Structural basis for targeted DNA cytosine deamination and mutagenesis by APOBEC3A and APOBEC3B

Ke Shi1,9, Michael A Carpenter1,9, Surajit Banerjee6, Nadine M Shaban1,4, Kayo Kurahashi1,3, Daniel J Salamango1,4, Jennifer L McCann1,4, Gabriel J Starrett1,4, Justin V Duffy1,3, Özlem Demir7, Rommie E Amaro7, Daniel A Harki2,8, Reuben S Harris1,5 & Hideki Aihara1,3

APOBEC-catalyzed cytosine-to-uracil deamination of single-stranded DNA (ssDNA) has beneficial functions in immunity and detrimental effects in cancer. APOBEC enzymes have intrinsic dinucleotide specificities that impart hallmark mutation signatures. Although numerous structures have been solved, mechanisms for global ssDNA recognition and local target-sequence selection remain unclear. Here we report crystal structures of human APOBEC3A and a chimera of human APOBEC3B and APOBEC3A bound to ssDNA at 3.1-Å and 1.7-Å resolution, respectively. These structures reveal a U-shaped DNA conformation, with the specificity-conferring –1 thymine base tipped out and the target cytosine inserted deep into the zinc-coordinating active site pocket. The –1 thymine base fits into a groove between flexible loops and makes direct hydrogen bonds with the protein, accounting for the strong 5′-TC preference. These findings explain both conserved and unique properties among APOBEC family members, and they provide a basis for the rational design of inhibitors to impede the evolvability of viruses and tumors.

Vertebrates encode variable numbers of active polynucleotide cytosine deaminase enzymes collectively called APOBECs1-2. All vertebrate species have activation-induced deaminase (AID), which is essential for antibody-gene diversification through somatic hypermutation and class-switch recombination3,4. Most vertebrates also have APOBEC1, which edits cytosine nucleobases in RNA and ssDNA and functions in regulating the transcriptome and probably also in blocking the spread of endogenous and exogenous mobile elements, including viruses5,6. The APOBEC3 subfamily of enzymes is specific to mammals, is activated upon signal-induced expression, and is involved in regulating the transcriptome and probably also in blocking the spread of endogenous and exogenous mobile elements, including retrotransposons L1 and Alu and retroviruses such as HIV-1 (refs. 2,7,8).

APOBEC enzymes, APOBEC3A through APOBEC3H (A3A–A3H, human designation, though most cells express subsets as a result of differential gene regulation9-12). The local substrate preference of each of these enzymes is an intrinsic property that has helped scientists elucidate several biological and pathological functions, including the identification of AID as an antibody-gene DNA deaminase3,4, the delineation of the subset of APOBEC3 enzymes responsible for HIV-1 hypermutation (A3D, A3F, A3G, and A3H)2,7,8, and, recently, the implication of at least one APOBEC3A family member in mutagenesis in a wide variety of cancers13-15. The nucleobase immediately 5′ of the target cytosine (−1 relative to the target cytosine at position 0) is the most important determinant of each enzyme’s intrinsic substrate preference16-19. AID preferentially deaminates ssDNA cytosine bases preceded by an adenine or guanine (5′-RC), corresponding to cytosine mutation spectra in immunoglobulin gene variable and switch regions. A3G uniquely targets cytosine bases preceded by another cytosine (5′-CC), a pattern that is evident in patient-derived HIV-1 sequences. APOBEC1 and the remaining APOBEC3 enzymes elicit preferences for cytosine bases preceded by a thymine (5′-TC), a pattern that is evident in patient-derived HIV-1 sequences. APOBEC3A through APOBEC3H (A3A–A3H, excluding A3E), though most cells express subsets as a result of differential gene regulation9-12.

The local substrate preference of each of these enzymes is an intrinsic property that has helped scientists elucidate several biological and pathological functions, including the identification of AID as an antibody-gene DNA deaminase3,4, the delineation of the subset of APOBEC3 enzymes responsible for HIV-1 hypermutation (A3D, A3F, A3G, and A3H)2,7,8, and, recently, the implication of at least one APOBEC3A family member in mutagenesis in a wide variety of cancers13-15. The nucleobase immediately 5′ of the target cytosine (−1 relative to the target cytosine at position 0) is the most important determinant of each enzyme’s intrinsic substrate preference16-19. AID preferentially deaminates ssDNA cytosine bases preceded by an adenine or guanine (5′-RC), corresponding to cytosine mutation spectra in immunoglobulin gene variable and switch regions. A3G uniquely targets cytosine bases preceded by another cytosine (5′-CC), a pattern that is evident in patient-derived HIV-1 sequences. APOBEC1 and the remaining APOBEC3 enzymes elicit preferences for cytosine bases preceded by a thymine (5′-TC), a pattern that is evident in patient-derived HIV-1 sequences. APOBEC3A through APOBEC3H (A3A–A3H, excluding A3E), though most cells express subsets as a result of differential gene regulation9-12.

Received 11 October; accepted 16 November; published online 19 December 2016; doi:10.1038/nsmb.3344
APOBEC family members prefer substrates with 5′-T<sub>C</sub> motifs, and the potential for additional specificity-conferring enzyme-nucleobase contacts is poorly understood. Thus corroborating data sets including expression profiles, biochemical activities, global mutation analysis, data from animal experiments, and clinical information need to be combined in order for the subset of enzymes actually responsible for the observed APOBEC signature mutations in cancer to be deduced. The leading candidates to explain APOBEC mutagenesis in cancer are A3A<sup>21–23</sup>, A3B<sup>12,20,24</sup>, and A3H<sup>20</sup>. However, to date, only A3B has met all experimental criteria, including clinical correlations between high expression levels and poor outcomes for multiple cancer types<sup>25–29</sup>.

Despite the importance of APOBEC3-mediated mutations in a variety of biological and pathological processes, the molecular mechanisms underlying global ssDNA-binding activity and local target selectivity are poorly understood. In particular, prior structural studies have yielded numerous apo-enzyme structures (e.g., refs. 30–40) as well as two structures with nonsubstrate nucleotides<sup>37,41</sup>, but complexes with relevant ssDNA substrates have proven elusive. Here we report crystal structures of human A3A and a variant of the human A3B catalytic domain in complex with optimal and commonly mutated ssDNA substrates, respectively. Together with supporting biochemical data and comparisons with prior apo-enzyme structures, our findings provide robust molecular explanations for why these enzymes selectively bind ssDNA and elicit strong preferences for cytosine nucleobases in specific 5′-T<sub>C</sub> contexts in viral and cancer genomes.

**RESULTS**

**Optimal target for A3A-catalyzed C-to-U deamination in ssDNA**

A3A is a globular enzyme with a single zinc-coordinating active site, and it has proven to be the most potent human DNA cytosine deaminase<sup>32,43</sup>. To elucidate the mechanisms underlying the binding of ssDNA and the preference for 5′-T<sub>C</sub> dinucleotide targets, we sought to solve the crystal structure of human A3A bound to ssDNA. Although prior studies have examined the next-nucleotide preferences of this enzyme<sup>44,45</sup>, none has determined the extended substrate requirements in an unbiased manner. Therefore, first we determined the optimal substrate for wild-type A3A catalysis by performing deep deamination experiments with ssDNA containing a fixed target cytosine flanked on the 5′ and 3′ sides by four and three randomized bases, respectively (Fig. 1). Enzymatic reactions were allowed to proceed to ~10% completion (single-hit kinetics), and after conversion to double-stranded DNA with adaptors, deep-sequencing was used to determine the preferred bases flanking C-to-U deamination events (detected as C-to-T events). An analysis of >10,000 reads and 641 mutagenic events showed near-complete enrichment for T at the 5′ −1 position relative to the target C (position 0), a strong preference for a cytosine or purine nucleobase at the −2 position, and an unexpected preference for G at all 3′ positions (+1 to +3) (Fig. 1). This analysis informed the design of an optimal ssDNA substrate for cocrystallization studies, 5′-AAAAAAATCGGGAAA.

