proteins<sup>32,33</sup>. However, dZ-ssDNA oligonucleotides are not drug-like and are unstable in blood and cells, therefore further modifications are required to be useful as tools to study biology of A3s in cells, or lead compounds for drug discovery and development campaigns. Therefore, further efforts are needed to better understand key determinants of dZ-ssDNA binding to A3s, and use that information to develop the next generation of potent A3 specific inhibitors.

In general, structural analysis is a powerful strategy to visualize how proteins engage their ligands and binding partners, thus providing a framework for understanding specificity and affinity of binding, as well as informing drug and tool compound design. For example, co-crystal and NMR structures of A3A<sup>34–36</sup> and chimeric A3B-CTD (catalytic domain)<sup>35</sup> with a ssDNA substrate bound have provided information with regard to the 5'-TC target sequence preference for A3A and A3B. The structure of A3G-CTD2 co-crystallized with substrate ssDNA provided information on how A3G-CTD specifically recognizes its 5'-TCCA preferred target sequence<sup>37</sup>. However, to date, there are no structures of an A3 enzyme complexed with a competitive active site inhibitor. Here, we present the 1.5 Å resolution co-crystal structure of a soluble variant of human A3G catalytic domain (A3G-CTD2) bound to a 9-nucleotide ssDNA inhibitor containing 2'-deoxy-zebularine at the target position, 5'-AATCCdZAAA (dZ-ssDNA). This high-resolution structure clearly showed that dZ was hydrated, generating a tight-binding tetrahedral intermediate that coordinates with the active site Zn<sup>2+</sup> ion. Thus, our structure captured the transition state, suggesting that hydrated dZ (dZ-H<sub>2</sub>O) mimics the transition state of the A3's deamination mechanism. Thus, our results lay the foundation for designing further inhibitors for A3 enzymes.

## Results

### Synthesis of dZ phosphoramidite and dZ-containing oligonucleotide

We synthesized the 2'-deoxy-zebularine (dZ) phosphoramidite according to previous literature<sup>32</sup>. Briefly, the 2-hydroxypyrimidine nucleobase was silylated and then glycosylated to 2'-deoxy-3',5'-di-*O*-p-toluoyl- $\alpha$ -D-erythro-pentofuranosyl chloride (Hoffer's chlorosugar). The 3'- and 5'- protecting group esters were removed using ammonia, to give an anomeric mixture of 2'-deoxy-zebularine nucleosides. After protecting the 5'-OH group with a dimethoxytrityl (DMT) group, we isolated the desired  $\beta$  anomer using column chromatography. This product was then 3'-*O*-phosphitylated using 2-cyanoethyl-*N,N*-diisopropylchlorophosphoramidite (see methods for further details and characterization). We then incorporated this dZ monomer into a short DNA oligonucleotide (5'-AATCCdZAAA) using an AKTA Oligopilot10 at a 12  $\mu$ mole synthesis scale in place of the target cytidine, in the

preferred 5'-CCCA- binding motif for A3G. The oligonucleotide sequence used in this work matches that of our previous work<sup>37,38</sup>.

### Activated 2'-deoxy-zebularine generates strong affinity to catalytically active A3G-CTD

We tested the binding affinity of the 9 nt dZ-oligonucleotide (DNA: 5'-AATCCdZAAA) for the catalytically active and soluble A3G-CTD variant (referred to throughout as A3G-CTD2)<sup>37</sup>, or the catalytically inactive A3G-CTD2 mutant (E259A mutation, referred to throughout as A3G-CTD2\*). For the active A3G-CTD2, the apparent dissociation constant,  $K_D$ , was determined to be  $0.27 \pm 0.07 \mu\text{M}$  (Table 1 and Supplementary Fig. 2a), whereas the  $K_D$  for the catalytically inactive A3G-CTD2 mutant was determined to be  $14.3 \pm 1.6 \mu\text{M}$ , representing a ~50 fold drop in affinity. Furthermore, the  $K_D$  for the equivalent DNA substrate containing 2'-deoxy-cytidine instead of dZ, (DNA: 5'-AATCCCAAA), to the inactive A3G-CTD2\* was determined to be  $5.8 \pm 0.8 \mu\text{M}$ .

We also tested the ability of the 5'-AATCCdZAAA oligonucleotide to inhibit deamination of the equivalent substrate by A3G-CTD2, using a <sup>1</sup>H-based NMR assay<sup>37,39</sup>. We determined the apparent  $K_m$  and  $V_{max}$  for the substrate at 200 nM enzyme concentration to be  $509 \pm 86 \mu\text{M}$  and  $33.4 \pm 4.4 \text{ min}^{-1}$  respectively (Table 2). In the presence of 50  $\mu\text{M}$  5'-AATCCdZAAA oligonucleotide inhibitor, the apparent  $K_m$  increased to  $8.8 \pm 1.8 \text{ mM}$  and the  $K_i$  was measured to be  $3.07 \pm 0.78 \mu\text{M}$  (Table 2). While the  $K_D$  value ( $0.27 \pm 0.07 \mu\text{M}$ ) and  $K_i$  value ( $3.07 \pm 0.78 \mu\text{M}$ ) are not exactly the same, this is likely due to variations in experimental conditions which include varied protein concentration and equilibration time. We confirmed the inhibition mode to be competitive through a Lineweaver–Burk plot (Supplementary Fig. 2b). These data indicate that the 5'-AATCCdZAAA oligonucleotide is a potent inhibitor for the deamination reaction of A3G.

### Co-crystal structure of A3G-CTD2 and 5'-AATCCdZAAA

We solved the high-resolution co-crystal structure of A3G-CTD2 bound to our dZ-oligonucleotide inhibitor characterized above (5'-AATCCdZAAA). The structure was determined to a resolution of 1.5 Å by molecular replacement using a previously determined structure of the catalytically inactive A3G-CTD2\*:ssDNA substrate complex (PDB 6BUX). The final refinement of the structure resulted in  $R_{work}/R_{free}$  of 0.17/0.19, respectively (Table 3). The structure was solved in the space group P2<sub>1</sub> and contains a single A3G-CTD2:dZ-ssDNA complex in the asymmetric unit, similar to the previous co-crystal structure of A3G-CTD2\* with ssDNA substrate (Fig. 1a). Previously for the inactive A3G-CTD2\*:ssDNA structure, we used 2'-deoxy-cytidine at the deamination target site (C<sub>0</sub>) and replaced the catalytic residue E259 with alanine to prevent catalysis<sup>37</sup>. In the new structure, we used active protein retaining E259, and replaced the target cytidine by a cytidine analog, 2'-deoxy-zebularine. This structure of an oligonucleotide inhibitor bound to fully catalytically competent (active) A3G, and, at 1.5 Å, represents the highest resolution A3 complex solved to date.

Given that our structure features the active A3G enzyme, we observed that 2'-deoxy-zebularine was converted to a tetrahedral hydration product; 4-(*R*)-hydroxy-3,4-dihydro-2'-deoxy-zebularine (dZ-H<sub>2</sub>O). In this hydration product, the N3 nitrogen of the zebularine nucleobase has become protonated, and a hydroxyl group added to the top face of the C4 carbon (Fig. 1b). Direct coordination of the C4-OH to the Zn<sup>2+</sup> ion is evident by the short Zn<sup>2+</sup>-O distance (2.01 Å) and

**Table 1 |  $K_D$  values from Micro Scale Thermophoresis**

Protein	Ligand	$K_D$
CTD2	5'-AAT CCdZ AAA	$0.27 \pm 0.07 \mu\text{M}$
CTD2*	5'-AAT CCdZ AAA	$14.3 \pm 1.6 \mu\text{M}$
CTD2*	5'-AAT CCC AAA	$5.8 \pm 0.8 \mu\text{M}$

