Crystal structure of the catalytic domain of HIV-1 restriction factor APOBEC3G in complex with ssDNA

Atanu Maiti, Wazo Myint, Tapan Kanai, Krista Delviks-Frankenberry, Christina Sierra Rodriguez, Vinay K. Pathak, Celia A. Schiffer & Hiroshi Matsuo

The human APOBEC3G protein is a cytidine deaminase that generates cytidine to deoxyuridine mutations in single-stranded DNA (ssDNA), and capable of restricting replication of HIV-1 by generating mutations in viral genome. The mechanism by which APOBEC3G specifically deaminates 5′-CC motifs has remained elusive since structural studies have been hampered due to apparently weak ssDNA binding of the catalytic domain of APOBEC3G. We overcame the problem by generating a highly active variant with higher ssDNA affinity. Here, we present the crystal structure of this variant complexed with a ssDNA substrate at 1.86 Å resolution. This structure reveals atomic-level interactions by which APOBEC3G recognizes a functionally-relevant 5′-TCCCA sequence. This complex also reveals a key role of W211 in substrate recognition, implicating a similar recognition in activation-induced cytidine deaminase (AID) with a conserved tryptophan.
Human APOBEC ("apolipoprotein B mRNA editing enzyme, catalytic polypeptide") proteins are single-stranded DNA (ssDNA) cytidine deaminases that catalyze the Zn-dependent deamination of a deoxy-cytidine, generating deoxy-uridine. The APOBEC family includes APOBEC1, APOBEC2, APOBEC3, APOBEC4, and activation-induced cytidine deaminase (AID). Among them, APOBEC3 proteins are genetically expanded in humans in response to the evolution of pathogens. As a result of this expansion, humans contain seven APOBEC3 proteins (APOBEC3A, 3B, 3C, 3D, 3F, 3G, and 3H), which are all encoded on chromosome 22. APOBEC3G (A3G) restricts human immunodeficiency virus type 1 (HIV-1)\(^{-}\), a finding that prompted extensive studies of APOBEC3 protein restriction of retroviruses and retrotransposons. Of the seven human APOBEC3 proteins, A3D, A3F, A3G, and A3H can restrict HIV-1, and hypermutation of the virus genomes by their deamination activity is the primary mechanism by which these A3 proteins restrict HIV-1. The APOBEC3 proteins catalyze the deamination of deoxy-cytidine introducing C-to-U modifications in newly synthesized ssDNA strands of the virus genome, which results in G-to-A mutations in (+)DNA as U is used as a template during (+)DNA strand synthesis. Although, there are many reports of deaminase-independent HIV restriction by APOBEC3 proteins, the significance of deaminase-independent mechanisms remains elusive and requires further study.

HIV-1 has developed a mechanism against APOBEC3 proteins by using one of its accessory proteins, namely viral infectivity factor or Vif. Vif physically interacts with HIV-relevant APOBEC3 proteins, and assembles host cellular proteins including an E3 ubiquitin ligase to trigger degradation of the APOBEC3 proteins through the ubiquitin–proteasome pathway. For A3D, A3F, and A3G, which contain two Zn\(^{2+}\)-binding motifs/domains, the catalytically inactive N-terminal domain (NTD) binds Vif as well as RNA, DNA, and other viral proteins. The C-terminal domain (CTD) of these APOBEC3 proteins is catalytically active, containing Zn\(^{2+}\)-binding motif HxE-x23,-C-x3,-C. The catalytic mechanism of cytosine/cytidine deamination has been studied biochemically and structurally using deaminases from Escherichia coli and yeast. Briefly, the hydroxide ion generated from a water molecule chelating Zn\(^{2+}\) attacks the C4 atom of cytosine, then the hydrogen is transferred to the carboxylate group of glutamic acid from the Zn\(^{2+}\)-binding motif; this hydrogen is ultimately transferred to the product ammonia.

Although all APOBEC3 proteins deaminate cytidines in ssDNA, they show differences in preferred hotspot sequences as 5′-CC for A3G and 5′-TC for other A3s (A3A can deaminate 5′-CC albeit to a lesser extent). A3G’s deamination mechanism may be more complicated than that of other A3 proteins because several groups have reported that A3G deaminates 5′-CC hotspots processively from the 3′-end to the 5′-end of ssDNA.

Three-dimensional structures of APOBEC3 proteins have emerged in the last 10 years as our laboratories along with others have solved nuclear magnetic resonance (NMR) and crystal structures of single domains of human APOBEC proteins. These structures are similar as they share the same secondary structure, including six helices and five β-strands, and one His-x25-Cys-x2-C zinc-binding motif. We and others have proposed several alternative ssDNA binding surfaces for A3G-CTD based on NMR and crystal structures of apo-form CTDs, yet none of these models are convincing because they lack atomic-level information of interactions between ssDNA and protein. Most recently, the crystal structures of A3A in complex with ssDNA containing a 5′-TC deamination motif have been reported by us and others. These A3A–ssDNA co-crystal structures revealed the interactions between A3A and the 5′-TC motif, and structural similarity with the crystal structure of Staphylococcus aureus tRNA adenosine deaminase (TadA) in complex with RNA.

In this study, we present the co-crystal structure of the A3G-CTD and ssDNA at 1.86 Å resolution. To overcome A3G-CTD weak DNA-binding affinity, we generated a catalytically enhanced variant of A3G-CTD that binds ssDNA stronger than wild-type. This A3G-CTD variant was co-crystallized with a 9 nucleotides ssDNA containing a 5′-TCCA target sequence with all nine nucleotides well resolved in the structure. The nucleotides within the 5′-TCCA target sequence show numerous interactions with protein, explaining the nucleotide specificity preferences. Furthermore, the backbone architecture of the protein changed upon ssDNA binding, enabling the target sequence to fit. These results provide fundamental insights into the mechanisms by which APOBEC3s recognize their specific substrate sequences.