**A3A–ssDNA structure reveals a novel U-shaped binding conformation**

The optimal ssDNA substrate for A3A was cocrystallized with human A3A purified from *Escherichia coli*. This protein represents the wild-type enzyme apart from a four-residue C-terminal truncation to improve solubility and a single-amino-acid substitution of the catalytic glutamate (E72A) to prevent substrate turnover and bacterial genotoxicity (Supplementary Fig. 1). The 3.1-Å-resolution A3A–ssDNA structure has four monomeric complexes in the asymmetric unit, and each shows clear electron density for either five (5′-ATCGG) or six (5′-ATCGGG) nucleotides centered on the target cytosine (Table 1 and Supplementary Fig. 2). The superposition view shows near-identical conformations of all four proteins as well as the positioning of the −1 T and the target C (0) nucleotides and, as expected, some variation in the locations of bases outside the 5′-T<sub>C</sub> dinucleotide core (Supplementary Fig. 2).

The A3A-bound ssDNA adopts a U-shaped conformation anchored by the target C and the −1 T, with upstream and downstream ssDNA

**Figure 1** Deep-deamination approach for determining an optimal human A3A substrate. An ssDNA library with a single target C and N’s on the 5′ and 3′ sides was reacted with human A3A (near-single-hit kinetics). The resulting pool containing C-to-U deamination products was annealed to a barcoded Illumina adapter (IA), and T4 DNA polymerase was used to produce a complementary DNA strand. The resulting intermediate was denatured, annealed to a 5′ IA, and converted to duplex DNA by Phusion thermostable high-fidelity DNA polymerase. Illumina MiSeq was used to generate reads for subsequent informatics analysis. A WebLogo representation of deamination products unique to A3A (bottom) shows enrichment for −1 T and +1 G, which informed the ssDNA sequence for cocrystallization experiments (n = 641; error bars represent twice the sample correction value). Source data are available in the online version of the paper.
bent away from the active site (Fig. 2a). At the bottom of the ‘U’, the target C and the 5′ T nucleobases are flipped out toward the protein, with the sugar-phosphate backbone rotated with respect to those of the flanking nucleotides (Fig. 2a–d and Supplementary Fig. 2). The two flipped-out nucleotides fit between loops 1 and 7 and are stabilized by extensive van der Waals contacts with Trp98 at the base of the groove and by hydrogen bonds to backbone phosphates on the 5′ and 3′ sides of the target C, respectively, by the side chain of Tyr130 in loop 7 and Asn57 preceding loop 3 (Fig. 2b–d). Across the ssDNA-binding groove and opposite Tyr130, His29 from loop 1 fits inside the ‘U’ and donates hydrogen bonds to the backbone phosphates of both the target C and the 5′ T. The simultaneous hydrogen-bonding of His29 suggests that this side chain interacts with DNA optimally when doubly protonated, in agreement with the reported pH dependence of His29 when doubly protonated, in agreement with the reported pH dependence of A3A and A3G ssDNA-deamination activity 46,47. The His29 side chain also stacks with the +1 base and makes van der Waals contacts with the nucleotide at the −2 position, where the +1 and −2 bases may be close enough to interact (Fig. 2d). Thus, His29 seems to serve as a scaffold to stabilize ssDNA substrates in the U-shaped conformation. The +2 and +3 bases linearly stack on the +1 base, analogously to a B-form double-stranded DNA (Fig. 2b,c).

Supplementary Table 1

| Ligand/ion | Protein/DNA | No. atoms | Water | B factors (Å²) | Protein/DNA | Water | R.m.s. deviation Bond lengths (Å) | Bond angles (°) |
|------------|-------------|-----------|-------|----------------|-------------|-------|---------------------------------|----------------|
| 24/4 (GOL/Zn²⁺) | 6,597 | 16 | 73.54 | 24/19/2 | 24/19/2 | 45.97 | 0.002 | 0.45 |
| 24/4 (EG/Zn²⁺/I⁻/Cl⁻) | 1,712 | 98 | 73.58 | 24/19/2 | 24/19/2 | 45.24 | 0.015 | 1.20 |

Values in parentheses are for highest-resolution shell. Each structure is from one crystal.

Figure 2 Crystal structure of human A3A bound to ssDNA with preferred 5′-TCG deamination target motif. (a) Ribbon schematic of A3A–ssDNA complex showing flipped-out target C and −1 T nucleotides, as well as the overall U-shaped binding conformation. (b) Molecular surface of the A3A active site with the surrounding loops color-coded, and superposed stick models of ssDNA bound to four different molecules in the crystal’s asymmetric unit. (c) A view of a bound ssDNA molecule similar to that shown in b, with nucleobases numbered and key amino acid side chains from active site loops labeled. (d) A wall-eyed stereo view of the A3A active site and the bound ssDNA molecule shown in sticks. Hydrogen bonds are indicated by yellow dashed lines.

A3B–ssDNA structure and the mechanism of local target recognition

To improve crystallographic resolution and increase the relevance to cancer, we next sought to solve a crystal structure of the human A3B catalytic domain bound to ssDNA. A3B is a nuclear-localizing enzyme strongly implicated in cancer mutagenesis on the basis of, for example, associations with poor clinical outcomes for estrogen-receptor-positive breast cancer, multiple myeloma, and lung cancer25–29. A3B is double-domain enzyme with an N-terminal pseudocatalytic domain and a C-terminal catalytic domain27,28. The catalytic domain shares 92% amino acid sequence identity with A3A, and the majority of differences occur in solvent-exposed surfaces including loop regions (Supplementary Fig. 1). We obtained crystals using a catalytic mutant derivative (E255A) of our previously described A3B variant with loop 1 from A3A and near-wild-type activity (A3Bctd-QMloop3-A3Aloop1, hereinafter called A3Bctd*37) and a 7-mer ssDNA (5′-TTTT CAT) containing the most frequently mutated APOBEC motif in cancer (5′-TCA)13–15. The 1.7-Å-resolution structure of this A3Bctd*–ssDNA complex has a single nucleoprotein complex in the asymmetric unit and clear electron density for four nucleotides (5′-TTCA) (Fig. 3a, Table 1, and Supplementary Fig. 3). Despite the different enzyme–ssDNA combination, the overall DNA conformation is also U-shaped and is highly