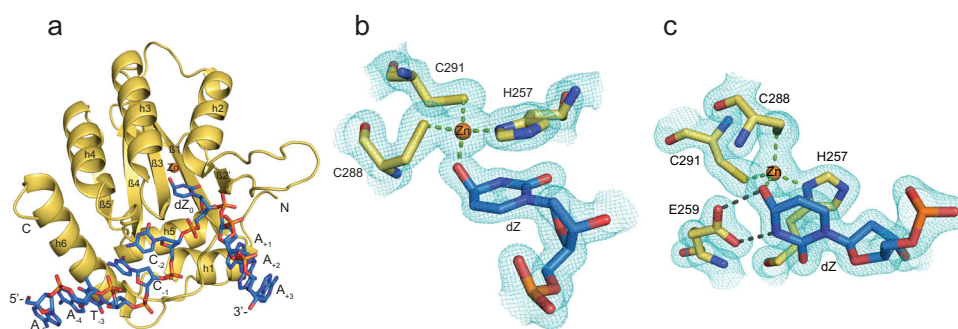
**Table 2 | Enzymatic parameters for A3G CTD2**

Substrate	Inhibitor	$K_m$	$V_{max}$	$K_i$
5'-AAT CCC AAA	n/a	$0.509 \pm 0.086 \text{ mM}$	$33.4 \pm 4.4 \text{ min}^{-1}$	n/a
5'-AAT CCC AAA	5'-AAT CCdZ AAA	$8.8 \pm 1.8 \text{ mM}$	$42.3 \pm 8.6 \text{ min}^{-1}$	$3.07 \pm 0.78 \mu\text{M}$

**Table 3 | Data collection and refinement statistics**

Data collection	
Space group	P2 <sub>1</sub>
Cell dimensions	
a, b, c (Å)	47.48, 47.12, 52.04
α, β, γ (°)	90.00, 103.18, 90.00
Resolution (Å)	50.00–1.50 (1.55–1.50)*
R <sub>merge</sub> (%)	8.6 (46.8)
R <sub>meas</sub> (%)	10.4 (56.1)
R <sub>pim</sub> (%)	5.8 (30.6)
I/σI	12.91 (1.72)
CC1/2	0.963 (0.752)
Completeness (%)	93.4 (98.9)
Redundancy	3.2 (3.2)
Refinement	
Resolution (Å)	34.50–1.50 (1.55–1.50)
No. of reflections	33634 (3534)
R <sub>work</sub> /R <sub>free</sub> (%)	17.25/19.44
No. of atoms	1933
Protein	1539
DNA	179
Ligand/ion	1 (Zn <sup>2+</sup> )
Water	212
B-factor	
Average B-factors (Å <sup>2</sup> )	36.70
Protein/DNA	35.70
Ligand/ion	27.60
Water	44.40
r.m.s deviations	
Bond lengths (Å)	0.005
Bond angles (°)	0.761

\*Values in parentheses are for highest-resolution shell



**Fig. 1 | Structure of active APOBEC3G catalytic domain (A3G-CTD2) with ssDNA containing 2'-deoxy-zebularine at the target site. a** The asymmetric unit contains one protein shown as cartoon (yellow) and one oligonucleotide inhibitor (blue) shown as sticks. N and C indicate the N- and C-terminal ends of the protein, 5'- and 3'- indicates 5' and 3' ends of the oligonucleotide. **b** A 2Fo-Fc electron density map contoured at 1.5  $\sigma$  is shown in cyan around the 2'-deoxy-zebularine hydrated intermediate (dZ-H<sub>2</sub>O), the Zn<sup>2+</sup> ion, and the Zn<sup>2+</sup> coordinating residues. The 3'-oxygen of the preceding nucleotide of dZ- is not shown. Zn<sup>2+</sup> is shown as an orange sphere. C, N, S and O atoms are colored yellow, navy blue, gold, and red respectively, for amino acid residues of the protein. Atoms in nucleotides are

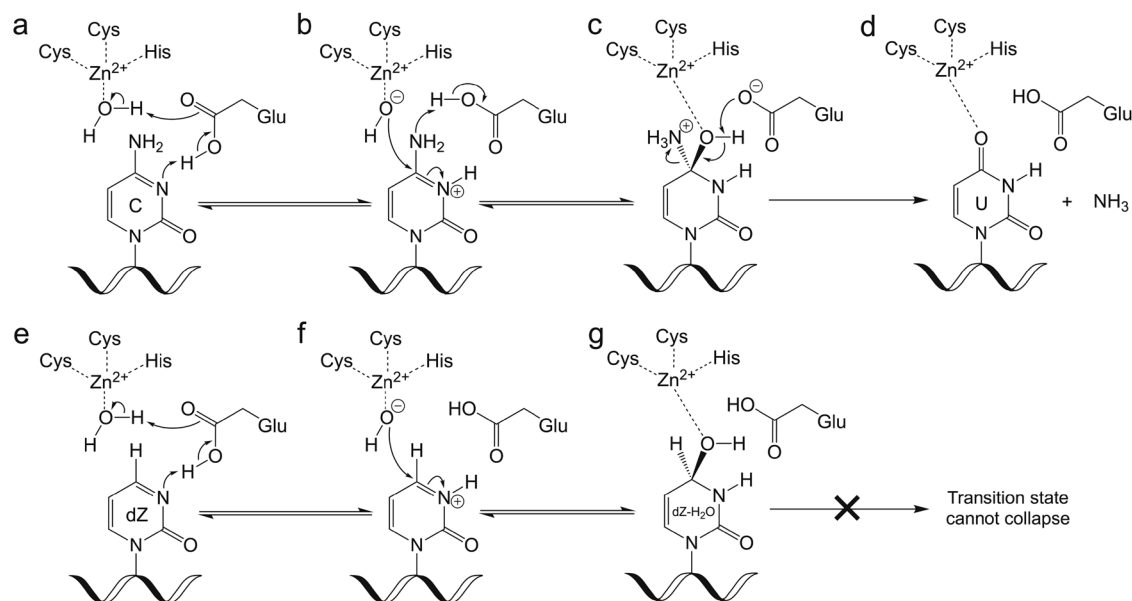
continuous electron density around the Zn coordination sites (Fig. 1b). Furthermore, we observed clear density of the key catalytic E259, with the sidechain forming hydrogen bonds with dZ-H<sub>2</sub>O (Fig. 1c). These results are in general agreement with the deamination mechanism proposed for bacterial cytidine deaminases (CDAs), related enzymes that deaminate single cytidine/2'-deoxycytidine<sup>40–50</sup>. Analogous to the CDAs, A3s also depend on a Zn<sup>2+</sup> located at the center of the substrate binding pocket (Fig. 1a), and the deamination reaction proceeds via a tetrahedral intermediate (Figs. 1c and 2). Therefore, our use of the active A3G-CTD2 for structural analysis has enabled the capture of the transition state, thus providing support for the proposed tetrahedral intermediate and offering an unprecedented view of the A3 catalytic mechanism<sup>49,50</sup>.

### Comparison between active A3G-CTD2:dZ-ssDNA and inactive A3G-CTD2\*:ssDNA structures

Backbone superposition of A3G-CTD2 (catalytically active, this structure) and A3G-CTD2\* (catalytically inactive, 6BUX) bound to ssDNA showed that the structures are almost identical, with a root mean square deviation (RMSD) for CA backbone atoms of 0.2 Å (superimposition is shown as Supplementary Fig. 3). Like our previous structure, all nine nucleotides of ssDNA including 2'-deoxy-zebularine (dZ) are well ordered and all five substrate specific nucleotides (5'-T<sub>-3</sub>C<sub>-2</sub>C<sub>-1</sub>dZ<sub>0</sub>A<sub>+1</sub>) form direct interactions with A3G. These interactions between the nucleotides flanking the active site (T<sub>-3</sub>, C<sub>-2</sub>, C<sub>-1</sub> and A<sub>+1</sub>) are essentially conserved (Fig. 3a, b). However, within the active site we see significant differences, including around the dZ nucleobase and Zn<sup>2+</sup> coordination residues (Fig. 3c).