**Results**

**Generation of a Hyperactive A3G-CTD Variant.** Although the CTD of A3G is catalytically active in vitro, detecting strong ssDNA-binding or purifying stable A3G-CTD–ssDNA complex has been challenging. To overcome this challenge, we designed and generated variants that are catalytically more active than the wild-type protein by introducing amino acid substitutions. The rationalization was that catalytically hyperactive variants may have increased affinity for the substrate, while retaining the intact structure and catalytic mechanism. A similar strategy worked well for us to generate a soluble A3G-CTD variant, namely CTD-2K3A, enabling our determination of the solution NMR structure. CTD-2K3A contains five amino acid substitutions including L234K, C243A, F310K, C321A, and C356A, and these substitutions altered neither catalytic activity, structure, nor HIV-1 restriction function, but increased solubility. Starting from CTD-2K3A, five additional substitutions were made, including P200A, N236A, P247K, Q318K, and Q322A. To increase the basicity of the region near the active site, we introduced lysine residues in loop3 and loop7; P247K and Q318K were chosen as they enhanced catalytic activity. Those lysine substitutions were combined with the other three substitutions, which we previously showed to increase catalytic activity. This variant, hereafter called CTD2, spans residues 191–384 of A3G, and contains P200A, L234K, N236A, C243A, P247K, F310K, Q318K, C321A, Q322A, and C356A substitutions (Fig. 1a).

The initial reaction speed of deamination of CTD2 was determined with that of wild-type A3G-CTD using a real-time NMR deamination assay we and others have previously used for enzyme kinetics analysis of A3G-CTD. The 9 nucleotides ssDNA 5′-AATCCCCAAA was used for the deamination assay, which contains the 5′-TCCA target sequence. 5′-TCCA is an optimized target sequence for A3G-CTD, and Yu et al. reported that 5′-CCCA is the preferential deamination sequence found in the minus strand of the HIV-1 genome. Representative NMR spectra (Supplementary Fig. 1a), illustrate CTD2 deaminates the 3′ C in the 5′-TCCA target sequence first, then the middle C is deaminated, but the 5′ TC was not acted upon by CTD2 as expected based on the preference of wild-type A3G. Indeed, the initial reaction speed was 20 times faster at pH 7.5 with 6.9 ± 0.2 reaction min\(^{-1}\) for CTD2 (red) and 0.3 ± 0.1 reaction min\(^{-1}\) for wild-type A3G-CTD (black) (Fig. 1b).

This catalytic activity also increased at lower pH 6.5 (Supplementary Fig. 1b) as we previously observed for A3G-CTD with 10.8 ± 0.3 reaction min\(^{-1}\) and 0.5 ± 0.1 reaction min\(^{-1}\) for CTD2 and A3G-CTD, respectively, suggesting that CTD2 retains the wild-type catalytic mechanism while increasing the reaction speed.
Since CTD2 exhibited greater catalytic activity than A3G-CTD (Fig. 1b), we tested whether it also exhibited increased binding affinity for ssDNA to the catalytically inactive E259A variant; this construct will be referred to as CTD2*. The 9 nucleotide substrate ssDNA 5’-AATCCCAAA was labeled with 6-FAM modification at the 3’-end for electrophoretic mobility shift assay (EMSA). Competitions with a non-specific (5’-AAAAA) or a specific (5’-AATCCCAAA) negative control ssDNA (fluorescent unprobed) showed that CTD2* specifically binds the substrate ssDNA (Fig. 1c). The affinity of CTD2* to the 9nt substrate ssDNA was further investigated by using microscale thermophoresis (MST) (Fig. 1d). The apparent dissociation constant, K_d, for the CTD2* was determined to be 55 ± 12 µM (Fig. 1d). An EMSA experiment of A3G-CTD for binding the same 9nt substrate ssDNA shows significantly weaker affinity than CTD2* as only a faint shifted band observed with 160 µM of A3G-CTD (Supplementary Fig. 1d), which is consistent with the observations previously reported by other laboratories. As expected from our previous studies, CTD2* appeared to have significantly less affinity to 5’-TCCdeoxy-UA (the first product) and 5’-TCdeoxy-Udeoxy-UA (final product) because K_d values for 9nt ssDNAs containing these product sequences were determined to be 150 ± 30 µM and 5.2 ± 0.8 mM, respectively (Supplementary Fig. 1c). Furthermore, we tested CTD2* for binding the 9nt RNA containing a target sequence 5’-rArArArUrCrCrArArA, and K_d was determined to be 1.5 ± 0.5 mM. Collectively, EMSA and MST experiments clearly indicate that CTD2* specifically binds the 5’-TCCCA target DNA sequence with high affinity, making this enzyme potentially amenable for structural studies.

To test whether full-length A3G containing wild-type NTD and CTD2 (NTD-CTD2) retained antiviral activity in vivo, we prepared HIV virus with increasing amounts of wild-type FLAG-A3G or FLAG-NTD-CTD2 in the presence or absence of HIV-1 Vif. Both FLAG-A3G and FLAG-NTD-CTD2 were functionally recognized by Vif and degraded as shown by the absence of A3G protein in the Vif+ lanes (Fig. 2a). As expected, neither FLAG-A3G nor FLAG-NTD-CTD2 blocked the infectivity of virus prepared in the presence of Vif in a single cycle replication assay (Fig. 2b). However, FLAG-A3G or FLAG-NTD-CTD2 potently inhibited viral infectivity when the HIV virus was prepared in the absence of Vif (Fig. 2c), and to similar extents when measuring comparable FLAG-A3G or FLAG-NTD-CTD2 protein expression levels from the producer cells (Fig. 2a) (t-test, p > 0.3). Introduction of mutation E259A into FLAG-NTD-CTD2 (FLAG-A3G), which abolishes a critical catalytic glutamate needed for deaminase activity, restored infectivity in the absence of Vif indicating that inhibition of HIV-1 replication by FLAG-NTD-CTD2 is largely deaminase-dependent (Supplementary Fig. 2a, b and c). Furthermore, no significant differences in viral infectivities were observed for virions encapsidating similar levels of FLAG-A3G or FLAG-NTD-CTD2 (t-test, p > 0.3) (Supplementary Fig. 2d and 2e). Overall, these data confirm that the full-length NTD-CTD2 blocks HIV-1 replication as potently as wild-type A3G.