Table 1 Data collection and refinement statistics

| Ligand/ion | Protein/DNA | No. atoms | Water | B factors (Å²) | Protein/DNA | Water | R.m.s. deviation Bond lengths (Å) | Bond angles (°) |
|------------|-------------|-----------|-------|----------------|-------------|-------|---------------------------------|----------------|
| 24/4 (GOL/Zn²⁺) | 6,597 | 16 | 73.54 | 24/19/2 | 24/19/2 | 45.97 | 0.002 | 0.45 |
| 24/4 (EG/Zn²⁺/I⁻/Cl⁻) | 1,712 | 98 | 73.58 | 24/19/2 | 24/19/2 | 45.24 | 0.015 | 1.20 |
Figure 3 Crystal structure of a variant of the human A3B catalytic domain bound to ssDNA with a 5′-TCA deamination target motif. (a) Ribbon schematic of the A3Bctd*–ssDNA complex showing the flipped-out target C (O) and −1 T, as well as the overall U-shaped binding conformation. (b) Superposition of the active site region of A3A (cyan) and A3Bctd* (magenta) with relevant ssDNA substrates in sticks, showing the near-identical positioning of the flipped-out target C and −1 T (representative DNA molecule from the A3A structure in light gray, and one from the A3Bctd* structure in yellow). (c,d) Composite omit 2Fo−Fc map contoured at 1.0σ, shown for the region surrounding the target C (c) or −1 T (d). (e) Deaminase activity of wild-type A3A on ssDNA substrates containing normal T or the indicated analogs at the −1 position, demonstrating that the 5-methyl group is unconstrained structurally. Uracil DNA glycosylase (UDG) readily excises dU and 5FdU from ssDNA and accounts for the 11-nt product in the absence of deamination. However, because of A3A activity on the target C and the 3′-end label, only the 10-nt product is apparent after deamination and gel fractionation. The uncropped gel image is provided in Supplementary Data Set 1; representative results from one of two independent experiments. (f) A wall-eyed stereo view of the A3Bctd* active site and the bound ssDNA molecule shown in sticks. Hydrogen bonds are indicated by yellow dashed lines. Water molecules are represented by small red crosses.

Interactions with nucleotides flanking the 5′-TC target motif

Outside of the central 5′-TC motif, A3A and A3Bctd* have limited direct interactions with nucleobases of the bound ssDNA, in agreement with the degeneracy of these positions in deep deamination experiments (Figs. 2 and 3). Arg28 in loop 1 of A3A interacts with an adenine at the −2 position, whereas the homologous Arg211 in A3Bctd* interacts with a thymine at the same position. This less specific base contact is consistent with A3A R28A and A3B R211A retaining robust ssDNA deaminase activity (data not shown and ref. 37) and the occurrence of multiple bases at the −2 position in the A3A deep deamination reaction (Fig. 1). In addition, Lys30 in A3A is positioned close to the major groove edge of the +1 guanine base, potentially contributing to the preference for a purine at this position. The corresponding residue Gln213 of A3B may similarly account for the reported +1 adenine preference12, although our A3Bctd*–ssDNA complex did not show this potential contact owing to the likely influence of crystal packing in the positioning of +1 adenine and the necessary
engineering of loop 1 residues to facilitate crystallization of the A3Bctd*–ssDNA complex.

To investigate whether interactions with any of the linearly stacked nucleobases at +1 to +3 positions are important in ssDNA engagement, we assayed A3A activity using normal versus 5-nitroindole-substituted ssDNA substrates. 5-nitroindole is a universal base analog that stacks like a canonical nucleobase but lacks hydrogen-bonding capabilities. A3A exhibits robust DNA cytosine deaminase activity on an ssDNA substrate containing 5-nitroindole substitutions at +1 to +3 positions, only about two-fold lower than that with an optimal ssDNA substrate with deoxyguanosines at the same positions (Supplementary Fig. 5). These data indicate that base-stacking or hydrophobic interactions between the nucleotides at +1 to +3 positions may be more relevant for the ssDNA deamination mechanism than nucleobase hydrogen-bonding with the enzyme.

The loop 3 region in both A3A and A3Bctd* complexes makes either direct or water-mediated hydrogen bonds with the backbone phosphate of the +1 nucleotide via the peptide main chain atoms (Figs. 2d and 3b,f). In addition, A3A loop 3 Lys60 points toward a midpoint between the backbone phosphate groups of +1 and +2 guanine nucleotides, suggesting a stabilizing interaction, although the ε amino group is not within direct hydrogen-bonding distance of either phosphate. The 5′ and 3′ nucleotides farther from the target cytosine, 5′ of −1 and 3′ of +1 positions, are out of the range of bonding potential with the enzyme, which is consistent with earlier biochemical footprinting studies48, HIV-1 hypermutation experiments (e.g., refs. 49–51), and cancer-mutation spectra analyses (e.g., refs. 24,52,53). However, ssDNA lengths greater than 3 nt seem to be required for full deaminase activity, indicating that nonspecific contacts may also be important54.

Comparisons of apo and ssDNA-bound structures
A comparison of the ssDNA-bound structures with previously reported A3A and A3Bctd crystal structures34,37 yielded additional mechanistic insights into how the activities of these enzymes are likely to be regulated (Fig. 4). The conserved active site and zinc-coordinating residues are virtually unchanged, consistent with the strong preferences of these enzymes (and other family members) for normal (unmodified) cytosine nucleobases in ssDNA, and indicating that the surrounding loop regions govern ssDNA-binding activity and local dinucleotide targeting. Indeed, comparisons of the native and ssDNA-bound structures revealed large side chain reorientations for His29 and Tyr132 of A3A, and loop 1 rearrangement and Tyr315 reorientation in A3B. As noted above, A3A His29 provides multiple contacts that enable the ssDNA to adopt the U-shaped conformation. The analogous histidine in the A3Bctd*–ssDNA structure has a similar scaffolding conformation, although it should be noted that this residue and most of loop 1 are derived from A3A (a necessary alteration for crystallization purposes). Therefore, further structural and biochemical studies will be needed to fully elucidate the mechanism of ssDNA engagement by wild-type A3B, in which one of the three arginines (Arg210, Arg211, or Arg212) in the wild-type A3B loop 1 may have a stabilizing function analogous to that of A3A His29.

Importantly, relative to apo-enzyme structures, the conserved tyrosine in loop 7, Tyr132 and Tyr315, is repositioned in ssDNA-bound A3A and A3Bctd* structures (bound and unbound comparisons of A3A and A3B are presented in Fig. 4a–c and Fig. 4d,e, respectively). This reorientation is important to confer dinucleotide target specificity through extensive van der Waals contacts with the −1 thymine (above). Indeed, loop 7 swap experiments have shown that the residues including A3A Tyr132 are critical for determining the preference of the −1 nucleobase17. For example, the 5′-CC dinucleotide specificity of A3G skewed toward 5′-TC after loop 7 was swapped in its entirety or after Asp317 was changed to a tyrosine in order to mimic the corresponding residue in A3A and A3B17. The Tyr315 reorientation is also likely to be an integral part of the overall mechanism that converts the closed active site conformation of A3Bctd37 to the

Figure 4 Comparison between apo-enzyme and ssDNA-bound A3A and A3B structures. (a,b) Molecular surface of A3A around the active site in the DNA-free (a; PDB 4XXO) and ssDNA-bound (b; this study) states showing reorientation of the side chains of His29 and Tyr132. (c) A superposition of the two conformations in a and b highlighting the repositioning of His29 and Tyr132 as well as shifting of loop 3 toward the bound ssDNA. (d,e) Molecular surface of A3Bctd in the DNA-free state (d; PDB 5COH) and A3Bctd* in the ssDNA-bound state (e; this study) around the active site, showing the large transition that is likely to occur between the closed (unbound) and open (ssDNA-bound) conformations.
To further validate these structural results, we compared wild-type A3A and structure-informed mutant derivatives in a series of biochemical experiments with extracts from 293T cells. Active site chemical experiments with extracts from 293T cells. Active site A3A and structure-informed mutant derivatives in a series of biochemical experiments with extracts from 293T cells. Active site