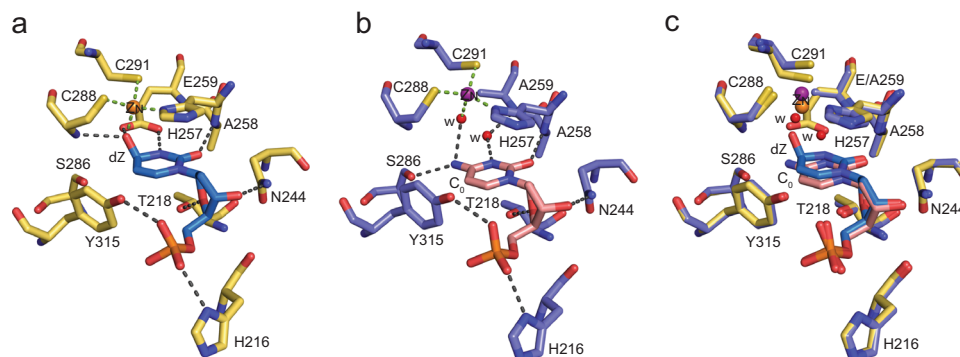
In the A3G-CTD2:dZ-ssDNA structure of the catalytically active complex, the essential catalytic residue E259 is oriented towards the zebularine nucleobase as expected (Fig. 1c), poised to carry out its acid/base role during catalysis. The side chain carboxylate of E259 makes two hydrogen bonds, to both the N3-H and Zn-coordinated C4-O (Fig. 3a). The oxygen of the C4-OH, is directly coordinated to the 4th Zn-coordination site, replacing the Zn-bound water molecule seen in the previous structure of inactive A3G. This provides direct structural evidence of the second step of the deamination mechanism (Fig. 2f), whereby the Zn-coordinated hydroxide has attacked the top face of the 2'-deoxy-zebularine C4 carbon, forming a tetrahedral transition state intermediate (Fig. 2g). To make this direct C4-O-Zn coordination

colored blue, navy blue, red, and orange for C, N, O, and P, respectively. Green dotted lines indicate the coordination bond with the Zn<sup>2+</sup> ion. Continuous electron density around the 2'-deoxy-zebularine hydrated intermediate, Zn<sup>2+</sup> ion and Zn<sup>2+</sup> coordinating residues support the formation of the 4-(R)-hydroxy-3,4-dihydro-2'-deoxy-zebularine intermediate where oxygen of C4-hydroxy directly coordinates with Zn<sup>2+</sup> ion. **c** The structure was rotated about 90 degrees from (b) along the axis of the 2'-deoxy-zebularine ring plane. A 2Fo-Fc electron density map showing E259 with the 2'-deoxy-zebularine hydrated intermediate, Zn<sup>2+</sup> and Zn<sup>2+</sup> coordinating residues. Gray dotted lines indicate hydrogen bonds between E259 and dZ-H<sub>2</sub>O.



**Fig. 2 | Mechanism of deamination by A3 enzymes on dC and dZ showing key tetrahedral intermediates.** Top panel, cytidine nucleobase; **a** Key active site Glu protonates the N3 nitrogen and absorbs a proton from the Zn-coordinated water molecule. **b** Zn-coordinated hydroxide ion attacks the activated C4 carbon as the Glu protonates the C4-amine. **c** The key tetrahedral intermediate collapses as the Glu absorbs the C4-OH proton, a carbonyl is formed, and ammonia leaves.

**d** Products formed (uracil and ammonia) after deamination is complete. Bottom panel, 2'-deoxy-zebularine nucleobase; **e** Same as for cytidine (panel **a**). **f** Same as for cytidine (panel **b**) but there is no C4-amine to be protonated. **g** Due to no C4-amine and the poor leaving group ability of H, the final product is hydrated-dZ (dZ-H<sub>2</sub>O), having added OH at C4 and H at N3) - a transition state mimic, which remains bound to the enzyme.



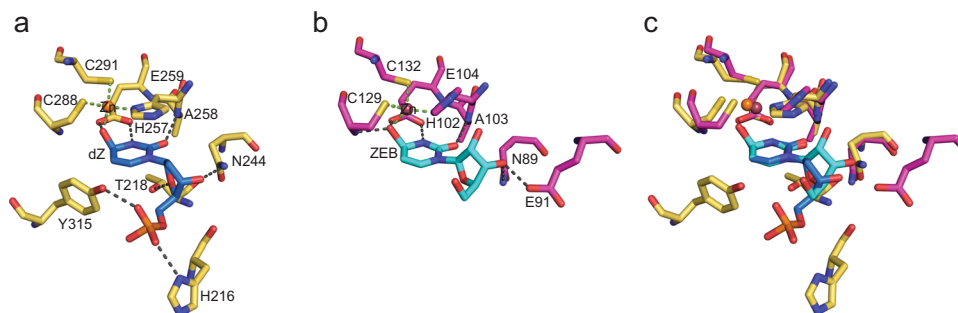
**Fig. 3 | Visualization of the target nucleotide in the active site and comparison with substrate (2'-deoxy-cytidine) bound to inactive enzyme.** The green and gray dotted lines indicate Zn<sup>2+</sup> coordination and hydrogen bonds, respectively. **a** Active site interactions of active A3G-CTD2 with the 2'-deoxy-zebularine hydrated intermediate (dZ-H<sub>2</sub>O) (PDB 7UXD). Zn<sup>2+</sup> is shown as an orange sphere. C, N, S and O atoms are colored yellow, navy blue, golden-yellow, and red, respectively, for amino acid residues of the protein. Atoms in nucleotides are colored blue, navy blue, red, and orange for C, N, O, and P, respectively. **b** Active site interactions of

inactive A3G-CTD2\* with 2'-deoxy-cytidine at target position (PDB 6BUX). The Zn<sup>2+</sup> and waters are shown as purple and red spheres, respectively. C, N, S and O atoms are colored light blue, navy blue, golden-yellow, and red, respectively, for amino acid residues of the protein. Atoms in nucleotides are colored pink, navy blue, red, and orange for C, N, O, and P, respectively. **c** Superimposition of two structures. Most of the active site interactions are conserved. However, significant differences in the plane of the target nucleotide's aromatic ring and Zn<sup>2+</sup> coordination are observed.

possible, the 2'-deoxy-zebularine nucleobase and Zn-coordinating residues adjust their positions accordingly. The pyrimidine ring of the hydrated 2'-deoxy zebularine (dZ-H<sub>2</sub>O) tilts upward -8° when compared to the target cytosine (C<sub>0</sub>) in our previous A3G-CTD2\*ssDNA structure. This tilt increases the distance between equivalent nucleobase atoms in the two structures gradually across the base from the anomeric C1' (0.36 Å) to the C4 atom (0.71 Å). In addition, the Zn<sup>2+</sup> ion is pulled down -0.8 Å from the inactive A3G-CTD2\*ssDNA structure. As a result, the Zn coordinating residues also shift down relative to the A3G-CTD2\*ssDNA structure, the SG atom of 288 C, SG atom of 291 C, and ND1 atom of 257H move down 0.34 Å, 0.99 Å and 0.31 Å, respectively (Fig. 3c). Overall, in the active A3G-CTD2:dZ-ssDNA structure the total Zn-O-C4 distance is 3.5 Å compared to a Zn-water-C4 distance of

5.4 Å in the inactive A3G-CTD2\*ssDNA structure. This variation makes sense as in the inactive structure the intact cytosine nucleobase is separated from the Zn<sup>2+</sup> ion by the water molecule, which becomes covalently attached to dZ in the active structure. The target cytosine is further stabilized in the active structure by the side chain of the catalytic E259 forming a direct hydrogen bond with the N3 nitrogen, in the inactive complex E259 is replaced with an ordered water coordinating between N3 and the backbone NH of A259 (Fig. 3b, c). Finally, the C4-NH<sub>2</sub> of the target cytosine also hydrogen bonds (2.8 Å) with the backbone carbonyl of S286, which is not possible with dZ due to lack of NH<sub>2</sub> group. However, the C4-OH of dZ-H<sub>2</sub>O instead hydrogen bonds (3.0 Å) with the backbone NH of C288. All other interactions within the active site of both structures are conserved.