Co-crystal Structure of CTD2* and ssDNA. The catalytically inactive CTD2 (CTD2*) was co-crystallized with the 9nt ssDNA containing a 5’-TCCCA target sequence (5’-AATCCCAAA). The
Remarkably, all five nucleotides in the target sequence interact with the protein in the co-crystal structure (Fig. 3b). As others have done, in the target sequence, nucleotides are numbered with the target cytidine at position 0 such as 5'-T_A3G_C_C_A3G-. Therefore, we describe here protein–DNA interactions for each nucleotide in the following sections.

The most remarkable interaction involving T_3 is the π–π stacking with W211 (Fig. 3c). T_3 also interacts with the following nucleotide C_2 by forming a hydrogen bond between the 5′-phosphate group of T_3 and the pyrimidine amino group of C_2. The Watson–Crick face of T_3 does not interact with the protein, whereas it forms a base pair with A_1 of the ssDNA in a neighboring asymmetric unit (Supplementary Fig. 3).

The nucleobase type of C_2 is recognized by the protein since Watson–Crick face of C_2 forms water-mediated hydrogen bonds between the pyrimidine carbonyl group and the mainchain amino proton of D316 as well as the guanidino group of R374 (Fig. 3c). In addition, the pyrimidine N3 atom of C_2 forms a hydrogen bond through an ordered water with the carboxyl group of D316 (Fig. 3c). Furthermore, the C_2 pyrimidine ring has a hydrophobic interaction with the indole ring of W211 (Fig. 3c), which creates a spatial restraint, favoring a pyrimidine nucleotide in this position.

Sugar puckers of nucleotides play a key role in shaping the structure of DNA and RNA strands. The deoxy-ribose of C_2 has a C3′-endo conformation, whereas all other eight nucleotides of the CTD2*-bound ssDNA have C2′-endo conformation. This is significant because DNA prefers the C2′-endo in aqueous solution, whereas RNA is predominantly in the C3′-endo conformation. The C3′-endo conformation of C_2 brings two neighboring backbone phosphorus atoms (which belong to C_1 and C_2) in close contact 5.8 Å, which is a typical distance for double-stranded RNAs in A-form. This spatial arrangement enables the 5′-phosphate group of C_2 to form a hydrogen bond with the guanidino group of R213. Interestingly, R213 is not conserved in other APOBEC3 proteins, except the charge is conserved with a lysine in AID, which we further describe in the Discussion.

The Watson–Crick face of C_1 has three direct interactions with the protein. The C2 carbonyl group forms a hydrogen bond with the mainchain amino proton of D317, the N3 atom forms a hydrogen bond with the mainchain amino proton of Q318, and the amino group forms a hydrogen bond with the mainchain amino proton of Q318. This hydrogen bond stabilizes the helix3 structure by forming an “N-cap” since F289 is located at the N-terminus of the helix. In addition, Q318 may provide further support in orienting D317 by interacting electrostatically as the ε-ammonium group of Q318 is located within 3.7 Å from the carboxyl group of D317 (Supplementary Fig. 4). Furthermore, the 5′-phosphate group of C_1 is supported by two water-mediated hydrogen bonds with the NE2 atom of H216 and the mainchain amino proton of R215 (Fig. 3d).

We observed electron density that fits a zinc ion (Zn^{2+}) chelated by H257, C288, C291, and additional density that fits a water molecule. The target cytosine (C_A3G) is tightly packed under the π–π interaction with Y315 (Fig. 3e). In addition to these π–π interactions, many hydrogen bonds support the position of target cytosine, including aromatic ring O2 to the mainchain amino proton of A258, and aromatic ring N3 to the mainchain amino proton of E259A through an ordered water molecule. Furthermore, the deoxy-ribose O3’ and O4 atoms form hydrogen bonds with the sidechain amino group of N244 and the hydroxyl group of T218, respectively, which
supports the 2′-endo conformation of deoxy-ribose of C0. The 5′-phosphate group of C0 is well-coordinated by interactions with the protein as it forms hydrogen bonds with the hydroxyl group of Y315 and ND1 atom of H216. Two hydrogen bonds provide key recognition of the amino group of C0, including one formed with the mainchain carbonyl group of S286, and another formed with the water carbonyl group coordinated by Zn2+. This Zn2+-bound water molecule is the key molecule to trigger the deamination by attacking the C4 position of cytosine18–20. These C0-interacting residues are conserved in all APOBEC3 proteins.
(except A3F has a serine instead of T218), and we and others have observed similar interactions in A3A–ssDNA complexes. Although the CTD2*–ssDNA complex showed that C2-endo sugar conformation of the target cytidine is required to fit in the catalytic pocket, the mechanism by which A3G discriminates RNA from deamination is not fully understood because the 2′ hydroxyl of the target cytidine can be replaced with a hydroxyl group without significant steric hindrance in the CTD2*–ssDNA complex structure.

The purine ring of A−1 stacks against the H216 imidazole ring (Fig. 3f). Since histidine forms stronger π–π stacking with a pyrimidine ring, this interaction selects purines in the complex structure. Hydrogen of the target cytidine can be replaced with a hydroxyl group without significant steric hindrance in the CTD2*–ssDNA complex structure.

In addition, the 5′ explanation for why 5′ purine ring than with a pyrimidine ring, this interaction selects purines in the complex structure. Hydrogen of the target cytidine can be replaced with a hydroxyl group without significant steric hindrance in the CTD2*–ssDNA complex structure. Overall, CTD2* recognizes 5′-C₂-C₁-C₀ through hydrogen bonds formed with the Watson–Crick faces, and T₃ and A₂₁ by using strong π–π interactions (Fig. 3g). Unusual sugar pucker of C₂₋₋ contributes in shaping the phosphate backbone to fit the ssDNA-binding site of CTD2*. The movements of CTD2* residues induced upon the ssDNA-binding will be discussed in the following section.