Biochemical analyses corroborate the ssDNA-bound A3A and A3Bctd* structures

We observed that the overall U-shaped ssDNA conformation and the positioning of the 5'-TC target dinucleotide are nearly identical in the two independent crystal structures (Fig. 3b). This superposition strongly implies that the observed ssDNA-binding and local targeting mechanisms are accurate reflections of the biological and pathological activities of these enzymes in virus and cancer mutagenesis. To further validate these structural results, we compared wild-type A3A and structure-informed mutant derivatives in a series of biochemical experiments with extracts from 293T cells. Active site alanine mutants were inactive in deaminating a 5'-TC-containing ssDNA substrate, as expected, including those with substitutions of the catalytic glutamate (E72A), the zinc-coordinating histidine and cysteines (H70A, C101A, and C106A), and the tryptophan lining a side of the active site pocket (W98A) (Fig. 5a and data not shown). The conserved cytosine-contacting residues, Ala71 (A71G, A71P), Ser99 (S99A, S99G, S99P), and Tyr130 (Y130A), also proved essential. The interaction between Thr31 and the 2'-deoxyribose of the target cytosine was only mildly compromised by an alanine substitution (T31A) and was fully disrupted by the introduction of a negative charge (T31D), consistent with data from previous studies indicating proximity of this residue to ssDNA58. All constructs were expressed similarly in immunoblots, indicating that the activity data are not due to poor expression or misfolding (Fig. 5a).

We conducted additional ssDNA deamination experiments to interrogate A3A contacts with the specificity-conferring −1 T nucleobase (Fig. 5b). We carried out these experiments essentially as described above, but we examined both activity and selectivity in parallel by systematically varying the −1 position of the ssDNA substrate. As a control, wild-type A3A has the most activity with ssDNA substrates with −1 T, intermediate activity with −1 C, and little activity with −1 A or G, whereas A3A-E72A has no activity. A comparison of Trp98 substitutions indicated that an aromatic residue is sufficient at this position in the structure to stabilize −1 T and the target C, as the W98A mutant was inactive but W98F and W98Y substitution mutants retained robust catalytic activity and near-wild-type dinucleotide preferences. The aromatic character of Tyr132 is similarly important in forming the hydrophobic pocket for −1 T, on the basis of near-wild-type activity for Y132F but not Y132A constructs. We observed parallel, albeit more severe, effects of phenylalanine and alanine substitutions for Tyr130. The overall greater importance of an aromatic side chain at position 130 was further indicated by contacts with the target C and the ssDNA backbone, as well as van der Waals interactions with Tyr132, which help to position the −1 T (Fig. 5c).

As alluded to above, Asp131 strongly influences the −1 preference, with a small nonpolar alanine substitution (D131A) loosening selectivity and showing near-equivalent activity with −1 T and −1 C substrates, a shorter hydroxylated residue (D131T) retaining selectivity

Figure 5 Corroborating biochemical data for human A3A. (a,b) DNA cytosine deamination by human A3A (WT, wild type) and the indicated mutant derivatives (S, substrate; P, product; TUB, tubulin). The corresponding anti-MYC (A3A) and anti-tubulin immunoblots indicate similar levels of A3A and soluble extract in each experiment relative to controls (A3A has two bands because of alternative translation initiation from Met1 or Met13). Reactions in a interrogated active site mutants using a 43-nt 5'-TC-containing ssDNA substrate, and reactions in b additionally interrogated the identity of the −1 position (A, C, G, or T) relative to the target cytosine. Representative results from one of two independent experiments; uncropped gel images are provided in Supplementary Data Set 1. (c) Graphs quantifying product accumulation in dose-response experiments for A3A and the indicated mutant derivatives (S, substrate; P, product; TUB, tubulin). Cytochrome c was used as a control, wild-type A3A has the most activity with ssDNA substrates with −1 T, intermediate activity with −1 C, and little activity with −1 A or G, whereas A3A-E72A has no activity. A comparison of Trp98 substitutions indicated that an aromatic residue is sufficient at this position in the structure to stabilize −1 T and the target C, as the W98A mutant was inactive but W98F and W98Y substitution mutants retained robust catalytic activity and near-wild-type dinucleotide preferences. The aromatic character of Tyr132 is similarly important in forming the hydrophobic pocket for −1 T, on the basis of near-wild-type activity for Y132F but not Y132A constructs. We observed parallel, albeit more severe, effects of phenylalanine and alanine substitutions for Tyr130. The overall greater importance of an aromatic side chain at position 130 was further indicated by contacts with the target C and the ssDNA backbone, as well as van der Waals interactions with Tyr132, which help to position the −1 T (Fig. 5c), as well as van der Waals interactions with Tyr132, which help to position the −1 T (Fig. 5c). As alluded to above, Asp131 strongly influences the −1 preference, with a small nonpolar alanine substitution (D131A) loosening selectivity and showing near-equivalent activity with −1 T and −1 C substrates, a shorter hydroxylated residue (D131T) retaining selectivity

Figure 6 Human A3A and Staphylococcus aureus TadA have similar U-shaped polynucleotide-binding conformations. (a) Ribbon schematics of A3A--ssDNA (this study) and TadA--IRNA (PDB 2B3J61) with the single zinc-coordinating active site regions positioned at similar angles for comparison. (b) Superposition of A3A--ssDNA and TadA--IRNA structures showing similar U-shaped binding conformations (predicted in a commentary60 on the original TadA--IRNA structure report61). A3A and TadA are colored in cyan and slate, respectively. The backbone and nucleobases of the bound ssDNA and RNA substrates in a are colored orange and yellow, respectively. In b, the A3A-bound ssDNA is shown in yellow, and TadA-bound RNA is shown with an orange backbone and slate nucleobases.
Figure 7 Structural comparison of the active sites of A3B and distantly related deaminase family members. (a–d) Active sites of (a) A3Bctd bound to ssDNA, (b) T4 bacteriophage 2′-deoxyxycytidylate deaminase bound to a dCMP analog (PDB 1VQ2 (ref. 63)), (c) mouse cytidine deaminase bound to cytidine (PDB 2F6R (ref. 64)), and (d) yeast cytosine deaminase bound to an analog of the free nucleobase cytosine (PDB 1P60 (ref. 65)). The catalytic glutamate (Glu255) was modeled into the A3Bctd*–ssDNA structure on the basis of its positioning in the apo-A3Bctd structure (PDB 5CQH), closely mimicking conformations of the corresponding residues in the T4, mouse, and yeast enzymes (Glu106, Glu67, and Glu64, respectively).

Gray spheres represent zinc ions. Smaller red spheres show the zinc-bound reactive water molecule. Sticks depict key residues contacting bound substrates, and ribbons represent protein backbones. A3Bctd and ssDNA in a are colored as in Figure 3f. Other deaminases are colored green, and substrates are in orange. The magenta and cyan loops in e are from adjacent subunits in the functional tetramer. A comparison of the structures shows similar zinc-coordination mechanisms and target cytosine positioning including conservation of surrounding aromatic residues, as well as substrate-specific interactions conferred by unique residues for each class of enzyme.