**Fig. 4 | Comparison of interactions between the catalytic site of A3G-CTD2 and *E. coli* CDA.** **a** Active site interactions of A3G-CTD2 with the 2'-deoxy-zebularine hydrated intermediate (dZ-H<sub>2</sub>O) (PDB 7UXD). Zn<sup>2+</sup> is shown as an orange sphere. C, N, S and O atoms are colored yellow, navy blue, gold, and red, respectively, for amino acid residues of the protein. Atoms in nucleotides are colored blue, navy blue, red, and orange for C, N, O, and P, respectively. **b** Active site interactions of *E. coli* CDA with ribo-zebularine-hydrated intermediate (ZEB-H<sub>2</sub>O) (PDB 1CTU). Zn<sup>2+</sup> is

shown as a raspberry sphere. C, N, S and O atoms are colored magenta, navy blue, gold and red, respectively, for amino acid residues of the protein. Atoms in nucleotides are colored cyan, navy blue and red for C, N and O respectively. **c** Superimposition of the 2'-deoxy-zebularine hydrated intermediate (dZ-H<sub>2</sub>O) and the ribo-zebularine hydrated intermediate (ZEB-H<sub>2</sub>O). Conserved amino acid residues are shown as stick models.

### Comparison between active A3G-CTD2:dZ-ssDNA and cytidine deaminase (CDA):ribo-zebularine structures

CDA enzymes do not bind single-stranded DNA/RNA, and have very low amino acid sequence homology when compared to A3 enzymes (Supplementary Fig. 1b). However, because both catalyze the deamination of cytosine nucleobases, key structures of active sites are conserved between CDA and A3 proteins, and their catalytic mechanisms are likely similar. This has led to the hypothesis that A3s employ the same tetrahedral transition state, as described for CDAs<sup>42,44</sup>, and our structure now confirms this to be the case. In comparing our new A3G-CTD2:dZ-ssDNA structure with the *E. coli* CDA-ribo-zebularine structure we note the two structures show similar interactions within the active site, when aligned by overlapping the transition state mimics (Fig. 4). In both structures, the His-Cys-Cys coordination of the Zn<sup>2+</sup> ion is essentially identical, including direct contact with the C4-OH of the zebularine nucleobase [Zinc coordination: H257, C288 and C291 in A3G-CTD2 (Fig. 4a), and H102, C129 and C132 in *E. coli* CDA (Fig. 4b)]. The side chain carboxylic acid group of the key catalytic residue (E259 in A3G-CTD2 and E104 in *E. coli* CDA) makes two hydrogen bonds to the nucleobase; one to the N3-proton and another to the C4-hydroxyl oxygen. Similar hydrogen bonding is also observed between the zebularine-C2 carbonyl and the backbone amide NH of A258 in A3G-CTD2 and A103 in *E. coli* CDA, as well as for the C4-hydroxyl oxygen and the backbone amide NH of C288 in A3G-CTD2 and C129 in *E. coli* CDA. Finally, a hydrogen bonding interaction between the C3'-O of both ribose/2'-deoxyribose sugars and the side chain amino group of N244 in A3G-CTD2 and N89 in *E. coli* CDA is conserved. The omit map of hydrated ribo-zebularine (ZEB-H<sub>2</sub>O) and hydrated-dZ (dZ-H<sub>2</sub>O) contoured at 3 $\sigma$  shows the difference of resolution in these structures and the improved model fitting of dZ-H<sub>2</sub>O (Supplementary Fig. 4). Thus our 1.5 Å resolution A3G structure provides clear evidence of hydration of dZ at the N3 and C4 positions resulting in N3-H and C4-OH, respectively, which was previously proposed as a possibility based on the 2.3 Å resolution structure of the *E. coli* CDA:ribo-zebularine.

### Discussion

APOBEC3 (A3) enzymes have been promising, but largely elusive, antiviral, and anticancer targets<sup>7,51,52</sup>. While small molecule inhibitor development has not been successful, modified ssDNA oligonucleotides with the C<sub>0</sub> changed to 2'-deoxy-zebularine or 2'-deoxy-zebularine analogs have been shown to inhibit A3s<sup>32,33</sup> although not with high potency. Zebularine was initially identified as a cytidine deaminase (CDA) inhibitor; however, CDA inhibitors do not inhibit A3s, because unlike CDAs that modify the single nucleoside, A3s

deaminate cytidine in the context of ssDNA<sup>53</sup>, suggesting they may require additional features beyond those needed for CDA binding. The high resolution (1.5 Å) A3G-CTD2:dZ-ssDNA structure we present here now helps understand both similarities between CDAs and A3s and the observed differences. Despite the similarities of the transition state and the interactions within the catalytic site, the A3G-CTD2:dZ-ssDNA structure differs from the *E. coli* CDA:ribo-zebularine structure in several ways. *E. coli* CDA appeared to bind a single ribo-zebularine as a dimer. In fact, all CDA structures from *E. coli* to human showed dimer or tetramer association of proteins<sup>42,44,54–56</sup>, whereas A3G-CTD2 bind dZ-ssDNA as a monomer. Furthermore, 2'-deoxy-zebularine embedded in our ssDNA inhibitor exhibits a C2'-endo DNA-like conformation, as was also observed for the C<sub>0</sub> in our previous work<sup>37</sup>, and is consistent with the finding that the target 2'-deoxy-cytidine favors a C2'-endo sugar conformation for substrate binding and deamination<sup>38</sup>. Indeed, in all four *E. coli* CDA co-crystal structures with ribose-based inhibitors bound (FZEB, ZEB-H<sub>2</sub>O, DHZ, and DAC), the sugar adopts a C3'-endo conformation<sup>42,44</sup>. It is likely that ribose-based CDA inhibitors must be converted to their C2'-endo, DNA-like, equivalents to be potent inhibitors of A3 enzymes when incorporated into ssDNA oligonucleotides. Taken together, our work provides a molecular explanation for why CDA inhibitors are inactive against A3s despite highly similar catalytic mechanisms<sup>37,38,42,44</sup>. The active A3G structure described here shows significant changes within the active site relative to the previously solved inactive A3G structure<sup>37</sup>, thus providing key insights into the transition state of A3 enzymes. The catalytic glutamic acid forms two hydrogen bonds with the transition state analog at both the N3 hydrogen and the C4-hydroxyl oxygen. This hydroxyl oxygen forms the 4th binding partner of the tetrahedrally coordinated catalytic Zn<sup>2+</sup>, replacing the coordinated water molecule seen in the inactive structure. Beyond the active site, the rest of the complex between our active and inactive A3G structures display very little conformational change as judged by the low RMSD of the backbone (0.2 Å). Furthermore, all the interactions of specific nucleotides T<sub>-3</sub>, C<sub>-2</sub>, C<sub>-1</sub> and A<sub>+1</sub> are essentially conserved in both structures, which suggests that we can accurately predict structural models of A3A or A3B with dZ-ssDNA inhibitors as the A3A:ssDNA and A3B:ssDNA structures are available<sup>34–36</sup>. These structural models with the activated dZ/H<sub>2</sub>O may be used for designing inhibitors with higher affinity and specificity to A3A or A3B, as the catalytic domain of A3G recognizes nucleotides containing CC, while A3A and A3B recognizes nucleotides containing TC. Since A3A and A3B cause mutations in the human genome in cancer cells<sup>18–27</sup>, inhibitors, with TdZ, can likely inhibit both A3A and A3B, while avoiding A3G, would be beneficial.

Our structure provides information about the inhibitor binding and the transition state of the A3s deamination mechanism. Therefore, this new structural work offers key information needed for A3 inhibitor development and helps rationalize recent activity data obtained from dZ-ssDNA based inhibitors<sup>32,33</sup>. For example, the substitution of the C5 hydrogen in the target dC or dZ by a methyl group (5MedC or 5MedZ) or fluorine (5FdC or 5FdZ) in ssDNA made both poorer substrates and poorer inhibitors for A3G when compared to dC or dZ respectively<sup>32,33</sup>. This observation was rationalized by steric effects, as Me and F are larger than H. In both of our A3G-CTD2:dZ-ssDNA and A3G-CTD2\*:ssDNA structures, the positioning of the aromatic cytosine or zebularine nucleobase is perpendicular to the aromatic ring of Y315 at C5 and exhibits  $\pi$ - $\pi$  stacking (this structure and<sup>37</sup>). This perpendicular arrangement likely stabilizes the substrate and transition state intermediate in the catalytic site. Introduction of Me or F at this C5 position likely destabilizes this interaction and hence precludes these from being good inhibitors. The A3G-CTD2:dZ-ssDNA structure provides clear structural evidence for the activation of dZ to a transition state analog and explains the stronger binding affinity and inhibitory activity seen for catalytically active A3 enzymes.