Discussion

Since we reported the first structure of the catalytic domain of A3G almost 10 years ago, structural studies of the A3G-CTD–ssDNA complex have been hampered by apparently weak binding of A3G-CTD to ssDNA. We have overcome this problem by generating a A3G-CTD variant that binds ssDNA with higher affinity than wild-type. P247K appears to be a key substitution that contributes to stabilizing the ssDNA-binding of CTD2* by providing an additional hydrogen bond to a backbone phosphate group located outside of the target sequence (Supplementary Fig. 3). Furthermore, non-Watson–Crick base pairs formed between neighboring asymmetric units in the CTD2*-ssDNA co-crystal may stabilize the crystallization of the complex (Supplementary Fig. 3). Critically, this highly active variant in the context of full-length A3G containing wild-type NTD restricted HIV-1 infection as potently as wild-type A3G in a Vif-dependent manner (Fig. 2).

A3G strongly prefers a cytidine at the −1 position, whereas A3A prefers a thymidine for that position. CTD2 and A3A form hydrogen bonds with the Watson–Crick face of the nucleotide at the −1 position to provide nucleobase specificity. For the CTD2*-ssDNA complex, the Watson–Crick face of C₁ forms three hydrogen bonds including amino group (NH₂) to carboxyl group of D316, carboxyl group to mainchain amino proton of D317, and N3 atom to mainchain amino proton of Q318k. If this cytidine is replaced by a thymidine, N3 and NH₂ would be replaced by NH and CO, respectively, resulting in the loss of two hydrogen bonds with the protein, which explains why A3G prefers cytidine over thymidine at the −1 position. Previously, Rausch et al. showed that N3 and NH₂ of cytosine ring of C−1 and C−2 are key for deamination efficiency, whereas 5-methyl deoxy-cytidine at C−1 or C−2 position are tolerated by A3G. This finding is consistent with the CTD2*-ssDNA structure showing that Watson–Crick faces of C₁ and C−2 interact with CTD2*, while the C5 positions do not have a contact with the protein (Fig. 3b). In addition to the recognition of the Watson–Crick face, the spatial coordination plays an important role in the recognition of C₁. Figure 4a, b shows a striking difference between the position of C₁ in the CTD2*-ssDNA complex and that of T₁ in the A3A-ssDNA complex. Interactions of T₃ and C₂ with CTD2* are important to position C₁ (Fig. 3b, c), whereas the A3A–ssDNA structures did not have interactions with nucleotides at −2 and −3 positions. A3A has a tyrosine (Y132) at the corresponding position to D317, and Fig. 3c shows that C₁ crashes into Y132 when we overlay the CTD2*-ssDNA structure onto the A3A–ssDNA structure. The significance of D316 and D317 in substrate specificity of A3G was originally reported by Holden and co-workers as they showed that D316R, D317R double substitutions enabled A3G to...

---

**Table 1 Crystal data collection and refinement statistics**

| Parameter                  | Value                  |
|----------------------------|------------------------|
| Space group                | P2₁                    |
| Cell dimensions (Å)        | 47.41, 47.24, 51.56    |
| α, β, γ (°)                | 90.00, 103.13, 90.00   |
| Resolution (Å)             | 50.00–1.86 (1.93–1.86) |
| Rmerge (%)                 | 8.3 (32.8)             |
| Rmeas (%)                  | 8.9 (35.3)             |
| Rp (%)                     | 3.3 (13.0)             |
| I/σ (°)                    | 24.1 (6.3)             |
| Completeness (%)           | 100.00 (100.00)        |
| Redundancy (%)             | 7.5 (7.3)              |
| Resolution (Å)             | 38.6–1.86 (1.92–1.86)  |
| No. of reflections         | 18,843 (1812)          |
| Rmerge/Rmeas (%)           | 17.89/21.11            |
| No. of atoms               | 1893                   |
| Protein                    | 1553                   |
| DNA                        | 179                    |
| Ligand (GOL)/Ion (Zn²⁺)    | 6/1                    |
| Water                      | 154                    |
| B-factor Average           | 33.9                   |
| Protein/DNA                | 32.6                   |
| Ligand/Ion                 | 43.3                   |
| Water                      | 36.6                   |
| ms deviations              | 0.006                  |
| Bond lengths (Å)           | 0.76                   |

*Values in parentheses are for the highest-resolution shell.*

---

Fig. 3 Structure of CTD2* in complex with ssDNA. a The asymmetric unit contains one protein (yellow) and one ssDNA (blue) molecule. A 2Fo-FC electron density map contoured at 1σ is shown in cyan around the ssDNA. Zn²⁺ ion is colored purple. N and C indicate the N- and C-terminal ends of the protein, respectively. b An enlarged view shows interactions between the 5′-TCCCA target sequence and the protein. Protein is colored yellow, 5′-TCCCA is blue, and amino acid sidechains interacting with DNA are shown as sticks. c–f Enlarged views show interactions between T₁, C₂, C₁, C₀, or A₃, and protein. C, N, and O atoms are colored yellow, blue, 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. Water molecules are shown as red spheres, and Zn²⁺ is shown as a purple sphere. Dotted lines indicate hydrogen bonds. In (c), the double arrow-headed line points to the neighboring backbone phosphorous atoms of C₁ and C₂. In (d), sidechains of F289 and Q318K are not shown. g Summary of the interactions between CTD2* and nucleotides in the 5′-TCCCA target sequence.
deaminate middle C and 3’ C of a 5’-CCG motif at the same reaction speed\textsuperscript{30}, whereas the wildtype A3G-CTD prefers 3’ C to the middle C by 45-fold\textsuperscript{41}. Another group showed that D317Y substitution changed A3G substrate preference to 5'-TC that is the A3A preferred target sequence\textsuperscript{49}. The co-crystal structure of CTD2* and ssDNA showed how D316 and D317 are involved in the recognition of C\textsubscript{−1}. Overall, spatial coordination as well as the interaction with the Watson–Crick face are critical for specific recognition of C\textsubscript{−1}.