(e) Proposed deamination mechanism (adapted from ref. 66): (i) Glu255 deprotonates the Zn2+-coordinated H2O for nucleophilic attack at cytosine with the residual, protonated Glu255 H-bonding to N3 to withdraw electron density from C4 of cytosine, thereby accelerating nucleophilic attack by hydroxide. (ii) Deprotonation of the alcohol on the tetrahedral intermediate by Glu255 ensues, followed by (iii) collapse of the tetrahedral intermediate due to elimination of ammonia, to which protonated Glu255 contributes a hydrogen. (iv) The resulting uracil base is likely to be stabilized by Zn2+ coordination in the enzyme active site.

for −1 T (probably by mimicking the hydrogen-bond-acceptor role of the aspartate), and a longer and acidic glutamate substitution (D131E) converting the preference at the −1 position to C (presumably by creating an opportunity for direct hydrogen-bonding with the amino group of the cytosine ring and simultaneously disrupting the hydrogen-bonding between the carboxyl group of the shorter aspartate side chain and the N3 hydrogen of thymine). As in the experiment described above, the mutants were expressed similarly in immunoblots with few exceptions, indicating that the activity data are not due to poor expression or misfolding (Fig. 5b). We confirmed the D131 results in quantitative dose-response experiments with purified A3A and single amino acid substitution derivatives (Fig. 5c). Overall, these biochemical data strongly support the observed conformation of 5′-TC-containing ssDNA bound to A3A in the crystal structure.

DISCUSSION

The U-shaped ssDNA-binding conformation, as revealed by cocystal structures and validated through biochemical analyses, suggests that the loop regions of DNA stem-loop structures may be hotspots for APOBEC signature mutations within loop regions of predicted stem-loop structures. The U-shaped ssDNA binding conformation of A3A and A3Bctd* resembles the conformation of RNA bound to the distantly related tRNA adenosine deaminase TadA (Fig. 6). Additional structural studies will be needed to determine whether the common ssDNA-binding mechanism observed here for A3A and A3Bctd* represents that of wild-type A3B with natural loop 1 residues or that of other APOBEC3 family members, including A3G. Nonetheless, the similarity between the A3A, A3Bctd* and TadA–nucleic acid binding conformations suggests that hairpin or hairpin-like ssDNA or RNA structures may be preferred substrates for many different polynucleotide deaminase family members. However, the vast majority of viral and genomic ssDNA APOBEC mutations are not found in predicted secondary structures and, instead, correlate with properties of DNA replication (single-stranded cDNA in retroviral reverse transcription and lagging-strand DNA in tumor cells), suggesting that the most critical feature may be simply single-strandedness.

The mechanism of ssDNA binding by human A3A and A3Bctd* observed here contrasts with prior models, the conformation of a short oligo(dT) cocry stallized with A3Gntd41 (Supplementary Fig. 6), and the mechanism of double-stranded RNA binding and adenosine deamination by ADAR2 (ref. 59).
Residues within the A3A and A3B active site pockets are highly conserved within the APOBEC3 family, as evidenced by close superpositioning of the active site region of crystal structures of several human APOBEC3 enzymes (A3A, A3B, A3C, A3F, and A3G) (Supplementary Fig. 7). The only minor exception is Thr31 of A3A (Thr214 in A3B), which is a Ser216 in A3F. The T31A substitution of A3A is well tolerated, but T31D abolishes deaminase activity, consistent with a potential for phosphoregulation at this position (Fig. 5a and refs. 56,60). The corresponding residue in more distantly related metabolic cytosine deaminases, which are likely to catalyze the same hydrolytic chemistry of the cytosine deamination reaction as APOBEC3 enzymes, is either valine or isoleucine and makes van der Waals contact with the target cytosine (Fig. 7). These enzymes also show interesting variations of the residues surrounding the target cytosine base, reflecting substrate-specific interactions. For instance, the aromatic residue stacked over the target cytosine corresponding to Tyr130 of A3A (Tyr313 of A3B) is tyrosine in the 2′-deoxycytidine-5′-monophosphate deaminase of bacteriophage T4, whereas it is phenylalanine in the nucleoside cytidine deaminase and the free cytosine deaminase, consistent with the lack of a 5′-phosphate group.

In contrast to the strict conservation of the catalytic residues in the APOBEC3 active sites, comparisons of the amino acid sequences and conformations of loops 1, 3, and 7 suggest that adjacent contacts with ssDNA substrates are either conserved or diverged among APOBEC3 family members1,2 (Supplementary Fig. 7). The conservation accounts for the strong preference of the APOBEC3 enzymes for ssDNA, and the divergence provides flexibility to evolve varying catalytic efficiencies and local sequence preferences in order to achieve overlapping but distinct functions in innate immunity. For instance, it is likely that the closed active site conformation of A3Bctd in the unbound state37, in spite of the similar modes of ssDNA interaction between A3A and A3Bctd⁎, reflects a need for a tight regulation of this enzyme’s activity in the nucleus. Given the clear roles of the APOBEC3 enzymes in virus and tumor evolution, most recently demonstrated for A3B as a driver of drug resistance in breast cancer26, the A3A–ssDNA and A3Bctd⁎–ssDNA crystal structures are expected to provide a foundation for rational design of small-molecule inhibitors able to impede virus and tumor evolvability.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank D. Largaespada and D. Yee for insightful comments, R. Moorthy for oligonucleotide sample preparations, and J. Sivers (Pharmacology and Molecular Sciences Department, Johns Hopkins University, Baltimore, Maryland, USA) for providing the human UNG2 expression construct and purification protocol. This work was supported by grants from the US National Institutes of Health (NIGMS R01-GM118000 to R.S.H. and H.A., NIGMS R01-GM110129 to D.A.H., NCI R21-CA236309 to R.S.H., and DP2-OD007237 and NIGMS P41-GM103426 to R.E.A.), the NSF (CHE066071N to R.E.A.), the Prospect Creek Foundation (R.S.H. and D.A.H.), and the University of Minnesota Masonic Cancer Center (Spore-Program-Project-Planning Seed Grant to R.S.H.). This work is based upon research conducted at the Northeastern Collaborative Access Team beamlines, which are funded by the US National Institutes of Health (NIGMS P41-GM103403). The Pilatus 6M detector on the 24-ID-C beamline is funded by an NIH-ORIP HEI grant (S10 RR029205). This research used resources of the Advanced Photon Source, a US Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under contract no. DE-AC02-06CH11357. R.S.H. is supported as the Margaret Harvey Schering Land Grant Chair for Cancer Research and as an Investigator of the Howard Hughes Medical Institute.

AUTHOR CONTRIBUTIONS

R.S.H. and H.A. conceived and designed the studies. K.S., N.M.S., K.K., J.V.D., and H.A. purified proteins and established crystallization conditions. S.B. collected X-ray diffraction data. K.S. solved the crystal structures. M.A.C., D.I.S., J.L.M., and G.I.S. performed the deep-deamination studies. M.A.C. performed biochemical experiments. D.A.H. designed modified DNA substrates. O.D. and R.E.A. provided computational and structural insights. K.S., M.A.C., N.M.S., R.S.H., and H.A. drafted the manuscript, and all authors contributed to revisions and figure preparation.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Conticello, S.G. The AID/APOBEC family of nucleic acid mutators. Genome Biol. 9, 229 (2008).

2. Harris, R.S. & Dudley, J.P. APOBECs and virus restriction. Virology 479–480, 131–145 (2015).

3. Di Noia, J.M. & Neuberger, M.S. Molecular mechanisms of antibody somatic hypermutation. Annu. Rev. Biochem. 76, 1–22 (2007).