Transition-state analogs have been developed to target many classes of enzymes, and these compounds typically achieve very strong binding affinity and improved specificity. In fact, transition-state analogs have become powerful drugs once optimized with structure-based drug design. These include oseltamivir (Tamiflu) which inhibits neuraminidase<sup>57–62</sup> and treats influenza or all of the HIV viral protease inhibitors<sup>63–72</sup>. Despite the emerging importance of A3s as potential drug targets, they have remained recalcitrant to small molecule drug discovery. Our structure captured the transition state and revealed critical interactions stabilizing the transition state of the A3's deamination mechanism, which is a critical insight needed for designing transition-state analogs. Our high-resolution co-crystal structure of A3G-CTD2:dZ-ssDNA thus opens further opportunities to develop a new class of A3 transition state-mimicking inhibitors with therapeutic potential in oncology and infectious diseases.

## Methods

### Synthesis and Characterization of 2'-deoxy-zebularine phosphoramidite

The dZ phosphoramidite was synthesized according to previous literature<sup>32</sup>, but was further purified by precipitation. For precipitation, the flash-purified solid (3.2 g) was dissolved in a small amount of anhydrous dichloromethane (~3 mL) and added dropwise to ~600 mL vigorously stirring cold hexane. The resulting solid was filtered and evaporated from chloroform to give the purified 2'-deoxy-zebularine phosphoroamidite, as an off-white crunchy foam. <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>, 500 MHz) was identical to that observed previously<sup>32</sup>, <sup>31</sup>P NMR (DMSO-*d*<sub>6</sub>, 202 MHz)  $\delta_{\text{ppm}}$  147.8, 147.5 (-1:1 diastereomers), plus a small amount of P(V) by-product (H-phosphonate) at 13.9 ppm. **HRMS (ESI)**: *m/z* Calcd. for C<sub>39</sub>H<sub>47</sub>N<sub>4</sub>O<sub>7</sub>P [M + H]<sup>+</sup> 715.3255; found, 715.3250. Full NMR and HRMS spectra shown in Supplementary Figs. 5–7.

### Synthesis and Characterization of a 2'-deoxy-zebularine containing oligonucleotide

The 9 nt dZ-ssDNA oligonucleotide was synthesized at 12- $\mu$ mole scale using an Akta OligoPilot synthesizer using standard methods on deoxy-adenosine (n-bz) 1000 Å 3'-Icaa controlled pore glass support (ChemGenes). DNA phosphoramidites were purchased commercially and diluted to 0.1 M in anhydrous acetonitrile (ChemGenes). The deoxy-zebularine phosphoramidite was synthesized as described above and diluted to 0.125 M in anhydrous acetonitrile. Detritylation was accomplished using 3% dichloroacetic acid in toluene (TEDIA). Oxidation was accomplished using 0.05 M iodine in pyridine/water (9:1 v/v, TEDIA). Benzylthiotetrazole was used as the activator (0.25 M in acetonitrile, TEDIA). Coupling time was 10 min. The oligonucleotide

was synthesized with 5'-DMT off, and treated with 10% diethylamine in acetonitrile while on the CPG support to remove cyanoethyl backbone protection. Deprotection was carried out using concentrated aqueous ammonium hydroxide for 8 h at room temperature. Longer exposure of dZ-containing oligonucleotides to ammonia resulted in visible side product formation by mass spectrometry. Deprotection was followed by immediate concentration using a centrifugal evaporator to remove ammonia, the aqueous solution of crude oligonucleotide was filtered (0.2  $\mu$ m), and then purified by ion-exchange on a preparative scale Agilent 1260 Infinity HPLC using an Agilent PL-SAX 10  $\mu$ m 1000 Å column (150 × 10 mm). HPLC was performed at 10 mL/min with a column temperature of 50 °C using a gradient of Buffer A: 70% water, 30% acetonitrile, and buffer B: 70% 1 M NaClO<sub>4</sub> in water containing 30% acetonitrile. UV absorbance was measured at 260 nm and the eluting oligonucleotide was fractionated and analyzed by LCMS. Fractions containing pure oligonucleotide were pooled and evaporated to dryness. The oligonucleotide was then desalted using a Nap-25 column (GE Healthcare) and evaporated to dryness again. The final dried oligonucleotide was resuspended in nuclease free water to a stock concentration of 4 mM for LCMS analysis, binding, inhibition, and crystallography experiments. LCMS characterization was carried out using a 6530 Accurate-Mass Q-TOF LC/MS (Agilent) linked to a pre-injection 1260 infinity HPLC (Agilent). Calc. MW for 5'-AATCCdZAAA oligonucleotide: 2660.8, Obs. MW: 2660.5. Full mass spectra shown in Supplementary Fig. 8.

### Protein expression and purification

All the variants of human A3G-CTD (A3G-CTD2 and A3G-CTD2\*) were expressed and purified as described previously<sup>37</sup>. Proteins used for crystallography and NMR experiment were expressed from pGEX6P-1 expression plasmid with Glutathione S-transferase (GST) tag and proteins used for MST assay were expressed from pET-28a plasmid with poly-Histidine tag (for Ni-NTA purification) in *E. coli* BL21(DE3) cells (Invitrogen). Cells were grown in LB media at 37 °C until reaching an optical density of 0.5–0.6 at 600 nm. Then, temperature was reduced to 17 °C and protein expression was induced for 18 h with 0.2 mM isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG). All the steps for protein purification were performed at 4 °C. *E. coli* cells were harvested by centrifugation and re-suspended in lysis buffer (either 50 mM sodium phosphate pH 7.3, 150 mM NaCl, 25  $\mu$ M ZnCl<sub>2</sub>, 2 mM DTT, and 0.002% Tween-20 for GST purification or 50 mM sodium phosphate pH 7.3, 150 mM NaCl, 50  $\mu$ M ZnCl<sub>2</sub>, 1 mM DTT, and 0.002% Tween-20 for Ni-NTA purification) and EDTA free protease inhibitor cocktail (Roche, Basel, Switzerland). The suspended cells were disrupted by sonication and then cell debris were separated by centrifugation at 48,384 g for 30 min. Supernatant containing desired protein was applied to either Glutathione-Sepharose resin (GE Healthcare Life Science) for GST purification or Ni-NTA Agarose resin (QIAGEN) for Ni-NTA purification, equilibrated with lysis buffer and agitated for about 2 h. For GST purification, protein-bound resin was washed with Pre-Scission Protease cleavage buffer (50 mM sodium phosphate, pH 7.5, 100 mM NaCl, 10  $\mu$ M ZnCl<sub>2</sub>, 2 mM DTT, and 0.002% Tween-20) and incubated with Pre-Scission protease (GE Healthcare Life Science) for 18 h. The supernatant containing the cleaved protein was separated from the resin by centrifugation and loaded on to HiLoad 16/600 Superdex 75 gel filtration column (GE Healthcare Life Science) equilibrated with 20 mM Bis-Tris (pH 6.5), 100 mM NaCl, 1 mM DTT, 0.01 mM ZnCl<sub>2</sub>, and 0.002% Tween-20. For Ni-NTA purification, protein-bound resin was washed with 50 mM sodium phosphate, pH 7.3, 1 M NaCl, 25  $\mu$ M ZnCl<sub>2</sub>, 1 or 2 mM DTT, and 0.002% Tween-20. Protein was eluted from resin in buffer containing 400 mM imidazole, 50 mM sodium phosphate, pH 7.3, 100 mM NaCl, 1 mM DTT, and 0.002% Tween-20. Eluted protein was loaded on to HiLoad 16/600 Superdex 75 gel filtration column equilibrated with 20 mM Bis-Tris pH 6.5, 100 mM NaCl, 1 mM DTT, 0.002% Tween-20,

and 20  $\mu\text{M}$   $\text{ZnCl}_2$ . For both GST and Ni-NTA purification, protein purity was analyzed by SDS-PAGE.