We and others have reported structures of A3G-CTD wild-type\textsuperscript{30,31,35}, and a soluble variant, namely CTD-2K3A\textsuperscript{25,25}. None of these structures were complexed with ssDNA, and we chose our previous crystal structure of CTD2K3A, solved at 2.25 Å resolution (PDB ID# 3IR2\textsuperscript{25}), as a representative of apo-form CTD2*. We and others have reported structures of A3G-CTD wild-type\textsuperscript{30,31,35}, and a soluble variant, namely CTD-2K3A\textsuperscript{25,25}. None of these structures were complexed with ssDNA, and we chose our previous crystal structure of CTD2K3A, solved at 2.25 Å resolution (PDB ID# 3IR2\textsuperscript{25}), as a representative of apo-form CTD2*. Superimposition of structures of ssDNA-bound CTD2* (yellow) and apo-CTD2-K3A (gray) (Fig. 5a) reveals that even with the additional substitutions in CTD2* the backbone structures are essentially unchanged. Regions that are changed include loops 1, 3, and 7, which are intrinsically dynamic in solution as shown in NMR structures of CTD2-K3A\textsuperscript{23,24}. Loops 1 and 7 contain amino acid residues, which form numerous interactions with ssDNA; therefore, the structural changes were likely induced upon ssDNA binding.

Loop1 migrates toward loop7 upon ssDNA binding, with W211 demonstrating the biggest change with its \(\alpha\) Ca atom moved by 3.9 Å from the position found in the apo-form CTD2-K3A (Fig. 5b). This backbone change enables W211 to have \(\pi\)–\(\pi\) stacking interaction with T\textsubscript{−1}. The sidechain of H216 showed a big rotamer change, enabling \(\pi\)–\(\pi\) stacking with A\textsubscript{+1} (Fig. 5b). These \(\pi\)–\(\pi\) interactions set both 5\textsuperscript{′} and 3\textsuperscript{′}-ends of the 5\textsuperscript{′}-TCCCA target sequence to the rim of the DNA-binding groove formed by loop1 of CTD2*. This rim is clearly visible by sidechains of W211, R213, and H216 in the surface representation of CTD2* (Fig. 5c). This loop1 rim interacts with the phosphate backbone of ssDNA, whereas loop7 faces nucleobases as depicted by the sidechains of Y315 and D316 (Fig. 5c). Residues in loop7, including Y315, D316, and D317, also alter main chain atom positions and sidechain rotamers upon ssDNA-binding (Fig. 5b). D317 significantly changed its mainchain position as its \(\alpha\) Ca atom moved 3.0 Å from the position found in the apo-form CTD2-K3A structure (Fig. 5b). These rearrangements of backbone atoms of the loop7 residues are particularly important for recognition of C\textsubscript{−1}, as mainchain amino protons form two hydrogen bonds with the Watson–Crick face of C\textsubscript{−1} (Fig. 3d). This dynamic property of CTD2* is a remarkable difference compared to A3A, which did not show significant changes in mainchain atom positions upon ssDNA-binding. For A3A, only the sidechains of R28 and H29 (R215 and H216 in CTD2, respectively) in loop1, and Y132 (Y315 in CTD2) in loop7 changes their rotamers\textsuperscript{29,39}.

Most recently, Ziegler et al. published a structure of A3G-CTD bound to an adenine nucleotide\textsuperscript{39}. In the A3G-CTD–adenine complex structure, the adenine nucleotide binds in the space that is similar to the T\textsubscript{−1} position found in the A3A–ssDNA complexes\textsuperscript{29,39}. Since C\textsubscript{−1} in the CTD2*–ssDNA complex does not occupy the T\textsubscript{−1} position (Fig. 4c), the protein–DNA interaction found in the A3G-CTD–adenine complex\textsuperscript{39} is different from the enzyme–substrate interaction revealed in our study. Ziegler et al. suggested that the A3G-CTD–adenine structure shows a non-specific interaction by which A3G-CTD scans ssDNA sequence\textsuperscript{39}. Interestingly, W211 rearranged its position to interact with the bound adenine\textsuperscript{39}, which may imply that W211 is important in the interaction with non-specific DNA as well as the target sequence.

As mentioned in the previous section, P247K is the only residue in loop3, which interacts with ssDNA, and the interaction likely contributes in changing the loop3 position (Supplementary Fig. 4a). In addition, other loop3 residues including E254 and R256 form hydrogen bonds with Q293 (helix3) and E323 (helix4) of a neighboring asymmetric unit, respectively, which support the position of loop3 and concurrently aids crystal formation. Noteworthy, the substitutions introduced to CTD2 did not change the structure of A3G-CTD as shown in the superimposed structures of CTD2* (this study) and wild-type A3G-CTD (PDB ID: 4ROV\textsuperscript{35}) (Supplementary Fig. 4b). CTD2* and 4ROV structures are well superimposed as indicated by the pairwise rms deviation, which is 0.9 Å.

We have previously found that A3G-CTD increases its catalytic activity at lower pH, and H216 plays a key role in this pH dependence\textsuperscript{41}. Alanine mutation of this residue abolished catalytic activity, whereas arginine mutation kept catalytic activity but lost the pH dependency, suggesting that positive charge and/or formation of hydrogen bonds of this residue are important for substrate binding\textsuperscript{41}. The CTD2*–ssDNA co-crystal structure indicates that protonation of the imidazole ring enables the formation of two hydrogen bonds with the 5’-phosphate groups of C\textsubscript{0} and C\textsubscript{−1}. H216 is conserved only in A3G and A3A among human APOBEC3 proteins, and similar interactions between the histidine and nucleotides at –1 (T\textsubscript{−1}) and 0 (C\textsubscript{0}) positions have been observed in co-crystal structures of A3A and ssDNA\textsuperscript{29,39}. These conserved interactions involving histidines provide an explanation for similar pH dependency of catalytic speeds of A3A and A3G\textsuperscript{31,50}.