4. Robbins, D.F. & Nussenzweig, M.C. Chromosome translocation, B cell lymphoma, and activation-induced cytidine deaminase. Annu. Rev. Pathol. 8, 79–103 (2013).

5. Fossil, N. & Tam, P.P. Re-editing the paradigm of cytidine (C) to uridine (U) RNA editing. RNA Biol. 11, 1233–1237 (2014).

6. Koito, A. & Ikea, T. Intrinsic immunity against retrotransposons by APOBEC cytidine deaminases. Front. Microbiol. 4, 28 (2013).

7. Malim, M.H. & Bienz, P.D. HIV restriction factors and mechanisms of evasion. Cold Spring Harb. Perspect. Med. 2, a006940 (2012).

8. Simon, V., Bloch, N. & Landau, N.R. Intrinsic host restrictions to HIV-1 and mechanisms of viral escape. Nat. Immunol. 16, 546–553 (2015).

9. Roiland, E.W. et al. Quantitative profiling of the full APOBEC3 mRNA repertoire in lymphocytes and tissues: implications for HIV-1 restriction. Nucleic Acids Res. 38, 4274–4284 (2010).

10. Koning, F.A. et al. Defining APOBEC3 expression patterns in human tissues and hematopoietic cell subsets. J. Virol. 83, 9474–9485 (2009).

11. Stenglein, M.D., Burns, M.B., Li, M., Lengyel, J. & Harris, R.S. APOBEC3 proteins mediate the clearance of foreign DNA from human cells. Nat. Struct. Mol. Biol. 17, 222–229 (2010).

12. Burns, M.B. et al. APOBEC3B is an enigmatic source of mutation in cancer. Nature 494, 366–370 (2013).

13. Helelad, Y., Eshtad, S. & Nik-Zainal, S. Mechanisms underlying mutational signatures in human cancers. Nat. Rev. Genet. 15, 585–598 (2014).

14. Roberts, S.A. & Gordenin, D.A. Hypermutation in human cancer genomes: footprints and mechanisms. Nat. Rev. Cancer 14, 786–800 (2014).

15. Swartout, C., McGranahan, N., Starrett, G.J. & Harris, R.S. APOBEC enzymes: mutagenic fuel for cancer evolution and heterogeneity. Cancer Discov. 5, 704–712 (2015).

16. Carpenter, M.A., Rajagurbandara, E., Wijesinghe, P. & Bhagwat, A.S. Determinants of sequence-specificity within human AID and APOBEC3G. DNA Repair (Amst.) 9, 579–587 (2010).

17. Rathore, A. et al. The local dinucleotide preference of APOBEC3G can be altered from 5′-CC to 5′-TC by a single amino acid substitution. J. Mol. Biol. 425, 4442–4454 (2013).

18. Kohli, R.M. et al. Local sequence targeting in the AID/APOBEC family differentially impacts retroviral restriction and antibody diversification. J. Biol. Chem. 285, 40956–40964 (2010).

19. Wang, M., Rada, C. & Neuberger, M.S. Altering the spectrum of immunoglobulin V gene somatic hypermutation by modifying the active site of AID. J. Exp. Med. 207, 141–153 (2010).

20. 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).

21. Cavel, V., Supsêne, R., Shapira, M., Vartanian, J.P. & Wain-Hobson, S. A prevalent cancer susceptibility APOBEC3A hybrid allele bearing APOBEC3B 3′ UTR enhances chromosomal DNA damage. Nat. Commun. 5, 5129 (2014).

22. Chan, K. et al. An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers. Nat. Genet. 47, 1067–1072 (2015).

23. Nik-Zainal, S. et al. Association of a germline copy number polymorphism of APOBEC3A and APOBEC3B with burden of putative APOBEC-dependent mutations in breast cancer. Nat. Genet. 46, 487–491 (2014).

24. Burns, M.B., Temiz, N.A. & Harris, R.S. Evidence for APOBEC3B mutagenesis in multiple human cancers. Nat. Genet. 45, 977–983 (2013).

ARTICLES

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.
25. Siewerts, A.M. et al. Elevated APOBEC3B correlates with poor outcomes for estrogen-receptor-positive breast cancers. *Horm. Cancer* **5**, 405–413 (2014).

26. Law, E.K. et al. The DNA cytosine deaminase APOBEC3G promotes tamoxifen resistance in ER-positive breast cancer. *Sci. Adv.* **2**, e1601737 (2016).

27. Cescon, D.W., Halbe-Kains, B. & Mak, T.W. APOBEC3 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).

28. Yan, S. et al. Increased APOBEC3B predicts worse outcomes in lung cancer: a comprehensive retrospective study. *J. Cancer* **7**, 618–625 (2016).

29. Walker, B.A. et al. Mutational spectrum, copy number changes, and outcome: results of a sequencing study of patients with newly diagnosed myeloma. *J. Clin. Oncol.* **33**, 3911–3920 (2015).

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

31. Holden, L.G. et al. Crystal structure of the DNA cytosine deaminase APOBEC3G, ACS Chem. Biol. **7**, 506–517 (2012).

32. Kitamura, S. et al. The APOBEC3G crystal structure and the interface for HIV-1 Vif binding. *Nat. Struct. Mol. Biol.* **19**, 1005–1010 (2012).

33. Li, M. et al. First-in-class small molecule inhibitors of the single-strand DNA cytosine deaminase APOBEC3G. *ACS Chem. Biol.* **7**, 480–485 (2012).

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

35. Byeon, I.J. et al. NMR structure of human restriction factor APOBEC3A reveals substrate binding and enzyme specificity. *Nat. Commun.* **4**, 1890 (2013).

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

37. Shi, K., Carpenter, M.A., Kurahashi, K., Harris, R.S. & Aihara, H. Crystal structure of the DNA deaminase APOBEC3F catalytic domain. *J. Biol. Chem.* **290**, 28120–28130 (2015).

38. Nakashima, M. et al. Structural insights into HIV-1 Vif-APOBEC3F interaction. *J. Virol.* **90**, 1034–1047 (2015).

39. Shaban, N.M., Shi, K., Li, M., Aihara, H. & Harris, R.S. 1.92 Ångstrom zinc-free APOBEC3F catalytic domain crystal structure. *J. Mol. Biol.* **428**, 2307–2316 (2016).

40. Byeon, I.J. et al. Nuclear magnetic resonance structure of the APOBEC3F catalytic domain: structural basis for substrate binding and DNA deaminase activity. *Biochemistry* **55**, 2944–2959 (2016).

41. Xiao, X., Li, S.X., Yang, H. & Chen, X.S. Crystal structures of APOBEC3G N-domain alone and its complex with DNA. *Nat. Commun.* **7**, 12193 (2016).

42. Carpenter, M.A. et al. Methylcytosine and normal cytosine deamination by the foreign DNA restriction enzyme APOBEC3A. *J. Biol. Chem.* **287**, 34801–34808 (2012).

43. Cavai, V. et al. Molecular basis of the attenuated phenotype of human APOBEC3B DNA mutator enzyme. *Nucleic Acids Res.* **43**, 9340–9349 (2015).

44. Chen, H. et al. APOBEC3A is a potent inhibitor of adenovirus-associated virus and retrotransposons. *Curr. Biol.* **16**, 480–485 (2006).