### Microscale Thermophoresis $K_D$ measurements

The affinity of A3G-CTD2\* and A3G-CTD2 for 5'-AATCCCCAAA substrate and 5'-AATCCdZAAA inhibitor were determined as dissociation constants,  $K_D$ , using a Monolith (Nano Temper Technologies, GmbH, Munich, Germany) microscale thermophoresis instrument. RED-tris-NTA fluorescent dye solution was prepared at 100 nM in the MST buffer (20 mM Bis-Tris pH 6.5, 100 mM NaCl, 1 mM DTT, 0.002% Tween 20). A3G-CTD2\* was mixed with dye at 100 nM and incubated for 30 min at room temperature followed by centrifugation at 15,000 g for 10 min. For the MST measurement, final concentrations of ssDNA solution was prepared at: 1 mM, 500  $\mu\text{M}$ , 250  $\mu\text{M}$ , 125.5  $\mu\text{M}$ , 62.5  $\mu\text{M}$ , 31.25  $\mu\text{M}$ , 15.62  $\mu\text{M}$ , 7.81  $\mu\text{M}$ , 3.91  $\mu\text{M}$ , 1.95  $\mu\text{M}$ , 976 nM, 488 nM, 244 nM, and 122 nM for 5'-AATCCCCAAA binding to A3G-CTD2\*; 995  $\mu\text{M}$ , 497.5  $\mu\text{M}$ , 248.75  $\mu\text{M}$ , 124.38  $\mu\text{M}$ , 62.19  $\mu\text{M}$ , 31.09  $\mu\text{M}$ , 15.55  $\mu\text{M}$ , 7.77  $\mu\text{M}$ , 3.89  $\mu\text{M}$ , 1.94  $\mu\text{M}$ , 971.68 nM, 485.84 nM, 242.92 nM, 121.46 nM, 60.73 nM for 5'-AATCCdZAAA binding to A3G-CTD2\*; 497.5  $\mu\text{M}$ , 248.75  $\mu\text{M}$ , 124.38  $\mu\text{M}$ , 62.19  $\mu\text{M}$ , 31.09  $\mu\text{M}$ , 15.55  $\mu\text{M}$ , 7.77  $\mu\text{M}$ , 3.89  $\mu\text{M}$ , 1.94  $\mu\text{M}$ , 971.68 nM, 485.84 nM, 242.92 nM, 121.46 nM, 60.73 nM, 3.04 nM, 1.52 nM, 759 pM, 380 pM, 190 pM, 95 pM, 47 pM, 24 pM, 12 pM for 5'-AATCCdZAAA binding to A3G-CTD2. Final protein concentration was at 50 nM. Measurements were performed using Nano Temper MST premium coated capillaries. Measurements were performed at 23 °C with 100% excitation power and 40% MST power. The experiment was repeated three times and data analysis was carried out using MO affinity analysis software (Nano Temper Technologies).

### Real-time NMR deamination and inhibition assays

The initial rates of deamination in the absence and presence of inhibitor were performed as previously described in Maiti et al. (2018) Nature Communications 9:2460. In brief, 200 nM of A3G-CTD2 was incubated with substrate in 50 mM  $\text{NaPO}_4$  pH 6.5, 100 mM NaCl, 0.002% Tween 20, 1 mM DTT, and 10  $\mu\text{M}$   $\text{ZnCl}_2$ . Formation of the uracil product was monitored by integration of the H5 uracil proton peak at 5.6 ppm from sequentially acquired  $^1\text{H}$  NMR spectra. Experiments were performed at 25 °C on a Bruker Avance III NMR spectrometer operating at a  $^1\text{H}$  Larmor frequency of 600 MHz. Initial rates were measured for 5'-AATCCCCAAA substrate concentrations of 100  $\mu\text{M}$ , 200  $\mu\text{M}$ , 400  $\mu\text{M}$ , 1 mM, and 2 mM.  $K_m$  and  $V_{\text{max}}$  for the substrate in the absence and presence of 50  $\mu\text{M}$  5'-AATCCdZAAA were determined from the slopes and the intercepts of Lineweaver-Burk plots, and the  $K_i$  was calculated from the  $K_m$  values.

### Crystal growth and data collection

Active A3G-CTD2 in 20 mM Bis-Tris pH 6.5, 100 mM NaCl, 1 mM DTT, 10  $\mu\text{M}$   $\text{ZnCl}_2$ , and 0.002% Tween-20 buffer was mixed with 50% molar excess of dZ-ssDNA (5'-AATCCdZAAA) and kept at 4 °C for overnight. The mixture was concentrated to ~450  $\mu\text{M}$  of protein using Amicon Ultra Centrifugal Filters unit 3KDa cutoff (Millipore Sigma). Crystals were grown at 4 °C, by sitting drop vapor diffusion method over a 300  $\mu\text{l}$  reservoir of 20% W/V PEG 6000, 50 mM di-sodium L-malate; pH 5.0 and 30 mM  $\text{CaCl}_2$  in a sitting drop 24 well crystallization plate from Hampton Research. Drops were set up by mixing 2  $\mu\text{l}$  of sample and 2  $\mu\text{l}$  of reservoir solution. Crystals were cryoprotected using reservoir solution containing 20% v/v glycerol and flash frozen in liquid nitrogen. X-ray diffraction data were collected at Southeast Regional Collaborative Access Team (SER-CAT) 22-ID beamline at the Advanced Photon Source, Argonne National Laboratory. The collected diffraction data were indexed, integrated, and scaled using HKL2000<sup>73</sup>. The crystals belong to the space group P2<sub>1</sub>.

### Structure determination and analysis

The structure was solved at 1.5 Å resolution by molecular replacement using the program Phaser<sup>74</sup>. Our previous structure of A3G-CTD2\* (PDB ID: 6BUX, ssDNA and solvents were removed) was used as search model. Model building of the protein and bound DNA were manually performed using the program Coot<sup>75</sup>. The DNA linking ligand for catalytically hydrated dZ; 4-(R)-hydroxy-3,4-dihydro-2'-deoxy-zebularine-5'-monophosphate {chemical formula: C<sub>9</sub> H<sub>15</sub> N<sub>2</sub> O<sub>8</sub> P and SMILES: Cl[C@@H]([C@H](O[C@H]IN2C=C[C@H](NC2=O)O)COP(=O)(O)O)}<sup>32</sup> was not used before in any crystal structure. However, we found a DNA linking ligand named 3,4-DIHYDRO-2'-DEOXYURIDINE-5'-MONOPHOSPHATE (ligand ID: DDN) in PDB Ligand Library with same chemical structure (Chemical formula: C<sub>9</sub> H<sub>15</sub> N<sub>2</sub> O<sub>8</sub> P and SMILES: Cl[C@@H]([C@H](O[C@H]IN2C=C[C@H](NC2=O)O)COP(=O)(O)O). We used the ligand DDN as our DNA linking ligand for catalytically hydrated dZ (4-(R)-hydroxy-3,4-dihydro-2'-deoxy-zebularine-5'-monophosphate) in our structure. The initial and final structure refinement were performed using the programs Refmac5<sup>76</sup> and Phenix<sup>77</sup>, respectively. The final model was refined to  $R_{\text{work}}/R_{\text{free}}$  values of 0.17/0.19 and was validated with the PDB validation tool and Molprobit<sup>78</sup>.

Pairwise rms deviation between A3G-CTD2 and A3G-CTD2\* structure were calculated using Dali<sup>79</sup>. Figures of structure models were generated by PyMOL (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.) and UCSF-Chimera<sup>80</sup>.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

The data that support this are available from the corresponding author upon request. The atomic coordinates and structural factors for the reported crystal structure have been deposited in the Protein Data Bank under the accession number 7UXD. Source data are provided with this paper.