Until this study, we have been puzzled how W211 contributes to catalysis as this residue is spatially far from the catalytic Zn\textsuperscript{2+} ion yet alanine substitution of W211 results in nearly complete loss of catalytic activity\textsuperscript{23}. Moreover, we observe large chemical shift changes for the indole amino proton of W211 upon mixing with ssDNAs\textsuperscript{23,51}. Interestingly, while W211 is not conserved in other members of human APOBEC3 proteins, tryptophan is conserved in AIM. AIM is a member of the human APOBEC
family, that plays important roles in antibody diversification and triggers both class switch recombination and somatic hypermutation.\(^{52-54}\) Qiao et al. recently published structures of the human AID protein co-crystallized with dCMP, cadocyclic acid or cytidine (PDB ID# 5W0U, 5W0R, 5W1C, respectively), and proposed a “substrate channel” composed by loop1 and loop7 of AID.\(^{40}\) AID appears to recognize long target sequences, similar to A3G recognition of a five nucleotide target sequence, and AID has nucleotide preferences in \(-2, -1, 0,\) and \(+1\) positions.\(^{55,56}\) Based on the CTD2*-ssDNA complex, AID likely uses the tryptophan residue corresponding to W211 for \(\pi-\pi\) stacking with the nucleotide at \(-3\) position, and supports specific recognition of the nucleotides at \(-2\) and \(-1\) positions in a manner similar to CTD2* use of W211. In addition, conservation of R213 of CTD2* as a lysine in AID suggests the similar use of the lysine for interaction with 5'-phosphate group of the nucleotide at \(-2\) position. Further experimental data and structures are necessary to elucidate the ssDNA interactions of AID.

This structure of the CTD2*-ssDNA complex reveals the mechanism at atomic-level resolution by which the catalytic domain of A3G uniquely binds ssDNA. Fundamental knowledge of this complex can guide the design of molecular-based therapeutics for AIDS by modulating A3G catalytic function. In addition, recently, Komor et al. showed that an APOBEC deaminase tethered to catalytically dead Cas9 can mutate DNA in a programmable manner, offering a new strategy for "gene editing".\(^{57}\) The CTD2 variant may be a better tool for "base editing" because it is more soluble, binds ssDNA stronger and catalyzes deamination faster than wild-type A3G-CTD.

Methods

Protein Expression and Purification. The CTD2 variant of human A3G CTD (residue 191–384) and its inactive variant CTD2* were expressed from pGEX6P-1 expression plasmid [P247K was introduced by synthesizing cDNA (GenScript), and the primer sequences used for E259A and Q318K substitutions are listed in Supplementary Table1 with Glutathione S-transferase (GST) tag (crystallography, EMSA, and real-time NMR deamination assay) for GST purification or from pET-28a plasmid with poly-Histidine tag (MST assay) 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 0.3–0.6 at 600 nm. Then, temperature was reduced to 17 °C and protein expression was induced for 18 h with 0.6 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\(_2\), 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\(_2\), 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 48,384g 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 2h. 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\(_2\), 2 mM DTT, and 0.002% Tween-20) and incubated with Pre-Scission

**Fig. 5** Comparison of structures of CTD2* with apo-CTD-2K3A. Proteins are in the same orientation in all three figures. \(\text{a}\) Superimposed yellow and gray cartoons show structural features of CTD2* (this study) and CTD-2K3A (PDB ID# 3JR2), respectively. W211, R213, H216, Y315, D316, and D317 sidechains of CTD2* are shown in sticks. Zn\(^{2+}\) ion is shown as a purple sphere. ssDNA is not shown. \(\text{b}\) An enlarged view of (\(\text{a}\)) that shows the repositioning of the critical residues W211, R213, and H216 of loop1, and Y315, D316, and D317 of loop7. Double headed arrows point to positions of C-D317. \(\text{c}\) Surface representation of ssDNA-bound CTD2* (this study). Locations of the loop1 and loop7 residues are labeled except D317 because D317 is not seen on the surface. The 5’-TCCCA target sequence is shown as sticks, and C, N, O, and P atoms are colored blue, dark blue, red, and orange, respectively.
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 ZnCl₂, and 0.002% Tween-20.

For Ni-NTA purification, protein-bound resin was washed with 50 mM sodium phosphate, pH 7.3, 1.1 M NaCl, 25 µM ZnCl₂, 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 µM ZnCl₂. For both GST and Ni-NTA purification, protein purity was analyzed by SDS-PAGE.

Crystal Growth and Collection. Samples used for crystallization contained about 9.5 mg/ml (415 µM) CTD2* and a 50% molar excess of ssDNA in 20 mM sodium phosphate, pH 7.3, 100 mM NaCl, 1 mM DTT, 0.01 mM ZnCl₂. For both GST and Ni-NTA purification, protein was concentrated to about 9.5 mg/ml (415 µM) CTD2* at 10 nM and 15 in 50 µl by mixing the 6-FAM-labeled ssDNA (5'-AACAAAA-AATCCCAA-AATCCCAA-AATCCCAAA-AATCCCAAA-AATCCCAAA, was obtained from Integrated DNA Technologies (IDT; Coralville, IA). Initial crystallization condition was identified using JScreen Nuc-Pro from MtGen. Crystals were grown at 4°C, by sitting drop vapor diffusion method over a 65 µl reservoir of 20% W/V PEG 6000, 50 mM di-sodium L-malate; pH 5.0 and 30 mM CaCl₂ in a sitting drop 2 well crystallization plates from Molecular Dimensions. Drops were set up by mixing 0.3 µl of CTD2*-ssDNA complex and 0.3 µl of reservoir solution using a robot. Mosquito Crystal from TTP Labtech. Crystals appeared after 1 week. Crystals grown at 4°C were melted at room temperature, and exactly similar crystal setup at 20°C did not produce any crystal.

Crystals were cryoprotected using reservoir solution containing 20% v/v glycerol and flash frozen in liquid nitrogen. X-ray diffraction data were collected at the Southeast Regional Collaborative Access Team (SER-CAT) 22-ID beamline at the Advanced Photon Source, Argonne National Laboratory. The crystals belong to the space group P2₁. The collected intensities were indexed, integrated, and scaled using HKL2000.