45. Logue, C.E. et al. A DNA sequence recognition loop on APOBEC3A controls substrate specificity. *PLoS One* **9**, e97706 (2014).

46. Harjes, S. et al. Impact of H216 on the DNA binding and catalytic activities of the HIV restriction factor APOBEC3G. *J. Virol.* **87**, 7008–7014 (2013).

47. Pham, P., Landolph, A., Mendez, C., Li, N. & Goodman, M.F. A biochemical analysis linking APOBEC3A to disparate HIV-1 restriction and skin cancer. *J. Biol. Chem.* **288**, 29294–29304 (2013).

48. Rausch, J.W., Chelico, L., Goodman, M.F. & Le Grice, S.F. Dissecting APOBEC3G substrate specificity by nucleotide analog interference. *J. Biol. Chem.* **284**, 7047–7058 (2009).

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

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

51. 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).

52. Roberts, S.A. et al. An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat. Genet.* **45**, 970–976 (2013).

53. Alexandrov, L.B. et al. Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).

54. Miltra, M. et al. Structural determinants of human APOBEC3A enzymatic and nucleic acid binding properties. *Nucleic Acids Res.* **42**, 1095–1110 (2014).

55. Fu, Y. et al. DNA cytosine and methylcytosine deamination by APOBEC3B: enhancing methylcytosine deamination by engineering APOBEC3B. *Biochem. J.* **471**, 25–35 (2015).

56. Demorest, Z.L., Li, M. & Harris, R.S. Phosphorylation directly regulates the intrinsic DNA cytidine deaminase activity of activation-induced deaminase and APOBEC3G protein. *J. Biol. Chem.* **286**, 26568–26575 (2011).

57. Holtz, C.M., Sadler, H.A. & Mansky, L.M. APOBEC3G cytosine deamination hotspots are defined by both sequence context and single-stranded DNA secondary structure. *Nucleic Acids Res.* **41**, 6139–6148 (2013).

58. Nik-Zainal, S. et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47–54 (2016).

59. Matthews, M.M. et al. Structures of human ADAR2 bound to dsRNA reveal base-flipping mechanism and basis for site selectivity. *Nat. Struct. Mol. Biol.* **23**, 426–433 (2016).

60. Shirakawa, K. et al. Phosphorylation of APOBEC3G by protein kinase A regulates its interaction with HIV-1 Vif. *Nat. Struct. Mol. Biol.* **15**, 1184–1191 (2008).

61. Losey, H.C., Ruthenburg, A.J. & Verdine, G.L. Crystal structure of Staphylococcus aureus tRNA adenosine deaminase TadA in complex with RNA. *Nat. Struct. Mol. Biol.* **13**, 153–159 (2006).

62. Conticello, S.G., Langlois, M.A. & Neuberger, M.S. Insights into DNA deaminases. *Nat. Struct. Mol. Biol.* **14**, 7–9 (2007).

63. Almog, R., Maley, F., Maley, G.F., Maccoll, R. & Van Roey, P. Three-dimensional structure of the R115E mutant of T4-bacteriophage 2′-deoxycytidylate deaminase. *Biochemistry* **43**, 13715–13723 (2004).

64. Teh, A.H. et al. The 1.48 Å resolution crystal structure of the homotetrameric cytidine deaminase from mouse. *Biochemistry* **45**, 7825–7833 (2006).

65. Iretion, G.C., Black, M.E. & Stoddard, B.L. The 1.14 Å crystal structure of yeast cytidine deaminase: evolution of nucleotide salvage enzymes and implications for genetic chemotherapy. *Structure* **11**, 961–972 (2003).

66. Ko, T.P. et al. Crystal structure of yeast cytosine deaminase. Insights into enzyme mechanism and evolution. *J. Biol. Chem.* **278**, 19111–19117 (2003).
ONLINE METHODS

Protein purification. Human A3A with single-amino-acid substitution (E72A) was expressed in the E. coli strain BL21(DE3) as a GST-fusion protein with the pGEXSP-1 vector. Transformed bacteria were grown to mid-log phase in LB medium, then supplemented with 100 µM ZnCl₂ and induced by the addition of IPTG at a final concentration of 0.5 mM. After overnight incubation at 18 °C, bacteria were collected by centrifugation at 3,000g for 30 min at 4 °C, resuspended in lysis buffer (20 mM Tris–HCl, pH 8.0, 0.5 M NaCl, 10 mM β-mercaptoethanol), and lysed by the addition of lysozyme and sonication. After centrifugation at 60,000g for 1 h at 4 °C, the cleared lysate was passed through a 0.22-µm filter, and the protein was captured with glutathione agarose resin (Pierce), eluted in lysis buffer supplemented with 10 mM reduced glutathione, cleaved by the human rhinovirus 3C protease to remove GST, and purified over a Superdex 75 (GE Healthcare) size-exclusion column. The purified protein contained A3A residues 1–195 (near full length) and additional vector-derived residues (GPGPSPEF) on the N terminus. The A3B construct used in this study was the previously reported A3Bctd-QMAloop–A3Aloop1, which has a substitution of the A3A loop 1 residues (GIGRHK) for A3B loop 1 (DPLVRQQR) and single serine for the A3B loop 3 residues (spanning Ala242 to Tyr250). A3Bctd-QMAloop–A3Aloop1 with the additional E255A amino acid substitution (referred to as A3BctdE) was expressed with a noncleavable C-terminal His₉-tag (LEHHHHHHH) in E. coli strain C41(DE3)pLysS (Lucigen) from a pET24a-based vector and was purified as reported. The final size-exclusion chromatography running buffer consisted of 20 mM Tris-HCl, pH 7.4, 0.2 M NaCl, and 0.5 mM TCEP. The proteins were >95% pure on the basis of visual inspection of Coomassie blue–stained polyacrylamide gels. The purified enzymes were concentrated by ultrafiltration for use in the crystallization experiments. The initial A3A and A3B protein sequences matched UniProt IDs P31941 and Q9UH17, respectively (aligned in Supplementary Fig. 1).

Crystallization and structure determination. Purified A3A and A3Bctd were mixed with ~1.5 molar excess ssDNA (see below) at a final protein concentration of 25–30 mg/ml. The A3A-ssDNA complex crystal was obtained in a sitting drop formed by mixing of the complex solution with an equal volume of 0.2 M NaF, 20% (v/v) PEG3350, 0.1 M Bis-Tris propane, pH 6.5, equilibrated via vapor diffusion against the same reservoir solution. The A3Bctd–ssDNA crystal was obtained similarly with a reservoir solution consisting of 0.45 M NaCl and 20% (v/v) PEG3350. The crystals were mounted directly from an Intellit-Plate 96 (Art Robbins Instruments) and flash-cooled in liquid nitrogen with glycerol or ethylene glycol used as a cryoprotectant. X-ray diffraction data were collected at the Advanced Photon Source Northeastern Collaborative Access Team beamline 24-ID-C using the selenium K absorption edge wavelength, and the data were processed with XDPRO. The A3A–DNA complex crystal is in the space group of P2₁2₁2₁, and Matthews coefficient calculation indicated that there were likely to be four monomers of A3A in an asymmetric unit. With monomeric A3A (PDB 4XXO) used as a search model, the molecular replacement calculations by PHASER located four copies of A3A in the asymmetric unit. The resulting electron density map clearly showed the presence of ssDNA bound to each A3A molecule (Supplementary Fig. 2). The A3Bctd–DNA complex crystal is in the space group of P6₁2₂, with one monomer in the asymmetric unit and diffracting to ~1.7 Å resolution. Because the crystallization condition contained a high concentration of iodide ions, the diffraction data showed strong anomalous signal, and a total of six iodine or zinc sites were located with SHELXD. The resulting single-wavelength anomalous dispersion–phased electron density map, after density modification using SHELXE, showed A3B monomer and the presence of 16 M4ssDNA molecule. The A3Bctd monomer (PDB 5QCK) was placed into the electron density map by molecular replacement using MOLREP. Subsequent iterative refinement with the PHENIX suite and manual model inspection and rebuilding with COOT resulted in final Rwork/Rfree values of 20.97%/26.30% and 18.11%/21.21% for A3A–DNA and A3Bctd–DNA complex, respectively. Each A3A or A3Bctd monomer is bound to one molecule of ssDNA substrate. A summary of X-ray data collection and model refinement statistics is reported in Table 1.