### References

- Harris, R. S. & Liddament, M. T. Retroviral restriction by APOBEC proteins. *Nat. Rev. Immunol.* **4**, 868–877 (2004).
- Coticello, S. G., Thomas, C. J., Petersen-Mahrt, S. K. & Neuberger, M. S. Evolution of the AID/APOBEC family of polynucleotide (deoxy) cytidine deaminases. *Mol. Biol. Evol.* **22**, 367–377 (2005).
- Chan, L. RNA editing: exploring one mode with apolipoprotein B mRNA. *Bioessays* **15**, 33–41 (1993).
- Teng, B., Burant, C. F. & Davidson, N. O. Molecular cloning of an apolipoprotein B messenger RNA editing protein. *Science* **260**, 1816–1819 (1993).
- Navaratnam, N. et al. The p27 catalytic subunit of the apolipoprotein B mRNA editing enzyme is a cytidine deaminase. *J. Biol. Chem.* **268**, 20709–20712 (1993).
- Malim, M. H. APOBEC proteins and intrinsic resistance to HIV-1 infection. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **364**, 675–687 (2009).
- Feng, Y., Baig, T. T., Love, R. P. & Chelico, L. Suppression of APOBEC3-mediated restriction of HIV-1 by Vif. *Front Microbiol.* **5**, 450 (2014).
- Coticello, S. G., Langlois, M. A., Yang, Z. & Neuberger, M. S. DNA deamination in immunity: AID in the context of its APOBEC relatives. *Adv. Immunol.* **94**, 37–73 (2007).
- Sheehy, A. M., Gaddis, N. C., Choi, J. D. & Malim, M. H. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* **418**, 646–650 (2002).



10. Warren C. J., Van Doorslaer K., Pandey A., Espinosa J. M. & Pyeon D. Role of the host restriction factor APOBEC3 on papillomavirus evolution. *Virus Evol.* **1**, (2015).
11. Warren, C. J. et al. APOBEC3A functions as a restriction factor of human papillomavirus. *J. Virol.* **89**, 688–702 (2015).
12. Jern, P., Russell, R. A., Pathak, V. K. & Coffin, J. M. Likely role of APOBEC3G-mediated G-to-A mutations in HIV-1 evolution and drug resistance. *PLoS Pathog.* **5**, e1000367 (2009).
13. Kim, E. Y. et al. Human APOBEC3 induced mutation of human immunodeficiency virus type-1 contributes to adaptation and evolution in natural infection. *PLoS Pathog.* **10**, e1004281 (2014).
14. Delviks-Frankenberry, K. A. et al. Minimal contribution of APOBEC3-induced G-to-A hypermutation to HIV-1 recombination and genetic variation. *PLoS Pathog.* **12**, e1005646 (2016).
15. Mohammadzadeh, N. et al. Role of co-expressed APOBEC3F and APOBEC3G in inducing HIV-1 drug resistance. *Heliyon* **5**, e01498 (2019).
16. Burns, M. B. et al. APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature* **494**, 366–370 (2013).
17. Alexandrov, L. B. et al. The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–101 (2020).
18. Swanton, C., McGranahan, N., Starrett, G. J. & Harris, R. S. APOBEC enzymes: mutagenic fuel for cancer evolution and heterogeneity. *Cancer Discov.* **5**, 704–712 (2015).
19. Walker, B. A. et al. APOBEC family mutational signatures are associated with poor prognosis translocations in multiple myeloma. *Nat. Commun.* **6**, 6997 (2015).
20. Cescon, D. W., Haibe-Kains, B. & Mak, T. W. APOBEC3B expression in breast cancer reflects cellular proliferation, while a deletion polymorphism is associated with immune activation. *Proc. Natl Acad. Sci. USA* **112**, 2841–2846 (2015).
21. Law, E. K. et al. The DNA cytosine deaminase APOBEC3B promotes tamoxifen resistance in ER-positive breast cancer. *Sci. Adv.* **2**, e1601737 (2016).
22. Starrett, G. J. et al. The DNA cytosine deaminase APOBEC3H haplotype I likely contributes to breast and lung cancer mutagenesis. *Nat. Commun.* **7**, 12918 (2016).
23. Du, Y. et al. APOBEC3B up-regulation independently predicts ovarian cancer prognosis: a cohort study. *Cancer Cell Int.* **18**, 78 (2018).
24. Kim, Y. S. et al. Clinical implications of APOBEC3A and 3B expression in patients with breast cancer. *PLoS ONE* **15**, e0230261 (2020).
25. Chen, T. W. et al. APOBEC3A is an oral cancer prognostic biomarker in Taiwanese carriers of an APOBEC deletion polymorphism. *Nat. Commun.* **8**, 465 (2017).
26. Roper, N. et al. APOBEC mutagenesis and copy-number alterations are drivers of proteogenomic tumor evolution and heterogeneity in metastatic thoracic tumors. *Cell Rep.* **26**, 2651–2666 e2656 (2019).
27. Cannataro, V. L. et al. APOBEC-induced mutations and their cancer effect size in head and neck squamous cell carcinoma. *Oncogene* **38**, 3475–3487 (2019).
28. Li, M. et al. First-in-class small molecule inhibitors of the single-strand DNA cytosine deaminase APOBEC3G. *ACS Chem. Biol.* **7**, 506–517 (2012).
29. Olson, M. E., Li, M., Harris, R. S. & Harki, D. A. Small-molecule APOBEC3G DNA cytosine deaminase inhibitors based on a 4-amino-1,2,4-triazole-3-thiol scaffold. *ChemMedChem* **8**, 112–117 (2013).
30. King, J. J. et al. Structure-based design of first-generation small molecule inhibitors targeting the catalytic pockets of AID, APOBEC3A, and APOBEC3B. *ACS Pharm. Transl. Sci.* **4**, 1390–1407 (2021).
31. Yan, X. et al. Aromatic disulfides as potential inhibitors against interaction between deaminase APOBEC3G and HIV infectivity factor. *Acta Biochim. Biophys. Sin. (Shanghai)* **54**, 725–735 (2022).
32. Kvach, M. V. et al. Inhibiting APOBEC3 activity with single-stranded DNA containing 2'-deoxyzebularine analogues. *Biochemistry* **58**, 391–400 (2019).
33. Kvach, M. V. et al. Differential inhibition of APOBEC3 DNA-mutator isozymes by fluoro- and non-fluoro-substituted 2'-deoxyzebularine embedded in single-stranded DNA. *ChemBiochem* **21**, 1028–1035 (2020).
34. Kouno, T. et al. Crystal structure of APOBEC3A bound to single-stranded DNA reveals structural basis for cytidine deamination and specificity. *Nat. Commun.* **8**, 15024 (2017).
35. Shi, K. et al. Structural basis for targeted DNA cytosine deamination and mutagenesis by APOBEC3A and APOBEC3B. *Nat. Struct. Mol. Biol.* **24**, 131–139 (2017).
36. Liu, Y., Lan, W., Wang, C. & Cao, C. Two different kinds of interaction modes of deaminase APOBEC3A with single-stranded DNA in solution detected by nuclear magnetic resonance. *Protein Sci.* **31**, 443–453 (2022).
37. Maiti, A. et al. Crystal structure of the catalytic domain of HIV-1 restriction factor APOBEC3G in complex with ssDNA. *Nat. Commun.* **9**, 2460 (2018).
38. Solomon, W. C. et al. Mechanism for APOBEC3G catalytic exclusion of RNA and non-substrate DNA. *Nucleic Acids Res* **47**, 7676–7689 (2019).
39. Furukawa, A. et al. Structure, interaction and real-time monitoring of the enzymatic reaction of wild-type APOBEC3G. *EMBO J.* **28**, 440–451 (2009).
40. Cohen, R. M. & Wolfenden, R. Cytidine deaminase from *Escherichia coli*. Purification, properties and inhibition by the potential transition state analog 3,4,5,6-tetrahydrouridine. *J. Biol. Chem.* **246**, 7561–7565 (1971).
41. Frick, L., Yang, C., Marquez, V. E. & Wolfenden, R. Binding of pyrimidin-2-one ribonucleoside by cytidine deaminase as the transition-state analogue 3,4-dihydrouridine and the contribution of the 4-hydroxyl group to its binding affinity. *Biochemistry* **28**, 9423–9430 (1989).
42. Betts, L., Xiang, S., Short, S. A., Wolfenden, R. & Carter, C. W. Jr. Cytidine deaminase. The 2.3 Å crystal structure of an enzyme: transition-state analog complex. *J. Mol. Biol.* **235**, 635–656 (1994).
43. Carlow, D. C., Smith, A. A., Yang, C. C., Short, S. A. & Wolfenden, R. Major contribution of a carboxymethyl group to transition-state stabilization by cytidine deaminase: mutation and rescue. *Biochemistry* **34**, 4220–4224 (1995).
44. Xiang, S., Short, S. A., Wolfenden, R. & Carter, C. W. Jr. Transition-state selectivity for a single hydroxyl group during catalysis by cytidine deaminase. *Biochemistry* **34**, 4516–4523 (1995).
45. Carlow, D. C., Short, S. A. & Wolfenden, R. Role of glutamate-104 in generating a transition state analogue inhibitor at the active site of cytidine deaminase. *Biochemistry* **35**, 948–954 (1996).
46. Xiang, S., Short, S. A., Wolfenden, R. & Carter, C. W. Jr. The structure of the cytidine deaminase-product complex provides evidence for efficient proton transfer and ground-state destabilization. *Biochemistry* **36**, 4768–4774 (1997).
47. Carlow, D. & Wolfenden, R. Substrate connectivity effects in the transition state for cytidine deaminase. *Biochemistry* **37**, 11873–11878 (1998).
48. Carlow, D. C., Short, S. A. & Wolfenden, R. Complementary truncations of a hydrogen bond to ribose involved in transition-state stabilization by cytidine deaminase. *Biochemistry* **37**, 1199–1203 (1998).
49. Leija, C. et al. Pyrimidine salvage enzymes are essential for de novo biosynthesis of deoxypyrimidine nucleotides in *Trypanosoma brucei*. *PLoS Pathog.* **12**, e1006010 (2016).
50. Frances, A. & Cordelier, P. The emerging role of cytidine deaminase in human diseases: a new opportunity for therapy? *Mol. Ther.* **28**, 357–366 (2020).