Structure Determination and Analysis. The structure was solved at 1.86 Å resolution by molecular replacement using the program Phaser and a previously determined structure of A3G-2K3A (PDB ID code 3R22, chain B was removed) as search model. Model building of the protein and bound DNA and RNA fragments were manually performed using the programs Coot and Phenix, respectively. The first 3 residues (Glu-Ile-Leu) and the last residue (Asn) were not modeled due to the lack of electron density. Due to the presence of extra positive density, Ser-368 and Ser-372 were modeled in two alternative conformations. The final model was refined to Rwork/Rfree values of 0.18/0.21 and was validated with the PDB validation tool and Molprobity. Statistics of Ramachandran analysis yielded 97.87% of the residues in the favored regions and 2.13% were found in the allowed regions. None of the residues were found in disallowed regions. Pairwise rms deviation between CTD2-CA3A and ssDNA-bound CTD2* and between ssDNA-bound A3A and ssDNA-bound CTD2* were calculated using Dool. Figures of structure models were generated by PyMOL.

Real-time NMR Deamination Assay. We determined initial rates of deamination reaction by using 1H NMR spectra. A 9nt ssDNA substrate (IDT), 5'-AATCCCAAA-AATCCCAA-AATCCCAA-AA-3', was used in the reaction. NMR spectra were acquired at 25°C on Bruker NMR spectrometers operating at 1H Larmor frequencies of 600 and 800 MHz. NMR samples contained 5% deuterium oxide with 200 nM protein, 200 µM ssDNA substrate, 100 mM NaCl, 0.002% Tween 20, 1 mM DTT, 10 µM ZnCl₂, and also included 50 mM sodium phosphate adjusted to pH 7.5. The concentration of deaminated product (5'-AATCCCAAA-UAA-3') was determined from integration of the H5 uracil proton peaks at 5.60 ppm. A series of 1H spectra were measured and the product concentrations as a function of the reaction times were used to determine the initial rate via linear regression.

EMSA. For EMSA, we used a 9nt ssDNA with 6-FAM at the 3' end and 5'-AATCCCAAA-FAM (IDT), and binding buffer containing 20 mM Bis-Tris pH 6.5, 100 mM NaCl, 10 µM ZnCl₂, and 0.002% Tween 20. To test the specificity of the DNA binding, we performed competition assays with fluorescent-unlabeled ssDNAs including 5'-AAAAA-5' and 5'-AATCCCAA-AATCCCAA-AATCCCAA-AATCCCAA, as a nonspecific and a specific ssDNA, respectively. Binding reactions were performed in 30 µl by mixing the 6-FAM-labeled ssDNA (5'-AATCCCAAA-AATCCCAA-AATCCCAA-AATCCCAA, was obtained from Integrated DNA Technologies (IDT; Coralville, IA). Initial crystallization condition was identified using JScreen Nuc-Pro from MtGen. Crystals were grown at 4°C, by sitting drop vapor diffusion method over a 65 µl reservoir of 20% W/V PEG 6000, 50 mM di-sodium L-malate; pH 5.0 and 30 mM CaCl₂ in a sitting drop 2 well crystallization plates from Molecular Dimensions. Drops were set up by mixing 0.3 µl of CTD2*-ssDNA complex and 0.3 µl of reservoir solution using a robot. Mosquito Crystal from TTP Labtech. Crystals appeared after 1 week. Crystals grown at 4°C were melted at room temperature, and exactly similar crystal setup at 20°C did not produce any crystal.

Real-time NMR Deamination Assay. We determined initial rates of deamination reaction by using 1H NMR spectra. A 9nt ssDNA substrate (IDT), 5'-AATCCCAAA-AATCCCAA-AATCCCAA-3', was used in the reaction. NMR spectra were acquired at 25°C on Bruker NMR spectrometers operating at 1H Larmor frequencies of 600 and 800 MHz. NMR samples contained 5% deuterium oxide with 200 nM protein, 200 µM ssDNA substrate, 100 mM NaCl, 0.002% Tween 20, 1 mM DTT, 10 µM ZnCl₂, and also included 50 mM sodium phosphate adjusted to pH 7.5. The concentration of deaminated product (5'-AATCCCAAA-UAA-3') was determined from integration of the H5 uracil proton peaks at 5.60 ppm. A series of 1H spectra were measured and the product concentrations as a function of the reaction times were used to determine the initial rate via linear regression.

Cell lysates were made using CeLyte M (Sigma) solution containing Protease Inhibitor Cocktail (Roche), followed by a 10 min, 10,000×g spin to remove cellular debris. The cell lysates were mixed with NuPAGE LDS sample buffer (Invitrogen) containing β-mercaptoethanol and heated for 5 min at 95°C. Samples were analyzed on 4–20% Tris-Glycine Gels (Invitrogen) using standard western blotting protocols. Proteins were detected with primary antibodies as follows: FLAG-A3G or FLAG-NTD-CTD2 (rabbit anti-FLAG polyclonal antibody, 1:5000 dilution, Sigma #F7425); Vif-HA (mouse anti-HA monoclonal antibody, 1:5000 dilution, Sigma #H8663); α-tubulin (mouse anti-α-tubulin antibody, 1:10,000 dilution, Sigma #T9026); Antibody against HIV-1 p24 (monoclonal, 1:10,000 dilution) was obtained through the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH: HIV-1 p24 Gag Monoclonal (24a-3) from Dr. Michael Malim (catalog #6458). An IRDye 800CW-labeled goat anti-rabbit secondary antibody (LI-COR #926-32211) was used at a 1:10,000 dilution to detect rabbit polyclonal antibodies and an IRDye 680CW-labeled goat anti-mouse antibody (LI-COR #926-68070) was used at a 1:1,000 dilution to detect mouse primary antibodies. Protein bands were visualized and quantified using an Odyssey Infrared Imaging System (LI-COR). Uncropped western blots are shown in Supplementary Figs. 6, 7, and 8.

Data Availability. Atomic coordinates and structural factors for the reported crystal structure have been deposited in the Protein Data Bank under the accession code 7WQ5.
References

1. Conticello, 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).

2. LaRue, R. S. et al. The artiodactyl APOBEC innate immune repertoire shows evidence for a multi-functional domain organization that existed in the ancestor of placental mammals. BMC Mol. Biol. 9, 104 (2008).