Deep deamination experiments to determine an optimal A3A target site. We reasoned that deep sequencing of a target ssDNA oligonucleotide with a single cytosine flanked by degenerate Watson–Crick nucleobases could be used to determine an optimal A3A target site. First, an ssDNA substrate oligonucleotide 5'-NNNNNNNNN flanked by cytosine-free 22-nt and 21-nt regions was synthesized (Integrated DNA Technologies). This yielded a pool with 16,384 unique substrate sequences. Second, wild-type human A3A was purified from semi-confluent 293T cells transfected with a pCDNA4/TO-A3Ai-2×Strep3×Flag (SF) expression vector. Cells were harvested 48 h post-transfection and lysed in 50 mM Tris–HCl, pH 8.0, 1% (v/v) NP-40, 150 mM NaCl, 0.5% (w/v) deoxycholate, 0.1% (w/v) SDS, 5 mM EDTA, RNase A, 1× EDTA-free Protease Inhibitor Cocktail (Roche); they were then further disrupted by sonication. A3A-SF was purified with Strep-tactin resin (IBA). Samples were washed in high-salt buffer (20 mM Tris–HCl, pH 7.5, 1.5 mM MgCl₂, 1 M NaCl, 0.5 mM DTT, and 5% glycerol) followed by low-salt buffer (same as the high-salt buffer except with 150 mM NaCl) and final wash buffer (100 mM Tris–HCl, pH 7.5, 150 mM NaCl) and eluted using 2.5 mM desthiobiotin. Purified protein was fractionated using 4–20% SDS–PAGE and quantified by staining with Coomassie brilliant blue (Sigma).

We carried out titration experiments with recombinant A3A-SF to determine single-hit reaction conditions. That amount of enzyme was then incubated with 8 pmol of the substrate oligonucleotide pool for 1 h at 37 °C in 50 mM Tris–HCl, pH 7.5, 75 mM NaCl 2 pmol of the treated pool was annealed to appropriate 3′-barcoded adaptor using T4 DNA polymerase at 12 °C for 20 min. Then we added a universal 3′ adaptor to the duplex using Phusion DNA polymerase (New England BioLabs, Ipswich, MA). Products were purified with a GeneJET PCR purification kit (Thermo Fisher Scientific), analyzed by 20% native PAGE, and diluted to appropriate concentrations for deep sequencing.

We analyzed the reaction products using 2 × 50-nt paired-end reads (Illumina HiSeq 2500, University of Minnesota Genomics Center). Reads were paired using FLASH (http://ccb.jhu.edu/software/FLASH/). Data processing was done with a locally installed FASTX toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). FASTX trimmer was used to trim the 5′ and 3′ constant regions from sequences. Trimmed sequences were then filtered for high-quality reads with the FASTQ quality filter. Sequences with a Phred quality score less than 30 (99.9% base-calling accuracy) at any position were eliminated. Preprocessed sequences were then further analyzed with the FASTAppter toolkit (http://burkelab.missouri.edu/fastappter.html). FASTAppter-Count was used to count the number of times each sequence was sampled from the population. Each sequence was then ranked and sorted on the basis of overall abundance, normalized to the total number of reads in each population, and directed into FASTAppter-Enrich. FASTAppter-Enrich calculates the fold-enrichment ratios from a starting population (no enzyme control) to a selected population (incubation with A3A or other enzymes). After the enrichment file had been generated, mutated sequences specific to the A3A reaction in comparison with control reactions were analyzed with WebLogo design software. Logo error bars represent twice the sample correction value.
A3A expression constructs and DNA deamination activity assays. The pcDNA3.1-A3Ai-myc-His expression construct is reported elsewhere 11, 33, 42. Derivatives were constructed by site-directed mutagenesis and were verified by DNA sequencing. The activities of a subset of the A3A mutant constructs reported here have been described previously 35, 40, 43–45, 54, 55, 75, 76. Semiconfluent 293T cells in six-well plates were transfected with 1 µg of plasmid and harvested after 48 h to allow time for enzyme expression. We prepared soluble whole cell extracts by pelleting the cells and resuspending them in HED buffer (20 mM HEPES, pH 7.4, 5 mM EDTA, 100 µg/ml RNase A, 1 mM DTT, 10% glycerol, and Roche Complete protease inhibitors). Resuspended cell pellets were freeze-thawed and then rotated for 1 h at room temperature, after which they were subjected to water-bath sonication for 20 min. Cell debris was pelleted, and the clarified lysate was used for DNA deaminase activity assays. 5 µl of whole cell extracts with the desired A3A construct (or control) were mixed with 5 µl of HED buffer containing 1.6 µM fluorescently labeled ssDNA (sequences above). Reactions were then allowed to progress for 1 h at 37 °C, after which they were treated with 120 nM recombinant human UNG2 (uracil DNA glycosylase) for 10 min at 37 °C and then with 100 mM NaOH for 10 min at 95 °C (refs. 33, 42). Reaction products were separated by 15% denaturing PAGE and scanned on a Typhoon FLA 7000 imager (GE Healthcare). A3A-myc-His expression was verified by immunoblotting with primary rabbit anti-cMYC at 1:3,000 (Sigma, C3956) and secondary goat anti-rabbit IgG-Alexa Fluor 680 at 1:10,000 (Life Technologies, A21076). Tubulin expression was used as a loading control, and staining was done with primary mouse anti-α-tubulin at 1:80,000 (Sigma, B512) followed by secondary goat anti-mouse IR-Dye 800CW at 1:10,000 (LI-COR, 926-32210). Washed immunoblots were imaged with a LI-COR Odyssey imaging system.

Recombinant A3A-myc-His and mutant derivatives were expressed in 293T cells and purified as described by nickel-affinity chromatography 11, 33, 42. Activity assays were conducted as outlined above, except reactions were initiated by the addition of 5 µl of enzyme (two-fold dilutions starting at 200 nM) to 5 µl of 50 mM NaCl, 10 mM HEPES buffer, pH 7.4, containing 1.6 µM fluorescently labeled ssDNA (sequences above). Reactions progressed for 1 h at 37 °C and were processed and quantified as described above.

Data availability. The coordinates and structure factors for the A3A—ssDNA and A3Bctd*—ssDNA complexes have been deposited in the Protein Data Bank (PDB) under accession codes 5SWW and 5TD5, respectively. Source data for Figures 1 and 5c are available in the online version of the paper. Other data are available upon request.