51. Petljak, M. et al. Characterizing mutational signatures in human cancer cell lines reveals episodic APOBEC mutagenesis. *Cell* **176**, 1282–1294 e1220 (2019).
52. Buisson R. et al. Passenger hotspot mutations in cancer driven by APOBEC3A and mesoscale genomic features. *Science* **364**, (2019).
53. Hou, S. et al. Structural basis of substrate specificity in human cytidine deaminase family APOBEC3s. *J. Biol. Chem.* **297**, 100909 (2021).
54. Johansson, E., Mejlhede, N., Neuhard, J. & Larsen, S. Crystal structure of the tetrameric cytidine deaminase from *Bacillus subtilis* at 2.0 Å resolution. *Biochemistry* **41**, 2563–2570 (2002).
55. Chung, S. J., Fromme, J. C. & Verdine, G. L. Structure of human cytidine deaminase bound to a potent inhibitor. *J. Med. Chem.* **48**, 658–660 (2005).
56. Teh, A. H. et al. The 1.48 Å resolution crystal structure of the homotetrameric cytidine deaminase from mouse. *Biochemistry* **45**, 7825–7833 (2006).
57. Colman, P. M., Varghese, J. N. & Laver, W. G. Structure of the catalytic and antigenic sites in influenza virus neuraminidase. *Nature* **303**, 41–44 (1983).
58. Varghese, J. N., Laver, W. G. & Colman, P. M. Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 Å resolution. *Nature* **303**, 35–40 (1983).
59. Varghese, J. N., McKimm-Breschkin, J. L., Caldwell, J. B., Kortt, A. A. & Colman, P. M. The structure of the complex between influenza virus neuraminidase and sialic acid, the viral receptor. *Proteins* **14**, 327–332 (1992).
60. Kim, C. U. et al. Influenza neuraminidase inhibitors possessing a novel hydrophobic interaction in the enzyme active site: design, synthesis, and structural analysis of carbocyclic sialic acid analogues with potent anti-influenza activity. *J. Am. Chem. Soc.* **119**, 681–690 (1997).
61. von Itzstein, M. The war against influenza: discovery and development of sialidase inhibitors. *Nat. Rev. Drug Disco.* **6**, 967–974 (2007).
62. Tokuda, K. et al. Hydration of ligands of influenza virus neuraminidase studied by the fragment molecular orbital method. *J. Mol. Graph Model* **69**, 144–153 (2016).
63. Roberts, N. A. et al. Rational design of peptide-based HIV proteinase inhibitors. *Science* **248**, 358–361 (1990).
64. Dorsey, B. D. et al. L-735,524: the design of a potent and orally bioavailable HIV protease inhibitor. *J. Med. Chem.* **37**, 3443–3451 (1994).
65. Kempf, D. J. et al. ABT-538 is a potent inhibitor of human immunodeficiency virus protease and has high oral bioavailability in humans. *Proc. Natl Acad. Sci. USA* **92**, 2484–2488 (1995).
66. Kaldor, S. W. et al. Viracept (nelfinavir mesylate, AG1343): a potent, orally bioavailable inhibitor of HIV-1 protease. *J. Med. Chem.* **40**, 3979–3985 (1997).
67. Sham, H. L. et al. ABT-378, a highly potent inhibitor of the human immunodeficiency virus protease. *Antimicrob. Agents Chemother.* **42**, 3218–3224 (1998).
68. Turner, S. R. et al. Tipranavir (PNU-140690): a potent, orally bioavailable nonpeptidic HIV protease inhibitor of the 5,6-dihydro-4-hydroxy-2-pyrone sulfonamide class. *J. Med. Chem.* **41**, 3467–3476 (1998).
69. Robinson, B. S. et al. BMS-232632, a highly potent human immunodeficiency virus protease inhibitor that can be used in combination with other available antiretroviral agents. *Antimicrob. Agents Chemother.* **44**, 2093–2099 (2000).
70. Koh, Y. et al. Novel bis-tetrahydrofuranylurethane-containing non-peptidic protease inhibitor (PI) UIC-94017 (TMC114) with potent activity against multi-PI-resistant human immunodeficiency virus in vitro. *Antimicrob. Agents Chemother.* **47**, 3123–3129 (2003).
71. De Meyer, S. et al. TMC114, a novel human immunodeficiency virus type 1 protease inhibitor active against protease inhibitor-resistant viruses, including a broad range of clinical isolates. *Antimicrob. Agents Chemother.* **49**, 2314–2321 (2005).
72. Surleraux, D. L. et al. Discovery and selection of TMC114, a next generation HIV-1 protease inhibitor. *J. Med. Chem.* **48**, 1813–1822 (2005).
73. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
74. Bunkoczi, G. et al. Phaser.MRage: automated molecular replacement. *Acta Crystallogr D. Biol. Crystallogr* **69**, 2276–2286 (2013).
75. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D. Biol. Crystallogr* **60**, 2126–2132 (2004).
76. Murshudov, G. N. et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D. Biol. Crystallogr* **67**, 355–367 (2011).
77. Adams, P. D. et al. The Phenix software for automated determination of macromolecular structures. *Methods* **55**, 94–106 (2011).
78. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol. Crystallogr* **66**, 12–21 (2010).
79. Labarga, A., Valentin, F., Anderson, M. & Lopez, R. Web services at the European Bioinformatics Institute. *Nucleic Acids Res.* **35**, W6–W11 (2007).
80. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput Chem.* **25**, 1605–1612 (2004).

## Acknowledgements

This work was supported in part by a grant from the U.S. National Institutes of Health R01AI150478 for C.A.S. and H.M. For A.M., W.M., V.B. and H.M., this project has been funded in part with federal funds from the National Cancer Institute, National Institutes of Health, under contract 75N91019D00024. W.M. was supported in part by the NIH Office of Intramural Training and Education's Intramural AIDS Research Fellowship. A.H. was supported in part by a predoctoral fellowship from the PhRMA Foundation. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

## Author contributions

A.M. performed the expression and purification of protein with support from V.B.; A.M. performed the protein crystallization, X-ray data collection, crystal structure solution and refinement; A.M., A.H., C.A.S., and H.M. performed the structural analysis and interpretation; W.M. performed the MST and NMR catalytic assays and analyzed the data; A.H. synthesized the 2'-deoxy-zebularine phosphoramidite and the ssDNA oligonucleotide containing 2'-deoxy-zebularine, and J.W. supervised A.H.; A.M., W.M., A.H., C.A.S., and H.M. wrote the paper; C.A.S. and H.M. conceived, coordinated and oversaw the project; C.A.S. and H.M. secured funding.

## Funding

Open Access funding provided by the National Institutes of Health (NIH).

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-34752-1>.

**Correspondence** and requests for materials should be addressed to Hiroshi Matsuo.

**Peer review information** *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022