3. Zhang, J. & Webb, D. M. Rapid evolution of primate antiviral enzyme APOBEC3G. Hum. Mol. Genet. 13, 1785–1791 (2004).

4. Jarmuz, A. et al. An anthropoid-specific loss of orphan C to U RNA-editing enzymes on chromosome 22. Genomics 79, 285–296 (2002).

5. Sherry, 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).

6. Mangeat, B. et al. Broad antiretroviral defence by human APOBEC3 through lethal editing of nascent reverse transcripts. Nature 424, 99–103 (2003).

7. Zhang, H. et al. The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. Nature 424, 94–98 (2003).

8. Liddament, M. T., Brown, W. L., Schumacher, A. J. & Harris, R. S. DNA deamination mediates innate immunity to retroviral infection. Cell 113, 803–809 (2003).

9. Harris, R. S. et al. DNA deamination mediates innate immunity to retroviral infection. Cell 113, 803–809 (2003).

10. Malim, M. H. APOBEC proteins and intrinsic resistance to HIV-1 infection. Philos. Trans. R. Soc. Lond. B Biol. Sci. 364, 675–687 (2009).

11. Chiu, Y. L. & Greene, W. C. The APOBEC cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements. Annu. Rev. Immunol. 26, 317–353 (2008).

12. Goila-Gaury, R., Delabays, R. & Schellekens, H. HIV-1 Vif, APOBEC, and innate immunity. Retrovirology 5, 51 (2008).

13. 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).

14. Yu, Q. et al. Single-strand specificity of APOBEC3G accounts for minus-strand deamination of the HIV genome. Nat. Struct. Mol. Biol. 11, 435–442 (2004).

15. Okada, A. & Iwatsuki, Y. APOBEC3G-mediated G-to-A hypermutation of the HIV-1 genome: the missing link in antiviral molecular mechanisms. Front. Microbiol. 7, 2027 (2016).

16. Yu, X. et al. Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cell-SCF adaptor complex. Science 302, 1056–1060 (2003).

17. Aydin, H., Taylor, M. W. & Lee, J. E. Structure-guided analysis of the human APOBEC3-HIV recombinase. Structure 22, 668–684 (2014).

18. Betts, L., Xiang, S., Short, S. A., Wolfenden, R. & Carter, C. W. Jr Cytidine deaminase. The 2.3 A crystal structure of an enzyme: transition-state analog reveals potential oligomerization interfaces. J. Mol. Biol. 235, 635–656 (1994).

19. 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).

20. 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).

21. Desimine, B. A. et al. Multiple APOBEC3 restriction factors for HIV-1 and one Vif to rule them all. J. Mol. Biol. 426, 1220–1245 (2014).

22. Chelico, L., Pham, P., Calabrèse, P. & Goodman, M. F. APOBEC3G DNA deaminase acts processively 5’→3’ on single-stranded DNA. Nat. Struct. Mol. Biol. 13, 392–399 (2006).

23. Chen, K. M. et al. Structure of the DNA deaminase domain of the HIV-1 restriction factor APOBEC3G. Nature 452, 116–119 (2008).

24. Harjes, E. et al. An extended structure of the APOBEC3G catalytic domain suggests a unique holoenzyme model. J. Mol. Biol. 389, 819–832 (2009).

25. Shandilya, S. M. et al. Crystal structure of the APOBEC3G catalytic domain reveals potential oligomerization interfaces. Structure 18, 28–38 (2010).

26. Bohn, M. F. et al. Crystal structure of the DNA cytidine deaminase APOBEC3F: the catalytically active and HIV-1 Vif-binding domain. Structure 21, 1042–1050 (2013).

27. Bohn, M. F. et al. The sDNA mutator APOBEC3A is regulated by cooperative dimerization. Structure 23, 903–911 (2015).

28. Kounou, T. et al. Structure of the Vif-binding domain of the antiviral enzyme APOBEC3G. Nat. Struct. Mol. Biol. 22, 485–491 (2015).

29. Kounou, 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).

30. Benson, J. D., Fass, D., Wolf, D. & Cosman, D. Programmed editing of a target base in genomic DNA without double-stranded DNA cleavage. Nat. Biotechnol. 53, 420–424 (2016).
Acknowledgments

The authors would like to acknowledge Drs. Alexander Wlodawer, Kylie Walters, George Lountos, Baohua Cao, Nese Kurt Yilmaz, and Ms. Marzena A. Dyba for their advice and suggestions. This work was supported by in part with grant from the U.S. National Institutes of Health R01GM118474. For A.M., T.K., and H.M., this project has been secured funding. 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.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-04872-8.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

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

© The Author(s) 2018

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) 2018

Figure 4: Structure of human APOBEC3F.

Figure 5: Screening assay images.

Figure 6: Proof of principle for HIV-1 restriction.

Figure 7: Kinetic analysis of HIV-1 restriction.

Figure 8: Summary of wild-type and mutants.

Figure 9: Structure of human APOBEC3G.

Figure 10: Structure of human APOBEC3C.

Figure 11: Structure of human APOBEC3B.

Figure 12: Structure of human APOBEC3A.

Figure 13: Structure of human APOBEC3D.

Figure 14: Structure of human APOBEC3E.

Figure 15: Structure of human APOBEC3H.

Figure 16: Structure of human APOBEC3I.

Figure 17: Structure of human APOBEC3J.

Figure 18: Structure of human APOBEC3K.

Figure 19: Structure of human APOBEC3L.

Figure 20: Structure of human APOBEC3M.

Figure 21: Structure of human APOBEC3N.

Figure 22: Structure of human APOBEC3O.

Figure 23: Structure of human APOBEC3P.

Figure 24: Structure of human APOBEC3Q.

Figure 25: Structure of human APOBEC3R.

Figure 26: Structure of human APOBEC3S.

Figure 27: Structure of human APOBEC3T.

Figure 28: Structure of human APOBEC3U.

Figure 29: Structure of human APOBEC3V.

Figure 30: Structure of human APOBEC3W.

Figure 31: Structure of human APOBEC3X.

Figure 32: Structure of human APOBEC3Y.

Figure 33: Structure of human APOBEC3Z.